UNIVERSIDAD POLITÉCNICA DE MADRID
ESCUELA TÉCNICA SUPERIOR DE INGENIEROS
INDUSTRIALES

**Digitalization Capacity for Knowledge Acquisition:
Learning from Health Monitoring**

TESIS DOCTORAL
DOCTORADO INTERUNIVERSITARIO EN ECONOMÍA Y
GESTIÓN DE LA INNOVACIÓN (DEGIN)

SHENGJING SUN, MSc (CS)

Supervisor:

Prof. JOAQUÍN ORDIERES MERÉ

2020

Tribunal nombrado por el Magfco. y Excmo. Sr. Rector de la Universidad Politécnica de Madrid, el día.....de.....de 2020.

Presidente:

Vocal:

Vocal:

Vocal:

Secretario:

Suplente:

Suplente:

Realizado el acto de defensa y lectura de la Tesis el día..... de.....de 2020 en la Escuela Técnica Superior de Ingenieros Industriales.

Calificación:

EI PRESIDENTE

LOS VOCALES

EL SECRETARIO

DOCTORATE IN ECONOMICS AND INNOVATION MANAGEMENT

The Interuniversity Master and PhD Programs in Economics and Innovation Management are implemented jointly by Universidad Autónoma de Madrid (UAM), Universidad Complutense de Madrid (UCM) and Universidad Politécnica de Madrid (UPM), enabling students to access the resources of the three universities, regardless of the center where they are enrolled. These programs offer teaching, professional competence, and initiation to advanced research in an area of rising influence in contact with leading European research teams. The main objectives of DEGIN are as follows:

- To train individuals who are able to address and solve problems of scientific and technological nature in the field of economics and innovation management through independent and original research work to expand the knowledge frontier further.
- To train researchers capable of disseminating research results through scientific publications with a recognized review process and impact index.
- To offer society a group of individuals who are able to elaborate, assess and critique new scientific and technological development research proposals in the field of economics and innovation management.
- To train researchers eligible for jobs requiring a PhD Degree as postdoctoral researcher, university professor, researcher at institutes or businesses and other jobs requiring creativity, leadership and management in the area of technology.
- To contribute to raise awareness of the scientific activity as a career and of its ethical implications for society.

Contents

Acknowledgment	I
Abstract	III
Resumen	V
1 Introduction	1
1.1 Research motivation: Human-centered society	3
1.2 People-Centric IoT: Health oriented	4
1.3 Formalization research objectives	9
2 Methodology	10
2.1 Research method	11
2.2 A guiding framework	11
2.2.1 Enabling technologies	15
2.2.2 Applied devices	20
2.2.3 Prototype: proof of concept	22
2.3 Study design	25
2.3.1 Study design 1	25
2.3.2 Study design 2	28
2.3.3 Study design 3	31
2.3.4 Study design 4	34
3 Collection of publications	36
3.1 Publication 1	38
3.2 Publication 2	57
3.3 Publication 3	83
3.4 Publication 4	103
3.5 Publication 5	116
3.6 Publication 6	133
3.7 Publication 7	155

4	Conclusions	179
4.1	Implications	185
4.2	Limitations	188
4.3	Future research	188
	Bibliography	202

Acknowledgment

I would like to express my sincere gratitude to my advisor Professor Joaquín Ordieres-Meré for all the support during my doctoral study and research. His kind and spirituality have been motivating me to act as a valuable person to the world.

I would like to thank my colleagues and collaborators Prof. Carlos A. Iglesias, Prof. Jack D. Spengler, Prof. José-Fernán Martínez-Ortega, Prof. Sokhn Maria, Dr. Gangao Zhu, Dr. Xiaochen Zheng, Dr. Bing Gong, Dr. Jesus Rodriguez, Dr. Yuanjiang Huang, Dr. Jie Yin, Dr. Julia Ayuso Sanchez, Dr. Zhan Liu, Mr. Geovanny Poveda, Miss Beatriz Bustos and Mr. Magalhaes Miguel. Working with you, has transformed me to be better person.

I would like to thank all my family members, especially my father Xicheng Sun and my husband Ning Li, for supporting me throughout my life. You have always been the most important motivation for me to continue and accomplish the doctoral study. Special gratitude goes to China Scholarship Council (CSC) for offering me the CSC scholarship, and a special thanks to Prof. Angel Alvarez who takes care of me like the father in Spain.

Last but not least, I would like to quote a poem to thanks to life itself.

Be thankful that you don't already have everything you desire, if you did, what would there be to look forward to?

Be thankful when you don't know something, for it gives you the opportunity to learn.

Be thankful for the difficult times, during those times you grow.

Be thankful for your limitations, because they give you opportunities for improvement.

Be thankful for each new challenge, because it will build your strength and character.

Be thankful for your mistakes, they will teach you valuable lessons.

Be thankful when you're tired and weary, because it means you've made a difference.

It's easy to be thankful for the good things. A life of rich fulfillment comes to those who are also thankful for the setbacks.

Gratitude can turn a negative into a positive. Find a way to be thankful for your troubles, and they can become your blessings.

Abstract

Digital transformation boosts the integration of intelligent data into all areas of society, from personal streams of life-span, an organization business workflows, to the whole ecosystem of different industries. The continuous connectivity and interaction in the digital world pave the way to learn knowledge from the variety of massive data. The Internet of Things (IoT) is a promising practice in the digitalization process. Its basic spirit is to thrust a paradigm that everything (machine and people) can be seamlessly connected into an IoT network by sensors.

Toward the next frontier society 5.0, it is aimed for a prosperous human-centered society where people can have a high quality of life. However, general IoT architectures and data value chain models are still device/platform-specific, which lacks necessary emphasis on people dimension. Health as the core aspect of people has a significant impact on their quality of life. The adverse health factors may cross the entire lifespan: from home to workplace, from commute to work or fitness, and even elderly people care homes. To center the research work, the research question was pinpointed: *How to accelerate and enhance people's health and well-being in the IoT data value chain?*

On the premise of research status-quo and the pinpointed research question, four specific research objectives were defined as follows. 1. *Enhance people dimension, especially health and well-being perspective in conventional IoT architectures.* 2. *Develop continuous long term monitoring solutions to support better health management.* 3. *Learn psychological and physical health impact from (e.g., workers) a group of people over the activities.* 4. *Accelerate health-related data sharing with security trust and privacy assurance.*

To fulfill the research objectives, the action research method was employed as a theoretical approach to analyze and implement the research; a guiding framework was designed to set up the research theme and context, the applied key concepts, enabling technologies, devices, and developed prototypes were introduced under the guiding framework; furthermore, several study designs were detailed presented including experimental setup, data collection, and relevant data analysis method.

Toward a human-centered society, the future of IoT is also seeking enhanced people-centered solutions. To accelerate and facilitate health-related data-driven knowledge acquisition and data value chain to society, organizations, and individuals, the research leverages advance technologies such as IoT, Smart Wearables, DLT, and machine learning techniques. To summarize, the

study focuses on the health dimension, the thesis generalizes highly the main contributions.

1. The study proposed IoT application architectures such as Healthy operator 4.0 architecture and proved their feasibility through real-world application cases in the industry. 2. Three long-term monitoring solutions were developed using low-cost IoT devices and successfully adopted in practical usage to continuously collect health-related parameters. 3. Different data mining approaches and machine learning methods were investigated and compared to learn health impacts over the activities from a group of people. The method chosen was proved to be capable of better understanding people's behavioral patterns and hidden rules, by the real-world empirical analysis conducted in both Spain and the USA. 4. A data sharing solution was designed in the study, that integrates DLT (IOTA Tangle) to IoT data management, by which data transparency and data ownership can be implemented under a secure, fee-less, and trust data sharing mechanism.

The value produced by the contributions is reflected on the individual level, organization level, and society level, which lies in societal aspects such as smart environment, industry 4.0, and smart city. With data-driven AI technology booming, big data analytic era comes. The future of work is now. The advance technology such as deep learning, Hadoop, Kubernetes, and Spark can be employed to dig knowledge out of data. The IoT big data analytic can achieve an improved understanding of data for individuals, organizations, and society, to make efficient and effective decisions.

RESUMEN

La transformación digital acelera la integración de la inteligencia de datos en todas las áreas de la sociedad, desde los flujos de datos personales relativos a las actividades diarias, pasando por los flujos de trabajo de procesos de negocio en las organizaciones, hasta alcanzar todo el ecosistema de relaciones inter-industrias. La conectividad universal continua y la interacción en el universo digital facilita la obtención de conocimiento desde una gran variedad de datos masivos. El internet de las cosas (IoT) es una práctica prometedora in la digitalización de procesos. Su espíritu primario es potenciar un paradigma en el que tanto las máquinas como la gente pueden ser conectados de un modo integrado en las redes IoT de sensores.

Manteniendo el foco en la siguiente frontera, "sociedad 5.0", se busca desarrollar una sociedad que potencie la dimensión humana, donde la gente podrá desarrollar una elevada calidad de vida. No obstante, tanto la arquitectura general del IoT como los flujos de valor de los datos obtenidos son aún específicos para los dispositivos o plataformas, lo que dificulta poner el énfasis en la dimensión humana general. La salud, como un aspecto central de interés para las personas tiene un impacto muy significativo en su calidad de vida. Factores negativos en relación a la salud pueden ser persistentes en el tiempo, tanto en ámbitos domésticos como laborales, y pueden ocurrir tanto en el transporte diario como en actividades deportivas o en el cuidado de personas de edad avanzada. Para centrar el trabajo de la tesis la pregunta de investigación ha sido resaltada: *¿Cómo acelerar y potenciar la dimensión salud y bienestar en la población a través de los flujos de valor correspondientes a esos datos?*

Sobre las premisas definidas pro el status-quo de investigación y por la pregunta de investigación cuatro objetivos específicos de investigación han sido definidos: *1.-Potenciar la dimensión humana, en particular salud y bienestar en las arquitecturas basadas en IoT. 2.- Desarrollar soluciones de monitorización de parámetros a largo plazo, que permitan una mejor gestión de los factores con influencia en la salud. 3.- Comprender el impacto en la salud tanto física como psicológica en grupos de población (por ejemplo trabajadores) durante el desempeño de su actividad. 4.- Acelerar el uso compartido de los datos relativos a la salud, cumpliendo criterios de confianza y privacidad suficientes.*

Para alcanzar los objetivos de investigación el método de investigación activa ha sido empleado como aproximación teórica para analizar e implementar la investigación. Se ha desarrollado un marco de referencia que guíe el proceso y permita establecer el tema de investigación y su

contexto, los conceptos clave, las tecnologías habilitantes, los dispositivos, así como desarrollar los prototipos de acuerdo al marco desarrollado. Adicionalmente, varios estudios de detalle han sido desarrollados, incluyendo su ciclo de vida de configuración, captura de datos y análisis de los mismos.

El futuro del IoT enfocado a una sociedad centrada en la dimensión humana busca también desarrollar soluciones centradas en la gente. Para acelerar y facilitar la generación y adquisición de conocimiento proveniente de los flujos de datos que pueda beneficiar a la sociedad (organizaciones e individuos), la investigación balancea tecnologías avanzadas, tales como IoT, dispositivos inteligentes, Libros Mayores Distribuidos y técnicas de aprendizaje automático. El estudio se centra en la dimensión de salud y de modo resumido sus principales contribuciones son: *1.- Proponer aplicaciones de arquitecturas IoT como Operario Saludable 4.0, demostrando su factibilidad a través de aplicaciones con casos de uso industriales. 2.- Desarrollar tres soluciones de monitorización de largo plazo usando dispositivos IoT de bajo coste con una implementación satisfactoria en las aplicaciones reales que permite la recogida de datos relativos a parámetros relacionados con salud. 3.- Investigar diferentes aproximaciones de minado de datos y métodos de aprendizaje automático, así como comparar sus resultados para comprender mejor las implicaciones sanitarias de comportamientos cuando se consideran agrupados. La selección del método elegido ha mostrado su utilidad en explicar los patrones de comportamiento de la gente y determinadas reglas no explícitas, al observar comportamientos de grupos de personas tanto en España como en E.E.U.U. 4.- Diseñar una solución de compartición de datos que integra la tecnología DLT (Tangle de IoTA) para facilitar el manejo de datos de IoT, de modo que transparencia y propiedad de los mismos estén aseguradas a través de mecanismos confiables y sin coste de transacción.*

El valor creado a través de las contribuciones de esta tesis se reflejan tanto a nivel individual como de sociedad en general, vinculado a gestión ambiental inteligente, industria 4.0 y ciudades inteligentes. Con las aplicaciones de inteligencia artificial aprovechando los flujos de datos en plena explosión, la era de creación de valor usando la analítica de datos masiva está en sus comienzos. Tecnologías avanzadas como el aprendizaje profundo, Hadoop, Kubernetes y Spark pueden ser empleadas para destilar conocimiento desde los datos. La analítica de datos masiva a partir de IoT se espera que contribuya a transformar el conocimiento tanto para individuos, como organizaciones y sociedad en general, facilitando tomas de decisiones más eficientes y efectivas.

Chapter 1

Introduction

Digitalization, also referred to as digital transformation, is about the integration of intelligent data into anything and everything that intersects our business workflows in different industries and personal streams of all life-span [1]. Data is at the heart of digitization, the lifeblood of the digital transformation process. Today, factories/enterprises are leveraging data to make employees and processes more productive and well-being; government and authorities are leveraging data to understand pollution exposure and climate change; companies are leveraging data to provide better caring to elders in an aging society and create new opportunities of competitive advantage-working toward the future of tomorrow.

The digital world owns its merits of continuous connectivity and communication, which will be always on, always tracking, always monitoring, always computing, always learning, and always adapting. This is because they are paving the way for knowledge learning benefiting from massive and diverse data. The term "epistemology" stems from the Greek words "episteme" and "logos". "Episteme" was translated as "knowledge" or "understanding" or "acquaintance", and "logos" was translated as "account" or "argument" or "reason" [2].

The Internet of Things (IoT) is a typical practice to boost the digitalization process in academics and industry. In 1999, the terminology 'Internet of things' was coined by Kevin Ashton of Procter & Gamble, when Radio-frequency identification (RFID) was viewed as essential to the IoT, enabling computers to connect and manage all things including individual things (including all existing things: machine and people). The basic idea of this concept is the 3A concepts: any-time, anywhere and any media [3], to reach a paradigm that everything around can be identified, interact with each other, and seamlessly integrated into the IoT network infrastructure. The IoT network enables physical object (machine/people) augmented with connectivity by sensors [4]. The wave of connectivity is towards smart homes/offices, smart wearables in healthcare/manufacturing, connected cars, connected manufacturing equipment, and smart cities. The total IoT connected devices are projected to reach the amount of 1.0 trillion worldwide by 2025, a five-fold augment in the coming five years [5].

The Internet of Things (IoT) is the next frontier in the digital revolution. More than a

decade ago, Michael E. Porter introduced the value-chain concept, depicting it as a series of activities that create and build value [6] in the field of Business Management. In the digitization world and the context of IoT, the data now is considered to be a vital strategic asset. The data value chain was defined by the European Commission: the 'center of the future knowledge economy, bringing the opportunities of the digital developments to the more traditional sectors (e.g. transport, health, manufacturing, public health, and financial services)' [7].

To understand the fundamental data value chain framework [8], the value streams of data can be described into gathering data from edges and endpoints, data storage, data processing/analytic, and promulgating information/the extracted knowledge back for consumption by end-users, including machines and people. This data value chain consists of four fundamental steps:

- Generation - sensing and capturing data leveraging the devices connected to things;
- Collection - collecting data, transmitting, validating and storing it via communication networks and various data storage techniques;
- Analytics - toward to different business and application domain, processing and analyzing the data to generate new insights and knowledge; and
- Exchange - Applying the outputs to use in practice for management improvement, triggering business opportunities, and boosting smart applications in society.

Figure 1.1: IoT value cycle and benefit paradigm.

The hyperconnected society, in essence, is based on hyperconnected IoT environments. A preliminary IoT data value chain model was produced by [9], which introduces the fundamental

IoT value cycle and benefit paradigm. Deriving from this preliminary IoT chain model, the adapted value chain model for this research work is introduced in Figure 1.1, where data flow and exchange across all IoT architectural layers. Linking the IoT value chain in Figure 1.1 and the four fundamental steps introduced aforementioned, the IoT data processes include the following procedures as a summary: data acquisition, data transmission, data processing, data integration, data storage, data filtering, data analysis/analytics, data usage, data exposure (openness), and data monetization.

1.1 Research motivation: Human-centered society

The goal of the next frontier society 5.0 is aiming for a prosperous human-centered society, in which people can achieve a high quality of life that is comfortable, healthy, and well-being [10]. The needs of diverse people, regardless of elders, employees, patients, individuals, etc, can be full-filled supported by the fusion of cyberspace with people. The digital replica of people can be established by IoT sensor-based data streams and corresponding virtual model built upon, that can be used for various purposes. The relevant concepts like digital twin [11] steaming from the manufacturing industry, shed a light to create a digital representation of holographic people's characteristics, that continuously learns and updates itself leveraging IoT, artificial intelligence, ontologies, and machine learning techniques. The digital twin of humans was proposed by [12], refers to collection and analysis of physical, physiological, and contextual data of human, to improve their quality of life and enhance well-being.

Different IoT architectures were proposed to enhance the IoT value chain in the next frontier hyper-connected society. The IoT umbrella covers different application fields, it appears that functional requirements and technologies used vary enormously [13]. The IoT is also a technological revolution that represents the future of computing and communications, and its development depends on dynamic technical innovation in fields, from connection, communications, storage, to future computing. IoT architecture was designed and defined by different research studies [14], depending on its booming development not only on the technology progress in fields of connection and future storage and computing, more on various new applications and successful business models [15]. It evolves from the early stage three/five-layer architecture [15,16], IoT architecture of cloud/fog computing [17,18], big data analytics architecture [18] to the novel Blockchain based architecture [19]. To seek autonomously adaptation to changing contexts without user involvement, many of these developments are still too device-centric or platform-specific [20,21], the fact that human context is often overlooked or is considered as an unpredictable element in their control loop. **General IoT architecture aforementioned lacks necessary emphasis on people dimension, which hinders its strengths and appeal to human-centered society.**

The concept of Industry 4.0 is not only applicable in factories, but also throughout the entire value chain system [22]. The RAMI4.0 model [23] was developed to build a common

reference architecture for all participants to share a standard concept and perspective. From the perspective of integration ability of the RAMI4.0 components [24], the integration of different sources of data includes RFID readers, Industrial IoT sensors, actuators, etc., but with limited consideration of human asset. Integration of human (operators) is only via enhanced Human Machine Interface (HMI) [25], which was also applied to the Cyber-Physical Systems architecture (5C) [26]. Industry 4.0 should carefully consider the central role of humans [27]. In particular, in industrial applications e.g, Industrial IoT, the long-term return of investment is imperative for promoting enterprise adoption, empowering people would pave the way to deliver values to all product life cycle.

People are an integral part of the IoT/IIoT ecosystem. The emerging concepts People-Centric IoT [21], Internet of People (IoP) [28], and human-in-the-loop (HITL) [29], are bridging the gap between IoT and the beneficiaries: humans. Humans and their interactions with smart devices can simultaneously be viewed as data sources for defining a people-centric IoT architecture [30]. In the IoT data value chain, the human context is useful that can augment a broad range of applications including environmental monitoring, occupational safety and health, aging care, among others. While existing studies primarily focus on how the IoT can assist people, the work that how people can cooperate with or assist IoT is rare in terms of closing the loop [21]. In this regard, personal smartphones/gadgets or wearable with a set of capabilities play a central role, that can improve the connection between people and the IoT in daily lives [28]. The approaches enabling people-centric IoT applications are needed to be enhanced, include aspects such as cognitive systems for smart homes [31], health monitoring, smart logistics, smart tracking [32], and behavior recognition [33], on basis of context-aware, condition-aware or mood-aware services.

1.2 People-Centric IoT: Health oriented

Health as the most important aspect of people has a significant impact towards to high-quality human-centered society. Health is vital to the entire lifespan: from home through the workplace, from commute to work to fitness and even old age/care homes. The fractions constituting the adverse health factors include dissatisfying living/working environment, anti-ergonomic body movements and postures, and psychological and physiological irregularity aroused by various activities. The health section is a fast-growing IoT market, the driver is to improve the quality and efficiency of various health services with new tools and technologies [34]. The networked sensors, either worn on the body or embedded in living environments, make it possible to gather indicative information of health-related parameters. To give the work a clear and specific focus, the research question is formulated: ***How to accelerate and enhance people's health and well-being in the IoT data value chain?***

Occupational health is a comprehensive practice of protecting and improving health conditions in the workplace. To ensure occupational health, regulatory agencies, e.g., Occupational Safety and Health Administration (OSHA) developed monitoring strategies to guarantee em-

employees' healthy working conditions. The environmental parameters such as chemical exposure and noise level are assessed by professionals sent from local regulatory agencies. However, the monitoring reference period is generally a short-term basis and it is regularly adopted to be 15 min in practice [35]. This setup raises challenges in terms of instrument calibration, labor time, and costs. Indoor working conditions vary greatly over time, and on spot short-term sampling or measurement in specific episodes could fail to provide a holistic assessment of the working environment [36]. The short-term monitoring strategy poses a potential risk to guarantee people's health.

Legislators, regulations, and standards were forced to intervene, which have brought to the forefront eliminating the occupational risk at the source. Considering the measurement strategy in regulatory agencies is short term basis due to limited resources, European Agency for Safety and Health at Work (EU-OSHA) [37] and the Partnership for European Research on Occupational Safety and Health (PEROSH) [38] pointed out the importance to explore using advanced ICT to ensure the safety and health of the worker, from a long-term perspective.

Continuous and persistent monitoring e.g. prevention, is no longer a word, with reaction to technological progress that applies sensor networks to support long-term monitoring [39]. In businesses flow, wherever in work office or industrial already understand that worker health is a major component of their financial success, as it ties to sickness absence, work-injure law-suite, productivity, and total quality. The study [36] has a deeper discussion of long-term monitoring benefits in terms of air quality management. To enhance long-term monitoring strategy, that will demand more monitoring capability and diagnostic centers that assist the decision-making process, which was pointed out by a systematic review of OSH in the industrial 4.0 era [39].

To ensure occupational health, working conditions are needed to be considered inside the workplace, as it is related to people's comfort and health. Working in poor air quality may lead to plentiful adverse health problems, such as headaches, nausea, sick building syndrome, skin irritation, kidney failure, and cancer [40]. IoT based environmental monitoring solution [41–43] shed a light to longer-term monitoring with low-cost sensors. Smart sensor networks for air quality monitoring applications [44, 45], AirCloud [46], ISSAQ [47], MAQS [48], hybrid [49] were developed to be able to detect temperature, humidity condition, gaseous pollutants (carbon dioxide, volatile organic compounds, nitrogen dioxide), particular matter, noise, or microbial contaminants (mold, fungus). The energy consumption and management, overall cost, data processing, sensor node, and sensor network infrastructure of previous research were summarised in study [47, 50]. However, customization and flexibility are lacking considering the monitoring solution is a block box for users, who could not configure and choose the sensors according to different monitoring environments. Regarding data processing, in existing approaches, less effort was made on data transparency and sharing among relevant stakeholders. **For monitoring solutions, the challenges in terms of customization, flexibility and transparency to all stakeholders (including workers themselves) are remaining.**

Health factors of people are not only related to their working environmental quality but also

psychological level and physical demands stimulated by job position. Wearable devices such as smartwatches, smart wristbands, smart insoles, and smart glasses are useful for ensuring people's psychological and physical health, as they can build cyber links to frontline workers in real working scenarios.

Wearable technology is becoming prominent and attractive in many applications, it also exhibits its significance on the quality of human life [51]. The use of wearable helps monitor the physical and physiological parameters of a person, to enable to monitor people's behaviors and activities and provide information awareness to ensure a healthy and wellness living environment and life. The wearable application consists of security, sports, and medical scenarios, where it is possible to monitor people's body temperature, heart rate, blood pressure, electrocardiogram (ECG), motion, brain activity, and skin response, to new a few [52,53]. Based on Futuresource research [54], the smart wearable sensors will change the people and their lives just like the personal computer did a few decades ago. The wearable technologies will impact every aspect of people's life, health, sport, fitness, eating habits, and medical diagnose, and redefine the relationship between consumers and producers, doctors and patients, etc. In particular, by 2017, in the healthcare market, the wearable sensor market grew to more than \$20 billion.

The basic architecture of wearable devices can be represented as shown in Fig 1.2. Based on different types of monitoring tasks, different sensors are used. The raw data measurement collecting from the sensors is sent to a processor (App in mobile phone, software platform installed in PC) for data multiplication. Then the processed data results are shown in displace to give user summarised data reports. **Nevertheless, the incompleteness of the presented model in Figure 1.2 remains, as learning capabilities are limited because of the non-persistent data and lack of explicit relationship (either time or activity-based).** Those commercial personal wearables only provide summarised data, as devices are linked to apps by Bluetooth, that mostly are designed to present the data to users and upload summaries to the cloud manufacturer [55]. The non-continuous and less-openness feature of commercial wearable hinders the learning ability benefiting from data-driven applications.

Figure 1.2: Block diagram of wearable sensing.

The human-centric cyber-physical systems (HCPSs) include broad concepts and technology to connect people with the cyber world with a sense of immersion. There is significant research

in the direction of empowering people including applying virtual and augmented reality applications to facilitate worker's activities by giving appropriate instructions, or for training needs [56]. The application of elements like positioning tags, safety gloves [57] and helmet, head-mounted display (HMD), and smart glasses [58] have started to be increasingly used in industrial applications. Different Industrial Wearable System (IWS) applied in industrial context pointed out the significance to consider wearable devices as contributors to the production system's performance, from its business value [59]. The industrial wearables market is projected to grow to \$2.78 billion by 2024 [60]. Wearable devices such as smart watch [61] and smart wristband can be useful for estimating operator's well-being (cognitive, psychological, and physical needs) in Human Cyber-Physical Systems (H-CPS) [62].

Towards occupational health, safety, and productivity of the well-being of workers, a conceptual framework for OSH risk management in smart working environments leveraging IoT technologies and ambient intelligence was proposed by [63]. In its foreseen smart working environments, worker's skill levels, cognitive abilities, and psychological states, as well as working ambient can be analyzed and adapted within building cyberspace. Healthy operator 4.0 was envisioned by [64, 65], where wearable trackers are used to measuring activity, stress, vital sign, biometrics data, and other health-related metrics. The cognitive or physical workload can be relieved and managed by assessing the operator's well-being via smart wearables such as skin conductance, heart rate, steps, motion, acceleration, and ergonomic metrics. Healthy Operator 4.0 aims to provide sustainable solutions for workers to support better job satisfaction and well-being.

Nevertheless, in current industrial practice, most applications are developed in isolated circumstances aimed at addressing specific problems with a single device (e.g. wearable). **Therefore, there is a gap in creating human-centered systems able to promote operators' learning context not only relying on single parameters but also providing a meaningful articulated set of relevant parameters both in the short and long term [66]. Besides, when the interest is to learn from the psychological and physical impact of different workers over the processes, the contribution is much less.** As the impact in terms of effectiveness depends on those cyber links, a higher level of data integration considering multiple subsystems (process and operator) is vital to achieving an effective and efficiency human-cyber-physical symbiosis [59]. Complex rules and knowledge can be learned or inferred from large scale and diverse data sources integration, in particular when significant variability sources are involved (operators).

When the data source is related to people, trust, security, and privacy concerns are more demanding as they are sensitive health data. "Protected Health Information" (PHI) has a higher value on the illegal market than credit card credentials. There is a global risk that cyber-attacks and data breaches to target health-related databases [67].

Although data management technologies exist in sensor networks or wearables devices, they are mostly based on centralized models. Data security and privacy face issues if the generated

data is stored in a centralized topology [68]. The limitation lies in many-to-one traffic flows (server-client model), which in turn increases delay and the problem of a single point of failure. Moreover, how to ensure the data has not been tampered/alterd/falsified, is a difficult question to answer in centralized architectures [69]. Centralized organizations managing sensitive user data may misuse them illegitimately, thus leading to a breach of privacy [70]. The number of networked machines, devices, and people includes in the myriad of IoT may increase rapidly with the prevalence of IoT. The developed IoT architecture must be scalable, however, the centralized based models issue to manage them globally [70].

Data streams/lakes/sources applied in the real world applications can be from many aspects. For instance, in an industrial case, the health-related parameters may come from the production system, industrial wearable system, and personal wearable. That accounts for different data sources that are needed to foster a holistic health-oriented solution. Furthermore, regarding data transparency to stakeholders, it is difficult for relevant stakeholders to access those data, as distinct data sources are fragmented in different data clouds. For instance, if a company needs to share its environmental qualify data to OSH regulatory agencies or other peer entities, the data sharing will be hard to implement as it needs to across different IT infrastructure to access the centralized database. Furthermore, relevant stakeholders such as OSH agencies, either can not benefit from the continuous long-term sensor-based monitoring solution. When diverse data sources are needed to be shared among stakeholders, the data lakes involved are across several data silos. A shared data management platform would be beneficial for all relevant stakeholders to follow up on all life-cycle information of product, in such a way data integrity and transparency can be ensured.

Data ownership is an additional challenge. When asked, 88% of U.S. internet users and 87% of British internet users would want to control the data being collected through smart devices [71]. The centralized data management solution issues data producers to reuse their own generated data outside the enterprise's internal environment. People lose data ownership and also the possible business opportunities stemming from those data, as people can sell their generating data to interested data consumers. Under this perspective, when referring to data related to people, the EU General Data Protection Regulation (GDPR) is an effective tool to add significant trust to this dimension, as users can now understand their rights and privacy is enhanced. **Therefore, a solution providing with salient merits: immutability, transparency, traceability, and reliability, is demanding to ensure trustworthiness in a scalable and robust way.**

Blockchain technology (BCT), as well-known technology that underpins the digital cryptocurrency like Bitcoin initiated by Satoshi Nakamoto [72], litecoin and Ethereum, has been recently used to provide security and privacy in peer-to-peer networks. The peer-to-peer network posses a similar topology to IoT [68, 73], which indicates BCT is also promising for IoT. BCT is based on a distributed network, which ensures scalability and robustness by applying resources of all participating nodes. It can eliminate many-to-one traffic flows and decrease

transmission delay even overcome the single point of failure problem. The Blockchain-based IoT architecture, in essence, is to use Blockchain technology to manage access control [70, 74, 75] or secure, decentralized data storage and sharing [76].

To ensure the transparency among trusted entities, integrate BCT in IoT may provide trustworthiness via its salient features [68]: immutability, transparency, traceability, auditability, reliability, and operational resilience [75, 77]. It realizes a secure database over untrusted parties: heterogeneous devices, organizations, etc., by a decentralized consensus. **The increasing development of Blockchain and its foundation Distributed Ledger Techniques (DLT), comparing the different DLT and choosing the proper one for the health monitoring scenario is a challenging task.**

1.3 Formalization research objectives

On the premise of research status-quo and the pinpointed research question, the research objectives in this study were defined. The aim is to accelerate and facilitate health-related data-driven knowledge acquisition and data value chain to society, organizations, and individuals, that leverage advanced technologies such as IoT, Smart Wearables, DLT, and machine learning techniques. To summarise, the primary research objectives shaped by the research question are formalized as follows.

1. *Enhance people dimension, especially health and well-being perspective in conventional IoT architectures.*
2. *Develop continuous long-term monitoring solutions to support better health management.*
3. *Learn psychological and physical health impact from (e.g., workers) a group of people over the activities.*
4. *Accelerate health-related data sharing with security trust and privacy assurance.*

Chapter 2

Methodology

For the methodology chapter, it composes the research method, a guiding research framework, and study designs. The research method was employed as a theoretical approach to analyze and implement the research objectives. Adhere to the applied research method, a guiding framework was designed to set up the research theme. It also helps visualize the theoretical research framework and depicts the key concepts, enabling technologies, tools, and devices. In the end, several study designs are introduced to achieve defined research objectives. Each study design includes experimental setup, data collection, and data analysis plan. About the research design, the holistic overview of the big picture is demonstrated in Figure 2.1.

Figure 2.1: Big picture: an integrated research design.

2.1 Research method

From a methodological point of view, this study employed an action research method. Action research method [78] is a theoretical method to test and verify research hypotheses in a real-world scenario. Action research depends on the collection, analysis, action, evaluation, and critical analysis of practices based on collected data and evidence aimed to bring improvements in real-world practices. This type of research is conducted by the participation and collaboration of some individuals with a common objective. Following the action research essence, the mission that looking for effective strategies to improve people's health and well-being was carried out in several real-world enterprises.

2.2 A guiding framework

Following the defined research method, a reference framework is designed as shown in Figure 2.2. It was used to build up the research context. The research framework crosses six IoT layers (Sensing, Network, Storage & Sharing, Modeling & Analysis, and Application layer), which was designed based on state-of-art IoT architectures.

Different IoT architectures were designed and defined by extensive research studies [14], depending on rapid development not only on the booming technology but also on diverse new coming applications and potential and vital business models [15]. The IoT umbrella covers different application fields, it appears that functional requirements and technologies used vary enormously [13]. The IoT architecture is also a technological revolution that represents the next frontier computing and communications, its development relies on dynamic technical advancement, from connection, communications, storage, to future computing.

Three-layer architecture [15] was introduced in very early stages. The three layers are the perception layer, network layer, and application layers. This architecture as the initial architecture and was extended by the following research work.

(i) The Perception layer is the sensing layer, which is used to identify physical objects and gather information. The perception layer includes smart connection objects such as RFID tags, cameras, sensor networks, and GPS. (ii) The Network layer is responsible for connecting smart objects to the IoT infrastructure network, is used for transmitting and processing the information gathered from the perception layer. (iii) The Application Layer is a fusion of IoT's social applications and industry demands. It envisions various applications in real-world applications such as smart manufacturing, smart homes, smart health, and smart cities.

The three-layer architecture expresses the main idea of IoT, but it is not enough for the whole IoT research aspects and future needs. Targeting different goals such as business-oriented, based on the three-layer architecture, more layer architecture was proposed to define finer and richer aspects of IoT.

Figure 2.2: A guiding framework and approach for implementation research.

Fiver layers architecture was proposed in [15, 16], with two more layers: processing and business layer added compared with three-layer architecture aforementioned. The processing layer as the middleware layer, it works to store and analyze a huge amount of data. The data management technologies such as databases, data processing, and data manipulating are employed in this layer. The business layer is oriented to profit business models by managing IoT systems or data enabling business applications. This is because the real success of the IoT architecture relies on good business models [15]. This business layer will be adopted to determine future business strategies towards a human-centered society.

Since a great number of connected devices is rapidly increasing in the IoT infrastructure network, so there is going to be generating tremendous data as well. Storing/processing that massive data locally and temporarily will not be feasible anymore. As cloud computing gains ground, there is research moving towards to cloud-centric IoT architecture. The user does not need to worry about the huge burden of maintenance or managing/processing all the resources locally.

IoT architecture of cloud computing The IoT and cloud computing working in integration make a new paradigm, as termed Cloud of Things (CoT). These architectures added cloud computing layer to get flexibility and scalability of data storage, data analytics, integrity, and privacy. The three key cloud computing models that interacted with cloud computing-based IoT architectures are Infrastructure as a Service (IaaS), Platform as a Service (PaaS), and Software as a Service (SaaS) [17, 79–82].

Cloud computing-based IoT architecture is a centralized computing model, most of the computations occur in the cloud [18]. Although the data computation and processing speed have risen rapidly, the network bandwidth has not increased appreciably accordingly, which may lead to long latency. The IoT latency-sensitive applications which require a short response time can not afford the delay caused by data transferring. The decision making in the cloud requires high-quality data, it is not necessary to transmit all the collected data to cloud considering its inefficient processing and storing. Raw data can be filtered, converted, or integrated before transmitting to the cloud. Besides, some decisions can be made locally without transmitting data to the cloud.

IoT architecture of fog computing. To overcome these issues, fog computing extends cloud service to the network edge. A new layer called fog layer is added between end devices and clouds. Fog computing architecture adds the fog layer between end devices and cloud to address these challenges such as low latency. It also enables high reliability and security, high performance, mobility, and interoperability [83, 84] to optimize IoT architecture.

In the big data era, cloud-centric/fog computing IoT architecture can not sufficient to fully implement the IoT paradigm. Big data is featured by its three main aspects: (a) volume, (b) variety, and (c) velocity, which were first popularized by Gartner to introduce the big data challenges [85]. Massive data is generated each second from IoT devices (sensors, actuators).

Those data are stored and managed in a cloud/fog computing platform or big data platform. However, how to fully make use of data to mine value information to support decision making is not answered in cloud/fog architecture. The current architecture up to now mainly focuses on IoT concerning data acquisition and data communications [86].

IoT architecture and big data analytic. With data-driven Artificial Intelligent (AI) technology booming, another new layer: big data analytic layer was popped out, where advance technology such as machine learning and deep learning, Hadoop, Kubernetes, and Spark can be employed to dig knowledge out of data, and forecast future patterns out of history data [86]. IoT big data analytic layer is aimed to assist business associations and organizations to achieve an improved understanding of data, by which they can find efficient and effective strategies [87]. The big data analytic layer powered by methods such as big data and distributed computing can also be used to improve the efficiency of IoT.

Data security and privacy are again challenged in the big data paradigm. Access management technologies do exist in IoT, but they are mostly based on centralized models that are hard to be managed on a global scale [70]. Data security and privacy face issues if all collected data is stored in a centralized topology [68]. To ensure the data has not been tampered/alterd/falsified is a tough question to answer within centralized architectures [69]. This brings about the need to verify that the information has never been modified or tampered. Centralized organizations managing sensitive user data may misuse them illegitimately, thus giving rise to a breach of privacy [70].

Of course, typical solutions such as authentication, authorization, and communication encryption by suffering excessive delays and overheads or computationally expensive could potentially compromise user privacy [88, 89]. In many circumstances, that may not inappropriate for latency-sensitive applications and many resource-limited IoT devices.

IoT architecture and Blockchain. Blockchain technology (BCT), as popular technology that underpins the digital cryptocurrency like Bitcoin initiated by Satoshi Nakamoto [72], has been extensively applied to implement security and privacy in peer-to-peer networks [68, 73]. Integrate BCT in IoT may provide trustworthiness via its merits such as [68]: immutability, transparency, traceability, auditability, reliability, and operational resilience [75, 77]. The advanced and recent technologies like smart contracts and consensus protocols [75, 90] pave the way for implementing those types of merits.

The BCT based IoT architecture brings scalability and robustness features by using resources of all participating nodes. Many-to-one traffic flows are eliminated under such architecture. In turn, it decreases transmission delay and addressed the issue of a single point of failure. It realizes a secure network over untrusted parties: heterogeneous devices, organization, etc., by a decentralized consensus. The BCT based IoT architecture is to adopt DLT as a technical foundation to manage access control [70, 74, 75] and secure, trust, and decentralized data storage and sharing [76].

The IoT can change the shape of the society and offer enormous benefits to the next frontier society 5.0 but it also faces many key challenges [75].

- *Challenge#1 – Distribution.* The devices on the IoT such as sensors, wearable devices, or other relevant types of devices or services are distributed; Furthermore, IoT infrastructure and services should enable the distributed users to be accessible and consume the services on the IoT.
- *Challenge#2 – Heterogeneity.* Interoperations ought to be supported among heterogeneous devices in the IoT network. Generally, the devices over IoT are with different operating systems (OS), data format, interfaces, and communication protocols. These devices can generate huge amounts of data, which usually come in different formats and meanings [91].
- *Challenge#3 – Trust.* It is impractical to rely on a common and single security control mechanism for all types of IoT devices and services. The trust of an IoT device or service is based on the security and trusts from all of the networked devices and services over the IoT infrastructure network.
- *Challenge#4 – Scalability.* The amount of networked and connected machine, device, service, and people in the myriad of an IoT will increase exponentially, as the rapid development of sensor technology and communication protocol. The developed IoT architecture and framework need to be scalable to adapt to the booming prevalent of IoT.
- *Challenge#5 – Resource constraints of IoT devices.* Devices over IoT such as sensors, smart wearable, RFID tags suffer from limited resources such as battery power, computing, and storage resources.

The research framework is functional as a guideline giving a broad overview of this research study, in terms of both technological, business, and application aspects. It took consideration of all significant merits of the state-of-art IoT architectures, as well as the key challenges confronted. More detailed information derived from the framework is introduced in the following subsections, that composes enabling technologies, applied IoT devices, and the developed prototypes.

2.2.1 Enabling technologies

Different technologies were used in connection with most of IoT data process flow, corresponding to different layers in the guiding framework. The study applied the sensor networks and smart wearables for health-related data collection. As a sensing layer, they are mainly identifying objects and gathering information. To extract knowledge in the modeling and analysis layer, machine learning and data mining are the tools for advanced data analysis and further decision making support. Ontology becomes the knowledge representation, with which multiple heterogeneous data sources from different domains can be fused, semantic-linked, and exchanged in

common understanding space, thanks to the standard representation format. Data usage, data exposure, and data monetization are important elements in the storage and sharing layer. DLT, specifically IOTA, is applied as a solution designed for IoT, which accelerates data storage, data sharing, and data reusability in a secure, fee-less, and scalable manner.

Sensor Networks. The number of sensors deployed in the IoT network is growing at a rapid pace. Sensor networks as an important aspect of IoT are functional in the sensing/perception layer. It may consist of an enormous scale of sensing nodes, which are addressed and connected within the protocol stack. The IoT devices such as sensors and wearables will bridge the gap between the physical and digital world to improve the quality and productivity of life, industries, and society ¹. Wireless sensors and wireless sensor networks (WSN) have many well-known applications: military applications, environmental monitoring, robotics, autonomous vehicles, smart agriculture, asset tracking, and smart building/home [92]. The typical sensors are: temperature sensor, motion sensors, pressure sensors, proximity sensors, noisy sensors, lighting sensors, and optochemical sensors [93] for air pollutants O₃, CO₂, CO. Environmental sensor networks with optochemical sensors were used in this study to monitor working ambient conditions in the workplace.

Context-awareness. Context-aware computing is proven to be successful in understanding the real-world situations. Since the early 1990S, context awareness has been studied as a core element in ubiquitous and pervasive computing sensor networks. The "context awareness" concept and its definition were proposed by Abowd et al. [94] in 1999, which has been widely accepted by the research community. To analyze the human's health such as abnormal physiological parameters of the people, the context information is required for understanding the physiological status occurrences under which adverse events are triggered. The premise that the environment affects people's physical activity, productivity, cognitive function, health, and well-being is well-established [95, 96]. The environmental conditions may act as stimuli to trigger irregular physiological status. To provide a healthy and high-performance environment, a holistic approach is needed. There is substantial knowledge of individual factors and their surrounding environmental factors as context, thus understanding how factors interact and what role occupants and environmental factors play in these interactions (both causative and receptive) is essential.

Data fusion. Building a complete picture of people's health could be achieved by the fusion of the data gathering from multiple sensor devices. To interpret the information and take appropriate decisions, different sources of sensory information is needed to be integrated into an information knowledge base. "Multisensor data fusion" was proposed by [97], it is defined as the technology concerned with the combination of how to combine data from multiple (and possible diverse) sensors to make inferences about a physical event, activity, or situation. The principles of multisensory integration are that crossmodal stimulus. Multisensor are integrated

¹<https://www.analyticsvidhya.com>

to find confidence in the data produced by the sensors. For instance, to give an integral vision of a specific human activity, it should be on a multisensor integration basis. The personal environmental exposure data would enhance additional knowledge of the human activity. After human movement data was classified into accurate activity groups, the environmental data can be fused to each activity data segmentation to allow a more complete view on the people's health condition.

Smart Wearable. In European countries, about 900,000 people die of sudden each year, caused by prevented cardiovascular diseases. Transient abnormalities can not always be captured because lots of cardiac diseases are episodic rather than continuous abnormalities. Those abnormalities are vital but their timing is hard to be predicted and detected. Life-threatening disorders can go undetected as they occur episodically or may never be captured objectively. Wearable technology is worn on a part of the body, focuses on tracking and improve safety, health, activity, and well-being. With the emergence of biosensor development, there is likely a rapid expansion of wearable biosensors in healthcare with the corresponding reduction of cost. The wearable devices enabling the monitoring includes smartwatch, smart glass, tracking devices (e.g. air pollution tracker), and mobile phones. Other common smart wearables include smart rings/pens, smart clothes, smart gloves, smart shoes, headphones, smart screens worn as the head display, which are normally for industrial applications. The ECG system is also emerging not only for patients but as general use in daily lives, like the Apple watch is embedded with this function. The EEG systems [98] with their functionalities to monitor brain actives, have started to be used in school, workplace, and industrial factories. In this study, the smart wearables: smartwatch, air pollution tracker, and smart glass were used to track people's stress level, concentration level, fatigue level, and personal pollution exposure level.

Ontology. Heterogeneity becomes a key characteristic in sensor networks. The sensors connected to IoT are from different hardware platforms and software architecture design. The data generated from sensors have different data types and data structures. Data interoperability is a vital design requirement to enable data interconnection and data integration across different devices and service platforms [99]. For high-level applications and services, data has to be interoperable and linked to other domain knowledge. Semantic modeling, such as ontologies based techniques have proved to be successful to deal with syntactic interoperability and semantic compatibility.

Ontology is a formal, explicit specification of a shared conceptualization [100]. In computer science, ontology is a formal representation of the knowledge by a set of concepts within a domain and the relationships between those concepts. Ontology was considered in many application domains to address data interoperability as it can be used to define a common understanding space by specifying the used concepts and their relationship in between, in a formal and standardized way. Ontology as considered in modern knowledge engineering provides standard domain vocabulary to maintain a common space of understanding. The metadata languages aligning with

ontology vocabulary would convey that kind of shared information in a landscape of information management. The representative metadata languages are XML, RDF, and OWL. Considering the balance between richness and usability, RDF achieved prospective usages in practical settings [101]. RDF is a graph-based data model that provides a grammar schema for data interchange on the web. RDF stream modal for storing environmental sensor data was proposed by [102].

Under sensor-based IoT context, the sensor-related data standards e.g, SSN ontology ² enabling data linking and annotating in IoT applications. The SSN ontology was developed to define core concepts and relations including sensors, features, properties, observations, and systems, etc. In health-related domains, extensive ontologies were constructed to represent the existing knowledge in a standard manner. Ontologies are used in health include upper-level ontologies (Basic Formal Ontology (BFO) [103] and Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) [104]), domain-dependent ontologies (gene ontology, vital sign ontology [105], and air quality ontology [106]), to new a few. More recently, ontologies and semantic technologies have been used for environmental modeling and representation. Indoor Environmental Quality ontology was proposed by [106], which can be used for indoor monitoring and control applications. In this regard, the data sources in this study are from several hardware and software architectures: air quality sensors, smartwatches, smart glasses, and positioning trackers. Corresponding ontologies were reused to sharing knowledge among entities in the value chain across knowledge bases, rather than just sharing data itself.

Machine Learning. Machine learning (ML) is a subset of artificial intelligence. The process is to build a mathematical model with training data, that perform the needed tasks: prediction, classifications, regression, and clustering among others. The ML approaches are divided into three broad categories: supervised learning (Support Vector Machines (SVM), logistic regression, naive Bayes, decision tree, K-nearest neighbor, multilayer perceptron), unsupervised learning (k-means, hierarchical clustering, principal component analysis, and autoencoders), and reinforcement learning. In recent years, deep learning (convolutional neural networks (CNNs) and deep belief networks (DBNs)) [107] has become a core methodology as it can automatically learn high-level features. For health-related data collected from IoT devices or wearable, ML algorithms such as association rule learning were explored to discover associated relations among diverse data sources in this study. To understand human behavior clustering or health status, classification/regression is helpful to trigger health alters, to model human digital twins from a long-term perspective.

Association rule learning. It is a rule-based machine learning method for discovering relations, patterns, and rules among variables in a given data set. An association rule is composed of two parts: an antecedent and a consequent. To identify interesting rules from a set of all possible ones, constraints on various measures of significance and interest apply. To identify the most

²<https://www.w3.org/TR/vocab-ssn/>

important relationships, the criteria include support, confidence, and lift. Support is a measure of how frequently a given itemset occurs in the whole data set. Confidence is the proportion of how often the rule has been found is correct. The lift of a rule is defined as the ratio of the observed support to that expected. The value of lift gives the degree that two occurrences are associated with each other when the lift is greater than one. Many algorithms have been proposed for association rule learning, such as the Apriori algorithm, Eclat algorithm, and FP-growth algorithm.

Cluster analysis. It also called classification analysis or numerical taxonomy, which is to group a set of data objects sharing common characteristics into relative groups/clusters. The computing distances are used to separate observations into different groups. There are more than 100 clustering algorithms, every methodology follows different criteria for defining the distances among data points. Some more popular ones which are used widely, they are connectivity models, centroid models, distribution models, and density models. The specific algorithm includes hierarchical clustering, k-means, and Density-based spatial clustering of applications with noise (DBSCAN).

Regression analysis. The regression model is a supervised technique, which is used to predicate a continuous value. In data sets, some variables are observed, such as independent variables (Characteristics that can be measured directly) and dependent variables (value depends on independent variables). In the statistical domain, the regression model is to estimate the relationship among those variables based on a set of statistical calculations. Types of regression models include simple linear regression, polynomial regression, support vector regression, decision tree regression, and random forest regression. Take the example of ergonomic parameters measured from wearable sensors. With the regression analysis used, the forecast ergonomic evolution is predicated in advance to prevent unnecessary anti-ergonomic movements or potential physical injuries.

Distributed Ledger Techniques. Generally, today's IoT solutions rely on centralized server-client methods of administering connectivity and interaction to various devices and services. With the increasing amount of IoT devices, deploying IoT solutions at scale will have to address complex issues such as security, latency-sensitive, connectivity, and cost. DLT is one of the technologies that can help if IoT solutions need to grow to scale. A distributed ledger is a database that is spread across different nodes. Changes to a distributed ledger require multiple authentications from these nodes [108]. This consensus protocol replaces a central administrator since all peers contribute to maintaining the integrity of the database [109]. This mechanism creates shared accountability and the pooling of resources to manage the ledger via a consensus protocol. The state of art consensus protocols was presented in [110] with a comprehensive review and analysis.

Currently, data is mostly limited to the control of specific entities and not available for the broad masses. IOTA presents a counterpart to the status quo, makes it possible to securely store,

sell, and access data streams. IOTA is a tangle-based consensus protocol designed specifically for the IoT industry. The IOTA was used to share data among engaged parties in the value chains, as its consensus protocol for building trustiness is superior to other blockchain's such as bitcoin and Ethereum, in terms of scalability, fees, and quantum attack [111]. The IOTA foundation also envisions a new model with which you can retain the data ownership. In a generic monitoring solution with IoT, the generated data is locked inside the organizations. Even the data from personal wearable, that is related to health like sensitive confidential information, data producers: the people do not own the control for their data. With the data ownership model in IOTA, data producers can sell the data they generated to other consumers to harvest economic benefits. The IOTA equipped with the smart contract by Qubic is a very important part for the future business processes like the data market. As IoT and sensor technology revolution, data is becoming increasingly important. Data is the new oil, the new gold. The immense growth of data marketplaces is an inevitable result.

2.2.2 Applied devices

The relevant IoT devices which were applied in our research work, are shown in Figure. 2.3.

Smart Watch: H Band. The device is a low-cost smart band containing sensors for measuring heart rate, blood pressure, step, and air pressure. Based on the App provided by the manufacturer, a customized App was developed for data collection [112]. The collected data can be transferred to an Android smartphone and then uploaded to a remote database. The frequency is every one-minute measuring once.

Smart Eyewear: JING-MEME. This is an eyeglass-type wearable device [113]. It applies three-point electrooculography sensors, accelerometer sensors to track peoples' body and mind so that to examine and analyze peoples' conditions (focus level, fatigue) in different activities. The JING-MEME logger App is used to record raw data for possible further analysis: eye blinks, eye movement, and eye rotation.

Kagoo Environmental Sensors. It is manufactured by Circulate that is used to measure environmental parameters [114]. Twelve environmental parameters are measured including chemical conditions and physical conditions. Different sensor combinations can be configured to be plugged into a workstation island, where each island five maximal sensors are loaded. The other hardware is a Raspberry Pi (a wifi router function supported). A wifi module works as data transmission media between sensors and Raspberry Pi. The Raspberry Pi is responsible to receive data from the sensor, manipulate data, and transmit data to the server. Moreover, a mobile phone is needed for configuring sensors to send data measurement to raspberry Pi via an App: EnvCon.

Shimmer Sensing. The solution provides wearable devices which could monitor human physiological response such as heart rate and galvanic skin conductance level [115]. It enables

Figure 2.3: Applied IoT devices in the study.

more than 30 people simultaneously monitoring, which powers large scale of human subjects participation. The real-time data can be streamed via Bluetooth to NeuroLynQ software platform where complex emotional response analysis results are provided to users. All the continuous data is stored into the local database and also logged into an embedded SD card.

MetaSensor. It refers to MetaMotionC (MMC) from mbientlab (<https://mbientlab.com/>), that enables us to monitor angles and acceleration and was used to track the arm angle through time. Similar to H Band data collection, the measurement is sent to a customized App on a smartphone, and then stored in a remote database when the phone reaches internet access.

ATMO TUBE Wearable Air Pollution Tracker. It is an air quality tracker and gives an air quality score. PM1, PM2.5, and PM10, VOCs; temperature and humidity; a barometer, and altimeter are measured by this tracker [116]. The environmental exposure can be monitored along with people's trajectory indoors and outdoors. The recorded data is stored Atmotube App installed on a mobile phone and can be exported to Google drive. The frequency is every one-minute measuring once.

Foobot Indoor Air Quality Monitor. It is an air quality monitor that sensing fine particles, total VOCs, temperature, and humidity [117]. It also provides CO₂ concentration level based on a calculation model. It is based on a Wi-Fi connection and every 5 min to store data to the cloud. The data collection can be done via request datapoints from API provided in their official website. The devices can also work with IFTTT ³ that can be configured to store data in Google drive sheet form.

Blukii Beacon. The smart beacon works with the Bluetooth Low Energy (BLE) standard and is Eddystone and iBeacon compatible [118]. The e-beacon data were broadcasted through Eddystone, an open-source beacon format. The smart beacon works together with an app: Owntracks [119] to track people indoors. The indoor tracking data were stored in a local database through the app.

2.2.3 Prototype: proof of concept

The proof of concept (PoC), earliest was coined by Bruce Carsten in 1984. The term now is widely defined as a realization of a certain method to demonstrate the proposed concept or theory whether it has practical potential in a real-world environment. As the working domain is related to system development, prototyping is working as an early demonstration built to test the feasibility of the proposed concept/architecture/framework. In this research work, four prototypes were developed to evaluate the usability of the corresponding framework and concept. The relevant source code could be found on GitLab [120]. Each prototype or combination of prototypes was used in different research actions, where each action is targeting one or multiple research objectives.

³<https://ifttt.com/>

Indoor air quality (IAQ) monitoring system [36]. Measuring IAQ conditions would be the first pace to ensure indoor conditions. If we can not measure it, we could not improve it. The system is designed as an IoT based monitoring system, which is developed using kagoo devices manufactured by Circulate, Raspberry Pi, and a central server. It supports real-time and long-term data collection, which will facilitate the research community and practitioners for long-term and integrated multiple pollutants data collection solutions. Several indoor environmental conditions are measured, including air quality, acoustic conditions, lighting, and thermal comfort. Nine sensors, particulate matter (PM), formaldehyde (HCHO), Total Volatile Organic Components (TVOC), benzene (C₆H₆), carbon dioxide (CO₂), carbon monoxide (CO), ozone (O₃), nitrogen dioxide (NO₂), and T.H.I.N (Temperature & Humidity & Illumination & Noise). Those sensors can be freely selected and plugged into an island, to support customization and flexibility for different monitoring environments. The system also contains an interface to connect to the data-sharing system, from which IAQ data can be shared among trusted entities/stakeholders based on DLT techniques. Meanwhile, the transparency among different stakeholders can be implemented via the secure data sharing system.

Location tracking system (indoors and outdoors) [121]. The location tracking system is a low-cost solution, which uses a personal smartphone and beacons to track people indoors. Beacons are deployed in different regions in indoor space. Different indoor regions are specified with identifier numbers of the beacons in the App: OwnTracks [119] installed on the phone. The solution fully employs personal smartphones to connect people to IoT in an economic way. The OwnTracks itself can also support outdoor positioning. The integrated functions were loaded into a docker container, enabling run-time on container level addressing platform or dependencies, configuration files control issue. The location tracking system servers to understand people's locations. In our study, it was used to learn how much time people staying in each micro-environment indoors to better assess their pollution exposure across micro-environments, combining their exposure outdoors as well.

Human physiological condition monitoring system [66]. The system is composed of a smart band, a mobile phone, and a local server. The smart band measures human physiological parameters like body's heart rate and blood pressure. The measurement is transmitted to the mobile phone via Bluetooth, and the data will be stored in the local database when the mobile phone has access to Wi-Fi or 4/5G, through an app running on the phone. The implemented system is useful to monitor human physiological conditions persistently and continuously. With the help of physiological information, makes it possible to assess people's workload, stress level, and mood along with timeline and activity diary. It would combine with more data sources such as data lakes from production and data streams from working ambient so that the holistic digital twin of humans can be built to improve job satisfaction and well-being.

IoT data sharing system based on DLT [111]. Trust is a concern to shared data with other counterpart entities in the value cycle. The designed data sharing system was based on

DLT, which is IOTA, is a distributed Ledger specially designed for IoT. It is the first open-source distributed ledger that is being built to power the future of the IoT with feeless microtransactions and data integrity for machines. IOTA presents counterpart to the status quo, makes it possible to securely store, sell, and access data streams. It supports secure, tamper-resistant, and highly-scalable data exchange based on Tangle data structure.

The Tangle-based cryptocurrency overcomes fundamental limitations of blockchain e.g., Ethereum, by using a directed acyclic graph (DAG) data structure for storing transactions. Two types of transactions are supported in the system, which is Masked Authenticated Messaging (MAM) based and message-based.

MAM is a data communication protocol which adds functionality to emit and access an encrypted data stream. It can be used in multiple ways to control visibility and access: public, private, and restricted. The combination of public and restricted MAM protocols could provide users granular control over the IoT data, which could bring great benefit to all stakeholders. For instance, when a workspace is needed to be OSH assessed, the manager can share with the accredited inspector the address and the side_key to the report air quality data streams from a certain time. Then the inspector will be able to fetch and evaluate all data streams. Afterward, the manager can change the key to revoke the authorization.

The modeled data can be also transmitted regularly to IOTA Tangle as a permanent message. To interact with the IOTA network, it is needed access to a node. A public node has been used to submit data from IOTA Tangle. The PyOTA is functional as IOTA python API library to implement data sending to Tangle. The IOTA JavaScript library is used for data retrieving since it is the most established IOTA libraries. By knowing the address/tag/bundle of the message, as well as the proper public key of the sender, will enable the user to consume the messages. The source code could be found in [122], which supports Python and NodeJS.

The data publisher can be any organization who produces continuous data such as environmental data. The data are published to the Tangle with MAM protocol, using specific encryption and privacy protocols. The data are published in their channels, and each channel has an address. The subscribers of a data channel will receive the newly published data. The published data are usually encrypted, and an extra decryption key may be necessary to decrypt the received data.

All the data are published and received through an IOTA node, which is a computer connected to the IOTA network. Users may use their node or use public nodes. A user can be a data publisher and a subscriber at the same time. For example, the employer can publish the workspace environmental data, and the employee/the accredited inspector from regulation agencies can subscribe to these data and review working environmental quality accordingly. Likewise, the inspector can publish the evaluation result to the Tangle, and the employer can subscribe to this channel and receive the result.

As many stakeholders are engaged in the value chain, each of them would need to access related data streams containing details of the whole workflow such as environmental footprint.

The data streams include production data, worker's health data, worker's activities, and working conditions data, that can be shared among stakeholders via the designed system.

2.3 Study design

To answer the defined research question, the four research objectives defined in the introduction parts play as drivers to implement the research study in practice. To fulfill the research objectives, several study designs were set up keeping the replication principle. In this thesis, we include four of them, considering the rests are either system/model-oriented or under the publishing process.

All the study designs are aimed to empower people's health and well-being within IoT applications. The included study design one and two are targeting research objective that is to apply continuous long-term IAQ monitoring solution to support better health management. The study design three and four, extending the long-term monitoring strategy, but mainly are aimed to learn psychological and physiological impact over people in different activities. The research objective that accelerates health-related data sharing with security trust and privacy assurance was mainly implemented via the aforementioned DLT based data sharing system.

Ethical approval: All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Furthermore, I have completed the ethical training program 'Social & Behavioral Research Investigators', under requirements set by Harvard University. All participating populations signed a consent form.

2.3.1 Study design 1

Exposure to indoor environments across the entire lifespan: from homes (bedroom, kitchen) through child care facilities, from schools, transport (bus, metro, airplane) to work; restaurant/bar, shopping malls, fitness room, hospitals, institutional buildings, and even old age/care homes. It ties to every entire life cycle and has profound impacts on many social aspects.

Environmental quality also plays a role in human behaviors indoors. The lighting system can induce eye strain or have impacts on circadian rhythm [123,124]. The increased CO₂ significantly decreased decision-making performance. Chronic exposure to NO₂ can lead to adverse respiratory effects such as airway inflammation and stressed respiratory symptoms to people with asthma. Exposure to Ozone can cause respiratory symptoms and lead to eye irritation for healthy people. It may increase risks for respiratory issues and visits to hospital admissions for asthma patients. A high concentration of formaldehyde will lead to increased eye blink frequency and subjective sensory irritation. Thermal effects on human performance in office environment

measured by [125], it indicated that there was a greater decrease in task performance due to thermal discomfort. The study [126] identified substantially elevated adverse health effects that come from overexposure to humidity have: fatigue, fainting, muscle cramps, and heat exhaustion.

Site selection. Different types of the workplace across continents were chosen to monitor their IAQ variations from a long-term perspective. These workplaces are relatively commonplace such as offices and factories, that have certain generality to understand the IAQ situation in common sense. The aim is to better manage IAQ in diverse workplaces as IAQ ties to people's comfort and health and has a huge impact on business perspective as well.

At each selected site, the developed IAQ monitoring system was used to record the IAQ situation with a frequency of 1 min. The data collection period is from one month to five months depending on the site availability and research plan schedule. The overall description of the study design is shown in Figure 2.4.

Figure 2.4: Study design: monitoring IAQ in diverse working conditions.

Data analysis. Two types of data analysis methods are used to extract meaningful data characteristics. The time series analysis method is employed to analyze the IAQ temporal characteristics together with standards and guidelines for OSH assessment. The OSH institutes such as NIOSH [35] and EPA [127] defined recommended exposure limit (REL) and Threshold Limit Value (TLV) for hundreds of workplace contaminants. These standard values are considered to assess the pollution level in a certain site.

Though the long-term monitoring process and time series analysis, IAQ conditions have significant variations from time to time due to changes in the working process, human activities,

and weather conditions. To fulfill legal compliance at the Occupational Safety and Health Act of 1970 (OSHAct), enterprises used to contract accredited inspectors to perform air-quality testing. The short-term sampling in the testing process would be possible to fail to catch the most effective sampling periods. To better select sample periods, a recommendation strategy was designed to give out reasonable time windows. The recommendation strategy is based on the continuous long-term measuring data and a voting strategy using a statistical analysis approach.

Managerial perspective. Indoor air quality is a major concern to businesses, managers, and employees as it impacts people’s health, comfort, productivity, and well-being. Studies conducted by EPA show that people spend up to 90% of their time indoors and indoor environments sometimes have higher levels of pollution than outdoors, occasionally, 100 times higher. In the US, productivity loss of 0.5 to 5 % per workplace, would bring 20 to 200 billion dollars loss annually.

Long-term IAQ monitoring ensures that the OSH dimension is given the same emphasis as other business objectives. Considering economic loss deriving from poor air quality, it is important to report indoor air quality daily and take it as one of the Key Performance Index (KPI) vectors. Moreover, the relevance of the KPIs increases when they are based on real-time measurements.

To quantify how a workplace affects productivity, creativity, and well-being, CBRE, a real-estate service and investment company, designed a science-based tool to measure specific criteria [128]. The related experiments proved that information awareness also affects employee performance. It shows if an employee understands they are working in a quality controlled environment, it is useful to improve their performance. The real-time long-term monitoring solution provides IAQ transparent empowerment to employees.

Both continuous collected IAQ raw data and statistical summary data are transmitted to IOTA’s Tangle through MAM or message. Different clients can register their interest to receive such data streams. According to relevant business models, different sharing mechanisms can be packaged to better serve all stakeholders. For example, the calculated avg pollutant value during each 15 min, can be uploaded and stored in the IOTA Tangle as the statistical summary data. Relevant accredited inspectors could assess this data in advance for better sampling period selections in OSH assessment.

The developed indoor air quality system can be also used to empower an indoor air quality audit management platform. The platform can help auditors and clients better understand the indoor air quality process, and save the cost of performing audits. It also supports to learn from historical experiences and suggest effective mitigation actions for ultimately improving indoor air quality condition. By bring healthier conditions inside buildings, the link among green procurement, sustainable building and housing, and energy-efficient management and control is actively fostered via the developed indoor air quality monitoring solution. A successful indoor air quality management practices for various building types and designs can also be accelerated

via the development of knowledge base powered by the proposed flexible, low-cost, and real-time IoT and DLT based air quality monitoring solution.

Meanwhile, there have been several trends in society that yield indoor air quality a priority to be addressed: economics growth press pressure to the environment, higher population densities, energy efficiency under the rise of air-conditioning; climate change; healing built-environment, cultural heritage protection, and cognitive improvement in school, new demanding construction materials. These all accentuate the indoor air quality research. Therefore, indoor air quality research is increasingly becoming an important focus in the scientific research community and practitioners.

2.3.2 Study design 2

Air pollution from both outdoor and indoor sources constitutes the greatest environmental risk to human health around the globe. About seven million premature deaths were attributed to air pollution in 2014, based on an estimate by the World Health Organization (WHO). It is projected that, by 2050, outdoor air pollution will be the number one cause of environment-related deaths worldwide. For these reasons, governments across the globe have started to monitor the levels of major air pollutants, especially in metropolitan and urban areas. Besides, as introduced aforementioned, indoor air pollution is equally detrimental, as statistics show that 4.3 million people per year die from exposure to household air pollution. Exposure to poor indoor air is a significant cause of productivity loss.

Monitoring Personal Air Pollution Exposure (PAPE), which refers to the amount of pollution being inhaled by an individual, has been a topic of growing interest worldwide, not only as a result of global health policies but more importantly, due to the interest in understanding its effects on various cardiovascular and respiratory diseases.

