



Universidad Politécnica
de Madrid



**Escuela Técnica Superior de
Ingenieros Informáticos**

Grado en Ingeniería Informática

Trabajo Fin de Grado

**Creación de una Base de Conocimiento
para el Cálculo Transparente de
Emisiones de Carbono**

Autor: Adrián Ramos Olivé

Tutor: Daniel Garijo Verdejo

Cotutora: María Poveda Villalón

Madrid, 14 enero 2025

Este Trabajo Fin de Grado se ha depositado en la ETSI Informáticos de la Universidad Politécnica de Madrid para su defensa.

Trabajo Fin de Grado
Grado en Ingeniería Informática

Título: Creación de una Base de Conocimiento para el Cálculo Transparente de Emisiones de Carbono
14 enero 2025

Autor: Adrián Ramos Olivé
Tutor: Daniel Garijo Verdejo
Departamento de Inteligencia Artificial
ETSI Informáticos
Universidad Politécnica de Madrid

Agradecimientos

En primer lugar, quería agradecer a mis padres por apoyarme durante todos estos años, en especial en estos cuatro años y medio durante mi carrera. Por apoyarme cuando más difícil lo tenía, cuando más me frustraba estudiando y demás momentos complicados.

A mi tutor en este Trabajo de Fin de Grado, Daniel Garijo Verdejo, por confiar en mí para darme la oportunidad de continuar su trabajo. Su orientación y conocimiento han sido fundamentales durante el desarrollo de esta hazaña.

A mis amigos de universidad, con quienes he compartido durante este camino una etapa muy bonita y he estrechado lazos de amistad resistentes.

A mis amigos de siempre, que han estado en todo momento dando ánimos y apoyo durante mi carrera. Su amistad y apoyo han sido esenciales para mantenerme motivado y enfocado.

No olvidarme de la ciudad de Budapest, donde estudié mi tercer curso con el programa Erasmus, que me ha ayudado a crecer como persona y en la que he hecho amistades por todo el país y a los que estoy muy agradecido.

Adrián Ramos Olivé

Resumen

El propósito del proyecto es el conseguir hacer un cálculo transparente de emisiones de carbono equivalente (CO₂e) generadas por múltiples actividades, desde la estancia en un hotel, la creación de modelos de IA, hasta en el ámbito de la agricultura. Dicho cálculo ha sido realizado mediante factores de conversión dependiendo del país de origen, año de consumición o del elemento químico de origen.

Hasta el momento se había conseguido tener una ontología definida en el que nos permitía obtener información de todos los elementos, países, fórmulas químicas e información extra que intervienen en el proceso. Para que no hubiese ningún problema con el idioma se pensó en poner enlaces de Wikidata del elemento para que cualquier persona pudiese entender la ontología sin que la barrera del idioma pudiese ser un problema.

No obstante, el proyecto desarrollado por los investigadores ofrecía ciertos detalles a mejorar, como la claridad en los datos o un formato homogeneizado. Uno de los desafíos identificados ha sido que todos los factores de conversión están basados en emisiones intermedias, en lugar de dióxido de carbono equivalente (CO₂e) directo, lo que complica la homogeneización.

El alumno ha desarrollado la automatización de limpieza y organización de los datos pasando de una capa raw a una capa gold en la que los datos ya están listos para su uso.

Los datos provienen de Reino Unido (UK) desde el departamento de Negocio, Energía e Industria Estratégica (BEIS – Business, Energy & Industrial Strategy) y del Departamento de Seguridad Energética & Cero Neto (DESNZ – Department for Energy Security & Net Zero). Estos departamentos nos ofrecen la información básica que como las unidades en las que está medido el factor de conversión, el elemento químico de origen, el valor del factor de conversión, entre otros.

Otro aspecto importante para mencionar ha sido la integración de una nueva fuente para el año 2024. La fuente seleccionada ha sido de la organización EPA (Environmental Protection Agency) que es una agencia estadounidense independiente del gobierno federal de los Estados Unidos que se encarga hacer cumplir los estándares de las leyes ambientales y la organización británica NAEI (National Atmospheric Emissions Inventory) que se encarga de recoger métricas sobre emisiones atmosféricas en Reino Unido e informa a otras agencias como Comisión Económica de las Naciones Unidas y Convención Marco de las Naciones Unidas sobre el cambio climático. Estas agencias nos proporcionan información bastante útil sobre factores de conversión de distintos elementos químicos a CO₂e, lo cual nos beneficia mucho en nuestra investigación ya que se asemejan a los datos que tenemos hasta el momento lo que fomenta la homogeneización de los datos.

Este Trabajo de Fin de Grado se centra en la mejora y expansión de contenidos del trabajo realizado por distintos investigadores de múltiples universidades europeas como de la University of Aberdeen, University of Oxford, de la

University of Glasgow y por supuesto de la Universidad Politécnica de Madrid de la mano del tutor Daniel Garijo Verdejo.

Abstract

This Bachelor's Thesis is focused on enhancing and expanding upon the work conducted by various researchers from multiple European universities, including the University of Aberdeen, the University of Oxford, the University of Glasgow, and, of course, the Polytechnic University of Madrid under the guidance of supervisor Daniel Garijo Verdejo.

The purpose of the project is to develop a transparent calculation for equivalent carbon emissions (CO₂e) generated by various activities, ranging from hotel stays and AI model creation to the agricultural sector. This calculation has been achieved through conversion factors based on the country of origin, year of consumption, or originating chemical element.

To date, a well-defined ontology has been developed that enables us to retrieve information about all elements, countries, chemical formulas, and supplementary data involved in the process. To avoid language barriers, it was decided to link Wikidata entries for each element, ensuring that anyone could understand the ontology without language being an obstacle.

However, the project developed by the researchers revealed certain areas for improvement, such as data clarity and format standardization. One of the challenges identified was that all conversion factors were based on intermediate emissions rather than direct carbon dioxide equivalent (CO₂e), complicating standardization.

Additionally, the student developed a data-cleaning and organizing automation process, transitioning from a raw layer to a gold layer, where the data is ready for use.

The data originates from the United Kingdom (UK) through the Department for Business, Energy & Industrial Strategy (BEIS) and the Department for Energy Security & Net Zero (DESNZ). These departments provide fundamental information, such as the units in which the conversion factor is measured, the originating chemical element, and the conversion factor value, among others.

Another important aspect to mention is the integration of a new data source for the year 2024. The selected source is the Environmental Protection Agency (EPA), an independent U.S. federal government agency responsible for enforcing environmental laws and the National Atmospheric Emissions Inventory (NAEI) which responsible to collect values regarding atmospheric emissions in the United Kingdom and informs other agencies such the United Nations Economic Commission and the United Nations Framework Convention about climate change. These agencies provide highly useful information on conversion factors for various chemical elements to CO₂e, which significantly benefits our research as it aligns closely with the data we have collected thus far, thereby promoting data standardization.

This Final Degree Project focuses on the improvement and expansion of the content of the work carried out by different researchers from multiple European universities such as the University of Aberdeen, University of

Oxford, the University of Glasgow and of course the Polytechnic University of Madrid led by tutor Daniel Garijo Verdejo.

Índice General

1. Introducción	1
1.1. Contexto y motivación del proyecto	1
1.2. Objetivos	1
1.3. Planificación del Proyecto	2
1.4. Estructura del Documento	3
2. Estado del Arte	4
2.1. Análisis del trabajo previo	4
2.2. Tecnologías y herramientas usadas en el desarrollo	6
2.2.1. Python	6
2.2.2. Wikidata	6
2.2.3. GitHub	7
2.2.4. YAML	7
2.2.5. RDF	7
3. Marco Teórico	8
3.1. Fundamentos sobre las Emisiones del CO ₂	8
4. Desarrollo	11
4.1. Diseño del Proyecto	11
4.1.1. Requisitos Funcionales	11
4.1.2. Requisitos no funcionales	11
4.2. Diseño de la Arquitectura	12
4.2.1. Flujo de Datos	12
4.3. Automatización Preprocesamiento Datos	12
4.4. Estructura del Proyecto	14
4.5. Integración de nuevas fuentes	16
4.5.1. United States Environmental Protection Agency (EPA)	16
4.5.2. National Atmospheric Emissions Inventory (NAEI)	17
4.6. Mapping	18
4.6.1. Estructura mapping YAML	18
4.6.2. Ventajas del uso de YAML en los mappings	18
4.6.3. Proceso de creación de las tripletas YAML de los mappings	19
4.6.4. Resultados obtenidos	19
4.7. Grafo de conocimiento	19

4.7.1	Estructura del grafo de conocimiento.....	20
4.7.2	Visualización del grafo	20
4.7.3	Impacto del grafo de conocimiento	20
4.8	Extensión de la ontología	20
5.	Resultados y conclusiones	23
5.1.	Resultados	23
5.2.	Limitaciones del proyecto	23
5.3.	Propuestas de mejora.....	24
5.4.	Conclusiones personales	24
6.	Análisis de Impacto	25
6.1.	Impacto general.....	25
6.1.1.	Impacto personal.....	25
6.1.2.	Impacto social	25
6.1.3.	Impacto económico.....	25
6.1.4.	Impacto medioambiental	26
6.2.	Objetivos de desarrollo sostenible.....	26
Bibliografía	27
Anexos	30
	Código Python perteneciente a la automatización de procesos	30
	Mappings.....	36
	Preprocesamiento de las nuevas fuentes	40
	Nueva ontología	46

Índice de Figuras

Figura 1. Diagrama de Gantt	3
Figura 2. Grafo de Conocimiento Inicial	5
Figura 3. Estructura RDF	7
Figura 4. Resumen de los Scopes [12]	9
Figura 5. Diagrama de la Arquitectura	12
Figura 6. Extracto de los datos del directorio data_raw	14
Figura 7. Extracto de los datos del directorio data, una vez procesados	14
Figura 8. Estructura inicial del grafo de conocimiento	21
Figura 9. Estructura actual del grafo de conocimiento	22

Capítulo 1

1. Introducción

1.1. Contexto y motivación del proyecto

El objetivo de hacer este Trabajo de Fin de Grado surge de la necesidad de dar voz a la transformación energética y ecológica que debemos hacer como sociedad. La creciente preocupación por el cambio climático ha instado a gobiernos y organizaciones de todo el mundo a investigar sobre cómo podemos reducir las emisiones de dióxido de carbono asociadas a diversas actividades como son las estancias en hoteles, la creación de modelos de IA, hasta en el ámbito de la agricultura. Sin embargo, el cálculo de las emisiones de dióxido de carbono equivalente (CO₂e) no es un desafío sencillo ya que no solo depende de la cantidad consumida, sino que también depende de factores como el país de emisión (depende de la capacidad de desarrollo energético del país) o del año de consumición, entre otros factores.

Este proyecto ayuda a resolver la falta de uniformidad de los resultados ya que, en la mayoría de los casos no viene dado por la emisión equivalente de CO₂e que no se calcula directamente como dióxido de carbono, sino que corresponde principalmente al metano (CH₄) o al amoníaco (NH₄), expresados en términos de su impacto equivalente al CO₂. Por lo que, se tiene que buscar una solución para mitigar este problema.

A nivel técnico, este proyecto supone un reto supone un desafío personal ya que se requieren conocimientos técnicos como web semántica, tecnologías de mappings, grafos de conocimientos, RDF, sparql que eran inicialmente desconocidos para mí. El hecho de ser capaz de aprender estos términos y poder aplicarlos a mi trabajo de fin de grado supone un reto satisfactorio para poner el broche a mis estudios de estos últimos cuatro años. Durante este tiempo, he desarrollado una sólida disciplina, una ambición por aprender nuevos conceptos, capacidad de trabajar de manera autónoma, que, junto a los conceptos técnicos, que puedo aplicar ahora en este proyecto.

1.2. Objetivos

El objetivo principal del proyecto es ayudar a mejorar la implementación del trabajo realizado por mi tutor y los investigadores con los que colabora, quienes han asentado una buena base en la ontología de las emisiones de dióxido de carbono. A través de mi aportación, se busca optimizar la comprensión de los resultados, aportando calidad y homogeneización en el dato, automatización de procesos para hacerlo más preciso y accesible.

Los principales objetivos de este Trabajo de Fin de Grado son:

- **Estudio del trabajo previo y herramientas utilizadas:** se espera realizar un estudio del trabajo que llevaban realizado los participantes del proyecto de las distintas universidades europeas ya mencionadas y las herramientas con las que han trabajado para llevar a cabo dicha tarea.
- **Automatización de procesos:** esto incluye la creación de scripts para poder dar formato a los datos para su posterior análisis durante el cálculo de su respectivo factor de conversión.
- **Extensión de la base de conocimiento:** se hace un estudio de las posibilidades de aumentar los nodos del grafo de conocimiento ya existente para poder incorporar nuevos datos que ayuden a la total comprensión de los resultados.
- **Homogeneización de los datos:** debido a la nueva implementación del grafo de conocimiento se requiere homogeneizar los datos para que todos contengan el mismo formato y no haya errores en su cálculo.

1.3. Planificación del Proyecto

En esta sección se incluye un análisis de la cronología de las tareas que se han llevado a cabo a lo largo del proyecto. Se incluye un diagrama de Gantt para mostrar las fechas estimadas de cada tarea:

1. Familiarización con las tecnologías
 - i. Investigación concepto grafos de conocimientos
 - ii. Familiarización ontología
 - iii. Estudio tecnología SPARQL
2. Automatización de procesos ingesta de datos
 - i. Estudio datos en bruto
 - ii. Script Python
 - iii. Redacción del apartado
3. Implementación nuevas fuentes
 - i. Estudio fuente adecuada
 - ii. Estudio homogeneización datos
 - iii. Script Python integración y homogeneización nuevos datos
 - iv. Redacción del apartado
4. Diseño e implementación nuevo grafo de conocimiento
 - i. Estudio nuevo grafo de conocimiento
 - ii. Implementación del nuevo grafo de conocimiento
 - iii. Ajuste de datos al nuevo grafo
 - iv. Redacción del apartado
5. Resultado y conclusiones
 - i. Redacción de la memoria
 - ii. Análisis de resultados
 - iii. Conclusiones y trabajo futuro

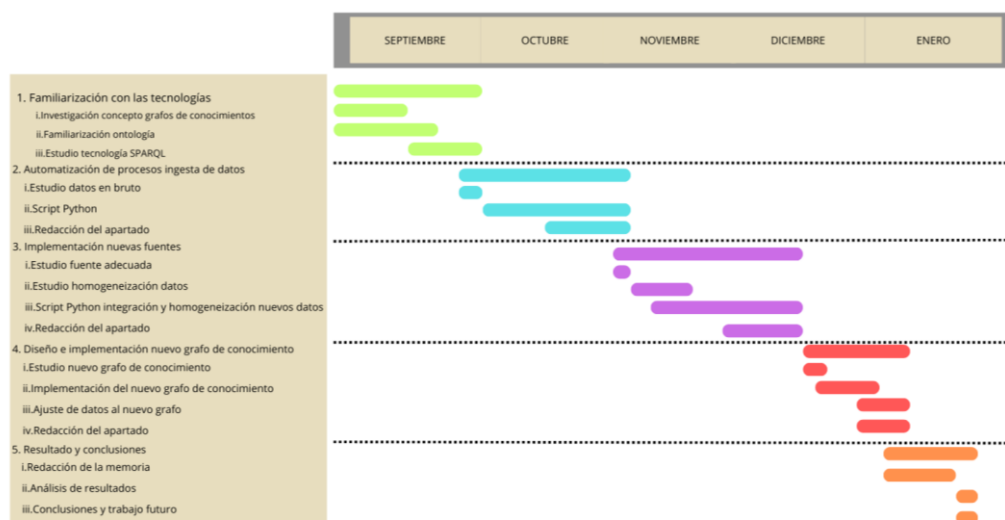


Figura 1. Diagrama de Gantt

1.4. Estructura del Documento

En este apartado se van a describir los capítulos que componen esta memoria de este Trabajo de Fin de Grado:

1. **Estado del Arte:** este capítulo se mencionan las herramientas y el trabajo previo. Además, se exponen las tecnologías y los recursos con las que se ha desarrollado este proyecto.
2. **Desarrollo:** esta parte contendrá los siguientes contenidos:
 - a. **Familiarización de las tecnologías:** se muestra una descripción de lo aprendido durante mi investigación de las tecnologías usadas como SPARQL, grafos de conocimiento y sobre la ontología y como se ha aplicado al proyecto.
 - b. **Automatización de procesos durante la ingesta de datos:** se explica cómo se realizó el tratado de datos para pasar de los datos en crudo a unos datos estructurados y homogéneos.
 - c. **Implementación de nuevas fuentes:** en este apartado se va a describir a detalle el proceso de la búsqueda de nuevas fuentes de datos que contengan información relevante para nuestro estudio.
 - d. **Estudio nuevo grafo de conocimiento:** se debe modificar el grafo de conocimiento ya existente para la adaptación de nuevos datos como el GWP para el mayor entendimiento de los resultados.