PAPE estimation. The study design used different sensors such as e-beacon and indoor pollution monitors and combined outdoor and indoor data with spatiotemporal activity patterns to estimate the Personal Air Pollution Exposure (PAPE) of an individual. The mobile-phone-based tracking app such as Moves [129] and OwnTracks provides the time and location data of the individual in the indoor and outdoor environment. For outdoor pollution data acquisition, respecting EU directive (Directive2008/50/EC) regarding the establishment of limits to major air pollutants, most large cities built their air quality monitoring stations. Take the example of Madrid, the air pollution data are open access and can be downloaded free of charge from the official website ⁴, the meteorological data ⁵ also like this.

For the indoor pollution monitors, the developed IAQ monitoring system or commercial Foobot indoor monitors can be used. Different indoor monitors can be placed in diverse micro-environments indoors if more precise indoor pollution exposure crossing micro-environments

⁴<https://www.madrid.es/portal/site/munimadrid>

⁵<https://opendata.aemet.es/centrodedescargas/inicio>

is needed to be estimated. Micro-environments (office, meeting room, and printing room) in the workplace could be replicated by deploying air monitoring devices in all available micro-environments. The indoor pollution data are minute-based measurements of PM, VOCs, CO₂, etc, which were retrieved from Foobot’s API. The e-beacon devices were placed in proximity to the indoor monitors. Together with the developed location tracking system, it helped us to determine whether the individual was within the indoor vicinity.

For location tracking, we used the developed Owntracks based location tracking system, in which time, location, and activity were recorded. The location tracking system is a low-cost solution, which uses a personal smartphone and beacons to track people indoors. Beacons are deployed in different regions in indoor space. The OwnTracks itself can also support outdoor positioning. The integrated functions were loaded into a docker container, enabling run-time on the container level. The location tracking system is functional to learn people’s locations, in both indoor and outdoor environments.

Data modeling. To model indoor pollution level, the indoor location tracking records people indoor moving area and timestamp while indoor air monitors stores the IAQ in each indoor area. The indoor modeling was done based on timestamp aggregation with e-beacon based indoor location tracking records and IAQ data from indoor air monitors. About outdoor modeling, geostatistical method [130], numerical-modeling based dispersion models [131] or big data, and machine learning-based models [132] can be used to estimate outdoor pollution exposure, with data collected from fixed-site monitor stations and people location tracking outdoors.

However, considering the computational costs and the number of potential users, we adopted some classical but still cost-effective approaches like the Kriging algorithms. Kriging methods estimate the value at a location by assigning a weight of the surrounding known values and calculating the weighted sum of the data. Interpolation technique e.g., Kriging, is a simple, intuitive, and fast method. Kriging surpasses others as it provides an interpolation error estimate, and it is an accurate interpolation.

The main data sources used to model indoor pollutions were the e-beacons and foobot indoor monitors. The timestamps recorded from the e-beacons provide the time when the individual was detected indoors. To obtain the corresponding pollution values during these periods, each of these timestamps was matched to the closest timestamp logged from the Foobot device. The pollution values were then aggregated in the indicated periods.

Ventilation measures the volume of gas inhaled by an individual. It varies with the type of activity. The type of activity or travel mode may have a significant effect on the exposure values. The types of activities in the tracking include “transport”, “walk”, “run”, and “cycle” and are based primarily on the speed of movement of the individual. Ventilation would be considered to the exposure calculation by multiplying a human inhalation rate.

Implications and discussion. This pilot case study presents evidence that PAPE can be estimated by employing indoor air quality monitors and e-beacon technology that have not

previously been used in similar studies and have the advantages of being low-cost and unobtrusive to the individual.

We did our best in providing a consistent perspective on how to model outdoor pollution, including the four major classes. Indeed, we provided a rather agnostic justification about the choice, including the potential added value due to data fusion, which can enable machine learning based models. Besides, potential limitations like roughness are discussed and some other limitations from other techniques like boundary conditions for Community Multi-scale Air Quality Model (CMAQ) are also discussed. We do believe we have argued to explain why the specific class of models has been chosen as well as their limitations. To this end, it makes sense to clarify that the contribution of the paper is not to promote the more accurate model but to provide an integrated approach, able to run dynamically and providing timely information to the users.

We must be aware of different uncertainty sources for both indoor and outdoor measurements. For sure, one of them is due to the outdoor modeling of pollution but it is not the largest one, and there is estimation for their accuracy. The paper does not claim that it proposes a full accurate estimation, but an estimation trying to promote a feasible and integrated solution to estimate the personal pollution exposure. Such exposure dosage is composed of different components, both from outdoor and indoor. Therefore, the uncertainty based on the outdoor model will impact in the same way to two different individuals being outdoor at the same point in time. Indeed, data fusion will contribute to increasing the precision specifically for interpolated based models. Finally, yet importantly, moderate uncertainty from the concentration of the pollution levels needs to be integrated through the exposition time according to the air ingestion rate (another source of uncertainty, depending on the physical activity being carried out). Because of those moderating factors, the total figure of uncertainty remains much less relevant than the aggregated exposition figure.

Variability across micro-environments in the same indoor environment has been investigated by numerical research work. They all demonstrated significant temporal and spatial variations in different micro-environments. Those micro-environments are like smoking and no-smoking area, underground and ground-level in the metro system, and office and photocopy. Besides, in households, although people spent less time in the kitchen than other rooms, the cooking preparation process takes place mostly using electric or gas cookers, or leaky stoves in developing countries such as China and India. Those could lead people to exposure to a high concentration of indoor pollutants than other rooms. Therefore, to differentiate micro-environments indoors, as shown in those applications mentioned above, has a significant implication for indoor pollution exposure assessment in terms of quality assurance and assessment accuracy.

We would like to point out that the current version is a conceptual paper in which a theoretical framework is promoted and a case study was made as a proof of concept to demonstrate the practical implications of the proposed IoT application. The research indicates the possible future

practical applications where our proposed framework and IoT application can be used. For instance, our IoT application can be extended to include prediction models, enabling dynamic feedback about PAPE risks. One important dimension that is possible to consider is the impact in terms of travel route decisions. According to the corresponding aggregated pollution and exposure values predicated, different routes that one individual can take when moving from one location to another. Greener and healthier routes can be selected. The proposed approach considers indoor pollution (waiting time inside subways for instance) when calculating the travel routes. Since indoor pollution data is currently not available in public transportation such as buses, trains, and subways, we were not able to fully test such an approach in our case study. Nevertheless, in our proposed framework, we acknowledged that indoor pollution is an integral part of estimating PAPE for such alternative travel routes.

We are aware of the limitations of including only 1 person in our case study. However, since this study is mainly interested in estimating the pollution exposure of an individual, concentrating on a single volunteer provides a fine-grained detail on the associated issues related to the technical feasibility of our proposed application. The authors [133] argue that "since we are interested in extracting routine behaviors of individual users we focused on long-term, fine-grain data acquisition from a few users rather than on coarse data from a large user population. Focusing on a single person over three months helped us identify the issues associated with smartphone momentary location tracking in very fine detail." Of course, in future work, we will try to increase the number of users so we can derive more robust conclusions about our proposed method.

Furthermore, PAPE data from this type of application could be useful for air quality policy development as well as in epidemiological studies that explore the effects of air pollution on certain diseases. Having a large scale air quality data, the benefit of a dense mesh of sensors becomes evident. Such a strategy is reported as being done in different cities around the world, such as Airbox in Taipei and Array of Things (AoT) sensor boxes in Chicago.

2.3.3 Study design 3

Recent advances in technology have empowered the widespread application of cyber-physical systems in manufacturing and fostered the Industry 4.0 paradigm. In the factories of the future, it is possible that all items, including operators, will be equipped with integrated communication and data processing capabilities.

In this context, a Healthy Operator 4.0(HO4.0) concept was proposed, aims to address the concerns regarding increasing workforce stress levels, the state of psychosocial health, and the new potential physical risks in Industry 4.0 paradigm. For the implementation of this relatively new concept, we constructed a unified architecture to support the integration of different enabling technologies. We designed an implementation model to facilitate the practical application of this

concept in the industry. Besides, a prototype system was developed, and relevant experiments were conducted to demonstrate the feasibility of the proposed system architecture.

The Health Operator (HO4.0) is presented as a virtual system: the HO4.0 digital twin, gathering all the health-related information from operators and potentially enabling the learning of rules from the different behaviors. The HO4.0 digital twin is powered by Industrial IoT networks, wearable technologies, ambient intelligence, and modeling technologies. It not only enables real-time health risk information, e.g., risk alert but also makes it possible to simulate the future behavior of operators; to infer and forecast the evolution from their behavior at mid/long term ranges.

Case selection. The case selected is connected with a logistic intra-facility in the steel manufacturing factory in Spain. The interest was in obtaining health-related knowledge of crane operators in their working environment, who was in charge of loading trucks that deliver manufactured steel rebar bundles to build concrete structures in the construction sector.

The case was selected due to different considerations. In the loading process, a lack of specific items could impact the delivery dates, and improper loading disposal sequence may negatively affect the scheduled work in the unloading process. The responsibility for such a decision relies on the crane operator, which may trigger their stressful situations. Furthermore, the crane operator was working in a rather complex environment. The production process in the plant includes cutting, blending, and welding may affect air quality, especially acoustic aspects. These complex environments may arise health threats to crane operators as well. Therefore, considering their working conditions, complex mobility schema, mental stressor, and high physical workload, these types of workers were decided to study subjects.

Subjects. Seven crane operators were involved in the experiment. They all maintain a healthy status during the experiment.

Experimental setup. At the site, the developed IAQ monitoring system was used to record the IAQ situation with a frequency of 1 min. Each crane operator wore a smartwatch, a positioning tag, and a MetaSensor during their working time. The developed location tracking system and the human physiological condition monitoring system can be employed for continuous data collection. The Ultra-wideband (UWB) technique based solution can also be considered for indoor positioning, as it demonstrates better performance compared with those BLE technique based [134]. The data collection period is from April 2019 to July 2019 depending on the schedule of the research plan. Some of the data collection starting time varies, that is because it took time to coordinate with some of the equipment suppliers. The overall description of the study design is shown in Figure 2.5.

Data analysis. When the interest was to analyze the similarity of the operating conditions and operator behavior through time, the application for building projectors with general non-linear dimension reduction was needed. The collected data were from multiple data sources. These data were fused into a 12-dimensional space according to time series. The nonlinear dimen-

Figure 2.5: Study design: healthy operator in industrial scenario.

sion reduction algorithms and tools such as Principal component analysis(PCA), T-distributed Stochastic Neighbor Embedding (t-SNE), and Uniform Manifold Approximation and Projection (UMAP) dimension reduction techniques can be used for data projection. They can optimize a low-dimensional graph from twelve dimensions into two, to keep close records that are close in a higher-dimensional space, and that preserved more of the global structure of the whole data set.

As several variables are associated with the operator's health, they are possibly from working conditions, mental status, or physically demanding. To understand human behaviors, e.g. higher blood pressure appeared in a specific situation, different machine learning techniques such as association rules can help explain the causes for such behaviors.

To understand the behavior of operators over time, different regression algorithms were applied to predicate future trends. The forecast method is either by using the time series approach or the multidimensional one. To identify similar behavior between operators, clustering algorithms were useful to learn psychological and physical health impacts over a group of people.

Implications. The implementation of a prototype to validate the HO4.0 framework was performed successfully, the layer structure helped in providing guidance to the required discussion to build the system. Data interconnection and interoperation further facilitated health action decisions making, aimed to improve the operator working conditions, where the actions were built on semantic-based data driver strategies.

The HO4.0 concept is toward a human-centric perspective on the fourth industrial revolution technologies. People as a core part of this large shift towards smart manufacturing or smart

factories. HO4.0 as a subtype of Operator 4.0 is capable to assist human workers in creating different types of healthy enhancements, from health monitoring that warn of risks, physical support to avoid anti-ergonomic postures and movements, to manage increased stress from working processes. Of course, the question of privacy should be properly addressed as health-related information is extremely sensitive. The ongoing DLT and its relevant technological advancements would be a suitable tool to manage health-related information reasonably.

It is possible to combine the organizational analysis based on patterns with the additional value created from the operator level, by allowing them to be aware of their values. This is another useful dimension of the HO4.0 concept, as awareness is a key element for empowerment. Useful discussion about the correlation between health impact with the working process is enabled among workers, as workers themselves can access their historical data from a built web-based platform.

2.3.4 Study design 4

Case selection. The study was undertaken in a controlled office in Escuela Técnica Superior de Ingenieros Industriales (ETSII), Universidad Politécnica de Madrid (UPM). The selected site was with dissatisfactory IAQ which was assessed via the developed long-term IAQ monitoring system.

Subjects. To maintain population generality, 12 healthy adults with different profiles and backgrounds participated in the study. They were invited from the ETSII campus or FOAF network.

Experimental setup. Each participant was required to visit the office once per day. The experimental procedure includes 3 min questionnaires, 10 min cognitive test, 1 min productivity test, 3 min creativity test, and another 3 min questionnaires.

The questionnaires compose of workplace subjective assessment (e.g., Subjective Assessment of Workplace Performance [135]), participant demographic information, fatigue status questionnaire (e.g., Chronic Fatigue Syndrome Attitude Test Questions [136]), and workload assessment (e.g., NASA TLX [137]).

The human activity was simulated via cognitive test, productivity test, and creativity test. The productivity ability is measured by typing (information processing work), designed by the Cambridge University Brain Science lab. The creativity is measured by the Taxonomy of Creative Thinking Method [138], that the simulation task is mind map (knowledge creation work) [139]. Cognitive performance is assessed by visual reaction time task, Stroop task, and visual backward digit span task provided by psytoolkit [140]. Meanwhile, the human response was measured by various wearable devices, where the focus was measured by smart glass manufactured by JINS MEME, the heart rate and blood pressure were measured by H Band based human physiological condition monitoring system.

One of the inevitable consequences of rapid urbanization is the compromised physiological and psychological health of urban dwellers as a result of limited access to nature and thus the accumulation of urban stress.

The biophilia hypothesis also named BET, which suggests that humans possess an innate tendency to seek connections with nature and other forms of life. The hypothesis was introduced by Edward O. Wilson in his book, *Biophilia* (1984) and became well-known and popularized. Biophilia is defined by him as "the urge to affiliate with other forms of life". Based on this hypothesis, emerging scientific evidence has shown that exposure to nature has significant physiological and psychological benefits, including stress recovery, improvement of cognitive functions, and an increase in positive affect.

The study design explored the biophilic effects on people's performance [141, 142]. Selected NASA plants were applied as biophilic design [141]. Hypothesis testing method: Paired t-test and Wilcoxon signed-rank test were used to test the significant differences between non-biophilic and biophilic scenarios.

Data analysis. The different statistical methods can be used to analyze similarities or differences in people's behaviors at the site. Besides, to understand physiological and psychological impact over the people's performance, a diverse machine learning method can be employed, especially when the data set reaches a certain scale.

Impact. The research results potentially have several benefits. It proved that dissatisfying working environments pose an adverse effect on people's performance, with a quantified and objective manner. Plus, the biophilic hypothesis was further assessed in the study via quantified biophilic design. If our research hypothesis is well verified, quantified biophilic design can be used to improve working environments, enhance economic benefits by improving worker productivity and performance, even in dissatisfying IAQ environments. It will bring tremendous benefits for people working or living in a dissatisfying air quality environment, especially in developing countries. Besides performance improvement, biophilic design is also low-cost, energy-saving, sustainable for all related stakeholders.

The research results will be applied in CBRE's clients, when applicable. The whole research study is a pilot experiment on a small scale and in preparation for our proposed global scale research project.

Chapter 3

Collection of publications

The main results have been published in eleven journal/conference articles shown in Table 3.1. There are seven journal articles and one conference article already published that systematically display the primary research outcomes, while the rest of them are to be put out soon. The major seven publications included in the collection are listing in this chapter beneath the table. The included articles in the chapter are not an infringement on copyright. Articles are licensed under open access, and the copyright is retained by the authors.

Table 3.1: List of publications from the study.

No.	5-year Impact Factor	Title	DOI
1	3.032 (Q1) in 'Instruments & Instrumentation'	<i>Indoor Air-Quality Data-Monitoring System: Long-Term Monitoring Benefits</i> [36]	10.3390/s19194157
2	2.948 (Q1) in 'Public, Environmental & Occupational Health'	<i>The Use of the Internet of Things for Estimating Personal Pollution Exposure</i> [121]	10.3390/ijerph16173130
3	3.032 (Q1) in 'Instruments & Instrumentation'	<i>Non-Invasive Ambient Intelligence in Real Life: Dealing with Noisy Patterns to Help Older People</i> [143]	10.3390/s19143113
4	4.924 (Q1) in 'digital health'	<i>Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies</i> [111]	10.2196/13583
5	3.032 (Q1) in 'Instruments & Instrumentation'	<i>Geometric Deep Lean Learning: Deep Learning in Industry 4.0 Cyber-Physical Complex Networks</i> [144]	10.3390/s20030763
6	3.032 (Q1) in 'Instruments & Instrumentation'	<i>Healthy Operator 4.0: A Human Cyber-Physical System Architecture for Smart Workplaces</i> [66]	10.3390/s20072011
7	3.032 (Q1) in 'Instruments & Instrumentation'	<i>Data Handling in Industry 4.0: Interoperability Based on Distributed Ledger Technology</i> [145]	10.3390/s20113046
8	–	<i>Biophilic Adaptation at the Workplace Multidimensional Impact on Human's Behavior</i> [146]	ISES-ISIAQ2019
9	–	<i>Applying Distributed Ledger Technology to Facilitate IIoT Data Exchange: an Approach based on IOTA Tangle</i> [147]	IESA2020-accepted
10	–	<i>Decentralized Industrial IoT Data Management Based on Blockchain and IPFS</i> [148]	APMS2020-accepted
11	8.763 (Q1) in 'Environmental Sciences'	<i>Physiological and Psychological Responses to Transitions Between Urban and Urban Nature Environments Using Immersive Visualization</i> [149]	Environment International-under review

3.1 Publication 1

- **Title:** *Indoor Air-Quality Data-Monitoring System: Long-Term Monitoring Benefits*
- **Authors:** Shengjing Sun, Xiaochen Zheng, Javier Villalba-Díez, and Joaquín Ordieres-Meré
- **Publication date:** 2019/9/25
- **Journal:** *Sensors*
3.031 (2018) ; 5-Year Impact Factor: 3.302 (2018)
Quartile: Q1 in ‘Instruments & Instrumentation’
ISSN: 1424-8220
Publisher: Multidisciplinary Digital Publishing Institute, BASEL, SWITZERLAND
- **Volume:** 19
- **Issue:** 19
- **Pages:** 4157
- **DOI:** <https://doi.org/10.3390/s19194157>

Article

Indoor Air-Quality Data-Monitoring System: Long-Term Monitoring Benefits

Shengjing Sun ^{1,2,*} , Xiaochen Zheng ^{1,3} , Javier Villalba-Díez ^{4,5} and Joaquín Ordieres-Meré ¹

¹ ETSII, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain; xiaochen.zheng@epfl.ch (X.Z.); j.ordieres@upm.es (J.O.-M.)

² Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

³ School of Engineering, Institute of Mechanical Engineering, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland

⁴ Fakultät fuer Management und Vertrieb, Campus Schwäbisch-Hall, Hochschule Heilbronn, 74081 Heilbronn, Germany; javier.villalba-diez@hs-heilbronn.de

⁵ ETSI Informática, Department of Artificial Intelligence, Universidad Politécnica de Madrid, Calle de los Ciruelos, Boadilla del Monte, 28660 Madrid, Spain

* Correspondence: shengjing.sun@alumnos.upm.es or shengjingsun@hsph.harvard.edu

Received: 25 August 2019; Accepted: 20 September 2019; Published: 25 September 2019

Abstract: Indoor air pollution has been ranked among the top five environmental risks to public health. Indoor Air Quality (IAQ) is proven to have significant impacts on people's comfort, health, and performance. Through a systematic literature review in the area of IAQ, two gaps have been identified by this study: short-term monitoring bias and IAQ data-monitoring solution challenges. The study addresses those gaps by proposing an Internet of Things (IoT) and Distributed Ledger Technologies (DLT)-based IAQ data-monitoring system. The developed data-monitoring solution allows for the possibility of low-cost, long-term, real-time, and summarized IAQ information benefiting all stakeholders contributing to define a rich context for Industry 4.0. The solution helps the penetration of Industrial Internet of Things (IIoT)-based monitoring strategies in the specific case of Occupational Safety Health (OSH). The study discussed the corresponding benefits OSH regulation, IAQ managerial, and transparency perspectives based on two case studies conducted in Spain.

Keywords: long-term monitoring benefits; indoor air quality; low cost; occupational safety and health; industry 4.0; IOTA tangle

1. Introduction

Indoor air pollution is a leading environmental risk, which affects people's working performance, comfort, health, and well-being [1–3]. People spend around 90% of their time indoors, and human exposure to indoor air pollutants may occasionally be more than 100 times higher than outdoor pollutant levels [4]. Exposure to poor indoor air is a significant cause of productivity loss, for the U.S., as productivity decreases 0.5–5% per workplace, generating a loss of 20 to 200 billion US dollars annually [5].

Indeed, exposure to poor indoor air also increases numerous adverse health problems, such as nausea, headaches, skin irritation, sick building syndrome, kidney failure, and even cancer [1,3,6]. The World Health Organization (WHO) estimates that 700,000 people per year die from poor breathing conditions [7]. Therefore, IAQ, has a significant impact on people's comfort, health, and performance.

IAQ have been investigated by research community and practitioners. However, two important gaps are identified by a systematic literature review. Details of literature review method and results are presented in Appendix A. The first gap is that long-term data-monitoring is lacking in the IAQ research community and in practice. The second gap is that IAQ data monitoring solution is hard to operate, expensive, and lacks transparency in terms of Occupational Safety Health (OSH) management.

Gap I. In most research work, sampling time period is relatively short term, and even sometimes lasts for less than one hour. Measurements were made based on assumption the IAQ conditions are the same during the year [3,8,9]. Several studies have shown that seasonal variation or non-heating and heating time period differences [10–12], considering IAQ variation by season climate change and human heating behavior. However, those studies only took a short time sampling each season. As IAQ varies from time to time due to changes in working conditions, human activity, and weather conditions etc., short-term or seasonal sampling could not cover all kinds of variations. Therefore, long-term monitoring becomes a need in the research community.

In addition to the research dimension, the practical applications, it is increasingly becoming important to gather indoor working conditions to evaluate and minimize adverse health problems. To ensure OSH, environmental regulatory agencies, e.g., Occupational Safety and Health Administrations (OSHA) and local authorities have developed monitoring strategies to assess employee exposure to indoor pollutants. Generally, the monitoring reference period is a short-term basis. A Short-Term Exposure Limit (STEL) is a term used in occupational health, industrial hygiene, and toxicology, and it is regularly adopted to be 15 min.

Long-term monitoring is challenging for environmental regulatory agencies considering instrument calibration, labor time, and costs. In regular monitoring strategies, the professional determination method is used to measure indoor contaminant level. For measuring Volatile Organic Components (VOCs), is to collect air samples, either based on whole air samples in SUMMA passivated stainless steel canisters or on solid adsorbent tubes. Subsequently, the VOCs are separated by gas chromatography and measured by mass-selective detector or multidetector techniques in a remote laboratory basis [13,14]. Due to obstacles in sampling pump technology, professional particulate matter meters such as personal environmental monitor (PEM) with Leland Legacy pump and sioutas personal cascade impactor sampler, only allow maximal 24-hour sampling [15]. Therefore, although those determination methods are accurate, in practice, it will not be possible to have a long-term monitoring strategy with professional instruments.

There are specific measurements considering different episodes, such as an eight-hour working period, night and day cycles, and seasonal variations. However, indoor working conditions vary greatly over time, and on spot short-term sampling or measurement in specific episodes could fail to provide a holistic assessment of the working environment. Long-term monitoring becomes a demand in the workplace and real applications such as OSH management.

Gap II. IAQ data-monitoring solution challenges are presented into two aspects. One challenge is that IAQ data sampling/collection is expensive, complex to customize and operate, and professional expertise dependency. To measure multiple pollutants, various devices should be bought from different manufactures [3,10,12]. For example, Aeroqual 200 to measure nitrogen dioxide (NO₂) and Total Volatile Organic Components (TVOC); Extech VPC300 to measure particulate matter (PM); htV-M to measure formaldehyde (HCHO); Q-Trak to measure carbon monoxide (CO). In addition, diffusive or passive samplers should be prepared and analyzed by chemical domain experts, and frequently replaced with new ones due to limited equipment lifespan [8,11,12], which is complicated to operate and time and labor consuming. Moreover, since different pollutants are measured with different instruments, the various collected pollutant data is complex to manage and process. Another challenge is that IAQ data sampling/collection lacks data sharing in terms of IAQ transparency to all stakeholders. IAQ transparency is vital for managers to regulate working conditions, because IAQ affects employee's productivity, comfort, and health [2]. IAQ transparency is significant for employees to enhance

worker empowerment. However, few data-monitoring solutions consider IAQ data-sharing from a transparent point of view.

As the graphical abstract shows in Figure 1, the research presented in this study will address those gaps by proposing an IoT and DLT based IAQ data-monitoring system. The system is a light, low-cost, long-term solution enabled by blooming development of IoT and sensor technology, and it does enable customization because of users can freely choose pollutant sensors by a simple plug-in. It supports multiple pollutants data-monitoring, which will facilitate research community and practitioners for long-term and integrated multiple pollutants data-monitoring solutions. The main contribution from the paper is not only the hardware layer involved, although it provides a solution ahead of what it is possible to find in the market these days, mainly because of the provided flexibility but also because of the data released over a public tangle in both ways, consolidated summaries as well as stream data flow, which overcomes the private clouds that providers used to adopt. Such approach is hard to handle in Europe regarding the General Data Protection Regulation (GDPR).

Figure 1. Graphical Abstract. IAQ: Indoor Air Quality; MAM: Masked Authenticated Messaging; OSH: Occupational Safety Health.

The paper promotes an open view of interaction from different layers of application, which enables different kinds of relationships and actually becomes kind of referential framework. The promoted system, which fits into the framework, applies IOTA distributed ledger techniques (concept introduced in Section 2.1) to enable data sharing, which benefits all stakeholders. It provides flexibility and low-cost, real-time, and summarized IAQ information to managers, aligns IAQ transparency to worker empowerment, and enables other benefits related to improving environments in the workplace. The relevant benefits are discussed and presented by two case studies conducted in Spain, from perspectives of OSH assessment, regulating working conditions, OSH transparency and data sharing with IOTA distributed ledger techniques.

A significant point highlighted by this paper is that although the paradigm of Industry 4.0 is being widely adopted by many industries in different sectors, there is a lack of penetration of these technologies in the specific case of the OSH [16]. Actually, the contribution from the IIoT can be considered as very significant because it can help in changing the implemented monitoring strategies. A contribution in this line is essentially the way adopted in in this paper.

2. Materials and Methods

2.1. Framework

The framework of the proposed data-monitoring architecture is presented in Figure 2. It is composed of five layers, including the sensing layer, network layer, IOTA-based data storage & sharing layer, data analysis layer, and application layer. (i) The sensing layer is the perception layer, which is the lowest layer of the conventional architecture of IoT. It contains different sensors for sensing and gathering IAQ information; (ii) The network layer is responsible for connecting to other smart things, network devices, and servers. It also facilitates transmitting and processing sensor data; (iii) The storage & sharing layer is based on DLT, the IOTA Tangle in this case, which supports secure and tamper-resistant data storage and sharing. The raw IAQ data can be shared through IOTA Tangle and in return the data publisher can receive monetary or other types of benefits. Moreover, the IAQ data analysis results, e.g., recommendation strategies, can also be shared through IOTA Tangle to help enhance regulation assessment; (iv) The analysis layer receives data from previous layers and use certain data processing techniques and machine learning models, to extract information and knowledge or even further means of wisdom; (v) The application layer is responsible for delivering specific services to corresponding users. It defines various scenarios in which the IoT and DLT can be deployed, for example, information transparency, IAQ management and assessment, gaining economic benefits for data providers. The proposed framework can also address low-cost sensor accuracy issue, as it supports persistently data aligning with OSH regulatory assessment conducted by high reliable professional instrument. The main enabling technologies involved in this framework are introduced as follows:

- **IoT.** The term "Internet of things" was coined by Kevin Ashton of Procter & Gamble in 1999, when he viewed Radio-frequency identification (RFID) as essential to the IoT, allowing computers to manage all individual things (all existing things). Presently, the IoT concept is that the pervasive presence of a variety of things or objects—such as RFID tags, sensors, actuators, mobile phones, etc.
- **Low-cost IoT-based sensing.** The low-cost IoT sensors enable the use of wireless communications and computing for interacting with the physical world. The relevant sensors could sense indoor environmental parameters such as IAQ, comfort, lighting, and acoustic conditions. Several systems [17–19] have been developed for monitoring indoor environmental conditions with low-cost sensors. The data quality generated by these sensors are often of questionable. The performance of different low-cost air-quality sensors vary from unit to unit, spatially and temporally, as it relies on different algorithms, the meteorological conditions and atmospheric composition [20]. The IAQ data-monitoring platform implemented in this study is low-cost sensor-based considering that high accuracy is not the top requirement for the targeted applications of this study. Instead, this platform is developed for purposes such as awareness raising and recommendation of sampling period selection for OSH legal compliance, which only demand the pollution level on a coarse scale. In addition, as shown in Figure 2, the accuracy of the proposed platform can be improved through data adjustment with professional instrument at each OSH regulatory spot-check in long periods, just by observing potential bias or sensor saturation.
- **Network.** The network e.g., IoT gateway, bridges sensor networks with the traditional communication networks. It settles the heterogeneity between various sensor networks, mobile communication networks, and the Internet (all computer networks) [21,22]. A single-board computer (SBC), such as Raspberry Pi, could provide low-cost and efficient gateway services based on emerging IoT standards.
- **DLT.** Blockchain, as the first DLT, was invented by Satoshi Nakamoto in 2008 to serve as the public transaction ledger of the cryptocurrency Bitcoin [23]. The main component of DLT is a distributed ledger, which is used as a distributed database maintained by a consensus protocol run by nodes in a peer-to-peer network. This consensus protocol replaces a central administrator, since all peers contribute to maintaining the integrity of the database [24]. With a

decentralized and consensus-driven nature, DLT could provide reliable solutions, such as blockchain [23], Ethereum [25] and IOTA Tangle [26], to enable secure and tamper-resistant data storage and sharing.

- **IOTA and the Tangle.** IOTA is a tangle-based cryptocurrency designed specifically for the IoT industry. The IOTA tangle naturally succeeds the blockchain as its next evolutionary step by overcoming some of its fundamental limitations, such as scalability, high transaction fees, and vulnerability to quantum attack [26]. The main feature of the tangle is that it uses a Directed Acyclic Graph (DAG) for storing transactions instead of sequential blocks. In the Tangle, users need to perform a small amount of computational work to approve two previous transactions to issue a new transaction. This new transaction will be validated by subsequent transactions [27].
- **Masked Authenticated Messaging (MAM).** The main data communication protocol used for data sharing in IOTA is MAM. It enables clients to emit and access an encrypted data streams over the IOTA Tangle, regardless of the size or cost of a device [28]. MAM uses channels (Public/Private/Restricted) for message spreading. IOTA users can create a channel and publish a message of any size, at any time. A small amount of proof-of-work is required to allow the data to propagate through the network and to prevent spamming. Other users can subscribe to this channel through its address, and receive a message that is published by the channel owner.

Figure 2. IAQ data-monitoring application framework supported by Internet of Things (IoT) and Distributed Ledger Technology (DLT). IAQ: Indoor Air Quality; MAM: Masked Authenticated Messaging; OSH: Occupational Safety Health.

The framework is also a functional guide following presented research work, for better understanding the IoT and DLT-based IAQ data-monitoring system and its long-term monitoring benefits.

2.2. Standards and Guidelines for OSH Assessment

The US National Institute for Occupational Safety and Health (NIOSH) defined a recommended exposure limit (REL) for hundreds of workplace chemical contaminants [29]. For NIOSH RELs,

a time-weighted average (TWA) concentration was measured for up to a 10 h workday during a 40 h workweek. They also define STEL as a TWA exposure that should not be exceeded at any time during a workday. For a ceiling REL, it is the ceiling value which should not be exceeded at any time. The STEL is as a legal limit for the exposure of an employee to a chemical substance. For chemicals, STEL assessments last for 15 min and are expressed in parts per million (ppm), or sometimes in milligrams per cubic meter (mg/m^3). Table 1 lists a couple of indoor chemical pollutants and their STEL, presented in NIOSH. We also list an averaging period of 24 h Threshold Limit Value (TLV) for PM referencing from EPA [30] and European Union (EU) air-quality standards [31], because NIOSH only provides TLV for a total particulate $10 \text{ mg}/\text{m}^3$ 8 h TWA.

Table 1. TLV for pollutant based on NIOSH, EPA and EU air-quality standards. STEL: Short-Term Exposure Limit.

Pollutant	STEL (15 min)	Average over 24 h
CO ₂	30,000 ppm ($54000 \text{ mg}/\text{m}^3$) STEL	
CO	200 ppm ($229 \text{ mg}/\text{m}^3$) ceiling	
Benzene	1 ppm ($3.2 \text{ mg}/\text{m}^3$) ceiling (15 min)	
Formaldehyde	0.1 ppm ($0.12 \text{ mg}/\text{m}^3$) ceiling (15 min)	
NO ₂	1 ppm ($0.18 \text{ mg}/\text{m}^3$) STEL	
O ₃	0.1 ppm ($0.2 \text{ mg}/\text{m}^3$) ceiling	
PM2.5		$50 \mu\text{g}/\text{m}^3$ (EPA)
PM10		$50 \mu\text{g}/\text{m}^3$ from EU air-quality standards

Considering the variations of workplace concentrations that originate from work patterns, processes (batch production or continuous process), human activity, and meteorological variations, several samples are required for the whole air-quality testing procedure. As shown in Table 2, the sampling duration and its related number of samples are presented, introduced by standard EU BS EN 689:1996.

Table 2. Minimum number of samples in relation to sampling duration:BS EN 689:1996.

Sampling Duration Time	Number of Samples
10 s	30
1 min	20
5 min	12
15 min	4
30 min	3
1 h	2
2 h	1

2.3. IoT and DLT-Based Data-Monitoring System

The research, unifying the proposed framework, has designed an IoT and DLT-based IAQ data-monitoring system (Figure 3), which is developed using kagoo devices manufactured by Circulate (Figure 3a,b), Raspberry Pi (Figure 3c), and IOTA Tangle (Figure 3d).

The kagoo devices (Figure 3a,b) were adopted as the sensing layer. Several indoor environmental conditions are measured with those devices, including air quality, acoustic conditions, lighting, and thermal comfort. Nine sensors, particulate matter (PM), formaldehyde (HCHO), TVOC, benzene (C_6H_6), carbon dioxide (CO_2), carbon monoxide (CO), ozone (O_3), nitrogen dioxide (NO_2), and T.H.I.N (Temperature & Humidity & Illumination & Noise), are used to measure indoor conditions. Those sensors can be freely selected and plugged into an island, where five maximal sensors are enabled. The list of mushroom sensors are presented in Table 3.

Figure 3. Deployment of Indoor Air-Quality (IAQ) data-monitoring system, enabling different type of clients. MAM: Masked Authenticated Messaging.

Table 3. Sensor specification.

No.	Sensor Name	Model	Functions	Range
1	PM	KG-PM2	PM2.5, PM10 Concentration Monitor	0–1000 µg/m ³
2	HCHO	KG-HO2	HCHO Concentration Monitor	0–7 mg/m ³
3	TVOC	KG-TV2	TVOC Concentration Monitor	0–3 mg/m ³
4	C ₆ H ₆	KG-C62	C ₆ H ₆ Concentration Monitor	0–320 mg/m ³
5	CO ₂	KG-C22	CO ₂ Concentration Monitor	0–0.5%
6	CO	KG-C12	CO Concentration Monitor	0–500 ppm
7	NO ₂	KG-N22	NO ₂ Concentration Monitor	0–20 ppm
8	O ₃	KG-O32	O ₃ Concentration Monitor	0–20 ppm
9	T.H.I.N	KG-TN2	Comfort Monitor (Temperature, humidity, illumination and noise)	T: −40–80°; H: 0–99.0% RH; I: 0–2000 Lux; N: 0–120 dB

The SBC Raspberry Pi (Figure 3c) was used as the network layer. A python program was run at Raspberry Pi to collect data measurements from sensors, parse data measurements, and calculate raw data measurement and statistic summaries. The raw data is collected with a frequency of 1 min. Usually, data measurements are stored every 1 min or 5 min in most IAQ monitoring devices [32,33]. The study took 1 min since smaller granularity data could be collected for data analysis. The relevant source code could be found on GitLab [34].

The continuous collected raw data is transmitted to IOTA Tangle through MAM communication protocol. The transmitted data will be broadcast in a streaming channel. Any IOTA user who knows the address of the channel (and the private key in case the channel is restricted) can consume the IAQ data.

The periodical statistic summary, such as 15 min (STEL basis) average indoor pollutant concentration, is uploaded and stored to IOTA Tangle as a message. Through different clients, all the users can review or use the real-time and long-term periodic statistic summary results, which could be applied for OSH regular assessment, IAQ management, and employee transparency for worker empowerment. For instance, when the organization needs to be OSH assessed, they can share with the regulatory authorities the data access from a certain time. Afterward, the regulatory agencies will be able to fetch data streams (e.g., statistic summary data report) for further assessment. Considering low-cost sensor's accuracy constrains, as the initial application, those data streams could be taken as pre-assessment to ensure spot-checking is conducted efficiently and effectively.

Indeed, larger populations could assess the data as scalability was considered to be a capability in the system.

2.4. Case Studies

Two case studies were conducted in Spain. The choice of location selection was mainly based on availability of sites; however, the location of each industrial building was carefully examined regarding the influence of the surroundings, urban/rural area, and green zone/heave traffic. Regarding their building structure, the selected locations were representative of the building stock of the country in terms of typology, construction techniques, and age. Site 1 is an advertising workshop located at a warehouse area next to a highway. Site 2 is a steel processing plant located at a rural area near a busy road, where the office section is monitored. The office represents the most common working environment and the advertising workshop represents a less common but with possible higher pollution environment. The site characteristics are listed in Table 4.

The chemical and physical parameters measured were PM_{2.5}, PM₁₀, HCHO, TVOC, C₆H₆, CO₂, CO, O₃, NO₂, temperature, humidity, illumination, and noise. The data-monitoring system was placed near worker activity area, with a height of 1.5 m above the ground. On Site 1, the data-monitoring period was from 15 October 2018 to 15 November 2018. On Site 2, the office section is from 7 December 2018 to 11 January 2019. 24 h with granularity 1-minute data measurements were collected inside two sites. All collected data are open access for researchers in OSF (<https://osf.io/t6rp8/>).

Table 4. Building site characteristics overview.

Characteristic	Site 1	Site 2
Section	workshop section	office section
Year of construction	35	46
Floor	1	1
Number of occupants	12	8
Total area (m ²)	200	100
Heating	No	Yes
Ventilation	Natural	Ventilation System
Windows	Single Glazing	Single Glazing
Floor covering	Coating	Coating
Facilities	One solvent printing machine, two caving machine, computers, furniture	Computers, furniture
Cleaning schedule	Once a week	Everyday
Working schedule	Flexible, 24 h, including weekends	Two shifts: 06:00–14:00; 14:00–22:00, only business days
Smoking	Yes	No
Nearby potential pollutant sources	No	No

3. Results and Discussion

3.1. Long-term Monitoring Benefits for OHS Assessment

The Occupational Safety and Health Act of 1970 (OSHAct) was passed to prevent workers from being harmed at work. The act created the OSHA, which enforces protective workplace safety and health standards. To fulfill legal compliance, employers used to contact accredited inspectors to perform air-quality testing. However, as presented in Table 2, some standards are defined to help regular short-term sampling. For example, the sampling duration time (15 min) is in relation to minimum 4 times samples established by statistical analysis and practical experience. However, it is still challenging for inspectors to select a sampling period. IAQ varies as working condition change such as different working process. As shown in Figure 4, heave printing work was done in afternoon in Site 1 leading to formaldehyde concentration rising and reaching to peak.

Figure 4. Time evolution of formaldehyde(HCHO) concentration in Site 1 during the day 2018-11-10. Dashed line refers to STEL threshold.

Practical research was conducted in Site 1, which has proved that regular spot checks are not sufficient to ensure employee health and safety. Suppose the inspector implemented a formaldehyde measuring path of 15 min sampling with stops of 30 min in between. If the inspector starts the testing by 09:00 a.m. considering that employee starts working at 08:00 a.m. the sampling period covers from 09:00 a.m. to 11:00 a.m. As depicted, formaldehyde evolution for the 10th of November in 2018 in Site 1, shown in Figure 4, there would not be any exceeding found. However, if the time span is expanded wider, exceedings appear continuously between 12:00 p.m. to 17:00 p.m. Therefore, it is proven that the inspector criteria, in full respect of regulation, still fails to guarantee employee health and safety. Although some surveys could be conducted, such as a review of work patterns, production processes, and exposure times that help select better sampling periods, they are always time-consuming and less accurate.

To address the sampling period selection bias, we designed a recommendation strategy which could give out a reasonable sampling period based on long-term measuring data and a statistical analysis approach. Considering all the historical data collected in the site for each indoor pollutant, a holistic statistical method was used to select better sampling periods. The recommendation strategy working flow is shown in Figure 5. The example of showing how the valid sampling duration for pollutant formaldehyde in Site 1 is selected based on the recommendation strategy is given. The data measurement from the whole data-monitoring period (15 October 2018–15 November 2018) are applied as input data. For each day in the monitoring period, all possible 15 min time intervals daily, with a time translation of 1 min are segmented, and the avg formaldehyde value during each segmentation is calculated. The 15 min segmentation with the max avg value is selected out in each day and the hour and two hours where the 15 min is located is marked and recorded, as shown in Figure 6b. Finally, based on a voting strategy of all monitoring days, those valid sampling durations could be recommended and selected out. Then, specific professional determination methods are used to measure indoor contaminant levels where the time windows are better selected.

Figure 5. Recommendation strategy flow diagram.

The recommendation strategy is evaluated with the case studies conducted in Site 1 and Site 2. Taking PM_{2.5} in Site 2 as an example, our recommendation framework enables us to demonstrate a distribution of PM_{2.5} maximal average value occurrences (sampling time: 15 min) in site. The detail is shown in Figure 6a. Each point representing the hour where the better 15 min is located in and each line segment representing the 15 min sampling duration covering 2 h, e.g., from 18:50 to 19:05. As shown in Figure 6a, most high PM_{2.5} values appear in the morning, before 08:00 a.m. Therefore, based on long-term measuring data analysis, for a regular PM_{2.5} check in Site 2, the sampling time period is recommended for the morning, before 08:00 a.m. As shown in Figure 6b, most of high formaldehyde values in Site 1 appear in the morning, between 08:00 a.m. to 12:00 p.m. Similarly, as shown in Figure 6c, the recommendation sampling time for benzene in Site 1 would be in the afternoon, around 14:00 p.m. to 18:00 p.m. With the recommendation strategy, exceeding and significant variation could be caught up to ensure employee's health and safety.

Figure 6. Distribution of indoor pollutant maximal average value occurrences (sampling time: 15 min) in site. (a) Distribution of PM_{2.5} maximal average value during monitoring period in Site 2. Dashed line refers to EPA TLV; (b) Distribution of HCHO maximal average value during monitoring period in Site 1. Dashed line refers to STEL threshold; (c) Distribution of C₆H₆ maximal average value during monitoring period in Site 1. Dashed line refers to STEL threshold.

3.2. Long-Term Monitoring Benefits for Regulating Working Conditions

The employer is legally responsible for ensuring the good working conditions of employees. Long-term IAQ monitoring enables employers to understand IAQ data and IAQ patterns, so they can take appropriate measures to regularly improve working conditions. As shown in Figure 7a, in Site 2, CO₂ concentration is constantly higher than TLV(0.1%). The TLV(0.1%) is based on a daily average, which obtained references from Circulate App: EnvCon [35]. The Circulate company set TLVs, taking references from China's air-quality standards. Measures should be taken to reduce CO₂ concentration in Site 2. The corresponding measures could be opening windows, adapting heating, ventilation, and air conditioning (HVAC) systems or with plants [36,37] to reduce CO₂ concentration.

Thermal comfort normally refers to temperature and humidity, and is the condition of mind that expresses satisfaction with the thermal environment and subjective evaluation (ANSI/ASHRAE Standard 55). The temperature and humidity ranges are 16–28° and 30–80% respectively, according to WHO. It is also important not to over-design illumination, which can induce adverse health effects such as headache frequency, stress, and increased blood pressure. For the work requiring perception of details, such as offices, sheet metal work, and bookbinding, the minimal illuminance is 100 lux, which is defined by EU standard. On Site 2, as shown in Figure 7b,c, displays very low humidity and illumination levels. Therefore, measures should be taken to improve employee comfort.

Figure 7. Working conditions demonstration in site 2. (a) Evolution of CO₂ hourly average concentration during monitoring period in site 2. Dashed line refers to Circulate TLV; (b) Evolution of humidity hourly average during monitoring period in site 2. Dashed line refers to WHO comfort recommendation; (c) Evolution of illumination hourly average during monitoring period in site 2. Dashed line refers to EU illuminance standard.

3.3. Long-Term Monitoring Benefits for OSH Transparency

Long-term IAQ monitoring ensures that OSH dimension is given the same emphasis as other business objectives. The economic benefits of high-quality indoor working conditions was demonstrated by [5]. The magnitude of productivity gains may be obtained by providing better indoor environments, a 20 to 50% reduction in sick building syndrome, saving between 10 and 100 billion US dollars. 8 to 25% fewer asthma-related absences save 1 to 4 billion US dollars. A 23 to 76% reduction in respiratory diseases saves 6 to 14 billion US dollars. The IAQ has a significant effect on economic benefits in enterprise management.

In building science studies, thermal comfort has been related to productivity and health. Office workers who are satisfied with their thermal environment are more productive [38]. For example, it is valuable to provide the correct light intensity for each task or environment. Otherwise, energy not only could be wasted, and over-illumination can lead to adverse health and psychological effects. Beyond the energy factors being considered, glare or excess light can decrease worker efficiency. The field of OSH comprises a variety of risks that need to be managed. Considering economic loss deriving from poor air quality, it is important to report IAQ daily and take it as a Key Performance Index (KPI) vector, along with other business objectives. Moreover, the relevance of the KPIs increases when they are based on real-time measurements.

To quantify how a workplace affects productivity, creativity, and well-being, CBRE, a real-estate services and investment company, designed a science-based tool to measure specific criteria [39]. The related experiments proved that information awareness also affects employee performance. That is to say, if an employee knows they are working in a good quality environment, it is helpful to improve their performance. The real-time long-term monitoring solution would provide IAQ transparent employee empowerment.

3.4. Long-Term Monitoring Benefits for Data Sharing by IOTA

Recently, with the tremendous development of Industry 4.0, DLT (e.g., IOTA) has attracted significant attention. With DLT, we will have the Internet of value. DLT has great potential to create new foundations for our economic and social systems by efficiently establishing trust among people and machines, reducing cost, and increasing the use of resources [40]. IOTA, as the most prominent distributed ledger project, whose goal is to become the very fundamental layer of such society, is challenging the looming paradigm shift.

Both continuous collected IAQ raw data and statistical summary data are transmitted to IOTA's Tangle through MAM or message. Different clients can register their interest to receive such data streams. According to relevant business models, different sharing mechanisms can be packaged to better serve all stakeholders [41]. For example, the calculated avg pollutant value during each 15 min, introduced in Section 3.1, can be uploaded and stored in the IOTA Tangle as the statistical summary data. Relevant accredited inspectors could assess this data in advance for better sampling period selections in OSH assessment.

The long-term IAQ monitoring solution could gain economic benefits for enterprises with the support of IOTA, with which enterprises could receive automatic, transparent, and frictionless payments from IAQ data consumers.

4. Conclusions

As proven by our study, long-term IAQ monitoring and data analytics have been lacking in the research community and in practice. Under this context, this research has contributed to address identified gaps by designing and testing a framework and a system, which is proven to be an effective solution for all stakeholder needs. For instance, through the monitoring system, managers can take measures to improve indoor working conditions. Some CO₂ exceeding values were found in Site 2. Managers could apply one basic green plant, an Areca palm (*Chrysalidocarpus lutescens*), which was

discovered by NASA to efficiently remove CO₂. In addition, this paper proves the enormous potential of the IIoT in the context of Industry 4.0 to contribute in bringing new insight in worker's environment, both by consistently monitoring such environment and to easily disseminate the collected information with minimum cost and infrastructure requirements. Besides aligning IOTA, a long-term monitoring solution provides continuous values to OSH assessment agencies, supporting IAQ transparency to employee empowerment, and bringing continuous economic value to enterprises through paid data sharing services.

The limitation of this research study is that precision could issue IoT based measurements, but the data-monitoring system could be re-adapted by the conducted professional spot-check by an accredited inspector, as indicated in Figure 2. At each spot-check, more accurate data measurements could be obtained with the professional instruments. The data-monitoring system would calibrate itself with those accurate measurement, therefore, in the long-term, the precision of the data-monitoring solution would be better improved.

The carried-out research results would be applied in CBRE in future work to assess IAQ influences on employee productivity, health, and well-being.

Author Contributions: S.S. has been responsible for the development of the case studies, and she has shared the efforts of implementation of the dissemination with X.Z. J.O.-M. contributed with S.S. and J.V.-D. to provide the industrial dimension context for OHS application. The IOTA selection has been responsibility of J.O.-M. and X.Z.

Funding: This research was partially funded by the European Commission through the RFCS program, grant number with ID 793505 (WISEST project).

Acknowledgments: The authors thank China Scholarship Council (CSC) (No.201206730038) for the support they provided for this work. Also, the authors want to thank the EU RFCS research program for the support of this research through the grant with ID 793505 (WISEST project).

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DLT	Distributed Ledger Technologies
GDPR	General Data Protection Regulation
IAQ	Indoor Air Quality
IoT	Internet of Things
IIoT	Industrial Internet of Things
OSH	Occupational Safety Health
TLV	Threshold Limit Value
TVOC	Total Volatile Organic Components

Appendix A

The literature review method is based on a guideline from [42]. The time span of the review is from Jan 1999 to Jan 2019. Only studies in English are considered. The IAQ monitoring and analytic are considered to be a whole, its sub-problems, such as the standalone monitoring system, exposure risk assessment, and micro-environment are not included in the literature review. The data sources include four digital libraries (ACM Digital Library, IEEE Xplore, Springer Link, and Science Direct), hand searching two conferences: the International Society of Indoor Air Quality and Climate (ISIAQ), the International Conference on Indoor Air Quality, Ventilation and Energy Conservation (IAQVEC) in Buildings. The search terms used are indoor (environmental/air), and quality (investigation/research/assessment/monitoring/analysis). After excluding irrelevant studies based on exclusion criteria and an analysis of their titles and abstracts, 65 studies were included based on full text screening, and 7 studies are selected from the reference lists. In the end, 72 studies were used as the final primary studies. Considering the objective of the research work is mainly from time and data-monitoring solution dimensions, the 23 most relevant studies were selected and listed

in Table A1, from pollutant (P), time period (T), seasonal (S), data collection (D), and transparency (Tr) aspects, respectively.

Table A1. Systematic literature review of IAQ.

Study	P	T	S	D	Tr
[3]	NO ₂ , TVOC, PM0.3-10	8h working time in five working days, September 2015	No	Aeroqual 200 (NO ₂ , TVOC), Extach VPC300 (PM0.3-10)	No
[10]	PM2.5, HCHO, CO ₂	PM2.5 and CO ₂ entire year, HCHO in four seasons (sampling time: 20 min)	Yes	A: on-line monitoring system with Ikaair (CO ₂) and Yun (PM2.5) sensors B: on-site measurement for HCHO by spectrophotometry	No
[43]	HCHO, CO ₂	4 h between 08:00 AM and 12:00 AM	No	A real-time occupational exposure monitoring system with Grove-HCHO and T6613C (CO ₂)	No
[8]	NO ₂ , O ₃ and 29 VOCs	One week between 20 and 27 December 2012	No	Diffusive samplers	No
[11]	34 VOCs, NO ₂ , O ₃	Summer: 24 and 28 May 2010; Winter: February 21 and 25, 2011	Yes	Passive samplers	No
[12]	Temperature, humidity, HCHO, C ₆ H ₆ , C ₂ HCl ₃ , Pinene, Limonene, NO ₂ , CO ₂ , CO, PM2.5, VOCs, Radon, O ₃	Monday to Friday, in both non-heating (26/09/2011-14/10/2011) and heating (23/01/2012-10/02/2012)	Yes	Diffusive samplers (HCHO, C ₆ H ₆ , C ₂ HCl ₃ , Pinene, Limonene, NO ₂ , O ₃); Telair 7001 (CO ₂), aeroQUAL (CO), PM2.5 (Derenda LVS3.1/PMS3.1-15)	No
[9]	PM2.5, PM10, CO ₂ , CO, HCHO, and VOCs, O ₃	1 h	No	Lighthouse handheld 3016 (PM, temperature, humidity), WolfSense (CO ₂ , CO, VOC and O ₃), htV-M (HCHO)	No
[44]	PAHs	One month in April	No	Passive sampler	No
[45]	VOCs, HCHO, acetone and O ₃	During 4 h with a 40-m frequency	No	PRO-EKOS AT. 401X (HCHO, O ₃), gas chromatograph Voyager (VOCs and acetone)	No
[46]	temperature, humidity, CO, CO ₂ , PM10, NO ₂ , HCHO, C ₆ H ₆ and toluene, bacteria and fungi	3–10 December	No	Passive bubblers (HCHO), passive bubbler (NO ₂), SKC passive sampler (VOCs)	No
[47]	PM, noise, temperature, humidity	May 2009 (hot season) and February 2010 (cold season)	Yes	–	No
[48]	Bacteria, fungi, dust, ammonia, and HCHO	2 h	No	Passive sampler	No
[49]	Eighteen PAHs	28 days (May–June 2014)	No	Passive sampler	No
[50]	PM	Pre-winter (November and early December 2013) and winter season (January and early February 2014)	Yes	MOUDI	No
[51]	17 VOCs	May 2015	No	Passive sampler	No
[52]	TVOC, 13 VOCs, PM2.5, NOx, O ₃	Two weeks (working and non-working days) which starts from early morning (08:00 a.m.) to late evening (20:00 p.m.) during winter season of 2014	No	Model EC 9810 series (O ₃), Model Ecotech Sernious 40 (NOx), Micro IV Single Gas Detector (CO), MiniVol™ TAS (PM2.5), PhoCheck 5000 photo-ionization detector (PID) (TVOC), NIOSH method (VOCs)	No
[53]	benzene, toluene, ethylbenzene m,p-xylene and o-xylene (BTEX)	Winter (from 9 December 2013 to 17 January 2014) and Spring (from 24 March to 17 April 2014)	Yes	Passive sampler	No
[54]	PM	Three weeks during the summer, autumn, and winter in 2014 and 2015	Yes	OPS; TSI model 3330	No
[55]	HCHO and C ₆ H ₆	45 min	No	Passive samplers	No
[56]	HCHO	Second semester of 2010 and first semester of 2011	No	Passive samplers	No
[57]	VOCs	24 h	No	Passive sampling	No
[58]	Temperature, humidity, fungi, dust, endotoxins, CHO, VOCs, CO ₂ , NO ₂	Two seasons: October–March; April–September	Yes	Radiello passive sampler (CHO and VOCs), Passam Ag passive sampler (NO ₂), Q-Trak (Temperature, humidity, CO ₂)	No
[59]	PM2.5, PM10	During rush hours (8:00 a.m.–12:00 p.m.) for one week per each season from June 2015–June 2016	Yes	Dust-Trak	No

References

- Ana, G.R.; Alli, A.S.; Uhiara, D.C.; Shendell, D.G. Indoor air quality and reported health symptoms among hair dressers in salons in Ibadan, Nigeria. *J. Chem. Health Saf.* **2019**, *26*, 23–30. [CrossRef]
- Tham, K.W. Indoor air quality and its effects on humans—A review of challenges and developments in the last 30 years. *Energy Build.* **2016**, *130*, 637–650. [CrossRef]
- Kiurski, J.S.; Aksentijević, S.M.; Mandarić, S.D. Statistical approach for characterization of photocopying indoor pollution. *Air Qual. Atmos. Health* **2018**, *11*, 867–881. [CrossRef]
- EPA. Why Indoor Air Quality is Important to Schools. Available online: <https://www.epa.gov/iaq-schools/why-indoor-air-quality-important-schools> (accessed on 24 September 2019).
- Fisk, W.J.; Rosenfeld, A.H. Estimates of Improved Productivity and Health from Better Indoor Environments. *Indoor Air* **1997**, *7*, 158–172. [CrossRef]
- Hadei, M.; Hopke, P.K.; Rafiee, M.; Rastkari, N.; Yarahmadi, M.; Kermani, M.; Shahsavani, A. Indoor and outdoor concentrations of BTEX and formaldehyde in Tehran, Iran: Effects of building characteristics and health risk assessment. *Environ. Sci. Pollut. Res.* **2018**, *25*, 27423–27437. [CrossRef] [PubMed]

7. WHO. Air Pollution and Health. Available online: <https://www.who.int/airpollution/en/> (accessed on 24 September 2019).
8. Can, E.; Üzmez, Ö.Ö.; Döğeroğlu, T.; Gaga, E.O. Indoor air quality assessment in painting and printmaking department of a fine arts faculty building. *Atmos. Pollut. Res.* **2015**, *6*, 1035–1045. [[CrossRef](#)]
9. Ramos, C.A.; Wolterbeek, H.T.; Almeida, S.M. Exposure to indoor air pollutants during physical activity in fitness centers. *Build. Environ.* **2014**, *82*, 349–360. [[CrossRef](#)]
10. Liu, J.; Dai, X.; Li, X.; Jia, S.; Pei, J.; Sun, Y.; Lai, D.; Shen, X.; Sun, H.; Yin, H.; et al. Indoor air quality and occupants' ventilation habits in China: Seasonal measurement and long-term monitoring. *Build. Environ.* **2018**, *142*, 119–129. [[CrossRef](#)]
11. Yurdakul, S.; Civan, M.; Özden, Ö.; Gaga, E.; Döğeroğlu, T.; Tuncel, G. Spatial variation of VOCs and inorganic pollutants in a university building. *Atmos. Pollut. Res.* **2017**, *8*, 1–12. [[CrossRef](#)]
12. Kalimeri, K.K.; Saraga, D.E.; Lazaridis, V.D.; Legkas, N.A.; Missia, D.A.; Tolis, E.I.; Bartzis, J.G. Indoor air quality investigation of the school environment and estimated health risks: Two-season measurements in primary schools in Kozani, Greece. *Atmos. Pollut. Res.* **2016**, *7*, 1128–1142. [[CrossRef](#)]
13. EPA. *Compendium of Methods for the Determination of Air Pollutants in Indoor Air*; Project Summary; EPA: Washington, DC, USA, 1990.
14. Ishizaka, T.D.; Kawashima, A.; Hishida, N.; Hamada, N. Measurement of total volatile organic compound (TVOC) in indoor air using passive solvent extraction method. *Air Qual. Atmos. Health* **2019**, *12*, 173–187. [[CrossRef](#)]
15. SKC. IP-10A Method Update. Available online: <https://www.skinc.com/catalog/pdf/instructions/1660.pdf> (accessed on 24 September 2019).
16. Mardonova, M.; Choi, Y. Toward Open-Source Hardware and Software for the Mining Industry: A Case Study of Low-Cost Environmental Monitoring System for Non-Metallic Underground Mines. *Min. Metall. Explor.* **2019**, *36*, 657–674. [[CrossRef](#)]
17. Zanni, S.; Lalli, F.; Foschi, E.; Bonoli, A.; Mantecchini, L. Indoor Air Quality Real-Time Monitoring in Airport Terminal Areas: An Opportunity for Sustainable Management of Micro-Climatic Parameters. *Sensors* **2018**, *18*, 3798. [[CrossRef](#)] [[PubMed](#)]
18. Zakaria, N.A.; Abidin, Z.Z.; Harum, N.; Hau, L.C.; Ali, N.S.; Jafar, F.A. Wireless Internet of Things-Based Air Quality Device for Smart Pollution Monitoring. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*. [[CrossRef](#)]
19. Benammar, M.; Abdaoui, A.; Ahmad, S.H.; Touati, F.; Kadri, A. A Modular IoT Platform for Real-Time Indoor Air Quality Monitoring. *Sensors* **2018**, *18*, 581. [[CrossRef](#)] [[PubMed](#)]
20. Castell, N.; Dauge, F.R.; Schneider, P.; Vogt, M.; Lerner, U.; Fishbain, B.; Broday, D.; Bartonova, A. Can commercial low-cost sensor platforms contribute to air quality monitoring and exposure estimates? *Environ. Int.* **2017**, *99*, 293–302. [[CrossRef](#)] [[PubMed](#)]
21. Zhu, Q.; Wang, R.; Chen, Q.; Liu, Y.; Qin, W. Iot gateway: Bridging wireless sensor networks into internet of things. In Proceedings of the 8th International Conference on Embedded and Ubiquitous Computing (EUC), Hong Kong, China, 11–13 December 2010; pp. 347–352.
22. Chen, H.; Jia, X.; Li, H. A brief introduction to IoT gateway. In Proceedings of the ET International Conference on Communication Technology and Application (ICCTA 2011), Beijing, China, 14–16 October 2011; pp. 610–613.
23. Nakamoto, S. Bitcoin: A Peer-to-Peer Electronic Cash System. Available online: <https://bitcoin.org/bitcoin.pdf> (accessed on 24 September 2019).
24. Brogan, J.; Baskaran, I.; Ramachandran, N. Authenticating health activity data using distributed ledger technologies. *Comput. Struct. Biotechnol. J.* **2018**, *16*, 257–266. [[CrossRef](#)] [[PubMed](#)]
25. Ethereum Foundation. Ethereum Blockchain App Platform. Available online: <https://www.ethereum.org/> (accessed on 24 September 2019).
26. IOTA. What Is IOTA? Available online: <https://www.iota.org/get-started/what-is-iota> (accessed on 24 September 2019).
27. Popov, S.; Saa, O.; Finardi, P. Equilibria in the Tangle. *Comput. Ind. Eng.* **2019**, *136*, 160–172.
28. Handy, P. *Introducing Masked Authenticated Messaging*; IOTA: Berlin, Germany, 2017.
29. NIOSH. *NIOSH Pocket Guide to Chemical Hazards*; U.S. Department of Health Human Services: Atlanta, GA, USA, 2016.

30. EPA. EPA Air Quality Standards. Available online: <https://www.epa.gov/criteria-air-pollutants/naaqs-table> (accessed on 24 September 2019).
31. EEA. European Union Air Quality Standards. 2019. Available online: <https://ec.europa.eu/environment/air/quality/standards.htm> (accessed on 24 September 2019).
32. Moreno-Rangel, A.; Sharpe, T.; Musau, F.; McGill, G. Field evaluation of a low-cost indoor air quality monitor to quantify exposure to pollutants in residential environments. *J. Sens. Sens. Syst.* **2018**, *7*, 373–388. [CrossRef]
33. AirVisual. How Often Does the AirVisual Pro's Sensor Take Measurements? What Is the Difference between "Continuous" and "Default" Mode? Available online: <https://support.airvisual.com/en/articles/3029368-how-often-does-the-airvisual-pro-s-sensor-take-measurements-what-is-the-difference-between-continuous-and-default-mode> (accessed on 24 September 2019).
34. Sun, S. Indoor Air Quality. Available online: <https://gitlab.com/sunshengjing/iaq> (accessed on 24 September 2019).
35. Circulate. EnvCon App. Available online: <https://apps.apple.com/ca/app/envcon/id1045648840> (accessed on 24 September 2019).
36. Gubb, C.; Blanus, T.; Griffiths, A.; Pfrang, C. Can houseplants improve indoor air quality by removing CO₂ and increasing relative humidity? *Air Qual. Atmos. Health* **2018**, *11*, 1191–1201. [CrossRef]
37. Meattle, K. Green Buildings: A Step Towards Sustainable Development. Available online: <https://www.nbmcw.com/tech-articles/tall-construction/30066-green-buildings-a-step-towards-sustainable-development.html> (accessed on 24 September 2019).
38. Sulewska, M.J.; Gładyszewska-Fiedoruk, K.; Sztulc, P. Analysis of the results of empirical research and surveys of perceived indoor temperature depending on gender and seasons. *Environ. Sci. Pollut. Res.* **2018**, *25*, 31205–31218. [CrossRef] [PubMed]
39. CBRE. Spain Occupier Report. Available online: <https://www.cbre.es/en/services/business-lines/global-workplace-solutions/workplace-strategy/cbre-lab> (accessed on 24 September 2019).
40. Yu, F.R.; Liu, J.; He, Y.; Si, P.; Zhang, Y. Virtualization for Distributed Ledger Technology (vDLT). *IEEE Access* **2018**, *6*, 25019–25028. [CrossRef]
41. Román, V.; Ordieres-Meré, J. [WiP] IoT Blockchain Technologies for Smart Sensors Based on Raspberry Pi. In Proceedings of the 11th Conference on Service-Oriented Computing and Applications (SOCA), Paris, France, 20–22 November 2018; pp. 216–220. [CrossRef]
42. Kitchenham, B.A.; Charters, S. *Guidelines for performing Systematic Literature Reviews in Software Engineering*; Technical Report EBSE-2007-01; Keele University: Keele, UK, 2007.
43. Fathallah, H.; Lecuire, V.; Rondeau, E.; Le Calvé, S. Development of an IoT-based system for real time occupational exposure monitoring. In Proceedings of the Tenth International Conference on Systems and Networks Communications, Barcelona, Spain, 15–20 November 2015.
44. Maragkidou, A.; Arar, S.; Al-Hunaiti, A.; Ma, Y.; Harrad, S.; Jaghbeir, O.; Faouri, D.; Hämeri, K.; Hussein, T. Occupational health risk assessment and exposure to floor dust PAHs inside an educational building. *Sci. Total Environ.* **2017**, *579*, 1050–1056. [CrossRef]
45. Kiurski, J.S.; Marić, B.B.; Aksentijević, S.M.; Oros, I.B.; Kecić, V.S.; Kovačević, I.M. Indoor air quality investigation from screen printing industry. *Renew. Sustain. Energy Rev.* **2013**, *28*, 224–231. [CrossRef]
46. Chiang, C.M.; Su, H.J.; Lai, C.M.; Chou, P.C.; Li, Y.Y.; Tu, Y.F. Indoor Air Quality Assessment in an Office Building in an Industrial Park in Taiwan. In Proceedings of the Asia-Pacific Conference on the Built Environment, Taipei, Taiwan, 1999. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.614.2821&rep=rep1&type=pdf> (accessed on 24 September 2019).
47. Loupa, G. Case study. Health hazards of automotive repair mechanics: Thermal and lighting comfort, particulate matter and noise. *J. Occup. Environ. Hyg.* **2013**, *10*, 35–46. [CrossRef] [PubMed]
48. Abdel Hameed, A.A.; Khoder, M.I.; Farag, S.A. Organic dust and gaseous contaminants at wood working shops. *J. Environ. Monit.* **2000**, *2*, 73–76. [CrossRef] [PubMed]
49. Keskin, S.S.; Sinan Keskin, S.; Dilmac, E. Indoor air particulate matter exposure of commuter bus passengers in Istanbul, Turkey. *Indoor Built Environ.* **2016**, *26*, 337–346. [CrossRef]
50. Goel, A.; Wathore, R.; Chakraborty, T.; Agrawal, M. Characteristics of Exposure to Particles due to Incense Burning inside Temples in Kanpur, India. *Aerosol Air Qual. Res.* **2017**, *17*, 608–615. [CrossRef]

51. Dai, H.; Jing, S.; Wang, H.; Ma, Y.; Li, L.; Song, W.; Kan, H. VOC characteristics and inhalation health risks in newly renovated residences in Shanghai, China. *Sci. Total Environ.* **2017**, *577*, 73–83. [[CrossRef](#)]
52. Singh, D.; Kumar, A.; Kumar, K.; Singh, B.; Mina, U.; Singh, B.B.; Jain, V.K. Statistical modeling of O₃, NO_x, CO, PM_{2.5}, VOCs and noise levels in commercial complex and associated health risk assessment in an academic institution. *Sci. Total Environ.* **2016**, *572*, 586–594.
53. Mainka, A.; Kozielska, B. Assessment of the BTEX concentrations and health risk in urban nursery schools in Gliwice, Poland. *AIMS Environ. Sci.* **2016**, *3*, 858–870. [[CrossRef](#)]
54. Kwon, S.B.; Jeong, W.; Park, D.; Kim, K.T.; Cho, K.H. A multivariate study for characterizing particulate matter (PM₁₀), PM_{2.5}, and PM₁) in Seoul metropolitan subway stations, Korea. *J. Hazard. Mater.* **2015**, *297*, 295–303. [[CrossRef](#)] [[PubMed](#)]
55. Huang, L.; Mo, J.; Sundell, J.; Fan, Z.; Zhang, Y. Health risk assessment of inhalation exposure to formaldehyde and benzene in newly remodeled buildings, Beijing. *PLoS ONE* **2013**, *8*, e79553. [[CrossRef](#)] [[PubMed](#)]
56. de M. Ochs, S.; de O. Grotz, L.; Factorine, L.S.; Rodrigues, M.R.; Pereira Netto, A.D. Occupational exposure to formaldehyde in an institute of morphology in Brazil: A comparison of area and personal sampling. *Environ. Sci. Pollut. Res.* **2012**, *19*, 2813–2819.
57. Guo, H.; Lee, S.C.; Chan, L.Y.; Li, W.M. Risk assessment of exposure to volatile organic compounds in different indoor environments. *Environ. Res.* **2004**, *94*, 57–66. [[CrossRef](#)]
58. Roda, C.; Barral, S.; Ravelomanantsoa, H.; Dusséaux, M.; Tribout, M.; Le Moullec, Y.; Momas, I. Assessment of indoor environment in Paris child day care centers. *Environ. Res.* **2011**, *111*, 1010–1017. [[CrossRef](#)] [[PubMed](#)]
59. Bolourchi, A.; Atabi, F.; Moattar, F.; Ehyaei, M.A. Experimental and numerical analyses of particulate matter concentrations in underground subway station. *Int. J. Environ. Sci. Technol.* **2017**, *15*, 2569–2580. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.2 Publication 2

- **Title:** *The Use of the Internet of Things for Estimating Personal Pollution Exposure*
- **Authors:** Keith April G. Arano, Shengjing Sun, Joaquin Ordieres-Mere, and Bing Gong
- **Publication date:** 2019/8/28
- **Journal:** *Int. J. Environ. Res. Public Health*
2.468 (2018) ; 5-Year Impact Factor: 2.948 (2018)
Quartile: Q1 in 'Public, Environmental & Occupational Health'
ISSN: 1660-4601
Publisher: MDPI, St. Alban-Anlage 66, 4052 Basel, Switzerland
- **Volume:** 16
- **Issue:** 17
- **Pages:** 3130
- **DOI:** <https://doi.org/10.3390/ijerph16173130>

Article

The Use of the Internet of Things for Estimating Personal Pollution Exposure

Keith April G. Arano ¹, Shengjing Sun ², Joaquin Ordieres-Mere ² and Bing Gong ^{3,*}

¹ Department of Management, Economics and Industrial Engineering, Politecnico di Milano, Via Lambruschini 4/B - (Building 26/B), 20156 Milan, Italy

² Department of Industrial Engineering, Business Administration and Statistics, E.T.S de Ingenieros Industriales, Universidad Politécnica de Madrid, Calle de Jose Gutierrez Abascal 2, 28006 Madrid, Spain

³ Jülich Supercomputing Center Forschungszentrum Jülich GmbH, Wilhelm-Wohnen-Str, 52425 Jülich, Germany

* Correspondence: b.gong@fz-juelich.de

Received: 1 June 2019; Accepted: 29 July 2019; Published: 28 August 2019

Abstract: This paper proposes a framework for an Air Quality Decision Support System (AQDSS), and as a proof of concept, develops an Internet of Things (IoT) application based on this framework. This application was assessed by means of a case study in the City of Madrid. We employed different sensors and combined outdoor and indoor data with spatiotemporal activity patterns to estimate the Personal Air Pollution Exposure (PAPE) of an individual. This pilot case study presents evidence that PAPE can be estimated by employing indoor air quality monitors and e-beacon technology that have not previously been used in similar studies and have the advantages of being low-cost and unobtrusive to the individual. In future work, our IoT application can be extended to include prediction models, enabling dynamic feedback about PAPE risks. Furthermore, PAPE data from this type of application could be useful for air quality policy development as well as in epidemiological studies that explore the effects of air pollution on certain diseases.

Keywords: Personal Air Pollution Exposure (PAPE); air pollution monitoring; IoT; Air Quality Decision Support System; health impact

1. Introduction

Pollution and various forms of ecosystem contamination continue to be pressing issues across the globe [1]. China's rapid increase in urbanization in the last three decades, for example, has resulted into environmental challenges where air pollution is the leading problem [2]. Protecting the environment, therefore, is a serious undertaking that faces businesses and governments today. In recent years, there has been increasing pressure on institutions to measure and report environment-related parameters [3]. For this reason, there has been a significant increase in the number of reporting instruments used globally, of which sustainability reporting instruments account for the largest share owing to government regulations [4].

Environmental sustainability now underpins the policy-building initiatives of government institutions and businesses alike. In developed countries such as those of the European Union (EU), air pollution damage, which brings about a direct threat to public health, is expected to rise in the next decade. This has compelled the EU governments to give priority to air pollution level reduction above any other climate change policy plans [5]. In developing countries, however, there are still inadequate air quality policies and environmental monitoring plans. This is a major concern primarily because these are the regions that are more susceptible to increasing levels of air pollution [6]. There is therefore a challenge in finding economical solutions to monitor pollution levels and other relevant health parameters.

Air pollution from both outdoor and indoor sources constitutes the greatest environmental risk to human health around the globe [7]. About seven million premature deaths were attributed to air pollution in 2014, based on an estimate by the World Health Organization (WHO). It is projected that, by 2050, outdoor air pollution will be the number one cause of environment-related deaths worldwide [8]. For these reasons, governments across the globe have started to monitor the levels of major air pollutants, especially in metropolitan and urban areas.

In addition, people spend around 90% of their time indoors, and human exposure to indoor air pollutants may occasionally be more than 100 times higher than outdoor pollutant levels, according to the United States (US) Environmental Protection Agency (EPA). Indoor air pollution is equally detrimental, as statistics show that 4.3 million people per year die from exposure to household air pollution [9]. Exposure to poor indoor air is a significant cause of productivity loss in the US, as productivity decreases by 0.5 to 5% per workplace, generating a loss of 20 to 200 billion US dollars annually [10].

Monitoring Personal Air Pollution Exposure (PAPE), which refers to the amount (μg) of pollution being inhaled by an individual, has been a topic of growing interest worldwide, not only as a result of global health policies, but more importantly, due to the interest in understanding its effects on various cardiovascular and respiratory diseases. These diseases have been documented widely in existing epidemiological studies [11,12]. Nevertheless, this traditional evaluation of PAPE has not been directly undertaken for individuals, but rather for groups of people exposed to the annual average concentrations of pollution that are indicated by a network of fixed-site outdoor monitors.

Existing studies on the association of air pollution with different diseases [13,14] recognize the importance of measuring PAPE among individuals. By monitoring activity patterns, it is possible to establish correlations between different populations or levels of socioeconomic status and PAPE. Although there are novel methods to measure and model these exposures, the great variability in PAPE remains a major challenge [15] and provides a compelling case for research on the effects of air pollution on health. Dias and Tchepel [16] suggest that, in order to assess personal exposure, not only the spatial-temporal variability of urban air pollution should be taken into account but also the indoor exposure and the individual time-dependent activities should be measured.

The recent advances in data technology are expected to play major roles in the next decade, permitting easier access and analysis of data [1]. The Internet of Things (IoT), which is defined as the network of various ubiquitous devices that are capable of computation and communication over the Internet, has been gaining recognition in the development of advanced applications in the healthcare sector [17]. This wave of digital innovation is driving the healthcare industry and paving the way for cheaper, smaller and more efficient wearable technologies that monitor health indicators in real time [18].

Health systems around the world are under pressure to come up with economical solutions to existing problems [18]. These sensor technologies, coupled with sophisticated analytics, have the means to improve process efficiency and achieve cost reductions [19]. With this kind of digital innovation, an economical and scalable application can be built to help developing countries to measure PAPE and other health parameters to improve their current health risk assessment systems.

Although there was initially a concern about sharing private information from sensor technologies or other similar devices, the general public has started to accept the use of digital services, even for sensitive information such as health data [20]. In fact, in a consumer survey performed by Price Waterhouse Coopers in 2015, 83% of respondents indicated that they were willing to share data to aid in the diagnosis and treatment of diseases [21]. With this collaborative support from the general public, exploring IoT applications to test sensor-driven projects can greatly facilitate advances in the healthcare industry [22]. Insights gained from the surveillance of vital health information, such as PAPE, can establish a foundation for predictive, preventive, and personalized healthcare systems [18].

In line with these trends in the healthcare industry, this paper seeks to propose a framework for an Air Quality Decision Support System (AQDSS) and to develop an IoT application that measures PAPE based on this framework. The four sections of this paper are organized as follows. The first section provides a literature review of studies on key air pollutants, determination methods for air pollutants, and PAPE estimation techniques as well as the current opportunities and challenges in the field. The second part discusses the methodology, which includes the proposed framework and an IoT application that was tested by means of a case study. This is followed in the third part by the analysis and a discussion of the results of the case study. The fourth part presents conclusions that highlight the study's important contributions and directions for future research.

2. Literature Review

2.1. Key Air Pollutants

The key air pollutants that are currently being monitored by agencies such as the WHO, the EPA in the United States, and the European Environment Agency (EEA) in Europe, are particulate matter (PM₁₀ and PM_{2.5}), ozone (O₃), nitrogen oxide (NO), nitrogen dioxide (NO₂), carbon monoxide (CO), sulfur dioxide (SO₂), volatile organic compounds (VOC), and benzene (C₆H₆). They are also frequently studied in academic research [23,24]. Although there is a substantial amount of monitoring data available for each of these pollutants, PM₁₀ and PM_{2.5} are considered to be the most widely studied air pollutants in the existing environmental risk and health literature. This is because PM poses one of the greatest risks to human health [25].

The indoor environment is a critical domain where an average person spends an estimated 90% of his or her time [26]. Thus, indoor air pollution is more likely to account for total population exposure than pollution from the outdoor environment [27]. While individuals are spending more and more time indoors, an assessment of the health impact of indoor air pollution has not been studied as extensively as the impact of outdoor air pollutants. One of the main reasons for this is the lack of indoor air quality monitoring information [28]. There are primary indoor air pollutants, which are recommended based on the EU (2008) directive for Clean Air and the WHO [28]. They are schematically listed as benzene, formaldehyde, naphthalene, nitrogen dioxide, polycyclic aromatic hydrocarbon, radon, trichloroethylene, and tetrachloroethylene. In the research community of indoor air quality monitoring and assessment, particulate matter, carbon dioxide, carbon monoxide, ozone, nitrogen oxide, formaldehyde, benzene, total volatile organic compound (TVOC), polycyclic aromatic hydrocarbon, and other VOCs have been extensively studied [29,30].

2.2. Determination Method for Air Pollutants

Most workspaces or industrial environments still apply traditional measuring strategies to assess occupational health and safety. These strategies are mainly based on the EPA Compendium of Methods [31] and the International Organization for Standardization (ISO) method, which rely on complex sampling and analysis techniques. These methods, such as Method-10A and IP-3A, require domain experts to prepare diffusive or passive samplers and are frequently replaced with new ones due to the limited equipment lifespan. Subsequently, the collected samplers are separated by gas chromatography and measured by mass-selective detector or multidetector techniques in a remote laboratory [32]. Moreover, to measure multiple pollutants, the equipment for each pollutant has to be prepared or bought from different manufacturers, which can lead to issues about data manipulation and integration. These aforementioned elements of the traditional measuring strategy restrict the sampling time to a short-term basis [33]. As indoor air quality varies from time to time due to changes in working conditions, human activity, and weather conditions, short term sampling cannot cover all kinds of variations. Therefore, long-term monitoring has become a need in the research community and practical applications such as Occupational Safety and Health (OSH) management.

The rapid development of IoT and sensor techniques enables light, low-cost, and real-time pollution monitoring solutions. The integration of IoT and the sensor network in air quality monitoring addresses the aforementioned gaps: short-term monitoring and complex air monitoring solutions. Recent studies on the development of indoor air quality monitoring systems have been undertaken on PM, carbon dioxide (CO₂), CO, and VOC. Moreover, IoT-based indoor air monitoring devices such as Foobot and AirVisual are already commercially available on the market.

2.3. IPAPE Measurement Techniques

As noted previously, there is a growing interest in measuring PAPE at the individual level. At present, there is a wide range of low-cost sensor technologies [34] that can be leveraged to implement large scale monitoring networks by means of complex measurement techniques [35]. PAPE requires tracking of a person's activity patterns to learn the time and location of their exposure to pollution concentrations as well as the duration of exposure and nature of the pollutants. This is necessary to understand the probable effects on health of the exposure [36].

The different PAPE measurement techniques that have been developed in the last decade can be grouped into three categories. The first group is the traditional method in which pollution data are collected from fixed-site outdoor monitors and assigned to the home address of the individual through spatial interpolation techniques. Examples include Land Use Regression (LUR) [37], Inverse Distance Interpolation [38], and the geostatistical Kriging algorithm [39]. Numerical models, such as the Community Multiscale Air Quality (CMAQ) model and the Urban Atmospheric Dispersion model (DAUMOD), were proposed for regional air pollution modeling prediction in previous studies [40]. However, the expensive computational cost and failure to capture pollution variability make them inadequate for the application of modeling in real time in urban areas where there are severe photochemical pollution conditions. Graz Lagrangian Model (GRAL) is another advanced mathematical model that can handle the motion of pollution in buildings and complex terrains [41]. A major drawback of these types of models, however, is the need to have accurate information about emissions, meteorological data, and the structural and geographical figures of the area, which may not always be available in high resolution [42]. While the performance of the spatial interpolation methods may significantly drop in dynamic terrains such as in urban environments, they have still been used widely in recent studies [43,44] of areas where the detailed information needed for complex numerical models (e.g., street-based monitoring) is still unavailable.

In summary, these methods are inadequate, as they do not address the issue of the individual's spatio-temporal PAPE variability [45] and neglect indoor air pollution. Accordingly, this has led researchers to explore new techniques that can provide more accurate measures of PAPE.

The second group of techniques, which is built on the traditional method but addresses the issue of exposure variability, takes into account the activity patterns by tracking an individual's location. It incorporates indoor pollution data based on the amount of time spent indoors. A commonly used indoor pollution measurement method is the indoor/outdoor ratio [37]. Other techniques, such as modeling based on data from vehicle type and emissions have also been proposed [46]. With respect to activity tracking, different tools have been used in studies to track the location and activity patterns of an individual. These include Global Positioning Systems (GPS) [35], public WiFi networks [37], and accelerometers [45]. A common characteristic shared by these activity tracking tools is the use of a mobile device, particularly a smartphone. This mobile technology has proved to be an enabling tool in the health industry with its ability to access data anytime from anywhere [47]. Although this group of PAPE measurement techniques is an improvement from the first group, it still faces the issue of pollution variability and measurement accuracy with its reliance on fixed-site outdoor monitors and indoor/outdoor ratios alone [48].

The last group of techniques stems from the two previously discussed groups but further captures the issue of indoor pollution measurement accuracy and the spatio-temporal resolutions of data from fixed-site outdoor monitors. The periodic measurements by these fixed-site outdoor

monitors by nature have low spatial resolution and do not address the issue of variability in pollution concentration [49]. Since the indoor environment has a much greater impact on human health than the outdoor environment [50], it is essential to have solutions that can provide more accurate measures of indoor air pollution instead of employing the traditional method of using the indoor/outdoor ratio.

Personal exposure measurements can be performed directly and indirectly [51]. Passive samplers are widely used in personal sampling, since they have the merits of being light, electricity-free, and wearable. Passive samplers exist for nitrogen dioxide, carbon monoxide, VOC, ozone, sulfur dioxide, and formaldehyde [52]. Due to sampler lifespan, the sampling time usually lasts from a few days up to one week [53].

On the other hand, using a micro-environmental model is an indirect way of assessing personal exposure. In daily life, people move around and are exposed to various levels of pollutants in various locations. The term “micro-environment” is defined as a chunk of air space with a homogeneous pollutant concentration [54]. Such a micro-environment can be an indoor location (bedroom, kitchen, etc.) or workplace location (meeting room, office, printing room, etc.). The spatio-temporal individual time-activity crossing in micro-environments is tracked through questionnaires or time-activity diaries (TADs).

The key to measuring individual pollution exposure is to track an individual’s activities in both the space and time dimensions. GPS technology is the ideal technology, and it has been used successfully for this purpose. Some well-designed integrations of GPS devices and portable pollution monitors have been proposed by some studies [35,55] to determine the potential exposure at the individual level. However, in indoor environments, GPS technology does not function as well as it does outdoors.

Therefore, more extensive approaches have been developed such as the use of mobile sensors (i.e., handheld, USB-pluggable smartphone sensors, wearable sensors) to monitor PAPE indoors. Studies [56,57] that have included this group of PAPE measurement techniques have managed to address the most relevant issues of pollution variability by employing mobile sensors. Beacon technology offers a promising solution for indoor location tracking. Furthermore, the use of indoor monitors instead of mobile sensors, which are often used in similar studies, eliminates the inconvenience of carrying a device around. The use of an indoor monitor and e-beacons also enables unobtrusive and low-cost collection of pollution data for multiple individuals, in contrast to a mobile sensor, which only collects data for a single individual.

2.4. Opportunities and Challenges

Although continuous technological advancements have enabled researchers to propose solutions that provide measures of PAPE, the issue about the cost and scalability of such methods remains to be addressed. The most recent approach, as discussed in the third group of PAPE measurement techniques, employs mobile sensors that the individual carries around. Although these mobile sensors are able to provide better spatial resolution of pollution data, the willingness of individuals to carry these sensors is still a challenge, in addition to the cost and scalability issues of the method.

The trade-off between cost and quality of pollution data continues to be a point of discussion among studies. The proposed PAPE measurement techniques that are currently available in the literature are limited to PAPE estimation alone and, therefore, fail to provide a more comprehensive view of the entire AQDSS. Thus, there is an opportunity to further explore the use of existing technologies to enable the development of a more comprehensive PAPE measurement technique that is able to provide a preventive, predictive, and personalized system.

Within similar studies on the measurement of PAPE, there are some proposed conceptual frameworks [35,36] and system architectures [56,58]. However, they are centered primarily on PAPE measurement and the potential health impacts. In this paper, we present a comprehensive framework that encompasses not only PAPE measurement but provides a holistic view of the entire AQDSS. As a proof of concept for this framework, we also develop a low-cost and unobtrusive IoT application for measuring PAPE that addresses the gaps in the currently available solutions.

3. Methodology

3.1. Framework

Figure 1 shows the proposed framework for an AQDSS. There are three key stakeholders identified, namely the individual, the healthcare industry, and the government. The three pillars at the center represent the elements that are directly linked to the government. They include pollution laws, sectorial regulations, and incentives, all of which make up the air quality monitoring policies of the government.

Figure 1. Proposed framework for the Air Quality Decision Support System (AQDSS).

As pointed out previously, these government regulations play an important role, as they largely support and drive policies that enable the measurement and access to air quality data.

The other two stakeholders are the individuals and the healthcare sector, which form the apex of the framework. They are supported by five layers of activities, as illustrated by the pillars on both sides. The first three on the left correspond to the analysis of past data to estimate PAPE and their related health impacts. The two pillars on the right represent future possibilities of forecasting PAPE and the associated health risks.

Although the estimation of PAPE is fundamental to the entire AQDSS, its associations with forecasting and as a predictive health risk assessment system are noteworthy. Air quality forecasting techniques are already being explored in current research [59,60] in environmental modeling literature,

where their important contributions to the development of control measures to prevent damage to human health have been highlighted.

This proposed framework can be adopted to aid in the development of an AQDSS and various IoT applications. For instance, consider a mobile application that allows an individual to select the best route to travel from home to work that minimizes the risk to pollution exposure. This could be achieved by employing different modeling techniques to continuously analyze real-time air quality data and forecast PAPE values for each of the possible routes to the destination. Actual PAPE data that are stored in the database can also be used for epidemiological studies and air quality policy development. This could be one of the applications of the proposed framework when all identified pillars are fully employed. In this paper, however, as a proof of concept, we focus mainly on PAPE measurement (pillar 2) in accordance with the system architecture that is illustrated on the base of the framework. In order to manage sensor data in an interoperable way, this implementation considers the Web Service Description Language provided by the Sensor Observation Service v2.1 (SOS) from the SOS-OGC consortium. This standard defines a Web service interface which allows observation queries, sensor metadata, as well as representations of observed features. Furthermore, this standard defines a means to register new sensors and to remove existing ones. Also, it defines operations to insert new sensor observations. The feasibility was assessed through the developed case study.

3.1.1. System Architecture

As indicated in the framework, the base shows the set of activities that are related to the gathering and management of all air quality and personal data. This groundwork is required for the entire system to function. There are five different sources of data. They are the (S1) outdoor pollution monitors, (S2) location tracking application, (S3) indoor pollution monitors, (S4) e-beacons, and (S5) meteorological monitors. S1 and S2 are intended for outdoor pollution modeling, and S3 and S4 are for indoor pollution modeling. We also consider meteorological data, as they are relevant for air quality prediction studies [60,61].

The data are extracted from the mentioned data sources and stored in a database management system. Data mining, numerical modeling, and geostatistics, as shown in the center of the framework are the key activities that support the entire system, as it is a continuous process to discover and analyze spatio-temporal data.

3.1.2. PAPE Measurement

There are essentially three elements to consider when measuring PAPE. They are (1) outdoor pollution, (2) indoor pollution, and (3) the individual's location pattern. With respect to the outdoor pollution, the mobile-phone-based tracking app provides the time and location data of the individual in the outdoor environment. An outdoor pollution map is created by using potentially different strategies, such as

- Numerical-modeling-based dispersion models [62,63],
- Big data, machine-learning-based models [60,64],
- Geostatistic-based techniques, like Kriging [65,66].

Each of these techniques has its specific advantages and limitations, and its consideration in a specific application will support its choice. Actually, the latest family of methods has specific advantages, as it is suitable for working with the fixed network of pollution stations the city has implemented. Indeed, it also deals with the limitations of sparse data, as Data fusion can increase the reliability of data as well as it can contribute to dealing with local effects like street canyons, etc., by using the street granularity-based IoT air quality stations some cities are deploying, such as Airbox in Taipei [67] and Array of Things (AoT) sensor boxes in Chicago [68].

For the duration of time that an individual is outdoors, the corresponding pollution data are estimated by superimposing the developed outdoor pollution map over the collected location pattern data.

For the indoor pollution, the e-beacons indicate the period when the individual is indoors, and the indoor air quality monitors provide the corresponding air quality data, when available. Failing this, outdoor information will be used by default. The integration of personal mobiles and fixed e-beacons located in different indoor micro-environments enables the individual's time-location information to be understood. The corresponding time-location knowledge combined with location-specific indoor air quality information collected from air monitoring devices can provide a detailed picture of personal exposure in the indoor environment.

Both outdoor and indoor data are then integrated, and statistical modeling techniques are employed to either estimate or forecast the individual's PAPE.

3.2. Madrid Case Study

In order to assess the feasibility of the proposed IoT application that measures PAPE and contributes to empowering users because of the relevant figures provided at the personal level, we conducted a case study to analyze significant functionalities.

3.2.1. Study Area

The study area was the City of Madrid, which is the capital of Spain, as well as its largest municipality. It was the first city in Spain to have air quality monitoring stations and has always been at the forefront of the fight against air pollution. In response to the most recent EU directive (Directive 2008/50/EC) regarding the establishment of limits to major air pollutants, the Madrid government has committed to maintaining acceptable pollution levels by continuous air quality monitoring.

The Madrid air pollution monitoring network consists of 24 fixed-site outdoor monitors (Figure 2). The hourly averaged measurements of SO₂, CO, NO, NO₂, PM₁₀, PM_{2.5}, C₆H₆, toluene (C₆H₅–CH₃), hexane (C₆H₁₄), propene (C₃H₆), m-xylene, o-xylene, and methane (CH₄) hydrocarbons can be downloaded free of charge from the official open data website of the *Ayuntamiento de Madrid* [69]. Meteorological data, such as temperature, humidity, ultraviolet radiation, pressure, solar radiation, rainfall, precipitation, diffuse solar radiation, global radiation, wind speed, and wind direction can also be accessed through the website of the *Agencia Estatal de Meteorología* [70].

Figure 2. Air Quality Monitoring Network of Madrid.

3.2.2. Data Collection

In this case study, S1, S2, S3, and S4 data sources were used and meteorological data were excluded (see Figure 1). We had one individual volunteer whose activities were monitored during the study period.

For this particular case, Madrid does not yet implement the street-based pollution monitoring strategy, but based on similar studies [43,44], the research team adopted the geostatistics-based approach, as it becomes linear scalable with time and is suitable for integrating additional data sources. Therefore, outdoor pollution figures were downloaded from the mentioned open data website of Madrid City Hall. For location tracking, we used the mobile app Moves [71] in which time, location, and activity were accessed through an open Application Programming Interface (API). Other similar open source mobile apps are widely available, such as OwnTracks [72], Miataru [73], and Geo2Tag [74].

For the indoor pollution, Foobot indoor monitors were used. One of them was placed in the individual's workplace, as this is where she spends most of her indoor time. The indoor pollution data were retrieved from Foobot's API [75]. The e-beacon devices were placed in proximity to the indoor monitors. They helped us to determine whether the individual was within the indoor vicinity. The e-beacon data were broadcasted through Eddystone, an open-source beacon format, and were retrieved through an app that we developed in Cordova [76]—a free and open-source platform for building mobile applications.

All of these data sources promote the scalability of the proposed IoT application, as most are publicly available without charge. The only costs incurred were for the indoor monitor and e-beacons. E-beacons, however, are low-cost, small enough to attach to any surface, and are finding an increasing number of location-based applications in various industries such as retail and transportation as well as in households [77]. Hence, beacon technology offers a promising solution for indoor location tracking.

All data that were collected from the mentioned data sources were processed as indicated in the Data in Brief Collection documents that were submitted to the journal for this paper. The developed code can be also found in a public repository [78].

The selection of the pollutants used for the PAPE estimation was based primarily on the data provided by the devices, which also agreed with the data on the most common air pollutants that have been widely studied previously [23,24]. Table 1 shows the available pollutants for each of the data sources used.

Table 1. Available Pollutants for each Data Source.

Source	Pollutant	Unit
Indoor Monitor	PM _{2.5}	µg/m ³
	CO ₂	ppm
	VOC	ppb
Outdoor Monitor	PM _{2.5}	µg/m ³
	PM ₁₀	µg/m ³
	CO	µg/m ³
	NO ₂	µg/m ³
	SO ₂	µg/m ³
	O ₃	µg/m ³
	NO _x	µg/m ³

3.2.3. Outdoor Pollution Modeling

Existing studies of PAPE essentially rely on modeling techniques in which data collected from fixed-site outdoor monitors are used to estimate pollution at specific geographic locations. To create an outdoor pollution map, there are several alternative methods. These include using micro meteorological numerical-based models (WRF, CMAQ, etc.) [79], or machine-learning-based models [64]. However, for the sake of simplicity and considering the computational costs and the

number of potential users, we adopted some classical but still cost-effective approaches like the Inverse Distance Weighting (IDW), Simple Kriging, Ordinary Kriging, and Co-Kriging algorithms.

Table 2 shows the formulae and main characteristics of these techniques in which $\hat{z}_0$ is the measured value at the prediction location, λ_i is the weight of the measured value at the i th location, and X_i is the measured value at the i th location. The parameters that were tuned are also indicated in the table.

All three methods estimate the value at a particular location by assigning a weight of the surrounding known values and calculating the weighted sum of the data. These techniques differ mainly in the calculation of the assigned weight λ_i . Kriging, which is a geostatistical method, offers advantages over other interpolation techniques, as it provides an interpolation error estimate, and it is an exact interpolation. The interpolations are based on weights that do not depend on data values [80]. The advantages of the deterministic interpolation technique IDW, on the other hand, are that it is simple, intuitive, and computes the interpolated values quickly [81]. We created the outdoor pollution map by employing these three interpolation techniques in R, an open-source statistical modeling software.

Table 2. Techniques Used for Outdoor Pollution Modeling.

Technique	IDW	Simple Kriging	Ordinary Kriging	Co-Kriging
Formula	$\hat{z}_0 = \sum_{i=0}^n \lambda_i X_i$	$\hat{z}_0 - c = \sum_{i=0}^n \lambda_i (X_i - c)$	$\hat{z}_0 = \sum_{i=0}^n \lambda_i X_i$	$\hat{z}_0 = \sum_{i=0}^n \lambda_i X_i + \sum_{j=0}^n \beta_j t_j$
Characteristics	The weight, λ_i , depends solely on the distance to the prediction location.	Assumes a constant and known mean c of the samples. The weight, λ_i , depends on the use of a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location.	Condition that $\sum_{i=0}^n \lambda_i = 1$ assumes a constant and unknown mean of the samples. The weight, λ_i , depends on the use of a fitted model to the measured points, the distance to the prediction location, and the spatial relationships among the measured values around the prediction location.	t_j is the secondary regionalized variable which is co-located with the target variable t_j . The weight β_j assigned to t_j varies between 0 to 1.
Parameters	Idp 0.1, 0.3, 0.5, 0.8, 2, 5	Sill 1, 10, 100, 250, 500, 600, 700, 800, 900 Range 0.1, 0.5, 1, 10, 20, 50, 80 Nugget 0.00001, 0.0001, 0.001, 0.01, 0.1, 1, 10, 100, 200, 50 Beta 0.05, 0.12, 0.2, 0.5, 0.9, 1.5, 3, 1 Variogram Model: Gaussian, Circular, Exponential		

1. Optimal Parameters and Model Selection

In order to select the optimal parameters and the best modeling technique for each of the hourly outdoor pollution datasets, a 5-fold cross validation was performed to avoid overfitting. For each of the 24-hourly datasets and each of the three modeling techniques and all combinations of their respective parameters, the selection of optimal values was based on the root-mean-squared-error (RMSE) metric. The dataset was separated into two parts, training and testing, which were used to fit the model and calculate errors, respectively. The parameters and the model that provided the least RMSE were selected.

For the Simple and Ordinary Kriging techniques, the weights λ_i were derived by fitting a covariance function or variogram. First, a graph of the empirical variogram was plotted and a model was fitted to the points based on this plot. Table 3 shows the different models and functions from which to choose when fitting a model to the empirical variogram. Based on the 5-fold cross validation, the Gaussian Model was selected as the optimal configuration.

2. Outdoor Pollution Map

Similar to [43], an hourly outdoor pollution map was created that was based on the identified optimal parameters and modeling technique for each respective hour. Figure 3 shows an example of the pollution maps based on the PM_{2.5} pollution data on 2017-03-24. It shows that, from midnight to the morning at around 6:00, the highest pollution levels consistently occurred in the southwestern part of the city and moved towards the north with maximum levels that ranged from 8 to 12 µg/m³. Concurrently, high pollution levels were also experienced in the northwestern part of the city at midnight and in the northeastern part at 01:00 in the morning.

The selection of time frequency (hourly-based in this case) also impacts the accuracy, depending on how spiky the pollution looks. In Madrid, the pollution sources are strongly related to traffic and then variations are smooth [82]. Therefore, hourly-based frequency is a rather convenient basis for calculations.

Table 3. Available Pollutants for each Data Source.

Model	Function
Circular	$\text{semivar}(d) = c_0 + c(1 - \frac{2}{\pi} \cos^{-1}(\frac{d}{\alpha}) + \sqrt{1 - \frac{d^2}{\alpha^2}}) \quad 0 < d \leq \alpha$ $\text{semivar}(d) = c_0 + cd > \alpha$ $\text{semivar}(0) = 0$
Spherical	$\text{semivar}(d) = c_0 + c(\frac{3d}{2\alpha} - \frac{1}{2}(\frac{d}{\alpha})^3) \quad 0 < d \leq \alpha$ $\text{semivar}(d) = c_0 + cd > \alpha$ $\text{semivar}(0) = 0$
Exponential	$\text{semivar}(d) = c_0 + c(1 - e^{-\frac{d}{r}}) \quad d > 0$ $\text{semivar}(0) = 0$
Gaussian	$\text{semivar}(d) = c_0 + c(1 - \exp(-\frac{d^2}{r^2})) \quad d > 0$ $\text{semivar}(0) = 0$

d = distance between two locations, c_0 = y-intercept, α = range.

3.2.4. Indoor Pollution Modeling

The main data sources used to model the indoor pollution were the e-beacons and indoor monitor. The timestamps recorded from the e-beacons provide the time when the individual was detected indoors.

In this study, we refer to “indoor” as the work location, since the indoor monitor was only present at the individual’s workplace. The “outdoor” environment, on the other hand, refers to any other location outside the workplace. To obtain the corresponding pollution values during these periods, each of these timestamps was matched to the closest timestamp logged from the Foobot device. As illustrated in Figure 4, the pollution values were then aggregated in time periods based on Equation (1) on the assumption that, if the difference between two sequential timestamps recorded on the e-beacons was more than 10 min, the individual was outdoors and a new indoor period would start. For instance, in Figure 4, from 2017-03-24 at 14:37:46 to 2017-03-24 at 14:44:59, the pollution levels were aggregated, since the timestamp immediately following 2017-03-24 at 14:44:59 is 2017-03-24 at 14:59:00 and the difference is longer than 10 min. Therefore, during the period between 2017-03-24 at 14:44:59 and 2017-03-24 at 14:59:00, the individual was outdoors and the new indoor period resumed at 2017-03-24 14:59:00.

Similarly, micro-environments (office, printing room, meeting room) in the workplace could be replicated by deploying e-beacons and air monitoring devices in all available micro-environments.

Figure 3. Co-kriging interpolation of PM_{2.5} on 24 March 2017.

The PAPE Exposure(p) in period p inhaled by the individual was calculated by multiplying the pollution value SZ(p) by the respective minute ventilation (VE) value using Equation (2).

$$SZ(p) = \sum_{t=i}^n \frac{1}{2} (Z_{t_{i+1}} + Z_{t_i}) (t_{i+1} - t_i) \quad (1)$$

where $t_{i+1} - t_i < 10$ mins, and $SZ(p)$ is the fully aggregated pollution value during the period from time t_i to t_n . This period is named p .

$$Exposure(p) = SZ(p) * VE \quad (2)$$

where $VE \in (VER, VEW, VERT, VEC)$

VER = VE for activity type “run”;

VEW = VE for activity type “walk”;

VERT = VE for activity types “rest” and “transport”;

VEC = VE for activity type “cycle”.

Figure 4. Aggregation of Indoor Pollution Values.

3.2.5. Indoor and Outdoor Pollution Integration

The individual's location was tracked through the Moves mobile application. The recorded data from this tracking app include the starting and ending times, latitude, longitude, and activity type, as shown in Table 4. To obtain the corresponding pollution values for these periods, the time and location records were matched against the interpolated values from the created outdoor pollution map. The resulting outdoor pollution data were then matched against the aggregated indoor pollution values in Figure 4, in which the outdoor data were replaced by the corresponding indoor data.

Table 5 shows the resulting individual's indoor and outdoor PAPE values for PM_{2.5} with the respective period (i.e., starting and ending times), location (i.e., longitude and latitude), environment type (i.e., indoor or outdoor), activity type (i.e., transport, rest, walk, run, cycle), and minute ventilation (VE). The PAPE values are indicated in its last column “Exposure”.

VE (m³/min) measures the volume of gas inhaled by an individual. It varies with the type of activity. The type of activity or travel mode may have a significant effect on the exposure values [83,84] and, hence, it is important to account for VE. We obtained the VE values from a study done by [85] on human inhalation rates. The types of activities in the tracking app include “transport”, “walk”, “run”, and “cycle” and are based primarily on the speed of movement of the individual. In this study, for the time periods that lack one of these types of activity data, we assumed that the individual was at “rest” (i.e., sleeping, sitting, etc.). Since VE is based primarily on the body movement of the individual, we used the same VE values for both activity types “transport” and “rest”.

3.2.6. Practical Application

To illustrate a possible IoT application [86,87] that can be developed using the proposed framework, we identified different travel routes and their corresponding forecasted PAPE values [60]

that give the individual an opportunity to select a travel route that minimizes the risk of exposure to pollution.

As an example, we selected an entry in Table 5 for the time period 12:04:30 to 12:23:44 on 24 March 2017, in which the individual was outdoors and in transport mode. During this selected period, by using the starting and ending location data that the tracking app provided, we identified alternative routes using the ggmap package in R.

From this package, the estimated travel time and route locations (i.e., latitude and longitude) were obtained. Then, based on these specific time and location data, the corresponding pollution values were taken from the previously interpolated outdoor pollution values.

Table 4. Data from Tracking App.

Start Time	End Time	Latitude	Longitude	Activity
2017-03-24 00:00:00	2017-03-24 11:55:37	40.4612	−3.7093	Rest
2017-03-24 11:55:37	3/24/2017 11:59:20	40.4592	−3.7106	Walk
2017-03-24 11:59:20	3/24/2017 12:04:30	40.4571	−3.7118	Rest
2017-03-24 12:04:30	3/24/2017 12:23:44	40.4486	−3.7006	Transport
2017-03-24 12:23:44	3/24/2017 19:52:00	40.4400	−3.6894	Rest
2017-03-24 19:52:00	3/24/2017 20:08:40	40.4506	−3.6994	Transport
2017-03-24 20:08:40	3/24/2017 21:13:07	40.4612	−3.7093	Rest
2017-03-24 21:13:07	3/24/2017 21:21:55	40.4594	−3.7105	Walk
2017-03-24 21:21:55	3/24/2017 22:32:59	40.4575	−3.7117	Rest
2017-03-24 22:32:59	3/24/2017 22:42:51	40.4594	−3.7105	Walk
2017-03-24 22:42:51	3/25/2017 00:00:00	40.4612	−3.7093	Rest

Table 5. Integrated Indoor and Outdoor Personal Air Pollution Exposure (PAPE).

Start	End	Latitude	Longitude	PM _{2.5} (µg/m ³ * min)	Environment	Activity	VE (m ³ /min)	Exposure (µg)
2017-03-24 0:00	2017-03-24 11:55	40.461	−3.709	5354.15601	Outdoor	Rest	0.00893	47.81261
2017-03-24 11:55	2017-03-24 11:59	40.459	−3.711	13.9543	Outdoor	Walk	0.01326	0.18503
2017-03-24 11:59	2017-03-24 12:04	40.457	−3.712	20.08333	Outdoor	Rest	0.00893	0.17934
2017-03-24 12:04	2017-03-24 12:23	40.449	−3.701	111.43482	Outdoor	Transport	0.00893	0.99511
2017-03-24 12:23	2017-03-24 14:03	40.43999	−3.68938	268.65655	Indoor	Rest	0.00893	2.3991
2017-03-24 14:03	2017-03-24 14:37	40.44	−3.689	347.06125	Outdoor	Walk	0.01326	4.60203
2017-03-24 14:37	2017-03-24 14:44	40.43999	−3.68938	28.72532	Indoor	Rest	0.00893	0.25652
2017-03-24 14:44	2017-03-24 14:59	40.44	−3.689	141.90761	Outdoor	Walk	0.01326	1.88169
2017-03-24 14:59	2017-03-24 15:58	40.43999	−3.68938	273.87957	Indoor	Rest	0.00893	2.44574
2017-03-24 15:58	2017-03-24 16:09	40.44	−3.689	80.2915	Outdoor	Walk	0.01326	1.06467
2017-03-24 16:09	2017-03-24 17:00	40.43999	−3.68938	275.25632	Indoor	Rest	0.00893	2.45804
2017-03-24 17:00	2017-03-24 17:12	40.44	−3.689	70.15613	Outdoor	Walk	0.01326	0.93027
2017-03-24 17:12	2017-03-24 17:41	40.43999	−3.68938	253.86892	Indoor	Rest	0.00893	2.26705
2017-03-24 17:41	2017-03-24 17:56	40.44	−3.689	63.02457	Outdoor	Walk	0.01326	0.83571
2017-03-24 17:56	2017-03-24 18:32	40.43999	−3.68938	278.09301	Indoor	Rest	0.00893	2.48337
2017-03-24 18:32	2017-03-24 18:46	40.44	−3.689	85.27723	Outdoor	Walk	0.01326	1.13078
2017-03-24 18:46	2017-03-24 19:19	40.43999	−3.68938	193.88363	Indoor	Rest	0.00893	1.73138
2017-03-24 19:19	2017-03-24 20:08	40.451	−3.699	182.9052	Outdoor	Transport	0.00893	1.63334
2017-03-24 20:08	2017-03-24 21:13	40.461	−3.709	626.14768	Outdoor	Rest	0.00893	5.5915
2017-03-24 21:13	2017-03-24 21:21	40.459	−3.711	68.61289	Outdoor	Walk	0.01326	0.90981
2017-03-24 21:21	2017-03-24 22:32	40.457	−3.712	414.90097	Outdoor	Rest	0.00893	3.70507
2017-03-24 22:32	2017-03-24 22:42	40.459	−3.711	51.34302	Outdoor	Walk	0.01326	0.68081
2017-03-24 22:42	2017-03-25 0:00	40.461	−3.709	6.56381	Outdoor	Rest	0.00893	0.05861

4. Results and Discussion

4.1. Outdoor Pollution Model Performance

The adopted modeling technique based on geostatistics [65,66] using hourly-based data [43] from the fixed network of pollution stations can be interpolated by using different techniques, and criteria for technique selection is needed. Therefore, a cross validation with the leave out strategy was adopted. Based on the 5-fold cross validation performed for each of the three modeling techniques, among the 24 hourly datasets of PM_{2.5} captured on 34 March 2017, the Simple Kriging technique proved to be the

best model with a selection occurrence of 13, followed by the Ordinary Kriging with 10, Co-kriging with 7, and IDW with 1, out of 24 datasets. Our results agree with previous studies such as [88], where Simple Kriging outperformed Co-Kriging, and [89], in which Simple Kriging turned out to be the best model for estimating NO_2 and PM_{10} .

It can be argued that some local effects like turbulence around buildings, roughness of constructions, and some other aspects impact the accuracy of the estimation. The techniques used in such dimensions can be those related to the integration of meteorological, chemical and transportation numerical modeling (WRF and CMAQ models), with the limitations of being able to precisely estimate the boundary conditions as well as to properly model the city configuration (buildings, trees, surface properties, etc.). When running with high spatial resolution, they produce good results, although the quality is slightly reduced and numerical stability becomes an issue [90]. Another potential contribution could be to use artificial-intelligence-based models to estimate pollution levels. In these fields, the authors have already made significant contributions. Actually, some papers [91] have shown the competitive advantage of these methods over those based on numerical simulations. However, to keep the implementation interoperable and extendable, interpolation was finally adopted, because it can easily be enriched with the data fusion option based on IoT-based, street level pollution sensors.

4.2. Device Performance

To validate the fully aggregated indoor pollution values (SZ(p)) obtained from the indoor monitor and e-beacon devices, they were matched against the pollution data that were measured simultaneously during the study period using a portable air pollution monitoring tool- Atmotube [92] that was carried by the individual.

Figure 5 shows the indoor VOC values measured from the Atmotube and the Foobot monitor on 2017-03-31. It can be seen that there is significant measurement variance between the two devices. Nevertheless, the measured values follow the same trend. There is no consistent Air Quality Index (AQI) provided for comparing the pollution values measured by each device. In agreement with [93], the AQI scales differ across countries, organizations, and devices, and this presents an obstacle for comparison and invalidates its usability, which emphasizes the need for a standardized awareness procedure.

Figure 5. Indoor VOC Values from Atmotube and Foobot for 2017-03-31.

Variance in measurements can be attributed to the differences in calibration and measurement methods that were used by these sensors (see Figure 5). However, this situation partially unveils the observed difficulties in getting people aware of the real importance of pollution, as someone can exhibit different figures for the same pollutants at the same place and point in time. Actually, it is another strong point to have a common framework such as the one proposed in this paper, because it mainly fosters transparency and then allows interpolated or modeled values for outdoor pollution over time at a particular place to be compared with local, privately owned sensors from both outdoor, and indoor locations. From such observations where different local sensors can indeed participate, a better understanding about outliers and commonalities and trends can be derived.

4.3. PAPE Values

Figure 6 illustrates a color map of the average PM_{2.5} levels ($\mu\text{g}/\text{m}^3$) for one day, in which the range of the specific values is presented on a color scale on the right. The location pins indicate the environment, activity type, time percentage (%), and the respective amount of PM_{2.5} (μg) that the individual was exposed to within the indicated time duration. It can be seen that the individual spent most of the day (62.78 %) outdoors (i.e., outside the workplace) on the northwest side of the city where the highest daily average pollution level of $12 \mu\text{g}/\text{m}^3$ was concentrated, and this resulted in a total PM_{2.5} exposure of 52.7 μg .

Figure 7 shows the one day PM_{2.5} exposure levels by activity type. Based on this plot, the individual spent most of the day (88.13%) at rest and was exposed to approximately 70 μg of PM_{2.5} during this period. PM_{2.5} exposure values within a selected time period on the same day are also plotted in Figure 8, which shows that the individual had the highest pollution exposure at 15:32 in the afternoon during this selected period.

Figure 6. One Day PM_{2.5} Exposure Per Location, Activity Type, and Time Percentage.

From this analysis, the value of people being able to figure out the distribution of the total intensity of pollutants based on their activity becomes evident, as this method can make them aware of the real dimension of the problem and avoid classical myths, like the idea that most of the pollution is acquired outdoors (see Figure 6). While there are similar studies such as in [87], where the authors demonstrated the cleanest air routing algorithm for path navigation by calculating the $PM_{2.5}$ exposure, they mainly focused on pollution acquired outdoors and not indoors.

Since information is the key aspect in having the opportunity to make proper decisions, the advantage of such an integrated framework that is able to integrate not only outdoor conditions but also indoor ones when available becomes more evident. This can also have an impact not only at the individual level by making everyone aware of their exposed pollution levels but at an aggregated level as well, because the public health dimension is impacted when buildings are seen as actionable regarding the indoor conditions. Therefore, KPIs can be adopted by considering the gradient between outdoor and indoor levels per area of occupancy of the buildings. By having systematic monitoring inside, the management dimension can be adopted.

Figure 7. One Day $PM_{2.5}$ Exposure by Activity Type Percentage.

Figure 8. Indoor $PM_{2.5}$ Exposure Values Across Time. (DeltaT = 1 min).

4.4. Alternative Travel Routes

Similar to [87,94], another non-neglectable dimension that is possible to consider is the impact in terms of transportation decisions. Figure 9 shows different routes that one individual can take when moving from one location to another, and the corresponding aggregated pollution and exposure values are provided in Table 6. These values were predicted on the basis of the individual's activity data for 24 March 2017 from 12:04:30 to 12:23:44. The most frequent one adopted by the user was labeled "Actual", while the other potential routes were named A to C.

In this example, the better individual route will be B, as it causes the least amount of PAPE at $0.769 \mu\text{g}$, which is 22.75% lower than the actual exposure of $0.995 \mu\text{g}$. However, the decision process can be more complex, because there will certainly be some time duration uncertainties, which will consequently result in uncertainty about the total PAPE value of each alternative route.

Figure 9. Alternative Travel Routes.

Although most of the tools that give routing solutions for transportation problems are based on duration, some of them have the capability of filtering them out based on pollution exposure outdoors [87,94]. In terms of added value, this contribution enables alternatives to be ranked based on estimated pollution levels both outdoors and indoors, provided that pollution data is also available inside public transportation modes such as trains, buses, and subways. In these cases, as forecast

for pollution is needed, machine-learning-based models that infer outdoor pollution values need to be used.

Table 6. Total PM_{2.5} and Exposure Values of the Different Routes..

Route	PM _{2.5} (µg/m ³ * min)	Exposure (µg)
A	89.34	0.80
B	86.08	0.77
C	100.17	0.89
Actual	111.43	0.99

4.5. Limitations

Due to the lack of publicly available air quality information for other indoor areas such as shops, buses, cars, metros, etc., outdoor pollution data from the fixed-site outdoors information must be used in such cases. If there are more available resources, additional monitoring IoT devices in other indoor areas will provide greater accuracy. In most cases, good results demand good inputs, and existing data are replaced whenever better data become available. Quality improvements can be expected from those actions. Smart city empowered data sharing platforms such as IOTA Tangle [95] would boost IoT-based indoor air quality resource availability.

Accuracy for outdoor pollution estimation is another known limitation, both because of the time frequency resolution of available data and because of the interpolation errors. It would be possible to implement Weather Research and Forecasting (WRF) models such as the CMAQ. This decision requires significant effort, not only because of using the appropriate Digital Elevation Model (DEM) required to represent the landscape and building configuration, which is a complex task, but because it requires the boundary conditions to be realistic. This means adopting pressure and wind speed conditions for all surfaces external to the volume of interest. These situations need to be updated regularly throughout the day, as environmental conditions change as well. Indeed, numerical stability conditions must be carefully managed in this case as well.

For future applications, the best solution for environments will come from both the increasing deployment of dense (e.g., street-level) IoT-based air quality sensors and the prosperity of the data sharing platform, which can increase the available data and, consequently, will increase the accuracy.

5. Conclusions

This paper (1) proposed a framework for an AQDSS and (2) developed an IoT application based on this framework. The feasibility of the IoT application in measuring PAPE was evaluated through a case study. In comparison to mobile sensors that were used in previous studies, this IoT application has higher scalability, because it involves minimal cost and intrusion to the individual. This pilot case study also presents evidence that PAPE can be estimated by employing indoor monitors and e-beacon technologies that have not been used previously in similar studies.

Using our proposed framework as a general guideline, the IoT application that we developed can be further extended to include prediction models that will allow an individual to make smart decisions when it comes to PAPE risk. Furthermore, PAPE data obtained from the application can be used in air quality policy development as well as in epidemiological studies to explore the correlations of PAPE with certain diseases.

We faced difficulties during the extraction and integration of data from multiple devices, which highlights the importance of choosing the right technologies to use when developing such IoT applications. There was an observed variance among the different devices, which can be attributed to device calibration and the measurement techniques used. Future research should, therefore, explore these issues and identify emerging technologies that permit seamless data integration and more accurate PAPE measurements.

Author Contributions: The conceptual framework was carried out by B.G. and J.O.-M., while implementation of software was carried out by B.G. Technical writing was carried out mainly by K.A.G.A. and S.S. Funding acquisition was done mainly by J.O.-M.

Funding: Financial support for this research was granted by China Scholarship Council and the European Commission through the EU funded project under the RFCS program with project number 793505 4.0 Lean system integrating workers and processes (WISEST).

Acknowledgments: The authors would like to acknowledge the City Hall of Madrid for the public data on air quality. Indeed, additional acknowledge is due to the Spanish Agencia Estatal de Investigación, through the research project with code RTI2018-094614-B-I00 into the “Programa Estatal de I+D+i Orientada a los Retos de la Sociedad”.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Initiative, G.R. *Sustainability and Reporting Trends in 2025: Preparing for the Future*; Technical Report; Global Reporting Initiative: Amsterdam, The Netherlands, 2015.
2. EY. *Rapid-Growth Markets*; Technical Report; EY: Oxford, UK, 2014.
3. KPMG. *Ten Emerging Trends in 2017*; Technical Report; KPMG: Zurich, Switzerland, 2017.
4. Bartels, W.; Fogerlberg, T.; Hoballah, A.; Van der Lugt, C.C.A. *Carrots & Sticks: Global Trends in Sustainability Reporting Regulation and Policy*; Technical Report; KPMG: Zurich, Switzerland, 2016.
5. Bollen, J.; Brink, C. Air pollution policy in Europe: Quantifying the interaction with greenhouse gases and climate change policies. *Energy Econ.* **2014**, *46*, 202–215. [[CrossRef](#)]
6. Shah, A.S.V.; Langrish, J.P.; Nair, H.; McAllister, D.A.; Hunter, A.L.; Donaldson, K.; Newby, D.E.; Mills, N.L. Global Association of Air Pollution and Heart Failure: A Systematic Review and Meta-Analysis. *Lancet* **2013**, *382*, 1039–1048. [[CrossRef](#)]
7. WHO. *Evolution of WHO Air Quality Guidelines: Past, Present and Future*; Technical Report; WHO: Copenhagen, Denmark, 2017.
8. OECD. *Environmental Outlook To 2050: The Consequences of Inaction Key Findings on Health and Environment*; Technical Report; OECD: Paris, France, 2012.
9. WHO. *Household (Indoor) Air Pollution*; WHO: Liverpool, UK, 2017.
10. Fisk, W.J.; Rosenfeld, A.H. Estimates of Improved Productivity and Health from Better Indoor Environments. *Indoor Air* **2004**, *7*, 158–172. [[CrossRef](#)]
11. Chen, R.; Chu, C.; Tan, J.; Cao, J.; Song, W.; Xu, X.; Jiang, C.; Ma, W.; Yang, C.; Chen, B.; et al. Ambient air pollution and hospital admission in Shanghai, China. *J. Hazard. Mater.* **2010**, *181*, 234–240. [[CrossRef](#)] [[PubMed](#)]
12. Micheli, G.J.; Farné, S. Urban railway traffic noise: Looking for the minimum cost for the whole community. *Appl. Acoust.* **2016**, *113*, 121–131. [[CrossRef](#)]
13. Kuo, C.Y.; Chan, C.K.; Wu, C.Y.; Phan, D.V.; Chan, C.L. The Short-Term Effects of Ambient Air Pollutants on Childhood Asthma Hospitalization in Taiwan: A National Study. *Int. J. Environ. Res. Public Health* **2019**, *16*. [[CrossRef](#)]
14. Saldiva, S.R.D.M.; Barrozo, L.V.; Leone, C.R.; Failla, M.A.; Bonilha, E.D.A.; Bernal, R.T.I.; Oliveira, R.C.D.; Saldiva, P.H.N. Small-Scale Variations in Urban Air Pollution Levels Are Significantly Associated with Premature Births: A Case Study in Sao Paulo, Brazil. *Int. J. Environ. Res. Public Health* **2018**, *15*. [[CrossRef](#)] [[PubMed](#)]
15. Laumbach, R.J.; Kipen, H.M. Respiratory health effects of air pollution: Update on biomass smoke and traffic pollution. *J. Allergy Clin. Immunol.* **2012**, *129*, 3–11. [[CrossRef](#)]
16. Dias, D.; Tchepel, O. Spatial and Temporal Dynamics in Air Pollution Exposure Assessment. *Int. J. Environ. Res. Public Health* **2018**, *15*. [[CrossRef](#)]
17. Philip, V.; Suman, V.K.; Menon, V.G.; Dhanya, K.A.A. Review on Latest Internet of Things Based Healthcare Applications. *Int. J. Comput. Sci. Inf. Security* **2017**, *15*, 248–255.
18. EY. *Health Reimagined: A New Participatory Health Paradigm*; Technical Report; EY: Melbourne, Australia, 2016.
19. Deloitte. *2015 Global Health Care Outlook: Common Goals, Competing Priorities*; Technical Report; Deloitte: London, UK, 2015.
20. Niedermans, F.; Biesdorf, S. *Healthcare's Digital Future*; Technical Report; McKinsey: Munich, Germany, 2014.

21. Barnes, K.; Isgur, B.; Tsouderos, T. *Top Health Industry Issues of 2016: Thriving in the New Health Economy*; Health Research Institute: USA, 2015; pp. 1–18.
22. Reh, G.; Korenda, L.; Boozer, C. *Will Patients and Caregivers Embrace Technology-Enabled Health Care?* Technical Report; Deloitte: London, UK, 2016.
23. Kim, K.H.; Kabir, E.; Kabir, S. A review on the human health impact of airborne particulate matter. *Environ. Int.* **2015**, *74*, 136–143. [[CrossRef](#)] [[PubMed](#)]
24. Szigeti, T.; Dunster, C.; Cattaneo, A.; Cavallo, D.; Spinazze, A.; Saraga, D.E.; Sakellaris, I.A.; de Kluizenaar, Y.; Cornelissen, E.J.; Hänninen, O.; et al. Oxidative potential and chemical composition of PM_{2.5} in office buildings across Europe—The OFFICAIR study. *Environ. Int.* **2016**, *92–93*, 324–333. [[CrossRef](#)] [[PubMed](#)]
25. Guerreiro, C.B.; Foltescu, V.; de Leeuw, F. Air quality status and trends in Europe. *Atmos. Environ.* **2014**, *98*, 376–384. [[CrossRef](#)]
26. Abraham, S.; Li, X. Design of A Low-Cost Wireless Indoor Air Quality Sensor Network System. *Int. J. Wirel. Inf. Netw.* **2016**, *23*, 57–65. [[CrossRef](#)]
27. WHO. Air Quality Guidelines. Global Update 2005. *Environ. Sci. Pollut. Res.* **2006**, *3*, 23.
28. Penney, D.; Benignus, V.; Kephelopoulous, S.; Kotzias, D.; Kleinman, M.; Verrier, A. *Guidelines for Indoor Air Quality*; Technical Report; WHO: Copenhagen, Denmark, 2010.
29. Anna Mainka, B.K. Assessment of the BTEX concentrations and health risk in urban nursery schools in Gliwice, Poland. *AIMS Environ. Sci.* **2016**, *3*, 858. [[CrossRef](#)]
30. Yurdakul, S.; Civan, M.; Ozden, O.; Gaga, E.; Dogeroglu, T.; Tuncel, G. Spatial variation of VOCs and inorganic pollutants in a university building. *Atmos. Pollut. Res.* **2017**, *8*, 1–12. [[CrossRef](#)]
31. EPA. *Compendium of Methods for the Determination of Air Pollutants in Indoor Air*; Technical Report; EPA: Springfield, VA, USA, 1990.
32. Ishizaka, T.D.; Kawashima, A.; Hishida, N.; Hamada, N. Measurement of total volatile organic compound (TVOC) in indoor air using passive solvent extraction method. *Air Qual. Atmos. Health* **2019**, *12*, 173–187. [[CrossRef](#)]
33. Topping, M. OCCUPATIONAL EXPOSURE LIMITS FOR CHEMICALS. *Occup. Environ. Med.* **2001**, *58*, 138–144. [[CrossRef](#)]
34. Kumar, P.; Morawska, L.; Martani, C.; Biskos, G.; Neophytou, M.; Sabatino, S.D.; Bell, M.; Norford, L.; Britter, R. The rise of low-cost sensing for managing air pollution in cities. *Environ. Int.* **2015**, *75*, 199–205. [[CrossRef](#)]
35. Reis, S.; Seto, E.; Northcross, A.; Quinn, N.W.; Convertino, M.; Jones, R.L.; Maier, H.R.; Schlink, U.; Steinle, S.; Vieno, M.; et al. Integrating modelling and smart sensors for environmental and human health. *Environ. Model. Softw.* **2015**, *74*, 238–246. [[CrossRef](#)]
36. Steinle, S.; Reis, S.; Sabel, C.E. Quantifying human exposure to air pollution—Moving from static monitoring to spatio-temporally resolved personal exposure assessment. *Sci. Total Environ.* **2013**, *443*, 184–193. [[CrossRef](#)] [[PubMed](#)]
37. Su, J.G.; Jerrett, M.; Meng, Y.Y.; Pickett, M.; Ritz, B. Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment. *Sci. Total Environ.* **2015**, *506–507*, 518–526. [[CrossRef](#)] [[PubMed](#)]
38. Milando, C.W.; Martenies, S.E.; Batterman, S.A. Assessing concentrations and health impacts of air quality management strategies: Framework for Rapid Emissions Scenario and Health impact ESTimation (FRESH-EST). *Environ. Int.* **2016**, *94*, 473–481. [[CrossRef](#)] [[PubMed](#)]
39. Mercer, L.D.; Szpiro, A.A.; Sheppard, L.; Lindström, J.; Adar, S.D.; Allen, R.W.; Avol, E.L.; Oron, A.P.; Larson, T.; Liu, L.J.S.; et al. Comparing universal kriging and land-use regression for predicting concentrations of gaseous oxides of nitrogen (NO_x) for the Multi-Ethnic Study of Atherosclerosis and Air Pollution (MESA Air). *Atmos. Environ.* **2011**, *45*, 4412–4420. [[CrossRef](#)] [[PubMed](#)]
40. Rojas, A.L.P. Simple atmospheric dispersion model to estimate hourly ground-level nitrogen dioxide and ozone concentrations at urban scale. *Environ. Model. Softw.* **2014**, *59*, 127–134. [[CrossRef](#)]
41. Oetl, D. *Documentation of the Lagrangian Particle Model GRAL (Graz Lagrangian Model)*; Amt der Steiermärk: Graz, Austria, 2016.
42. Marjovi, A.; Arfire, A.; Martinoli, A. High Resolution Air Pollution Maps in Urban Environments Using Mobile Sensor Networks. In Proceedings of the 2015 International Conference on Distributed Computing in Sensor Systems, Fortaleza, Brazil, 10–12 June 2015; pp. 11–20. [[CrossRef](#)]

43. Gomez-Losada, A.; Santos, F.M.; Gibert, K.; Pires, J.C. A data science approach for spatiotemporal modelling of low and resident air pollution in Madrid (Spain): Implications for epidemiological studies. *Comput. Environ. Urban Syst.* **2019**, *75*, 1–11. [CrossRef]
44. Montero, J.M.; Fernández-Avilés, G. Functional kriging prediction of atmospheric particulate matter concentrations in Madrid, Spain: Is the new monitoring system masking potential public health problems? *J. Clean. Prod.* **2018**, *175*, 283–293. [CrossRef]
45. de Nazelle, A.; Seto, E.; Donaire-Gonzalez, D.; Mendez, M.; Matamala, J.; Nieuwenhuijsen, M.J.; Jerrett, M. Improving estimates of air pollution exposure through ubiquitous sensing technologies. *Environ. Pollut.* **2013**, *176*, 92–99. [CrossRef] [PubMed]
46. Yu, T.C.; Lin, C.C.; Chen, C.C.; Lee, W.L.; Lee, R.G.; Tseng, C.H.; Liu, S.P. Wireless sensor networks for indoor air quality monitoring. *Med. Eng. Phys.* **2013**, *35*, 231–235. [CrossRef] [PubMed]
47. Schlesinger, J.; Burris, S.; Tippmann, C. *Health and Mobility: Realizing the Power of Mobile Technology*; Technical Report; EY: Los Angeles, CA, USA, 2015.
48. Tong, Z.; Chen, Y.; Malkawi, A.; Adamkiewicz, G.; Spengler, J.D. Quantifying the impact of traffic-related air pollution on the indoor air quality of a naturally ventilated building. *Environ. Int.* **2016**, *89–90*, 138–146. [CrossRef] [PubMed]
49. Mead, M.; Popoola, O.; Stewart, G.; Landshoff, P.; Calleja, M.; Hayes, M.; Baldovi, J.; McLeod, M.; Hodgson, T.; Dicks, J.; et al. The use of electrochemical sensors for monitoring urban air quality in low-cost, high-density networks. *Atmos. Environ.* **2013**, *70*, 186–203. [CrossRef]
50. Saad, S.M.; Mohd Saad, A.R.; Kamarudin, A.M.Y.; Zakaria, A.; Shakaff, A.Y.M. Indoor air quality monitoring system using wireless sensor network (WSN) with web interface. In Proceedings of the 2013 International Conference on Electrical, Electronics and System Engineering (ICEESE), Selangor, Malaysia, 4–5 December 2013; pp. 60–64. [CrossRef]
51. Ott, W.R. Concepts of human exposure to air pollution. *Environ. Int.* **1982**, *7*, 179–196. [CrossRef]
52. Nash, D.G.; Leith, D. Use of Passive Diffusion Tubes to Monitor Air Pollutants. *J. Air Waste Manag. Assoc.* **2010**, *60*, 204–209. [CrossRef] [PubMed]
53. Monn, C. Exposure assessment of air pollutants: A review on spatial heterogeneity and indoor/outdoor/personal exposure to suspended particulate matter, nitrogen dioxide and ozone. *Atmos. Environ.* **2001**, *35*, 1–32. [CrossRef]
54. Klepeis, N. Modeling Human Exposure To Air Pollution. Available online: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.460.8304&rep=rep1&type=pdf> (accessed on 15 February 2017).
55. Gu, Y.; Yim, S. The air quality and health impacts of domestic trans-boundary pollution in various regions of China. *Environ. Int.* **2016**, *97*, 117–124. [CrossRef] [PubMed]
56. Predic, B.; Yan, Z.; Eberle, J.; Stojanovic, D.; Aberer, K. ExposureSense: Integrating Daily Activities with Air Quality Using Mobile Participatory Sensing. In Proceedings of the 2013 IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops, San Diego, CA, USA, 18–22 March 2013; pp. 303–305.
57. Steinle, S.; Reis, S.; Sabel, C.E.; Semple, S.; Twigg, M.M.; Braban, C.F.; Leeson, S.R.; Heal, M.R.; Harrison, D.; Lin, C.; et al. Personal exposure monitoring of PM_{2.5} in indoor and outdoor microenvironments. *Sci. Total Environ.* **2015**, *508*, 383–394. [CrossRef] [PubMed]
58. Liu, R.; Wang, Y.; Shu, M. Internet of Things Healthcare Cloud System Based on IEEE 802.15.4. *J. Appl. Sci.* **2013**, *13*, 1582–1586. [CrossRef]
59. Stojić, A.; Maletić, D.; Stojić, S.S.; Mijić, Z.; Šoštarić, A. Forecasting of VOC emissions from traffic and industry using classification and regression multivariate methods. *Sci. Total Environ.* **2015**, *521–522*, 19–26. [CrossRef] [PubMed]
60. Gong, B.; Ordieres-Mere, J. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of Hong Kong. *Environ. Model. Softw.* **2016**, *84*, 290–303. [CrossRef]
61. Diaz-Robles, L.A.; Ortega, J.C.; Fu, J.S.; Reed, G.D.; Chow, J.C.; Watson, J.G.; Moncada-Herrera, J.A. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban areas: The case of Temuco, Chile. *Atmos. Environ.* **2008**, *42*, 8331–8340. [CrossRef]
62. Beevers, S.D.; Kitwiroon, N.; Williams, M.L.; Carslaw, D.C. One way coupling of CMAQ and a road source dispersion model for fine scale air pollution predictions. *Atmos. Environ.* **2012**, *59*, 47–58. [CrossRef] [PubMed]

63. Raducan, G. Pollutant dispersion modelling with OSPM in a street canyon from Bucharest. *Roman. Report Phys.* **2008**, *60*.
64. Gong, B.; Ordieres-Mere, J. Reconfiguring existing pollutant monitoring stations by increasing the value of the gathered information. *Environ. Model. Softw.* **2017**, *96*, 106–122. [CrossRef]
65. Farhi, A.; Boyko, V.; Almagor, J.; Benenson, I.; Segre, E.; Rudich, Y.; Stern, E.; Lerner-Geva, L. The possible association between exposure to air pollution and the risk for congenital malformations. *Environ. Res.* **2014**, *135*, 173–180. [CrossRef] [PubMed]
66. Nikzad, N.; Verma, N.; Ziftci, C.; Bales, E.; Quick, N.; Zappi, P.; Patrick, K.; Dasgupta, S.; Krueger, I.; Rosing, T.V.; et al. CitiSense: Improving Geospatial Environmental Assessment of Air Quality Using a Wireless Personal Exposure Monitoring System. In *Proceedings of the Conference on Wireless Health*; ACM: New York, NY, USA, 2012; pp. 11:1–11:8. [CrossRef]
67. Government, T.C. Air Boxes PM2.5. Available online: <https://smartcity.taipei/posts/3?locale=en> (accessed on 30 April 2019).
68. Catlett, C. Array of Things. Available online: <https://www.anl.gov/mcs/array-of-things> (accessed on 30 April 2019).
69. Council, M.C. Air Quality: Real-time data. Available online: <https://www.madrid.es/portal/site/munimadrid> (accessed on 13 February 2017).
70. AEMET. Open data from AEMET. Available online: <https://opendata.aemet.es/centrodedescargas/inicio> (accessed on 13 February 2017).
71. Welsh, B.; Baird, T.; Zhao, J.; Block-Schachter, D. Web App Design to Implement Travel Behavioral Nudging Using “Moves”. In *Proceedings of the Transportation Research Board 93rd Annual Meeting*, Washington, DC, USA, 12–14 January 2014.
72. OwnTracks. OwnTracks: Your Location Companion. Available online: <https://owntracks.org/> (accessed on 13 February 2017).
73. Kirstenpfad, D. Miataru - Be Found: OpenSource Location Tracking. Available online: <http://miataru.com/ios/> (accessed on 13 February 2017).
74. Geo2tag. TAG EVERYWHERE(EVERYTHING) WITH GEO2TAG. Available online: <http://www.geo2tag.com/> (accessed on 13 February 2017).
75. Loh, M.; Sarigiannis, D.; Gotti, A.; Karakitsios, S.; Pronk, A.; Kuijpers, E.; Annesi-Maesano, I.; Baiz, N.; Madureira, J.; Oliveira Fernandes, E.; et al. How Sensors Might Help Define the External Exposome. *Int. J. Environ. Res. Public Health* **2017**, *14*. [CrossRef] [PubMed]
76. Mahesh, B.R.; Kumar, M.B.; Manoharan, R.; Somasundaram, M.; Karthikeyan, S.P. Portability of mobile applications using PhoneGap: A case study. In *Proceedings of the International Conference on Software Engineering and Mobile Application Modelling and Development (ICSEMA 2012)*, Chennai, India, 19–21 December 2012; pp. 1–6. [CrossRef]
77. Danova, T. These Are Some Of The Top Emerging Business Applications For Beacons. Available online: <https://www.businessinsider.com/top-emerging-applications-for-beacons-2014-4> (accessed on 3 March 2017).
78. Gong, B. masak1112/IoT-for-Estimating-Personal-Pollution-Doses 0.1.0. Available online: <https://zenodo.org/record/818074#.XWVKMpNKiYU> (accessed on 25 June 2017).
79. Vedrenne, M.; Borge, R.; Lumbreras, J.; Rodriguez, M.E. Advancements in the design and validation of an air pollution integrated assessment model for Spain. *Environ. Model. Softw.* **2014**, *57*, 177–191. [CrossRef]
80. Andria, G.; Cavone, G.; Lanzolla, A.M. Modelling study for assessment and forecasting variation of urban air pollution. *Measurement* **2008**, *41*, 222–229. [CrossRef]
81. Li, L.; Losser, T.; Yorke, C.; Piltner, R. Fast Inverse Distance Weighting-Based Spatiotemporal Interpolation: A Web-Based Application of Interpolating Daily Fine Particulate Matter PM2.5 in the Contiguous U.S. Using Parallel Programming and k-d Tree. *Int. J. Environ. Res. Public Health* **2014**, *11*, 9101–9141. [CrossRef]
82. Council, M.C. Web Portal of Air Quality of the Madrid City Council. Available online: <https://www.madrid.es/portal/site/munimadrid> (accessed on 15 June 2019).
83. de Nazelle, A.; Bode, O.; Orjuela, J.P. Comparison of air pollution exposures in active vs. passive travel modes in European cities: A quantitative review. *Environ. Int.* **2017**, *99*, 151–160. [CrossRef]
84. O'Donoghue, R.; Gill, L.; McKevitt, R.; Broderick, B. Exposure to hydrocarbon concentrations while commuting or exercising in Dublin. *Environ. Int.* **2007**, *33*, 1–8. [CrossRef]

85. Allan, M.; Richardson, G.M.; Jones-Otazo, H. Probability Density Functions Describing 24-Hour Inhalation Rates for Use in Human Health Risk Assessments: An Update and Comparison. *Human Ecol. Risk Assess. Int. J.* **2008**, *14*, 372–391. [CrossRef]
86. Dirks, K.N.; Salmond, J.A.; Talbot, N. Air Pollution Exposure in Walking School Bus Routes: A New Zealand Case Study. *Int. J. Environ. Res. Public Health* **2018**, *15*. [CrossRef] [PubMed]
87. Mahajan, S.; Tang, Y.S.; Wu, D.Y.; Tsai, T.C.; Chen, L.J. CAR: The Cleanest Air Routing Algorithm for Path Navigation with Minimal PM_{2.5} Exposure on the Move. In Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, Munich, Germany, 10–15 June 2018; ACM: New York, NY, USA, 2018; pp. 532–532. [CrossRef]
88. Krivoruchko, K. *Using Linear and Non-Linear Kriging Interpolators to Produce Probability Maps*; Environmental Systems Research Institute: Redlands, CA, USA, 2001.
89. Halimi, M.; Farajzadeh, M.; Zarei, Z. Modeling spatial distribution of Tehran air pollutants using geostatistical methods incorporate uncertainty maps. *Pollution* **2016**, *2*, 375–386. [CrossRef]
90. San José, R.; Pérez, J.; González, R. Very High Resolution Urban Simulations with WRF/UCM and CMAQ over European Cities. In *Urban Environment*; Rauch, S., Morrison, G., Norra, S., Schleicher, N., Eds.; Springer: Dordrecht, The Netherlands, 2013; pp. 293–301.
91. Fernando, H.; Mammarella, M.; Grandoni, G.; Fedele, P.; Marco, R.D.; Dimitrova, R.; Hyde, P. Forecasting PM₁₀ in metropolitan areas: Efficacy of neural networks. *Environ. Pollut.* **2012**, *163*, 62–67. [CrossRef]
92. Atmotube. Atmotube Makes You Smarter about the Air You Breathe. Available online: <https://atmotube.com/> (accessed on 13 February 2017).
93. Lemes, S. Air Quality Index (AQI)—Comparative Study And Assessment Of An Appropriate Model For B&H. In Proceedings of the 12th Scientific/Research Symposium with International Participation “Metallic And Nonmetallic Materials” MNM 2018, Zenica, Bosnia and Herzegovina, 19–20 April 2018.
94. Luo, J.; Boriboonsomsin, K.; Barth, M. Reducing pedestrians’ inhalation of traffic-related air pollution through route choices: Case study in California suburb. *J. Trans. Health* **2018**, *10*, 111–123. [CrossRef]
95. Tangle, I. Meet the Tangle. Available online: <https://www.iota.org/research/meet-the-tangle> (accessed on 13 February 2017).