Resultado y conclusiones: en este último capítulo se presentan los resultados obtenidos tras el desarrollo. Se analizan las nuevas implementaciones sobre el trabajo previo y el impacto que ha supuesto. Se exponen los conocimientos adquiridos y los desafíos encontrados durante el desarrollo. Por último, se proponen nuevas implementaciones para seguir desarrollando el proyecto.

Capítulo 2

2. Estado del Arte

2.1. Análisis del trabajo previo

Antes de empezar con el desarrollo del proyecto es fundamental conocer el trabajo que ya ha sido implementado. Para ello, se ha hecho una búsqueda de conocimiento en el ámbito de las emisiones del dióxido de carbono equivalente (CO₂e) buscando como objetivo los factores que puede intervenir en una mayor emisión de CO₂e, los elementos de emisión más influyentes, etc. En este punto se comentan las conclusiones que he extraído:

- **Diferencia de CO₂ y CO₂e:** el CO₂ es el gas de efecto invernadero (GEI) con mayor emisión por las actividades humanas que se liberan a la atmósfera como resultado de distintos procesos industriales en los que destacamos la producción de combustibles fósiles (petróleo, carbón, gas natural ...) que supone el 80% del total del efecto del calentamiento global. No obstante, el CO₂ solo representa uno de los tantos GEI que existen. La función del CO₂e es la de incluir el resto de los gases contaminantes que permite comparar y sumar el impacto climático fruto de las distintas actividades que generan dichos gases.

El CO₂e no tiene sentido si no mencionamos el Potencial de Calentamiento Global (en inglés Global Warming Potencial – GWP) relaciona el potencial de cada gas con el CO₂ en un periodo de tiempo (en este caso utilizaremos un intervalo de tiempo de 100 años).

Esto significa que, si el origen de la emisión es NH₄, el factor de conversión que tenemos hasta ahora representa la cantidad de NH₄ de CO₂e y no de CO₂e directo como se contemplaba en un primer momento. Aquí entra en juego la labor de GWP, esta variable nos permitirá relacionar el valor que ya tenemos con el verdadero factor de conversión que estamos buscando. Para poder actualizar la ontología deberemos añadir un nuevo nodo GWP del que se añade la información relativa esta variable como el valor, el origen, etc.

Usaremos como referencia la cantidad de CO₂e emitido para cuantificar el impacto de proyectos climáticos y dejar reflejado que no solo es CO₂ lo que se emite, sino también otros gases de efecto invernadero, que a menudo son pasados por alto. [1]

- **Gases con mayor efecto invernadero:** existen multitud de gases de efecto invernadero y su contribución al calentamiento global no es uniforme. Es por ello la necesidad de conocer cuáles son los que suponen una mayor amenaza.

Entre estos gases destacamos:

- **CO₂:** producido mayormente por la respiración de los animales y a la descomposición de materia orgánica y también puede entrar a la atmósfera por actividades relacionadas con la quema de combustibles fósiles y reacciones químicas

- **CH4 (metano):** sus emisiones proceden de la producción y transporte de carbón, gas natural y petróleo, así como actividades agrícolas, pesca o la silvicultura.
- **N2O (óxido nitroso):** procedente de la quema de madera, combustión de materiales fósiles, residuos sólidos y, sobre todo, en la trata de aguas residuales
- **Gases fluorados:** la presencia de emisiones de gases en presencia del flúor es notoria. Podemos destacar los gases hidrofluorocarburos (presente en los frigoríficos, aerosoles, etc.), perfluorocarburos, hexafluoruro de azufre, etc.
- **País de origen:** una de las conclusiones que he obtenido que esta investigación es que, dependiendo del país de origen, la misma actividad y cantidad de emisión producirá distinto CO2e. No va a ser lo mismo quemar 1 kg de CH4 en un país desarrollado donde se tengan los mecanismos y herramientas para para generar energía de manera más eficiente que en uno menos desarrollado donde no se tiene la infraestructura necesaria para generar la misma cantidad de energía con la misma eficiencia.

Esto demuestra cómo las condiciones y los recursos disponibles en cada región afectan directamente al impacto ambiental de las emisiones de CO2e.

En relación con el trabajo realizado por Daniel Garijo Verdejo y los investigadores con los que colabora podemos obtener las siguientes conclusiones [2]:

- **Grafo de conocimiento:** por el momento tenemos un grafo de conocimiento con la siguiente forma [3].

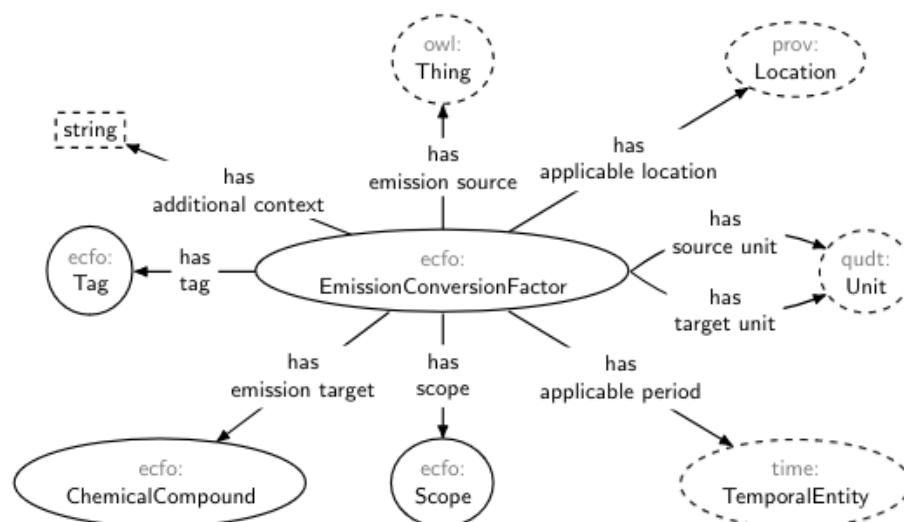


Figura 2. Grafo de Conocimiento Inicial

El grafo de conocimiento se centra en EmissionConversionFactor, que representa la variable objetivo que será el resultado del factor de conversión entre los gases implicados en la actividad realizada, que se

conecta con entidades clave como Location, Scope, ChemicalCompound, TemporalEntity, que representan la ubicación, el tipo de emisión y su “cadena de valor”, el compuesto químico formada por la unión de dos o más elementos y el intervalo temporal, respectivamente.

- **Datos actuales:** actualmente se limpian los datos de manera manual, es decir por cada set de datos que obtenemos directamente de las entidades y organismos públicos de los gobiernos estadounidenses y británicos se realiza un script en yaml para ajustar los datos en crudo a una capa de datos gold para su conveniente uso.
- **Tripletas tratamiento de datos:** una vez se hace el mapping con los datos en crudo, se obtiene un grafo basado en tripletas que contiene en un formato sujeto, predicado y objeto, toda la información necesaria para entender el resultado del factor, esto es todos los nodos del grafo de conocimiento.

Para el análisis del proyecto se ha tenido en cuenta la web TEC Toolkit donde se especifica todo lo relacionado con el trabajo realizado previamente. [4]

2.2. Tecnologías y herramientas usadas en el desarrollo

2.2.1. Python

Python es un lenguaje de programación muy polivalente que te permite de manera extensa y eficiente. Su amplia variedad de bibliotecas lo convierte en una herramienta ideal para el tratamiento y la manipulación de grandes cantidades de datos. La biblioteca que mejor se adapta a mis necesidades es *Pandas* la cual permite organizar los datos en DataFrames, lo cual supone una alternativa ideal para almacenar, analizar y realizar operaciones con una gran cantidad de datos a partir de archivos Excel (.xlsx) y CSV. [5] [6]

2.2.2. Wikidata

Gracias a Wikidata somos capaces de internacionalizar los elementos que componen el proceso del cálculo del factor de conversión. La idea de este proyecto es la que cualquier persona del planeta sea capaz de entender e interpretar a la perfección los resultados, para ello, no debemos dar por hecho que el lector domina inglés (que es el idioma predominante y sobre el que está desarrollado el proyecto). Es por ello por lo que gracias a Wikidata podemos incluir los enlaces correspondientes a los elementos participantes para que el lector sea capaz de entender la ontología sin que el idioma sea una barrera. [7]

2.2.3. GitHub

Para este proyecto, ha sido fundamental para dejar actualizado el repositorio con todos los avances del código. En caso de querer retroceder en alguna versión del código, la plataforma GitHub te lo permite. [8]

Aquí se encuentra el repositorio de este proyecto: [9]

2.2.4. YAML

YAML es un lenguaje intuitivo y legible por humanos para mapear datos. Proporciona un formato estandarizado para representar datos estructurados de forma que sea fácil de entender por humanos. La sintaxis de YAML está basada en parejas clave-valor y está pensado para el intercambio de datos entre distintos sistemas. [10]

2.2.5. RDF

RDF (Resource Description Framework) sirve para proporcionar información descriptiva sobre los recursos que se encuentran en una web, permitiendo el intercambio de información facilitando la reutilización de recursos. Una característica fundamental de RDF es la combinación de distintas fuentes de datos, lo que hace que este proyecto sea fácilmente extensible. Presenta la siguiente estructura:



Figura 3. Estructura RDF

Capítulo 3

3. Marco Teórico

3.1. Fundamentos sobre las Emisiones del CO2

En este capítulo se detallan conceptos fundamentales para el entendimiento del ámbito de las emisiones de gases de efecto invernadero. Se abordan las principales clasificaciones, categorías y definiciones que estructuran este campo, así como Scopes, los distintos niveles (nivel 1, nivel 2, nivel 3 y nivel 4), la fuente de emisiones o el objetivo de emisiones, entre otros conceptos. Este conocimiento es primordial para el completo entendimiento del trabajo realizado y los cambios que se han necesitado para completar el grafo de conocimiento.

- **Id:** identificador que permite localizar todos los factores de conversión de manera sencilla y segura.
- **Scope:** supone una medida que identifica las distintas actividades que producen las emisiones de gases de efecto invernadero según el protocolo GHG creado bajo el protocolo de Kioto en 1997. Se dividen en tres tipos de scopes dependiendo de la actividad:
 - **Scope 1:** según el ministerio de Transición Ecológica y el Reto Demográfico del Gobierno de España, en este alcance entran en juego el consumo de combustibles en instalaciones fijas de una empresa, cualquier medio de transporte (terrestre, marítimo o aéreo), maquinaria, etc.
 - **Scope 2:** agrupa todas las emisiones indirectas asociadas al consumo eléctrico en edificios y vehículos, energía térmica, energía renovable, entre otros. En el caso de la fabricación de un coche, podríamos agrupar en este alcance de impacto, todas las emisiones generadas por la electricidad consumida al hacer funcionar las instalaciones, las máquinas de café de las oficinas o la luz de la fábrica de montaje.
 - **Scope 3:** este nivel de alcance supone el resultado de las actividades que no son propiedad por la entidad informadora, es decir, todas aquellas gestiones que no están incluidas en los scopes 1 y 2 [11]. Podemos agrupar estas actividades en *upstream* y *downstream activities*.
 - **Upstream activities:** son todas aquellas relacionadas con los bienes y servicios adquiridos o comprados a un distribuidor. Podemos encontrar actividades como: transporte y distribución, transporte de empleados, etc.
 - **Downstream activities:** son todas aquellas relacionadas con los bienes y servicios vendidos a un cliente cuando ya dejan de ser propiedad de la empresa. Podemos encontrar actividades como: inversiones, creación de franquicias, etc. [12].

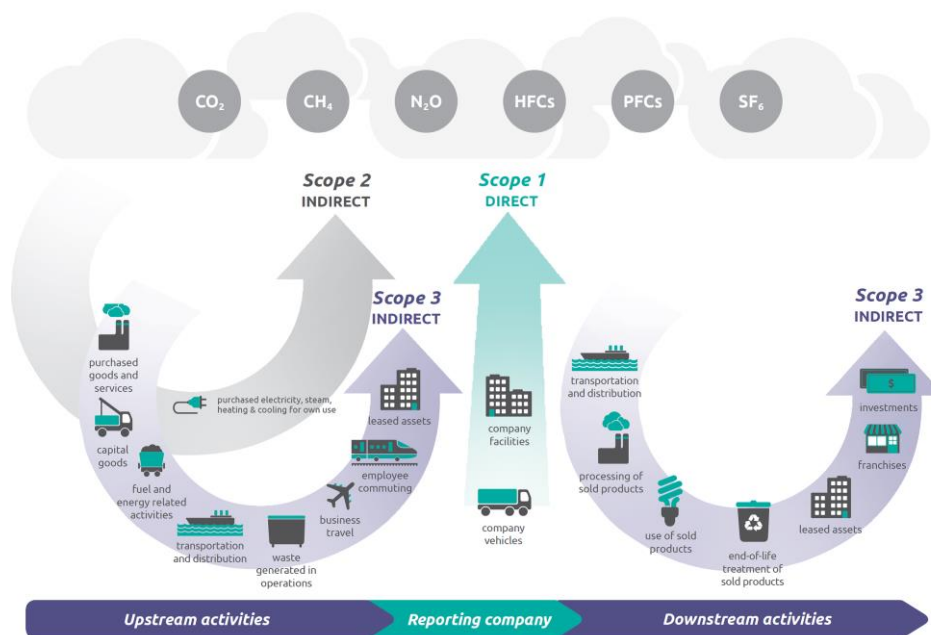


Figura 4. Resumen de los Scopes [13]

Estas clasificaciones suponen un aspecto fundamental en la clasificación de las actividades que producen emisiones de gases de efecto invernadero desde la producción de productos y servicios, la energía consumida hasta la venta de estos. [14]

- **Level:** especifica por capas las fuentes de emisiones de CO₂, que generan las emisiones. Hay cuatro niveles:
 - **Level 1:** clasificación general (por ejemplo: combustibles, vehículos de transporte, etc.).
 - **Level 2:** subcategoría que defina la clasificación general (por ejemplo: combustible gaseoso, camiones refrigerados (solo diésel), etc.).
 - **Level 3:** desglose más específico de la subcategoría que, prácticamente termine de definir la fuente de emisión (por ejemplo: butano, antracita, etc.).
 - **Level 4:** ofrece detalles sobre el nivel 3 (por ejemplo: clase turista, business, etc.).
- **Column Text:** ofrece información extra sobre la fuente de emisión. Una breve descripción que termine de definir el tipo de emisor.
- **Emission Source:** genera un id de la fuente de emisión uniendo el *Level 2* y *Level 3*.
- **UOM:** indica la unidad de medición que se utiliza para cuantificar las emisiones (por ejemplo: toneladas, litros, kilovatios por hora, etc.)
- **GHG/Unit:** indica la unidad y el elemento químico que produce la actividad indicada (por ejemplo: kg CO₂e, kg CO₂e of N₂O)
- **Emission Target Formula:** selecciona la formula del elemento químico que produce la emisión de gases de efecto invernadero. (por ejemplo: CO₂e of N₂O, CO₂e of CO₂, etc.)