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.3 Publication 3

- **Title:** *Non-Invasive Ambient Intelligence in Real Life: Dealing with Noisy Patterns to Help Older People*
- **Authors:** Miguel Ángel Antón, Joaquín Ordieres-Meré, Unai Saralegui, and Shengjing Sun
- **Publication date:** 2019/07/14
- **Journal:** *Sensors*
3.031 (2018) ; 5-Year Impact Factor: 3.302 (2018)
Quartile: Q1 in ‘Instruments & Instrumentation’
ISSN: 1424-8220
Publisher: Multidisciplinary Digital Publishing Institute, BASEL, SWITZERLAND
- **Volume:** 19
- **Issue:** 14
- **Pages:** 3113
- **DOI:** <https://doi.org/10.3390/s19143113>

Article

Non-Invasive Ambient Intelligence in Real Life: Dealing with Noisy Patterns to Help Older People

Miguel Ángel Antón ¹ , Joaquín Ordieres-Meré ^{2,3,*} , Unai Saralegui ¹ and Shengjing Sun ² ¹ Tecnia, Parque Científico y Tecnológico de Gipuzkoa, Mikeletegi Pasealekua, 2. 20009 San Sebastián, Spain² Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain³ i3-crg, École Polytechnique, Route de Saclay, 91128 Palaiseau, France

* Correspondence: j.ordieres@upm.es; Tel.: +34-91-06-77-107

Received: 7 May 2019; Accepted: 11 July 2019; Published: 14 July 2019

Abstract: This paper aims to contribute to the field of ambient intelligence from the perspective of real environments, where noise levels in datasets are significant, by showing how machine learning techniques can contribute to the knowledge creation, by promoting software sensors. The created knowledge can be actionable to develop features helping to deal with problems related to minimally labelled datasets. A case study is presented and analysed, looking to infer high-level rules, which can help to anticipate abnormal activities, and potential benefits of the integration of these technologies are discussed in this context. The contribution also aims to analyse the usage of the models for the transfer of knowledge when different sensors with different settings contribute to the noise levels. Finally, based on the authors' experience, a framework proposal for creating valuable and aggregated knowledge is depicted.

Keywords: smart building; IoT; machine learning; ambient intelligence; ambient assisted living

1. Introduction

It has been largely demonstrated that huge improvements in human health and the increased attention being paid to the impacts of different factors (e.g., food and habits (smoking, sports, etc.)) on quality of life are resulting in the increased longevity of humans. According to a United Nations report [1], it is expected that there will be more than one billion people aged over 65 years by 2030, which is equivalent to one in eight people on Earth. Significantly, the most rapid increases in the 65-and-older population are occurring in developing countries, which will see a jump of 140% by 2030. This effect will dramatically impact the required infrastructures (e.g., homes). Therefore, it will be necessary to possess smarter ways to evaluate people's behaviour and needs. There are different approaches for handling this, starting from ambient intelligence (AmI), which refers to electronic environments that are sensitive and responsive to the presence of people. It is possible to identify "ambient assisted living" (AAL) aiming to help individuals to continue to carry out their daily activities in their homes, for as long as possible and with the least amount of assistance.

Although there are research works aimed at tracking the movements of individuals with high precision, in order to gain knowledge regarding how the interactions develop for different tasks, as a wide variety of Internet of Things (IoT) devices are used [2,3], this paper aims to discuss how to exploit the advantages of the knowledge gathered in much more real contexts with non-intrusive sensors. The ambition of this work is to accept high levels of noise in the dataset, which is not very rich in terms of detail, and to try to derive useful behavioural rules to be applied in the same or other environments. The noise arises because of the usage on general non-intrusive sensors in real ambiances, where many uncontrolled factors such as not properly closed room door, not foreseen requests, etc.,

can affect the measures from those sensors. As far as those features are not directly related to the outcome of the model, they can be seen as noise for the learning process. Our major motivation is the fear that most users report the unwanted intrusive behaviour of devices that are already in use. Therefore, we are looking to still derive useful knowledge from the less specific (non-intrusive) sensor network easily available at Smart Homes.

According to the TRUSTe report [4], 83% of U.S. Internet users and 84% of British Internet users were concerned about the idea of information being collected through smart devices, and 87% of U.S. Internet users compared with 84% of British Internet users were concerned about the type of information collected by smart devices (see Figure 1). Although the IoT literature still lacks studies on the behavioural aspect that explains the customers' perception of IoT adoption and focuses more on technological aspects, an IoT technology trust model is proposed in [5] that covers a varied set of factors related to consumer trust, which affects the decision to adopt IoT technology, including three main dimensions and eight domains. The three dimensions are: security of products and service-related factors, product-related factors, and social-influence-related factors. According to their survey analysis, the security and privacy of IoT products and services are among the highest priorities to ensure consumers' trust; however, they remain a challenge in IoT technology. For this reason, this paper has the goal of promoting the development of useful behavioural models by learning from non-intrusive general-purpose sensors that are already in place to cover other functions. Additionally, because of the intrinsic uncertainties, it becomes relevant to analyse to what extent the uncertainty in the underlying datasets impacts the quality of the models.

Figure 1. Circle of trust regarding the Internet of Things. Research carried out by TRUSTe. Source: [4].

A non-invasive environment for tracking the behaviour of individuals is particularly useful for the living environments of the ageing population, although some concerns regarding usability and user friendliness have been raised [6]. There is a wide range of solutions for monitoring the mobility of this population, but many proposals are based on the installation of new dedicated sensors [7]. However, it is also possible to take advantage of the already-existing infrastructure, designed mainly for other purposes, to infer information associated with people's behaviour [6,8,9].

According to the World Health Organisation (WHO), over 47 million people live with dementia worldwide [10]. These figures are increasing, and, by 2050, the prevalence of Alzheimer's-related diseases will quadruple [11]. Unusual behaviour and mobility patterns in the older population may be early symptoms of dementia [12]. The detection of these patterns would help provide a better intervention regarding the living environments of people in the early stages of dementia to improve both their well-being and that of their caregivers, and could help to delay institutionalisation [13].

On the other hand, analysing the behavioural changes in the routines of already-diagnosed people helps to evaluate their long-term changes and enables the implementation of effective assistance and palliative care [14]. Although specialised equipment and expertise are needed for accurate diagnosis, smart environments with ambient intelligence may raise the alarm and allow for further investigation to be made.

From the previous analysis different cohorts of population can benefit from soft monitoring systems able to analyse trends in different temporal scale, including elaborated information from the existing sensors. Such cohorts include risk attention to daily activities in the kitchen, initial steps of some diseases, and many other regular cases.

To show that these goals can be effectively reached, the remaining structure of the paper develops as follows. Section 2 reviews different relevant contributions in the field. Section 3 presents the promoted framework enabling the development of user-centred AmI applications and it defines the real context used to learn from behaviours. Later, in Section 4, the modelling effort is coherently developed for the context defined in the previous section. In Section 5, the results for the modelling activities are presented. Then, Section 6 discusses the results found in relation to the expectations, and it derives the impact of noise effects. Finally, Section 7 summarises the findings, establishes limitations and points out related topics suitable for further research.

2. Related Work Review

Older people want to remain independent while they have no disabilities that prevent them from doing so, or while their level of disability is manageable. In the U.S. alone, about 80% of those over 65 are living with at least one chronic disease, and an estimated 5.4 million senior citizens are suffering from Alzheimer's disease [15]. To this end, AmI is a very useful tool, as it enables people to continue to live in their homes with sufficient confidence and security. Regarding the AmI, there are some interesting contributions, such as a theoretical one looking to establish an ontology between the two dimensions of AmI and smart buildings [16]. In other cases, the contribution is very narrow, and the studies emphasise a particular aspect, as in [17], which aims to control an LED system for lighting under more user-centred criteria, or the work in [18], which promotes learning from user behaviour to reduce energy usage.

Sometimes the AmI proposal involves specific components, such as a fuzzy embedded agent-based approach for realising AmI in intelligent inhabited environments for controlling the heating, ventilation and air conditioning (HVAC) subsystem of the smart building [19]. In other cases, the contribution is much more referential and it proposes a new way of understanding the phenomena, such as the proposal of a framework aiming to coordinate a set of agents for the management of the energy in a smart building [20]. There are also contributions that are very specific with regard to comparing features based on the IoT with people's behaviour. One example is found in [21], which is concentrated on devices that measure by actigraphy instead of polysomnography and is looking to identify similarities between people with regard to their movements during periods of sleep. It is focused on a specific ethnographic community of people, and, therefore, could face issues to do with privacy.

There are also several interesting surveys, such as the one in [22], which presents its findings under five groups: privacy and security, sensing, reasoning, acting and human–computer interaction. There are some others, such as the one in [23], which emphasises the relationship between AmI and artificial intelligence (AI) to help identify and learn about user profiles, recognizing recurrent users' behaviours, etc. Unfortunately, almost all the mentioned contributions were related to rather simplistic configurations with high levels of intrusive screening, including charge-coupled device (CCD)-based tracking, etc. Indeed, there are authors that seek to identify the most convenient tools, which emerged from the literature review as the most suitable and widespread, by comparing them to identify the interaction mechanisms that end users appreciate most [24]. To this end, there are also relevant

contributions coming from ontology fields looking to provide insights to different contexts, such as hotels [25], etc.

Some researchers have identified smart homes as instances of AAL technologies. They have suggested that these AAL technologies can facilitate care, and consequently improve the quality of service provided to users [26]. Almost all the works in this field are based on the principle of event-condition-action (ECA), mainly due to the use of user interfaces for triggering actions from a program. AAL systems must represent many different types of contextual information, such as sensor information, activity structure, user profiles, temporal information such as environmental data, and spatial information such as the residence layout and its surroundings [26]. In this paper, the relevant AAL areas are identified as activity recognition, context modelling, anomaly detection, location and identification, and planning. For instance, planning makes it possible to schedule daily plans and to create flexible daily reminders to help dementia patients to carry out their daily activities [27]. Planning and scheduling techniques can be categorised into classical approaches such as forward and backward searches or graph-based analysis, decision-theoretic techniques such as Markov decision processes, hierarchical techniques, as well as those that are suggested by rule-based systems.

It is worth highlighting the contribution of Hao et al. [28], who focused on activity recognition as one of the most important prerequisites for smart home applications. They clearly identified that it is a challenging topic due to the high requirements for reliable data acquisition and efficient data analysis, which also emphasises the noise issue we have already highlighted. The critical factors that were identified as aggravating the complexity of recognition are the heterogeneous layouts of smart homes, the number of residents, and varied human behavioural patterns. Therefore, most human activity recognition systems are based on an unrealistic assumption that there is only one resident who is performing all the activities.

Most of the contributions to anomaly activity detection have been used for detecting anomalies in daily activities or medication compliance, by using rule-based techniques, similarity-based techniques or temporal relation discovery techniques. There was no transfer of knowledge involved as the learning activity recognition came from the same context. Therefore, learning in real noisy environments, identifying significant patterns and looking not just to fire rules to drive actions, but to increase the existing knowledge by using a general-purpose non-invasive sensing strategy could be interesting. It can be valuable to infer long-term bias with some degree of anticipation. To the best of our knowledge, this approach has not been analysed in depth previously.

Regarding the detection of human activity indoors, several approaches are currently being studied. Regarding human presence/movement detection, three main technologies are the current focus of analysis in the literature: environmental data-fusion, image/video and Wi-Fi/Bluetooth. Image-video-based solutions have been stated to obtain high-accuracy results; however, they suffer from privacy-related issues. Applications of this kind are presented in [29], where image-based detection is used to determine the occupancy of various university buildings. Wireless communication signals are also being used in two different ways. Firstly, they are used to count users with connected devices by analysing the footprints of the detected devices, as in [30]. Secondly, Wi-Fi signal perturbations are being used to infer presence and movement, as stated in [31,32]. Finally, environmental data-fusion refers to the use of environmental values such as temperature, relative humidity, CO₂, etc., to infer additional information by using data-fusion and other data mining techniques, as shown in [33,34].

AAL technologies have also been studied to develop alarm-based monitoring systems to notify users of non-ordinary actions and behaviours in older people and people with physical disabilities. In the Italian project HDOMO, they developed an automated domestic environment that made it possible for people to continue being autonomous in their private environment. The behaviour of home inhabitants was monitored by using low-cost passive sensors alongside depth cameras. They developed a series of alarms that detected unusual behaviours such as falls, sleep apnoea, incorrect feeding, etc. [35]. Other approaches, such as the one presented in [36], define systems that

learn from room-to-room transitions and permanence to detect behaviour changes that may be related with health problems, which are communicated via alarms. Apart from sensors installed in built environments, the application of wearable devices has also been studied to detect attention-requiring actions (e.g., low blood pressure), enabling monitoring to continue outside homes [37].

Based on the findings from the literature review, the authors agreed with the the major limitations reported, which are that recognition activities are based on unrealistic hypotheses, and mainly used in highly monitored environments that are difficult to generalise. Therefore, this study selected a multi-human environment with different behavioural patterns as the source for learning user practices (see Section 3). Indeed, we suggest the idea of using AI techniques to deal with non-intrusive and non-human dedicated sensing activities to contribute to certain human activity recognition, when some context exists which provides insight for location identification in order to make it possible to use transfer learning by applying classification pattern analysis to help in anomaly detection for other contexts. In the next section, more details are provided regarding the selected context.

After analysis of the contributions, we can identify a gap related to the lack of consideration of identification of behavioural rules people adopt when different time scales are involved. Our vision is that, because such rules involve created knowledge, bias from the regular behaviour of people can be identified, and, by understanding their significance, potential negative effects or risk can be mitigated. This is of crucial importance in the case of ageing people living alone, as it can lead to early warning of potential approaching difficulties such as memory loss, unsafe usage of facilities, etc. Therefore, this paper aims to contribute to the research in this field.

3. Materials and Methods

To evaluate the potential of such rule discovering capabilities, we selected a convenient context. The research environment was carried out in a care home for ageing residents with dementia who are in long-term care.

The Technology Acceptance questionnaire was not administered as such, because the non-invasive design for the experience and the characteristics of the residents. Instead, an informed consent was given by each participant or a responsible family member stating that only anonymous data would be used for epidemiological, public health, educational or research purposes.

The care home allowed us to install and monitor non-intrusive sensors in six different rooms of the psychogeriatric ward located in the second floor of the building: the Main Hall, the living room, the dining room, and three bedrooms spread over the same independent cohabitation unit. The three monitored rooms were chosen because the residents with major behaviour problems resided there and, therefore, they were in need of more assistance and attention. Among the incidences that sometimes occur are agitation, alteration, screams, verbal aggression, physical aggression and inappropriate behaviour. The site layout plan with the monitored rooms depicted on a grey background is shown in Figure 2.

Thirteen residents were living in the cohabitation unit during the year in which data were collected. The residents were seven men and six women with an average age of 84 years. A predetermined timetable of the care home provided clues about the residents' normal location, making abnormalities stand out more clearly. Residents were in bed between 22:00 and 08:00. From 8:00 on, residents were woken up, cleaned and dressed by the nursing assistants. They spent most of the day between the Main Hall and the living room, connected by a narrow passage. They usually only accessed the Dining Room during meals (8:30–10:30, 13:00–14:00 and 19:00–20:00). Bedrooms were occupied at night, although they may have occasionally been used during the day (e.g., nap after lunch), and since residents can receive visits from relatives and the bedrooms closest to the living room are not locked. Almost half of the residents need to be moved from one space to another because they do not move on their own. On the other hand, other residents roam freely.

Figure 2. Area of interest: A cohabitation unit in the psychogeriatric ward of an elderly care home.

The six monitored rooms were equipped with identical wireless sensors to measure and report levels of ambient temperature ($^{\circ}\text{C}$), relative humidity RH (%), CO_2 concentration (ppm) and passive infrared (PIR) technology-based motion detection. The temperature, relative humidity and CO_2 levels were collected by multisensor devices mounted on the wall of each room in the “breathing zone”. The “breathing zone” was located at a height of 1.5 m above the floor in the Main Hall, the living room and the dining room, and at a height of 1.2 m, next to the bedside tables, in the bedrooms. The PIR sensors were mounted on the ceiling, in the middle of the shared lounges, and above the bed area in the bedrooms. These sensors were primarily designed for the control of lighting, heating, ventilation, and air conditioning systems (HVAC) and characterised as being easily installable and non-invasive. During normal operation, the sensors harnessed light energy from the environment through integrated solar cells and transmitted the measured data to a central node that uploaded them into the cloud. At night or in low light, backup batteries provided additional power to ensure the proper operation of the sensors. All sensors support the EnOceanTM standard protocol (ISO/IEC 14543-3-10) for low-power wireless communications.

The data acquisition was carried out with respect to the national laws regarding non-intrusiveness, privacy and confidentiality over the period of one year. The multisensor devices collected environmental data regarding the temperature, relative humidity and CO_2 levels at a regular frequency that could decrease during extended periods of low light due to low battery charge. However, the ceiling-mounted PIR motion sensors operated as room occupancy gauges by detecting when a room was occupied or vacant. These PIR sensors did not send the motion detection data periodically but when the room occupation status changed with minimum pre-configured delays. In this way, when motion was detected, the data indicating the occupied statuses were transmitted immediately and no more data were sent in the 2 min that followed. Then, the PIR sensor transmitted an occupancy message upon detection of a new motion. On the contrary, if no motion was detected for an extended period of 10 min, an unoccupied message was sent. These pre-configured delays optimised the energy consumption of the PIR sensors without affecting the effective control of the HVAC and lighting systems; however, they made data analysis more challenging as no information regarding the room occupancy was sent between delays. Therefore, the sampling frequency of the environmental sensors was finally established to be 10 min because that frequency matched the restrictions of the PIR sensors, that do not send an unoccupied message until ten minutes have elapsed since the last motion detection. In addition, the power generation requirements under normal lighting conditions do not allow the frequency to be much higher than ten minutes and indoor environmental values usually remain stable without sudden changes.

The installed sensors provided information on the quality of the environmental parameters in which residents lived, as well as information regarding their movements. A deeper analysis to discover behaviour and mobility patterns of the residents and the integration of additional information, such as the timetable of the care home, could reveal aspects associated with the evolution of the disease.

In general, the identification of either sudden or progressive changes in their routines helps to provide specific care to the residents.

The data collection provided more than 50,000 patterns, considering the raw sensor data. Careful labelling process was carried out for 16,900 patterns to be used for the presence data model, and about 560 patterns for the window opening process, as it is a less frequent event and limited unbalanced datasets were used. The patterns used for the presence model were randomly segmented into 11,323 for training and 5577 for testing. In a similar way, the 560 patterns for the opening window model were segmented into 392 for training and 168 for testing. It became clear that specific effort was required for labelling the patterns, therefore it was indicated that the chosen case study had a partially labelled dataset.

After the description of the site, the sensing strategy adopted, and the data collection carried out, the method used for data preparation is presented here. Note that we were interested in developing software-based sensors that were able to determine and formalise new knowledge. Two different models were proposed, the first one was regarding the presence or absence of people inside a room during a significant period of time, only taking into account the patterns of CO₂, temperature and humidity evolution. The second model was to detect when the window in the room was opened or closed. As described above, the environment was very noisy, not only because of the occasional failure of the sensors but due to additional unforeseen behaviours such as nurses leaving doors opened, forced ventilation rules, diffusion of air (as well as the related physical properties) between spaces, which all occurred without any clear pattern.

To be able to support the conclusions found in an unbiased context, data preparation adopted the removal of sequences of days from the datasets for learning purposes. These removed data became the elements that were used to test the knowledge gathered after the training processes had ended. For the remaining data in the datasets, as several configurations required testing, a cross-validation folder approach was adopted to decide the most suitable modelling configuration.

Indeed, regarding the context awareness discussed in the state of the art AAL approaches, it was decided to incorporate some degree of semantic description of the spaces, making it possible to be responsive not only to the indoor variations of parameters, but also to those from the next room, thus adding information about the building. The usage of Web Ontology technology enables to setup a clear relationship between the sensor data and the home context towards defining the activity context. It enables not only the description of the entities themselves, but also their relationships in a formal way suitable for being exploited in the model construction [38]. Indeed, services, home owners but also family relatives can be easily defined, as well as time references. Based on the established relationships both geometric but also functional properties can be named. Geometric properties can help in improving the modelling task but functional are in the basis for the rule adoption, as depicted in the Discussion Section.

4. Modelling

The first aspect to be analysed is what is known in the cross-industry standard process for data mining (CRISP-DM) methodology as “data understanding”, in the context of the business understanding. To this end, some pictures of the variables per monitored space based on the expected seasonality, which is a daily behaviour, can help to understand the particularities. From the analysis, it became evident that the relative position of the sensors for the same type of spaces (e.g., bedrooms) was significant in terms of the absolute values. Indeed, the geographical orientation of the spaces inside the building also became relevant, as the relative humidity (RH) and temperature levels were significantly different for different orientations when the bedrooms were considered.

Presence sensors provide very noisy information, as their operational design makes the movement signal last for a long period. On the contrary, if people do not move at all, the sensors indicate the absence of people, which is not true. Indeed, the movements of nurses and other working people

contribute to their activation. In the end, and after various filtering strategies, it was decided that the PIR information was mainly useless because of its high noise level.

However, those sensors allowed us to appreciate when ventilation occurs during room cleaning in the morning as RH spiked and temperature also became impacted. Additionally, a reduction of CO₂ was seen when users left a room.

Overall, it was evident that RH and temperature were strongly regulated by the building control system and they were, therefore, not very informative. However, the CO₂ sensors were still useful as they provide information regarding the room occupancy. The basic principle of its behaviour is that the CO₂ concentration increases over time, as the person sleeping inside is continuously producing it. It is also expected that when the room becomes empty its concentration will be lower than when it is in use. Obviously, real behaviour is slightly more complex because sometimes the door to the room remains open when the room's user is inside and diffusion happens, which moderates or reverses the growth in CO₂ concentration.

By inspecting Figure 3, it can be seen that the users of bedrooms 1 and 2 spent some time resting after lunch and that this was not the case for the user of bedroom 3. The figure also shows that, in bedroom 3, the CO₂ level began to increase after dinner time (20:30), and during dinner, the CO₂ flowed from the dining room to bedroom 3 when the room's door remained open due to its proximity. These effects are relevant and reinforce the need for the semantic description of the building. By observing the same levels during night time, it becomes clear that the CO₂ concentration does not increase in bedrooms (e.g., bedroom 2) in the first part of the night, whereas some increases occur in the living room and in the main hall. This behaviour changes around 07:00 when the nurse visits the room and the bedroom's door is closed. Then, CO₂ starts to increase again. If the focus becomes the dining room, it is possible to identify breakfast time (short period of time around 08:30), lunch time (larger period of time around 13:00), and dinner time (at around 19:00). The Figure 4 shows a rather similar behaviour for the same spaces, but in a different point in time.

Figure 3. Sensor variations per room on 31 October 2017.

Figure 4. Sensor variations per room on 25 October 2018.

The previous analysis makes it evident that some knowledge can be derived from the sensors' data. Mainly, the rules that were found were related to the room occupancy for users, where the aggregation of data from sensors over time must be significant in order to deal with the previously highlighted local variations.

Therefore, we were interested in building models that inferred the use of the spaces, by considering the level of CO₂, the slope of its variation, the values from the neighbouring rooms, and the RH variations. To adequately learn from the data, a smooth data function was required, and therefore a moving average of five values was used.

Then, a model for presence in bedrooms was trained by preparing a dataset that involved the following variables:

- CO₂ concentration level in the room.
- CO₂ concentration level in the contiguous room.
- Slope of the regression linear model for RH level in the room considering three points before the present time and three points after.
- Slope of the regression linear model for CO₂ level in the room considering three points before the present time and three points after.
- Slope of the regression linear model for CO₂ level in the contiguous room considering three points before the present time and three points after.

Labels for the outcome variable were adopted according to the reported information available from the care home.

For the model related to the window status, the variables under consideration were:

- CO₂ concentration level in the room.
- Relative humidity in the room (RH).

- Slope of the regression linear model for RH level in the room considering three points before the present time and three points after.
- Slope of the regression linear model for CO₂ level in the room considering three points before the present time and three points after.

As already indicated in the last part of the Materials and Methods Section, to preserve the independence for validating the learned model, the information from January 2018, after the 10 January 2018, was removed from the dataset and reserved for testing purposes. The remaining part of the dataset that was used for learning purposes involved 11,323 samples with an unbalanced ratio between classes of 66%/33% (i.e., presence/absence from the bedroom).

The goal when addressing artificial intelligence techniques is not just to find models that can produce accurate predictions, but also can be used to extract knowledge in an intelligible way, i.e., can be understood as being able to explain the reasoning logic behind the provided outcome.

Although there are many techniques and algorithms helpful for this purpose, tree-based methods stand as one of the most effective and useful method, capable to produce both reliable and understandable results, on mostly any kind of data [39]. Therefore, although different methods were explored such as naive Bayes, support vector machine and feed forward back propagation neural networks, the final decision was to use random forest technique, as a good representative technique for the tree based classification algorithms. As a summary for their benefits are:

- Decision trees are non-parametric. They can model arbitrarily complex relations between inputs and outputs, without any a priori assumption. Indeed, these methods do not require variable standardisation and other transformations, which can hinder future usage as values can go out of the adopted range.
- Decision trees handle heterogeneous data (ordered or categorical variables, or a mix of both). This become an interesting characteristic for classification problems.
- Decision trees intrinsically implement feature selection, making them robust to irrelevant or noisy variables (at least to some extent). Actually, it is a convenient characteristic that was used in this study to identify the most contributing variables. Indeed, they help to understand the path for the provided outcome.
- Decision trees are robust to outliers or errors in labels. This characteristic becomes critical in real life applications, when sensors can have issues and provide either wrong or biased measures.

Therefore, the adopted modelling approach was the random forest classifier [40] because of its flexibility, ensemble capabilities, and bagging properties. During the tuning process, different parameters were tested, such as the number of variables considered for the trees (between two and five were analysed and two was the value finally adopted), the number of trees considered for the ensemble (between 100 and 600, where 500 was selected) and a strategy of a 10-fold cross-validation was adopted to avoid bias due to specific pattern selection for training. The same strategy was used for both models and the adopted parameters were the same.

Creating a classification and regression tree (CART), involves selecting input variables and split points on those variables until a suitable tree is constructed. The selection of which input variable to use and the specific split or cut-point is chosen using a greedy algorithm to minimise a cost function. It is a recursive binary splitting and tree construction ends using a predefined stopping criterion, such as a minimum number of training instances assigned to each leaf node of the tree.

The random forest technique relies on the consensus carried out by different trees (the forest). When the training set for the current tree is drawn by sampling with replacement, about one-third of the cases are left out of the sample. Each tree is constructed using a different bootstrap sample from the original data. About one-third of the cases are left out of the bootstrap sample and not used in the construction of the k th tree. These out-of-bag data are used to get a running unbiased estimate of the classification error as trees are added to the forest. It is also used to get estimates of variable importance.

In random forests, there is no need for cross-validation or a separate test set to get an unbiased estimate of the test set error. It is estimated internally, during the run of the algorithm, which makes it even less dependent on the configuration decisions.

After each tree is built, all of the data are run down the tree, and proximities are computed for each pair of cases. If two cases occupy the same terminal node, their proximity is increased by one. At the end of the run, the proximities are normalized by dividing by the number of trees, and they become useful in replacing missing data and locating outliers [40].

The objective here is not to develop a specific study regarding the capabilities of machine learning to represent the knowledge but to use almost-standard and sufficiently robust techniques to capture and properly represent this new knowledge. The new knowledge distilled is represented by the two models built, and as far as such information was not measured but just inferred by the models from the general purpose sensors, they are seen as soft sensors. This soft sensors become key elements regarding different problems, as part of the assessment for the construction of the behavioural rule.

5. Results

The two selected models were trialled under a strategy of 10-fold cross-validation with minimal adjustments for the selected variables as per the split of the standard recursive partitioning algorithm used, which starts with all the data and performs an exhaustive search looking for possible split points to find the one that best explains the entire data, or, in formal terms, that reduces the node impurity the most. Cross-validation ensures that the averaged performance being considered is independent from the set of data.

The results show that, for this particular case, it was best to only consider two variables, as depicted in Table 1, where the accuracy represents the percentage of correctly classified instances out of all the instances, and Cohen's Kappa essentially depicts the classification accuracy, except it is normalised at the baseline of random chance on the dataset.

Table 1. Cross-validated performance by considering different numbers of variables regarding the presence model.

Number of Variables	Accuracy	Kappa
1	0.9935522	0.9851646
2	0.9963665	0.9916528
3	0.9960815	0.9909948
4	0.9961171	0.9910764
5	0.9954046	0.9894418

Validation of the trained models was carried out using data from the same room but measured during a period after the one used for training. Therefore, full independence of the data was adopted and the results are presented in Table 2.

Table 2. Contingency table for the presence classifier.

	Real Room Absence	Real Room Presence
Predicted room absence	2510	18
Predicted room presence	86	2963

From the results in this table, obtained from data not previously seen as the data came from the part of the dataset that was omitted for validation, it becomes clear that the accuracy of the classifier was high (F1-score: 0.984).

After validating the coherence of the model, it was then necessary to determine which variables the created knowledge was based upon, which is presented in Table 3.

Table 3. Relevance of variables for the presence model. Gini coefficient is related to Mean Decrease in Impurity.

Variable	Absence (%)	Presence (%)	GINI Decrease
CO ₂ in Bedroom 2	31.25	38.47	4644.5
CO ₂ in Living Room	25.03	32.24	3467.6
RH slope	20.53	13.14	2627.8
CO ₂ slope in Bedroom	11.40	8.87	746.9
CO ₂ slope in Living Room	11.79	7.28	733.6

To this end, it becomes clear that the presence model has a high accuracy level, and the highest impact for the decision-making process regarding the presence/absence of the user in the bedroom was the CO₂ concentration (with a significant contribution from the variation rate (i.e., the slope) of this concentration). There were some other relevant effects we were unable to consider, such as the effective contribution of the forced ventilation system with regard to the variation in bedroom temperature, etc., as such data were not accessible. The window model was also considered and the results are presented in Table 4.

Table 4. Contingency table for the opened window classifier.

	Real Room Absence	Real Room Presence
Predicted room absence	97	8
Predicted room presence	5	58

For this model, the accuracy metrics (F1-score) was equal to 0.935 and it was considered acceptable. Indeed, it was possible to understand upon which variables the knowledge created was based, which is presented in Table 5.

Table 5. Relevance of variables for the window model.

Variable	Closed (%)	Opened (%)	GINI Decrease
CO ₂ in Bedroom	58.79	48.55	73.44
RH in Bedroom	50.67	54.98	69.06
RH slope	55.75	48.47	85.07
CO ₂ slope in Bedroom	63.22	50.96	100.75

The built models allowed for additional knowledge by means of the new features to be obtained. These features, which are not directly measured by the existing sensors, allow for estimation of the room occupancy time and the ventilation routine of the room to be carried out on a daily basis. Indeed, in Table 6, the models were applied to the data collected during several days and the new information was obtained.

Table 6. New knowledge creation for bedroom 1.

Date	Inside from	Inside Until	Opened Window
26 January 2018	20:35	08:05	08:10/09:15
27 January 2018	21:05	08:15	08:20/09:05
...	...	...	...

Based on the information obtained from the models, it becomes possible to estimate frequent behaviours, such as the routine for ventilation based on the time of opening the window in the room (see Figure 5).

Figure 5. Frequency of room ventilation based on the opening of the window.

Furthermore, it can be estimated how long such process lasted, as shown in Figure 6.

Figure 6. Duration of ventilation activities.

This information can be refined with regard to a specific period such as in summer or in winter, or for working days and weekends. Therefore, behaviour can be compared against that which was exhibited during the preceding days and it can be determined when it becomes unexpected. For instance, one rule could read as follows, “if during last three days the ventilation last more than the maximum threshold then raise an alert” or “if more than three times during last ten days there are absences from the room lasting more than 10 min each between 22:00 and 07:00, then alert”. This is because such behaviours are understood as potential signals for behavioural changes that needed to be analysed. All such rules can be implemented in the reasoning layer, named “High Level Rules” according to the proposed framework in Figure 7.

To provide a consistent context, capable of managing the structure proposed in the depicted framework, both for handling the complexity of different rules, and their updating process, and for enabling the regular accuracy monitoring of the built soft sensors and launched the reinforced learning processes when required, it is proposed to adopt the usage of Web Ontology Language (OWL), in a particular semantic context expression, given the basis of satisfying context-aware reasoning. Particular attention is deserved to the rules designed with Semantic Web Rule Language (SWRL) [38].

Figure 7. Referential framework proposed. AAL, ambient assisted living; AI, artificial intelligence.

The approach will typically enable the implementation of effective strategies to help in the management of different use cases in a flexible way, as the resulting ontologies are essentially shared knowledge models that enhance the capabilities of automated processing and the level of automation by allowing machines or agents to interpret data/information and reason against ontological contexts, thus enabling knowledge based intelligent decision support [41]. Applications have also been reported in the context of IoT, as in [9], where similar architecture was proposed for Cruise cabins.

6. Discussion

As explained in the previous sections, the a priori effect of a human presence on the temperature and humidity of a room is limited by artificial regulation systems. However, we were able to detect window openings in the form of high immediate increases in humidity, which of course were influenced by the climate in the monitored zone where normal humidity values were between 70% and 80%.

In contrast with temperature and humidity, CO₂ values were shown to be highly correlated with human presence and, that being the case, this was the parameter used to determine room occupancy in this study. Not only were the on-site values used to determine human presence but measurements in nearby large spaces were also used, as they had a great effect on the contiguous rooms. This was especially important when the doors were opened and the diffusion of air between spaces was more notable. In this way, information regarding the physical distribution of the building was added to the models.

The aim of this work was to enable the detection of regular behaviours to aid in the monitoring of patterns, such as occupancy or ventilation, based on normal environmental sensors which are already present in many spaces for HVAC control, etc. When this purpose was achieved, the information gathered by these sensors could be modelled to gain additional information without adding more sensors or infrastructure to the building. To validate our proposal, additional sensors were installed in the form of PIR sensors to gather basic knowledge of binary presence (i.e., occupied/unoccupied) in the rooms. The aim here was to avoid installing these supplementary devices while being able to

obtain the knowledge they provide. Indeed, this was a significant conclusion because the sensor does not regard presence but movement.

The viability of using machine learning models, in this particular case by using the random forest method, for this problem was demonstrated for different problems, obtaining high accuracy values (see Tables 1 and 4), when a 10-fold cross-validation strategy was adopted.

All the possible configurations of the data that could be used in the models were tested, and, finally, the one that obtained the most accurate results was the model using trees with two variables in the forest. The trained model with just two variables was validated with previously unseen data, obtaining an accuracy score of 98% (the confusion matrix is shown in Table 2) for the room occupancy and 93% for the ventilation model. With the results obtained, the presented approach is considered to be valid to determine the presence in rooms in which the data were used to train the models. However, further work is needed to generate a unique model that could be transferred to other rooms.

Based on the knowledge gathered, it should be possible to transfer the knowledge obtained in one room to others, by using the model to estimate the presence of people in other bedrooms. However, because of the relevance established in Table 3 for the different types of variables, it becomes clear that absolute levels of CO₂ have strong relevance. Therefore, the model cannot be attributed to specific rooms and it provided unacceptable failure rates when applied to different rooms.

When transferability becomes a requirement for the knowledge gathered, a different strategy must be adopted (e.g., variable normalisation), in order to make it possible to compare scenarios. The size of the rooms could also be needed for such normalisation, as the effect of one person's presence does not affect a bedroom or the Main Hall in the same way due to their size difference. As this aspect is a relevant factor, specific tests were performed to analyse the effects when learning was transferred. To this end, a normalisation between 0 and 1 for the CO₂ and RH values per room was performed (as effective relative volumes of rooms were not available). The same variables were defined, including the slopes of the normalised variables and the same type of random forest model was built, with exactly the same set of values for training and with the same strategy for validation. Therefore, the results can be interpreted using the same contingency table as Table 2 in comparison with the actual contingency table (see Table 7), and it was concluded that the quality was similar for both models.

Table 7. Contingency table for the normalised classifier.

	Real Room Absence	Real Room Presence
Predicted room absence	2517	18
Predicted room presence	79	2963

Subsequently, it was possible to transfer the model to Bedroom 1. This was particularly significant, as the behaviour observed in Bedroom 1 was less defined than in the other bedrooms (see Figure 3). The contingency table describing the performance of the model is presented in Table 8.

Table 8. Contingency table for the “transferred” normalised classifier.

	Real Room Absence	Real Room Presence
Predicted room absence	9976	774
Predicted room presence	468	16279

Indeed, some improvements could be made when reinforcement learning was considered. However, even though the performance was slightly reduced, the value of transferring knowledge can still be easily appreciated. The F1 score as averaged value per class was reduced from 0.984 in the proper context up to 0.84 when transferring was considered.

It is also necessary to recognise that the use of these sorts of models requires a certain level of abstraction because, in the creation of the synthetic variables, a time summarising approach must be

adopted. As previously commented, values within $(t - 30 \text{ min}, t + 30 \text{ min})$ were used to calculate the slopes for moment t . Therefore, the decision was made based on hourly behaviour instead of having a prediction at the same frequency as the sampling rates.

In addition to transferring models between contexts, which sometimes requires the application of reinforced learning from a small amount of data coming from the present room in addition to the already trained model in another space, an additional use can be derived from the created knowledge. This involves the establishment of rules by comparing the learned long-term behaviour with the present one. As an example, by comparing the starting routine for performing room ventilation (see Figure 5) with the recent and short-term (for instance last week) pattern, potential advice can be derived and, when combined with other factors, such as variations in the room usage, etc., they can alert against early effects that can require extra investigation. Another potential rule could also consider the ventilation time used in connection with the outdoor weather conditions. The criteria for raising alerts will depend on firing several of these rules at the same time, in order to collect stronger evidence for any change in behavioural patterns.

Through the studied case it becomes pretty clear that the quality of the derived knowledge strongly depends on the labelling processes required both for training at the beginning and for reinforced learning in the case of update of the behaviour or transference of the models to a different environment. Actually, this dependency becomes an opportunity to offer the people being monitored to participate and to be at the same time users but also co-creators of the Aml technology. If a convenient and simple phone app is enabled to ask the participants willing to adopt such active role, the regular labelling process can be enriched dramatically, and full set of benefits can be derived, at the system level but also at individual level. To foster such dynamics and provide a smart system just querying when relevant events happen, and also for making a consistent management of the produced rules possible, the assistance from the semantic perspective becomes essential. Therefore, the natural evolution for using such soft sensors is to integrate them into a semantic schema able to provide a holistic perspective.

7. Conclusions

The ability to determine human presence in the spaces of a building without adding additional sensors in a non-intrusive way could be applied to detect behaviours that are not understood to be regular in cases where regular patterns can be established, such as in the case of elderly people living alone or in cases such as the one described in this paper, where patients display very regular behaviour day-to-day. The acquired knowledge allows the authors to propose the reference framework depicted in Figure 7.

Either from the direct model outcome or by using transferring strategies, the newly created knowledge can be seen as a kind of aggregated soft sensor where rules can be derived, such as the number of hours per weekday the user sleeps or the number and duration of absence times during the night time per weekday. Such newly derived knowledge can be used to establish alerts or triggers. This is significant as these alerts can inform the user of any unexpected long-term deviation from regular behaviour (e.g., month-over-month). This kind of soft surveillance will help to identify mental diseases or certain other difficulties in their early stages. The applications of the proposed framework will be further analysed in future research.

Although limitations exist because of the sample sizes and the strict context needed for labelling the samples helping in producing the supervised learning, the research carried out still shows strong capabilities for building soft sensors, and their advantage to establish variations in a different time scale, which can be useful for higher level behavioural rules. To this end, integration with semantic schema need to be deeply investigated in future research, in particular the rule management and updating procedures according to the confidence estimations.

In addition, in terms of limitations, it is relevant to emphasise that there is a bias towards the benefits and the positive outcome of the AAL technologies in particular in the case of aging people.

This field is named as gerontechnology [42]. In this sense, the present paper revolves on the same idea of becoming useful to the relatives and the older people as well. However, it is worth recognising the lack of enough social studies with an agnostic lens trying to identify negative impacts of AAL over such collective, either because of the opt-out or just because of disruptions because of unsuccessful integration among components or AAL programs. Indeed, derived effects such as increased levels of social isolation because of the technology adoption will require further attention.

Author Contributions: Conceptualisation, M.Á.A. and J.O.-M.; Methodology, U.S.; Modelling, J.O.-M. and S.S.; Discussion, J.O.-M. and S.S.; Validation, all authors; Formal Analysis, U.S.; Data Curation, U.S.; Writing—Original Draft Preparation, J.O.-M. and S.S.; Writing—Review and Editing, M.Á.A. and U.S.; and Project Administration, M.Á.A.

Funding: This research was partially funded by Fundación Tecnalia Research & Innovation, and J.O.-M. also wants to recognise the support obtained from the EU RFCS program through project number 793505 ‘4.0 Lean system integrating workers and processes (WISEST)’ and from the grant PRX18/00036 given by the Spanish Secretaría de Estado de Universidades, Investigación, Desarrollo e Innovación del Ministerio de Ciencia, Innovación y Universidades.

Acknowledgments: The authors want to recognise the helpful contribution from the Editor and anonymous reviewers for the help provided.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AmI	Ambient Intelligence
AAL	Ambient Assisted Living
IoT	Internet of Things
HVAC	Heating, Ventilation and Air Conditioning
AI	Artificial Intelligence
RH	Relative Humidity

References

1. United Nations Department of Economic and Social Affairs. *World Population Ageing (Highlights) (ST/ESA/SER.A/397)*; United Nations Department of Economic and Social Affairs: New York, NY, USA, 2017.
2. Kamienski, C.; Jentsch, M.; Eisenhauer, M.; Kiljander, J.; Ferrera, E.; Rosengren, P.; Thestrup, J.; Souto, E.; Andrade, W.S.; Sadok, D. Application development for the Internet of Things: A context-aware mixed criticality systems development platform. *Comput. Commun.* **2017**, *104*, 1–16. [[CrossRef](#)]
3. Roccotelli, M.; Nolic, M.; Fanti, M.P.; Ukovich, W. Internet of things and virtual sensors for electromobility. *Int. Technol. Lett.* **2018**, *1*, e39. [[CrossRef](#)]
4. TrustArc Inc. *Internet of Things Industry Brings Data Explosion, but Growth Could Be Impacted by Consumer Privacy Concerns*; TrustArc Inc.: San Francisco, CA, USA, 2014.
5. AlHogail, A. Improving IoT Technology Adoption through Improving Consumer Trust. *Technologies* **2018**, *6*, 64. [[CrossRef](#)]
6. Spoladore, D.; Sacco, M. Semantic and Dweller-Based Decision Support System for the Reconfiguration of Domestic Environments: RecAAL. *Electronics* **2018**, *7*, 179. [[CrossRef](#)]
7. Scanail, C.N.; Carew, S.; Barralon, P.; Noury, N.; Lyons, D.; Lyons, G.M. A review of approaches to mobility telemonitoring of the elderly in their living environment. *Ann. Biomed. Eng.* **2006**, *34*, 547–563. [[CrossRef](#)] [[PubMed](#)]
8. Saralegui, U.; Antón, M.A.; Ordieres-Mere, J. Taking advantage of an existing indoor climate monitorization for measuring occupancy. *WSEAS Trans. Environ. Dev.* **2017**, *13*, 327–334.
9. Nolic, M.; Spoladore, D.; Carciotti, S.; Buqi, R.; Sacco, M. Cabin as a Home: A Novel Comfort Optimization Framework for IoT Equipped Smart Environments and Applications on Cruise Ships. *Sensors* **2019**, *19*, 1060. [[CrossRef](#)]

10. Alzheimer's, A. 2015 Alzheimer's disease facts and figures. *Alzheimers Dement. J. Alzheimers Assoc.* **2015**, *11*, 332.
11. Brookmeyer, R.; Johnson, E.; Ziegler-Graham, K.; Arrighi, H.M. Forecasting the global burden of Alzheimer's disease. *Alzheimers Dement.* **2007**, *3*, 186–191. [[CrossRef](#)]
12. Van Ooteghem, K.; Musselman, K.; Gold, D.; Marcil, M.N.; Keren, R.; Tartaglia, M.C.; Flint, A.J.; Iaboni, A. Evaluating Mobility in Advanced Dementia: A Scoping Review and Feasibility Analysis. *Gerontologist* **2018**. [[CrossRef](#)]
13. de Vugt, M.E.; Verhey, F.R. The impact of early dementia diagnosis and intervention on informal caregivers. *Prog. Neurobiol.* **2013**, *110*, 54–62. [[CrossRef](#)] [[PubMed](#)]
14. Mahurin, R.K.; DeBettignies, B.H.; Pirozzolo, F.J. Structured assessment of independent living skills: Preliminary report of a performance measure of functional abilities in dementia. *J. Gerontol.* **1991**, *46*, P58–P66. [[CrossRef](#)] [[PubMed](#)]
15. Rashidi, P.; Mihailidis, A. A survey on ambient-assisted living tools for older adults. *IEEE J. Biomed. Health Inform.* **2013**, *17*, 579–590. [[CrossRef](#)]
16. Stavropoulos, T.G.; Vrakas, D.; Vlachava, D.; Bassiliades, N. BOnSAI: A smart building ontology for ambient intelligence. In Proceedings of the 2nd International Conference on Web Intelligence, Mining and Semantics, Craiova, Romania, 6–8 June 2012; p. 30.
17. Tan, Y.K.; Huynh, T.P.; Wang, Z. Smart Personal Sensor Network Control for Energy Saving in DC Grid Powered LED Lighting System. *IEEE Trans. Smart Grid* **2013**, *4*, 669–676. [[CrossRef](#)]
18. Dounis, A.I. Artificial intelligence for energy conservation in buildings. *Adv. Build. Energy Res.* **2010**, *4*, 267–299. [[CrossRef](#)]
19. Ahmad, M.W.; Mourshed, M.; Yuce, B.; Rezgui, Y. Computational intelligence techniques for HVAC systems: A review. In *Building Simulation*; Springer: Berlin, Germany, 2016; Volume 9, pp. 359–398.
20. Stavropoulos, T.G.; Rigas, E.S.; Kontopoulos, E.; Bassiliades, N.; Vlahavas, I. A multi-agent coordination framework for smart building energy management. In Proceedings of the 2014 25th International Workshop on Database and Expert Systems Applications, Munich, Germany, 1–5 September 2014; pp. 126–130.
21. Iglesias Martínez, M.E.; García-Gómez, J.M.; Sáez, C.; Fernández de Córdoba, P.; Alberto Conejero, J. Feature Extraction and Similarity of Movement Detection during Sleep, Based on Higher Order Spectra and Entropy of the Actigraphy Signal: Results of the Hispanic Community Health Study/Study of Latinos. *Sensors* **2018**, *18*, 4310. [[CrossRef](#)]
22. Cook, D.J.; Augusto, J.C.; Jakkula, V.R. Ambient intelligence: Technologies, applications, and opportunities. *Pervasive Mob. Comput.* **2009**, *5*, 277–298. [[CrossRef](#)]
23. Sadri, F. Ambient intelligence: A survey. *ACM Comput. Surv. (CSUR)* **2011**, *43*, 36. [[CrossRef](#)]
24. Caivano, D.; Fogli, D.; Lanzilotti, R.; Piccinno, A.; Cassano, F. Supporting end users to control their smart home: Design implications from a literature review and an empirical investigation. *J. Syst. Softw.* **2018**, *144*, 295–313. [[CrossRef](#)]
25. Spoladore, D.; Arlati, S.; Carciotti, S.; Nolich, M.; Sacco, M. RoomFort: An ontology-based comfort management application for hotels. *Electronics* **2018**, *7*, 345. [[CrossRef](#)]
26. Amiribesheli, M.; Bouchachia, H. A tailored smart home for dementia care. *J. Ambient Intell. Humaniz. Comput.* **2018**, *9*, 1755–1782. [[CrossRef](#)]
27. Pollack, M.E.; Brown, L.; Colbry, D.; McCarthy, C.E.; Orosz, C.; Peintner, B.; Ramakrishnan, S.; Tsamardinos, I. Autominder: An intelligent cognitive orthotic system for people with memory impairment. *Robot. Auton. Syst.* **2003**, *44*, 273–282. [[CrossRef](#)]
28. Hao, J.; Bouzouane, A.; Gaboury, S. Recognizing multi-resident activities in non-intrusive sensor-based smart homes by formal concept analysis. *Neurocomputing* **2018**, *318*, 75–89. [[CrossRef](#)]
29. Davis, J.A., III; Nutter, D.W. Occupancy diversity factors for common university building types. *Energy Build.* **2010**, *42*, 1543–1551. [[CrossRef](#)]
30. Wang, Y.; Shao, L. Understanding occupancy pattern and improving building energy efficiency through Wi-Fi based indoor positioning. *Build. Environ.* **2017**, *114*, 106–117. [[CrossRef](#)]
31. Sobron, I.; Del Ser, J.; Eizmendi, I.; Velez, M. Device-Free People Counting in IoT Environments: New Insights, Results and Open Challenges. *IEEE Internet Things J.* **2018**, *5*, 4396–4408. [[CrossRef](#)]
32. Ohara, K.; Maekawa, T.; Matsushita, Y. Detecting state changes of indoor everyday objects using Wi-Fi channel state information. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2017**, *1*, 88. [[CrossRef](#)]

33. Chen, Z.; Zhu, Q.; Masood, M.K.; Soh, Y.C. Environmental Sensors-Based Occupancy Estimation in Buildings via IHMM-MLR. *IEEE Trans. Ind. Inform.* **2017**, *13*, 2184–2193. [[CrossRef](#)]
34. Nesa, N.; Banerjee, I. IoT-Based Sensor Data Fusion for Occupancy Sensing Using Dempster–Shafer Evidence Theory for Smart Buildings. *IEEE Internet Things J.* **2017**, *4*, 1563–1570. [[CrossRef](#)]
35. Frontoni, E.; Pollini, R.; Russo, P.; Zingaretti, P.; Cerri, G. HDOMO: Smart Sensor Integration for an Active and Independent Longevity of the Elderly. *Sensors* **2017**, *17*, 2610. [[CrossRef](#)]
36. Eisa, S.; Moreira, A. A Behaviour Monitoring System (BMS) for Ambient Assisted Living. *Sensors* **2017**, *17*, 1946. [[CrossRef](#)] [[PubMed](#)]
37. Lloret, J.; Canovas, A.; Sendra, S.; Parra, L. A smart communication architecture for ambient assisted living. *IEEE Commun. Mag.* **2015**, *53*, 26–33. [[CrossRef](#)]
38. Huang, X.; Yi, J.; Zhu, X.; Chen, S. A semantic approach with decision support for safety service in smart home management. *Sensors* **2016**, *16*, 1224. [[CrossRef](#)] [[PubMed](#)]
39. Abdar, M.; Zomorodi-Moghadam, M.; Das, R.; Ting, I.H. Performance analysis of classification algorithms on early detection of liver disease. *Expert Syst. Appl.* **2017**, *67*, 239–251. [[CrossRef](#)]
40. Breiman, L. Random forests. *Mach. Learn.* **2001**, *45*, 5–32. [[CrossRef](#)]
41. Chen, L.; Hoey, J.; Nugent, C.D.; Cook, D.J.; Yu, Z. Sensor-based activity recognition. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **2012**, *42*, 790–808. [[CrossRef](#)]
42. Blackman, S.; Matlo, C.; Bobrovitskiy, C.; Waldoch, A.; Fang, M.L.; Jackson, P.; Mihailidis, A.; Nygård, L.; Astell, A.; Sixsmith, A. Ambient assisted living technologies for aging well: A scoping review. *J. Intell. Syst.* **2016**, *25*, 55–69. [[CrossRef](#)]

© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.4 Publication 4

- **Title:** *Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies*
- **Authors:** Xiaochen Zheng, Shengjing Sun, Raghava Rao Mukkamala, Ravi Vatrappu, and Joaquín Ordieres-Meré
- **Publication date:** 2019/06/06
- **Journal:** *The Journal of Medical Internet Research (JMIR)*
4.924 (2019)
Quartile: Q1 in 'digital health'
ISSN: 1438-8871
Publisher: MIR Publications open access resource for evidence-based digital health research, head office, Canada.
- **Volume:** 21
- **Issue:** 6
- **Pages:** 13583
- **DOI:** <https://doi.org/10.2196/13583>

Original Paper

Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies

Xiaochen Zheng¹, DrPH; Shengjing Sun¹, MSci; Raghava Rao Mukkamala^{2,3}, DPhil; Ravi Vatrappu^{2,3}, DPhil; Joaquín Ordieres-Meré¹, DPhil

¹Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, Madrid, Spain

²Centre for Business Data Analytics, Department of Digitalization, Copenhagen Business School, Copenhagen, Denmark

³Department of Technology, Kristiania University College, Oslo, Norway

Corresponding Author:

Joaquín Ordieres-Meré, DPhil

Escuela Técnica Superior de Ingenieros Industriales

Universidad Politécnica de Madrid

José Gutiérrez Abascal 2

Madrid, 28006

Spain

Phone: 34 910677107

Fax: 34 913363005

Email: j.ordieres@upm.es

Abstract

Background: Huge amounts of health-related data are generated every moment with the rapid development of Internet of Things (IoT) and wearable technologies. These big health data contain great value and can bring benefit to all stakeholders in the health care ecosystem. Currently, most of these data are siloed and fragmented in different health care systems or public and private databases. It prevents the fulfillment of intelligent health care inspired by these big data. Security and privacy concerns and the lack of ensured authenticity trails of data bring even more obstacles to health data sharing. With a decentralized and consensus-driven nature, distributed ledger technologies (DLTs) provide reliable solutions such as blockchain, Ethereum, and IOTA Tangle to facilitate the health care data sharing.

Objective: This study aimed to develop a health-related data sharing system by integrating IoT and DLT to enable secure, fee-less, tamper-resistant, highly-scalable, and granularly-controllable health data exchange, as well as build a prototype and conduct experiments to verify the feasibility of the proposed solution.

Methods: The health-related data are generated by 2 types of IoT devices: wearable devices and stationary air quality sensors. The data sharing mechanism is enabled by IOTA's distributed ledger, the Tangle, which is a directed acyclic graph. Masked Authenticated Messaging (MAM) is adopted to facilitate data communications among different parties. Merkle Hash Tree is used for data encryption and verification.

Results: A prototype system was built according to the proposed solution. It uses a smartwatch and multiple air sensors as the sensing layer; a smartphone and a single-board computer (Raspberry Pi) as the gateway; and a local server for data publishing. The prototype was applied to the remote diagnosis of tremor disease. The results proved that the solution could enable costless data integrity and flexible access management during data sharing.

Conclusions: DLT integrated with IoT technologies could greatly improve the health-related data sharing. The proposed solution based on IOTA Tangle and MAM could overcome many challenges faced by other traditional blockchain-based solutions in terms of cost, efficiency, scalability, and flexibility in data access management. This study also showed the possibility of fully decentralized health data sharing by replacing the local server with edge computing devices.

(*J Med Internet Res* 2019;21(6):e13583) doi: [10.2196/13583](https://doi.org/10.2196/13583)

KEYWORDS

Internet of Things; distributed ledger technologies; data sharing; health information interoperability; IOTA Tangle; masked authenticated messaging; blockchain; intelligent healthcare

Introduction

Internet of Things and Intelligent Health Care

Internet of Things (IoT) has been developing explosively in recent years. It is believed to be the next revolutionary technology and bring great benefits to various domains of the society including health care [1]. The health care industry has been dramatically changed because of the information technology revolution that started in the last century. New technologies such as telemedicine, digital hospital, electronic health and mobile health have been widely applied during the past decades, and now, the rapidly development of IoT is promoting health care from digital into intelligent [2].

The advances of IoT have resulted in rapid emergence of smart environments such as smart home [3]. Sensors in these environments can measure the values of various environmental factors including temperature, humidity, air quality, and noise [4].

As an important aspect of IoT, wearable technology has also shown a surge in the past decade. Different types of wearable devices containing various embedded sensors such as smartphone, smart watch, smart band, and smart glasses have been used in health care applications to realize various health-related applications such as remote diagnosis [5], disease monitoring [6], and elderly people caring [7].

Challenges of Health Care Data Sharing

Large amounts of health-related data are generated by these smart devices including environmental data from stationary sensors and activity data from wearable devices. These data are valuable resources for health care applications, research, and commercial projects. Properly sharing these health data can benefit all related stakeholders including the device users, patients, researchers, and companies and improve the public health care system.

Currently, most data generated by IoT devices are controlled by different service providers, device manufacturers, or scattered in different health care systems [8,9]. These siloed and segmented data make it impossible or very difficult to share data outside their own closed environments, and this leads to enormous quantities of wasted data [10]. Besides, it puts data security and privacy at risk as these centralized data stores and authority providers are attractive targets for cyberattacks [11].

With the increasing concern about data privacy and security issue from public and private users, data protection regulations will become stricter. For example, the European Union has published the General Data Protection Regulation [12] to protect individual data. Such regulations make data sharing even more difficult.

Besides the complex data protection regulations, another main obstacle to freely flowing of big data is that, although data sharing is becoming cheaper from a technological perspective, it is prohibitively expensive to transfer fine, granular data in real time because of intermediary fees [13]. Another barrier is the lack of ensured authenticity and audit trails of data. Traditional data transmission protocols and databases are

susceptible to various attacks, including *man-in-the-middle* attacks and data tampering [14].

To overcome these barriers that hinder the full use of valuable health data, it is necessary to develop advanced systems to accelerate secure, fee-less, tamper-resist, and high-scalable health data sharing.

Distributed Ledger Technologies and Blockchain

A distributed ledger is a distributed database, maintained by a consensus protocol run by nodes in a peer-to-peer network. This consensus protocol replaces a central administrator, as all peers contribute to maintaining the integrity of the database [15].

As one of the most widespread DLT, the blockchain, has gained substantial popularity in recent years, primarily in the financial field because of the cryptocurrencies. For example, Bitcoin was first introduced in 2008 [16] and ever since has attracted the attention of the research community from diverse academic fields [17-19] and gained mainstream popularity because of its unique characteristics such as the absence of centralized control, an assumed high degree of anonymity, and distributed consensus over decentralized networks. Blockchain solutions could reduce data breach risks by utilizing threshold encryption of data together using public key infrastructure, where cooperation of multiple parties is required to decrypt data, and asymmetric cryptography is used to authenticate communication with system participants [20].

Limitations of Blockchain

Specialized distributed consensus protocols based on DLT have enabled novel decentralized applications such as cryptographic currencies [16] and smart contracts [21]. The rise and success of Bitcoin during the last 6 years proved that blockchain technology has real-world value. However, these block-based protocols, such as blockchain and Ethereum, also have several drawbacks that prevent them from being used as a generic platform for IoT data sharing.

Scalability

A blockchain has an inherent transaction rate limit because all participants agree on the longest chain and discard forks and side branches [22]. Common practice is to wait for 6 blocks to be added to the longest chain before reaching a high level of confidence that a transaction is final on the Bitcoin network [20,23]. As an example, it took on average 9.3 min to confirm a Bitcoin transaction at the end of December 2018 [24]. Applications that require exchange of value and low latency cannot be certain that their transactions are final in a shorter time frame and must trust the payer to not double spend [20]. The current incentive schemes that allow these protocols to spread virally make inefficient use of computational resources while constraining the transaction rate on the network. The transaction rate of Bitcoin protocol has been lower than 5 transactions per second in the whole network during most of the time in the year 2018 [25]. Similarly, the Ethereum protocol currently processes about 6 transactions per second across the entire network [26]. This low throughput cannot fill the requirements of data sharing in many health care scenarios.

Fees

Another notable drawback is the concept of a transaction fee for transactions of any value. For example, the Bitcoin protocol requires a fee that may exceed US \$0.30 each transaction [27] according to the statistics of January 10, 2019. To use a distributed ledger at scale for financial or other industrial use cases, this low throughput and high fee model will not suffice. The importance of micropayments will increase in the rapidly developing IoT technology and paying a fee that is larger than the amount of value being transferred is not logical. Furthermore, it is not easy to get rid of fees in the blockchain infrastructure as they serve as an incentive for the creators of blocks [28].

Centralization

Lots of computing power is required to maintain the blockchain, and mining power has become centralized to some extent. The latest statistic shows that the 6 largest mining pools control 75.76% of the of the network's mining power (BTC.com 21.5%, AntPool 14.9%, SlushPool 11.01%, ViaBTC 10.65%, BTC.TOP 9.67%, F2Pool 8.03%) [29].

Vulnerable to Quantum Attack

Bitcoin and other proof-of-work-based blockchains are susceptible to being broken by quantum computers. Quantum computers, although still a hypothetical construct as of today, could be very efficient for handling problems that rely on trial and error to find a solution [28]. The process of finding a nonce to generate a Bitcoin block is a good example of such a problem.

As of today, one must check an average of 2^{68} nonce to find a suitable hash that allows a new block to be generated. Theoretically, a quantum computer would need $\theta(\sqrt{N})$ operations to solve a problem that is analogous to the Bitcoin puzzle stated above [30]. This same problem would need $\theta(\sqrt{N})$ operations on a classical computer. Therefore, a quantum computer would be around 17 billion ($\sqrt{2^{68}}$) times more efficient at mining the Bitcoin blockchain than a classical computer. It would make possible of gaining control of over 51% of computing power of the whole blockchain network, which would enable attackers to double spend and break the entire network.

IOTA and the Tangle

IOTA is a tangle-based cryptocurrency designed specifically for the IoT industry where a machine-to-machine micropayment system is required. The tangle naturally succeeds the blockchain as its next evolutionary step by overcoming some of its previously mentioned fundamental limitations [31]. The main feature of the tangle is that it uses a directed acyclic graph for storing transactions instead of sequential blocks. In the Tangle, users must perform a small amount of computational work to approve 2 previous transactions to issue a new transaction. This new transaction will be validated by some subsequent transactions [28].

This structure enables the Tangle with high scalability. There is no maximum throughput, as the more activities in the Tangle, the faster transactions can be confirmed. In addition, with this 'pay-it-forward' system of validations, there is no need to offer financial rewards. Transacting with IOTA can be free of charge.

Moreover, IOTA has no miners, therefore it is truly decentralized.

The IOTA tangle is designed to be quantum resistant. The number of nonce that one needs to check to find a suitable hash for issuing a transaction is around 3^8 on average, which is not unreasonably large. The gain of efficiency for an "ideal" quantum computer would therefore be of order $3^4=81$, which is already quite acceptable [28]. More importantly, the algorithm used in the IOTA implementation is structured such that the time to find a nonce is not much longer than the time needed for other tasks that are necessary to issue a transaction. The latter part is much more resistant against quantum computing compared with the traditional blockchain.

Masked Authenticated Messaging

The main data communication protocol used in the proposed system is Masked Authenticated Messaging (MAM). It enables to emit and access an encrypted data stream over the Tangle regardless of the size or cost of a device [32]. MAM uses channels for message spreading. IOTA users can create a channel and publish a message of any size at any time. A small amount of proof-of-work is required to allow the data to propagate through the network and to prevent spamming. Other users can subscribe this channel through its address and receive a message that is published by the channel owner.

Merkle Hash Tree

MAM uses a signature scheme based on Merkle Hash Tree (MHT) [33-35] to sign the cipher digest of an encrypted message [32]. The *address* of a channel is the *root* of this Merkle tree, which itself is created using the *seed* of the user.

As the MHT example shown in Figure 1, private keys (A, B, C, D) are generated according to the *seed*, *index*, and *security level* [36]. The corresponding *addresses*, also called *leaves* (A', B', C', D'), can be generated respectively [37,38]. By applying the hash functions to narrow the addresses, the *root* of the Merkle tree can be obtained. In a MAM stream, a single MHT only lasts for a short period of time, each message contains the root of the next Merkle tree (or the future direction of the channel) [32].

Each message is signed with the one-time signature (OTS) scheme. Each leaf in the MHT corresponds to 1 OTS scheme. This means that each tree can produce the same number of messages as the number of leaves in the MHT [15].

In an MHT, the set of complementary hashes of a given leaf are the *siblings* of this leaf. As shown in Figure 1, the siblings of leaf A' (in red color) are B' and Hash(C', D'). By combining a given leaf and its siblings, the *root* of an MHT can be calculated.

A complete MAM transaction should include a signature section and the masked message section. The signature is created from one of the private keys corresponding to one of the leaves. The masked message consists of the raw data that need to be shared, the root of the next MHT, the index of the chosen leaf (branch index), and the siblings of this leaf. Figure 2 shows an example of a MAM stream with 2 transactions.

Figure 1. An example of Merkle Hash Tree with 4 leaves. **Figure 2.** Structure of a Masked Authenticated Messaging stream with 2 transactions.

Privacy and Encryption Modes

MAM has 3 privacy and encryption modes to control the visibility and access of a channel: public, restricted, and private.

In public mode, the root of MHT is directly used as MAM transaction address and channel key. Therefore, any user who receives a message randomly or intentionally can then decode it by using the address of the message.

In private mode, the hash of the MHT root is used as the address, and the message is decrypted using the *root*. This prevents random users from decrypting the message if they stumble across it as they are unable to derive the *root* from the hash.

In restricted mode, an authorization key, named as *sideKey* in this study, is added based on private mode. The address used to attach to the network is the hash of the *sideKey* and the *root* (according to the current MAM source code [38], only the hash of root is used, which differs from the introduction of IOTA website [32,36]). It enables a message publisher to revoke access to future messages from subscribers by changing the *sideKey*.

To consume a MAM message, the receiver needs to use the *root* to calculate the address of the transaction and fetch the masked message. Then, use the *root*, and *sideKey* in restricted mode, to decrypt the masked message.

In a MAM channel, the current message contains the address of the following message, whereas the previous ones are not referenced. This adds the forward secrecy character to a channel. When users are authorized with the correct decryption key, they could follow a MAM stream from the current transaction, but there is no way to read previous messages.

Objective of This Paper

The objective of this study was to integrate IOTA Tangle with IoT to develop a health data sharing system, which could support secure, fee-less, tamper-resist, high-scalable, and granular-controllable health data exchange. The data source

could include both wearable devices and stationary sensors in a smart environment such as smart home. The feasibility of the proposed system needs to be verified with a prototype system and its application in a practical case.

Methods

System Architecture

The architecture of the proposed health care data sharing system is presented in Figure 3. There are 2 roles involved in this system, data publisher and data subscriber. The publisher can be an individual, a family, or any other organization who possesses smart devices and sensors. These devices, sensors, and their owners produce health-related data, which are then published to the Tangle using specific encryption and privacy protocols. The data are published in their own channels, and each channel has an address. The subscribers of a data channel will receive the new published data. The published data are usually encrypted, and an extra decryption key may be necessary to decrypt the received data.

All the data are published and received through an IOTA node, which is a computer connected to the IOTA network. Users may use their own node or use public nodes. A user can be a data publisher and a subscriber at the same time. For example, a patient can publish his or her health data, and his or her doctor can subscribe these data and make evaluation accordingly. Afterwards, the doctor can publish the evaluation result to the Tangle, and the patient can subscribe this channel and receive the result.

Due to the limitations of size, power supply, and computing capability, most wearable devices and environmental sensors cannot publish or receive data directly to or from the Tangle. In this case, a gateway layer will be necessary, which could be a computer, a smartphone, or a single-board computer such as Raspberry Pi [39].

Figure 3. Architecture of the proposed health data sharing system based on IOTA Tangle. GPS: Global Positioning System; MAM: Masked Authenticated Messaging.

Implementation

To verify the feasibility of the proposed health data sharing system and demonstrate the implementation process, a prototype has been developed. The structure of the prototype is shown in Figure 4.

A portable human movement monitoring system using smartwatches was developed previously for the remote diagnosis of essential tremor (ET) [40,41]. The Pebble smartwatch [42] in this system could measure the triaxial acceleration data for tremor evaluation and activity recognition. The customized apps in the smartphone allows users to report their location, activity name, tremor level, self-evaluation about the disease, and other factors related to the disease, such as medication, alcohol, and coffee intake. These data, after integrated with other sensor data generated by the smartphone, will be compressed and uploaded to the remote server via internet for analysis using machine learning techniques.

In addition to the movement monitoring system, we added an environmental monitoring system composed of Kagoo air quality sensors [43] and Raspberry Pi [39]. The Kagoo sensors could measure various environmental factors such as temperature, humidity, noise, and the content of pollutions in the air,

including particulate matter, formaldehyde, total volatile organic components, benzene, carbon dioxide, carbon monoxide, ozone, and nitrogen dioxide. These sensors can be freely combined and plugged into a motherboard, which can communicate with the single-board computer Raspberry Pi through wired or Wi-Fi connection. A Python program running on Raspberry Pi could fetch and preprocess the environmental data from air sensors. More technical details, including the hardware manual and software codes, are openly accessible [44].

We use these 2 data collecting systems to represent wearable devices and stationary context sensors. The combination of these 2 data sources could provide a more complete understanding about users' health-related information.

In this prototype, the Pebble smartwatch, Android smartphone, and air quality sensors compose the sensing layer; the smartphone and Raspberry Pi play the role of gateway corresponding to the architecture of the proposed system. The data collecting frequency varies among different devices. The acceleration data from the smartwatch are recorded with a frequency of 25 Hz and uploaded every minute in a batch. The frequency of the data from smartphone depends on user's habit and usually is less than once per hour. The Kagoo sensors record environmental data once per minute.

Figure 4. Prototype of the health data sharing system using smartwatches, smartphone, air sensors, and Raspberry Pi. MAM: Masked Authenticated Messaging.

In this prototype, instead of directly published to the Tangle through the gateway as shown in Figure 1, the raw data are first sent to a local server for processing. The reasons are 2-fold. First, the frequency of the acceleration data from smartwatch is much higher than the other 2 data sources. Publishing these raw data to the Tangle will lead to a long lagging period. Therefore, on the server side, the acceleration data will pass through a tremor evaluation module based on deep learning approaches [41]. The output will be a shorter message per minute with a time stamp and a tremor score based on the classification result. This shorter message will be published to the Tangle. The raw acceleration data will be saved in a private database for future use. The second reason is to simplify the experiment of testing the average waiting time of publishing messages. The data from different sources are all gathered in the server and published through the same node in a concentrated period to obtain a more reliable result.

In terms of privacy and encryption modes, the environmental data are published in public MAM mode, whereas the patient report data and tremor evaluation data are published in restricted MAM mode.

To consume the published data over the Tangle, subscribers only need to know the address of the channel if the data are published in public mode, whereas an extra decryption key is needed for the data in restricted mode as introduced previously. Both data publishing and data receiving were realized through

the JavaScript programs, which are introduced with details in the following experiment and results section.

Results

Experiment

An experiment was conducted to prove the feasibility of the proposed system, which can broadcast and receive combined health data from both wearable devices and stationary environmental sensors. In general, 3 types of data are tested, including tremor level based on smartwatch acceleration data, patient reports from smartphone, and environmental data from air sensors. The environmental data were broadcasted using public MAM protocol, whereas the other 2 types of data were broadcasted using restricted MAM protocol. The authentication keys of restricted mode were changed during a broadcast stream to demonstrate how a user could revoke access to the data they generate in future.

All the data were broadcasted and received in JSON format. For each type of data, 100 trials of broadcasting were realized to test the average waiting time. The data were published using a computer equipped with a 4-core Intel Core i5-4460 3.2 GHz CPU, a 12 GB of RAM memory, and the Ubuntu Linux 18 64-bit version operating system. The data were published through a public IOTA node [45,46]. The memory usage was 50%, and the number of neighbors was 12 when connected to this node during the experiment. The complete scripts for

publishing and receiving JSON data over the Tangle are openly available [47].

Experiment Outcomes

Figure 5 shows an example of published environmental data over the Tangle using public MAM mode. It displays that the *address* of the channel is the same as the *root* of the MHT. Any user who knows the *address* could fetch the message and decrypt it with *root*, which is the same to the *address*.

Figure 6 presents an example of patient report data published in restricted mode. In this case, the *address* is the hash of the MHT *root*, which is totally different. Subscribers need to know both the *address* and the extra encryption key (*side_key*) to fetch and decrypt the message. In restricted mode, the publisher can send a subscriber the *address* and *side_key* to grant him or her

access to the current and future messages in the data stream. To revoke the authorization, the publisher just needs to change the *side_key* when publishing a new message, and subscribers without the new *side_key* will lose the access to this message and future ones.

The combination of public and restricted MAM protocols could provide users granular control over their health data, which could bring great benefit to the health care system. For instance, in our prototype, when a patient wants to be diagnosed, he or she can share with the neurologist the *address* and the *side_key* to the report data and tremor evaluation data streams from a certain time. Afterwards, the neurologist will be able to fetch all 3 data streams as the environmental data are published in public mode. After the diagnosis, the patient can change the key to revoke the authorization, as shown in Figure 7.

Figure 5. Environmental data published to the Tangle with public Masked Authenticated Messaging mode.

```
Root: YCHCPSUSUSMSODZI9RESQDYGGBYGJCVLJHYQWAUGMECGZXQZDGQUDSFWAIBHWJBXLLAJDVS IKWFZOZDZE
Address: YCHCPSUSUSMSODZI9RESQDYGGBYGJCVLJHYQWAUGMECGZXQZDGQUDSFWAIBHWJBXLLAJDVS IKWFZOZDZE
waiting_time:37852
location: Celsa Group Office, timestamp: 2019-01-09 00:02:00, pm2_5: 12.286, pm10: 13.143, t
voc: 0.036, co2: 0.2, temperature: 26.7, humidity: 14.913, illumination: 0.0, noise: 66.897,
hcho: 0.02, co: 0, c6h6: 0.0, no2: 0, o3: 0
```

Figure 6. Patient report data published to the Tangle with restricted Masked Authenticated Messaging mode.

```
Root: LCMNXLYG9MBPSFUZRLJPQ09YQZAGWVWCQNNNGWDGCXSTZAQGCJVGWVCOHYNPOGQICSWQIELBPLUEDVCUW
Address: ZWSYJNEKDGWLOTBSJUKNLOVPOHFNHBRILTOWDPKKKBXFRGDWKNTPAJGYIOHSCQQQMCVEVRECEXQWMPBB
waiting_time:33040
date: 2016-06-09T17:13:41.000Z, alcohol: No, caffee: No, Medicine: Sumial Mysoline, t1: 2016-
06-09T17:05:02.000Z, ArmExtendL: 4, ArmExtendR: 4, t2: 2016-06-09T17:06:03.000Z, TouchNoseL:
4, TouchNoseR: 4, t3: 2016-06-09T17:06:52.000Z, Writting: 4, t4: 2016-06-09T17:08:39.000Z,
```

Figure 7. Granular access control over messages published over the Tangle by combing public and restricted Masked Authenticated Messaging protocol.

Table 1. Result of the data broadcasting experiment using Masked Authenticated Messaging (MAM) protocol.

Data	MAM mode	Size (bytes)	Waiting time per message (seconds)		
			Mean (SD)	Maximum	Minimum
Air quality	Public	260	20.41 (8.11)	55.41	8.81
Tremor level	Restricted	29	17.35 (5.37)	33.41	7.81
Personal report	Restricted	570	19.99 (7.61)	55.17	10.10
Air quality from Raspberry Pi	Public	260	22.56 (5.97)	40.36	12.87

The summary of the waiting time for publishing the 3 types of data to the Tangle based on 100 trials is presented in Table 1. The result shows that there is no obvious difference among these 3 types of data in terms of waiting time for publishing to the Tangle, although their message length and encryption modes are different. This is because of the fact that, in IOTA Tangle, the size of a transaction is 2673 trytes, which is about 1650 bytes. It means that as far as a message is shorter than this limit, the waiting time of publishing such messages should be similar on the same node. The actual waiting time depends on the computing capability of the node to perform proof of work, and it may vary from a few seconds to more than half minute according to our tests as shown in Table 1.

Currently, the bottleneck of the data publishing speed is the total number of nodes connected to the Tangle network and the condition of the specific node used by the publishing device. It is expected that the time to publish data from a local server or from the single board computer should be similar. Aiming to verify this consumption, an extra test using Raspberry Pi to publish the air quality data was conducted in addition to the experiment of publishing 3 types of data using a local server. The result is presented in Table 1. It shows that there is no obvious difference regarding the waiting time for publishing a message, which verified the aforementioned consumption.

Discussion

Principal Findings

This study explored the application of emerging distributed ledger technology in the health care domain. We proposed a health data sharing system by converging IoT, IOTA Tangle, and MAM protocol. It makes possible of a reliable marketplace for the individuals to share their health-related data with hospitals, researcher, industry companies, or any other organizations in a secure and controllable way. In return, individuals can get benefit from their own data in monetary, medical services, or other forms. On the other hand, researchers and companies will be able to gather relevant data for their studies, clinical trials, or product development.

Most existing studies about the applications of DLT and IOT in health care either focused on the conceptual design of health data sharing systems or discussed relevant policies from managerial perspectives. In comparison, this study not only proposed an application framework supported by DLT and IOT technologies but also implemented a prototype system in practice from technical perspective.

Through an experiment based on a prototype system, we demonstrated how the health-related data are collected and

published to the Tangle in different encryption and privacy options. Our experiment showed that combining public and restricted MAM data streams, individuals are enabled to define granular access controls to different data consumers. The proposed system could facilitate the development of fee-less, secure, and efficient health data sharing marketplace to handle the big data generated by numerous IoT devices, and hence, pave way to the promising intelligent health care.

Although the current implementation of IOTA Tangle and its MAM protocol are already usable, they are still under development and are evolving rapidly. The current waiting time for attaching a message to the Tangle may vary from a few seconds up to more than 1 min. Although this is faster than other block-based protocols, there is still large room for performance improvements, as the more nodes connected to the Tangle network, the faster a transaction can be approved.

Limitations

The feasibility of the proposed health data sharing system using IOTA Tangle and MAM protocol was verified through the experiments based on a prototype system in a controlled environment. There are a few limitations worth to be mentioned.

First, in the prototype, a local server was introduced between the gateway layer and the IOTA nodes. The aim was to handle the large amount of raw acceleration data and simplify the testing process. In practical application, this local server can be excluded. The sensor data can be published to the Tangle directly from IoT devices or through a gateway such as smartphone or Raspberry Pi. This could enable the real machine-to-machine communication and make it easier for large scale implementation.

Another limitation of this pilot study is that a public IOTA node was used for publishing and receiving data over the Tangle. The lagging time varied depending on the workload of that public node, which is always stable. In practical implementation, a private node should be set up according to the practical requirements.

Conclusions

IOTA Tangle, together with the MAM communication protocol, could provide a fee-less, secure, highly scalable, and quantum-immune data sharing platform. The fast development of IoT is upgrading health care industry from digital to intelligent. The converging of IOTA Tangle, MAM, and IoT could significantly accelerate the health data sharing and pave way to realizing the vision of intelligent health care. The proposed solution in this study overcomes many of the challenges faced by other traditional block-based solutions in

terms of cost, efficiency, scalability, and flexibility in data access management. It could be applied in many scenarios of health care, such as remote diagnosis, chronic disease monitoring, and elderly caring, as introduced in the previous ET diagnosis experiment. Patients can publish their own health data to the Tangle with different encryption options and authorize medical experts to access to the tremor and activity data during a period. Experts can also share the diagnosis result with patients or their relatives.

This solution could be useful in many other areas such as rehabilitation, sports and fitness, and labor health protection in

workplaces, which indicates the directions for future work. For example, wearable devices can be used to monitor workers' positions, activities, working load, and health indicators such as heart rate and blood pressure. Environmental sensors can be used to monitor the working conditions including the air quality, temperature, humidity, noise, and illumination. All these data or the periodical statistic results can be published to the Tangle and authorize access to different stakeholders such as Environmental Health and Safety experts, production managers, and government audit departments to better understand the health status of workers and avoid overfatigue or injuries.

Conflicts of Interest

None declared.

References

- Gubbi J, Buyya R, Marusic S, Palaniswami M. Internet of Things (IoT): a vision, architectural elements, and future directions. *Future Gen Comput Syst* 2013 Sep;29(7):1645-1660. [doi: [10.1016/j.future.2013.01.010](https://doi.org/10.1016/j.future.2013.01.010)]
- Zheng X, Rodríguez-Monroy C. The development of intelligent healthcare in China. *Telemed J E Health* 2015 May;21(5):443-451. [doi: [10.1089/tmj.2014.0102](https://doi.org/10.1089/tmj.2014.0102)] [Medline: [25671730](https://pubmed.ncbi.nlm.nih.gov/25671730/)]
- Risteska Stojkoska B, Trivodaliev K. A review of Internet of Things for smart home: challenges and solutions. *J Clean Product* 2017 Jan;140:1454-1464. [doi: [10.1016/j.jclepro.2016.10.006](https://doi.org/10.1016/j.jclepro.2016.10.006)]
- Ni Q, García Hernando AB, Pau de la Cruz I. A context-aware system infrastructure for monitoring activities of daily living in smart home. *J Sens* 2016;2016:1-9. [doi: [10.1155/2016/9493047](https://doi.org/10.1155/2016/9493047)]
- Son D, Lee J, Qiao S, Ghaffari R, Kim J, Lee JE, et al. Multifunctional wearable devices for diagnosis and therapy of movement disorders. *Nat Nanotechnol* 2014 May;9(5):397-404. [doi: [10.1038/nnano.2014.38](https://doi.org/10.1038/nnano.2014.38)] [Medline: [24681776](https://pubmed.ncbi.nlm.nih.gov/24681776/)]
- Zheng X, Vieira A, Marcos S, Aladro Y, Ordieres-Meré J. Activity-aware essential tremor evaluation using deep learning method based on acceleration data. *Parkinsonism Relat Disord* 2019 Jan;58:17-22. [doi: [10.1016/j.parkreldis.2018.08.001](https://doi.org/10.1016/j.parkreldis.2018.08.001)] [Medline: [30122598](https://pubmed.ncbi.nlm.nih.gov/30122598/)]
- Gao Y, Li H, Luo Y. An empirical study of wearable technology acceptance in healthcare. *Industr Mngmnt Data Syst* 2015 Oct 19;115(9):1704-1723. [doi: [10.1108/IMDS-03-2015-0087](https://doi.org/10.1108/IMDS-03-2015-0087)]
- Varshney U. Pervasive healthcare and wireless health monitoring. *Mobile Netw Appl* 2007 Jul 12;12(2-3):113-127. [doi: [10.1007/s11036-007-0017-1](https://doi.org/10.1007/s11036-007-0017-1)]
- Zhang J, Xue N, Huang X. A secure system for pervasive social network-based healthcare. *IEEE Access* 2016;4:9239-9250. [doi: [10.1109/ACCESS.2016.2645904](https://doi.org/10.1109/ACCESS.2016.2645904)]
- Manyika JM. McKinsey & Company. By 2025, Internet of things applications could have \$11 trillion impact URL:<https://tinyurl.com/y4jcejlg> [accessed 2019-02-01] [WebCite Cache ID 75rGpVMDI]
- Peterson K, Deeduvanu R, Kanjamala P, Boles K. A blockchain-based approach to health information exchange networks. 2016 Presented at: NIST Workshop Blockchain Healthcare; 2016; Gaithersburg, MD URL:<https://www.colleaga.org/sites/default/files/12-55-blockchain-based-approach-final.pdf>
- EUR-Lex: EU law. Brussels: European Parliament; 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46 URL:<https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679> [accessed 2019-05-21] [WebCite Cache ID 78Y7eDQns]
- Sønstebo D. IOTA Blog. IOTA Data Marketplace URL:<https://blog.iota.org/iota-data-marketplace-cb6be463ac7f?gi=12ef095fee1f> [accessed 2019-02-01] [WebCite Cache ID 75rH5vOj9]
- Callegati F, Cerroni W, Ramilli M. Man-in-the-middle attack to the HTTPS protocol. *IEEE Secur Privacy Mag* 2009 Jan;7(1):78-81. [doi: [10.1109/MSP.2009.12](https://doi.org/10.1109/MSP.2009.12)]
- Brogan J, Baskaran I, Ramachandran N. Authenticating health activity data using distributed ledger technologies. *Comput Struct Biotechnol J* 2018;16:257-266 [FREE Full text] [doi: [10.1016/j.csbj.2018.06.004](https://doi.org/10.1016/j.csbj.2018.06.004)] [Medline: [30101004](https://pubmed.ncbi.nlm.nih.gov/30101004/)]
- Nakamoto S. ResearchGate. 2008. BitcoA peer-to-peer electronic cash system URL:https://www.researchgate.net/publication/228640975_Bitcoin_A_Peer-to-Peer_Electronic_Cash_System [accessed 2019-05-18] [WebCite Cache ID 78T2oB3dp]
- Böhme R, Christin N, Edelman B, Moore T. Bitcoin: economics, technology, and governance. *J Econ Perspect* 2015 May;29(2):213-238. [doi: [10.1257/jep.29.2.213](https://doi.org/10.1257/jep.29.2.213)]
- Ali ST, Clarke D, McCorry P. Bitcoin: Perils of an Unregulated Global P2P Currency. School of Computing Science Technical Report Series 2015 May 01:283-293 University of Newcastle upon Tyne.

19. Harlev M, Sun YH, Langenheldt K. Breaking Bad: De-Anonymising Entity Types on the Bitcoin Blockchain Using Supervised Machine Learning. 2018 Presented at: 51st Hawaii International Conference on System Sciences; January 3-6, 2018; Hawaii p. 3497-3506 URL: <https://core.ac.uk/download/pdf/143481278.pdf>
20. Mamoshina P, Ojomoko L, Yanovich Y, Ostrovski A, Botezatu A, Prikhodko P, et al. Converging blockchain and next-generation artificial intelligence technologies to decentralize and accelerate biomedical research and healthcare. *Oncotarget* 2018 Jan 19;9(5):5665-5690 [FREE Full text] [doi: [10.18632/oncotarget.22345](https://doi.org/10.18632/oncotarget.22345)] [Medline: [29464026](https://pubmed.ncbi.nlm.nih.gov/29464026/)]
21. Ethereum F. Ethereum.org. 2018. URL: <https://www.ethereum.org/dao> [accessed 2019-02-01] [WebCite Cache ID [75rHBYTtW](https://www.webcitation.org/75rHBYTtW)]
22. IOTA Blog. 2018. Meet the Tangle URL: <https://www.iota.org/research/meet-the-tangle> [accessed 2019-02-01] [WebCite Cache ID [75rHI9zXS](https://www.webcitation.org/75rHI9zXS)]
23. Bitcoin. Confirmation URL: <https://en.bitcoin.it/wiki/Confirmation> [accessed 2019-02-01] [WebCite Cache ID [75rHLWc13](https://www.webcitation.org/75rHLWc13)]
24. Statista. 2017. Average confirmation time of Bitcoin transactions from January to December (in minutes) URL: <https://www.statista.com/statistics/793539/bitcoin-transaction-confirmation-time/> [WebCite Cache ID [75rHcAn3d](https://www.webcitation.org/75rHcAn3d)]
25. Blockchain. 2019. Transaction Rate URL: <https://www.blockchain.com/en/charts/transactions-per-second> [accessed 2019-02-01] [WebCite Cache ID [75rHlcZRI](https://www.webcitation.org/75rHlcZRI)]
26. Etherscan. 2019. Ethereum transaction chart URL: <https://etherscan.io/chart/tx> [accessed 2019-02-01] [WebCite Cache ID [75rHqWE2T](https://www.webcitation.org/75rHqWE2T)]
27. BitInfoCharts. 2019. Bitcoin Avg. Transaction Fee historical chart URL: <https://bitinfocharts.com/comparison/bitcoin-transactionfees.html#1y> [accessed 2019-02-01] [WebCite Cache ID [75rIcoXMM](https://www.webcitation.org/75rIcoXMM)]
28. Popov S. Contentful. 2018 Apr 30. The Tangle URL: https://assets.ctfassets.net/r1dr6vzfxfhev/2t4uxvsIqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota_1_4_3.pdf [accessed 2019-05-18] [WebCite Cache ID [78T3lkPA3](https://www.webcitation.org/78T3lkPA3)]
29. BTC.com. 2019. Pool Distribution URL: https://btc.com/stats/pool?pool_mode=year [accessed 2019-02-01] [WebCite Cache ID [75rHwwivP](https://www.webcitation.org/75rHwwivP)]
30. Brassard G, Høyer P, Tapp A. Quantum cryptanalysis of hash and claw-free functions. *Latin American Symposium on Theoretical Informatics*: Springer; 1998 Presented at: LATIN'98: Theoretical Informatics; April 20-24, 1998; Campinas, Brazil URL: <https://link.springer.com/chapter/10.1007/BFb0054319> [doi: [10.1007/BFb0054319](https://doi.org/10.1007/BFb0054319)]
31. IOTA Blog. 2018. What is IOTA? URL: <https://www.iota.org/get-started/what-is-iota> [accessed 2019-02-01] [WebCite Cache ID [75rI5aNS4](https://www.webcitation.org/75rI5aNS4)]
32. Handy P. IOTA Blog. 2017. Introducing Masked Authenticated Messaging URL: <https://blog.iota.org/introducing-masked-authenticated-messaging-e55c1822d50e?gi=97f611b532fb> [accessed 2019-02-01] [WebCite Cache ID [75rIDpJCj](https://www.webcitation.org/75rIDpJCj)]
33. Adam L. ImperialViolet. 2013. Hash based signatures URL: <https://www.imperialviolet.org/2013/07/18/hashsig.html> [WebCite Cache ID [75rII7DYN](https://www.webcitation.org/75rII7DYN)]
34. Merkle R. A certified digital signature. *Conference on the Theory and Application of Cryptology*: Springer; 1989 Presented at: CRYPTO'89; August 20-24, 1989; Santa Barbara URL: https://link.springer.com/chapter/10.1007/0-387-34805-0_21
35. Merkle R. A digital signature based on a conventional encryption function. *Conference on the theory and application of cryptographic techniques*: Springer; 1987 Presented at: Conference on the Theory and Application of Cryptographic Techniques; August 16-20, 1987; Santa Barbara p. 369-378 URL: https://link.springer.com/chapter/10.1007/3-540-48184-2_32
36. ABmushi. Medium. IOTA: MAM Eloquently Explained URL: <https://medium.com/coinmonks/iota-mam-eloquently-explained-d7505863b413> [accessed 2019-02-01] [WebCite Cache ID [75rIKrSNp](https://www.webcitation.org/75rIKrSNp)]
37. ABmushi. Medium. IOTA: Signature and Validation URL: <https://medium.com/@abmushi/iota-signature-and-validation-b95b3f9ec534> [accessed 2019-02-01] [WebCite Cache ID [75rIOTiI8](https://www.webcitation.org/75rIOTiI8)]
38. GitHub. IOTA URL: <https://github.com/iotaledger> [accessed 2019-02-01] [WebCite Cache ID [75rIRmm28](https://www.webcitation.org/75rIRmm28)]
39. Raspberry Pi. URL: <https://www.raspberrypi.org/> [accessed 2019-02-01] [WebCite Cache ID [75rI8dVyO](https://www.webcitation.org/75rI8dVyO)]
40. Zheng X, Vieira Campos A, Ordieres-Meré J, Balseiro J, Labrador Marcos S, Aladro Y. Continuous monitoring of essential tremor using a portable system based on smartwatch. *Front Neurol* 2017;8:96 [FREE Full text] [doi: [10.3389/fneur.2017.00096](https://doi.org/10.3389/fneur.2017.00096)] [Medline: [28360883](https://pubmed.ncbi.nlm.nih.gov/28360883/)]
41. Zheng X, Vieira A, Marcos S, Aladro Y, Ordieres-Meré J. Activity-aware essential tremor evaluation using deep learning method based on acceleration data. *Parkinsonism Relat Disord* 2019 Jan;58:17-22. [doi: [10.1016/j.parkreldis.2018.08.001](https://doi.org/10.1016/j.parkreldis.2018.08.001)] [Medline: [30122598](https://pubmed.ncbi.nlm.nih.gov/30122598/)]
42. FitBit. 2016. The Original Uncomplicated Pebble Classic URL: <https://www.pebble.com/pebble-smartwatch-features> [accessed 2019-04-15] [WebCite Cache ID [77eeWUHKU](https://www.webcitation.org/77eeWUHKU)]
43. Circulate Health Science And Technology. 2019. URL: <http://www.circulateoao.com/> [accessed 2019-02-01] [WebCite Cache ID [75rIWNn2b](https://www.webcitation.org/75rIWNn2b)]
44. Sun S. GitLab. 2019. Indoor air quality project URL: <https://gitlab.com/sunshengjing/iaq> [WebCite Cache ID [75rLI4HKH](https://www.webcitation.org/75rLI4HKH)]
45. IOTA. 2019. Public nodes list URL: [https://www.tangle-nodes.com/?sorts\[load\]=1&sorts\[tls\]=-1](https://www.tangle-nodes.com/?sorts[load]=1&sorts[tls]=-1) [WebCite Cache ID [75rNVsSJV](https://www.webcitation.org/75rNVsSJV)]
46. IOTA Nodes List. URL: <https://pow.iota.community:443> [accessed 2019-05-22] [WebCite Cache ID [78YaiXKKJ](https://www.webcitation.org/78YaiXKKJ)]

47. Zheng X. GitHub.: Github repository; 2019. iota mam data sharing URL:https://github.com/zhengxiaochoen/iota_mam_data_sharing[WebCite Cache ID 75rNHxYIF]

Abbreviations

DLT: distributed ledger technologies
ET: essential tremor
IoT: Internet of Things
MAM: Masked Authenticated Messaging
MHT: Merkle Hash Tree
OTS: one-time signature

Edited by K Clauson, P Zhang; submitted 01.02.19; peer-reviewed by C Reis, E Minvielle, A Bello-Garcia; comments to author 31.03.19; revised version received 16.04.19; accepted 27.04.19; published 06.06.19

Please cite as:

Zheng X, Sun S, Mukkamala RR, Vatrappu R, Ordieres-Meré J

Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies

J Med Internet Res 2019;21(6):e13583

URL: <https://www.jmir.org/2019/6/e13583/>

doi: [10.2196/13583](https://doi.org/10.2196/13583)

PMID: [31172963](https://pubmed.ncbi.nlm.nih.gov/31172963/)

©Xiaochen Zheng, Shengjing Sun, Raghava Rao Mukkamala, Ravi Vatrappu, Joaquín Ordieres-Meré. Originally published in the Journal of Medical Internet Research (<http://www.jmir.org>), 06.06.2019. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <http://www.jmir.org/>, as well as this copyright and license information must be included.

3.5 Publication 5

- **Title:** *Geometric Deep Lean Learning: Deep Learning in Industry 4.0 Cyber-Physical Complex Networks*
- **Authors:** Javier Villalba-Díez, Martin Molina, Joaquín Ordieres-Meré, Shengjing Sun, Daniel Schmidt, and Wanja Wellbrock
- **Publication date:** 2020/01/30
- **Journal:** *Sensors*
3.031 (2018) ; 5-Year Impact Factor: 3.302 (2018)
Quartile: Q1 in ‘Instruments & Instrumentation’
ISSN: 1424-8220
Publisher: Multidisciplinary Digital Publishing Institute, BASEL, SWITZERLAND
- **Volume:** 20
- **Issue:** 3
- **Pages:** 763
- **DOI:** <https://doi.org/10.3390/s20030763>

Article

Geometric Deep Lean Learning: Deep Learning in Industry 4.0 Cyber–Physical Complex Networks

Javier Villalba-Díez ^{1,2,3,*} , Martin Molina ² , Joaquín Ordieres-Meré ³ , Shengjing Sun ^{3,4} , Daniel Schmidt ^{3,5} and Wanja Wellbrock ¹

¹ Fakultät fuer Management und Vertrieb, Campus Schwäbisch-Hall, Hochschule Heilbronn, 74523 Schwäbisch-Hall, Germany; wanja.wellbrock@hs-heilbronn.de

² Department of Artificial Intelligence, Universidad Politécnica de Madrid, Campus de Montegancedo, 28660 Boadilla del Monte, Madrid, Spain; martin.molina@upm.es

³ Escuela Técnica Superior de Ingenieros Industriales, Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain; j.ordieres@upm.es (J.O.-M.); shengjing.sun@alumnos.upm.es (S.S.); daniel.schmidt@saueressig.de (D.S.)

⁴ Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

⁵ Lead Developer Quality Inspection, Matthews International GmbH, Gutenbergstraße 1-3, 48691 Vreden, Germany

* Correspondence: javier.villalba.diez@hs-heilbronn.de

Received: 31 December 2019; Accepted: 29 January 2020; Published: 30 January 2020

Abstract: In the near future, value streams associated with Industry 4.0 will be formed by interconnected cyber–physical elements forming complex networks that generate huge amounts of data in real time. The success or failure of industry leaders interested in the continuous improvement of lean management systems in this context is determined by their ability to recognize behavioral patterns in these big data structured within non-Euclidean domains, such as these dynamic sociotechnical complex networks. We assume that artificial intelligence in general and deep learning in particular may be able to help find useful patterns of behavior in 4.0 industrial environments in the lean management of cyber–physical systems. However, although these technologies have meant a paradigm shift in the resolution of complex problems in the past, the traditional methods of deep learning, focused on image or video analysis, both with regular structures, are not able to help in this specific field. This is why this work focuses on proposing geometric deep lean learning, a mathematical methodology that describes deep-lean-learning operations such as convolution and pooling on cyber–physical Industry 4.0 graphs. Geometric deep lean learning is expected to positively support sustainable organizational growth because customers and suppliers ought to be able to reach new levels of transparency and traceability on the quality and efficiency of processes that generate new business for both, hence generating new products, services, and cooperation opportunities in a cyber–physical environment.

Keywords: Industry 4.0; IIoT; geometric deep learning; lean management

1. Introduction

Today it seems almost a truism to talk about the fact that data surround us. According to recent studies, by 2025 humanity will have created about 163 zettabytes of information [1]. However, the alarming thing is not that we are going to be flooded with data, but that these data will be very different from the data with which we are used to dealing in classical disciplines such as signal or image processing, statistics, or machine learning. Beyond this, the data we will face are data that will emerge from the trillions of objects connected to the Internet of Things (IoT). In many cases, including the industrial IoT (IIoT), these data are produced by distributed sources, such as thousands

of sensors in factories, i.e., data are distributed over networks. Managing large amounts of data in these ever-expanding networks raises nontrivial concerns about the efficiency of data collection, processing, analysis, and security [2,3]. Currently, data from processes and systems are collected and stored without a clear strategy, and this can be a barrier to implementing paradigms such as “social manufacturing” [4]. In addition to being distributed, these data may be unstructured, and therefore cannot generally be encapsulated in one table. A defined strategy is therefore needed on what kind of data to collect at the technical and the organizational level. Finally, in addition to numerical, data can be ordinal, categorical, or other. The aim of this work is to introduce the reader to a series of concepts that pave the way for processing these data by means of adapted deep-learning techniques [5].

The purpose of this work is to study the possibility of providing Industry 4.0 leaders with a theoretical model that allows for the extraction of relevant patterns embedded within their organizations by means of artificial intelligence. Specifically, the goal of this work is to provide the reader with mathematical models that adapt convolutional and pooling deep-learning operations, hence describing the possible use of geometric deep-learning architectures on non-Euclidean Industry 4.0 complex cyber–physical networks. The structure of this work is as follows: First, Section 2 provides relevant background information, clarification, and definitions. Second, Section 3 provides a framework of previous relevant concepts regarding deep learning, specifically geometric deep learning. Third, Section 4 provides mathematical models to compute geometric-deep-learning algorithms over Industry 4.0 lean-management complex-networked cyber–physical systems. Finally, Section 5 outlines the conclusions and managerial implications of this model, and its applications in the field.

2. Background

This brief section presents and defines fundamental preliminary concepts to the comprehensive understanding of the presented content in the following sections of this work:

- **Industry 4.0.** The term Industry 4.0 has gained large traction since it was first publicized [6], stating the need for a paradigm shift towards a less centrally controlled manufacturing structure. It is seen as the Fourth Industrial Revolution, with the first three being mechanization through steam power, mass production through electrically operated engineering, and the digital revolution through the integration of electronics and IT. Industry 4.0 enables more production autonomy as technology becomes more interconnected, and machines are able to influence each other by creating a cyber–physical system.
- **Cyber–Physical Systems.** The term “cyber–physical system” in the context of Industry 4.0 refers to the tight conjoining of and co-ordination between computational and physical resources. The impact on the development of such systems is a new paradigm of technical systems based on collaborative embedded software systems [7].
- **Lean Management.** Lean-management systems in an Industry 4.0 cyber–physical context have been described as sociotechnical entities that aim to systematically reduce the variability of value-creation processes [8–13]. These two fundamental dimensions, the social and the technical, are subsequently meant to symbiotically support each other to maximize value creation through the systematic elimination of activities that do not add value for the client. A series of models were presented by scholars that allow the analysis and quantification of these systems as complex networks [14,15].
- **Complex-Networked Organizational Design.** Under the organizational-network paradigm, modern Industry 4.0 cyber–physical lean-management-oriented organizations can be understood as a symbiotic sociotechnical ecosystem of social networks [16] that interacts with increasingly complex-networked physically distributed interconnected sensors [17], whose readings are modeled as time-dependent signals on the vertices, human or cyber–physical, respectively. This means that, on the nodes of the network, attributes can be found that describe them as having the form of a given time series.

Within this framework, a complex network is defined as a graph with nontrivial topological features that do not occur in simple graphs such as lattices and random networks [18]. For any given time t , lean complex cyber–physical networks can be formally described by time-dependent graphs $\Omega(t) = [N(t); E(t)]$ that can be understood as lists of $N(t)$ nodes and $E(t) \subset (N(t) \times N(t))$ edges that represent its human and cyber–physical nodes, and its standard communication edges [19]. Given the static graph in t , $\Omega(t)$, each node and edge can be characterized by a series of typically two-dimensional signals $x = [x_1, \dots, x_n] \in (\mathbb{R}^n \times \mathbb{R}^m)$, where n relevant parameters of the node or axis are described as the time series of m elements. In the case of nodes, signals typically represent demographic, sociological, or competence information. In case that the nodes are human, and in the case of a cyber–physical node, relevant information on the state of the cyber–physical node expressed in time series of several key performance indicators. In the case of edges, signals typically represent information referring to the quality of measurable relationships of the individual with other stakeholders of the organization; in the case of human–human or cyber–physical-to-human edges, of the time series associated with relevant key performance indicators being reported to other stakeholders. Specifically, snapshots for the time-dependent graph can be built, that is, the time-dependent graph is considered as an ordered pair of potentially different sets. A time-dependent graph considered as a sequence of static graphs is given by Expression 1.

$$\Omega = [\Omega(t_1), \Omega(t_2), \dots, \Omega(t_k)] \quad (1)$$

This method is most commonly used for modeling discrete time-dependent graphs, and is suitable for the time-dependent graph with a specific time structure, especially in real-time networks such as complex-networked cyber–physical systems [20]. This modelling method is assumed here, and the time sequence of static graphs is not explicitly mentioned when referring to time-dependent graphs.

As a consequence of these references, it can be stated that cyber–physical complex-networked lean-management systems in an Industry 4.0 context can be understood as management systems that systematically try to reduce the intrinsic variability of industrial value-creation processes by understanding them as complex networks of computational and physical elements.

3. Related Work

Within this framework, the work approaches the interpretation of strategic information contained in Industry 4.0 cyber–physical complex-networked lean-management systems from two main vectors: social and technical strategic organizational design complexity. As shown in the research overview in Table 1, these two research directions have been intensively examined at three (micro-, meso-, and macroscopic) levels of complexity. A better visualization of these organizational levels is in the graphical abstract of Figure 1 for clarity purposes, but it should be noted that this classification is purely synthetic; in reality, cyber–physical systems in an Industry 4.0 context present the continuous complexification of networks arranged in nested hierarchies. This by no means suggests that one level of aggregated complexity is more difficult to deal with than a less aggregated one. In fact, the opposite is often true. For example, in the study of value-creating cyber–physical systems, the study of shop-floor management has been done for decades with almost solely qualitative methods and common sense [21–23]. Deep learning has been recently used to extract statistical patterns from cyber–physical systems at certain microscopic local levels [24,25]; however, there is an urgent need for algorithms to be developed that ensure a holistic understanding of cyber–physical systems at the meso- and macroscopic level of complex-networked aggregation.

Figure 1. Macroscopic, mesoscopic, and microscopic levels of organizational sociotechnical complexity.

Table 1. Research overview.

	Social	Technical	Socio Technical
Micro	Imai, 2012 [26]; Stock and Seliger, 2016 [27]	Takeda, 2009 [28]; Francis and Bian, 2019 [29]; Jabeur et al., 2015 [17]; Li et al., 2017 [30]; Aazam et al., 2018 [31]; Tao et al., 2018 [32]; Mushtaq and Haq, 2019 [33]; Shevchik et al., 2019 [34]; Al-Jaroodi and Mohamed, 2019 [35]; Sun et al., 2019 [36]	Villalba-Diez et al., 2015 [23]; Villalba-Diez et al., 2019 [24].
Meso	Rother, 2010 [37]; Villalba et al., 2018 [13]; Birkel et al., 2019 [38]	Takeda, 2011 [39]; Davis et al., 2012 [40]; Gomez et al., 2015 [41]; Culot, 2019 [42]; Jimenez et al., 2016 [43]; Wang et al., 2018 [44]; Villalba-Diez et al., 2019 [25]; Jang et al., 2019 [45]; Ordieres-Mere et al., 2019 [46]	Villalba-Diez and Ordieres-Mere, 2015 [10]; Villalba-Diez et al., 2015 [9]; Villalba-Diez and Ordieres-Mere, 2016 [11]; Villalba-Diez et al., 2017 [47]; Davies et al., 2017 [48]; Kumar et al., 2019 [49].
Macro	Womack and Jones, 2003 [50]; Toyota, 2014 [51]; Burton et al., 2015 [52]; Covey, 2004 [53]; Rabelo et al., 2019 [54]; Romero et al., 2017 [55]; Wang et al., 2019 [56]; Guo and Jyang, 2019 [57]	Lee et al., 2015 [58]; Wang et al., 2015 [59]; Goodfellow et al., 2016 [5]; Sisini et al., 2018 [60]; Zheng et al., 2018a [61]; Lu and Xu, 2019 [62]	Stock and Seliger, 2016 [27]; Villalba-Diez, 2017 [15]; Villalba-Diez, 2017 [14]; Kiel et al., 2017 [63]; Stock et al., 2016 [64]; Shang et al., 2019 [65].

Subsequently, a research hypothesis can be formulated. Due to the high potential shown by deep learning in a wide range of applications, we could hypothesise that deep learning can be used to find patterns within Industry 4.0 lean-management complex-networked cyber-physical systems, which takes us to the concept of geometric deep lean learning. The analogy of networks proposed in this work, as well as the global analysis of the evolving networks and, through the geometric deep lean learning of the local relations between agents, provide an adequate context to establish which data to collect, and how to structure their analysis in a general and systematic way.

Within this context, there are two main resource-organizing classes for integrating deep learning in Industry 4.0 cyber-physical contexts with regard to different assumptions on data acquisition:

- Offline training, and decision-support learning and predicting from a global and integrated way, for example, by extracting relevant information from an organization by means of deep-learning algorithms that analyze previously labeled text in organizational categories [66]. Alternatively, by combining deep learning with other computing methods that allow for more balanced datasets and, hence, better deep-learning performance [67].
- Digital twin and augmented reality. Creating virtual environments that, by recording, visualization, and interaction with cyber-physical assets, are capable of generating necessary tagged information in real time that is fed to deep-learning algorithms [68]. The creation of digital twins in combination with deep-learning algorithms was also proposed to enable the parallel control of cyber-physical value-creating processes [69].

Deep-learning algorithms are built by stacking data-processing filters—layers—in deep architectures [5]. These layers extract increasingly accurate representations of the data fed into them through a series of algebraic operations, such as convolution (learning local patterns of feature maps) and pooling (downsampling of feature maps). A key reason for the success of these classical deep-learning applications on time-series, images, or video processing, is on its underlying Euclidean

or gridlike data-structure space. The ability to leverage statistical properties of such data through local statistics is possible because of the shift invariance, local connectivity and the multi-resolution of the dataset. For instance, in a color image, pixels are placed together (shift invariance), present local properties (local connectivity), and present a red–green–blue-layered color structure (multiresolution). The use of convolution and pooling imposes conditions on the dataset while extracting local features shared throughout images that make it suitable for the problem without sacrificing the expressive capacity of the network. In fact, the graph's Laplacian $\mathcal{L} = D - A$ that supports the information contained in the images is constant [70], where D and A represent the degree and adjacency matrix of the graph, respectively [19]. This allows a series of mass algebraic operations that make the magic of deep learning possible. However, at an organizational level, networks associated with Industry 4.0 lean-management cyber–physical systems are, by definition, dynamic and do not present these characteristics.

The fundamental idea of deep learning is that it is assumed that data to be studied came from the combination of different attributes at multiple hierarchical levels. An important underlining concept in this context is that of the manifold. A manifold can be intuitively understood as locally Euclidean space. Earth, for example, can be understood as a gigantic ellipsoid, but to a human at a point on its surface, it appears to be a plane. In other words, the manifold is an interconnected region: a series of points associated with its surrounding environment. From any of these points, the manifold appears to be locally Euclidean. Formally speaking, differentiable X manifold of dimension d is a topological space in which each point x has an environment that is homeomorphic to a Euclidean space of dimension d called tangent space $T_x X$ [71]. If the manifold is equipped with a Riemannian metric, such as an inner product $\langle \cdot, \cdot \rangle_{T_x X} : (T_x X) \times (T_x X) \rightarrow \mathbb{R}$, then the manifold is called a Riemannian manifold. The set of tangent spaces at all points is known as tangent bundle TX and is assumed to be smoothly dependent on the x position. It is precisely this feature that is exploited by machine-learning algorithms. The condition for this is the implicit assumption that interesting points occur only in a collection of manifolds in directions tangent to the TX planes, and with statistically interesting variations happening only when switching manifolds.

In other words, manifolds are topological spaces locally homeomorphic to Euclidean spaces. Complex networks, the object of this study, can be described by complexes of nodes and edges (i.e., triangles) that can be treated as discrete types of manifolds [72]. As has been described before [73–75], these can be understood as manifolds in order to explain the problems related to evolutionary manifolds using the theory of complex evolutionary networks. Specifically, deep learning applied to graphs usually considers these as manifolds; for this reason, we can consider deep lean learning as a manifold learning challenge. In the following sections, the consideration of graphs as manifolds is not geometrically rigorous, and might not be as smooth as previously defined. Classical applications of deep learning to graphs [76] focuses on static networks, but cyber–physical systems represented by complex networks are dynamic in nature, as nodes (both human and cyber–physical) and sociotechnical relations between them are constantly evolving.

For this reason, in order to discover statistical patterns within lean-management cyber–physical systems by means of deep learning, it is necessary to either transform existing data into figures that can be interpreted by classical approaches, or to generalize the concept of deep learning to dynamic networks. The first strategy was successfully implemented by one of the authors [13]. The second strategy follows in the footsteps of geometric deep learning.

Geometric deep learning is an emerging technique to generalize deep-learning models to non-Euclidean domains, such as certain graphs and manifolds [70]. The wide variety of domains in which geometric deep learning has so far been useful can be summarized in four categories:

- Graphwise classification. For instance, in the classification of molecules [77]. In this model, atoms represent the nodes, and chemical bonds are the edges of a graph. Research aims to extract certain features that predict certain properties of the molecule. This is relevant, for instance, to the pharmaceutical companies that are in the business of drug design. Some of these properties are

toxicity and water solubility. Given a graph, researchers aim to classify a molecule graph. This is analogous to classical deep-learning-based visual image classification [78].

- Vertexwise classification. For example, in a social-network domain in which nodes are people of which we have certain demographic information, a researcher aims to predict how these people will vote in the next election. The analogy in computer vision is semantic image segmentation [79] in which the pixels of an image are labeled as belonging to a certain category.
- Graph dynamics. There are also domains that are described by fixed graphs, and others in which the graph changes with time [70]. Complex-networked cyber-physical systems belong to the second class.
- Known vs. unknown domain. In some cases, the graph can be known; in others, it is only partially known, *noisy*, or not known at all and needs to be learned. In these cases, the researcher aims to not only learn the graph features, but also the graph itself [80].

Existing approaches to implement geometric deep learning can be classified into two broad categories: spectral and local filtering methods.

- Spectral filtering methods.

Spectral filtering methods make use of the spectral eigendecomposition of the Laplacian graph to elegantly mathematically define convolution-like operators. The fundamental limitation of the spectral construction is that it can only be used to single and static domains. This is because filter coefficients are dependent on the eigenvector- and eigenvalue-decomposition basis of the Laplacian graph, which is highly dependent on network architecture [70]. This approach is not suitable for our needs because of the dynamic characteristics of Industry 4.0 lean-management cyber-physical complex systems and their associated complex networks.

- Local filtering methods.

Local filtering methods, on the other hand, are not topology-dependent, fall within the frame of signalling processing on graphs [81], and are more suitable in this setting, in particular, in order to define an operation similar to convolution in this domain [82].

4. Geometric Deep Lean Learning Over Industry 4.0 Lean-Management Complex-Networked Cyber-Physical Systems

According to Immanuel Kant, a science is not a science until there is a relation to mathematics. Although this characterization is provocative, and few would discuss such absolute numbers today, the implicit main question remains valid: can we find mathematical expressions that explain, process, and learn from network data, especially from complex-networked cyber-physical systems? This question is the motivator of this work, both for its practical and theoretical interest. On the one hand, empirically speaking, the processing of signals on graphs from complex cyber-physical networks has exponential importance due to the unstoppable emergence of technologies such as the IIoT and blockchain. On the other hand, the theoretical field of artificial intelligence constantly needs to develop new algorithms and computational architectures to later allow its practical application.

Applied to the analysis of complex-networked cyber-physical systems in the context of Industry 4.0, this leads to two classes of problem formulations that geometric deep lean learning theoretically solves:

- Strategic organizational design. Performing classical inference problems [76].

Recently, it has been shown that this classification can be considerably improved by using information about the proximity environment [83,84]. Analyzing signals on graph vertices and edges could potentially help to learn inherent structures of the graphs, such as organizational clusters, with better accuracy than that provided by topological information alone—this is a strategic challenge to which organizational design tries to respond.

- Trust and power structures. Learning hidden organizational properties.

Although deep learning has been employed in a wide variety of fields of knowledge, such as modeling social influence [85] and computer vision [86–88], it is important to incorporate knowledge about the domain to be treated in the model. For example, in order to build a deep-learning model for the study of a network of sensors in a cyber–physical system of industry 4.0, it might be useful, in a first approximation, to choose the edge weights of the graph as a decreasing function of the distance between nodes, as this would lead to a smooth graph signal model [89]; however, this would not be suitable for a lean structural network, because adjacency does not necessarily mean similarity [14]. For this reason, the model of the graph to be used can be superimposed on other structures, instead of being a pure unconnected abstraction. In other words, the graph that represents the complex-networked cyber–physical system in an Industry 4.0 context, can be studied from different perspectives, superimposing it to a specific sociotechnical environment that helps to better understand the statistical information that it contains. As a consequence, the integration of these priors is a fundamental challenge for the success of geometric deep learning. Some examples are the structures of power or trust between the different actors of an organization that are fundamental variables that influence the success of an organization, but remain elusive, since they often cannot be directly measured. Geometric deep learning could be applied to learn these parameters as weights between the nodes of the complex organizational network.

These problems reduce to fitting a time-dependent tensor $A(t)$, so that $\Omega(t+1) \approx A(t) \cdot \Omega(t)$ [90]. The hypothesis underlying this objective is that $x(t+1) \approx A(t) \cdot x(t)$ where $A(t)$ is constant in a window of time. The reason why we can take this assumption as true is that complex networks associated with cyber–physical systems in Industry 4.0 environments do not have very high variability [14]. As a result, a sufficiently small time window can always be found in which the hypothesis is sufficiently true.

Generalizing deep-learning models to dynamic structured data in complex graphs requires a detailed description of the non-Euclidean equivalents to the basic elements of deep learning (convolutional layers and downsampling “pooling”), locally applied to each of the graph elements [70]:

- Convolution on non-Euclidean complex-networked cyber–physical graph time-dependent signals.

As expressed in Expression 1, for weighted time-dependent directed graph $\Omega(t)$, a series of signals $x = [x(1), \dots, x(n)] \in (\mathbb{R}^n \times \mathbb{R}^m)$ expressed on its human and cyber–physical nodes, and on its standard communication edges, are considered, in which components of x_a reside in or are protruding from node a .

For each node, we define a proximity environment given by group $N_a = \{b : (b, a) \in E\}$ that represents set of nodes b connected with a . This N_a set is characterized by an $\mathbb{R}^{N \times N}$ matrix S called the network-translation matrix operator that defines the manifold metric. We defined S as the graph adjacency matrix, the Laplacian of the graph, or any other normalization of it, as a linear transformation to encode the structure of a graph. Without loss of generality, the singularity problem of the adjacency matrix, which is nontrivial, was not considered in this work [91]. As shown in Figure 2, group N_a represents the manifold upon which the convolution acts.

Figure 2. Local manifold upon which graph convolution acts.

The Fourier decomposition of graph $\Omega(t)$ is expressed by $\hat{x} = U^{-1} \cdot x$, where $S = \mathcal{U} \cdot \Lambda \cdot \mathcal{U}^{-1}$ and autovalues Λ describe the frequencies of the graph [92]. Now, we can directly filter x from the spectral domain by means of function $f : \mathbb{C} \rightarrow \mathbb{R}$ that allows to compute convolution $\hat{z} = f(\Lambda) \cdot \hat{x}$ by means of point-by-point multiplication in the spectral domain between filter $f(\Lambda)$ and the Fourier transform of the graph in x . Therefore, by inverting the Fourier transform of the graph, we obtain the extension of the convolutional operation to the non-Euclidean time-dependent graph in Equation (2).

$$z = \mathcal{P}(S)x \text{ and } \mathcal{P}(S) = \mathcal{U}f(\Lambda)\mathcal{U}^{-1} \quad (2)$$

The filter operation can be directly described on the node, resulting in an alternative formulation given by Equation (3), where scalar parameter $\phi_{a,b}$ is a representation of the information weights coming from neighbour node b into or from node a .

$$z_a = \sum_{b \in N_a \cup a} \phi_{a,b} \cdot x_b \quad (3)$$

Due to the local properties of S , z_a can be obtained in the domain of the node through local-information exchange. This means that the initial signal on the node is recursively transformed by S a K number of times until decomposition is obtained that determines z_a as the convolution between the network filter with a polynomial transfer function and x_b .

By means of the Fourier transform of the network, the screening operation of Equation (3) has the transfer function given by Equation (4):

$$h(\Lambda) = \sum_{k=0}^{\kappa} \phi_k \cdot \Lambda^k \quad (4)$$

This filter, based on local-information exchanges, captures information in K -radius proximity from the node representing the depth of the geometric-deep-learn-learning algorithm.

Taking into account this convolutional operation given by Equation (3), we are able to compute the f th level feature produced as output of the l th layer:

$$y_f^l = \sigma^l \cdot \left(\sum_{g=1}^{l-1} \mathcal{P}_{f,g}^l \cdot y_g^{l-1} \right) \quad (5)$$

where:

- σ^l represents the nonlinear activation function (i.e., ReLU); and
- $\mathcal{P}_{f,g}^l \cdot y_g^{l-1}$ indicates the graph structure relating the g th input y_g^{l-1} to the f th output y_f^l .

Now, we simply combine two cases to model the mechanism of a convolutional network applied to a non-Euclidean graph in each time slot: the case in which edges vary, and that in which nodes vary. This can be combined into a single expression to describe $\mathcal{P}_{f,g}^l$ given by Equation (6):

$$\mathcal{P}_{f,g}^l(S) = \sum_{k=1}^K Y_{f,g}^{l,(k:1)} + \sum_{k=0}^K \left(\prod_{m=0}^k Y_d^{(m)} + \phi_k \cdot \tau^k \right) \quad (6)$$

where:

- $\sum_{k=1}^K Y_{f,g}^{l,(k:1)}$ represents the edge-varying case, in which
 - * $Y_{f,g}^{l,(k:1)}$ acts as a shift operator, and therefore represents a learning paradigm for data embedded within complex graphs, whose weights are known to some degree of ambiguity, are only partially known, or are unknown.
 - $\sum_{k=0}^K \left(\prod_{m=0}^k Y_d^{(m)} + \phi_k \cdot \tau^k \right)$ represents the node-varying case, in which
 - * $d \subset E$ is a special set of nodes (i.e., nodes with a degree above a certain threshold, nodes with a certain level of hierarchy in the organization, or any other relevant feature),
 - * $\phi_k \in [0, 1]^{N \times d}$ is a binary matrix, and
 - * τ^k is a vector describing the node parameters in d .
- Pooling in non-Euclidean complex-networked cyber–physical graph time-dependent signals.

As introduced earlier, downsampling pooling layers in classical deep-learning architectures that extract information from Euclidean domains such as speech, images, or videos typically report the maximal output within rectangular proximity [93]. In this way, it is possible to extract local characteristics that are shared by other areas of the images, thus considerably reducing the number of parameters that the deep network has to learn without sacrificing its learning capacity. Pooling can be described as a progressive coarsening of the graph. A simple way to do this is to collapse edges and reduce the size of the graph through a standard max-pooling operation on the nodes by just taking the maximum of each one of the feature tensors on each of the nodes being coarsened. This can be represented as a binary-tree structure of node indices. These pooling modules on graphs can be inserted between the convolutional modules in order to extract high-level graph representations, and thus be able to perform effective graph classification.

Some alternatives in this field have not been to try to pool the whole network, but different hierarchies of the complex network in order to be able to learn which node groups have similar characteristics [94]. Once these groups are learned, clusters are made, and network pooling is carried out as described above or with an alternative method. This process is repeated for each of the network layers; thus, its classification is obtained. This presupposes, however, prior knowledge of the network structure.

The extraction of shared local characteristics is not possible through this method in time-varying non-Euclidean domains, i.e., complex-networked cyber–physical graphs, because no stationarity or shift invariance can be found within these domains. Wu et al. [95], and Lee et al. [96] provided state-of-the-art surveying overview of this interesting open research question.

5. Conclusions and Management Implications

Geometric deep lean learning at a strategic level is expected to ensure sustainable organizational growth because customers and suppliers are able to reach new levels of transparency and traceability on the quality and efficiency of the processes, which generates new business opportunities for both, and new products, services, and co-operation opportunities in a cyber-physical environment. In a world of limited resources, increasing business volume can only be achieved by increasing the depth of integrated intelligence capable of successfully handling the emerging complexity in value streams. The future implications of geometric deep lean learning at an organizational level are yet to be fully deployed, but it is expected that the field of analysis of complex-networked cyber-physical systems in Industry 4.0 environments will attract intense attention from both industry and scholars who could develop tools to interpret, classify, and better understand the behavioral patterns of such networks through the application of this very exciting field of artificial intelligence.

Managerial implications of geometric deep lean learning on a mesoscopic level should try to integrate geometric deep lean learning in whole-value-stream processes to substantially improve resource optimization. Geometric deep lean learning at a value-stream level is expected to impact lead time and on-time delivery. At a mesoscopic level, producing only what the customer needs, when they need it, in the required quality, the integration of deep-learning technologies is expected to not only allow the systematic improvement of complex value chains, but also the better use and exploitation of resources, thus reducing the environmental impact of Industry 4.0 processes. This technology could also be implemented at the customer side to increase defect-detection accuracy on products themselves. Such analyses provide sensitivity about operations and operational conditions, which also impacts value-stream-related efficiency and effectiveness.

The theoretical implications of the application of these geometric-deep-lean-learning models to data embedded within complex-networked datasets support researchers in departing from “crafted” features in modeling machine-learning models when dealing with geometric data. In the context of Industry 4.0 cyber-physical systems, these could be drone-positioning and decision-making algorithms, and the proper interpretation of wearable devices (i.e., physical sensors) on human or cybernetic process owners. Until now, models dealing with such problems required a certain amount of prior knowledge (e.g., the isometric-shape-deformation model), and often did not capture the full complexity and wealth of data. Geometric-deep-lean-learning methodologies could bring a breakthrough to the field and be the first indications of a coming paradigm shift by, for instance, expanding existing social-manufacturing knowledge into unknown territory through the contextual self-organizing of mass-individualization processes under a social-manufacturing paradigm through a cyber-physical-social system approach.

Some of the main potential applications can be clustered four categories:

- Graphwise classification. The classification of complex cyber-physical graphs by deep lean learning, thus creating product families and allowing automated decision making in real time in which products are developed, produced, and channeled to the final customer.
- Vertexwise classification. The classification of certain crucial nodes in the value-creation process by means of deep-lean-learning models that allows an improvement of organizational design to assure an increase of overall process performance.
- Graph dynamics. Learning complex-networked cyber-physical graph dynamics is of great interest when dealing with change management within non-Euclidean sociotechnical systems.
- Known vs. unknown domains. The learning, generation, and semisupervised design of value streams by learning the most suitable complex cyber-physical graphs for certain types of products, thus potentially generating high customization with high efficiency and effectiveness in resource use.