- **Emission Target:** supone la nomenclatura de *emisión target formula*. (por ejemplo: Carbon dioxide equivalent of Nitrous Oxide, Carbon dioxide equivalent of Methane)
- **GHG:** selecciona la unidad básica de medida como, por ejemplo: kilogramos, libras, etc.)
- **Valid from:** indica el comienzo de la validez del factor de conversión en cuestión en formato YYYY-MM-DD.
- **Valid to:** indica el final de la validez del factor de conversión en cuestión en formato YYYY-MM-DD.
- **Region:** indica la región desde donde se recogen los datos. En caso de los datos recogidos por BEIS, DESNZ y NAEI será Reino Unido y en caso de EPA será Estados Unidos.
- **GWP:** la agencia IPCC lo define la relación entre la liberación instantánea de 1 kg de una sustancia traza en relación con el de 1 kg de un gas de referencia. Cada 7 años el IPCC ajusta los datos con el objetivo de cada año ser más exactos, para ello crea los *Assessments Reports* (AR), en el 2007 publicaron el AR4, en 2014 el AR5 y en 2021 el AR6. Tienen como objetivo calcular la relación del GWP entre distintos emisores de gases de efecto invernadero como del metano (CH₄), óxido nitroso (N₂O), fluoruros, entre otros elementos destacados. En nuestro proyecto obtendremos los datos pertenecientes al óxido nitroso y al metano. [15]
- **GWP id:** id único para cada tipo de GWP dependiendo del emisor y del año de validez (por ejemplo: ch4_co2e_2014, ch4_co2e_2022, n2o_co2e_2022, etc.)
- **Value:** es el valor del factor de conversión en bruto, es decir, sin haber aplicado el GWP.
- **Conversion Factor:** es el valor del factor de conversión una vez aplicado el GWP.

Además de todas estas variables para tener en cuenta, también se ha añadido en todos los DataFrames una columna que contiene las URL de WikiData de todos los datos de las columnas UOM, emission target, GHG y region, para su entendimiento en cualquier idioma y su correcta definición.

Capítulo 4

4. Desarrollo

4.1. Diseño del Proyecto

En esta función se describen los requisitos funcionales y no funcionales para el desarrollo del proyecto, así como las diferentes decisiones que se han ido tomando para su correcto funcionamiento.

4.1.1. Requisitos Funcionales

- **Gestión de Datos de Emisiones**
 - Identificación y almacenamiento de bases de datos de factores de conversión de CO₂e.
 - Integración de los factores de conversión de emisión en el grafo de conocimiento.
- **Construcción del Grafo de Conocimiento**
 - Extensión del grafo de conocimiento TEC Toolkit con la introducción del concepto de GWP.
- **Interoperabilidad**
 - Capacidad para exportar e importar datos en formato RDF.
 - Enlace a recursos externos como Wikidata
- **Transparencia y Validación**
 - Registro de fuentes y procesos de automatización para auditoría.
 - Generación de informes sobre el cálculo de emisiones.
- **Soporte Multilingüe**
 - Implementación de soporte para explorar la integración en cualquier idioma.

4.1.2. Requisitos no funcionales

- **Escalabilidad**
 - La estructura del grafo debe estar preparada para recibir nuevos datos en un futuro y debe ser capaz de integrar nuevas funcionalidades.
- **Mantenibilidad**
 - Debe haber un código estructurado y documentado para facilitar las futuras modificaciones.
- **Estandarización**
 - Cumplimiento sobre las normativas internacionales sobre el cálculo de emisiones
 - Uso de estándares abiertos como AR5 y AR6. [16]

- **Integración y Dependencias**
 - Uso de bibliotecas como rdflib y SPARQL

4.2 Diseño de la Arquitectura

La arquitectura del proyecto ha sido creada para ser modular y, sobre todo, escalable, facilitando la integración de nuevas fuentes de datos en el futuro. A continuación, se expone el flujo de datos de la aplicación.

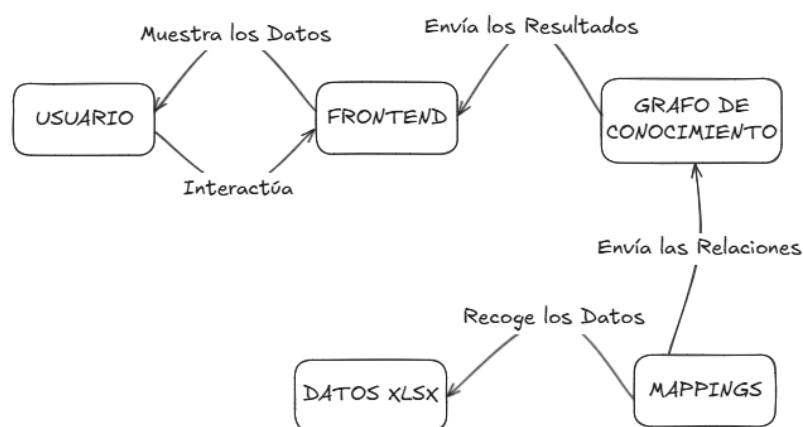


Figura 5. Diagrama de la Arquitectura

4.2.1 Flujo de Datos

El flujo de datos del proyecto sigue estos pasos:

1. Los usuarios interactúan con el frontend realizando los factores de conversión.
2. El grafo de conocimiento envía la información pertinente relacionada con los factores de conversión al frontend.
3. Se envían las relaciones generadas por las tripletas de los mapping al grafo de conocimiento.
4. Los datos pertenecientes al mapping se obtienen de los archivos .xlsx de las fuentes de datos.

4.3 Automatización Preprocesamiento Datos

Para esta sección se detalla cómo se ha automatizado el proceso de pasar los datos en crudo de las fuentes previamente identificadas, es decir, según se recibe desde la fuente de datos, a unos datos más organizados listos para su uso en el grafo de conocimiento.

Para ello, en primer lugar, se ha leído un archivo XLSX, que como se ha explicado en el apartado 3.2.1, contiene la información relevante de cada año de los datos proporcionados por el gobierno británico (BEIS y DESNZ). Para leer y procesar los datos se ha hecho uso de la herramienta **Pandas**, que es una

biblioteca de Python que proporciona estructuras de datos (DataFrame) y herramientas de análisis de datos. Con los datos ya en un DataFrame, se procede a la reestructuración de los datos según la especificación dada por el tutor y sus compañeros, entre otras indicaciones podemos destacar:

- **Integración links Wikidata:** para una mayor estandarización y explicación de conceptos más técnicos, se ha procedido a incluir enlaces a cada elemento para su mayor explicación y un mayor soporte multilingüe, ya que Wikidata facilita la integración de datos de manera independiente al idioma, permitiendo que las referencias y conceptos puedan ser comprendidos sin importar el idioma original, facilitando así su entendimiento desde cualquier rincón del mundo. Para ello se tuvo que crear un nuevo archivo XLSX para generar un diccionario en el que se incluyera el elemento y su respectivo enlace de Wikidata.
 - **Generación de XLSX con enlaces a Wikidata:** para la generación de dicho documento se ha procedido a realizar un proceso de web scraping en el que en función del elemento se obtenía su enlace de Wikidata pertinente teniendo en cuenta el área en el que se está buscando (elementos químicos, unidades de medida o regiones del mundo).

Una vez generado el archivo donde se van a almacenar los datos de los elementos y sus respectivos enlaces, se procesa y se crea un diccionario con ambas columnas y en función del dato que haya en cada columna se busca su respectivo enlace y se incluye en una nueva columna del DataFrame.

- **Formato del documento:** como especificaciones se incluía que ciertas columnas deberían llevar un determinado formato como son el uso de mayúsculas, la inclusión del año del reporte, una columna debía ser la concatenación de dos columnas originales con un guion bajo ('_') como separador, limpiar los valores nulos, etc.

Inicialmente contábamos con 50644 datos repartidos en 8 años, desde 2016 hasta 2023, y, tras el filtrado, se han obtenido 44461 datos, lo que supone un 87.79% de los datos originales.

$$\% \text{ Datos Finales} = \frac{N^{\circ} \text{ Datos Finales}}{N^{\circ} \text{ Datos Iniciales}} * 100\%$$

A continuación, podemos observar una comparativa de como recibimos los datos de su respectivo origen (Figura 5) y como tras un proceso de automatización se transforman hasta obtener una estructura ordenada y completa (Figura 6).

ID	Scope	Level 1	Level 2	Level 3	Level 4	Column Text	UOM	GHG/Unit	GHG Conversion Factor 2023
10	1_100_1000_15_4	Scope 1	Fuels	Gaseous fuels	Butane		tonnes	kg CO2e of N2O per unit	1.60067
11	1_100_1000_8_1	Scope 1	Fuels	Gaseous fuels	Butane		litres	kg CO2e	1.74293
12	1_100_1000_8_2	Scope 1	Fuels	Gaseous fuels	Butane		litres	kg CO2e of CO2 per unit	1.74296
13	1_100_1000_8_3	Scope 1	Fuels	Gaseous fuels	Butane		litres	kg CO2e of CH4 per unit	0.00144
14	1_100_1000_8_4	Scope 1	Fuels	Gaseous fuels	Butane		litres	kg CO2e of N2O per unit	0.00092
15	1_100_1000_7_1	Scope 1	Fuels	Gaseous fuels	Butane		kg CO2e		0.24107
16	1_100_1000_7_2	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Net CV)	kg CO2e of CO2 per unit	0.24074
17	1_100_1000_7_3	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Net CV)	kg CO2e of CH4 per unit	0.000102
18	1_100_1000_7_4	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Net CV)	kg CO2e of N2O per unit	0.00012
19	1_100_1000_6_1	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Gross CV)	kg CO2e	0.22241
20	1_100_1000_6_2	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Gross CV)	kg CO2e of CO2 per unit	0.22219
21	1_100_1000_6_3	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Gross CV)	kg CO2e of CH4 per unit	0.00019
22	1_100_1000_6_4	Scope 1	Fuels	Gaseous fuels	Butane		KWh (Gross CV)	kg CO2e of N2O per unit	0.00012
23	1_100_1001_15	Scope 1	Fuels	Gaseous fuels	CNG		tonnes	kg CO2e	2552.7441
24	1_100_1001_15_2	Scope 1	Fuels	Gaseous fuels	CNG		tonnes	kg CO2e of CO2 per unit	2557.33000
25	1_100_1001_15_3	Scope 1	Fuels	Gaseous fuels	CNG		tonnes	kg CO2e of CH4 per unit	3.85389
26	1_100_1001_15_4	Scope 1	Fuels	Gaseous fuels	CNG		tonnes	kg CO2e of N2O per unit	1.19251
27	1_100_1001_8_1	Scope 1	Fuels	Gaseous fuels	CNG		litres	kg CO2e	0.44485
28	1_100_1001_8_2	Scope 1	Fuels	Gaseous fuels	CNG		litres	kg CO2e of CO2 per unit	0.44757
29	1_100_1001_8_3	Scope 1	Fuels	Gaseous fuels	CNG		litres	kg CO2e of CH4 per unit	0.00087
30	1_100_1001_8_4	Scope 1	Fuels	Gaseous fuels	CNG		litres	kg CO2e of N2O per unit	0.00031
31	1_100_1001_7_1	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Net CV)	kg CO2e	0.20267
32	1_100_1001_7_2	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Net CV)	kg CO2e of CO2 per unit	0.20229
33	1_100_1001_7_3	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Net CV)	kg CO2e of CH4 per unit	0.00011
34	1_100_1001_7_4	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Net CV)	kg CO2e of N2O per unit	0.00010
35	1_100_1001_6_1	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Gross CV)	kg CO2e	0.18299
36	1_100_1001_6_2	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Gross CV)	kg CO2e of CO2 per unit	0.18256
37	1_100_1001_6_3	Scope 1	Fuels	Gaseous fuels	CNG		KWh (Gross CV)	kg CO2e of CH4 per unit	0.00028

Figura 6. Extracto de los datos del directorio data_raw

1	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	
1	Scope	Level 1	Level 2	Level 3	Level 4	Column Text	emission source	UOM	GHG/Unit	ion target	ion target	ion target	ion target	GHG	GHG_wd	valid from	valid to	region	region_wd	GWP	Value 2023	GWP_id	Conversion Factor 2023	
2	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	tonnes	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
3	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	tonnes	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
4	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	tonnes	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
5	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	tonnes	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
6	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	litres	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
7	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	litres	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
8	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	litres	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
9	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	litres	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
10	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
11	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
12	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
13	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
14	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
15	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
16	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
17	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
18	Scope1	Fuels	Gaseous fuels	Butane			Gaseous_Fuels_Butane	KWh	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
19	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	tonnes	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
20	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	tonnes	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
21	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	tonnes	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
22	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	litres	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
23	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	litres	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
24	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	litres	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
25	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	litres	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
26	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
27	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
28	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
29	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
30	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
31	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
32	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of CH4	CO2e of CH4	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
33	Scope1	Fuels	Gaseous fuels	CNG			Gaseous_Fuels_CNG	KWh	https://www.kg CO2e of N2O	CO2e of N2O	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
34	Scope1	Fuels	Gaseous fuels	LNG			Gaseous_Fuels_LNG	tonnes	https://www.kg CO2e	CO2e	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093
35	Scope1	Fuels	Gaseous fuels	LNG			Gaseous_Fuels_LNG	tonnes	https://www.kg CO2e of CO2	CO2e of CO2	Carbon dioxide equ	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	https://www.kg	2023	3093	3093	3093

Figura 7. Extracto de los datos del directorio data, una vez procesados

4.4 Estructura del Proyecto

Para cada una de las fuentes se ha seguido la siguiente organización de directorios y archivos que cumplen funciones específicas:

- Data_raw: contiene la información en crudo, lista para su automatización (descrita en la sección 4.3).
- Data: contiene los datos procesados, con el formato correspondiente.
- Mappings: archivos yaml para generar las tripletas.
- py: contiene los scripts Python desde donde se obtienen los archivos de data_raw para transformarlos a archivos de la carpeta data.
- GWP: contiene la tabla con la información correspondiente al GWP.