The data needed to implement these mathematical concepts are enormous and fall within the field of big data. The acquisition of data associated with the cyber-physical systems of Industry 4.0 is costly and of great strategic value to the involved organizations, which is why systems that

increase the confidence of the involved actors and guarantee the security of these IIoT data, as the distributed ledger technology, are essential for the practical application of the exposed concepts. The quality of the obtained data essentially depends on the trust that the various value-creating actors have in each other. Achieving the necessary high degrees of confidence and successfully managing these parameters in an environment of interdependent supplier and customer networks is one of the challenges in the immediate future, and ought to be met by several blockchain and distributed ledger protocols. The Constrained Application Protocol is excellent for use with limited devices and low-power networks, such as those preferred in IIoT. To ensure greater security, applications known by the more important User Datagram Protocol, such as Voice over IP/Session Initiation Protocol, Datagram Transport Layer Security, can be run on User Datagram Protocol instead of Transmission Control Protocol. The Rivest–Shamir–Adleman hybrid algorithm can also be good, with high efficiency, better security and privacy protection, and is suitable for the end-to-end encryption requirements of the future IIoT. Future IIoT research within an Industry 4.0 complex-networked cyber–physical context should focus on, among others, the following characteristics: the open security system, the way in which individual privacy is protected, terminal-security function, and laws related to IIoT security. It is undeniable that IIoT security requires a set of policies, laws, and regulations, and a perfect security-management system for mutual collocation to ensure the success of this exciting and fruitful research endeavor.

Author Contributions: For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used conceptualization, J.V.-D., M.M. and J.O.-M.; methodology, J.V.-D.; formal analysis, J.V.-D.; investigation, J.V.-D.; resources, S.S. and D.S.; writing—original draft preparation, J.V.-D.; writing—review and editing, J.V.-D., M.M. and J.O.-M.; visualization, J.V.-D.; supervision, J.V.-D.; project administration, J.V.-D.; funding acquisition, J.V.-D. and W.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the European Commission through the RFCS program, grant number ID 793505. This research was partially funded by the Ministerium für Wissenschaft, Forschung, und Kunst Baden-Württemberg (MWK), Germany, as part of idea competition “Mobility concepts for the emission-free campus ‘NaMoCa’.” The authors thank the China Scholarship Council (CSC) (no. 201206730038) for the support they provided for this work. The authors would also like to acknowledge the Spanish Agencia Estatal de Investigación, through research project code RTI2018-094614-B-I00 into the “Programa Estatal de I+D+i Orientada a los Retos de la Sociedad”.

Acknowledgments: J.V.D. would like to thank Paloma García-Lázaro (RIP), María Jesús Sánchez-Naranjo, and Eva Sánchez-Mañes for their magnificent job as professors of mathematics at the Escuela Técnica Superior de Ingenieros Industriales of the Universidad Politécnica de Madrid, for their inspiration, vitality, and energy in transmitting knowledge. All thanks to you, for teaching me these things that have greatly influenced my life and my work.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

IIoT Industrial Internet of Things

References

1. Reinsel, D.; Gantz, J.; Rydning, J. The Digitization of the World. From Edge to Core. 2018. Available online: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (accessed on 30 January 2020).
2. Ahmed, E.; Yaqoob, I.; Hashem, I.A.T.; Khan, I.; Ahmed, A.I.A.; Imran, M.; Vasilakos, A.V. The role of big data analytics in Internet of Things. *Comput. Netw.* **2017**, *129*, 459–471. [CrossRef]
3. Bhattacharjya, A.; Zhong, X.; Wang, J.; Li, X. Security Challenges and Concerns of Internet of Things (IoT). In *Cyber-Physical Systems: Architecture, Security and Application*; Guo, S., Zeng, D., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 153–185. [CrossRef]
4. Jiang, P.; Leng, J.; Ding, K.; Gu, P.; Koren, Y. Social manufacturing as a sustainable paradigm for mass individualization. *Proc. Inst. Mech. Eng. Part B J. Eng. Manuf.* **2016**, *230*, 1961–1968. [CrossRef]

5. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*, 1st ed.; Adaptive Computation and Machine Learning Series; The MIT Press: Cambridge, MA, USA, 2016.
6. Hermann, M.; Pentek, T.; Otto, B. Design Principles for Industrie 4.0 Scenarios. In Proceedings of the 2016 49th Hawaii International Conference on System Sciences (HICSS), Koloa, HI, USA, 5–8 January 2016; pp. 3928–3937. [\[CrossRef\]](#)
7. Mosterman, P.J.; Zander, J. Industry 4.0 as a Cyber-Physical System study. *Softw. Syst. Model.* **2016**, *15*, 17–29. [\[CrossRef\]](#)
8. Shah, R.; Ward, P.T. Lean manufacturing: Context, practice bundles, and performance. *J. Oper. Manag.* **2003**, *21*, 129–149. [\[CrossRef\]](#)
9. Villalba-Diez, J.; Ordieres-Mere, J.; Chudzick, H.; Lopez-Rojo, P. NEMAWASHI: Attaining Value Stream alignment within Complex Organizational Networks. *Procedia CIRP* **2015**, *37*, 134–139. [\[CrossRef\]](#)
10. Villalba-Diez, J.; Ordieres-Mere, J. Improving manufacturing operational performance by standardizing process management. *Trans. Eng. Manag.* **2015**, *62*, 351–360. [\[CrossRef\]](#)
11. Villalba-Diez, J.; Ordieres-Meré, J. Strategic Lean Organizational Design: Towards Lean World-Small World Configurations through Discrete Dynamic Organizational Motifs. *Math. Probl. Eng.* **2016**, *2016*, 1–10. [\[CrossRef\]](#)
12. Villalba-Diez, J.; Ordieres-Mere, J.; Rubio-Valdehita, S. Lean Learning Patterns. (CPD)nA vs. KATA. *Procedia CIRP* **2016**, *54*, 147–151. [\[CrossRef\]](#)
13. Villalba-Diez, J.; Ordieres-Meré, J.; Molina, M.; Rossner, M.; Lay, M. Lean Dendrochronology: Complexity Reduction by Representation of KPI Dynamics Looking at Strategic Organizational Design. *Manag. Prod. Eng. Rev.* **2018**, *9*, 3–9. [\[CrossRef\]](#)
14. Villalba-Diez, J. *The Lean Brain Theory. Complex Networked Lean Strategic Organizational Design*; CRC Press: Boca Raton, FL, USA, 2017.
15. Villalba-Diez, J. *The HOSHIN KANRI FOREST. Lean Strategic Organizational Design*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2017.
16. Cross, R.L.; Singer, J.; Colella, S.; Thomas, R.J.; Silverstone, Y. (Eds.) *The Organizational Network Fieldbook: Best Practices, Techniques and Exercises to Drive Organizational Innovation and Performance*, 1st ed.; Jossey-Bass: San Francisco, CA, USA, 2010.
17. Jabeur, N.; Sahli, N.; Zeadally, S. Enabling Cyber Physical Systems with Wireless Sensor Networking Technologies, Multiagent System Paradigm, and Natural Ecosystems. *Mob. Inf. Syst.* **2015**, *2015*, 15. [\[CrossRef\]](#)
18. Saleh, M.; Esa, Y.; Mohamed, A. Applications of Complex Network Analysis in Electric Power Systems. *Energies* **2018**, *11*, 1381. [\[CrossRef\]](#)
19. Barabási, A.-L. *Network Science*; Cambridge University Press: Cambridge, UK, 2016.
20. Wang, Y.; Yuan, Y.; Ma, Y.; Wang, G. Time-Dependent Graphs: Definitions, Applications, and Algorithms. *Data Sci. Eng.* **2019**. [\[CrossRef\]](#)
21. Suzuki, K. *The New Shopfloor Management*; The Free Press: New York, NY, USA, 1993.
22. De Leeuw, S.; van der Berg, J.P. Improving operational performance by influencing shopfloor behavior via performance management practices. *J. Oper. Manag.* **2011**, *29*, 224–235. [\[CrossRef\]](#)
23. Villalba-Diez, J.; Ordieres-Meré, J.; Nuber, G. The HOSHIN KANRI TREE. Cross-Plant Lean Shopfloor Management. *Procedia CIRP* **2015**, *32*, 150–155. [\[CrossRef\]](#)
24. Villalba-Diez, J.; Zheng, X.; Schmidt, D.; Molina, M. Characterization of Industry 4.0 Lean Management Problem-Solving Behavioral Patterns Using EEG Sensors and Deep Learning. *Sensors* **2019**, *19*, 2841. [\[CrossRef\]](#)
25. Villalba-Diez, J.; Schmidt, D.; Gevers, R.; Ordieres-Meré, J.; Buchwitz, M.; Wellbrock, W. Deep Learning for Industrial Computer Vision Quality Control in the Printing Industry 4.0. *Sensors* **2019**, *19*, 3987. [\[CrossRef\]](#)
26. Imai, M. *Gemba Kaizen: A Commonsense Approach to a Continuous Improvement Strategy*, 2nd ed.; McGraw-Hill Professional: New York, NY, USA, 2012.
27. Stock, T.; Seliger, G. Opportunities of Sustainable Manufacturing in Industry 4.0. *Procedia CIRP* **2016**, *40*, 536–541. [\[CrossRef\]](#)
28. Takeda, H. *Intelligent Automation Textbook*; Nikkan Kogyo Shimbun: Tokyo, Japan, 2009.
29. Francis, J.; Bian, L. Deep Learning for Distortion Prediction in Laser-Based Additive Manufacturing using Big Data. *Manuf. Lett.* **2019**, *20*, 10–14. [\[CrossRef\]](#)

30. Li, B.H.; Hou, B.C.; Yu, W.T.; Lu, X.B.; Yang, C.W. Applications of artificial intelligence in intelligent manufacturing: A review. *Front. Inf. Technol. Electron. Eng.* **2017**, *18*, 86–96. [\[CrossRef\]](#)
31. Aazam, M.; Zeadally, S.; Harras, K.A. Deploying Fog Computing in Industrial Internet of Things and Industry 4.0. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4674–4682. [\[CrossRef\]](#)
32. Tao, F.; Qi, Q.; Liu, A.; Kusiak, A. Data-driven smart manufacturing. *J. Manuf. Syst.* **2018**, *48*, 157–169. [\[CrossRef\]](#)
33. Mushtaq, A.; Haq, I.U. Implications of Blockchain in Industry 4.0. In Proceedings of the 2019 International Conference on Engineering and Emerging Technologies (ICEET), Lahore, Pakistan, 21–22 February 2019; pp. 1–5. [\[CrossRef\]](#)
34. Shevchik, S.A.; Masinelli, G.; Kenel, C.; Leinenbach, C.; Wasmer, K. Deep Learning for In Situ and Real-Time Quality Monitoring in Additive Manufacturing Using Acoustic Emission. *IEEE Trans. Ind. Inform.* **2019**, *15*, 5194–5203. [\[CrossRef\]](#)
35. Al-Jaroodi, J.; Mohamed, N. Blockchain in Industries: A Survey. *IEEE Access* **2019**, *7*, 36500–36515. [\[CrossRef\]](#)
36. Sun, S.; Zheng, X.; Villalba-Diez, J.; Ordieres-Meré, J. Indoor Air-Quality Data-Monitoring System: Long-Term Monitoring Benefits. *Sensors* **2019**, *19*, 4157. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Rother, M. *Toyota Kata. Managing People for Improvement, Adaptiveness, and Superior Results*; McGraw-Hill: New York, NY, USA, 2010.
38. Birkel, S.H.; Veile, W.J.; Müller, M.J.; Hartmann, E.; Voigt, K.I. Development of a Risk Framework for Industry 4.0 in the Context of Sustainability for Established Manufacturers. *Sustainability* **2019**, *11*, 384. [\[CrossRef\]](#)
39. Takeda, H. *LCIA—Low Cost Intelligent Automation*, 3rd ed.; Finanzbuch Verlag GmbH: Munich, Germany, 2011.
40. Davis, J.; Edgar, T.; Porter, J.; Bernaden, J.; Sarli, M. Smart manufacturing, manufacturing intelligence and demand-dynamic performance. *Comput. Chem. Eng.* **2012**, *47*, 145–156. [\[CrossRef\]](#)
41. Gómez, A.; Cuiñas, D.; Catalá, P.; Xin, L.; Li, W.; Conway, S.; Lack, D. Use of Single Board Computers as Smart Sensors in the Manufacturing Industry. *MESIC Manuf. Eng. Soc. Int. Conf.* **2015**, *132*, 153–159. [\[CrossRef\]](#)
42. Culot, G.; Orzes, G.; Sartor, M. Integration and Scale in the Context of Industry 4.0: The Evolving Shapes of Manufacturing Value Chains. *IEEE Eng. Manag. Rev.* **2019**, *47*, 45–51. [\[CrossRef\]](#)
43. Jiménez, P.; Villalba-Diez, J.; Ordieres-Meré, J. HOSHIN KANRI Visualization with Neo4j. Empowering Leaders to Operationalize Lean Structural Networks. *Procedia CIRP* **2016**, *55*, 284–289.
44. Wang, J.; Ma, Y.; Zhang, L.; Gao, R.; Wu, D. Deep learning for smart manufacturing: Methods and applications. *J. Manuf. Syst.* **2018**. [\[CrossRef\]](#)
45. Jang, S.H.; Guejong, J.; Jeong, J.; Sangmin, B. *Fog Computing Architecture Based Blockchain for Industrial IoT. Computational Science—ICCS 2019*; Rodrigues, J.M.F., Cardoso, P.J.S., Monteiro, J., Lam, R., Krzhizhanovskaya, V.V., Lees, M.H., Dongarra, J.J., Sloot, P.M., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 593–606.
46. Ordieres-Meré, J.; Villalba-Diez, J.; Zheng, X. *Challenges and Opportunities for Publishing IIoT Data in Manufacturing as a Service Business*; Cyber Physical Manufacturing; Procedia Manufacturing; Elsevier: Chicago, IL, USA, 2019.
47. Villalba-Diez, J.; DeSanctis, I.; Ordieres-Meré, J.; Ciarapica, F. Lean Structural Network Resilience. In *Complex Networks and Its Applications VI: Proceedings of Complex Networks 2017 (The Sixth International Conference on Complex Networks and Their Applications)*; Cherifi, C., Cherifi, H., Karsai, M., Musolesi, M., Eds.; Number 689 in Studies in Computational Intelligence; Springer International Publishing: Lyon, France, 2017; pp. 609–619.
48. Davies, R.; Coole, T.; Smith, A. Review of Socio-technical Considerations to Ensure Successful Implementation of Industry 4.0. *Procedia Manuf.* **2017**, *11*, 1288–1295. [\[CrossRef\]](#)
49. Kumar, K.; Zindani, D.; Davim, J.P. Socio-technical Considerations. In *Industry 4.0: Developments towards the Fourth Industrial Revolution*; Kumar, K., Zindani, D., Davim, J.P., Eds.; Springer: Singapore, 2019; pp. 43–51. [\[CrossRef\]](#)
50. Womack, J.P.; Jones, D.T. Introduction. In *Lean Thinking*, 2nd ed.; Simon and Schuster: New York, NY, USA, 2003; pp. 15–28.
51. Toyota Motor Corporation. *Toyota Motor Corporation. Sustainability Report 2013*; Sustainability Report; Toyota Motor Corporation: Tokyo, Japan, 2014.
52. Burton, R.M.; Øbel, B.; Håkansson, D.D. *Organizational Design: A Step-by-Step Approach*, 3rd ed.; Cambridge University Press: Cambridge, UK, 2015.
53. Covey, S.R. *The 8th Habit. From Effectiveness to Greatness*; Free Press: New York, NY, USA, 2004.

54. Rabelo, R.J.; Zambiasi, S.P.; Romero, D. Collaborative Softbots: Enhancing Operational Excellence in Systems of Cyber-Physical Systems. In *Collaborative Networks and Digital Transformation*; Camarinha-Matos, L.M., Afsarmanesh, H., Antonelli, D., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 55–68.
55. Romero, D.; Wuest, T.; Stahre, J.; Gorecky, D. Social Factory Architecture: Social Networking Services and Production Scenarios Through the Social Internet of Things, Services and People for the Social Operator 4.0. In *Advances in Production Management Systems. The Path to Intelligent, Collaborative and Sustainable Manufacturing*; Lödding, H., Riedel, R., Thoben, K.D., von Cieminski, G., Kiritsis, D., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 265–273.
56. Wang, F.; Shang, X.; Qin, R.; Xiong, G.; Nyberg, T.R. Social Manufacturing: A Paradigm Shift for Smart Prosumers in the Era of Societies 5.0. *IEEE Trans. Comp. Social Syst.* **2019**, *6*, 822–829. [[CrossRef](#)]
57. Guo, W.; Jiang, P. Product Service Systems for Social Manufacturing: A new service system with multi-provider. *IFAC-PapersOnLine* **2019**, *52*, 749–754. [[CrossRef](#)]
58. Lee, J.; Bagheri, B.; Kao, H.A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [[CrossRef](#)]
59. Wang, L.; Toerngren, M.; Onori, M. Current status and advancement of cyber-physical systems in manufacturing. *J. Manuf. Syst.* **2015**, *37*, 517–527. [[CrossRef](#)]
60. Sisinni, E.; Saifullah, A.; Han, S.; Jennehag, U.; Gidlund, M. Industrial Internet of Things: Challenges, Opportunities, and Directions. *IEEE Trans. Ind. Inform.* **2018**, *14*, 4724–4734. [[CrossRef](#)]
61. Zheng, P.; Wang, H.; Sang, Z.; Zhong, R.Y.; Liu, Y.; Liu, C.; Mubarak, K.; Yu, S.; Xu, X. Smart manufacturing systems for Industry 4.0: Conceptual framework, scenarios, and future perspectives. *Front. Mech. Eng.* **2018**, *13*, 137–150. [[CrossRef](#)]
62. Lu, Y.; Xu, X. Cloud-based manufacturing equipment and big data analytics to enable on-demand manufacturing services. *Robot. Comput. Integr. Manuf.* **2019**, *57*, 92–102. [[CrossRef](#)]
63. Kiel, D.; Müller, J.M.; Arnold, C.; Voigt, K.I. Sustainable Industrial Value Creation: Benefits and Challenges of Industry 4.0. *Int. J. Innov. Manag.* **2017**, *21*, 1740015. [[CrossRef](#)]
64. Stock, T.; Obenaus, M.; Kunz, S.; Kohl, H. Industry 4.0 as enabler for a sustainable development: A qualitative assessment of its ecological and social potential. *Process Saf. Environ. Prot.* **2018**, *118*, 254–267. [[CrossRef](#)]
65. Shang, X.; Shen, Z.; Xiong, G.; Wang, F.Y.; Liu, S.; Nyberg, T.R.; Wu, H.; Guo, C. Moving from mass customization to social manufacturing: A footwear industry case study. *Int. J. Comput. Integr. Manuf.* **2019**, *32*, 194–205. [[CrossRef](#)]
66. Leng, J.; Jiang, P. A deep learning approach for relationship extraction from interaction context in social manufacturing paradigm. *Knowl. Based Syst.* **2016**, *100*, 188–199. [[CrossRef](#)]
67. Leng, J.; Chen, Q.; Mao, N.; Jiang, P. Combining granular computing technique with deep learning for service planning under social manufacturing contexts. *Knowl. Based Syst.* **2018**, *143*, 295–306. [[CrossRef](#)]
68. Subakti, H.; Jiang, J.-R. Indoor Augmented Reality Using Deep Learning for Industry 4.0 Smart Factories. In Proceedings of the 2018 IEEE 42nd Annual Computer Software and Applications Conference (COMPSAC), Tokyo, Japan, 23–27 July 2018; Volume 2, pp. 63–68. [[CrossRef](#)]
69. Tao, F.; Qi, Q.; Wang, L.; Nee, A.Y.C. Digital Twins and Cyber-Physical Systems toward Smart Manufacturing and Industry 4.0: Correlation and Comparison. *Engineering* **2019**, *5*, 653–661. [[CrossRef](#)]
70. Bronstein, M.M.; Bruna, J.; LeCun, Y.; Szlam, A.; Vandergheynst, P. Geometric Deep Learning: Going beyond Euclidean data. *IEEE Signal Process. Mag.* **2017**, *34*, 18–42. [[CrossRef](#)]
71. Ma, Y.; Fu, Y. *Manifold Learning Theory and Applications*; CRC Press: Boca Raton, FL, USA, 2011.
72. Keller, M. Curvature, Geometry and Spectral Properties of Planar Graphs. *Discret. Comput. Geom.* **2011**, *46*, 500–525. [[CrossRef](#)]
73. Wu, Z.; Menichetti, G.; Rahmede, C.; Bianconi, G. Emergent complex network geometry. *Sci. Rep.* **2015**, *5*, 10073. [[CrossRef](#)]
74. Bianconi, G.; Rahmede, C.; Wu, Z. Complex quantum network geometries: Evolution and phase transitions. *Phys. Rev. E Stat. Nonlinear Soft Matter Phys.* **2015**, *92*, 022815. [[CrossRef](#)]
75. Bianconi, G.; Rahmede, C. Complex Quantum Network Manifolds in Dimension $d > 2$ are Scale-Free. *Sci. Rep.* **2015**, *5*, 13979. [[CrossRef](#)]
76. Zhang, Z.; Cui, P.; Zhu, W. Deep Learning on Graphs: A Survey. *arXiv* **2018**, arXiv:1812.04202.

77. Duvenaud, D.; Maclaurin, D.; Aguilera-Iparraguirre, J.; Gómez-Bombarelli, R.; Hirzel, T.; Aspuru-Guzik, A.; Adams, R.P. Convolutional Networks on Graphs for Learning Molecular Fingerprints. *arXiv* **2015**, arXiv:1509.09292.
78. Krizhevsky, A.; Sutskever, L.; Hinton, G. Imagenet classification with deep convolutional neural networks. In Proceedings of the 25th International Conference on Neural Information Processing Systems, Lake Tahoe, CA, USA, 3–8 December 2012; pp. 1106–1114.
79. Chen, L.C.; Barron, J.T.; Papandreou, G.; Murphy, K.; Yuille, A.L. Semantic Image Segmentation with Task-Specific Edge Detection Using CNNs and a Discriminatively Trained Domain Transform. *arXiv* **2015**, arXiv:1511.03328.
80. Cao, S.; Lu, W.; Xu, Q. Deep Neural Networks for Learning Graph Representations. In Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence, Phoenix, AZ, USA, 12–17 February 2016; pp. 1145–1152.
81. Shuman, D.; Narang, S.; Frossard, P.; Ortega, A.; Vandergheynst, P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. *IEEE Signal Process. Mag.* **2013**, *30*, 83–98. [[CrossRef](#)]
82. Henaff, M.; Bruna, J.; LeCun, Y. Deep convolutional networks on graph-structured data. *arXiv* **2015**, arXiv:1506.05163.
83. Defferrard, M.; Bresson, X.; Vandergheynst, P. Convolutional Neural Networks on Graphs with Fast Localized Spectral Filtering. *arXiv* **2016**, arXiv:1606.09375.
84. Kipf, T.N.; Welling, M. Semi-Supervised Classification with Graph Convolutional Networks. *arXiv* **2016**, arXiv:1609.02907.
85. Qiu, J.; Tang, J.; Ma, H.; Dong, Y.; Wang, K.; Tang, J. DeepInf: Social Influence Prediction with Deep Learning. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, London, UK, 19–23 August 2018; ACM: New York, NY, USA, 2018; pp. 2110–2119. [[CrossRef](#)]
86. Garcia, V.; Bruna, J. Few-Shot Learning with Graph Neural Networks. *arXiv* **2017**, arXiv:1711.04043.
87. Jain, A.; Zamir, A.R.; Savarese, S.; Saxena, A. Structural-RNN: Deep Learning on Spatio-Temporal Graphs. *arXiv* **2015**, arXiv:1511.05298.
88. Marino, K.; Salakhutdinov, R.; Gupta, A. The More You Know: Using Knowledge Graphs for Image Classification. *arXiv* **2016**, arXiv:1612.04844.
89. Sandryhaila, A.; Moura, J.M.F. Discrete Signal Processing on Graphs. *IEEE Trans. Signal Process.* **2013**, *61*, 1644–1656. [[CrossRef](#)]
90. Harris, K.D.; Aravkin, A.; Rao, R.; Brunton, B.W. Time-varying Autoregression with Low Rank Tensors. *arXiv* **2019**, arXiv:1905.08389.
91. Sciriha, I. A characterization of singular graphs. *Electron. J. Linear Algebra* **2007**, *16*, 451–462. [[CrossRef](#)]
92. Ortega, A.; Frossard, P.; Kovačević, J.; Moura, J.M.; Vandergheynst, P. Graph Signal Processing: Overview, Challenges, and Applications. *Proc. IEEE* **2018**, *106*, 808–828. [[CrossRef](#)]
93. Zhou, Y.T.; Chellappa, R. Computation of optical flow using a neural network. In Proceedings of the IEEE 1988 International Conference on Neural Networks, San Diego, CA, USA, 24–27 July 1988; Volume 2, pp. 71–78. [[CrossRef](#)]
94. Ying, Z.; You, J.; Morris, C.; Ren, X.; Hamilton, W.; Leskovec, J. Hierarchical Graph Representation Learning with Differentiable Pooling. In *Advances in Neural Information Processing Systems 31*; Bengio, S., Wallach, H., Larochelle, H., Grauman, K., Cesa-Bianchi, N., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2018; pp. 4800–4810.
95. Wu, Z.; Pan, S.; Chen, F.; Long, G.; Zhang, C.; Yu, P.S. A Comprehensive Survey on Graph Neural Networks. *arXiv* **2019**, arXiv:1901.00596.
96. Lee, J.; Lee, I.; Kang, J. Self-Attention Graph Pooling. In Proceedings of the 36th International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; Chaudhuri, K., Salakhutdinov, R., Eds.; PMLR: Long Beach, CA, USA, 2019; Volume 97, pp. 3734–3743.

3.6 Publication 6

- **Title:** *Healthy Operator 4.0: A Human Cyber-Physical System Architecture for Smart Workplaces*
- **Authors:** Shengjing Sun, Xiaochen Zheng, Bing Gong, Jorge García Paredes and Joaquín Ordieres-Meré
- **Publication date:** 2020/04/03
- **Journal:** *Sensors*
3.031 (2018) ; 5-Year Impact Factor: 3.302 (2018)
Quartile: Q1 in ‘Instruments & Instrumentation’
ISSN: 1424-8220
Publisher: Multidisciplinary Digital Publishing Institute, BASEL, SWITZERLAND
- **Volume:** 20
- **Issue:** 7
- **Pages:** 2011
- **DOI:** <https://doi.org/10.3390/s20072011>

Article

Healthy Operator 4.0: A Human Cyber–Physical System Architecture for Smart Workplaces

Shengjing Sun ^{1,2} , Xiaochen Zheng ^{1,3} , Bing Gong ⁴ , Jorge García Paredes ¹ and Joaquín Ordieres-Meré ^{1,*}

¹ Escuela Técnica Superior de Ingenieros Industriales (ETSII), Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain; shengjing.sun@alumnos.upm.es (S.S.); xiaochen.zheng@epfl.ch (X.Z.); j.gparedes@alumnos.upm.es (J.G.P.)

² Exposure, Epidemiology, and Risk Program, Department of Environmental Health, Harvard T.H. Chan School of Public Health, Boston, MA 02115, USA

³ ICT for Sustainable Manufacturing, SCI-STI-DK, École polytechnique fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

⁴ Jülich Supercomputing Center, Forschungszentrum Jülich GmbH, Wilhelm-Wohnen-Str, 52428 Jülich, Germany; b.gong@fz-juelich.de

* Correspondence: j.ordieres@upm.es

Received: 18 March 2020; Accepted: 1 April 2020; Published: 3 April 2020

Abstract: Recent advances in technology have empowered the widespread application of cyber–physical systems in manufacturing and fostered the Industry 4.0 paradigm. In the factories of the future, it is possible that all items, including operators, will be equipped with integrated communication and data processing capabilities. Operators can become part of the smart manufacturing systems, and this fosters a paradigm shift from independent automated and human activities to human–cyber–physical systems (HCPSS). In this context, a Healthy Operator 4.0 (HO4.0) concept was proposed, based on a systemic view of the Industrial Internet of Things (IIoT) and wearable technology. For the implementation of this relatively new concept, we constructed a unified architecture to support the integration of different enabling technologies. We designed an implementation model to facilitate the practical application of this concept in industry. The main enabling technologies of the model are introduced afterward. In addition, a prototype system was developed, and relevant experiments were conducted to demonstrate the feasibility of the proposed system architecture and the implementation framework, as well as some of the derived benefits.

Keywords: healthy operator 4.0; human–cyber–physical system; industrial internet of things; industry 4.0; smart workplaces

1. Introduction

The continuous technological innovations in the domains of information technology (IT), the Internet of Things (IoT), and artificial intelligence (AI), among others, have significantly changed production systems [1–3]. Recent advances of these technologies have enabled a systematical implementation of cyber–physical systems (CPS) in manufacturing, which has significantly improved the efficiency of production systems and made them perform more resiliently and collaboratively. These advanced technologies are transforming the manufacturing industry to the Industry 4.0 paradigm [4,5].

In the Industry 4.0 era, all items in a smart factory will be equipped with integrated communication and certain data processing capabilities, by the so called Industrial Internet of Things (IIoT). The IIoT scenario enables a close connection between the physical and digital worlds, and the paradigm of cyber–physical equivalence or digital twin emerges [6], where the vision of the digital twin

itself refers to a comprehensive physical and functional description of a component, product, or system, which includes more or less all information that could be useful in all lifecycle phases [7]. This represents a seamless integration between both worlds, meaning that the digital part can virtually replicate the behavior of the physical counterpart and, henceforth, be used to create new added value services in both directions.

The smart “artifacts”, including people, will be connected to the CPS [8,9], which is different from computer integrated manufacturing (CIM), as the intention of Industry 4.0 paradigm is not creating unmanned production facilities. In contrast, it provides a great opportunity for operators to become part of the smart manufacturing systems in such a way where the individual skills and talents of the operators can be better realized [10–12].

The advanced technologies, such as IoT and CPS, in the Industry 4.0 paradigm provide new forms of interaction between operators and machines that will produce new intelligent workforces and will bring significant impacts on the nature of manufacturing [13,14]. Human-centricity is one of the most critical focuses in the transformation towards Industry 4.0. It allows for a paradigm shift from independent, automated, and human activities to human–cyber–physical systems (HCPSs) where machines are not designed to replace the skills and talents of humans, but rather to co-exist with, and to assist humans in being more efficient and effective [15,16]. A clear opportunity in such a path is to align the human central view in the decisions for process improvement, where improvement in the classical lean management, e.g., (CPD)_nA [17], was performed through a strategy enabling taking advantage of the IoT devices, and looking to increase the knowledge analysis capabilities of the responsible worker. Therefore, the aims of HCPSs can be described [12,15] as:

- Empowering people to dynamically interact with machines in both the cyber and physical worlds to fit the operators’ cognitive and physical needs supported by intelligent human–computer interaction techniques;
- Improving the physical, sensing, and cognitive capabilities of people taking advantages from diverse technologies, such as IoT and wearable technologies.

On the basis of HCPSs, the concept of Operator 4.0 was proposed, aiming at improving the cooperation between humans and machines [13,15]. Operator 4.0 represents “a new design and engineering philosophy for adaptive production systems where the focus is on treating automation as a further enhancement of the human’s physical, sensorial, and cognitive capabilities” [15].

According to different enabling technologies and targeting aspects, Operator 4.0 can be empowered by different technologies [12,13], such as,

- Virtual operator, enabled by virtual reality (VR)/augmented reality (AR) [18],
- Super-strength operator, enabled by exoskeletons [19],
- Smarter operator, enabled by intelligent personal assistant (IPA)-based solutions [20],
- Healthy operator, enabled by wearable technologies combined with advanced data analytic techniques [21],
- etc.

As one of the core sub-types of Operator 4.0, the Healthy Operator 4.0 (HO4.0) aims to address the concerns regarding increasing workforce stress levels, the state of psycho-social health [22,23], and the new potential physical risks [24,25] in the cyber–physical production environments caused by the introduction of new Industry 4.0 technologies, including autonomous and collaborative robots, AR, and VR.

Although occupational health has been a popular topic for decades in both academic research and industrial applications, it has rarely had the advantage of real time quantitative measurements, mainly due to the limited capabilities for directly measuring such factors [21,26–31].

The recent advances in sensing technologies and IoT, especially wearable technologies, provide new solutions for real-time monitoring of an operator’s activities, locations, vital signs, etc., as well as

the status of the surrounding workplace environment [30–32]. These technologies make it possible to develop health-related applications, such as alerting operators of possible exposure to hazardous environments, avoiding collisions with moving heavy equipment, and preventing anti-ergonomic body movements and postures [21,33]. However, in current industrial practice, most applications are developed in isolated circumstances aimed at addressing specific problems. Therefore, there is a gap in creating human-centered systems able to promote operators' learning context not only relying on single parameters but also providing a meaningful articulated set of relevant parameters both in the short and long term.

To cope with this challenge, a unified architecture for HO4.0 is required to support the integration of different enabling technologies, thus guiding the implementation of this concept in the context of occupational safety and health. The focus of the paper is established to:

- Formally define the HO4.0 concept, by carefully analyzing the contributions from previous research in relevant fields,
- Propose a HO4.0 application framework enabling relevant data integration in highly dynamic work environments, supported by the IIoT and wearable technology. According to the definitions, we presented a multi-layer architecture and an implementation framework to guide the their application in industry;
- Provide a prototype system, developed according to the proposed framework, where the assessment of its capabilities can be derived; and
- Conduct experiments to verify the capabilities of the prototype system to provide new knowledge on real applications.

From the methodological point of view, this study followed the critical action research method, in which the researcher and a field operator collaborated in the diagnosis of the problem and in the development of a solution based on the diagnosis [34]. Critical action research is based on the analysis, action, evaluation, and critical analysis of practices based on collected data, in order to introduce improvements in the relevant practices. This type of research is facilitated by the participation and collaboration of a number of individuals with a common purpose where the research focuses on specific situations and their context.

The rest of this paper is organized as follows: Section 2 introduces the formal definition for HO4.0, enabling us to build over the proposed framework, in Section 3. Section 4 introduces the application case supporting the prototype creation. The experiments and outcomes are presented in Section 5. Then Section 6 summarizes the findings and provides our interpretation in the context of the proposed HO4.0 framework. Finally, Section 7 provides a more strategic perspective of the knowledge gains, including the limitations and further research directions.

2. HO4.0 Concept

As the sub-type of Operator 4.0, the HO4.0 concept was first envisioned in [12], where it was roughly depicted as *Operator + Wearable Tracker = Healthy Operator (physical and cognitive interaction)*. In this stage, individual wearable devices, such as smartwatches, were used to track an operator's health-related metrics. However, an isolated and separate monitoring dimension can not fully represent comprehensive health aspects. Therefore, a unified health view of the operator is needed to enable holistic workforce health management and analytics. In this study, aligned with the Operator 4.0 definition [13,15], and extending the former concept, the HO4.0 is formally defined as a system focused on health and operator well-being, looking to facilitate the operators' empowerment by enabling relevant knowledge creation, including modeling their behaviors.

It is relevant that our adopted perspective differs from the one chosen in [12,21], where the concept was closer to the operator. It was introduced as a specific type or vision for the Operator 4.0 concept. Instead, the ambition here was to consider the relevant information gathered, when appropriately fused, as a digital image of the operator's behavior, and due to this, the digital twin

approach comes in naturally. To facilitate the integration of concepts, the natural approach is to consider the HO4.0 as a system instead of a type of operator. Under this interpretation, the HO4.0 can be also presented as a virtual system: the HO4.0 digital twin, gathering all the health related information from operators and potentially enabling the learning of rules from the different behaviors. This does include benchmarks regarding the health impact of different operator routines, etc.

The HO4.0 digital twin is powered by IIoT networks, wearable technologies, ambient intelligence, and modeling technologies. It not only enables real-time health risk information, e.g., risk alert, but also make it possible to simulate the future behavior of operators; to infer and forecast the evolution from their behavior at mid-long term ranges, aiming to reduce the operator's cognitive and physical workload, and increase the operator's well-being such as Occupational Safety and Health (OSH), job satisfaction, work-related affect, and enhance the workforce productivity in the Industry 4.0 context.

3. HO4.0 Framework

The proposed framework has two main components, one is the architecture, and the other involves the required technologies.

3.1. Proposed Architecture

A traditional CPS structure usually includes five levels (5C structure), including Smart Connection, Data-to-information Conversion, Cyber, Cognition and Configuration [4]. Based on the Operator 4.0 definition, we adapted the 5C structure of CPS and designed a four-level architecture for HO4.0, compatible with the 5C CPS structure, as the proposed architecture shown in Figure 1 is composed of the Smart Connection layer, the Integration and Communication layer, the Modeling layer, and the Cognition layer.

- **Smart Connection layer:** Accurate and reliable data from operators, machines, the ambient environment, and other parts of the production system are the foundation of the HO4.0 system. The sensing layer collected these data utilizing wearable devices, ambient sensors, and the industrial sensors of manufacturing systems etc. Wearable devices, such as smart watches, smart bands, smart glasses, smart shoes, and smart helmets, contain a variety of sensors that can measure the operator's location, movements, heart rate, blood pressure, body temperature, concentration level (e.g., fatigue), foot pressure, etc. Stationary or portable ambient sensors, such as air quality, thermal (temperature and humidity) sensors, and acoustic and light sensors, are used to monitor the conditions of the workplace environment surroundings, such as the indoor air quality level, temperature, humidity, pollutant concentration, light intensity, and noise levels. The indoor positioning system (IPS) was adopted to detect the movement and location of the operators, vehicles, work pieces, machines, etc. in a workplace.
- **Integration and Communication layer:** This layer was mainly composed of edge computing devices or gateways, which could include a smartphone, tablet, router, or single-board computer (SBC) etc. This layer was required due to the limitations of the size, power supply, and computing capabilities of most wearable devices and ambient sensors from the sensing layer. They are not capable of performing operations such as data filtering and integration. This layer aimed to filter and integrate these diverse data, and if necessary, to convert and pre-process the raw data based on different data modeling approaches. The sensors used in the sensing layer are from many domains, e.g., monitoring human psychological behaviors and detecting environmental conditions. This layer also facilitated the data uploading to the server storage for further data analysis. The edge computing device in this layer can be applied to provide real-time health warnings that are critical for a human that interacts with machines and HCPSs.
- **Modeling layer:** The health related data was fused and modeled in this layer. The twin model of operator health was created and relevant simulations can be run to apply the derived high lever rules. This is typically conducted in a local or remote/cloud server with a higher computing

capability that is able to care about data streams from different communication devices. Semantic data description enables the interoperability and usage of semantic models, and knowledge graphs might also be employed to facilitate the knowledge extraction. Supported by advanced machine learning techniques and semantic-driven data fusion, useful knowledge, such as risk alerts, improved advice, and rules, can be extracted.

- **Cognition layer:** Based on the results from the model layer, the cognition layer can provide hints and insights from cyber space to physical space and acts as a monitoring system for the preventive decisions from operators, machines, or ambient environments. On this layer, analyzing the results can support the decision-making and can be presented with proper data visualization techniques. Certain user interfaces, applets, or web-based services can be also implemented. The alerts, advice, and orders can be delivered to different stakeholders to help them conduct corresponding actions. The target is to ensure work–life health, safety, and satisfaction.

Figure 1. Healthy Operator 4.0 (HO4.0) architecture.

3.2. Enabling Technologies

The successful implementation of HO4.0 architecture requires the support of a series of advanced enabling technologies. Some of the ones that were mentioned in the model and used in the prototype are introduced as follows.

- **Wearable technology:** Wearable technology or wearable computing is the study or practice of inventing, designing, building, or using miniature body-worn computational and sensory devices [35]. With the rapid development of sensing technologies, various types of sensors can be embedded in wearable devices. According to their features, they can be divided into four major groups [36]:

1. Environmental sensors, such as light sensors, temperature sensors, sound sensors, humidity sensors, air quality sensors (e.g., CO₂ sensors, particulate matter (PM) sensors, and volatile organic components (VOCs) sensors), and barometric sensors.
2. Biosensors, including body temperature sensors, heart-rate-monitoring sensors, electrocardiogram (ECG), electroencephalography (EEG), electromyography (EMG) sensors, blood pressure sensors, galvanic skin response (GSR) sensors, eye tracking, weight insole, and glucose level sensors.
3. Location tracking sensors, such as GPS, altimeters, magnetometers, compasses, and accelerometers.
4. Other sensors, including camera sensors, communication sensors, ultrasonic sensors, infrared receiver (IR), sensors.

The application of wearable technologies in HCPSSs provides rich information related to the operators and the surrounding environment.

- **Indoor Positioning System:** IPSs locate and track objects within a closed environment usually based on triangulation and multilateration methods using light, ultrasound, or radio signals to provide positional information [37]. The tracking of operators and production equipment can be significantly improved by the application of IPS [13]. Several wireless technologies might be used in IPS depending on the application situations by means of different positioning algorithms [38], such as Global Positioning System (GPS), Radio Frequency IDentification (RFID), Cellular networks, Ultra-wideband (UWB), wireless local area network (WLAN), Bluetooth etc. The adoption of IPS makes possible of a context-aware system for HO4.0 to help operators monitor the locations of machines, vehicles and workpieces to improve productivity and avoid potential collision risks.
- **Ambient environment monitoring:** The condition of the ambient environment plays a crucial role in a HO4.0 system. It may impact the operators' working performance or even harm their health. For example, indoor air pollution is one of the leading environmental risks, and indoor air quality (IAQ) is proven to have significant impacts on human comfort, health, and performance [29]. The advancement of low-cost IoT sensors, in recent years, has enabled the use of wireless communications and computing for interacting with the physical world. These sensors can measure indoor environmental parameters including IAQ, comfort, lighting, and acoustic conditions [39–41]. Many of these low-cost ambient sensors allow portable deployment and different sensors can be easily packed in one board to fit the requirements of different application scenarios. The data generated by such sensors can be collected by certain edge computing devices, such as single board computers (SBC). As an example, an ambient environment monitoring system based on a sensor island and a Raspberry Pi SBC was developed in a previous study [29]. The sensor island used in this system was composed of nine different ambient sensors to monitor different aspects of the ambient environment. More details of this system are explained in the prototyping section.
- **Knowledge engineering:** A semantic data model is a high-level representation of knowledge. It is designed to capture the meaning of data attributes and relationships, and obtain the meaning of each instance in an application context. The standard data model and knowledge engineering facilitates information exchange and data interpretation in a very heterogeneous data sources. Fusion and integration data based on rich semantic context gravitates toward applications, such as effective decision making processes, by fully making use of the various heterogeneous data. Ontologies, stemming from the Semantic Web, are used to define standard and common terms, vocabulary, and relationships for a particular area/domain. Aligning on common interpretations through ontologies, data/knowledge graphs can be generated to refer to the relationships among different data sources from different domains. Existing or adapted ontologies [42,43]

are recommended to be used for semantic modeling aiming to derive action rules toward the operator's health management.

- **Machine learning (ML):** The data gathered from the sensing layer can contain valuable information, which is the core of the HO4.0 system. Indeed, when scaled up appropriately, it can become a large volume of data; whereas, a global analysis can help in providing better interpretations for the impact of different behaviors. Therefore, advanced data modeling tools, such as machine learning, can be useful to create or emphasize such new knowledge. The wide applications of the IoT have promoted the concept of big data, which on the other hand, has fostered the explosive involvement of the advanced machine learning techniques represented by deep learning. Despite the great achievement of machine learning in many fields, its power has not been fully and systematically explored in the industry sector, particularly in the human-centric systems. Here the benefits from the adopted concept for HO4.0, such as the system proposed in this research (which enables the digital twin approach) facilitates modeling capabilities to detect patterns and to add value in the real world for human-centric applications. For example, deep learning can be used to monitor an operator's activities based on the motion data collected from wearable devices [9,44,45], but when combined with health related information and environmental conditions, this can provide a health index for such activities. In addition to the self learning process based on the real time data, the inter-comparison between workers would help to better understand the physiological and mental demands, but it can also help to find different patterns and to learn about the sustainability of them for the long term [46].

4. The Application Case and Experiments

To validate the proposed architecture, an application case was carried out, enabling the construction of a prototype helping with different practical implementations.

4.1. Case Selection

The case is connected with a logistic intra-facility, where the interest was in obtaining health related knowledge of crane operators in their working environment, who were in charge of loading trucks that moved manufactured steel rebar to build singular concrete structures.

Figure 2a shows how an operator moved around, looking for the pieces to be loaded, then attaching them and loading them into the trucks, where they went up and down a specific platform as presented in Figure 2b, which was designated to facilitate access to the truck itself and to maintain awareness on how the material was placed inside the truck. The preparation of the materials on trucks must be carefully planned as the unload process typically needs to visit different construction sectors to deliver specific sets of rebar. Therefore, a specific disposal sequence of items as per the layers in the truck must also be carefully planned and is prone to errors and stressful situations, where there are doubts about what was actually loaded into the truck. This is even more sensitive for sets involving small numbers of light rebar, where weight variations are not informative enough about errors in the loading process.

Outside of their specific job, the crane operator was working in a rather complex environment, with potential environmental quality concerns. This environment was a plant that manufactures steel rebar, where the production of specific items by cutting, blending, and welding may affect the air quality. Due to the working environment, as shown in Figure 2a, acoustic and illumination aspects were be other health threats.

Due to the different working conditions for crane operators and their complex mobility schema, their mental working environment, and physical workload, and because of the continuous over-pressure to quickly deliver products to trucks, we determined that these workers were suitable for inclusion in the validation for the HO 4.0 architecture.

(a) Crane operator at work.

(b) Operator unloading platform.

Figure 2. Crane operator working conditions.

4.2. Sensing Layer

In this application, based on the specific characteristics of the workers, the sensing layer was designed to include different dimensions, including the:

- **Environmental conditions**, including temperature, humidity, and noise, as they can impact the operators' health, as well as their performance.
- **Heart rate**, as related to their physical or mental effort.
- **Blood Pressure**, also as related to their physical or mental effort.
- **Arm angle**, referring to the horizontal line, as a reference for ergonomic working activities.
- **Position**, as a reference for understanding the movements and knowledge requirements in order to complete the work.
- **Steps**, as a reference for the physical effort.
- **Crane position**, in relation to the crane operator movements, allowing us to identify the effective working time.

Therefore, different sensing devices need to be attached to the crane operator to collect the information in relevant frequency patterns, although they must be as lightweight as possible, in order to be feasible for daily operations.

For the prototype, we decided to include three devices attached to the crane operators and one environmental station deployed on the shopfloor. Two of them were based on low cost devices, a MetaSensor (see Figure 3a) from mbientlab (<https://mbientlab.com/store/metamotionr/>), enabled us to monitor angles and acceleration and was used to track the arm angle through time. Another sensor we included was a smartband from HBand (see Figure 3b), able to track the heart rate, blood pressure, and steps.

For positioning the crane operator and the crane hook, we decided to use Ultra-wideband (UWB) technology with six stations able to track those positions [47]. The technology named *uRTLSTM*, is a high precision real-time location solution (RTLS) based on UWB technology. The solution used several channels available at 3–7 GHz using the Decawave UWB chipset in compliance with IEEE 802.15.4.

The indoor environmental quality substantially affects a operator's health, comfort, working performance, and well-being. In this study, the working environment for operators was monitored with an IoT based environmental quality monitoring system developed by [29]. The measured indoor environment conditions included: the chemical environmental parameters: particulate matter (PM), formaldehyde (HCHO), Total Volatile Organic Compounds (TVOC), benzene (C₆H₆), carbon dioxide (CO₂), carbon monoxide (CO), ozone (O₃), nitrogen dioxide (NO₂), and the physical parameters: temperature, humidity, illumination, and noise.

Although these specific devices were selected for this implementation, without a lack of generalization, certain other devices, such as smart insoles, smart helmets, and skin conductivity sensors, could be also adequate, depending on the application case under consideration.

(a) Angle measurement device.

(b) Smartband device.

Figure 3. Wearable devices used in this application.

4.3. Integration and Communication Layer

In this application, the usage of several technologies were required in order to properly handle the data streams. There were data collected from the factory that benefited from using the existing extranet to upload the datasets to the data repository. This is the case for the environmental monitoring station, which was based on a raspberry pi delivering the data throughout its internal network card. This was also the case for the positioning system, which was acquired from the tracktio company (<https://tracktio.com/>). This system had its own data repository, therefore an additional middleware was required in order to enable data integration.

There were other data streams that were collected only by bluetooth, as the devices only had this interface. To handle these streams, an Android based app was developed [48], making it convenient to integrate the features and enabling a flexible data collection architecture [49]. Figure 4 presents the main details for this application, which collected data according to a specific time sequence. For this application, a minute-based frequency was adopted for health and environmental parameters, and a second-based frequency for the position, which provided convenient granularity for the learning application. Indeed, it was relevant that the developed app showed real time values and informed the operator regarding significant values. Although it was possible to order smartband vibrations when specific conditions were met, this was not activated as it could distract the operators from their main task and possibly become an unacceptable safety risk.

(a) Main screen.

(b) Device list.

Figure 4. The app developed for data collection.

The overall context for the data handling of the applied wearable in this study is presented in Figure 5.

Figure 5. Overview of the applied wearable data flow.

4.4. Modeling Layer

With all the data streams ingested, this layer facilitated the fusion and time alignment as per the device media access control (MAC) address and time. Therefore, the record-sets were derived providing context for the digital twin, according to the chosen application case. The data fusion could integrate data streams that in some cases belonged to the same individual or in other cases belonged to properties of the area, affecting several operators on the shop-floor. The collected data streams allowed us to consider semantic annotation to enable a machine based data handling strategy.

The existing Vcard ontology for people [50] was used to model each operator. The indoor environmental quality ontology: AIR_POLLUTION_Onto, proposed by [51], was used to model the operator working environmental exposure, risk, and control applications. The physiology factor of the operator was semantically represented with HUMAN STRESS ONTOLOGY [52], considering the data sources steaming from the physiological parameters, such as the heart rate and blood pressure. The indoor navigation ontology introduced in the ILONA system [53] was applied for the indoor model positioning of the operators. Therefore, the considered set of references, based on the integration of such ontologies was as shown in Table 1.

Table 1. Data streams with semantic annotation.

Entity	Message
crane operator	{ "Ontology": " https://bit.ly/2OxeEkO ", "object": "vcard", "fn": "José Luis Fernández", "nickname": "Jose", "hasEmail": "mailto:perpalper@gmail.com", "Gender": "Male", "bday": "1988-06-23", "adr": "José Gutiérrez Abascal 2, 28006, Madrid" }
environmental situation	{ "Ontology": "AIR_POLLUTION_Onto", "object": "airPollutants", "deviceId": "Airmonitor1", "PM": "25 ug/m ³ ", "CO2": "0.04%", "VOC": "0.4 mg/m ³ ", "NOX": "0", "Timestamp": "09/08/2019 09:10:00" }
stress factor	{ "Ontology": "HUMAN STRESS ONTOLOGY", "object": "Measurements", "deviceId": "Hband1", "Timestamp": "09/08/2019 09:10:00", "stressPhysiology": [{ "heartRate": "80", "bloodPressureHigh": "120", "bloodPressureLow": "70" }] }
position	{ "Ontology": "ILONA", "object": "Position", "deviceId": "Cranel1", "Coordinate": [{ "x": "40.342712", "y": "-3.123472", "z": "605.85" }], "Timestamp": "09/12/2019 10:10:00" }

Based on the specific interest, the annotated data streams were combined to create the contextual datasets, where their structure was not fixed at all. Instead, it was problem related. Therefore, when the interest was to analyze the similarity of the operating conditions and operator behavior through time, the application for building projectors with general non-linear dimension reduction was needed.

Then, this layer will include different algorithms and tools, such as Principal component analysis (PCA), T-distributed Stochastic Neighbor Embedding (t-SNE), or Uniform Manifold Approximation and Projection (UMAP) dimension reduction techniques, which optimize a low-dimensional graph from twelve dimensions into two, in order to keep close records that are close in a higher dimensional space, and that preserved more of the global structure of the whole dataset.

When the interest is to look at specific behaviors that can be described by particular variations in variables, such variables were created and the ranges calculated. Based on those derived new variables, different Machine Learning techniques, such as association rules, can bring interesting rules explaining the causes for such behaviors.

When the case is to understand the behavior of operators through time, different regression algorithms were applied in order to forecast future trends, either by using the time series approach or the multidimensional one. Indeed, if the interest was to identify similar behavior between operators, different clustering algorithms were helpful, based on the medium term records of their digital twins.

The previous examples illustrated the strong possibilities that this layer can provide when the interest is to derive new knowledge for operators at different level of aggregation and for different time scales.

5. Results

The working environment of operators for this particular case study is dissatisfying as shown in Figure 6, due to different aspects where the actual exposition levels exceeded the threshold limit value (TLV).

The TLV of a environmental substance (chemical or physical) is defined to be a level to which a worker can be exposed day after day for a working lifetime without adverse effects. The recommended comfort TLVs for temperature and humidity are 16–28 °C and 30–80%, respectively, according to World Health Organization (WHO) [54]. For work that requires the perception of details, such as offices, sheet metal work, or bookbinding, the minimal luminance TLV is 100 Lux, which is defined by the European Union (EU) standard. The noise levels defined by WHO are 85–90 dB(A) daily average [55], the smallest value (85 dB (A)) was taken as the TLV to ensure acoustic health to the maximal extent. The TLV for PM_{2.5} and HCHO are a 25 µg/m³ average per day and a 0.016 ppm (0.02 mg/m³) average per 8 hours as defined by WHO [56] and National Institute for Occupational Safety and Health (NIOSH) [57], respectively. The TLV of CO₂ (0.1%) and TVOC (0.60 mg/m³ average per 8 h) are based on a daily average, which obtained references from Circulate App: EnvCon [58]. The Circulate company set TLVs by taking references from China's air-quality standards.

In this application case, the noise level was rather high consistently over time, as demonstrated in Figure 6a. In the vast majority of the working time, the noise level was higher than 85 dB(A) as defined by WHO [55], as the primary working activity is steel operations, which produces a huge amount of noise. Although operators are wear industrial headsets to protect their ears, constantly working in very noisy environments presents potential damage and hearing related problems, e.g., tinnitus. It also affects concentration at work, which was the case for crane operators. The environment was also rather dry and dim as depicted in Figure 6a (Humidity and Lighting). The chemical environmental conditions were inadequate as shown in Figure 6b. The PM_{2.5} and HCHO exceeded the contaminant level from time to time, and the TVOC is approached limits periodically. Therefore, these systems also help to identify dimensions where measures can be taken to improve the working environment, which would have a significant impact on the operators.

(a) Physical environmental conditions.

(b) Chemical environmental conditions.

Figure 6. Operator exposure to the environmental condition on site. The red line refers to the threshold limit value (TLV).

It is relevant to identify different clusters of behavior from the collected data as shown in Figure 7. In fact, this Figure presents a non-linear two dimensional projection from the originally 12-dimensional space of fused data from two different crane operators was embedded. It was decided to use the

UMAP projection technique to provide a clear vision regarding the sample distribution where the inter-distance between samples was maintained and good preservation of the data's global structure was granted [59].

Figure 7. Projected patterns from the digital twin.

The projection result, shown in Figure 7, depicts the whole dataset's distributions and clusters phenomenon. Then, we decided to segment the whole dataset into five groups of the same length, to explore time dependency of clusters (see Figure 7a). The five groups were segmented via homogeneous distribution through time, and marked with different colors (labels 1–5). As far as the same color appears in different clusters, that means that they are not related to time mainly, but related to operator's behaviour. In Figure 7b, the behaviors from the two monitored crane operators are presented. The projection of the whole dataset was also segmented into two groups: operator 1 and operator 2 marked with two different colors. It becomes clear that there is not operator specificity as per cluster, therefore the existence of many of them is due to intrinsic reasons. On the other side, different operators behaved in different ways, as different colors appeared in different clusters, although similarities due to the common activities were also there. Indeed, there was room to identify each of the clusters and to enrich the collection with larger datasets and, thus, derive knowledge from such clusters. The evolution of operator health such as stress level can be identified via clusters and amount of time on each, which makes it possible to adapt individual healthy requirement from management point of view. The objective measures can be taken, e.g., to allocate specific operator to less demanding trucks in order to provide better wellness or convenient recovery process.

The interest in the prototype was also to evaluate its functionality to derive specific knowledge when considering digital twins. In particular, the interest was to analyze the stress levels and the impacts. We decided to focus on situations where large differences between high and low blood pressure happen. This is because no matter whether an operator can suffer hypertension, an unwanted situation related to high levels of stress will have locally larger differences between the high and the low pressures, and the risk for cardiovascular disease, diabetes, chronic kidney disease, high cholesterol levels, and aneurysms will be larger [60]. A segmentation in different discrete categories between Low to Intense for the original variables as well for the derived ones (mainly differences and ranges), was carried out, including Δ BloodPressure, where the threshold for a large difference was adopted to be 45 mmHg. After the data preparation an a priori algorithm for the association rules discovery was applied, which provided some rules like those presented in Table 2.

Table 2. The derived rules from the application of machine learning techniques looking at high differences in blood pressure and explaining the intense values.

Rule Antecedent	Confidence	Support	Lift
'Arm angle:UP FRONT' AND 'High Blood Pressure:Intense' AND 'Heart Rate:Moderate'	1.0	5.9%	3.86
'Sound Level:Noisy' AND steps:Low AND 'Arm angle:UP FRONT AND 'High Blood Pressure:Intense'	1.0	6.27%	3.86
'steps:Low' AND 'HCHO:Relative Moderate' AND 'High Blood Pressure:Intense'	0.84	10.0%	3.85

Only relevant rules were highlighted as the algorithm discovered many others related to normal conditions, including normal environmental conditions and normal body indicators, which imply a normal range for Δ BloodPressure, etc. All of them had strong support and moderately high confidence however the lift values were close to 1. Indeed, the biggest interest was to check the rules with high confidence values, high lift figures, and moderate or low support. Those rules became excellent candidates to create knowledge (maximum confidence); however, due to their relatively low support, they remain mostly unknown.

During the validation of this prototype there was an unexpected high interest from the operators to obtain the knowledge about their parameters (to learn from their digital twins), and as far as it was possible and it was understood it could also provide benefits to the operators as they can become more aware of the reasons for their parameters; therefore, a significant effort was conducted to address the increase of awareness.

The operator were able to have the real-time information of their health parameters, as shown in Figure 8.

(a) Biometric parameters. (b) Ergonomic parameters.

Figure 8. Informative phone screens.

When lean management techniques are being used at the organization level, it can also help with the management of processes to communicate to the operators their figures on a daily basis, as described above for the $(CPD)_nA$ method [17]. Therefore, an application for the trelloTM system was created at the Cognition layer, where the detailed messages can be seen, as in Figure 9, bringing a way to deeply dig into the specific graphics Figure 10, making it possible to promote the operator self-learning process.

(a) Graphical view sent to user account.

(b) Summary screen sent as message to trello.

Figure 9. Informative user phone screens where time granularity can be configured. **Figure 10.** The graphical detailed view of the collected data at an individual level.

6. Discussion

The implementation of a prototype to validate the HO4.0 framework was performed successfully, where the layer structure helped in providing guidance to the required discussion in order to build the system. Data interconnection and interoperability further facilitated the making of health action decisions, aiming to improve the operator working conditions, and the strategies and actions were built on semantic-based data driver strategies.

The data from the sensing layer was collected via the data networking layer and was multimodal and heterogeneous when considering the different sensing instruments and data collection solutions. To make decisions, especially for the long-term policy in terms of health management, the data coming from different domains have to be interoperable and linked. We also needed to emphasize the meaning of each type of data streams, as well as the frequency of sampling, aligned with the relevant business model.

To further support decision making for health operator applications, different ontologies were applied in this study to make the data interoperable and semantically accessible while crossing domains. Although the already presented combination was retained for this prototype, there was no limitation other than to properly describe the entities and their inter-relationships to maintain a standard domain vocabulary and common understanding. Therefore, they can also be created on purpose. The metadata languages aligning with the ontology vocabulary convey that kind of shared information in a landscape of information management.

With the data inter-operation features and semantic annotation, a holistic view can be built up to enable the machine learning application to create new knowledge, as already presented. This can be operationalized and communicated looking to improve the operator health and safety.

Additionally, it is possible to combine the organizational analysis based on patterns with the additional value created from the operator level, by allowing them to be aware of their own values. This is another useful dimension of the HO4.0 concept, as awareness is a key element for empowerment [61].

7. Conclusions

Empowering people has significant implications for factories under the Industry 4.0 paradigm. Health, as one of main concerns affects operator's job satisfaction and well-being. However, it was rarely investigated in the context of HCPSs. Aiming to reduce this gap, holistic view of operator health dimension was proposed, where the main contributions of this paper are summarized as follows:

- We explored the HO4.0 concept.
- Based on the classical five-level CPS structure, we proposed a four-layer framework for the implementation of the HO4.0 concept.
- We investigated of the enabling technologies for HO4.0.
- We developed a prototype system to showcase the application process of the proposed framework.
- We applied advanced machine learning and preparation with semantic engineering technologies to manage and analyze heterogeneous data from different sources, bringing this support to the digital twin perspective.

The research scope is not limited to the outcome of conducted experiments, but to provide an integrated framework providing vertical and horizontal integration capabilities. The proposed architecture supports easy integration of different data sources as from Industrial Internet of Things, wearables such as smart helmet, smart insoles, smart glasses, to name a few. From a practical perspective, it also successfully addresses the need for data integration from different devices, as manufacturers aim to provide their own data lakes, which are not flexible enough to allow user defined sampling frequency or convenient restful based data access. In addition, enhancing the capability for monitoring different devices with the same middleware (phone, tablet, etc.) including reconnecting capabilities, makes the data collection process stronger, which is really valuable in industrial environments.

The gathered data sources with the defined data fusion techniques, including semantic interoperability, produce multidimensional datasets, which are suitable to be analyzed in real time with advanced modeling technologies (big analytics). By learning from operators' daily activities and parameters, inference about their evolution in long term perspective can be derived through their digital-twins. Such approach can have managerial value as adopted strategies can relay on transparent

measurement system and they can better preserve knowledge, experience, and well-being for longer, as well as company enlarges its social sustainability dimension.

In addition, it was also possible to realize the high degree of interest that such applications open to workers themselves, and how it becomes a way of discussion between them, as they try to identify patterns and explanations. We have found it very interesting for using it in a smarter way of management, taking advantage of such interest in order to increase the empowerment and engagement of workers into their processes.

Further than that, based on the adopted methodology, we concluded that the designed framework for HO4.0, as depicted above, based on the HO4.0 concept, was adequate and easy to follow in the integration of different sources of data at any time.

As a limitation, the prototype was applied in only one company with a wide range of operator profiles, but where the size and culture can also have an impact, and these dimensions were not analyzed. Another limitation is based on appropriate sensor availability, as there are safety regulations that can limit the application of specific sensors, and there are a lack of sensors for certain tasks, such as human concentration. However, the range of applications was very large as the integration between the operator's health and smart monitoring is just starting.

As a further research ambition, the prototype will be extended by integrating weight measurement through smart insoles, giving the information of the loads being carried by operators along the working time. This will be implemented by extending the 'Other devices' option from the Android app, following the created implementation rules. Indeed, further research is needed to better categorize the identified clusters, its distribution and stability against sample size. Another research line is connected to the information that could be derived from the digital-twins when more data populate them. It seems to be very promising. Finally, but yet importantly, to connect the HO4.0 with production processes can bring a truly integrated management perspective which demands further investigations.

Author Contributions: Data curation, S.S.; Investigation, S.S., X.Z. and J.O.-M.; Methodology, X.Z. and S.S.; Project administration, J.O.-M.; Resources, J.O.-M.; Software, J.G.P.; Validation, B.G.; Visualization, B.G.; Writing—original draft, X.Z., J.O.-M. and S.S.; Writing—review and editing, B.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was partially funded by the European Commission through the RFCS program, grant number with ID 793505 (WISEST project), and by the Spanish Agencia Estatal de Investigación, through the research project with code RTI2018-094614-B-I00.

Acknowledgments: The authors thank the EU RFCS research program for the support of this research through the grant with ID 793505 (WISEST project) as well as the industrial partners. Indeed, Joaquín Ordieres-Meré wants to acknowledge the Spanish Agencia Estatal de Investigación, through the research project with code RTI2018-094614-B-I00 into the "Programa Estatal de I+D+i Orientada a los Retos de la Sociedad".

Conflicts of Interest: The authors declare no conflict of interest. The founders did not have any influence either in selecting the topic or in the experiment design.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence.
CIM	Computer Integrated Manufacturing.
CPS	Cyber-Physical Systems.
ECG	Electrocardiogram.
EEG	Electroencephalography.
EMG	Electromyography.
GPS	Positioning System.
HCPSs	Human-Cyber-Physical Systems.
HO4.0	Healthy Operator 4.0.
IAQ	Indoor Air Quality.
IIoT	Industrial Internet of Things.
IoT	Internet of Things.

IPA	Intelligent Personal Assistant.
IPS	Indoor Positioning System.
IR	Infrared Receiver.
IT	Information Technology.
MAC	Media Access Control
ML	Machine Learning
OSH	Occupational Safety and Health
RFID	Radio Frequency IDentification.
SBC	Single-Board Computer.
TLV	Threshold Limit Value
UWB	Ultra-wideband.
VR	Virtual Reality.
WLAN	wireless local area network.