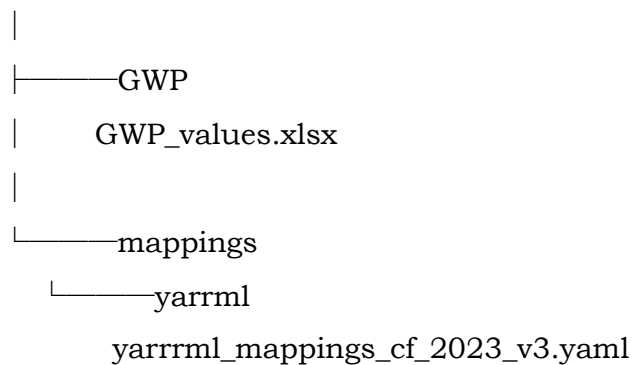
La estructura de directorios para cada fuente de datos es la siguiente:

```
C:
| cf_2023_v3.ini
| data_preparation_script.txt
| generate_ttl.bat
| nt_to_ttl.py
```

```

| README.md
| wikidata_labels.ttl
|
|-----data
| |-----v2
| |   conversion_factors_2023_modified_v2.csv
| |
| |-----v3
| |   Conversion_Factor_2023.csv
| |   Conversion_Factor_2023.xlsx
|
|-----data_processing
|   processing.py
|
|-----data_raw
|   ghg-conversion-factors-2023-flat-file-update.xlsx
|   readme.md
|
|-----graphs
| |-----v2
| | |-----nt
| | |   out_cf_2023_v2.nt
| | |
| | |-----ttl
| | |   out_cf_2023_v2.ttl
| | |
| |-----v3
| | |-----nt
| | |   out_cf_2023_v3.nt
| | |
| | |-----ttl
| | |   out_cf_2023_v3.ttl

```



En este caso obtenemos la estructura de directorios para la fuente DESNZ.

4.5 Integración de nuevas fuentes

Durante este apartado se explicará el proceso de obtención de nuevas fuentes para conseguir una mayor riqueza y precisión en los datos integrados.

Durante mi investigación encontré diversas organizaciones que aportaban datos sobre las emisiones de distintas actividades de CO₂, no obstante, me he decantado por escoger dos organizaciones estatales que, a mi parecer, contienen una mayor calidad en los datos aportados como son *United States Environmental Protection Agency (EPA)* y *National Atmospheric Emissions Inventory (NAEI)*.

4.5.1 United States Environmental Protection Agency (EPA)

La Agencia de Protección Ambiental (EPA en inglés [17]), constituye una institución ejecutiva independiente del gobierno federal de Estados Unidos que se encarga de la protección ambiental. Fundada en 1970 está dirigida por un administrador que es designado por el presidente y aprobado por el Senado. Tiene como principales responsabilidades reducir los riesgos medioambientales, garantizar los estándares de diversas leyes medioambientales, aplicando multas y sanciones por su incumplimiento, ciertos productos químicos del mercado son revisado de manera periódica para garantizar su seguridad.

Esta agencia proporciona datos anualmente sobre las emisiones de CO₂ de diversas actividades en el territorio americano. Para nuestra investigación nos ceñiremos a los años que nos faltan (en comparación con las fuentes ya integradas BEIS y DESNZ) que en este caso sería 2024 [18]. Entre las distintas actividades descritas en el reporte podemos destacar electricidad consumida, combustión móvil, transporte de aguas, transporte de productos, etc., todas estas actividades se enmarcan en los diferentes *Scopes* definidos para la medición y categorización de emisiones de CO₂.

A diferencia de las fuentes ya integradas, para esta fuente se ha tenido que hacer una etapa preliminar de preprocesamiento de los datos ya que no se encontraban en crudo donde los datos están sin procesar y así poder aplicar el

script generado para las anteriores fuentes. Para ello, se ha tenido que analizar cada apartado, correspondiente a una actividad generadora de CO₂. En esta fase se han tenido que tomar decisiones como el *Scope* que pertenece, los niveles e GHG/Unit y la unidad en la que se emite la actividad. Se han seguido utilizando la herramienta Pandas de Python. Una vez ya obtenida la capa *bronze*, se procede a aplicar el mismo procedimiento que con las fuentes BEIS y DESNZ hasta obtener los datos procesados y homogeneizados, listos para su respectivo mapping.

4.5.2 National Atmospheric Emissions Inventory (NAEI)

Esta agencia británica se encarga de recoger métricas sobre emisiones atmosféricas en Reino Unido e informa a otras agencias como Comisión Económica de las Naciones Unidas (UN/ECE en inglés) y Convención Marco de las Naciones Unidas sobre el Cambio Climático (UNFCCC en inglés) con el objetivo desarrollar, publicar y actualizar regularmente los datos de las emisiones de los gases de efecto invernadero utilizando los reportes del Grupo Intergubernamental de Expertos sobre el Cambio Climático (IPCC en inglés) [19].

Esta agencia proporciona datos anualmente sobre las emisiones de CO₂ de diversas actividades. Para nuestra investigación nos ceñiremos a los años pertenecientes entre 2016 y 2022. Entre las distintas actividades descritas en el reporte podemos destacar industria, combustión industrial, refinerías, estaciones de tren o agricultura, entre otros. Todas estas actividades se enmarcan en los diferentes *Scopes* definidos para la medición y categorización de emisiones de CO₂.

Esta nueva fuente nos aporta nuevos datos que enriquecen y complementan el grafo de conocimiento, permitiendo un análisis más detallado y preciso. Para este conjunto de datos hemos integrado el concepto de *CO₂ como C* (*CO₂ as C en inglés*) que se define como la cantidad de dióxido de carbono emitido en términos de los átomos de carbono. Científicos de todo el mundo utilizan esta expresión para simplificar cálculos o estandarizar la comparación de emisiones ya que en la mayoría de los gases de efecto invernadero, el carbono es el componente principal. Es importante conceptualizar las diferencias entre el CO₂ y el *CO₂ como C* ya que no suponen el mismo componente, mientras el CO₂ hace referencia a todo el compuesto, el *CO₂ como C* hace hincapié en el aislamiento del carbono. Un ejemplo para ver el enfoque del *CO₂ como C* sería el siguiente:

$$\text{Conversión molar C} = \frac{\text{Masa molar C}}{\text{Masa molar CO}_2} = \frac{12 \text{ g/mol}}{44 \text{ g/mol}} = 12/44$$

Por lo tanto, para expresar CO₂ como C, se utiliza el factor de conversión:

$$C = \text{CO}_2 \cdot 12/44$$

Por lo tanto, si una actividad genera 44 toneladas de CO₂, generará:

$$C = 44 \text{ T} \cdot 12/44 = 12 \text{ T de C}$$

Con este ejemplo podemos observar con un ejemplo práctico como 44 toneladas de CO2 no son lo mismo que 44 toneladas de *CO2 como C*.

4.6 Mapping

En esta sección se describirá el apartado correspondiente a los mapping de los datos, que supone un paso fundamental en la estructuración de la información de manera eficiente dentro del grafo de conocimiento.

Los mappings permiten transformar y organizar la información en crudo en tripletas RDF bajo el uso de archivos YAML. Se representa bajo un modelo sujeto – predicado – objeto que es esencial para que las relaciones sean sencillas y comprensibles.

4.6.1 Estructura mapping YAML

El formato de YAML se rige por la siguiente sintaxis: sujeto – predicado – objeto que lo caracteriza como una estructura sencilla y una buena capacidad para organizar los datos de manera jerárquica, ya que establece una relación padre-hijo entre los elementos anidados, lo que simplifica la definición de las tripletas. A continuación, se describen los principales componentes de las tripletas:

- **Sujeto:** supone la cabecera de la tripleta, que puede ser un elemento químico, un GWP o un factor de conversión. Todos ellos, están representados por su url correspondiente dentro del grafo de conocimiento.
- **Predicado:** define la propiedad que relaciona el sujeto y el objeto.
- **Objeto:** indica el valor o entidad relacionada con el sujeto

Un ejemplo de una tripleta RDF puede ser el siguiente:

```
<https://w3id.org/ecfkg/i/USA/EPA/2024/CF_224>  
<http://www.w3.org/1999/02/22-rdf-syntax-ns#value>  
"7.22038"^^<http://www.w3.org/2001/XMLSchema#float> .
```

En el que la primera parte, el sujeto, representa el factor de conversión numero 224 de la fuente EPA, la segunda parte, el predicado, indica que existe una relación que indica el valor numérico del factor de conversión, que se refleja en la tercera y última parte, el objeto, que indica que el valor obtenido es un numero decimal, float, con valor 7.22038.

4.6.2 Ventajas del uso de YAML en los mappings

- **Legibilidad:** su simple formato facilita la edición y revisión de los datos ya que evita que el usuario tenga que especificar todos sus parámetros en una línea de comandos.

- **Flexibilidad:** el uso de ficheros yaml permite al usuario una mejor lectura de los datos y facilita la escritura.
- **Escalabilidad:** YAML permite números tipos de datos como diccionarios, matrices, escalares, etc. Lo que permite mantener una estructura limpia y organizada.

4.6.3 Proceso de creación de las tripletas YAML de los mappings

El proceso de creación de tripletas se puede dividir en varias etapas:

- **Identificación de datos clave:** se determinan que datos van a formar parte de cada componente, como el factor de conversión, la unidad de medida, etc.
- **Definición de tripletas:** se establecen las relaciones mediante la asociación cada elemento al modelo RDF.
- **Validación de los mapping:** se realiza una comprobación de su correcto funcionamiento de los valores obtenidos.

Para la creación de las tripletas es fundamental un archivo intermedio que ejecute el archivo YAML. Aquí entra en juego el archivo .ini, que indica el origen del archivo YAML a procesar y la salida del archivo. Tiene la siguiente estructura:

```
[CONFIGURATION]
output_file=out_cf_2024_v3

[DataSourceCSV]
mappings=mappings/yarrml/yarrml_mappings_cf_2024_v3.yaml
```

4.6.4 Resultados obtenidos

El uso de los mappings ha permitido la transformación de los datos para todas las fuentes, logrando un grafo de conocimiento robusto gracias a las relaciones precisas y homogéneas. Además, esta metodología ha optimizado el tiempo de carga de los datos y ha mejorado la calidad de estos.

4.7 Grafo de conocimiento

El grafo de conocimiento es una estructura que permite representar los datos disponibles mediante relaciones, utilizando un enfoque semántico para aportar coherencia.

En este proyecto presenta un papel fundamental para organizar y relacionar las fuentes de información sobre los factores de conversión de las distintas actividades que emiten Gases de Efecto Invernadero (GEI) con otras entidades relevantes como los países de origen, los elementos químicos, actividades emisoras, unidades de medida, entre otros.

4.7.1 Estructura del grafo de conocimiento

Como se ha explicado en el apartado Mapping, el grafo de conocimiento se compone de tripletas RDF con un sujeto – predicado – objeto que establecen las relaciones semánticas entre las entidades. Esta estructura permite que los datos estén interconectados, formando una red navegable y fácil de explorar.

4.7.2 Visualización del grafo

El grafo puede ser consultado mediante herramientas como:

- **SPARQL:** que es un lenguaje de consultas diseñado para interactuar con grafos de conocimiento basados en RDF. Este lenguaje permite recuperar información específica mediante filtros y condiciones generando reportes personalizados y realizar análisis detallados de las conexiones entre entidades.
- **Interfaces de usuario:** así como implementación de aplicaciones web que permiten explorar visualmente las relaciones de manera más intuitiva.

4.7.3 Impacto del grafo de conocimiento

- **Homogeneización de los datos:** al tener datos de varias fuentes, se consigue unificar los distintos formatos en un único modelo homogeneizado, eliminando redundancias y garantizando la consistencia de los datos.
- **Escalabilidad:** el grafo de conocimiento permite integrar nuevas fuentes de información de manera sencilla.
- **Interoperabilidad:** los datos pueden ser reutilizados gracias a los estándares ofrecidos por RDF y SPARQL.

4.8 Extensión de la ontología

En este apartado se detalla cómo se ha tenido que modificar el grafo de conocimiento de integrar nuevos conceptos con el objetivo de homogeneizar y nutrir de información al grafo de conocimiento. La incorporación de nuevas fuentes ha permitido localizar los nuevos términos para poder representar de manera más precisa las relaciones entre las entidades involucradas en el cálculo de emisiones de carbono equivalente (CO₂e).

Al comienzo del proyecto presentábamos el siguiente esquema relacionado con el grafo de conocimiento:

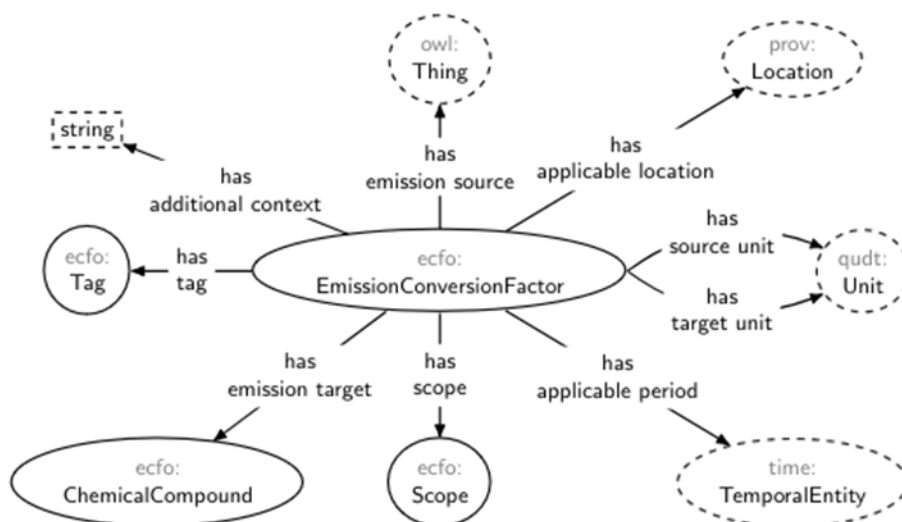


Figura 8. Estructura inicial del grafo de conocimiento

Pero debido a la incorporación del concepto del Potencial de Calentamiento Global (GWP, por sus siglas en inglés), el modelo se ha ampliado para incluir información detallada sobre el impacto relativo de cada gas de efecto invernadero en comparación con el CO2 en un periodo de tiempo que, por los datos ofrecidos por las distintas organizaciones, se estandariza a 100 años, por lo que nos queda un grafo como el que se muestra a continuación.

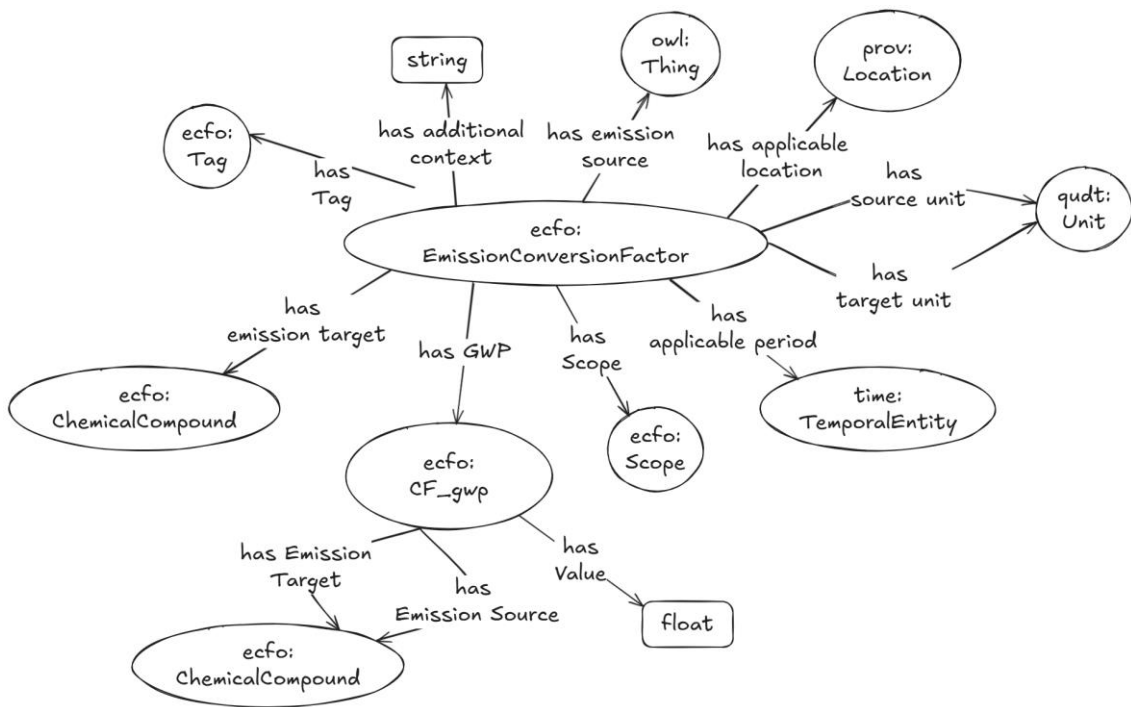


Figura 9. Estructura actual del grafo de conocimiento

Como se puede observar, se ha añadido el término GWP que contiene un valor, que indica el valor inicial del factor de conversión previo al procesamiento, y los elementos químicos relacionados con el factor de emisión y su resultado final.

Capítulo 5

5. Resultados y conclusiones

En este capítulo se presentan los resultados y las conclusiones personales que se han obtenido a lo largo de este Trabajo de Fin de Grado (TFG).

5.1. Resultados

Los resultados obtenidos durante este TFG son múltiples y abarcan desde el desarrollo técnico hasta investigación aplicada en el ámbito de la emisión de gases de efecto invernadero y se pueden encontrar en el siguiente repositorio [20]. Dichos resultados se describen a continuación:

- **Automatización de procesos:** bajo un proceso de ETL se ha conseguido optimizar y agilizar los procesos de recolección, limpieza y transformación, reduciendo así la intervención humana y minimizando el error humano. Este enfoque ha permitido obtener resultados más estandarizados que refuerzan la consistencia e integridad de los datos.
- **Integración de nuevas fuentes:** se ha conseguido integrar las siguientes fuentes EPA (por parte del gobierno estadounidense) y NAEI (del gobierno británico). Esto ha permitido enriquecer el grafo de conocimiento existente proporcionando factores de conversión más actualizados y diversos, ofreciendo nuevas actividades y localizaciones geográficas y temporales.
- **Extensión del grafo de conocimiento:** debido a la adición del término GWP se ha tenido que extender el grafo de conocimiento para terminar de completar la ontología. Esto ha permitido mayor interoperabilidad al vincular los datos ya existentes con los nuevos datos lo que fortalece el sistema y permite incorporar nuevos conceptos de manera eficiente y manteniendo el esquema actual.
- **Enlazado de la base de conocimiento con Wikidata:** uno de los procesos de la automatización de procesos era la conexión de los conceptos con Wikidata para su total entendimiento. Para ello se ha creado una tabla con el concepto y la url correspondiente, para poder enlazarlo en los scripts y así obtener un archivo homogeneizado y completo.

5.2. Limitaciones del proyecto

- **Dependencia de fuentes externas:** la precisión de los datos depende directamente de la actualización periódica de los datos proporcionados por las agencias mencionadas.

- **Complejidad semántica:** la representación de relaciones complejas en el grafo puede dificultar la visualización de alguno de los factores de conversión.

5.3. Propuestas de mejora

- **Integración de nuevas fuentes:** incorporar datos de nuevas regiones para ampliar la riqueza de datos del sistema.
- **Desarrollo de una interfaz de usuario:** crear una herramienta visual que permita explorar el grafo y realizar el cálculo transparente de manera interactiva.

5.4. Conclusiones personales

A lo largo de este Trabajo de Fin de Grado (TFG) he aprendido numerosos conceptos y ha sido un desafío y experiencia enriquecedora. A continuación, presento mis conclusiones personales:

- **Aplicación de conocimientos:** la realización de este TFG me ha impulsado a poner en práctica los conocimientos teóricos adquiridos durante la carrera, especialmente en el área del desarrollo de software. Ha sido un reto muy enriquecedor que he disfrutado bastante.
- **Desarrollo de habilidades:** algunas de las aptitudes que he conseguido desarrollar durante la realización de este proyecto son gestión de problemas de un proyecto, la correcta toma de decisiones o la adaptación a nuevas tecnologías y conceptos. Todos estos conceptos me van a ayudar a crecer en el ámbito profesional.
- **Aprendizaje del uso de grafos de conocimiento:** previo a la realización del TFG no tenía nociones de uso de grafos de conocimientos. Una vez realizado el proyecto, he adquirido notables competencias en el entendimiento y gestión de un grafo de conocimiento y la actualización de los datos que hay dentro.
- **Conocimientos técnicos en el área:** este proyecto me ha obligado a estudiar competencias técnicas en el ámbito de la emisión de gases de efecto invernadero (GEI). Alguno de los conceptos mencionados son el Scope, los distintos niveles o el uso del GWP. Sin el estudio previo de estos conceptos no hubiera sido posible la realización correcta de este proyecto.

Capítulo 6

6. Análisis de Impacto

Este capítulo se dedica a examinar el impacto general de los resultados obtenidos durante la realización del Trabajo de Fin de Grado (TFG). Además, se evalúa cómo el trabajo realizado puede contribuir a los Objetivos de Desarrollo Sostenible (ODS) establecidos en la Agenda 2030 de las Naciones Unidas.

6.1. Impacto general

Este Trabajo de Fin de Grado (TFG) ha tenido un impacto exponencial en varios ámbitos, incluido el contexto personal, social, económico y medioambiental. A continuación, se detalla el impacto positivo de cada uno de ellos.

6.1.1. Impacto personal

Realizar este TFG ha sido una gran oportunidad para mejorar mis habilidades de programación y crecer profesionalmente, al poner en práctica los conocimientos adquiridos durante mi formación académica. Este proyecto también me ha fortalecido habilidades esenciales como la gestión del tiempo, ya que he tenido que compatibilizar el *Prácticum* con el proyecto, documentación técnica, durante la elaboración de la memoria final, y otros conocimientos técnicos, como los grafos de conocimiento o las ontologías, las cuales desconocía hasta el momento.

6.1.2. Impacto social

En cuanto al impacto social, es *vox populi* que los gases de efecto invernadero (GEI) tiene grandes efectos negativos en la atmósfera. El aumento de las temperaturas, consecuencia directa de las actividades que generan estos gases, pueden crear fenómenos atmosféricos como sequías, incrementación del nivel del mar, etc. Aunque el proyecto no pretende solucionar el problema del cambio climático, su contribución proporciona unos cálculos más transparentes y precisos sobre las emisiones de gases de efecto invernadero y obtener una estimación más fiable sobre los impactos ambientales asociados a las distintas actividades descritas en los datos, lo que facilita la toma de decisiones.

6.1.3. Impacto económico

Desde una perspectiva económica, este proyecto puede ayudar a las empresas a acceder a certificaciones de sostenibilidad (como las que ofrece AENOR [21]) e incluso oportunidades de inversión por parte de fondos de inversión sostenibles.

6.1.4. Impacto medioambiental

Como se ha adelantado en el Impacto social, la emisión de carbono a la atmósfera tiene grandes repercusiones sobre el medioambiente. Este proyecto ofrece cálculos más transparentes lo que permite a las organizaciones identificar las actividades con mayor emisión y controlarlas de manera más eficiente. Este control permite evitar la degradación de ecosistemas, reduciendo así la alteración de ciclos naturales y pérdida de biodiversidad.

6.2. Objetivos de desarrollo sostenible

Este Trabajo de Fin de Grado (TFG) también tiene potencial para ayudar a varios Objetivos de Desarrollo Sostenible (ODS) de la Agenda 2030. Se analiza a continuación algunos de los objetivos que cumple este proyecto:

- **ODS 7: Energía Asequible y No Contaminante:** los resultados de los factores de conversión pueden ser un limitante para las organizaciones para emitir gases de efecto invernadero (GEI) como bien estipulan algunas de las leyes que limitan la emisión de dichos gases (por ejemplo: Ley de Cambio Climático y Transición Energética, Protocolo de Kioto o el Acuerdo de París), lo que lleva a las organizaciones a tomar acciones ecologistas y tomar conciencia del problema.
- **ODS 13: Acción por el Clima:** como se viene describiendo anteriormente, la gran mayoría de las actividades que se recogen en los datos emiten gases de efecto invernadero (GEI), que en una cantidad desmesurada puede conllevar el desgaste de la atmósfera. Por eso es necesario actuar de inmediato para evitar más desastres naturales causados a raíz del cambio climático.
- **ODS 14: Vida Submarina:** el océano supone el 75% de la superficie de nuestro planeta y contienen el 97% del agua de la Tierra y nos proporcionan múltiples recursos naturales esenciales para subsistir. Las emisiones de carbono afectan directamente a los océanos. Se estima que el 23% de las emisiones anuales de CO₂ generadas por actividades humanas son absorbidas por los océanos y contribuyen a mitigar los efectos del cambio climático, además absorben el 90% del exceso de calor, como podemos observar en las subidas del nivel del mar.
- **ODS 15: Vida de Ecosistema Terrestre:** como consecuencia directa de las emisiones de gases de efecto invernadero (GEI) como el carbono, se produce un aumento de la temperatura, que produce cambios en múltiples ecosistemas. Sin irnos muy lejos, en España tenemos el ejemplo de las migraciones de las aves ibéricas. Se estima que, debido al aumento de las temperaturas en 2024, la cigüeña blanca llega a España 36 días antes de su fecha normal. Como consecuencia en ecosistema puede llevar a una mayor demanda de alimentos y un desbalance de la disponibilidad de los recursos para otras especies. [22]

Bibliografía

- [1] ClimatePartner GmbH, «Equivalente de dióxido de carbono (CO2e),» ClimatePartner GmbH, [En línea]. Available: <https://www.climatepartner.com/es/conocimientos/glosario/equivalent-e-de-dioxido-de-carbono>.
- [2] M. Markovic, D. Garijo, S. Germano y I. Naja , «TEC: Transparent Emissions Calculation Toolkit. In: Payne, T.R., et al. The Semantic Web – ISWC 2023. ISWC 2023. Lecture Notes in Computer Science, vol 14266. Springer, Cham.,» [En línea]. Available: https://doi.org/10.1007/978-3-031-47243-5_5.
- [3] [En línea]. Available: <https://w3id.org/ecfo#>.
- [4] «TEC Toolkit,» [En línea]. Available: <https://tec-toolkit.github.io/>.
- [5] «Python,» Web de Python, [En línea]. Available: <https://es.python.org/>.
- [6] «Pandas,» Web de Pandas, [En línea]. Available: <https://pandas.pydata.org/>.
- [7] «Wikidata,» Web de Wikidata, [En línea]. Available: https://www.wikidata.org/wiki/Wikidata:Main_Page.
- [8] «GitHub,» Web de GitHub, [En línea]. Available: <https://github.com/>.
- [9] A. Ramos Olivé, Repositorio del proyecto, [En línea]. Available: <https://github.com/TEC-Toolkit/cfkg>.
- [10] T. Khan y M. Goodwin, «IBM,» 11 Diciembre 2023. [En línea]. Available: <https://www.ibm.com/es-es/topics/yaml#:~:text=YAML%20es%20un%20lenguaje%20vers%C3%A1til,de%20interpretar%20para%20las%20m%C3%A1quinas..>

- [11] «United States Environmental Protection Agency,» [En línea]. Available: <https://www.epa.gov/climateleadership/scope-3-inventory-guidance#:~:text=Scope%203%20Resources-,Description%20of%20Scope%203%20Emissions,affects%20in%20its%20value%20chain..>
- [12] «unilog.company,» [En línea]. Available: <https://unilog.company/blog-articles/upstream-vs-downstream-supply-chain/#:~:text=Upstream%20supply%20chain%20is%20the,they%20do%20affect%20each%20other..>
- [13] EPA. [En línea]. Available: <https://www.epa.gov/climateleadership/scope-1-and-scope-2-inventory-guidance>.
- [14] «BBVA,» [En línea]. Available: <https://www.bbva.com/es/sostenibilidad/que-son-las-emisiones-de-alcance-1-2-y-3/>.
- [15] «IPCC,» [En línea]. Available: [https://archive.ipcc.ch/ipccreports/tar/wg1/247.htm#:~:text=The%20GWP%20has%20been%20defined,reference%20gas%20\(IPCC%2C%201990\)%3A](https://archive.ipcc.ch/ipccreports/tar/wg1/247.htm#:~:text=The%20GWP%20has%20been%20defined,reference%20gas%20(IPCC%2C%201990)%3A).
- [16] U. E. O. o. R. a. D. (ORD), «IPCC,» [En línea]. Available: <https://catalog.data.gov/dataset/ipcc-ar4-ar5-and-ar6-20-100-and-500-year-gwps>.
- [17] «EPA,» [En línea]. Available: <https://www.epa.gov/aboutepa/our-mission-and-what-we-do>.
- [18] «EPA,» [En línea]. Available: <https://www.epa.gov/climateleadership/ghg-emission-factors-hub>.
- [19] «NAEI,» [En línea]. Available: <https://naei.energysecurity.gov.uk/>.
- [20] A. R. Olivé, «GitHub,» [En línea]. Available: <https://github.com/adrianramosolive/cfkg>.
- [21] AENOR. [En línea]. Available: <https://www.aenor.com/conocenos/historia>.
- [22] I. Pintado, «ConSalud,» 21 Agosto 2023. [En línea]. Available: <https://www.consalud.es/animalcare/salud-animal/cambio-climatico->

adelanta-migracion-aves-ciguena-llega-espana-hasta-36-dias-
antes_133576_102.html.

Anexos

Código Python perteneciente a la automatización de procesos

El siguiente código pertenece a una fuente. Todas las fuentes presentan el mismo código para este apartado.

```
import pandas as pd
from datetime import datetime
import os

def get_formula(text):
    formulas = text.split(' per ')[0].strip() # Takes the part before
'per'
    if len(formulas) > 1:
        if 'kg' in formulas:
            formulas = formulas.replace('kg', '').strip() # Removes 'kg'
if present
        return formulas
    else:
        return formulas[0] # If not, returns the first

# Function to remove net/gross CV to add the Wikidata URL later
def clean_parentheses(text):
    if '(' in text:
        return text.split('(')[0].strip() # Takes the part before the
parentheses and removes spaces
    else:
        return text.strip() # If there are no parentheses, returns the
original text without additional spaces

# Modify the 'region' column based on 'emission_source'
def get_region(emission_source):
    if emission_source.startswith('Hotel_stay'):
        # Remove the 'Hotel_stay_' prefix and replace '_' with spaces
        return emission_source.replace('Hotel_stay_', '').replace('_', '
')
    else:
        return 'United Kingdom'

def nomenclature(formula):
    if formula is None or pd.isna(formula): # Checks if the value is
None or NaN
```

```

        return None
    if 'of' in formula: # Checks if 'of' is in the formula
        formula1, formula2 = formula.split(' of ') # Splits the string
into two parts
        # Perform mapping using the tuplas_nomen dictionary
        formula1_mapped = tuplas_nomen.get(formula1.strip(),
formula1.strip())
        formula2_mapped = tuplas_nomen.get(formula2.strip(),
formula2.strip())
        result = f"{formula1_mapped} of {formula2_mapped}"
        return result
    else:
        # Maps directly if the formula does not contain 'of'
        return tuplas_nomen.get(formula.strip(), formula.strip())

def calculate_gwp(row, year_dataset, value):
    # Filter by year
    relevant_gwp = df_GWP[(df_GWP['start'] <= year_dataset) &
(df_GWP['end'] >= year_dataset)]
    # Create a dictionary
    gwp_dict = dict(zip(relevant_gwp["emission_source"],
relevant_gwp["value"]))
    return gwp_dict.get(row["emission_target_formula_aux"], value)

def id_gwp(row, year_dataset, value):
    # Filter by year
    relevant_gwp = df_GWP[(df_GWP['start'] <= year_dataset) &
(df_GWP['end'] >= year_dataset)]
    # Create a dictionary
    gwp_dict = dict(zip(relevant_gwp["emission_source"],
relevant_gwp["id"]))
    return gwp_dict.get(row["emission_target_formula_aux"], value)

# Path to the Excel files directory
script_dir = os.path.dirname(os.path.abspath(__file__))
path = os.path.join(script_dir, "../data_raw/ghg-conversion-factors-2022-
flat-format.xlsx")
# Read the Excel file
df_raw = pd.read_excel(path, sheet_name='Factors by Category',
engine='openpyxl')

i = 0
headers = []
while 'Scope' not in headers:
    i += 1

```

```

    headers = df_raw.iloc[i].tolist()
df_raw = pd.read_excel(path, sheet_name='Factors by Category',
engine='openpyxl', header=i+1)

# Check if there is no 'ID' column and add it if missing
if not any(col.lower() == 'id' for col in headers):
    headers.insert(0, 'id')
df_final = pd.DataFrame(columns=headers)
if 'GHG/Unit' in headers:
    ghg_label = 'GHG/Unit'
else:
    ghg_label = 'GHG'

df_final.pop(ghg_label)

# Populate df_final with data from df_raw
for j in range(len(df_raw)):
    row = df_raw.iloc[j].values.tolist()
    if 'id' in df_final.columns: # If 'id' was manually added
        row.insert(0, None) # Insert None at the beginning to align with
'id'
    row = row[:len(headers)] # Ensure the row has the same length as
'headers'
    df_final.loc[j, headers[:len(row)]] = row
print("Initial lines", df_final.shape[0])

df_raw[ghg_label] = df_raw[ghg_label].astype(str)
column_name = [col for col in df_raw.columns if 'GHG Conversion Factor'
in col][0]
year_dataset = int(column_name.split()[-1])

df_final['Scope'] = df_final['Scope'].astype(str).str.replace(' ', ''),
regex=False)
df_final.insert(7, "emission_source", None)
df_final['emission_source'] = df_final['Level
2'].astype(str).str.replace(" ", "_", regex=False) + "_" +
df_final['Level 3'].astype(str).str.replace(" ", "_", regex=False)

# Load the Excel file with labels and URLs
path_labels = os.path.join(script_dir,
"../../auxiliary_op/unique_values_wikidata_urls.xlsx")
df_labels = pd.read_excel(path_labels, engine='openpyxl')
df_labels['label_url'] = df_labels['label_url'].str.strip()

# Create a dictionary of labels and their URLs

```

```

tuplas = dict(zip(df_labels['label'], df_labels['label_url']))

# Extract URL values for columns in df_final ending with '_wd'
for col in df_final.columns:
    if col.endswith('_wd'):
        # Get the base column without '_wd' and replace it with URLs from
        the dictionary
        base_col = col[:-3]
        df_final[col] = df_final[base_col].map(tuplas).fillna('-')

# ----- Other transformations -----

df_final['UOM'] = df_final['UOM'].fillna('').astype(str)
df_final['UOM'] = df_final['UOM'].apply(clean_parentheses)
print("Lines 1", df_final.shape[0])

# Create 'UOM_wd' column by mapping values of 'UOM' to the dictionary of
labels and URLs (tuplas)
df_final['UOM_wd'] = df_final['UOM'].map(tuplas)
# Reorder columns so 'UOM_wd' appears right after 'UOM'
columns = list(df_final.columns)
index_uom = columns.index('UOM')
columns.insert(index_uom + 1, columns.pop(columns.index('UOM_wd')))
df_final = df_final[columns]

# Extract and clean 'emission_target_formula'
df_final['emission_target_formula'] =
df_raw[ghg_label].apply(get_formula)
df_final['emission_target_formula'] =
df_final['emission_target_formula'].apply(lambda x: x.replace('kg',
'').strip() if isinstance(x, str) else x)
df_final['emission_target_formula'] =
df_final['emission_target_formula'].replace('nan', None)
# Clean parentheses values in both columns
df_final['emission_target_formula'] =
df_final['emission_target_formula'].str.replace("(Net CV)", "",
regex=False).str.replace("(Gross CV)", "",
regex=False).str.replace("(net)", "", regex=False).str.replace("kWh ",
"kWh", regex=False)
df_final[ghg_label] = df_raw[ghg_label].str.replace("(Net CV)", "",
regex=False).str.replace("(Gross CV)", "",
regex=False).str.replace("(net)", "", regex=False)

# Load the Excel file with formulas and their corresponding nomenclatures
path_nomen = os.path.join(script_dir, ".././auxiliary_op/formulas.xlsx")

```

```

df_nomen = pd.read_excel(path_nomen, engine='openpyxl')

#-----
-----

df_final['emission_target_formula_aux'] =
df_final['emission_target_formula'].str.split().str[-1]
# Create a dictionary of labels and URLs
tuplas_nomen = dict(zip(df_nomen['formula'], df_nomen['nomenclature']))
df_final['emission_target'] =
df_final['emission_target_formula'].apply(nomenclature)
df_final = df_final.dropna(subset=['emission_target_formula'])

df_final['emission_target_wd'] =
df_final['emission_target_formula'].map(tuplas)
print("Lines 2", df_final.shape[0])
#-----
-----

# Extract 'GHG' from the 'GHG/Unit' field
df_final['GHG'] = df_raw[ghg_label].str.split().str[0]
df_final['GHG'] = df_final['GHG'].replace('nan', None).str.replace(' per
unit', '', regex=False)
# Identify rows where df_final['GHG'] is equal to 'kWh'
rows_to_delete = df_final[df_final['GHG'] == 'kWh'].index
# Remove those rows
df_final = df_final.drop(rows_to_delete)
df_final['GHG_wd'] = df_final['GHG'].map(tuplas)
#-----
-----

# Set date validity columns
df_final['valid_from'] = datetime(year_dataset, 1, 1, 0, 0,
0).isoformat()
df_final['valid_to'] = datetime(year_dataset, 12, 31, 23, 59,
59).isoformat()
print("Lines 4", df_final.shape[0])
#-----
-----

# Apply the function to the 'region' column
df_final['region'] = df_final['emission_source'].apply(get_region)
df_final['region_wd'] = df_final['region'].map(tuplas)

```

```

#-----
#-----
df_final = df_final.rename(columns={f'GHG Conversion Factor
{year_dataset}': f'Conversion Factor {year_dataset}'})
# Ensure values are of string type to use .str.replace() and convert to
numeric
df_final[f'Conversion Factor {year_dataset}'] = df_final[f'Conversion
Factor {year_dataset}'].astype(str)
df_final[f'Conversion Factor {year_dataset}'] =
pd.to_numeric(df_final[f'Conversion Factor
{year_dataset}'].str.replace(', ', '', regex=False), errors='coerce')
print("Lines 5", df_final.shape[0])
#-----
#-----
# Drop rows with NaN in 'Conversion Factor {year_dataset}'
df_final = df_final.dropna(subset=[f'Conversion Factor {year_dataset}'])
print("Lines 6", df_final.shape[0])
# Move 'Conversion Factor {year_dataset}' to the last position
conversion_factor_column = f'Conversion Factor {year_dataset}'
columns = [col for col in df_final.columns if col !=
conversion_factor_column] + [conversion_factor_column]
df_final = df_final[columns]
#-----
#-----
# Rename columns and ensure they are stripped of whitespace
df_final.columns = df_final.columns.str.strip()
#-----
#-----
# GWP
path_GWP_values = os.path.join(script_dir, "../../GWP/GWP_values.xlsx")
df_GWP = pd.read_excel(path_GWP_values, engine='openpyxl')

df_final['GWP'] = df_final.apply(lambda row: calculate_gwp(row,
year_dataset, 1), axis=1)
df_final[f'Value {year_dataset}'] = df_final[f'Conversion Factor
{year_dataset}'] / df_final['GWP']
df_final['GWP'] = df_final.apply(lambda row: calculate_gwp(row,
year_dataset, ''), axis=1)
df_final['GWP_id'] = df_final.apply(lambda row: id_gwp(row, year_dataset,
''), axis=1)

```

```

#-----
#-----
# If 'id' was manually added in 'headers', generate ID values
if 'id' in df_final.columns and df_final['id'].isnull().all():
    df_final['id'] = range(1, len(df_final) + 1)

# Remove 'emission_target_formula_aux' column
df_final = df_final.drop(columns=['emission_target_formula_aux'])
#-----
#-----

# Save to Excel
final_path = os.path.join(script_dir,
f"../data/v3/Conversion_Factor_{year_dataset}_v3.csv")
df_final.to_csv(final_path, index=False)
print("Final lines", df_final.shape[0])

```

Mappings

Para cada año y cada fuente se presenta el mismo código yaml. A continuación, se presenta el mapping perteneciente a la fuente DESNZ y el año 2023.

```

prefixes:
  ecfo: "https://w3id.org/ecfo#"
  qudt: "http://qudt.org/schema/qudt/"
  skos: "http://www.w3.org/2004/02/skos/core#"
  prov: "http://www.w3.org/ns/prov#"
  time: "http://www.w3.org/2006/time#"
  schema: "https://schema.org/"
  ecfo_instance: "https://w3id.org/ecfkg/i/UK/DESNZ/2023/"
  ecfo_gwp: "https://w3id.org/ecfkg/i/GWP/"
  dc: "http://purl.org/dc/elements/1.1/"
  rdf: "http://www.w3.org/1999/02/22-rdf-syntax-ns#"

sources:
  main-source: ['data/Conversion_Factor_2023.xlsx']
  gwp-source: ['GWP/GWP_values.xlsx']
mappings:
  CF:
    sources:
      - main-source
    s: ecfo_instance:CF_$(id)
  po:
    - [a, ecfo:EmissionConversionFactor]

```

```

- [ecfo:hasAdditionalContext, $(Column Text), en~lang]
- [ecfo:hasScope, https://w3id.org/ecfo#$(Scope)~iri]
- [rdf:value, $(Value 2023),xsd:float]
- [ecfo:hasTag, https://w3id.org/ecfkg/i/$(Level 1)~iri]
- [ecfo:hasTag, https://w3id.org/ecfkg/i/$(Level 2)~iri]
- [ecfo:hasTag, https://w3id.org/ecfkg/i/$(Level 3)~iri]
- [ecfo:hasTag, https://w3id.org/ecfkg/i/$(Level 4)~iri]
- [ecfo:hasSourceUnit, $(UOM_wd)~iri]
- [ecfo:hasEmissionSource,
https://w3id.org/ecfkg/i/$(emission_source)~iri]
- [ecfo:hasTargetUnit, $(GHG_wd)~iri]
- [ecfo:hasEmissionTarget, $(emission_target_wd)~iri]
- [ecfo:hasApplicablePeriod,
ecfo_instance:applicablePeriod/$(valid_from)/$(valid_to)~iri]
- [ecfo:hasApplicableLocation, $(region_wd)~iri]
- [ecfo:hasGWP, ecfo_gwp:$(GWP_id)~iri]
- [dc:publisher, https://w3id.org/ecfkg/i/Organization/DESNZ~iri]
- [prov:wasDerivedFrom,
https://assets.publishing.service.gov.uk/media/649c536906179b00113f74a9/g
hg-conversion-factors-2023-flat-file-update.xlsx, xsd:anyURI]

```

Agent:

sources:

- main-source

s: "https://w3id.org/ecfkg/i/Organization/DESNZ"

po:

- [a, prov:Organization]
- [rdfs:label, 'Department of Energy Security and Net Zero (DESNZ)', en~lang]
- [rdfs:seeAlso,

'https://www.gov.uk/government/organisations/department-for-energy-security-and-net-zero',xsd:anyURI]

TimeStart:

sources:

- main-source

s: ecfo_instance:applicablePeriod/\$(valid_from)/\$(valid_to)/start

po:

- [a, time:Instant]
- [time:inXSDDate, \$(valid_from),xsd:dateTime]

TimeEnd:

sources:

- main-source

s: ecfo_instance:applicablePeriod/\$(valid_from)/\$(valid_to)/end

```

    po:
      - [a, time:Instant]
      - [time:inXSDDate, $(valid_to),xsd:dateTime ]

TemporalEntity:
  sources:
    - main-source
  s: ecfo_instance:applicablePeriod/$(valid_from)/$(valid_to)
  po:
    - [a, time:TemporalEntity]
    - [time:hasBeginning,
ecfo_instance:applicablePeriod/$(valid_from)/$(valid_to)/start~iri]
    - [time:hasEnd,
ecfo_instance:applicablePeriod/$(valid_from)/$(valid_to)/end~iri]

EmissionSource:
  sources:
    - main-source
  s: https://w3id.org/ecfkg/i/$(emission_source)
  po:
    - [a , skos:Concept]
    - [rdfs:label, $(emission_source), en~lang]

EmissionTarget:
  sources:
    - main-source
  s: $(emission_target_wd)
  po:
    - [a , ecfo:ChemicalCompound]
    - [rdfs:label, $(emission_target), en~lang]
    - [schema:molecularFormula, $(emission_target_formula), en~lang]

L1:
  sources:
    - main-source
  s: https://w3id.org/ecfkg/i/$(Level 1)
  po:
    - [a , skos:Concept]
    - [rdfs:label, $(Level 1), en~lang]

L2:
  sources:
    - main-source
  s: https://w3id.org/ecfkg/i/$(Level 2)
  po:

```

```

- [a , skos:Concept]
- [rdfs:label, $(Level 2), en~lang]

L3:
sources:
- main-source
s: https://w3id.org/ecfkg/i/$(Level 3)
po:
- [a , skos:Concept]
- [rdfs:label, $(Level 3), en~lang]

L4:
sources:
- main-source
s: https://w3id.org/ecfkg/i/$(Level 4)
po:
- [a , skos:Concept]
- [rdfs:label, $(Level 4), en~lang]

sourceUnit:
sources:
- main-source
s: $(UOM_wd)
po:
- [a , qudt:Unit]
- [rdfs:label, $(UOM), en~lang]

targetUnit:
sources:
- main-source
s: $(GHG_wd)
po:
- [a , qudt:Unit]
- [rdfs:label, $(GHG), en~lang]

GWP:
sources:
- gwp-source
s: ecfo_gwp:$(id)
po:
- [rdf:value, $(value)]
- [ecfo:hasEmissionSource, $(emission_source_wd)~iri]
- [ecfo:hasEmissionTarget, $(emission_target_wd)~iri]
- [prov:wasDerivedFrom, $(source)~iri]

```

```
- [ecfo:hasApplicablePeriod,  
ecfo_instance:applicablePeriod/$(start)/$(end)~iri]
```

Para esta parte del proyecto, al alumno ha tenido que implementar la parte correspondiente al GWP y sus respectivas relaciones.