References

- Schmidt, R.; Möhring, M.; Härting, R.C.; Reichstein, C.; Neumaier, P.; Jozinović, P. Industry 4.0-potentials for creating smart products: empirical research results. In *International Conference on Business Information Systems*; Springer: London, UK, 2015; pp. 16–27.
- Pereira, A.; Romero, F. A review of the meanings and the implications of the Industry 4.0 concept. *Procedia Manuf.* **2017**, *13*, 1206–1214. [[CrossRef](#)]
- Zheng, X.; Sun, S.; Mukkamala, R.R.; Vatrappu, R.; Ordieres-Meré, J. Accelerating health data sharing: A solution based on the internet of things and distributed ledger technologies. *J. Med. Internet. Res.* **2019**, *21*, e13583. doi:10.2196/13583. [[CrossRef](#)] [[PubMed](#)]
- Lee, J.; Bagheri, B.; Kao, H.A. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manuf. Lett.* **2015**, *3*, 18–23. [[CrossRef](#)]
- Chen, S.L.; Chen, Y.Y.; Hsu, C. A new approach to integrate internet-of-things and software-as-a-service model for logistic systems: A case study. *Sensors* **2014**, *14*, 6144–6164. [[CrossRef](#)] [[PubMed](#)]
- Glaessgen, E.; Stargel, D. The Digital Twin Paradigm for Future NASA and US Air Force Vehicles. In Proceedings of the 53rd AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference 20th AIAA/ASME/AHS Adaptive Structures Conference 14th AIAA, Honolulu, HI, USA, 23–26 April 2012; p. 1818.
- Boschert, S.; Rosen, R. Digital twin—The simulation aspect. In *Mechatronic Futures*; Springer: London, UK, 2016; pp. 59–74.
- Hao, Y.; Helo, P. The role of wearable devices in meeting the needs of cloud manufacturing: A case study. *Robot. Comput.-Integr. Manuf.* **2017**, *45*, 168–179. [[CrossRef](#)]
- Zheng, X.; Wang, M.; Ordieres-Meré, J. Comparison of data preprocessing approaches for applying deep learning to human activity recognition in the context of industry 4.0. *Sensors* **2018**, *18*, 2146. [[CrossRef](#)] [[PubMed](#)]
- Putnik, G. Advanced manufacturing systems and enterprises: Cloud and ubiquitous manufacturing and an architecture. *J. Appl. Eng. Sci.* **2012**, *10*, 127–134.
- Gorecky, D.; Schmitt, M.; Loskyll, M.; Zühlke, D. Human-Machine-Interaction in the Industry 4.0 Era. In Proceedings of the 2014 12th IEEE International Conference on Industrial Informatics (INDIN), Porto Alegre, RS, Brazil, 27–30 July 2014; pp. 289–294.
- Romero, D.; Stahre, J.; Wuest, T.; Noran, O.; Bernus, P.; Fast-Berglund, Å.; Gorecky, D. Towards an Operator 4.0 Typology: A Human-Centric Perspective on the Fourth Industrial Revolution Technologies. In Proceedings of the International Conference on Computers and Industrial Engineering (CIE46) Proceedings, Tianjin, China, 29–31 October 2016.
- Ruppert, T.; Jaskó, S.; Holczinger, T.; Abonyi, J. Enabling technologies for operator 4.0: A survey. *Appl. Sci.* **2018**, *8*, 1650. [[CrossRef](#)]
- Lorenz, M.; Rüßmann, M.; Strack, R.; Lueth, K.L.; Bolle, M. *Man and Machine in Industry 4.0: How Will Technology Transform the Industrial Workforce through 2025*; Boston Consulting Group: Boston, MA, USA, 2015; Volume 2.

15. Romero, D.; Bernus, P.; Noran, O.; Stahre, J.; Fast-Berglund, Å. The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *IFIP International Conference on Advances in Production Management Systems*; Springer: London, UK, 2016; pp. 677–686.
16. Tzafestas, S. Concerning human-automation symbiosis in the society and the nature. *Int. J. Fact. Autom. Robot. Soft Comput.* **2006**, *1*, 16–24.
17. Villalba-Díez, J.; Ordieres-Meré, J. Improving manufacturing operational performance by standardizing process management. *Trans. Eng. Manag.* **2015**, *62*, 351–360. [[CrossRef](#)]
18. Horváth, L.; Rudas, I.J. Role of Information Content in Multipurpose Virtual Engineering Space. In Proceedings of the 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMII), Her'ľany, Slovakia, 26–28 January 2017; pp. 000099–000104.
19. Sylla, N.; Bonnet, V.; Colledani, F.; Fraisse, P. Ergonomic contribution of ABLE exoskeleton in automotive industry. *Int. J. Ind. Ergon.* **2014**, *44*, 475–481. [[CrossRef](#)]
20. Myers, K.; Berry, P.; Blythe, J.; Conley, K.; Gervasio, M.; McGuinness, D.L.; Morley, D.; Pfeffer, A.; Pollack, M.; Tambe, M. An intelligent personal assistant for task and time management. *AI Mag.* **2007**, *28*, 47.
21. Romero, D.; Mattsson, S.; Fast-Berglund, Å.; Wuest, T.; Gorecky, D.; Stahre, J. Digitalizing occupational health, safety and productivity for the operator 4.0. In *IFIP International Conference on Advances in Production Management Systems*; Springer: London, UK, 2018; pp. 473–481.
22. Buffet, M.; Gervais, R.; Liddle, M.; Eeckelaert, L. Well-being at work: Creating a positive work environment. In *Literature Review*; European Agency for Safety and Health at Work: Bilbao, Spain, 2013.
23. Salanova, M.; Del Líbano, M.; Llorens, S.; Schaufeli, W.B. Engaged, workaholic, burned-out or just 9-to-5? Toward a typology of employee well-being. *Stress Health* **2014**, *30*, 71–81. [[CrossRef](#)] [[PubMed](#)]
24. Vasic, M.; Billard, A. Safety Issues in Human-Robot Interactions. In Proceedings of the 2013 IEEE International Conference on Robotics and Automation, Karlsruhe, Germany, 6–10 May 2013; pp. 197–204.
25. Matthias, B.; Oberer-Treitz, S.; Staab, H.; Schuller, E.; Peldschus, S. Injury Risk Quantification for Industrial Robots in Collaborative Operation with Humans. In Proceedings of the ISR 2010 (41st International Symposium on Robotics) and ROBOTIK 2010 (6th German Conference on Robotics), Munich, Germany, 7–9 June 2010; pp. 1–6.
26. Page, K.M.; Vella-Brodrick, D.A. The ‘what’, ‘why’ and ‘how’ of employee well-being: A new model. *Soc. Indic. Res.* **2009**, *90*, 441–458. [[CrossRef](#)]
27. Podgorski, D.; Majchrzycka, K.; Dąbrowska, A.; Gralewicz, G.; Okrasa, M. Towards a conceptual framework of OSH risk management in smart working environments based on smart PPE, ambient intelligence and the Internet of Things technologies. *Int. J. Occup. Saf. Ergon.* **2017**, *23*, 1–20. [[CrossRef](#)]
28. Bernal, G.; Colombo, S.; Al Ai Baky, M.; Casalegno, F. Safety++: Designing IoT and Wearable Systems for Industrial Safety through a User Centered Design Approach. In Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments, Rhodes, Greece, 21–23 June 2017; pp. 163–170.
29. Sun, S.; Zheng, X.; Villalba-Díez, J.; Ordieres-Meré, J. Indoor air-quality data-monitoring system: Long-term monitoring benefits. *Sensors* **2019**, *19*, 4157. [[CrossRef](#)]
30. Kassner, L.; Hirmer, P.; Wieland, M.; Steimle, F.; Königsberger, J.; Mitschang, B. The Social Factory: Connecting People, Machines and Data in Manufacturing for Context-Aware Exception Escalation. In Proceedings of the 50th Hawaii International Conference on System Sciences, Village, HI, USA, 4–7 January 2017.
31. Romero, D.; Wuest, T.; Stahre, J.; Gorecky, D. Social factory architecture: social networking services and production scenarios through the social internet of things, services and people for the social operator 4.0. In *IFIP International Conference on Advances in Production Management Systems*; Springer: London, UK, 2017; pp. 265–273.
32. Pavón, I.; Sigcha, L.; Arezes, P.; Costa, N.; de Arcas, G.; Lopez-Navarro, J. Wearable technology for occupational risk assessment: Potential avenues for applications. *Occup. Saf. Hyg.* **2018**, *VI*, 447–452.
33. Perera, C.; Liu, C.H.; Jayawardena, S. The emerging internet of things marketplace from an industrial perspective: A survey. *IEEE Trans. Emerg. Top. Comput.* **2015**, *3*, 585–598. [[CrossRef](#)]
34. Bell, E.; Bryman, A.; Harley, B. *Business Research Methods*; Oxford University Press: Oxford, UK, 2018.
35. Mann, S. Wearable computing. In *The Encyclopedia of Human-Computer Interaction*, 2nd ed.; International Design Foundation: New York, NY, USA, 2013.

36. Mardonova, M.; Choi, Y. Review of wearable device technology and its applications to the mining industry. *Energies* **2018**, *11*, 547. [CrossRef]
37. Koyuncu, H.; Yang, S.H. A survey of indoor positioning and object locating systems. *IJCSNS Int. J. Comput. Sci. Netw. Secur.* **2010**, *10*, 121–128.
38. Liu, H.; Darabi, H.; Banerjee, P.; Liu, J. Survey of wireless indoor positioning techniques and systems. *IEEE Trans. Syst. Man Cybern. Part C (Appl. Rev.)* **2007**, *37*, 1067–1080. [CrossRef]
39. Zanni, S.; Lalli, F.; Foschi, E.; Bonoli, A.; Mantecchini, L. Indoor air quality real-time monitoring in airport terminal areas: An opportunity for sustainable management of micro-climatic parameters. *Sensors* **2018**, *18*, 3798. [CrossRef] [PubMed]
40. Zakaria, N.A.; Abidin, Z.Z.; Harum, N.; Hau, L.C.; Ali, N.S.; Jafar, F.A. Wireless Internet of Things-based air quality device for smart pollution monitoring. *Int. J. Adv. Comput. Sci. Appl.* **2018**, *9*, 65–69. [CrossRef]
41. Benammar, M.; Abdaoui, A.; Ahmad, S.; Touati, F.; Kadri, A. A modular IoT platform for real-time indoor air quality monitoring. *Sensors* **2018**, *18*, 581. [CrossRef]
42. Industrial Ontologies Foundry (IOF). Available online: <https://sites.google.com/view/industrialontologies/home> (accessed on 20 March 2020).
43. Sampath Kumar, V.R.; Khamis, A.; Fiorini, S.; Carbonera, J.; Olivares-Alarcos, A.; Habib, M.; Gonçalves, P.; Li, H.; Olszewska, J. Ontologies for Industry 4.0. *Knowl. Eng. Rev.* **2019**, *34*. doi:10.1017/S0269888919000109. [CrossRef]
44. Ward, J.A.; Lukowicz, P.; Troster, G.; Starner, T.E. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2006**, *28*, 1553–1567. [CrossRef]
45. Shoaib, M.; Bosch, S.; Incel, O.; Scholten, H.; Havinga, P. Complex human activity recognition using smartphone and wrist-worn motion sensors. *Sensors* **2016**, *16*, 426. [CrossRef]
46. Ordieres-Meré, J.; Prieto Remón, T.; Rubio, J. Digitalization: An opportunity for contributing to sustainability from knowledge creation. *Sustainability* **2020**, *12*, 1460. [CrossRef]
47. Macoir, N.; Bauwens, J.; Jooris, B.; Van Herbruggen, B.; Rossey, J.; Hoebeke, J.; De Poorter, E. Uwb localization with battery-powered wireless backbone for drone-based inventory management. *Sensors* **2019**, *19*, 467. [CrossRef]
48. Mobile Operating System Market Share Worldwide. Available online: <https://gs.statcounter.com/os-market-share/mobile/worldwide> (accessed on 20 March 2020).
49. García Paredes, J.; Ordieres-Meré, J. Healthy Operator 4.0. 2020. Available online: <https://github.com/jbmeré/HealthOperator4.0/releases> (accessed on 20 March 2020).
50. vCard Ontology - for describing People and Organizations. Available online: <https://www.w3.org/TR/vcard-rdf/> (accessed on 20 March 2020).
51. Oprea, M.M. AIR_POLLUTION_Onto: An ontology for air pollution analysis and control. In *Artificial Intelligence Applications and Innovations III*; Maglogiannis, I., Karpouzis, K., Eds.; Springer: London, UK, 2009; pp. 135–143.
52. Khoozani, E.N.; Hadzic, M. Designing the human stress ontology: A formal framework to capture and represent knowledge about human stress. *Aust. Psychol.* **2010**, *45*, 258–273. doi:10.1080/00050061003664811. [CrossRef]
53. Kun, D.P.; Varga, E.B.; Toth, Z. Ontology Based Navigation Model of the ILONA System. In Proceedings of the 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), Herl'any, Slovakia, 26–28 January 2017; pp. 479–484. doi:10.1109/SAMI.2017.7880357. [CrossRef]
54. Head, K.; Clarke, M.; Bailey, M.; Livinski, A.; Ludolph, R.; Singh, A. Report of the systematic review on the effect of indoor heat on health. In *WHO Housing and Health Guidelines*; World Health Organization: Geneva, Switzerland, 2018.
55. Noise Guidelines. Available online: http://www.euro.who.int/__data/assets/pdf_file/0008/383921/noise-guidelines-eng.pdf?ua=1 (accessed on 20 March 2020).
56. WHO Guidelines for Indoor Air Quality. Available online: http://www.euro.who.int/__data/assets/pdf_file/0009/128169/e94535.pdf (accessed on 20 March 2020).
57. NIOSH Pocket Guide to Chemical Hazards: Formaldehyde. Available online: <https://www.cdc.gov/niosh/npg/npgd0293.html> (accessed on 20 March 2020).
58. EnvCon App. Available online: <https://itunes.apple.com/tr/app/envcon/id1045648840?mt=8> (accessed on 20 March 2020).

59. McInnes, L.; Healy, J.; Saul, N.; Grossberger, L. UMAP: Uniform Manifold Approximation and Projection. *J. Open Source Softw.* **2018**, *3*, 861. [[CrossRef](#)]
60. Soni, R.K.; Porter, A.C.; Lash, J.P.; Unruh, M.L. Health-related quality of life in hypertension, chronic kidney disease, and coexistent chronic health conditions. *Adv. Chronic Kidney Dis.* **2010**, *17*, e17–e26. [[CrossRef](#)] [[PubMed](#)]
61. Konijn, A.M.; Lay, A.M.; Boot, C.R.; Smith, P.M. The effect of active and passive occupational health and safety (OHS) training on OHS awareness and empowerment to participate in injury prevention among workers in Ontario and British Columbia (Canada). *Saf. Sci.* **2018**, *108*, 286–291. [[CrossRef](#)]

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

3.7 Publication 7

- **Title:** *Data Handling in Industry 4.0: Interoperability Based on Distributed Ledger Technology*
- **Authors:** Shengjing Sun, Xiaochen Zheng, Javier Villalba-Díez and Joaquín Ordieres-Meré
- **Accepted date:** 2020/05/24
- **Journal:** *Sensors*
3.031 (2018) ; 5-Year Impact Factor: 3.302 (2018)
Quartile: Q1 in ‘Instruments & Instrumentation’
ISSN: 1424-8220
Publisher: Multidisciplinary Digital Publishing Institute, BASEL, SWITZERLAND
- **Volume:** 20
- **Issue:** 11
- **Pages:** 3046
- **DOI:** <https://doi.org/10.3390/s20113046>

Article

Data Handling in Industry 4.0: Interoperability Based on Distributed Ledger Technology

Shengjing Sun ^{1,*}, Xiaochen Zheng ^{1,2}, Javier Villalba-Díez ^{1,3,4}, and Joaquín Ordieres-Meré ^{1*}

¹ Escuela Técnica Superior de Ingenieros Industriales (ETSII), Universidad Politécnica de Madrid, José Gutiérrez Abascal 2, 28006 Madrid, Spain

² ICT for Sustainable Manufacturing, SCI-STI-DK, École polytechnique fédérale de Lausanne (EPFL), 1015 Lausanne, Switzerland

³ Fakultät für Management und Vertrieb, Campus Schwäbisch-Hall, Hochschule Heilbronn, 74081 Heilbronn, Germany

⁴ Escuela Técnica Superior de Ingenieros Informáticos (ETSIInf), Universidad Politécnica de Madrid, Calle de los Ciruelos, Boadilla del Monte, 28660 Madrid, Spain

* Correspondence: j.ordieres@upm.es

Version May 24, 2020 submitted to Sensors

Abstract: The information-intensive transformation is vital to realize Industry 4.0 paradigm, where processes, systems and people should be in a connected environment. Current factories must combine different sources of knowledge with different technological layers. Taking into account data interconnection and information transparency, it is necessary to enhance the existing frameworks. This paper proposes an extension to an existing framework, which enables access to knowledge about the different data sources available, including data from operators. To develop the interoperability principle, a specific proposal to provide a (public and encrypted) data management solution to ensure information transparency is presented, which enables semantic data treatment and provides an appropriate context to allow data fusion. This proposal is designed also considering the Privacy by Design option. As a proof of application case, an implementation was carried out regarding the logistics of the delivery of industrial components in the construction sector, where different stakeholders may benefit from shared knowledge under the proposed architecture.

Keywords: Industry 4.0; Reference Architecture Model; Interoperability; Digital Twin; Distributed Ledger Technology; GDPR; RAMI 4.0; LASFA.

1. Introduction

The recent advances in Information Technology, Internet of Things (IoT) and Cyber-Physical Systems (CPS), among others, have enabled digitization and automation of production processes and led to the definition of the fourth industrial revolution, also known as Industry 4.0 (I4.0) [1–3]. In manufacturing domain, the I4.0 vision has promoted the smart manufacturing and smart factories concepts by augmenting all assets with sensor-based connectivity [4]. Intelligent sensors such as positioning tags, safety gloves [5], head-mounted displays (HMDs), and smart glasses [6] have been widely applied in industrial applications. These intelligent sensors generate large volume of industrial data and it remains a challenging task to collect, store, analyze, and exploit them in business, including simulation, virtual reality, digital twins and so on [7].

Human factors, such as fatigue indicators, have significant effects on product quality and factory productivity in manufacturing activities [8]. It is clear that increasing the integration of data (including wearable information) can help to increase understanding of the performance of a business [9].

Indeed, when there is interest in increasing the existing information, a higher level of data integration considering both sub-systems (process and operator) is vital to achieve an effective and efficient human cyber–physical symbiosis [10]. Therefore, pervasive data interconnection and data-driven dynamic decision-making are the key points of Industrial Internet of Things (IIoT) infrastructures, in order to deliver a significant increase in business performance.

To ensure all participants involved in I4.0 share a common perspective and understanding, the *Reference Architecture Model for Industry 4.0* (RAMI 4.0) [11] has been conceived as a standard architecture to describe the fundamental aspects of I4.0. It helps to leverage the process of transitioning classical manufacturing systems to I4.0. However, this model, as well as some existing frameworks [12,13], insufficient focus on those wearable devices in the integration layer, especially those related to human bio-sensing dimensions; in such a way, the value extended from data fusion from wearable devices is weak, from both functional and business points of view. Such imperfections mainly linked to aspects of privacy, data ownership, data silos and a lack of interoperability.

Figure 1. Lower level components of LASFA architecture model [14].

Recently, a new architecture model based on RAMI 4.0, named LASFA (LASim Smart Factory), has been proposed [14]. It adopts the hierarchy of the layers from RAMI 4.0 and focuses on communication among the distributed systems in smart factories, which can help manufacturing companies to transform their manufacturing processes and systems towards the Industry 4.0 paradigm. Figure 1 presents an example of the lower level of such a local production system, as well as the data flows among different components. Compared with RAMI 4.0, the LASFA approach provides a simple configuration for smart factory-enabling agents and manufacturing units in different layers. Due to these advantages, the LASFA architecture model was selected for further development in this work.

Integration of data requires data fusion capabilities, where data fusion from different systems demands interoperability capacities of such systems, as this accounts for the capability of a system to work with other products or systems; either for implementation or access without significant restrictions [15,16].

Interoperability is related to heterogeneity in information theory. Therefore, classical taxonomies have identified Semantic Interoperability, Structural Interoperability, Syntactic Interoperability, and System Interoperability [17]. The focus of this work is to enable the syntactic and structural dimensions of interoperability by fostering terminology transparency and context-sensitive information processing.

This is due to the lack of common data standards (i.e., models, dictionaries, and conventions) which enable the interpretation of data across sub-systems inside an organization, hindering effective data re-usability under a standard format outside of the provided data environment. Therefore, addressing such a gap will help automatic information processes through CPS. The paradigm of cyber–physical equivalence (or digital twin) has emerged [18] to provide a comprehensive physical and functional description of a component, product, or system, in such a way that its information can be useful in any of its life-cycle phases [19].

When a value proposal requires the integration of different stakeholders having different Information Technology (IT) capabilities, such as customers or specific providers, granting access to a public database having immutable but encrypted data is a significant proposal. To this end, different cases can be figured out; on one hand, customers wish to have access not only to the product itself, but also to the production data of the product [20] for different reasons, such as knowledge of the environmental footprint, logistics, and so on.

On the other hand, providers of specific services (e.g., certification bodies) require certified production rates, environmental performance, or continuous improvement, which is another clear example of requiring access to such production-related data streams. Organizations such as regulatory agencies expect holistic and transparent information exchange with peer companies or factories. Furthermore, when asked, 88% of U.S. internet users and 87% of British internet users wish to control the data being collected through smart devices [21]. Encrypted but public databases enable data ownership of producers (especially those related to humans), which can ensure trust and business values resulting from data re-usability by other data consumers.

In modern interconnected societies, where different stakeholders aim to gather different views of different data flows and implement their workflows to make use of them, companies need to consider their value proposals in order to attract current and future stakeholders with relevant data, as well as being able to properly interpret them [22,23]. There is a need for clients, such as middleware platforms, end-users, services, and applications, which efficiently and effectively retrieve IIoT data resources, enabling the integrated and interoperable usage of their data streams.

Due to the different requirements that different stakeholders have in relationship to production-related data encoded into public databases, but also due to the different ways of processing such data, the convenience of integrating semantic meaning into the data itself has arisen. Thus, different stakeholders can easily integrate data processing into their own workflows, and the whole process becomes more resilient to changes in format, the creation of new attributes or entities, and so on. These aspects are linked to data transparency, understood (in the sense of [24–26]) as the degree of visibility and accessibility of information, which includes a component for providing an institutional infrastructure.

Of the secure and non-intensive technologies requiring third-party data access as previously described, the company itself may benefit from providing such standardized production process data flows; particularly when a lean management strategy is implemented. In the case of the usage of the (CPD)_nA method [27–29], the expansion of KPI at different levels through time can render traceability useful for the external certification of auditing processes. In addition, workflows can be activated based on their progress, as per the Value Stream Mapping method [30,31]. Therefore, the management dimension is also relevant for this proposal.

Under the described context, this paper aims to promote the principles of interoperability and data transparency in the I4.0 context by enabling the integration of data coming from different IIoT devices, including wearable devices, as well as enabling the integration from different data sources, with the aim of enhancing data fusion among processes from different stakeholders.

The contributions from this paper are the following:

- An enhanced framework providing flexibility to accommodate different data sources and digital twin, as well as an integrated mechanism to disseminate the desired content while limiting exposition of the related IT systems. This creates a common understanding of interoperability.

- Integration of ontologies with the data sets, enabling data fusion techniques and a higher degree of flexibility for data manipulation, enabling automation in machine learning applications.
- Integration of data dissemination in an encrypted way which is immutable and isolated from other IT areas of the company, through the usage of Distributed Ledger Technology (DLT) solutions. Such automation accounts for the transparency principle. Obviously, its contribution must be understood as not being attached to the any particular DLT ledgers, but as a convenient data management approach to show its capabilities.
- Validation of the proposal through Proof of Concept, as deployed in an industrial case involving a real manufacturing company.

Therefore, in Section 2, we review some background principles and contributions related to attributes connected with the value proposal. Section 3 introduces the proposed architecture, including the reference updated framework and its significant components. Section 4 presents the scenario for proof of concept, involving several devices with different semantic meanings in such a way that their data flows are integrated. In Section 5, a discussion of the business dimension is carried out. Finally, the last section is devoted to presenting our Conclusions.

2. Literature review

There are several principles that need to be considered when a new or improved framework is presented. These principles are:

- *Data silos.* At present, data produced by IIoT deployment devices and wearable devices are controlled or owned by different device manufacturers. Most of the existing value proposals for the IIoT deployment or wearable devices include one platform to manage the physical devices in an integrated way, as well as collecting the data from them by placing it into a private cloud for processing [32,33]. Benefiting from Radio-frequency identification (RFID) and sensor network technologies, common physical objects can be connected and are able to be monitored and managed by a single system [34]. However, as increasing types of IIoT and wearable devices from different vendors become available, more data silos are generated. The existence of such data silos not only jeopardize potential data-driven services and applications; the limited and restricted features provided also dramatically hinder the learning and inference capability derived from the higher requirements of data integration. Most importantly, all related stakeholders in I4.0 are looking forward to a transparent and shared information platform including production and operator data, which is difficult to realize due to the security constraints when passing between various data silos.
- *Data ownership.* It is an additional challenge for data producers (normally referred to as operators) to reuse (e.g., monetize) their data outside the data provider's environments. The data producer cannot benefit from the data they generate, as the data is locked inside of the internal data environment of the enterprise. They lose their data ownership and the business opportunities stemming from those data are outside of the data owner's control. Under this perspective, when referring to data related to humans, the EU General Data Protection Regulation (GDPR) is an effective tool to add significant trust to this dimension, as users are now able to understand their rights and privacy.
- *Privacy.* When data are related to people, such as devices which are linked to apps by bluetooth, the apps are mostly designed to present the data to users and upload summaries to the cloud manufacturer [35]. However, data integration (especially when related to people) has limitations related to privacy and misuse [36,37]. Cyber security is vulnerable, as wearable device manufacturers have reduced their safety protocols and safety stack layers to enable cheap products, as they have been understood as only serving the end user in a local context. Therefore, there is a clear demand regarding the concept of building and embedding security and privacy controls into connected products, as well as the infrastructure itself. This is one of the implementations of the Privacy by Design (PbD) concept [38,39].

- *Interoperability.* IIoT devices, including wearable devices, are highly heterogeneous in terms of the underlying communication protocols, data formats, and technologies from different vendors. Such heterogeneous infrastructures, devices, and configurations have become a strong limitation for data integration and interoperability. By 2021, 25 billion sensor-enabled objects are expected to be connected to the IIoT, as reported by Gartner [40].

Traditional factory-floor control and interconnection data management solutions are mostly based on centralized systems. The German Federal Ministry of Economic Affairs and Energy has provided specifications for the exchange of information within the administration shell. The Open Platform Communications Unified Architecture (OPC UA)[41] is the core communication standard for I4.0-compliant communications [42]. Its adoption was a significant step towards the interconnection of devices, according to RAMI 4.0.

RAMI 4.0 is based on a three-dimensional co-ordinate system consisting of the Layers, Life-Cycle & Value Stream, and Hierarchy Levels dimensions, according to the IEC 62890 and IEC 62264/IEC 51512. It is derived from the Smart Grid Architecture Model (SGAM), which is a key outcome of the EU Mandate M/490 Reference Architecture group for the purposes of communication in networks of renewable energy sources.

In addition, a number of other proposals aimed at contributing to such common perspectives among stakeholders have been proposed by different institutions. Therefore, The National Institute of Standards and Technology (NIST) has promoted the Smart Manufacturing vision as collaborative manufacturing systems which are able to cope with the challenges of quality, efficiency, and personalization that manufacturing companies are presently facing. China has proposed the National Intelligent Manufacturing System Architecture (IMSA), looking to guide the upgrade of Chinese manufacturing towards intelligent manufacturing. However, wearable devices have been neither well-considered as connected elements in the Integration Layer of RAMI4.0 [13], nor in the 5C architecture [12].

After RAMI 4.0, the Industrial Internet Reference Architecture (IIRA) for IIoT Systems has been introduced, promoting IIoT architects to use a standard-based architecture framework and adopt a common vocabulary based on ISO 42010. All of these frameworks have been discussed in [42], although such architectures did not specifically include the operator's contributions.

Compared with RAMI 4.0, the LASFA model can not only enable more reliable and simple modeling of smart factories, but also cover the functionality of ERP, MES, and PLM sub-systems with new digital twins, which are integrated digital agents supported by artificial intelligence. Approaches such as cloud-based data storage methods have created the centralized approach, shifting the way in which franchisers interact with franchisees [32,33]. However, such systems (e.g., PLCs) are expected to be too limited in their accessibility to keep up with the growing demand for higher level integration [43] and information transparency for all relevant stakeholders [44,45].

Distributed ledger technologies (DLTs), such as blockchain, provide a convenient replacement for the central administrator in guaranteeing the integrity of a database. Blockchain-based data sharing systems [46] were proposed to address data security and privacy issues and have gained trust from data owners [47,48]. However, blockchain-based protocols still have several drawbacks that hinder their usage in generic data IoT data sharing platforms [49]. The inherent transaction rate limit leads to low throughput, which brings about scalability issues. The concept of a transaction fee for any value transaction is difficult to get rid of in blockchain-based IIoT infrastructures and some concerns have been raised regarding the susceptibility of blockchains to quantum computers [49].

The lack of interconnected data and shared meaning is another limitation for most DLT-based IIoT data platforms [50,51]. Most sensor data sources, including wearable devices, are characterized by a high degree of heterogeneity, and the implementation of their provided interfaces is highly dependent on the underlying device hardware [52]. These data silos with heterogeneous data jeopardize potential data-driven services and applications. Delicato et al. [53] pointed out that discovering and retrieving data from IoT is a challenge, which is worsened by a lack of standardization of formats in representing

resources. However, efforts are still slow-paced to create interoperable platforms to bridge such huge data silos.

To ensure data privacy and to protect data ownership, the data need to be digitally signed and encrypted; even in a distributed platform supported by DLT. Besides blockchain-structured protocols, some recent DLTs have been developed which are specifically designed for the IoT industry. For example, the Directed Acyclic Graph (DAG)-structured IOTA surpassed conventional blockchains, in terms of its scalability, fee-less, and quantum-resistant features [54]. In the IOTA network (named the Tangle), a new transaction needs to approve two previous transactions to attach itself to the network; this transaction will then be validated by some subsequent transactions [49,54]. This mechanism theoretically eliminates throughput caps as the more transactions that are added, the faster the approval speed is.

IIoT requires interconnected data for predictive maintenance, industrial automation, operational efficiency, and better decision-making. Extending the ability to use sensor data integrated from a wide variety of heterogeneous sources facilitates the building of multiple applications and services, which is a vital step toward the success of the IIoT vision and I4.0. However, to effectively address the interconnection of data from heterogeneous entities ranging from simple sensing devices to complex robotic devices, service agents, or human actors in a consistent way, it is necessary to pay attention to the meanings of the individual data streams [55]. The functions of the information layer in LASFA+ also demand consistent integration of higher-quality data, with the provisioning of structured data [56].

Aimed at the data interconnection issue, semantic modeling (e.g., ontology-based techniques) has proven successful in dealing with data interoperability and semantic compatibility [57,58] in the manufacturing domain. The Blockchain ontology with dynamic extensibility (BLONDiE) ontology [59] has been developed for Bitcoin and Ethereum but is still in its initial phase. Upper level ontologies, such as the Basic Formal Ontology (BFO) [60] and DOLCE [61], work as hub-and-spokes strategies for bringing about interoperability throughout a domain. Ontologies, thus, are in development in various areas. BFO is currently being used in industry IT contexts to gain interoperability in digital manufacturing. Based on BFO, the aim of the Industry Ontology Foundry (IOF) [62] initiative is to create a suite of interoperable and high-quality ontologies covering the domain of industrial (especially manufacturing) engineering [63].

Integration not only happens among the data streams of wearable devices, but higher levels of knowledge can also be derived in industrial contexts when integration between production processes and operator data is considered. This allows more integral access to the infrastructures by means of digital twins [64].

To overcome the aforementioned concerns, this study improves upon an interoperable data handling architecture. The semantic modeling module encompasses heterogeneous resource handling by applying an ontology-based information model, where the ontologies take advantage of well-established and standardized concepts and relationships derived from existing ontologies, in order to semantically represent data resources in IIoT [65].

3. Proposed Architecture

To complement the RAMI 4.0 reference model, and taking advantage of the LASFA architectural model, we integrate all the key elements and their interconnections in a two-dimensional platform. It includes several production layers, as well as a business process. The production layers represent the manufacturing company in terms of production lines, production cells, warehouses, and manual workplaces, creating the local production systems (which are treated as distributed systems). The business layer of the LASFA model includes the company's strategy, business planning, and the monitoring and the delivery of orders [14].

As indicated in Section 1, our main ambition is to extend the proposal of the existing, modern, and valuable framework, LASFA. The general structure of the LASFA architectural model was presented

in [14], and its components and attributes were compared with those of RAMI 4.0 in Section 4 of the above reference.

It is evident that LASFA is much more specific and offers the end-user a simple visualization of the entire architecture of the smart factory, with the definitions of the exact locations and functions of the digital twins and agents, as well as the specific information and data flows between them. The model shows, very clearly, the distribution and autonomy of every single building block in the smart factory, from the product to the management.

In general terms, the LASFA architecture fits with the expectations raised in this work. However, it does have several limitations and, so, this paper adds some extensions to create the improved LASFA model, which we call LASFA+ (see Figure 2).

Figure 2. Improved lower layer for the LASFA+ architectural model.

The first difference is the wider perspective given to the elements participating in production, where not just traditional assets are considered but also extended ones, providing complementary information to the production process. Some of those services rely either on local or (more frequently) on provider-based clouds outside of the company boundaries. Therefore, the reference model needs to be consistent with such realistic conditions. Obviously, such data flows should be integrated with the traditional ones to build the digital twin.

Another difference is the role of the MES and ERP systems, where the MES used to be closer to production and as a support for visualization of the data of the production process.

Machines, devices, and people are equipped with sensors in order to connect and communicate with each other. This proposed architecture enables the connection of people by low-cost wearable devices which can be conveniently obtained in the market, such as personal health-tracking devices. This architecture makes it possible to not only collect physical features and events but also virtual ones, as a result of the model's output. This becomes useful, when integrated, to develop even more complex models.

The information from various sensors can be further processed in edge devices, such as local servers and single-board computers, where such processing includes data fusion, semantic modeling, and data encryption.

Data fusion is relevant to different aspects, and can include data summaries over certain periods of time from a certain type of sensor. In order to provide a higher view of data, raw data may need to be statistically summarized. Extracting data patterns or classification of raw data to high-level knowledge could also be included in data fusion. Machine learning (ML), deep learning (DL), or inference mechanisms can be applied for this purpose. A local server is preferable in this context, considering the high computation demands of ML or DL methods. The motion/positioning data

collected from sensors can be segmented and classified into different types of activity, such as running, walking, and so on.

To enable data interoperability, semantic information is annotated in each data message. Existing ontologies in any domain (besides industrial) are able to be reused under this architecture. Ontology-based structured data can be conveniently integrated with other data sets, in order to empower more complex applications (e.g., data-driven AI algorithms). Ontologies aligned with existing ones or newly proposed ontologies are also supportive in this semantic modeling module.

As shown in Figure 3, an application scenario was developed which enables interoperability among different data sources, further providing a secure, distributed database that empowers data access by different stakeholders. The main components of the prototype include a heterogeneous data source, ontology-based interoperability modeling, and incorporation within the DLT module for data dissemination.

Figure 3. Prototype of data handling in an industrial context.

The corresponding ontology schema is functional as an information model for the automatic understanding of data during retrieval procedures. Data on the DLT, such as the IOTA Tangle, can be retrieved by data consumers. Various data transactions can be retrieved by having the address/tag/bundle or integrated based on defined ontology schema. In this way, the dominant paradigm of keeping data hidden from public knowledge by having protected access is broken and the data becomes part of the product being delivered for further usage. Data consumers may define their own ontology fitting their specific requirements, where this data will be seamlessly integrated.

Although data published to a DLT (e.g., IOTA Tangle), is protected in a secure infrastructure, the message within each transaction confronts a data security issue. If the message is not encrypted, once the transaction address is spread in the Web, any party who has the tag, address of the receiver, or transaction ID is able to read the contents.

Here, different policies can be adopted, such as:

- *Public.* This is the already-described mechanism, where anyone knowing the address of the receiver has the right to freely access to the content.

- *Restricted*. This is the case where sensitive information needs to be shared among only a limited number of stakeholders. To prevent unauthorized entities from reading the data, the message content is obfuscated by encrypting it before uploading it to the Tangle. DLT access at message level, including the encryption/decryption process, is demonstrated in Figure 6.
- *Private*. In this case, due to the specific content of the data, to prevent unauthorized entities from reading the data and to respect the GDPR enforcement rules, the message content is obfuscated by encrypting it before uploading it to the Tangle by using a private certificate, making it Private by Design (PbD).

For messages which are determined to be *Private*, the framework proposed in this paper can be extended in such a way that data owners are able to share a concrete data set defined by matching some conditions (e.g., a period of time) with specific entities. The proposed solution is that the data owner installs a web-service which is able to handle such authorized data requests. The signature key of the data structure will be a public one, preserved under a smart contract and consumed by a web service which is in charge of providing the requested information. The designed data structure for the request can be provided in JSON format for the REST-type implemented protocol, which includes the following elements,

- The DLT address, representing the interested entity or device;
- The DLT tag, representing the interested data set;
- The public key, for the user requesting the access to the DLT stored data;
- Selection criteria for the required data, such as from [DateTime] to [DateTime]; and
- Proof of worth for accessing the data.

Figure 4. Global schema for Web Services operation including certificate usage.

When either people or agents request the web service to access the specific data according to the above structure, when they have the right of access, the web-service will access the DLT repository under the specific criteria provided and verify the ownership of the requested information. This verification will be based on consuming the smart-contract to obtain access to the symmetric encryption key used by the node, thus being able to decrypt it and compare signatures. Then, such data can

be aggregated (according to the requested period of time) and be sent to the requester by the web service after encrypting it using the provided public key of the user/agent. The data consumers will receive back the data and handle it by using its private key, in such a way that only they themselves are allowed to consume it.

The whole schema can be seen in Figure 4, where the role of different certificates is shown clearly. In this way, no third parties are aware of the encrypting/decrypting certificate, as the data is made available to them by re-encrypting it with their public asymmetric certificates, which only enables the third party to decrypt its content.

This type of solution can help companies from at least three different perspectives. They can redesign their adopted business models related to the IIoT and its integration into the management dimension of the organization [66,67]. However, it may also become useful to increase the knowledge of the existing processes in the existing business model, or may just help in the update process of such business models themselves.

In the case of public messages, there are no requirements for encryption. Under the restricted approach, the encryption certificate can be shared among the interested stakeholders, in such a way they can collect the information on their own.

Not only due to the transparency provided, but also because such data streams are delivered automatically by the different digital twins, the trust of shareholders and consumers will increase immediately. In addition, such knowledge will enable the management needs of owners, being more consistent with its potential value when shared.

4. Proof of Concept: an industrial scenario

To validate the proposed architecture, a data-handling system based on a DLT (IOTA), and domain ontologies was designed to improve data interoperability, with integrated privacy criteria through message encryption and data sharing under the Industry 4.0 paradigm.

The implemented system was adopted in a practical industrial scenario. The aim was to verify that our proposed architecture has practical potential in real-world applications, leveraging the implemented system to demonstrate its feasibility. The approach, which adopted a positivist view of research, relied on the literature and empirical data coming from the case itself, as well as on the insights of the researcher to build incrementally more powerful theories.

Figure 5. Application scenario of the proof of concept implementation.

The application case was conducted in a Spanish factory manufacturing steel rebars to reinforce concrete in the construction sector. The general workflow is as shown in Figure 5. Its production process (operation) includes cutting, blending, and welding rebars into different size configurations, packing them into rebar bundles according to customer's order, and loading the rebar bundles into trucks for logistics. For each process, there are different data flows that may affect the production efficiency and effectiveness; the data flows came from data sources such as production line, IIoT devices such as industrial wearable systems (IWSs), and personal wearable devices.

Consider the truck loading process as an example. The requirement is that the right rebar bundles are loaded into the truck in a specific disposal sequence, which needs to be well managed as the bundles are delivered and distributed to different sites. As a lack of specific items could impact the delivery dates, an improper loading disposal sequence may negatively affect the scheduled work in the unloading process. In the present situation, the responsibility for such a decision relies on the crane operators, who take charge of loading the rebar bundles to the truck buffer using a crane machine.

To effectively assess this internal logistics, different data sources should be taken into consideration. Such an analysis requires knowledge about the order and manufacturing sequences, which are handled by the Enterprise resource planning (ERP) system, Manufacturing Execution System (MES), or Product Lifecycle Management (PLM). The logistics information is required, such as where and when the items are loaded into the truck, in order to understand whether the items were loaded and well-placed in a proper storage buffer in the truck.

Moreover, more details, such as the crane operator profile, their working conditions, and working status (e.g., health parameters and movement trajectory/speed) are also relevant. Body-related parameters need to be integrated, mainly as the stress of the crane operators may affect the loading process. The movement trajectory/speed need to be considered, as analysis can be made to understand whether the loading process requires excessive physical efforts.

When different stakeholders need to have access to specific process-related information, non-tampered data are required due to certification principles; therefore, an open system is a convenient tool, reducing the IT barriers and which is robust against facility ownership changes.

The system prototype will adapt the proposal presented in Section 3 by adopting specific mappings over the particular characteristics of the company and its processes.

Different devices, systems, and data sources composed the configured prototype, as indicated below:

- Production related information from ERP/MES/PLM system;
- Ultra wide band (UWB) indoor position system (IPS) to track crane position and crane operator movements, to better define the location for rebar bundles; and
- Smart band to monitor crane operator's heart rate and blood pressure.

The MES manages sequence planning and the bottom-up data flow on the shop-floor [68]. It provides the current state of the process, identifying the optimal resources allocation by taking consideration of the workstation capacity and organization-level requirements (e.g., manufacturing order). All data generated by manufacturing processes are stored in a single database in the company local cloud, based on a Microsoft SQL server database.

The UWB indoor positioning system (IPS) from the TracktioTM company¹ was used to track the crane hook and crane operators. Rebar bundle locations in different buffer areas of production or positions in the truck for customer delivery were derived from the dynamic behavior of the process. The position of crane operator was tracked as a reference for understanding the movement trajectory and speed. The IPS had its own data repository; the data were acquired from a web service.

A smart band was worn by the crane operator, in order to monitor their stress factors, such as blood pressure and heart rate. Every minute, data was collected and uploaded to a smart phone (as an

¹ <https://tracktio.com/>

example of edge computing); then, it was reprocessed and transmitted to a MongoDB database located in the local cloud.

For data interoperability, all data sources in this industrial scenario were semantically modeled, fostering their linkage to other domain knowledge. To this end, different existing domain ontologies were reused in this study. The data sources were in the three aforementioned sectors: production system, IPS, and individual wearable devices. The existing and shared/published ontologies for each sector were collected and one was selected to model the data source in this study, according to the mapping of data structure. The details of available ontologies for reusability are listed in Table 1.

Table 1. Ontology Reusability Selection.

Ontology List		
IPS	Wearable	Production System
Positioning Ontology [69,70]	MIMU-Wear Ontology [71]	MES ontology [72–74]
IndoorGML [75]	SmartBAN Ontology [76]	ERP ontology [77,78]
Navigation ontology [70,79,80]	HealthIoT Ontology [81]	PLM ontology [73]
Indoor space ontology [82,83]	Fitbit Ontology [84]	
	Vital Sign Ontology [85]	

To map the generated data sources in this industrial scenario, three ontologies were chosen for data modeling. The selection criteria were chosen on basis of the degree of matching between the ontological schema and data source structure. After deeper analysis of each ontology's schema, as listed in Table 1, the vital sign ontology [85] was used to model heart rate and blood pressure data from the Smart band; the rebar bundle and crane operator's position were modeled with the positioning ontology [69]; and the MES ontology [72] was applied for the model production system data source.

Table 2. Data semantic transformation based on ontology and Karma.

Semantic Modeling			
Data Source	Data Storage	Ontology	JSON-LD
MES (ISA-95)	Microsoft SQL database	MES ontology [72]	<pre>{ "@context": { "mes": "http://mesontology/schema", "gr": "http://purl.org/goodrelations/v1#", "pto": "http://www.productontology.org/id/", "xsd": "http://www.w3.org/2001/XMLSchema#", "gr:serialNumber": { "@type": "xsd:int" }, "gr:description": { "@type": "xsd:string" }, "gr:amountOfThisGood": { "@type": "xsd:int" }, "gr:eligibleRegions": { "@type": "xsd:string" }, "gr:weight": { "@type": "xsd:float" } "mes:planificationId": "2270077239", "gr:serialNumber": "AAAE3452XX", "gr:description": "bundle of rebars", "gr:amountOfThisGood": "20", "gr:eligibleRegions": "DE(Germany)", "gr:weight": "300", "gr:includes": { "@type": ["gr:Individual", "pto:Rebar"] } }</pre>
Smart band	MongoDB	Vital Sign Ontology [85]	<pre>{ "@context": { "vso": "https://bioportal.bioontology.org/ontologies/VSO", "xsd": "http://www.w3.org/2001/XMLSchema#", "vso:pulserate": { "@type": "xsd:int" }, "vso:systolicbloodpressure": { "@type": "xsd:int" }, "sumo": "http://www.adampease.org/OP/SUMO.owl", "sumo:timepoint": { "@type": "xsd:dateTime" }, "vso:pulserate": "79", "vso:systolicbloodpressure": "111", "sumo:timepoint": "2020-04-08T11:20:00Z" }</pre>
IPS (tracktio)	CSV file	Positioning Ontology [69]	<pre>{ "@context": { "positionpoint": "http://positioningontology/schema", "xsd": "http://www.w3.org/2001/XMLSchema#", "axis": "http://data.ign.fr/def/ignf#CoordinateSystemAxis", "axis:listofaxes": { "@type": "xsd:list" }, "sumo": "http://www.adampease.org/OP/SUMO.owl", "sumo:timepoint": { "@type": "xsd:dateTime" }, "positionpoint": { "axis:listofaxes": "7.299998 1.140002 1.499998" }, "sumo:timepoint": "2020-04-09T10:00:00Z" }</pre>

The mapping tool Karma ² provides a graphical user interface which automates the semantic modeling process. Karma learns to recognize the mapping of data to the chosen ontology classes and proposes a model that can generate JSON-LD for large data sets in a batch mode. JSON-LD is a lightweight Linked Data format, which was designed based on the concept of "context", linking object properties in JSON to concepts in an ontology. The JSON-LD data type was selected as it is lightweight and interoperable at Web-scale, but also provides embedded semantics. The data format is based on JSON and ontological schemas, maintaining a common space of understanding and supporting the evolution of schemes over time without requiring data consumers to change format. The data source, applied ontology schema, and the transformed JSON-LD format (containing semantic annotation) are listed in Table 2.

In order to distribute information about the process interaction, the IOTA network is required. For this, access to a node is needed and a public node (<https://nodes.thetangle.org:443>) was selected for data submission to the IOTA Tangle. The message sent to the IOTA Tangle is encrypted. A python script that implements the RSA (Rivest–Shamir–Adleman) public-key cryptosystem was used for encryption and validation of messages. PyOTA, an IOTA python API library, was used to implement data sending and retrieval to and from the IOTA Tangle. The source code can be found at [86], which supports Python and NodeJS.

A data message describes relevant information combined from different data sources, in which the details about loading a rebar bundle are encrypted and delivered to the IOTA Tangle by the IIoT system, after integrating the meaningful information. In this application, every bundle loaded into a truck and delivered to a construction site was encoded as a message. As the unit is the truck, every truck has a specific encryption certificate, which could be shared with interested stakeholders (e.g., construction workers, transport agencies, insurance companies, or end consumers).

As shown in Figure 6, the loading truck activity is labeled with the tag `[LOADTRUCKTESTCASE]` in IOTA Tangle. Related transactions could be looked up and fetched by tags or bundles. After data retrieval, a data message should be validated and decrypted by stakeholders, where the semantic meaning of the data are encoded in the message itself. This design fosters data reusability by learning processes and growth of models, as they can select the appropriate meaning of data sets or ask for automatic preprocessing before accomplishing further transformations.

5. Discussion

The needs of the near-future dimensions of I4.0 require an increasing level of integration of data from different sources, including data from not only workers, but also from different providers and sources. This enforces PbD requirements, due to both company security rules and regulatory requirements. A major aspect of the GDPR are the so-called legal grounds for lawfully processing personal data, one of which is consent. In several IoT applications where consent is used, it may even need to be explicit consent.

There still remain, after the GDPR, some problems with personal data, as follows:

- Data consistency and the use of different sources of information and/or different time periods or geographical positions;
- Data storage: although properly scrambled, masked, or blurred, the related persons probably are not aware such pieces of information do exist related to them; and
- Problems related to EU citizens when requesting products or services outside of the EU.

The harder problem is the second one, mainly due to the absence of accountability: as a user never knows who is gathering information about them, they cannot ask for access, erasure, or modification. Thus, there are two types of accountability:

² <https://usc-isi-i2.github.io/karma/>

Figure 6. Data handling process supported by IOTA Tangle; loading activity as example.

- Accountability regarding users; but also,
- Accountability within the organization.

In our validation case, the delivery process to the customer has limited information from the process itself, as truck loading is mainly a human-driven activity. Therefore, it risks exposure to a significant time variability and existence of errors in missed or items wrongly included inside the truck. From a business perspective, adding transparency by maximization in the amount of automation involved is worthwhile for stakeholders. Furthermore, current private and centralized data management approaches limit the high-level integration of multi-modal data sources, as well as data ownership and re-usability (e.g., monetizing health data generated from data producers).

The deployed architecture, as proof of concept, was able to provide additional information, integrated in a convenient way in order to help to understand the type of items frequently creating problems, configurations of truck layout consistent with specific delivery sequences, the movements used to load particular item shapes, and so on.

There exist potential contributions in the direction of empowering operators of I4.0 [87], including virtual and augmented reality applications to facilitate the operator's activities and to reduce unwanted mistakes or damage by giving appropriate instructions for complex tasks or working environments. They are also useful for job role allocation and/or for determining training needs, especially in dynamic work environments with changing conditions [88]. Therefore, the impact of the physiological conditions of workers can be considered as relevant, especially the influences of tiredness or stress.

Obviously, to have significance in such an analysis, large periods of data collection and close integration are required. Indeed, path dependency [89] does exist; however, such analysis exceeds the scope of the present research.

Another critical enabling technology for data interoperability adopted in this study is semantic modelling and ontology engineering. Ontologies provide standardized definitions for different data sources and make possible of reasoning and autonomous decision-making, by formalizing the structure of the knowledge. Existing ontologies, i.e. vital sign ontology, positioning ontology, and the MES ontology, are reused in this study to accelerate the development. To obtain a higher level of interoperability, such ontologies can be further exploited and adapted to keep align with widely recognized upper-level ontology like BFO [60] and to be attached to industrial ontology initiatives like IOF [62]. In this way, the developed ontologies will be able to interact with other domain ontologies thus to achieve wider range of interoperability in I4.0 context.

As one of the fundamental enabling technologies of I4.0, the digital twin concept is briefly discussed in this study but not explored with details as it is not the main focus of this paper. Nevertheless, the proposed data integration and handling approaches are important basis for implementation of digital twins. For example, in some existing digital twin studies [90,91], similar semantic modelling techniques have been adopted to improve data interoperability data heterogeneous issues. Aiming to identify the dynamics of virtual model evolution and enhancing the decision-making capabilities, the next generation of digital twin, i.e. cognitive twin concept, has been proposed in a recent study [92], which integrated augmented semantic capabilities with conventional digital twins. The work presented in this paper will have great potential in these digital twin and cognitive twin domains.

6. Conclusions

In this paper, our main contributions were to extend and complement the I4.0 reference model LASFA, with reference to some more general ones (e.g., RAMI4.0), by adding additional data flows to complete the production process description and including different provider add-ons for digital twin construction. The proposed update enables higher level integration and more efficient process description. By considering the PbD design for exchange of information over public DLT systems, it also accounts for information transparency and data ownership for related stakeholders.

The adopted configuration, which enables the semantic enrichment of data, also make it possible to implement rules enforcing interaction mechanisms and to derive new properties related to the employed ontologies. Although such a direction was not implemented in the proof of concept carried out, it may be useful for stakeholders accessing the data.

The presented proposal benefits from the integration of semantically annotated data from different sources, including human health-related wearable device data, in an industrial context. When information related to people has been included, a signature layer is used to provide a PbD solution, in full accordance with the GDPR regulations and to enable positive-Sum [39].

The core idea of the proposed architecture is to enable the sharing of different obfuscated data streams over a public immutable network-oriented database with quick-answer capabilities and good scalability characteristics. The proposed architecture enables extensions, in order to facilitate higher levels of control and empowerment of data owners by integrating an explicit delegation of access to specific data sets and time windows through the use of web services.

Considering a simple application case, the benefits from a business perspective appear evident, with the integration of different dimensions not previously considered. There also exists additional value with the possibility of further performance analysis of procedures, when the proposed architecture is combined with artificial intelligence (AI) techniques. The theoretical implications of this research allow for a convenient framework providing a high level of control on the data produced by each agent, while also providing a platform to share data coherently with different business models. It also provides a flexible way to interconnect semantically powered data streams, which is a significant contribution for data driven ML/AI applications.

This paper also makes a point from a practical application perspective; in particular, by providing an interesting implementation of data handling strategies aligned with Industry 4.0 principles enabling

the integration of data coming from workers. Such tools can have a significant impact on advanced lean manufacturing and lean management implementations, as they go a step beyond classical digitization approaches. This could be the case for the standardization of formal communications under Lean Management (CPD)_nA.

There are several aspects that provide us with further research paths, as we require long data collection approaches and the definition of consistent key performance indicators for individual applications. Furthermore, the definition of related business models for micro (smart contract-based) data monetization and traceability applications also provides avenues for further research.

A clear limitation of the existing research is the adoption of a single and specific DLT technology; namely, IOTA. The proposal is not conceptually dependent on IOTA; it simply served as a driver to implement our architecture in the proof of concept. However, some other technologies, such as Obyte, may be considered as potential substitutes.

Another limitation is the required digital infrastructure needed in the companies to successfully implement the digital twin concepts, which is capable to manage requirement of both internal and external stakeholders.

Future research steps will be to deeply analyze the meaning of the integrated data built, in terms of value for business, looking to identify behavioral patterns or to create forecasting models, and its ability to better adjust operation times and, most importantly, timely delivery.

Another interesting perspective to be considered is that wearable devices, such as smart watches [88] and smart wristbands, can be useful for estimating the operator's well-being (i.e., cognitive, psychological, and physical needs) in Human Cyber-Physical Systems (H-CPS) [93]. Wearable monitoring provides opportunities for health diagnostics and operator well-being in industrial environments, as has also been identified in [94,95].

Author Contributions: SS was responsible of the development of the case study, and she shared the efforts of implementation of the dissemination with XZ. JOM contributed with SS and JV to provide the industrial dimension context for the application. The IOTA selection was the responsibility of JOM and XZ.

Funding: This research was partially funded by the European Commission through the RFCS program, grant number with ID 793505.

Acknowledgments: The authors thank China Scholarship Council (CSC) (No.201206730038) for the support they provided for this work. Also the authors want to thank the EU RFCS research program for the support of this research through the grant with ID 793505 (WISEST project). Indeed, JOM wants to acknowledge the Spanish Agencia Estatal de Investigacion, through the research project with code RTI2018-094614-B-I00 into the "Programa Estatal de I+D+i Orientada a los Retos de la Sociedad".

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
DAG	Directed Acyclic Graph
DL	Deep Learning
DLT	Distributed Ledger Technology
ERP	Enterprise Resource Planning
GDPR	General Data Protection Regulation
I4.0	Industry 4.0 paradigm
IIoT	Industrial Internet of Things
IoT	Internet of Things
IPS	Indoor Position System
IT	Information Technology
IWS	Industrial Wearable System
JSON	JavaScript Object Notation
JSON-LD	JavaScript Object Notation for Linked Data
KPI	Key Performance Indicator
LASFA	Lasim Smart Factory reference model
MES	Manufacturing Execution System
ML	Machine Learning Techniques
OPC UA	Open Platform Communications Unified Architecture
PbD	Privacy by Design
PLM	Product Lifecycle Management
UWB	Ultra Wide Band

References

- Piccarozzi, M.; Aquilani, B.; Gatti, C. Industry 4.0 in management studies: A systematic literature review. *Sustainability* **2018**, *10*, 3821.
- Maresova, P.; Soukal, I.; Svobodova, L.; Hedvicakova, M.; Javanmardi, E.; Selamat, A.; Krejcar, O. Consequences of industry 4.0 in business and economics. *Economies* **2018**, *6*, 46.
- Villalba-Díez, J.; Molina, M.; Ordieres-Meré, J.; Sun, S.; Schmidt, D.; Wellbrock, W. Geometric Deep Lean Learning: Deep Learning in Industry 4.0 Cyber-Physical Complex Networks. *Sensors* **2020**, *20*, doi:10.3390/s20030763.
- Grangel-González, I.; Halilaj, L.; Coskun, G.; Auer, S.; Collarana, D.; Hofmeister, M. Towards a Semantic Administrative Shell for Industry 4.0 Components. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)* **2016**, pp. 230–237.
- Scheuermann, C.; Heinz, F.; Bruegge, B.; Verclas, S. Real-Time Support During a Logistic Process Using Smart Gloves. *Smart SysTech 2017; European Conference on Smart Objects, Systems and Technologies*, 2017, pp. 1–8.
- Hao, Y.; Helo, P. The role of wearable devices in meeting the needs of cloud manufacturing: A case study. *Robotics and Computer-Integrated Manufacturing* **2017**, *45*, 168 – 179. Special Issue on Ubiquitous Manufacturing (UbiM), doi:https://doi.org/10.1016/j.rcim.2015.10.001.
- Vrchota, J.; Pech, M. Readiness of Enterprises in Czech Republic to Implement Industry 4.0: Index of Industry 4.0. *Applied Sciences* **2019**, *9*, 5405.
- Neumann, W.; Kolus, A.; Wells, R. Human Factors in Production System Design and Quality Performance - A Systematic Review. *IFAC-PapersOnLine* **2016**, *49*, 1721–1724. doi:10.1016/j.ifacol.2016.07.830.
- Sun, S.; Zheng, X.; Gong, B.; Paredes, J.G.; Ordieres-Meré, J. Healthy Operator 4.0: A Human Cyber-Physical System Architecture for Smart Workplaces. *Sensors* **2020**, *20*, 2011. doi:10.3390/s20072011.
- Kong, X.T.R.; Yang, X.; Huang, G.Q.; Luo, H. The impact of industrial wearable system on industry 4.0. *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, 2018, pp. 1–6. doi:10.1109/ICNSC.2018.8361266.

11. Zvei, E. Plattform Industrie 4.0: The reference Architecture Model of Industrie 4.0 (2015). Available online: <https://www.zvei.org/en/subjects/industrie-4-0/the-reference-architectural-model-rami-4-0-and-the-industrie-4-0-component/> (Accessed on 25 April 2020).
12. Lee, J.; Bagheri, B.; Kao, H.A. A Cyber-Physical Systems architecture for Industry 4.0-based manufacturing systems. *Manufacturing Letters* **2015**, *3*, 18–23. doi:<https://doi.org/10.1016/j.mfglet.2014.12.001>.
13. Alcácer, V.; Cruz-Machado, V. Scanning the Industry 4.0: A Literature Review on Technologies for Manufacturing Systems. *Engineering Science and Technology, an International Journal* **2019**, *22*, 899–919. doi:<https://doi.org/10.1016/j.jestch.2019.01.006>.
14. Resman, M.; Pipan, M.; Simic, M.; Herakovic, N. A new architecture model for smart manufacturing: A performance analysis and comparison with the RAMI 4.0 reference model. *Advances in Production Engineering & Management* **2019**, *14*, 153–165. doi:10.14743/apem2019.2.318.
15. Wiemann, S.; Bernard, L. Spatial data fusion in spatial data infrastructures using linked data. *International Journal of Geographical Information Science* **2016**, *30*, 613–636.
16. Gagnon, M. Ontology-based integration of data sources. 2007 10th International Conference on Information Fusion. IEEE, 2007, pp. 1–8.
17. Noura, M.; Atiquzzaman, M.; Gaedke, M. Interoperability in Internet of Things Infrastructure: Classification, Challenges, and Future Work. IoT as a Service; Lin, Y.B.; Deng, D.J.; You, I.; Lin, C.C., Eds.; Springer International Publishing: Cham, 2018; pp. 11–18.
18. Glaessgen, E.; Stargel, D. The digital twin paradigm for future NASA and US Air Force vehicles. 53rd AIAA/ASME/ASCE/AHS/ASC structures, structural dynamics and materials conference 20th AIAA/ASME/AHS adaptive structures conference 14th AIAA, 2012, p. 1818.
19. Boschert, S.; Rosen, R. Digital twin—the simulation aspect. In *Mechatronic futures*; Springer, 2016; pp. 59–74.
20. Kagermann, H. Change through digitization—Value creation in the age of Industry 4.0. In *Management of permanent change*; Springer, 2015; pp. 23–45.
21. Groopman, J. Circles of Trust for Smart Devices for US citizens. Source: Privacy & The Internet of Things: The Importance of Transparency in Accounting for What We Can't See. Available online: <https://www.trustarc.com/blog/tag/iot-summit/> (Accessed on 25 April 2020).
22. Frysak, J.; Krenn, F.; Kaar, C.; Sary, C. Decision-Making Support for View-Oriented I4.0 System Architecture Design. 2018 IEEE 20th Conference on Business Informatics (CBI), 2018, Vol. 01, pp. 186–195.
23. Bakhshandeh, M.; Antunes, G.; Mayer, R.; Borbinha, J.; Caetano, A. A Modular Ontology for the Enterprise Architecture Domain. 2013 17th IEEE International Enterprise Distributed Object Computing Conference Workshops, 2013, pp. 5–12.
24. Zhu, K. Information transparency in electronic marketplaces: Why data transparency may hinder the adoption of B2B exchanges. *Electronic markets* **2002**, *12*, 92–99.
25. Flatt, H.; Schriegel, S.; Jasperneite, J.; Trsek, H.; Adamczyk, H. Analysis of the Cyber-Security of industry 4.0 technologies based on RAMI 4.0 and identification of requirements. 2016 IEEE 21st International Conference on Emerging Technologies and Factory Automation (ETFA), 2016, pp. 1–4.
26. Müller, J.M.; Kiel, D.; Voigt, K.I. What Drives the Implementation of Industry 4.0? The Role of Opportunities and Challenges in the Context of Sustainability. *Sustainability* **2018**, *10*. doi:10.3390/su10010247.
27. Villalba-Diez, J.; Ordieres-Meré, J. Improving manufacturing performance by standardization of interprocess communication. *IEEE Transactions on Engineering Management* **2015**, *62*, 351–360.
28. Villalba-Diez, J.; Ordieres-Meré, J.; Nuber, G.. The HOSHIN KANRI TREE. Cross-Plant Lean Shopfloor Management. *Procedia CIRP*; Elsevier: Bochum, Germany, 2015; Vol. 32, pp. 150–155. doi:10.1016/j.procir.2015.02.120.
29. Villalba-Diez, J.; Ordieres-Meré, J.; Chudzick, H.; Lopez-Rojo, P. NEMAWASHI: Attaining Value Stream alignment within Complex Organizational Networks. *Procedia CIRP*. Elsevier, 2015, Vol. 37, pp. 134–139. doi:10.1016/j.procir.2015.08.021.
30. Villalba-Diez, J. *The Lean Brain Theory. Complex Networked Lean Strategic Organizational Design*; CRC Press. Taylor and Francis Group LLC: Boca Raton, FL, USA, 2017.
31. Villalba-Diez, J. *The HOSHIN KANRI FOREST. Lean Strategic Organizational Design*, 1st ed.; CRC Press. Taylor and Francis Group LLC: Boca Raton, FL, USA, 2017.

32. Lee, J.; Ardakani, H.D.; Yang, S.; Bagheri, B. Industrial Big Data Analytics and Cyber-physical Systems for Future Maintenance & Service Innovation. *Procedia CIRP* **2015**, *38*, 3 – 7. doi:<https://doi.org/10.1016/j.procir.2015.08.026>.
33. Liu, Y.; Xu, X. Industry 4.0 and Cloud Manufacturing: A Comparative Analysis. *Journal of Manufacturing Science and Engineering* **2016**, *139*. doi:10.1115/1.4034667.
34. Jiang, L.; Da Xu, L.; Cai, H.; Jiang, Z.; Bu, F.; Xu, B. An IoT-oriented data storage framework in cloud computing platform. *IEEE Transactions on Industrial Informatics* **2014**, *10*, 1443–1451.
35. Stergiou, C.; Psannis, K.E.; Kim, B.G.; Gupta, B. Secure integration of IoT and cloud computing. *Future Generation Computer Systems* **2018**, *78*, 964–975.
36. D’Antrassi, P.; Prenassi, M.; Rossi, L.; Ferrucci, R.; Barbieri, S.; Priori, A.; Marceglia, S. Personally Collected Health Data for Precision Medicine and Longitudinal Research. *Frontiers in medicine* **2019**, *6*, 125.
37. Antón, M.Á.; Ordieres-Meré, J.; Saralegui, U.; Sun, S. Non-Invasive Ambient Intelligence in Real Life: Dealing with Noisy Patterns to Help Older People. *Sensors* **2019**, *19*. doi:10.3390/s19143113.
38. Bednar, K.; Spiekermann, S.; Langheinrich, M. Engineering Privacy by Design: Are engineers ready to live up to the challenge? *The Information Society* **2019**, *35*, 122–142.
39. Cavoukian, A. Privacy by Design: The 7 foundational principles. Available online: <https://www.ipc.on.ca/wp-content/uploads/resources/7foundationalprinciples.pdf> (Accessed on 25 April 2020).
40. Gartner. Gartner Identifies Top 10 Strategic IoT Technologies and Trends. Available online: <https://gtmr.it/2JLiEbA> (Accessed on 06 February 2020).
41. OPC, U. Specification: Part 1–Concepts, Version 1.0 or later. *OPC Foundation* **2009**.
42. Fraile, J.; Sanchis, J.; Poler, J.; Ortiz, J. Reference Models for Digital Manufacturing Platforms. *Applied Sciences* **2019**, *9*, 4433. doi:10.3390/app9204433.
43. Glebke, R.; Henze, M.; Wehrle, K.; Niemietz, P.; Trauth, D.; Mattfeld, M.; Bergs, T.; others. A Case for Integrated Data Processing in Large-Scale Cyber-Physical Systems. *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019.
44. Al-Jabri, I.M.; Roztock, N. Adoption of ERP systems: Does information transparency matter? *Telematics and Informatics* **2015**, *32*, 300 – 310. doi:<https://doi.org/10.1016/j.tele.2014.09.005>.
45. Ghobakhloo, M. The future of manufacturing industry: a strategic roadmap toward Industry 4.0. *Journal of Manufacturing Technology Management* **2018**, *29*, 910–936.
46. Zhang, L.; Luo, M.; Li, J.; Au, M.H.; Choo, K.K.R.; Chen, T.; Tian, S. Blockchain based secure data sharing system for Internet of vehicles: A position paper. *Vehicular Communications* **2019**, *16*, 85–93. doi:<https://doi.org/10.1016/j.vehcom.2019.03.003>.
47. Wang, L.; Guo, S. Blockchain Based Data Trust Sharing Mechanism in the Supply Chain. *Security with Intelligent Computing and Big-data Services*; Yang, C.N.; Peng, S.L.; Jain, L.C., Eds. Springer International Publishing, 2020, pp. 43–53.
48. Shen, B.; Guo, J.; Yang, Y. MedChain: Efficient Healthcare Data Sharing via Blockchain. *Applied Sciences* **2019**, *9*. doi:10.3390/app9061207.
49. Zheng, X.; Sun, S.; Mukkamala, R.R.; Vatrappu, R.; Ordieres-Meré, J. Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies. *J Med Internet Res* **2019**, *21*, e13583. doi:10.2196/13583.
50. Ledwaba, L.P.I.; Hancke, G.P.; Isaac, S.J.; Venter, H.S. Developing a Secure, Smart Microgrid Energy Market using Distributed Ledger Technologies. 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 2019, Vol. 1, pp. 1725–1728. doi:10.1109/INDIN41052.2019.8972018.
51. Nikander, P.; Autiosalo, J.; Paavolainen, S. Interledger for the Industrial Internet of Things. 2019 IEEE 17th International Conference on Industrial Informatics (INDIN), 2019, Vol. 1, pp. 908–915. doi:10.1109/INDIN41052.2019.8972167.
52. Gomes, P.; Cavalcante, E.; Batista, T.; Taconet, C.; Conan, D.; Chabridon, S.; Delicato, F.C.; Pires, P.F. A semantic-based discovery service for the Internet of Things. *Journal of Internet Services and Applications* **2019**, *10*, 10. doi:10.1186/s13174-019-0109-8.
53. FC, D.; PF, P.; T, B. Resource management for the Internet of Things. *Springer International Publishing AG* **2017**. doi:<https://doi.org/10.1016/j.vehcom.2019.03.003>.