Preprocesamiento de las nuevas fuentes

Para el preprocesamiento de la nueva fuente EPA, se ha tenido que pasar por un proceso de ajuste de formato, para así poder aplicar el Código Python perteneciente a la automatización de procesos para su correcto funcionamiento. El código descrito es el siguiente:

```
import re  
import pandas as pd  
import warnings  
import os  
import numpy as np  
  
# Function to remove text inside parentheses  
def clean_parentheses(text):  
    if '(' in text:  
        return text.split('(')[0].strip()  
    else:  
        return text.strip()  
  
# Ignore specific warnings  
warnings.filterwarnings("ignore")  
  
# Path to the Excel file  
script_dir = os.path.dirname(os.path.abspath(__file__))  
path = os.path.join(script_dir, "../data_raw/ghg-emission-factors-hub-  
2024.xlsx")  
df_raw = pd.read_excel(path, sheet_name='Emission Factors Hub',  
engine='openpyxl')  
  
# -----  
# -----  
# TABLE 1 (Scope 1)  
df = df_raw.iloc[15:90, 2:] # Extract Table 1  
df.columns = df_raw.iloc[14, 2:] # Set header  
df = df.reset_index(drop=True)  
  
# Initialize `ghg_units` with the units GHG/Unit
```

```

ghg_units = df_raw.iloc[15, 2:].tolist()

# Create empty final DataFrame
df_final = pd.DataFrame(columns=['id', 'Scope', 'Level 1', 'Level 2',
'Level 3', 'Level 4',
                                'Column Text', 'UOM', 'GHG/Unit',
'Conversion Factor 2024'])

# Temporary variable to store the current category for Level 2
current_level_2 = None

# Process rows of Table 1
for index, row in df.iterrows():
    if pd.isna(row.iloc[0]): # If the row starts empty
        ghg_units = row.iloc[1:].tolist() # Update `ghg_units`
    elif pd.notna(row.iloc[0]) and row.iloc[1:].isna().all():
        current_level_2 = row.iloc[0] # Update Level 2 category
    else:
        # Create rows for each value in `GHG/Unit`
        if pd.notna(row.iloc[0]):
            for i, ghg_unit in enumerate(ghg_units):
                conversion_factor = row.iloc[i + 1] if
pd.notna(row.iloc[i + 1]) else None
                new_row = {
                    'id': None,
                    'Scope': "Scope 1",
                    'Level 1': "Stationary Combustion",
                    'Level 2': current_level_2 if current_level_2 else
"",
                    'Level 3': row.iloc[0],
                    'Level 4': None,
                    'Column Text': None,
                    'UOM': 'Metric Tons',
                    'GHG/Unit': ghg_unit,
                    'Conversion Factor 2024': conversion_factor
                }
                df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

# Assign unique IDs
df_final['id'] = range(1, len(df_final) + 1)
print(df_final['id'].max())
# -----
# -----
# TABLE 3 (Scope 1)

```

```

df3 = df_raw.iloc[123:235, 2:] # Extract Table 3
df3.columns = df_raw.iloc[122, 2:]
df3 = df3.loc[:, ~df3.columns.isna()] # Remove empty columns

current_level_1 = None
current_level_2 = 'On-Road Gasoline Vehicles'
column_text = None
last_id = df_final['id'].max() # Get last used ID

# Process rows of Table 3 and add directly to df_final
for index, row in df3.iterrows():
    if pd.notna(row.get('Vehicle Type')):
        try:
            column_text, current_level_1 = row['Vehicle Type'].split(' ',
1)
        except ValueError:
            column_text = row['Vehicle Type']
            current_level_1 = ""

    level_3 = row.get('Model Year')

    # Create rows for each GHG/Unit
    for ghg_unit, conversion_factor in zip(df3.columns[-2:], row[-2:]):
        ghg_unit = re.search(r'\((.*?)\)', ghg_unit).group(1)
        new_row = {
            'id': df_final['id'].max() + 1,
            'Scope': "Scope 1",
            'Level 1': current_level_1,
            'Level 2': current_level_2,
            'Level 3': level_3,
            'Level 4': None,
            'Column Text': column_text,
            'UOM': 'Metric Tons',
            'GHG/Unit': ghg_unit,
            'Conversion Factor 2024': conversion_factor
        }
        df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

print(df_final['id'].max())
# -----
# -----
# TABLE 4 (Scope 1)
df4 = df_raw.iloc[241:275, 2:] # Extract Table 4
df4.columns = df_raw.iloc[240, 2:]

```

```

df4 = df4.loc[:, ~df4.columns.isna()] # Remove empty columns

level_1 = None
level_2 = 'On-Road Diesel and Alternative Fuel Vehicles'
level_3 = None
column_text = None

for index, row in df4.iterrows():
    if pd.notna(row.get('Vehicle Type')):
        level_1 = row.get('Vehicle Type')
    if pd.notna(row.get('Fuel Type')):
        column_text = row.get('Fuel Type')
    if pd.notna(row.get('Model Year')):
        level_3 = row.get('Model Year')
    else:
        level_3 = ''

    for ghg_unit, conversion_factor in zip(df4.columns[-2:], row[-2:]):
        ghg_unit = re.search(r'\((.*?)\)', ghg_unit).group(1)
        new_row = {
            'id': df_final['id'].max() + 1,
            'Scope': "Scope 1",
            'Level 1': level_1,
            'Level 2': level_2,
            'Level 3': level_3,
            'Level 4': None,
            'Column Text': column_text,
            'UOM': 'Metric Tons',
            'GHG/Unit': ghg_unit,
            'Conversion Factor 2024': conversion_factor
        }
        df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

print(df_final['id'].max())
# -----
# TABLE 5 (Scope 1)
df5 = df_raw.iloc[282:322, 2:] # Extract Table 5
df5.columns = df_raw.iloc[281, 2:]
df5 = df5.loc[:, ~df5.columns.isna()] # Remove empty columns

level_1 = None
level_2 = 'Non-Road Vehicles'
column_text = None

```

```

for index, row in df5.iterrows():
    if pd.notna(row.get('Vehicle Type')):
        level_1 = row.get('Vehicle Type')
    if pd.notna(row.get('Fuel Type')):
        column_text = row.get('Fuel Type')

    for ghg_unit, conversion_factor in zip(df5.columns[-2:], row[-2:]):
        ghg_unit = re.search(r'\((.*?)\)', ghg_unit).group(1)
        new_row = {
            'id': df_final['id'].max() + 1,
            'Scope': "Scope 1",
            'Level 1': level_1,
            'Level 2': level_2,
            'Level 3': None,
            'Level 4': None,
            'Column Text': column_text,
            'UOM': 'Metric Tons',
            'GHG/Unit': ghg_unit,
            'Conversion Factor 2024': conversion_factor
        }
        df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

print(df_final['id'].max())
# -----
# TABLE 6 (Scope 2)
df6 = df_raw.iloc[330:359, 1:7]
columns_primary = df_raw.iloc[329, 1:7]
columns_secondary = df_raw.iloc[330, 1:7]
df6 = df6.loc[:, ~df6.columns.isna()]
df6.columns = pd.MultiIndex.from_tuples(zip(columns_primary,
columns_secondary))

level_2 = None
level_3 = None
for index, row in df6.iterrows():
    if pd.notna(row.get(('eGRID Subregion Name', np.nan))):
        level_3 = row[('eGRID Subregion Name', np.nan)]
        secondary_headers = df6.columns.get_level_values(1)

        for ghg_unit, conversion_factor in
zip(secondary_headers[~pd.isna(secondary_headers)][-3:], row[-3:]):
            ghg_unit = re.search(r'\((.*?)\)', ghg_unit).group(1)

```

```

new_row = {
    'id': df_final['id'].max() + 1,
    'Scope': "Scope 2",
    'Level 1': "Electricity",
    'Level 2': level_2,
    'Level 3': level_3,
    'Level 4': None,
    'Column Text': 'Total Output Emission Factors',
    'UOM': 'Metric Tons',
    'GHG/Unit': ghg_unit,
    'Conversion Factor 2024': conversion_factor
}
df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

print(df_final['id'].max())
# -----
# -----
# TABLE 8 (Scope 3)
df8 = df_raw.iloc[411:418, 2:]
df8.columns = df_raw.iloc[410, 2:]
df8 = df8.loc[:, ~df8.columns.isna()]

level_1 = None
level_2 = 'Downstream Transportation and Distribution'

for index, row in df8.iterrows():
    if pd.notna(row.get('Vehicle Type')):
        level_1 = row.get('Vehicle Type')

    for ghg_unit, conversion_factor in zip(df8.columns[-4:], row[1:4]):
        unit = row[df8.columns[-1]]
        ghg_unit = re.search(r'\((.*?)\)', ghg_unit).group(1)
        new_row = {
            'id': df_final['id'].max() + 1,
            'Scope': "Scope 3",
            'Level 1': level_1,
            'Level 2': level_2,
            'Level 3': None,
            'Level 4': None,
            'Column Text': None,
            'UOM': unit,
            'GHG/Unit': ghg_unit,
            'Conversion Factor 2024': conversion_factor
        }

```

```

        df_final = pd.concat([df_final, pd.DataFrame([new_row])],
ignore_index=True)

print(df_final['id'].max())
# -----
# EXPORT
df_final = df_final.sort_values(by='id').reset_index(drop=True)
final_path = os.path.join(script_dir, f"../data_raw/EPA_raw.xlsx")
df_final.to_excel(final_path, index=False)

```

Nueva ontología

Se han tenido que añadir algunos detalles para completar la nueva ontología.

```

@prefix : <https://w3id.org/ecfo#> .
@prefix org: <http://www.w3.org/ns/org#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix xml: <http://www.w3.org/XML/1998/namespace> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix foaf: <http://xmlns.com/foaf/0.1/> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@base <https://w3id.org/ecfo> .

<https://w3id.org/ecfo> rdf:type owl:Ontology ;
                        owl:versionIRI <https://w3id.org/ecfo/1.0.0> ;
                        <http://purl.org/dc/elements/1.1/description>
"\"This ontology defines a vocabulary for describing carbon emission
conversion factors (CF). These are values typically used to calculate
carbon emissions where the CF multiplies a quantified estimate of the
energy (e.g., kWh of electricity, liters of fuel, etc.) used by a
particular activity.\"@en ;
                        <http://purl.org/dc/elements/1.1/title> "The
Emission Conversion Factor Ontology"@en ;
                        <http://purl.org/dc/terms/abstract> "With the
Net Zero agenda gaining significant traction across the world,
organisations are often required to report carbon emissions associated
with their operation. However, calculating emissions is a non-trivial
task and reported scores can differ depending on the choices made by
those performing the calculations or the software used to assist with
this task. Emission conversion factors are used to calculate greenhouse

```

```

emissions for different business operations. This ontology aims to
describe these conversion factors and their associated properties (e.g.,
origin, applicable location, etc.) to enhance the transparency of the
carbon footprint calculations."@en ;
    <http://purl.org/dc/terms/created> "April 20th,
2023"@en ;
    <http://purl.org/dc/terms/creator>
<https://orcid.org/0000-0001-6634-3266> ,
    <https://orci
d.org/0000-0001-6993-0618> ,
    <https://orci
d.org/0000-0002-5477-287X> ,
    <https://orci
d.org/0000-0003-0454-7145> ;
    <http://purl.org/dc/terms/issued> "May 8th,
2023"@en ;
    <http://purl.org/dc/terms/license>
<https://creativecommons.org/licenses/by/4.0/> ;
    <http://purl.org/dc/terms/modified> "May 8th,
2023"@en ;
    <http://purl.org/dc/terms/publisher> "TEC-
Toolkit on Github"@en ;
    <http://purl.org/ontology/bibo/status> "TEC
Toolkit specification (stable)"@en ;
    <http://purl.org/vocab/vann/preferredNamespacePr
efix> "ecfo"@en ;
    <http://purl.org/vocab/vann/preferredNamespaceUr
i> "https://w3id.org/ecfo#" ;
    rdfs:comment "1.0.0 release of the Emission
Conversion Factor Ontology"@en ;
    owl:priorVersion <https://w3id.org/ecfo/0.0.2> ;
    owl:versionInfo "1.0.0"@en ;
    <https://schema.org/imag
e> <https://raw.githubusercontent.com/TEC-
Toolkit/ECFO/21b4380e1f84f85f7cbc59f724157498e9d72094/Logo%20ECFO.svg> ;
    <https://schema.org/includedInDataCatalog>
<https://lov.linkeddata.es/dataset/lov/vocabs/ecfo> ;
    <https://schema.org/logo>
<https://raw.githubusercontent.com/TEC-
Toolkit/ECFO/21b4380e1f84f85f7cbc59f724157498e9d72094/Logo%20ECFO.svg> .

#####
#   Annotation properties
#####

```

```
### http://purl.org/dc/elements/1.1/description
<http://purl.org/dc/elements/1.1/description> rdf:type
owl:AnnotationProperty .

### http://purl.org/dc/elements/1.1/title
<http://purl.org/dc/elements/1.1/title> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/abstract
<http://purl.org/dc/terms/abstract> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/created
<http://purl.org/dc/terms/created> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/creator
<http://purl.org/dc/terms/creator> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/description
<http://purl.org/dc/terms/description> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/issued
<http://purl.org/dc/terms/issued> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/license
<http://purl.org/dc/terms/license> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/modified
<http://purl.org/dc/terms/modified> rdf:type owl:AnnotationProperty .

### http://purl.org/dc/terms/publisher
<http://purl.org/dc/terms/publisher> rdf:type owl:AnnotationProperty .

### http://purl.org/ontology/bibo/status
<http://purl.org/ontology/bibo/status> rdf:type owl:AnnotationProperty .
```

```

### http://purl.org/vocab/vann/preferredNamespacePrefix
<http://purl.org/vocab/vann/preferredNamespacePrefix> rdf:type
owl:AnnotationProperty .

### http://purl.org/vocab/vann/preferredNamespaceUri
<http://purl.org/vocab/vann/preferredNamespaceUri> rdf:type
owl:AnnotationProperty .

### http://www.w3.org/2008/05/skos#prefLabel
<http://www.w3.org/2008/05/skos#prefLabel> rdf:type
owl:AnnotationProperty ;
                                rdfs:subPropertyOf rdfs:label
.

### http://www.w3.org/ns/org#memberOf
org:memberOf rdf:type owl:AnnotationProperty .

### http://xmlns.com/foaf/0.1/homepage
foaf:homepage rdf:type owl:AnnotationProperty .

### http://xmlns.com/foaf/0.1/name
foaf:name rdf:type owl:AnnotationProperty .

### https://schema.org/image
<https://schema.org/image> rdf:type owl:AnnotationProperty .

### https://schema.org/includedInDataCatalog
<https://schema.org/includedInDataCatalog> rdf:type
owl:AnnotationProperty .

### https://schema.org/logo
<https://schema.org/logo> rdf:type owl:AnnotationProperty .

#####
#   Datatypes
#####

```

```

### http://www.w3.org/1999/02/22-rdf-syntax-ns#HTML
rdf:HTML rdf:type rdfs:Datatype .

### http://www.w3.org/2001/XMLSchema#date
xsd:date rdf:type rdfs:Datatype .

#####
# Object Properties
#####

### http://purl.org/dc/elements/1.1/publisher
<http://purl.org/dc/elements/1.1/publisher> rdf:type owl:ObjectProperty ;
                                             rdfs:comment "An entity
responsible for making the resource available."@en ;
                                             rdfs:isDefinedBy
<http://purl.org/dc/elements/1.1/> ;
                                             rdfs:label "publisher"@en .

### http://www.w3.org/2006/time#hasBeginning
<http://www.w3.org/2006/time#hasBeginning> rdf:type owl:ObjectProperty ;
                                             rdfs:domain
<http://www.w3.org/2006/time#TemporalEntity> ;
                                             rdfs:range
<http://www.w3.org/2006/time#Instant> ;
                                             rdfs:comment "Beginning of a
temporal entity."^^xsd:anyURI ;
                                             rdfs:isDefinedBy
<http://www.w3.org/2006/time> ;
                                             rdfs:label "has
beginning"^^xsd:anyURI .

### http://www.w3.org/2006/time#hasEnd
<http://www.w3.org/2006/time#hasEnd> rdf:type owl:ObjectProperty ;
                                       rdfs:domain
<http://www.w3.org/2006/time#TemporalEntity> ;
                                       rdfs:range
<http://www.w3.org/2006/time#Instant> ;
                                       rdfs:comment "End of a temporal
entity."@en ;
                                       rdfs:isDefinedBy
<http://www.w3.org/2006/time> ;

```

```

                                rdfs:label "has end"@en .

### http://www.w3.org/ns/prov#atLocation
<http://www.w3.org/ns/prov#atLocation> rdf:type owl:ObjectProperty ;
                                rdfs:domain
<http://www.w3.org/ns/prov#Entity> ;
                                rdfs:range
<http://www.w3.org/ns/prov#Location> ;
                                rdfs:comment "A location can be an
identifiable geographic place (ISO 19112), but it can also be a non-
geographic place such as a directory, row, or column. As such, there are
numerous ways in which location can be expressed, such as by a
coordinate, address, landmark, and so forth."@en ;
                                rdfs:isDefinedBy
<https://www.w3.org/ns/prov-o> ;
                                rdfs:label "atLocation"@en .

### http://www.w3.org/ns/prov#wasDerivedFrom
<http://www.w3.org/ns/prov#wasDerivedFrom> rdf:type owl:ObjectProperty ;
                                rdfs:domain
<http://www.w3.org/ns/prov#Entity> ;
                                rdfs:range
<http://www.w3.org/ns/prov#Entity> ;
                                rdfs:comment "A derivation is
a transformation of an entity into another, an update of an entity
resulting in a new one, or the construction of a new entity based on a
pre-existing entity."@en ;
                                rdfs:isDefinedBy
<http://www.w3.org/ns/prov#> ;
                                rdfs:label "was derived
from"@en .

### https://w3id.org/ecfo#hasApplicableLocation
:hasApplicableLocation rdf:type owl:ObjectProperty ;
                                rdfs:subPropertyOf
<http://www.w3.org/ns/prov#atLocation> ;
                                rdfs:domain :EmissionConversionFactor ;
                                rdfs:comment "A geographic area (e.g., a country)
or a physical location (e.g., a farm) to which a specific
ecfo:ConversionFactor may be applied."@en ;
                                rdfs:label "has applicable location"@en .

```

```

### https://w3id.org/ecfo#hasApplicablePeriod
:hasApplicablePeriod rdf:type owl:ObjectProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range
<http://www.w3.org/2006/time#TemporalEntity> ;
    rdfs:comment "A time period for which the emission
conversion factor is considered to be most relevant."@en ;
    rdfs:label "has applicable period"@en .

### https://w3id.org/ecfo#hasEmissionSource
:hasEmissionSource rdf:type owl:ObjectProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range owl:Thing ;
    rdfs:comment "Source of the emission in a conversion
factor (e.g., butane, refrigerator, herd of animals, etc.)."@en ;
    rdfs:label "has emission source"@en .

### https://w3id.org/ecfo#hasEmissionTarget
:hasEmissionTarget rdf:type owl:ObjectProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range :ChemicalCompound ;
    rdfs:comment "Target chemical compound of a conversion
factor (e.g., CO2e). The target compound is associated with the target
unit"@en ;
    rdfs:label "has emission target"@en .

### https://w3id.org/ecfo#hasScope
:hasScope rdf:type owl:ObjectProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range :Scope ;
    rdfs:comment "A property linking conversion factor to the
description of the emission scope."@en ;
    rdfs:label "has scope"@en .

### https://w3id.org/ecfo#hasSourceUnit
:hasSourceUnit rdf:type owl:ObjectProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range <http://qudt.org/schema/qudt/Unit> ;
    rdfs:comment "The unit corresponding to the source data
that is multiplied by the ecfo:ConversionFactor"@en ;
    rdfs:label "has source unit"@en .

```

```

### https://w3id.org/ecfo#hasTag
:hasTag rdf:type owl:ObjectProperty ;
      rdfs:domain :EmissionConversionFactor ;
      rdfs:range :Tag ;
      rdfs:comment "A property to assign a specific category/tag to
ecfo:ConversionFactor."@en ;
      rdfs:label "has tag"@en .

### https://w3id.org/ecfo#hasTargetUnit
:hasTargetUnit rdf:type owl:ObjectProperty ;
      rdfs:domain :EmissionConversionFactor ;
      rdfs:range <http://qudt.org/schema/qudt/Unit> ;
      rdfs:comment "The unit corresponding to the result of
multiplying input data by the ecfo:ConversionFactor"@en ;
      rdfs:label "has target unit"@en .

#####
# Data properties
#####

### http://qudt.org/schema/qudt/abbreviation
<http://qudt.org/schema/qudt/abbreviation> rdf:type owl:DatatypeProperty
;
      rdfs:range xsd:string ;
      rdfs:comment "An abbreviation
for a unit is a short ASCII string that is used in place of the full name
for the unit in contexts where non-ASCII characters would be problematic,
or where using the abbreviation will enhance readability. When a power of
abase unit needs to be expressed, such as squares this can be done using
abbreviations rather than symbols. For example, sq ft means square foot,
and cu ft means cubic foot."@en ;
      rdfs:isDefinedBy
<https://qudt.org/schema/qudt/> ;
      rdfs:label "abbreviation"@en .

### http://www.w3.org/1999/02/22-rdf-syntax-ns#value
rdf:value rdf:type owl:DatatypeProperty ;
      rdfs:comment "Idiomatic property used for structured values"@en
;
      rdfs:isDefinedBy rdf: ;

```

```

    rdfs:label "value"@en .

### http://www.w3.org/2006/time#inXSDDate
<http://www.w3.org/2006/time#inXSDDate> rdf:type owl:DatatypeProperty ;
    rdfs:domain
<http://www.w3.org/2006/time#Instant> ;
    rdfs:range xsd:date ;
    rdfs:comment "Position of an
instant, expressed using xsd:date" ;
    rdfs:isDefinedBy
"http://www.w3.org/2006/time" ;
    rdfs:label "in XSD date" .

### https://schema.org/molecularFormula
<https://schema.org/molecularFormula> rdf:type owl:DatatypeProperty ;
    rdfs:comment "The empirical formula
is the simplest whole number ratio of
all the atoms in a molecule."@en ;
    rdfs:isDefinedBy
<https://schema.org/> ;
    rdfs:label "molecular formula"@en .

### https://w3id.org/ecfo#hasAdditionalContext
:hasAdditionalContext rdf:type owl:DatatypeProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range xsd:string ;
    rdfs:comment ""Conversion factors often have
subteltlies that contextualize the emission source. For example the
emission source may be a transportation method (car), but additional
context may indicate that the CF is for \"large cars\", or lars with more
than 120 horse power.
These subteltlies are captured in the \"hasAdditionalContext\"
property""@en ;
    rdfs:label "has additional context"@en .

### https://w3id.org/ecfo#hasUsageNotes
:hasUsageNotes rdf:type owl:DatatypeProperty ;
    rdfs:domain :EmissionConversionFactor ;
    rdfs:range xsd:string ;
    rdfs:comment "Property used to describe any limitations,
disclaimers or usage notes associated with a CF. For example \"This CF is

```

```

an estimation based on local data and should be superseded by
governmental sources\""@en ;
    rdfs:label "has usage notes"@en .

#####
#   Classes
#####

### http://qudt.org/schema/qudt/Unit
<http://qudt.org/schema/qudt/Unit> rdf:type owl:Class ;
    rdfs:comment "A unit of measure, or
unit, is a particular quantity value that has been chosen as a scale for
measuring other quantities the same kind (more generally of equivalent
dimension). For example, the meter is a quantity of length that has been
rigorously defined and standardized by the BIPM (International Board of
Weights and Measures). Any measurement of the length can be expressed as
a number multiplied by the unit meter. More formally, the value of a
physical quantity Q with respect to a unit (U) is expressed as the scalar
multiple of a real number (n) and U, as \\(Q = nU\\)." ;
    rdfs:isDefinedBy
<http://qudt.org/schema/qudt/> ;
    rdfs:label "Unit" .

### http://www.w3.org/2006/time#Instant
<http://www.w3.org/2006/time#Instant> rdf:type owl:Class ;
    rdfs:comment "A temporal entity
with zero extent or duration" ;
    rdfs:isDefinedBy
<http://www.w3.org/2006/time> ;
    rdfs:label "Time instant" .

### http://www.w3.org/2006/time#TemporalEntity
<http://www.w3.org/2006/time#TemporalEntity> rdf:type owl:Class ;
    rdfs:comment "A temporal
interval or instant."^^xsd:anyURI ;
    rdfs:isDefinedBy
<http://www.w3.org/2006/time> ;
    rdfs:label "Temporal
entity"^^xsd:anyURI .

### http://www.w3.org/2008/05/skos#Concept

```

```

<http://www.w3.org/2008/05/skos#Concept> rdf:type owl:Class ;
                                         rdfs:comment "An abstract idea
or notion; a unit of thought." ;
                                         rdfs:isDefinedBy
<http://www.w3.org/2008/05/skos> ;
                                         rdfs:label "Concept" .

### http://www.w3.org/ns/prov#Agent
<http://www.w3.org/ns/prov#Agent> rdf:type owl:Class ;
                                   rdfs:comment "An agent is something
that bears some form of responsibility for an activity taking place, for
the existence of an entity, or for another agent's activity." ;
                                   rdfs:isDefinedBy
<https://www.w3.org/ns/prov-o> ;
                                   rdfs:label "Agent" .

### http://www.w3.org/ns/prov#Entity
<http://www.w3.org/ns/prov#Entity> rdf:type owl:Class ;
                                   rdfs:comment "An entity is a physical,
digital, conceptual, or other kind of thing with some fixed aspects;
entities may be real or imaginary." ;
                                   rdfs:isDefinedBy
<https://www.w3.org/ns/prov-o> ;
                                   rdfs:label "Entity" .

### http://www.w3.org/ns/prov#Location
<http://www.w3.org/ns/prov#Location> rdf:type owl:Class ;
                                       rdfs:comment "A location can be an
identifiable geographic place (ISO 19112), but it can also be a non-
geographic place such as a directory, row, or column. As such, there are
numerous ways in which location can be expressed, such as by a
coordinate, address, landmark, and so forth." ;
                                       rdfs:isDefinedBy
<https://www.w3.org/ns/prov-o> ;
                                       rdfs:label "Location" .

### http://xmlns.com/foaf/0.1/Organization
foaf:Organization rdf:type owl:Class .

### https://w3id.org/ecfo#ChemicalCompound

```

```

:ChemicalCompound rdf:type owl:Class ;
                  rdfs:comment "Pure chemical substance consisting of two
or more different chemical elements (source:
https://www.wikidata.org/wiki/Q11173)"@en ;
                  rdfs:label "Chemical compound"@en .

### https://w3id.org/ecfo#EmissionConversionFactor
:EmissionConversionFactor rdf:type owl:Class ;
                          rdfs:subClassOf
<http://www.w3.org/ns/prov#Entity> ;
                          rdfs:comment "A concept representing the
coefficient value used in greenhouse gas (GHG) emission calculations
(activity data x emission conversion factor = GHG emissions) and
additional metadata (e.g., applicable location, publisher, etc.)." ;
                          rdfs:label "Emission Conversion Factor" .

### https://w3id.org/ecfo#Scope
:Scope rdf:type owl:Class ;
        rdfs:comment "Type of emissions a company creates in its own
operations and in its wider 'value chain'. Typically emissions fall under
three categories - Scope 1, Scope 2 and Scope 3." ;
        rdfs:label "Scope" .

### https://w3id.org/ecfo#Tag
:Tag rdf:type owl:Class ;
     rdfs:subClassOf <http://www.w3.org/2008/05/skos#Concept> ;
     rdfs:comment "A generic class to describe a category/tag that
ecfo:EmissionFactor. Use ecfo:hasTag to link the category/tag and
rdf:value to provide the name of the category. A single
ecfo:EmissionFactor may refer to multiple tags. To identify individual
conversion factors a union of tags should be considered." ;
     rdfs:label "Tag" .

### https://w3id.org/ecfo#GWPConversionFactor
:GlobalWarmingPotentialConversionFactor rdf:type owl:Class ;
    rdfs:comment "Concept representing a coefficient value used to
transform chemical substances according to their Global Warming
Potential"@en ;
    rdfs:label "GWP Conversion Factor"@en .

### https://w3id.org/ecfo#hasGWP
:gwp rdf:type owl:ObjectProperty ;

```

```

    rdfs:domain :EmissionConversionFactor ;
    rdfs:range :GlobalWarmingPotentialConversionFactor ;
    rdfs:comment "GWP property"@en ;
    rdfs:label "has GWP"@en .

#####
#   Individuals
#####

### https://w3id.org/ecfo#Scope1
:