54. Popov, S. The Tangle. Available online: https://assets.ctfassets.net/r1dr6vzfxhev/2t4uxvslqk0EUau6g2sw0g/45eae33637ca92f85dd9f4a3a218e1ec/iota1_4_3.pdf (Accessed on 25 April 2020).
55. Sahinel, D.; Akpolat, C.; Görür, O.C.; Sivrikaya, F. Integration of Human Actors in IoT and CPS Landscape. 2019 IEEE 5th World Forum on Internet of Things (WF-IoT), 2019, pp. 485–490.
56. Zezulka, F.; Marcon, P.; Vesely, I.; Sajdl, O. Industry 4.0 – An Introduction in the phenomenon. *IFAC-PapersOnLine* **2016**, *49*, 8 – 12. doi:<https://doi.org/10.1016/j.ifacol.2016.12.002>.
57. Giannetti, C.; Ransing, M.R.; Ransing, R.S.; Bould, D.C.; Gethin, D.T.; Sienz, J. Organisational Knowledge Management for Defect Reduction and Sustainable Development in Foundries. *Int. J. Knowl. Syst. Sci.* **2015**, *6*, 18–37. doi:10.4018/ijkss.2015070102.
58. Cheng, H.; Zeng, P.; Xue, L.; Shi, Z.; Wang, P.; Yu, H. Manufacturing Ontology Development Based on Industry 4.0 Demonstration Production Line. 2016 Third International Conference on Trustworthy Systems and their Applications (TSA), 2016, pp. 42–47. doi:10.1109/TSA.2016.17.
59. Blockchain Ontology with Dynamic Extensibility. Available online: <https://github.com/hedugaro/Blondie> (Accessed on 06 February 2020).
60. Smith, B.; Grenon, P.; Goldberg, L. Biodynamic Ontology: Applying BFO in the Biomedical Domain. *Studies in Health and Technology Informatics* **2004**, *102*, 20–38.
61. Gangemi, A.; Guarino, N.; Masolo, C.; Oltramari, A.; Schneider, L. Sweetening Ontologies with DOLCE. Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web; Gómez-Pérez, A.; Benjamins, V.R., Eds. Springer Berlin Heidelberg, 2002, pp. 166–181.
62. Industrial Ontologies Foundry (IOF). Available online: <https://sites.google.com/view/industrialontologies/home> (Accessed on 25 April 2020).
63. Yun, J.; Ahn, I.Y.; Song, J.; Kim, J. Implementation of Sensing and Actuation Capabilities for IoT Devices Using oneM2M Platforms. *Sensors* **2019**, *19*, 4567.
64. Cimino, C.; Negri, E.; Fumagalli, L. Review of digital twin applications in manufacturing. *Computers in Industry* **2019**, *113*, 103130.
65. Jha, S.B.; Babiceanu, R.F.; Seker, R. Formal modeling of cyber-physical resource scheduling in IIoT cloud environments. *Journal of Intelligent Manufacturing* **2019**, pp. 1–16.
66. Ikävalko, H.; Turkama, P.; Smedlund, A. Value creation in the internet of things: Mapping business models and ecosystem roles. *Technology Innovation Management Review* **2018**, *8*.
67. Zhang, Y.; Wen, J. The IoT electric business model: Using blockchain technology for the internet of things. *Peer-to-Peer Networking and Applications* **2017**, *10*, 983–994.
68. D’Antonio, G.; Macheda, L.; Sauza Bedolla, J.; Chiabert, P. PLM-MES Integration to Support Industry 4.0. Product Lifecycle Management and the Industry of the Future; Ríos, J.; Bernard, A.; Bouras, A.; Foufou, S., Eds. Springer International Publishing, 2017, pp. 129–137.
69. Carbonera, J.L.; Fiorini, S.R.; Prestes, E.; Jorge, V.A.M.; Abel, M.; Madhavan, R.; Locoro, A.; Gonçalves, P.; Haidegger, T.; Barreto, M.E.; Schlenoff, C. Defining positioning in a core ontology for robotics. 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2013, pp. 1867–1872.
70. Kun, D.P.; Varga, E.B.; Toth, Z. Ontology based navigation model of the ILONA system. 2017 IEEE 15th International Symposium on Applied Machine Intelligence and Informatics (SAMI), 2017, pp. 479–484. doi:10.1109/SAMI.2017.7880357.
71. Villalonga, C.; Pomares, H.; Rojas, I.; Banos Legran, O. MIMU-Wear: ontology-based sensor selection for real-world wearable activity recognition. *Neurocomputing* **2017**, pp. 1–25. doi:10.1016/j.neucom.2016.09.125.
72. Long, W. Construct MES Ontology with OWL. 2008 ISECS International Colloquium on Computing, Communication, Control, and Management, 2008, Vol. 1, pp. 614–617.
73. Khedher, A.B.; Henry, S.; Bouras, A. Integration between MES and Product Lifecycle Management. ETFA2011. IEEE, 2011. doi:10.1109/etfa.2011.6058993.
74. Fumagalli, L.; Pala, S.; Garetti, M.; Negri, E. Ontology-Based Modeling of Manufacturing and Logistics Systems for a New MES Architecture. Advances in Production Management Systems. Innovative and Knowledge-Based Production Management in a Global-Local World; Grabot, B.; Vallespir, B.; Gomes, S.; Bouras, A.; Kiritsis, D., Eds. Springer Berlin Heidelberg, 2014, pp. 192–200.

75. OGC. OGC® IndoorGML. Available online: <https://www.ogc.org/standards/indoorgml> (Accessed on 06 February 2020).
76. ETSI. Smart Body Area Networks (SmartBAN) Unified data representation formats, semantic and open data model. Available online: https://www.etsi.org/deliver/etsi_ts/103300_103399/103378/01.01.01_60/ts_103378v010101p.pdf (Accessed on 25 April 2020).
77. Nach, H.; Lejeune, A. Implementing ERP in SMEs: Towards an Ontology Supporting Managerial Decisions. 2008 International MCETECH Conference on e-Technologies (mcetech 2008), 2008, pp. 223–226.
78. Hepp, M.; Wechselberger, A. OntoNaviERP: Ontology-Supported Navigation in ERP Software Documentation. The Semantic Web - ISWC 2008; Sheth, A.; Staab, S.; Dean, M.; Paolucci, M.; Maynard, D.; Finin, T.; Thirunarayan, K., Eds. Springer Berlin Heidelberg, 2008, pp. 764–776.
79. Dudas, P.M.; Ghafourian, M.; Karimi, H.A. ONALIN: Ontology and Algorithm for Indoor Routing. 2009 Tenth International Conference on Mobile Data Management: Systems, Services and Middleware, 2009, pp. 720–725.
80. Stevenson, G.; Ye, J.; Dobson, S.; Nixon, P. LOC8: A Location Model and Extensible Framework for Programming with Location. *IEEE Pervasive Computing* **2010**, *9*, 28–37.
81. Rhayem, A.; Mhiri, M.B.A.; Gargouri, F. HealthIoT Ontology for Data Semantic Representation and Interpretation Obtained from Medical Connected Objects. 2017 IEEE/ACS 14th International Conference on Computer Systems and Applications (AICCSA), 2017, pp. 1470–1477.
82. Maheshwari, N.; Srivastava, S.; Rajan, K.S. Development of an Indoor Space Semantic Model and Its Implementation as an IndoorGML Extension. *ISPRS International Journal of Geo-Information* **2019**, *8*. doi:10.3390/ijgi8080333.
83. Khruahong, S.; Kong, X.; Sandrasegaran, K.; Liu, L. Develop An Indoor Space Ontology For Finding Lost Properties for Location-Based Service of Smart City. 2018 18th International Symposium on Communications and Information Technologies (ISCIT), 2018, pp. 54–59.
84. universAAL IoT. Fitbit Ontology. Unified data representation formats, semantic and open data model. Available online: <https://github.com/universAAL/ontology/issues/479> (Accessed on 25 April 2020).
85. Goldfain, A.; Smith, B.; Arabandi, S.; Brochhausen, M.; Hogan, W. Vital Sign Ontology. In *The 14th Annual Bio-Ontologies Meeting; 15-16th July* **2011**.
86. Sun, S. Interoperable Data Handling for Industry 4.0. Available online: <https://gitlab.com/sunshengjing/interoperable-data-usage-for-industry-4.0> (Accessed on 25 April 2020).
87. Longo, F.; Nicoletti, L.; Padovano, A. Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering* **2017**, *113*, 144–159.
88. Aehnelt, M.; Urban, B. Follow-Me: Smartwatch Assistance on the Shop Floor. HCI in Business; Nah, F.F.H., Ed. Springer International Publishing, 2014, pp. 279–287.
89. Chapin, F.S.; Folke, C.; Kofinas, G.P. A framework for understanding change. In *Principles of ecosystem stewardship*; Springer, 2009; pp. 3–28.
90. Kharlamov, E.; Martin-Recuerda, F.; Perry, B.; Cameron, D.; Fjellheim, R.; Waaler, A. Towards semantically enhanced digital twins. 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018, pp. 4189–4193.
91. Gómez-Berbís, J.M.; de Amescua-Seco, A. SEDIT: Semantic Digital Twin Based on Industrial IoT Data Management and Knowledge Graphs. International Conference on Technologies and Innovation. Springer, 2019, pp. 178–188.
92. Lu, J.; Zheng, X.; Gharaei, A.; Kalaboukas, K.; Kiritsis, D. Cognitive Twins for Supporting Decision-Makings of Internet of Things Systems. *arXiv preprint arXiv:1912.08547* **2019**.
93. Zhou, J.; Zhou, Y.; Wang, B.; Zang, J. Human-Cyber-Physical Systems (HCPs) in the Context of New-Generation Intelligent Manufacturing. *Engineering* **2019**, *5*, 624 – 636. doi:https://doi.org/10.1016/j.eng.2019.07.015.
94. Roda-Sanchez, L.; Garrido-Hidalgo, C.; Hortelano, D.; Olivares, T.; Ruiz, M.C. OperaBLE: An IoT-Based Wearable to Improve Efficiency and Smart Worker Care Services in Industry 4.0. *Journal of Sensors* **2018**, *2018*, 1–12. doi:10.1155/2018/6272793.

- 817 95. Awolusi, I.; Marks, E.; Hallowell, M. Wearable technology for personalized construction safety
818 monitoring and trending: Review of applicable devices. *Automation in Construction* **2018**, *85*, 96 – 106.
819 doi:<https://doi.org/10.1016/j.autcon.2017.10.010>.

820 © 2020 by the authors. Submitted to *Sensors* for possible open access publication under the terms and conditions
821 of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

Chapter 4

Conclusions

Toward a high-quality human-centered society, the future of IoT is also seeking people-centered solutions. The study focuses on the health dimension, its contributions across the whole IoT value chain life cycle. To answer the research question 'How to accelerate and enhance people's health and well-being in the IoT data value chain?', the action research method was employed to analyze and implement the research. The research aim was defined to accelerate and facilitate health-related data-driven knowledge acquisition and data value chain to society, organizations, and individuals, that leverages advanced technologies such as IoT, Smart Wearables, DLT, and machine learning techniques. Specifically, the research aim was shaped via four research objectives. Firstly, to enhance people's dimension, especially health and well-being perspective in conventional IoT architectures. Secondly, to develop continuous long term monitoring solutions to support better health management. Thirdly, to learn psychological and physical health impacts from (e.g., workers) a group of people over the activities. Fourthly, to accelerate health-related data sharing with security, trust, and privacy assurance.

To implement the defined research objectives, a guiding framework was designed to set up the research structure, where key concepts, mainly enabling technologies, applied devices, and the developed tools were introduced. To reach the research aims, four study designs were presented in the study that detailed describe the experimental setup, data collection mechanism, and data analysis approach.

Eleven journal/conference articles have been published concentrating on the research question. Each publication has its significant contributions, which are aligned to the four research emphasis, respectively. Figure 4.1 gives a systematic view that shows how each publication's value is streaming into those four research aims.

- The publication titled *Indoor Air-Quality Data-Monitoring System: Long-Term Monitoring Benefits* [36] addressed short-term monitoring bias by providing a low-cost IoT based monitoring system, in the specific case of Occupational Safety and Health (OSH). The proposed system supports long-term, multiple pollutants data-monitoring and enables

Figure 4.1: Map of research outcomes.

customization in choosing pollutant sensors by a simple plug-in. To this end, temporal variations of various physical and chemical pollutants in the mid-long term can be tracked and analyzed leveraging the proposed system. The bringing in benefits was discussed in the individual and organization levels. For individuals, health awareness was enabled to empower people in the workplace in terms of performance improvement such as productivity. About the organization, the real-time working environmental data will be possible to be considered as other business objectives in KPIs, because working conditions affect people's performance. The developed solution also enables the organization to take measures to improve the working environment based on collected data indicators. The OSH regulatory agency can benefit from the long-term monitoring solution in terms of better catching up significant variations and exceeding via recommendation strategy powered by the proposed system. Transparency is enhanced as well via DLT to establish trust among organizations. Peer companies can share data streams with OSH regulatory agencies to fulfill legal compliance.

- The publication titled *The Use of the Internet of Things for Estimating Personal Pollution Exposure* [121] provides a holistic solution to monitor personal exposure tracking temporal and spatial variations. A research framework for air quality decision making support system was proposed to connect personal exposure assessment to different stakeholders namely individual, healthcare industry, and public health linked to the government. The proposed approach for personal exposure assessment has significant strength in cost and non-invasive aspect by combining air quality monitors with e-beacon technology. That can

boost personal exposure data collection scope to a large scale, especially in those developed countries.

- The publication titled *Non-Invasive Ambient Intelligence in Real Life: Dealing with Noisy Patterns to Help Older People* [143] contributes to discovering non-intrusive 'software' sensors for providing a healthy living environment for older people or dementia people. The built 'software' sensors make it possible to assist people when assistance and care are needed, to evaluate people's long-term changes for an accurate diagnosis. But most importantly, the non-intrusive feature is useful to fuse sensors seamlessly into people's routines, meanwhile effectively addressing people's trust concerns.
- The publication titled *Accelerating Health Data Sharing: A Solution Based on the Internet of Things and Distributed Ledger Technologies* [111] empowers health data sharing with security and privacy endorsement. The adopted distributed ledger technologies, which is Tangle based, overcame the limitations of traditional blockchain shows better suitability to be used as the data-sharing approach. Through the proposed data-sharing platform, the big health data paradigm can be fulfilled by fully utilizing and exploiting the data silos fragmented in different levels of (individual and organizational) storage.
- The publication titled *Geometric Deep Lean Learning: Deep Learning in Industry 4.0 Cyber-Physical Complex Networks* [144] was an exploration in handling complexity in the scale of big data value streams. A theoretical model integrating geometric deep-learning was proposed to tackle the needs of massive and multiple types of data networks. The inspired geometric deep lean learning injects a strategic level of transparency and traceability on processes inside organizations, which impacts business improvement in the product life-cycle.
- The publication title *Healthy Operator 4.0: A Human Cyber-Physical System Architecture for Smart Workplaces* [66] enhanced the human dimension compared with the transitional Cyber-Physical System architecture, especially in a health perspective. As health-related factors may come from different places such as working environment and activity executed, in the short, or long term, a holographic health monitoring mechanism ought to be built to ensure people's health and further wellness. To this end, the healthy operator 4.0 (HO4.0) concept was defined as a holistic system focused on the health and well-being aspect. The proposed HO4.0 architecture provides a reference framework to facilitate various sensing streams integration, human behavior modeling, and knowledge creation based decision making, towards a healthier smart workplace. Furthermore, a real-world validation of the proposed architecture was carried out in a Spanish factory.
- The publication title *Data Handling in Industry 4.0: Interoperability Based on Distributed Ledger Technology* [145] is towards improving data interconnection and information transparency in the Industry 4.0 paradigm. An extension to the existing framework was pro-

posed which is to enhance data access of different data sources including data from people. The research work demonstrated how different data sources can be interoperated and shared align with the extended framework. The study envisioned and instantiated an enhanced data management solution leveraging ontology modeling and DLT, also considered the Privacy by Design option. The proposed framework and solution would play a vital role to realize Industry 4.0 paradigm where processes, systems, and people are connected in cyberspace.

- The publication titled *Biophilic Adaptation at the Workplace: Multidimensional Impact on Human's Behavior* [146], in this study, how the working environment affects people's behavior was explored. The working space with limited air quality was detected by the developed air monitoring system after a mid-term monitoring period. In that workplace, different profile people's cognitive performance, productivity, focus, and creativity were accessed by designed tests and wearable devices. For example, the focus/concentration level was monitored by a smart glass. A comparison experiment was conducted as well to explore how biophilic design may help improve people's performance in a dissatisfy working space.
- The publication titled *Physiological and Psychological Responses to Transitions Between Urban and Urban Nature Environments Using Immersive Visualization* [149] was an extension study to explore how to improve human health and well-being leveraging nature space. This study enrolled 173 healthy adults via the Harvard Decision Science Lab in Boston. Their physiological and psychological responses such as heart rate and skin conductance level were measured continuously by Shimmer sensing. All participants experienced a total of four immersive outdoor environments (two urban and two urban nature) provided by 12k digital big screen. The result indicates that transitions from urban to urban nature is beneficial to residents for stress relief, that quantified research result would help urban planning decision making.
- The publication titled *Applying Distributed Ledger Technology to Facilitate IIoT Data Exchange: an Approach based on IOTA Tangle* [147] and *Decentralized Industrial IoT Data Management Based on Blockchain and IPFS* [148] made extension research based on DLT study in health data sharing. The application of DLT in Industrial IoT data exchange, especially when large data volume is involved, was studied. The InterPlanetary File System (IPFS) combining IOTA was proposed to provide a suitable complementary data management approach to address large data exchange.

As one of the three pillars in the sustainability concept, the role of people should be considered carefully to build a sustainable ecosystem. In a future society, the long-term return of investment is imperative to promote enterprise transformation, empowering people will act as a key step delivering values to all product life cycles. In industry 4.0 era, a full range of vari-

ables and steps including people can provide a broader perspective that impacts a production process. Collecting people's data e.g. health-related would enable a holistic measure space to boost overall performance effectiveness. In light of the continual improvement in management, continuous feedback from the process is significant to increase the effectiveness and efficiency of the organization to fulfill its goals. Gradual improvements are based on many small changes, in the long run, therefore, continuous monitoring is seen as meta-process to enable continual improvement management systems. Artificial Intelligence and machine learning are impacting many areas of our lives, as one of the key areas, they can also create values in the acceleration of the decision-making process. The digital twin of people powered by those advance technologies makes it possible to simulate the future behavior of worker; to infer and forecast the evolution from their behavior at mid-long term ranges, aiming to reduce the worker's cognitive and physical workload, and increase the worker's well-being such as OSH, job satisfaction, and work-related effect. In the current value stream, it is vital that organizations such as regulatory agencies can have a holistic and transparent information exchange with peer companies or factories. Providers of specific services (e.g., certification bodies) require certified production rates, indoor air quality compliance, environmental footprint, or production-related data tracing like a work-related injury. When a value proposal requires the integration of those different stakeholders, granting access to a public database having immutable but encrypted data is a significant proposal. In model interconnected societies, different stakeholders aim to gather different views of data flows, organizations can consider that approach to attract current and future stakeholders.

- The study proposed people-centric IoT application architectures such as Healthy operator 4.0 (HO4.0) architecture and proved their feasibility through real-world application cases in the industry. The proposed application architecture promotes a new design and engineering philosophy for adaptive production systems where human's physiological, physical, psychological, sensorial, and cognitive capabilities were enhanced in automation and adaptation. The digital twin of humans concept, specifically H04.0 digital twin in the study, came naturally. It envisions high data integration in dynamic work environments, with the human context in the loop. Due to this context-awareness e.g. health awareness, the feeling of safety and confidence is increased and stress can be reduced somehow. It also makes it possible to detect the life-threatening conditions or of no signs or symptoms underlying disease in advance. People can achieve a high quality of working conditions, that improve worker's job satisfaction and well-being. The proposed architectures are not only applicable for workers since they are people-centric oriented, but are also play a guiding role in other societal applications such as healthy buildings and smart aging.
- To enable low-cost and unobtrusive solution data transmission in daily lives, personal smartphones with a set of capabilities such as Bluetooth protocol were applied to improve the connection between people and the IoT. Ergonomic and physiological data streams were collected by Bluetooth, an Android-based app embedded in the smartphone was developed,

making it convenient to integrate the features and enabling a flexible data transmission mechanism connected to people's daily working environment.

- Three long-term monitoring solutions were developed with low-cost IoT devices and successfully adopted into practical usage to continuously collect health-related parameters. The continuous and long-term supportive monitoring provides the possibility to catch temporal and spatial variations of health-related parameters. The rich data sources offer multiple health dimensions, that can reflect a more holistic health digital replica. Individuals' health awareness is enabled in the workplace. The information awareness affects employee performance, e.g., the awareness of working in a good air quality control workplace improves people's productivity. It was possible to realize the high degree of interest that such applications open to workers themselves, and how it becomes a way of discussion between them, as they try to identify patterns and explanations of health impacts over working processes. Working conditions and worker's health tie to business flow. The continuous monitoring solution enables managers to understand working conditions, thus they may take measures e.g., mitigation actions to improve them accordingly. The developed solution also supports data sharing and data transiency vertically and horizontally, which is leveraged by DLT. For instance, OSH regulatory agencies can access indoor environmental quality data with the DLT based data sharing system, so better time windows can be selected to catch exceeding and significant variation for better-ensuring worker's health and safety.
- Different data mining approaches and machine learning methods were investigated and compared to learn health impacts over the process from a group of people. The method chosen was proved to better understand people's behavioral patterns and hidden rules, by the real-world empirical analysis in both a Czech Republic factory and a Spanish factory, as well as at Harvard in the USA. Those derived rules or knowledge can assist in decision making in terms of improving people's health and well-being. When the sample size reaches a certain scale, the population demographics would be more generalized to enable more complex and diverse analyses such as culture affects.
- A data sharing solution was designed in the study, that integrates DLT (IOTA Tangle), semantic modeling to IoT data management, by which data transparency, data interoperability, and data ownership, especially related to sensitive health data, can be implemented under a common-understanding, secure, fee-less, and trust data sharing mechanism. The proposed data sharing and data handling solutions empower data usages on different aspects. For each individual, the data owner can deliver to the beneficiary the requested data. Take the stress data collected from an industrial operator for instance, since the operator themselves as data producer, they can also control the data, they may choose to share with clinical or other relevant organizations. This data sharing solution applied ontology modeling design fosters data reusability by learning processes and growth of models,

as data consumers can select the appropriate meaning of datasets.

To summarize, every defined research objective in the research study was being carried out. The thesis also generalized highly the main contributions to respond to the research question 'How to accelerate and enhance people's health and well-being in the IoT data value chain?'. The value produced by the contributions is reflected on the individual level, organization level and society level in the means of enhancing personal health awareness, releasing stress and increasing performance, detecting life-threatening activities, driving business flows and opportunities, facilitating OSH regulatory check, enabling learning ability from a group of people, and a transparent and trust data management mechanism. To accelerate and enhance people's health and well-being in the IoT data value chain, the study involved data flow and exchange across all IoT architectural layers and IoT data value cycle. The creation of new values and their research implications lie in many societal aspects such as industry 4.0, smart environment, and smart city.

4.1 Implications

To accelerate and enhance people's health and well-being in the IoT data value chain, it involves most of the data processes inside the IoT data value cycle. In the study, the main components of the IoT data value cycle and benefit paradigm includes data processes: data acquisition, data integration, data analysis, and data usage.

- Data acquisition. Several IoT based health monitoring systems were developed to collect health-related data including environmental data, ergonomic data, and physiological parameters. The continuous and long-term supportive monitoring provides the possibility to catch temporal and spatial variations of health-related parameters. The rich data sources offer multiple health dimensions, that can reflect a more holistic health digital replica.
- Data integration. As there are different data lakes used, data heterogeneity would issue data fusion and data integration. Ontology-based syntactic interoperability was explored in the study to make each data source interoperable, integrable, and exchangeable among entities. The applied approach enables multiple health-related data sources interoperable and linked, which is considered as the significant foundation for mining knowledge from data.
- Data analysis. To extract knowledge and rule from collected data sources, diverse machine learning and data mining method were used to derive knowledge from the collected data parameters. Those derived rules or knowledge can assist decision making in terms of improving people's health and well-being such as locating the causes of monitored irregular behaviors. Different prediction methods will help detect potential life-threatening disease and improve workers' welfare and reduce personal and organization healthcare costs. Similar

workers are filtered out to learn physiological and psychological health impacts over a group of people, powered by clustering algorithms. Decision making for better human resources management is enabled via job position re-allocation, organized job training, reconfiguration of the manufacturing system for stress reduction, etc.

- **Data usage.** When a value proposal requires the integration of different stakeholders having different Information Technology (IT) capabilities, a data-sharing platform with immutability and reliability provides a significant approach. The collected health data is uploaded to a secure DLT based public database to enable data transparency and sharing with all relevant stakeholders. As data was generated by people, their data ownership can be ensured with the DLT platform where data producers also can monetize their data with the support of the smart contract. In this way, people are motivated to use health monitoring in their daily lives. The developed DLT based data sharing platform has high relevance for the scope and the objectives in the principle of accountability and the obligation of privacy by design.

The value produced by the contributions is reflected in the individual level, organization level, and society level. The contributions of this dissertation derive from improving the human health and well-being knowledge acquiring and sharing, as well as an exploration of usage of health-related information and knowledge in different contexts (living, working, learning, and commuting), in different environments (residential houses, offices, hospitals, industrial facilities, and education facilities) but also different levels of interest (individual/group, corporate/managerial, regulatory agencies/social).

At a personal level, individuals' health awareness is enabled in the workplace via real-time monitoring solutions. It helps prevent health-threatening activities, detect chronic disease, and minimize personal costs. The continuous data can also be shared with clinics and hospitals for better and accurate health diagnoses or instance companies to recover their healthcare cost. Moreover, user produced data can be monetized via the proposed data-sharing framework. Economic benefits are triggered for any individual, by sharing their live data streams with interested data consumers. In the data market, users can monetize their own generated data set with the support of the smart contract. In this way, people are motivated to use health monitoring in their daily lives.

At the organization level, working conditions and worker's health tie to business flow considering its huge impacts on worker's performance, health, and productivity. The continuous monitoring solution makes it possible for managers to learn working conditions, so they may take measures to better manage the air quality. These parameters are vital as they can be used by companies to develop their employee welfare programs to reduce healthcare costs [60]. Besides, continuous monitoring enables organizations to take health-related conditions as one of the Key Performance Index (KPI) vectors, considering economic loss deriving from a dissatisfy working environment. Furthermore, the link between sustainable management, energy-efficient manage-

ment, and green procurement design is fostered as well by the developed long-term indoor air quality monitoring solution.

The indoor air quality monitoring solution can also empower an air-quality audit management platform, such as keep tracking the environmental footprint. Several societal priorities such as climate change and CO₂ emissions-related issues can be addressed thanks to the developed air-quality monitoring solution. Besides, continuous air-quality monitoring is linked to new demanding societal needs, as it is pouring values to healing built-environment in an aging society, cultural heritage protection, and economic pressure pressing to the environment.

For organizations and business associations, the core technology DLT not only supports health data sharing, thanks to its salient metrics such as immutability, traceability, and reliability, the developed platform can also potentially be used to realize full visibility and effective coordination of global trade such as supply chains. At this critical moment, the coronavirus disease (COVID-2019) pandemic shakes the world. The sudden explosion and uncontrolled world-wide spread exposed the limitations of handling public health emergencies. The DLT based solution such as IOTA and blockchain, would helpful for governments and society in aspects of fast-tracking drug delivery, tracking citizens' phones with privacy assurance to curb the spread of the COVID-2019.

At the society level, there is a challenge in finding economical and non-intrusive solutions to monitor health-related parameters, especially in developing countries [121]. Environment Pollution and Climate Change have become a global concern. A great number of people are working and living in a polluted environment. The WHO recently reported that nine out of 10 people breathe polluted air. While the cost of pollution monitoring is always a vital topic, to reinforce scalability, low-cost would be on the promise. A low-cost and unobtrusive solution for estimate personal air pollution exposure (PAPE) was proposed employing indoor air quality monitors, personal phones, and e-beacon technology. The PAPE data is helpful for air quality policy-making, climate change, and epidemiological study. For example, the interest in understanding air pollution effects on various cardiovascular and respiratory diseases. Furthermore, the quantified research on stress relief benefiting from urban to urban nature transitions study, which would assist the urban planning decision-making process to achieve a better well-being for urban residents. Organized and structured urban green space can be planned accordingly to reduce urban dwellers' accumulated stress in the rapid urbanization process. The study findings can be used indoor as well. Many studies show that connecting people to nature is an important aspect of wellness, especially for children and patients. Spending time in green spaces is beneficial for mental and physical health, yet not everyone has equal access to such spaces. With the Virtual Reality and Augment Reality technology exponential development, recorded green space with the immersive experiences is promising to restore worker's attention and concentration level, as indicated by Kaplan's Attention Restoration Theory (ART) [150].

4.2 Limitations

Some conditions and influences that place restrictions on the research design/methodology, the findings of this study have to be seen in the light of some limitations.

- Due to sensor availability and safety regulations, the applied sensors are limited. It may affect building a worker's health digital twin under an intact digital replica. There is a lack of sensors for certain purposes, such as human concentration.
- Different sensors were used to measure the same parameter such as indoor air quality, there was an observed variance among the different devices. The distinction can be attributed to device calibration and the measurement techniques used, which should be acknowledged as potential 'selection bias'.
- When referring to people's health and well-being, cultural backgrounds can also have an impact. It has been clearly demonstrated that cultural differences do exist and these preferences may be found at the organizational or individual level. Some ways of measuring cultural values were studied, but considering its complexity and time constraint, the culture dimension was not analyzed in the study.
- Another limitation of this pilot study is that a public IOTA node was used for publishing and receiving data over the Tangle. The lagging time varied depending on the workload of that public node, in practical implementation, a private node should be set up according to the practical requirements.
- Because of time constraints, the integration between healthy operator's health and smart monitoring is at the initial stage. When data size reaches a big scale, advance big data analytic can be used to discover more complex rules or patterns from data.

4.3 Future research

The future of work is now. The IoT connects any "thing" that generates data without stopping, massive data is generated each second from IoT devices (sensors, actuators). The business advantages with IoT comes to Big data era. Big data in IoT will demand more robust and scalable platforms, finer analytical tools, and more powerful data storage systems.

With data-driven AI technology booming, big data analytic was come out, where advanced technology such as machine learning and deep learning, Hadoop, and Spark can be employed to dig knowledge out of data and predicate future out of history data [86]. IoT big data analytic aims to assist business associations and organizations to achieve an improved understanding of data to make efficient and effective decisions [87]. The next step work is to explore big data analytic platforms, that can uncover additional value from the monitoring health data with efficiency

and effectiveness. With a wider variety of sensors are integrated, a holographic digital twin of humans is built to empower people's health and well-being as a whole. Increasing data sources are fused into the analytical platform to foster diverse applications in real-world scenarios.

Machine/deep learning is an effective tool to dig knowledge from massive data. Human activity recognition (HAR) is one of the most important prerequisites for health-related applications such as smart factories, remote health monitoring, assistant living, and green building energy management. Deep learning techniques showed their merits on obtaining high-quality activity types from raw sensor data. Sensor-based activity recognition approaches were categorized into two main classes [151]: Data-Driven Approaches (DDA) and Knowledge-Driven Approaches (KDA). The DDA approaches are machine learning basis requiring preexistent large datasets for building an activity model. They are capable to deal with uncertainty and dynamic situations through training and learning processes. Deep learning has shown surpassing performance in HAR. The deep learning approaches: Convolutional Neural Network (CNN) can be used to build classification models, powered by TensorFlow. The bringing value stream and benefits derived from big data and machine/deep learning can be better enhanced and extended to be poured into future state for individual, organization, and society, toward the next high-quality society 5.0.

Bibliography

- [1] IDC. The digitization of the world: From edge to core. Available online: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf> (Accessed on 06 May 2020).
- [2] Matthias Steup and Ram Neta. Epistemology. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, spring 2020 edition, 2020.
- [3] L Srivastava. Pervasive, ambient, ubiquitous: the magic of radio. *In: Proceedings of European Commission Conference: From RFID to the Internet of Things*, 2006.
- [4] Irlán Grangel-González, Lavdim Halilaj, Gökhan Coskun, Sören Auer, Diego Collarana, and Michael Hofmeister. Towards a Semantic Administrative Shell for Industry 4.0 Components. *In 2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pages 230–237, 2016.
- [5] Gartner. Gartner identifies top 10 strategic iot technologies and trends. Available online: <https://gtnr.it/2JLiEbA> (Accessed on 06 May 2020).
- [6] Michael E. Porter. *Competitive strategy: Techniques for analyzing industries and competitors*. Free Press, 1980.
- [7] European Commission DG CONNECT. A european strategy on the data value chain. Available online: ec.europa.eu/information_society/newsroom/cf/dae/document.cfm?doc_id=34881 (Accessed on 06 May 2020).
- [8] GSMA. The data value chain. Available online: https://www.gsma.com/publicpolicy/wp-content/uploads/2018/07/GSMA_Data_Value_Chain_June_2018.pdf (Accessed on 06 May 2020).
- [9] H2020 CREATE-IoT Project. Iot data value chain model. Available online: https://european-iot-pilots.eu/wp-content/uploads/2017/10/D05_03_WP05_H2020_CREATE-IoT_Final.pdf (Accessed on 06 May 2020).

- [10] Mayumi Fukuyama. Society 5.0: Aiming for a new human-centered society. Available online: https://www.jef.or.jp/journal/pdf/220th_Special_Article_02.pdf (Accessed on 06 May 2020).
- [11] Werner Kritzinger, Matthias Karner, Georg Traar, Jan Henjes, and Wilfried Sihm. Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11):1016 – 1022, 2018.
- [12] A. El Saddik. Digital twins: The convergence of multimedia technologies. *IEEE MultiMedia*, 25(2):87–92, 2018.
- [13] Martin Bauer, Mathieu Boussard, Nicola Bui, Francois Carrez, Christine (SIEMENS, Jourik (ALUBE, Carsten (SAP, Stefan Meissner, Andreas IML, Alexis Olivereau, Matthias (SAP, Walewski Joachim, Julinda Stefa, and Alexander Salinas. Internet of things – architecture iot-a deliverable d1.5 – final architectural reference model for the iot v3.0. *Technical Report*, 07 2013.
- [14] Mehedi Masud Omar Said. Towards internet of things: Survey and future vision. *International Journal of Computer Networks*, 5(1):1 – 17, 2013.
- [15] Miao Wu, Ting-Jie Lu, Fei-Yang Ling, Jing Sun, and Hui-Ying Du. Research on the architecture of internet of things. In *2010 3rd International Conference on Advanced Computer Theory and Engineering(ICAETE)*, volume 5, pages V5–484–V5–487, Aug 2010.
- [16] R. Khan, S. U. Khan, R. Zaheer, and S. Khan. Future internet: The internet of things architecture, possible applications and key challenges. In *2012 10th International Conference on Frontiers of Information Technology*, pages 257–260, 2012.
- [17] Jayavardhana Gubbi, Rajkumar Buyya, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems*, 29(7):1645 – 1660, 2013.
- [18] Pengfei Hu, Sahraoui Dhelim, Huansheng Ning, and Tie Qiu. Survey on fog computing: architecture, key technologies, applications and open issues. *Journal of Network and Computer Applications*, 98:27 – 42, 2017.
- [19] Hong-Ning Dai, Zibin Zheng, and Yan Zhang. Blockchain for internet of things: A survey. *IEEE Internet of Things Journal*, 6(5):8076–8094, Oct 2019.
- [20] Marco A. Calderon, Saul E. Delgadillo, and J. Antonio Garcia-Macias. A more human-centric internet of things with temporal and spatial context. *Procedia Computer Science*, 83:553 – 559, 2016.
- [21] J. S. Silva, P. Zhang, T. Pering, F. Boavida, T. Hara, and N. C. Liebau. People-centric internet of things. *IEEE Communications Magazine*, 55(2):18–19, 2017.

- [22] Industry 4.0. The 4th industrial revolution is here. Available online: <http://catalonia.com/trade-with-catalonia/industry-40.jsp> (Accessed on 06 May 2020).
- [23] RAMI 4.0. Plattform industrie 4.0: The reference architecture model of industrie 4.0 (2015). Available online: <https://www.zvei.org/en/subjects/industry-4-0/the-reference-architectural-model-rami-40-and-the-industrie-40-component/>. (Accessed on 06 May 2020).
- [24] Christian Manzei, Ronald Heinze, and Linus Schleupner. *Industrie 4.0 im internationalen Kontext: Kernkonzepte, Ergebnisse, Trends*. VDE-Verlag, 2016.
- [25] F. Zezulka, P. Marcon, I. Vesely, and O. Sajdl. Industry 4.0 – an introduction in the phenomenon. *IFAC-PapersOnLine*, 49(25):8 – 12, 2016.
- [26] Jay Lee, Behrad Bagheri, and Hung-An Kao. A cyber-physical systems architecture for industry 4.0-based manufacturing systems. *Manufacturing Letters*, 3:18 – 23, 2015.
- [27] Xiang T. R. Kong, Hao Luo, George Q. Huang, and Xuan Yang. Industrial wearable system: the human-centric empowering technology in Industry 4.0. *Journal of Intelligent Manufacturing*, 30(8):2853–2869, 2019.
- [28] J. Miranda, N. Mäkitalo, J. Garcia-Alonso, J. Berrocal, T. Mikkonen, C. Canal, and J. M. Murillo. From the internet of things to the internet of people. *IEEE Internet Computing*, 19(2):40–47, 2015.
- [29] V. Petrov, K. Mikhaylov, D. Moltchanov, S. Andreev, G. Fodor, J. Torsner, H. Yanikomeroglu, M. Juntti, and Y. Koucheryavy. When iot keeps people in the loop: A path towards a new global utility. *IEEE Communications Magazine*, 57(1):114–121, 2019.
- [30] Fernando Boavida, Andreas Kliem, Thomas Renner, Jukka Riekk, Christophe Jouvray, Michal Jacovi, Stepan Ivanov, Fiorella Guadagni, Paulo Gil, and Alicia Triviño. People-centric internet of things - challenges, approach, and enabling technologies. In *IDC*, volume 616, pages 463–474. Springer, 2015.
- [31] S. Feng, P. Setoodeh, and S. Haykin. Smart home: Cognitive interactive people-centric internet of things. *IEEE Communications Magazine*, 55(2):34–39, 2017.
- [32] C. Martella, M. Cattani, and M. van Steen. Exploiting density to track human behavior in crowded environments. *IEEE Communications Magazine*, 55(2):48–54, 2017.
- [33] R. Parada and J. Melia-Segui. Gesture detection using passive rfid tags to enable people-centric iot applications. *IEEE Communications Magazine*, 55(2):56–61, 2017.
- [34] Prasant Misra, Yogesh Simmhan, and Jay Warrior. Towards a practical architecture for the next generation internet of things, 2015.

- [35] NIOSH. Niosh pocket guide to chemical hazards. Available online: <https://www.cdc.gov/niosh/npg/pgintrod.html> (Accessed on 06 May 2020).
- [36] Shengjing Sun, Xiaochen Zheng, Javier Villalba-Díez, and Joaquín Ordieres-Meré. Indoor air-quality data-monitoring system: Long-term monitoring benefits. *Sensors*, 19(19):4157, 2019.
- [37] EU-OSHA. European agency for safety and health at work. Available online: https://europa.eu/european-union/about-eu/agencies/eu-osha_en (Accessed on 06 May 2020).
- [38] PEROSH. The partnership for european research in occupational safety and health. Available online: <https://perosh.eu/about-perosh/the-partnership-for-european-research-in-occupational-safety-and-health-perosh/> (Accessed on 06 May 2020).
- [39] Adel Badri, Bryan Boudreau-Trudel, and Ahmed Saâdeddine Souissi. Occupational health and safety in the industry 4.0 era: A cause for major concern? *Safety Science*, 109:403 – 411, 2018.
- [40] Rui Pitarma, Gonçalo Marques, and Filipe Caetano. Monitoring indoor air quality to improve occupational health. In Álvaro Rocha, Ana Maria Correia, Hojjat Adeli, Luis Paulo Reis, and Marcelo Mendonça Teixeira, editors, *New Advances in Information Systems and Technologies*, pages 13–21, Cham, 2016. Springer International Publishing.
- [41] Sara Zanni, Francesco Lalli, Eleonora Foschi, Alessandra Bonoli, and Luca Mantecchini. Indoor air quality real-time monitoring in airport terminal areas: An opportunity for sustainable management of micro-climatic parameters. *Sensors*, 18(11), 2018.
- [42] Nurul Azma Zakaria, Zaheera Zainal Abidin, Norharyati Harum, Low Chen Hau, Nabeel Saleh Ali, and Fairul Azni Jafar. Wireless internet of things-based air quality device for smart pollution monitoring. *International Journal of Advanced Computer Science and Applications*, 9(11), 2018.
- [43] Mohieddine Benammar, Abderrazak Abdaoui, Sabbir H.M. Ahmad, Farid Touati, and Abdullah Kadri. A modular iot platform for real-time indoor air quality monitoring. *Sensors*, 18(2), 2018.
- [44] O. A. Postolache, J. M. Dias Pereira, and P. M. B. Silva Girao. Smart sensors network for air quality monitoring applications. *IEEE Transactions on Instrumentation and Measurement*, 58(9):3253–3262, 2009.
- [45] S. Bhattacharya, S. Sridevi, and R. Pitchiah. Indoor air quality monitoring using wireless sensor network. In *2012 Sixth International Conference on Sensing Technology (ICST)*, pages 422–427, 2012.

- [46] Yun Cheng, Xiucheng Li, Zhijun Li, Shouxu Jiang, Yilong Li, Ji Jia, and Xiaofan Jiang. Airccloud: A cloud-based air-quality monitoring system for everyone. In *Proceedings of the 12th ACM Conference on Embedded Network Sensor Systems*, SenSys '14, page 251–265. Association for Computing Machinery, 2014.
- [47] J. Kim, C. Chu, and S. Shin. Issaq: An integrated sensing systems for real-time indoor air quality monitoring. *IEEE Sensors Journal*, 14(12):4230–4244, 2014.
- [48] Yifei Jiang, Kun Li, Lei Tian, Ricardo Piedrahita, Xiang Yun, Omkar Mansata, Qin Lv, Robert P. Dick, Michael Hannigan, and Li Shang. Maqs: a mobile sensing system for indoor air quality. In *UbiComp '11*, 2011.
- [49] Y. Xiang, R. Piedrahita, R. P. Dick, M. Hannigan, Q. Lv, and L. Shang. A hybrid sensor system for indoor air quality monitoring. In *2013 IEEE International Conference on Distributed Computing in Sensor Systems*, pages 96–104, 2013.
- [50] Jagriti Saini, Maitreyee Dutta, and Gonalo Marques. A comprehensive review on indoor air quality monitoring systems for enhanced public health. *Sustainable Environment Research*, 30, 12 2020.
- [51] A. Nag, S. C. Mukhopadhyay, and J. Kosel. Wearable flexible sensors: A review. *IEEE Sensors Journal*, 17(13):3949–3960, July 2017.
- [52] J. Edwards. Wireless sensors relay medical insight to patients and caregivers [special reports]. *IEEE Signal Processing Magazine*, 29(3):8–12, May 2012.
- [53] K. Malhi, S. C. Mukhopadhyay, J. Schnepfer, M. Haefke, and H. Ewald. A zigbee-based wearable physiological parameters monitoring system. *IEEE Sensors Journal*, 12(3):423–430, March 2012.
- [54] Futuresource. Medical-grade technology brings renewed excitement to the wearables market. Available online: <https://www.futuresource-consulting.com/blog/posts/2018/june/medical-grade-technology-brings-renewed-excitement-to-the-wearables-market/> (Accessed on 06 May 2020).
- [55] Christos Stergiou, Kostas E Psannis, Byung-Gyu Kim, and Brij Gupta. Secure integration of iot and cloud computing. *Future Generation Computer Systems*, 78:964–975, 2018.
- [56] Francesco Longo, Letizia Nicoletti, and Antonio Padovano. Smart operators in industry 4.0: A human-centered approach to enhance operators' capabilities and competencies within the new smart factory context. *Computers & industrial engineering*, 113:144–159, 2017.
- [57] C. Scheuermann, F. Heinz, B. Bruegge, and S. Verclas. Real-time support during a logistic process using smart gloves. In *Smart SysTech 2017; European Conference on Smart Objects, Systems and Technologies*, pages 1–8, June 2017.

- [58] Yuqiuge Hao and Petri Helo. The role of wearable devices in meeting the needs of cloud manufacturing: A case study. *Robotics and Computer-Integrated Manufacturing*, 45:168 – 179, 2017. Special Issue on Ubiquitous Manufacturing (UbiM).
- [59] X. T. R. Kong, X. Yang, G. Q. Huang, and H. Luo. The impact of industrial wearable system on industry 4.0. In *2018 IEEE 15th International Conference on Networking, Sensing and Control (ICNSC)*, pages 1–6, March 2018.
- [60] Research and Markets. Industrial wearable devices market: Global industry trends, share, size, growth, opportunity and forecast 2019-2024. Available online: <https://www.researchandmarkets.com/reports/4775698/industrial-wearable-devices-market-global?> (Accessed on 06 May 2020).
- [61] Mario Aehmelt and Bodo Urban. Follow-me: Smartwatch assistance on the shop floor. In Fiona Fui-Hoon Nah, editor, *HCI in Business*, pages 279–287. Springer International Publishing, 2014.
- [62] Ji Zhou, Yanhong Zhou, Baicun Wang, and Jiyuan Zang. Human-cyber-physical systems (hcpss) in the context of new-generation intelligent manufacturing. *Engineering*, 5(4):624 – 636, 2019.
- [63] Daniel Podgorski, Katarzyna Majchrzycka, Anna Dąbrowska, Grzegorz Gralewicz, and Małgorzata Okrasa. Towards a conceptual framework of osh risk management in smart working environments based on smart ppe, ambient intelligence and the internet of things technologies. *International Journal of Occupational Safety and Ergonomics*, 23(1):1–20, 2017.
- [64] David Romero, Johan Stahre, Thorsten Wuest, Ovidiu Noran, Peter Bernus, Åsa Fast-Berglund, and Dominic Gorecky. Towards an operator 4.0 typology: a human-centric perspective on the fourth industrial revolution technologies. In *International conference on computers and industrial engineering (CIE46) proceedings*, 2016.
- [65] David Romero, Sandra Mattsson, Åsa Fast-Berglund, Thorsten Wuest, Dominic Gorecky, and Johan Stahre. Digitalizing occupational health, safety and productivity for the operator 4.0. In *IFIP International Conference on Advances in Production Management Systems*, pages 473–481. Springer, 2018.
- [66] Shengjing Sun, Xiaochen Zheng, Bing Gong, Jorge García Paredes, and Joaquín Ordieres-Meré. Healthy operator 4.0: A human cyber-physical system architecture for smart workplaces. *Sensors*, 20(7), 2020.
- [67] Ibrahim Sadek, Shafiq Ul Rehman, Josué Codjo, and Bessam Abdulrazak. Privacy and security of iot based healthcare systems: Concerns, solutions, and recommendations. In José Pagán, Mounir Mokhtari, Hamdi Aloulou, Bessam Abdulrazak, and María Fernanda

- Cabrera, editors, *How AI Impacts Urban Living and Public Health*, pages 3–17, Cham, 2019. Springer International Publishing.
- [68] Ali Dorri, Salil Kanhere, and Raja Jurdak. Blockchain in internet of things: Challenges and solutions. *arxiv*, 08 2016.
- [69] Ana Reyna, Cristian Martín, Jaime Chen, Enrique Soler, and Manuel Díaz. On blockchain and its integration with iot. challenges and opportunities. *Future Generation Computer Systems*, 88:173 – 190, 2018.
- [70] O. Novo. Blockchain meets iot: An architecture for scalable access management in iot. *IEEE Internet of Things Journal*, 5(2):1184–1195, April 2018.
- [71] Jessica Groopman. Circles of trust for smart devices for us citizens. source: Privacy & the internet of things: The importance of transparency in accounting for what we can’t see. Available online: <https://www.trustarc.com/blog/tag/iot-summit/> (Accessed on 06 May 2020).
- [72] Satoshi Nakamoto. Bitcoin: A peer-to-peer electronic cash system. Available online: <http://www.bitcoin.org/bitcoin.pdf> (Accessed on 06 May 2020).
- [73] A. Sherly Alphonse and M.S. Starvin. Chapter 12 - blockchain and internet of things: An overview. In Saravanan Krishnan, Valentina E. Balas, E. Golden Julie, Y. Harold Robinson, S. Balaji, and Raghvendra Kumar, editors, *Handbook of Research on Blockchain Technology*, pages 295 – 322. Academic Press, 2020.
- [74] S. Huh, S. Cho, and S. Kim. Managing iot devices using blockchain platform. In *2017 19th International Conference on Advanced Communication Technology (ICACT)*, pages 464–467, Feb 2017.
- [75] W. Viriyasitavat, L. Da Xu, Z. Bi, and A. Sapsomboon. New blockchain-based architecture for service interoperations in internet of things. *IEEE Transactions on Computational Social Systems*, 6(4):739–748, Aug 2019.
- [76] Hossein Shafagh, Lukas Burkhalter, Anwar Hithnawi, and Simon Duquennoy. Towards blockchain-based auditable storage and sharing of iot data. In *Proceedings of the 2017 on Cloud Computing Security Workshop*, page 45–50, New York, NY, USA, 2017. Association for Computing Machinery.
- [77] Alfonso Panarello, Nachiket Tapas, Giovanni Merlino, Francesco Longo, and Antonio Puliato. Blockchain and iot integration: A systematic survey. *Sensors*, 18(8), 2018.
- [78] Emma Bell, Alan Bryman, and Bill Harley. *Business research methods*. Oxford university press, 2018.

- [79] M. Aazam, I. Khan, A. A. Alsaffar, and E. Huh. Cloud of things: Integrating internet of things and cloud computing and the issues involved. In *Proceedings of 2014 11th International Bhurban Conference on Applied Sciences Technology (IBCAST) Islamabad, Pakistan, 14th - 18th January, 2014*, pages 414–419, Jan 2014.
- [80] Y. Xu and A. Helal. Scalable cloud–sensor architecture for the internet of things. *IEEE Internet of Things Journal*, 3(3):285–298, June 2016.
- [81] J. Rui and S. Danpeng. Architecture design of the internet of things based on cloud computing. In *2015 Seventh International Conference on Measuring Technology and Mechatronics Automation*, pages 206–209, June 2015.
- [82] Jiehan Zhou, T. Leppanen, E. Harjula, M. Ylianttila, T. Ojala, Chen Yu, Hai Jin, and L. T. Yang. Cloudthings: A common architecture for integrating the internet of things with cloud computing. In *Proceedings of the 2013 IEEE 17th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*, pages 651–657, June 2013.
- [83] Shanhe Yi, Cheng Li, and Qun Li. A survey of fog computing: Concepts, applications and issues. In *Proceedings of the 2015 Workshop on Mobile Big Data*, page 37–42. Association for Computing Machinery, 2015.
- [84] I. Stojmenovic and S. Wen. The fog computing paradigm: Scenarios and security issues. In *2014 Federated Conference on Computer Science and Information Systems*, pages 1–8, Sep. 2014.
- [85] Alexandros Labrinidis and H. V. Jagadish. Challenges and opportunities with big data. *Proc. VLDB Endow.*, 5(12):2032–2033, August 2012.
- [86] M. Marjani, F. Nasaruddin, A. Gani, A. Karim, I. A. T. Hashem, A. Siddiqua, and I. Yaqoob. Big iot data analytics: Architecture, opportunities, and open research challenges. *IEEE Access*, 5:5247–5261, 2017.
- [87] Chun-Wei Tsai, Chin-Feng Lai, Han-Chieh Chao, and Athanasios V. Vasilakos. Big data analytics: a survey. *Journal of Big Data*, 2:1–32, 2015.
- [88] A. F. Skarmeta, J. L. Hernández-Ramos, and M. V. Moreno. A decentralized approach for security and privacy challenges in the internet of things. In *2014 IEEE World Forum on Internet of Things (WF-IoT)*, pages 67–72, March 2014.
- [89] Hannes Gross, Marko Hölbl, Daniel Slamanig, and Raphael Spreitzer. Privacy-aware authentication in the internet of things. In Michael Reiter and David Naccache, editors, *Cryptology and Network Security*, pages 32–39, Cham, 2015. Springer International Publishing.
- [90] Yang Xiao, Ning Zhang, Wenjing Lou, and Yiwei Thomas Hou. A survey of distributed consensus protocols for blockchain networks. *ArXiv*, abs/1904.04098, 2019.

- [91] Yongmin Zhang, S. He, and J. Chen. Data gathering optimization by dynamic sensing and routing in rechargeable sensor networks. In *2013 IEEE International Conference on Sensing, Communications and Networking (SECON)*, pages 273–281, June 2013.
- [92] T. Arampatzis, J. Lygeros, and S. Manesis. A survey of applications of wireless sensors and wireless sensor networks. In *Proceedings of the 2005 IEEE International Symposium on, Mediterrean Conference on Control and Automation Intelligent Control, 2005.*, pages 719–724, June 2005.
- [93] Masanori Ando. Recent advances in optochemical sensors for the detection of h₂, o₂, o₃, co, co₂ and h₂o in air. *TrAC Trends in Analytical Chemistry*, 25(10):937 – 948, 2006.
- [94] Gregory D. Abowd, Anind K. Dey, Peter J. Brown, Nigel Davies, Mark Smith, and Pete Steggles. Towards a better understanding of context and context-awareness. In Hans-W. Gellersen, editor, *Handheld and Ubiquitous Computing*, pages 304–307. Springer Berlin Heidelberg, 1999.
- [95] N. Kamarulzaman, A.A. Saleh, S.Z. Hashim, H. Hashim, and A.A. Abdul-Ghani. An overview of the influence of physical office environments towards employee. *Procedia Engineering*, 20:262 – 268, 2011.
- [96] Olli Seppanen. Some quantitative relations between indoor environmental quality and work performance or health. *HVAC&R Research*, 12:957 – 973, 10 2006.
- [97] Bahador Khaleghi, Alaa Khamis, Fakhreddine O. Karray, and Saiedeh N. Razavi. Multisensor data fusion: A review of the state-of-the-art. *Information Fusion*, 14(1):28 – 44, 2013.
- [98] Ramani Kannan, Syed Saad Azhar Ali, Abdulrehman Farah, Syed Hasan Adil, and Amjad Khan. Smart wearable eeg sensor. *Procedia Computer Science*, 105:138 – 143, 2017. 2016 IEEE International Symposium on Robotics and Intelligent Sensors, IRIS 2016, 17-20 December 2016, Tokyo, Japan.
- [99] Luigi Atzori, Antonio Iera, and Giacomo Morabito. The internet of things: A survey. *Computer Networks*, 54(15):2787 – 2805, 2010.
- [100] Thomas R. Gruber. Toward principles for the design of ontologies used for knowledge sharing? *International Journal of Human-Computer Studies*, 43(5):907 – 928, 1995.
- [101] Miguel-Angel Sicilia. Metadata, semantics, and ontology: Providing meaning to information resources. *Int. J. Metadata Semant. Ontologies*, 1(1):83–86, January 2006.
- [102] Truong Duy, Gerald Quirchmayr, A Min Tjoa, and Hanh Hoang. A semantic data model for the interpretation of environmental streaming data. *2017 Seventh International Conference on Information Science and Technology (ICIST)*, pages 376–380, 04 2017.

- [103] Barry Smith, Pierre Grenon, and Louis Goldberg. Biodynamic ontology: Applying bfo in the biomedical domain. *Studies in Health and Technology Informatics*, 102:20–38, 2004.
- [104] Aldo Gangemi, Nicola Guarino, Claudio Masolo, Alessandro Oltramari, and Luc Schneider. Sweetening ontologies with dolce. In Asunción Gómez-Pérez and V. Richard Benjamins, editors, *Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web*, pages 166–181. Springer Berlin Heidelberg, 2002.
- [105] Albert Goldfain, Barry Smith, Sivaram Arabandi, Mathias Brochhausen, and William Hogan. Vital sign ontology. In *The 14th Annual Bio-Ontologies Meeting; 15-16th July*, 01 2011.
- [106] Jude A. Adeleke and Deshendra Moodley. An ontology for proactive indoor environmental quality monitoring and control. In *Proceedings of the 2015 Annual Research Conference on South African Institute of Computer Scientists and Information Technologists*. Association for Computing Machinery, 2015.
- [107] L. Zhang, L. Zhang, and B. Du. Deep learning for remote sensing data: A technical tutorial on the state of the art. *IEEE Geoscience and Remote Sensing Magazine*, 4(2):22–40, 2016.
- [108] Machnation. Distributed ledger technology (dlt): 7 benefits for iot applications. Available online: <https://www.machnation.com/2019/03/13/distributed-ledger-technology-dlt-7-benefits-for-iot-applications/> (Accessed on 06 May 2020).
- [109] James Brogan, Immanuel Baskaran, and Navin Ramachandran. Authenticating health activity data using distributed ledger technologies. *Computational and Structural Biotechnology Journal*, 16:257–266, 2018.
- [110] Yang Xiao, Ning Zhang, Wenjing Lou, and Y. Thomas Hou. A survey of distributed consensus protocols for blockchain networks. *IEEE Communications Surveys & Tutorials*, page 1–1, 2020.
- [111] Xiaochen Zheng, Shengjing Sun, Raghava Rao Mukkamala, Ravi Vatrapu, and Joaquín Ordieres-Meré. Accelerating health data sharing: A solution based on the internet of things and distributed ledger technologies. *J Med Internet Res*, 21(6):e13583, Jun 2019.
- [112] HRBPData. Wearable smart band (hband). Available online: https://drive.google.com/file/d/1lh8M3xXWWZEA2DexDjipUIM_b0uMvjTQ/view (Accessed on 06 May 2020).
- [113] JING-MEME. Wearable eyewear. Available online: <https://jins-meme.com/en/> (Accessed on 06 May 2020).
- [114] Circulate Health Science and Technology. Smart air quality monitor suit. Available online: <https://bjsklt.en.alibaba.com/> (Accessed on 06 May 2020).

- [115] Shimmer Sensing. Consensys ecg and gsr development kits. Available online: <http://www.shimmersensing.com/> (Accessed on 06 May 2020).
- [116] ATMOUBE. Wearable air pollution tracker. Available online: <https://atmotube.com/?view=es> (Accessed on 06 May 2020).
- [117] Foobot. Air quality monitor by foobot. Available online: <https://foobot.io/> (Accessed on 06 May 2020).
- [118] Blukii. Smart beacon. Available online: <https://www.blukii.com/en/hardware/smart-beacon> (Accessed on 06 May 2020).
- [119] OwnTracks. Owntracks allows you to keep track of your own location. Available online: <https://github.com/owntracks/owntracks> (Accessed on 06 May 2020).
- [120] Shengjing Sun. Available online: <https://gitlab.com/sunshengjing/> (Accessed on 06 May 2020).
- [121] Keith April G. Arano, Shengjing Sun, Joaquin Ordieres-Mere, Gong, and Bing. The use of the internet of things for estimating personal pollution exposure. *International Journal of Environmental Research and Public Health*, 16(17), 2019.
- [122] Shengjing Sun. Interoperable data handling for industry 4.0. Available online: <https://gitlab.com/sunshengjing/interoperable-data-usage-for-industry-4.0> (Accessed on 06 May 2020).
- [123] P T. Stone. A model for the explanation of discomfort and pain in the eye caused by light. *Lighting Research & Technology*, 41:109–121, 05 2009.
- [124] J M Fugate. Physiological basis for discomfort glare. *Am. J. Optom. Arch. Am. Acad. Optom.*, 34(7):377–387, July 1957.
- [125] Li Lan, Pawel Wargocki, and Zhiwei Lian. Thermal effects on human performance in office environment measured by integrating task speed and accuracy. *Applied Ergonomics*, 45(3):490 – 495, 2014.
- [126] A V Arundel, E M Sterling, J H Biggin, and T D Sterling. Indirect health effects of relative humidity in indoor environments. *Environ. Health Perspect.*, 65:351–361, March 1986.
- [127] EPA. Epa air quality standards. Available online: <https://www.epa.gov> (Accessed on 06 May 2020).
- [128] CBRE. Available online: [url={https://www.cbre.com/report-download?PUBID=1ad503c5-8566-4f7f-bce8-3ea7121c75f4}](https://www.cbre.com/report-download?PUBID=1ad503c5-8566-4f7f-bce8-3ea7121c75f4) (Accessed on 06 May 2020).
- [129] T.; Zhao J Welsh, B.; Baird and D Block-Schachter. Web app design to implement travel behavioral nudging using "moves". In *Proceedings of the Transportation Research Board 93rd Annual Meeting*, 2014.

- [130] Adel Farhi, Valentina Boyko, Jonatan Almagor, Itzhak Benenson, Enrico Segre, Yinon Rudich, Eli Stern, and Liat Lerner-Geva. The possible association between exposure to air pollution and the risk for congenital malformations. *Environmental research*, 135:173–80, 2014.
- [131] Sean D. Beevers, Nutthida Kitwiroon, Martin L. Williams, and David C. Carslaw. One way coupling of cmaq and a road source dispersion model for fine scale air pollution predictions. *Atmospheric Environment*, 59:47 – 58, 2012.
- [132] Bing Gong and Joaquín Ordieres-Meré. Prediction of daily maximum ozone threshold exceedances by preprocessing and ensemble artificial intelligence techniques: Case study of hong kong. *Environmental Modelling & Software*, 84:290 – 303, 2016.
- [133] Jason G. Su, Michael Jerrett, Ying-Ying Meng, Melissa Pickett, and Beate Ritz. Integrating smart-phone based momentary location tracking with fixed site air quality monitoring for personal exposure assessment. *Science of The Total Environment*, 506-507:518 – 526, 2015.
- [134] Nicola Macoir, Jan Bauwens, Bart Jooris, Ben Van Herbruggen, Jen Rossey, Jeroen Hoebeke, and Eli De Poorter. Uwb localization with battery-powered wireless backbone for drone-based inventory management. *Sensors*, 19(3), 2019.
- [135] CASBEE Tools. Casbee for buildings. Available online: <http://www.ibec.or.jp/CASBEE/english/downloadE.htm> (Accessed on 06 May 2020).
- [136] DePaul Research Team. The chronic fatigue syndrome attitudes test questions. Available online: <https://condor.depaul.edu/ljason/cfs/measures.html> (Accessed on 06 May 2020).
- [137] Human Performance Group at NASA’s Ames Research Center. Nasa-tlx. Available online: <https://en.wikipedia.org/wiki/NASA-TLX> (Accessed on 06 May 2020).
- [138] MRC Cognition and Brain Sciences Unit. Using cognitive theory and innovations in neuroscience to understand and improve mental wellbeing across the lifespan. Available online: <http://www.mrc-cbu.cam.ac.uk/> (Accessed on 06 May 2020).
- [139] Ali Balaid. A comprehensive review of knowledge mapping techniques. *Journal of Information Systems Research and Innovation*, 3:71–76, 01 2012.
- [140] PsyToolkit. Cognitive-psychological experiments and survey. Available online: <https://www.psychtoolkit.org/> (Accessed on 06 May 2020).
- [141] Julia [Ayuso Sanchez], Toshiharu Ikaga, and Sergio [Vega Sanchez]. Quantitative improvement in workplace performance through biophilic design: A pilot experiment case study. *Energy and Buildings*, 177:316 – 328, 2018.

- [142] Jie Yin, Shihao Zhu, Piers MacNaughton, Joseph G. Allen, and John D. Spengler. Physiological and cognitive performance of exposure to biophilic indoor environment. *Building and Environment*, 132:255 – 262, 2018.
- [143] Miguel Ángel Antón, Joaquín Ordieres-Meré, Unai Saralegui, and Shengjing Sun. Non-invasive ambient intelligence in real life: Dealing with noisy patterns to help older people. *Sensors*, 19(14), 2019.
- [144] Javier Villalba-Díez, Martin Molina, Joaquín Ordieres-Meré, Shengjing Sun, Daniel Schmidt, and Wanja Wellbrock. Geometric deep lean learning: Deep learning in industry 4.0 cyber–physical complex networks. *Sensors*, 20(3), 2020.
- [145] Shengjing Sun, Xiaochen Zheng, Javier Villalba-Díez, and Joaquín Ordieres-Meré. Data handling in industry 4.0: Interoperability based on distributed ledger technology. *Sensors*, 20(11), 2020.
- [146] Shengjing Sun, Joaquín Ordieres-Meré, Yin Jie, and Xiaochen Zheng. Biophilic adaptation at the workplace: Multidimensional impact on human’s behavior. *ISES ISIAQ 2019*, 8 2019.
- [147] Xiaochen Zheng, Shengjing Sun, Joaquín Ordieres-Meré, Jinzhi Lu, and Dimitris Kiritsis. Applying distributed ledger technology to facilitate iiot data exchange: an approach based on iota tangle. *IESA 2020 Accepted*, 11 2020.
- [148] Xiaochen Zheng, Jinzhi Lu, Shengjing Sun, and Dimitris Kiritsis. Decentralized industrial iot data management based on blockchain and ipfs. *APMS 2020 Accepted*, 9 2020.
- [149] Marina Chen, Jie Yin, Shengjing Sun, Charlotte H. Gabel, Chia-Pin Yu, and John D. Spengler. Physiological and psychological responses to transition between urban and urban nature environments using immersive visualization. *Environment International Under Review*, 2020.
- [150] Stephen Kaplan. The restorative benefits of nature: Toward an integrative framework. *Journal of Environmental Psychology*, 15(3):169 – 182, 1995.
- [151] L. Chen, J. Hoey, C. D. Nugent, D. J. Cook, and Z. Yu. Sensor-based activity recognition. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 42(6):790–808, Nov 2012.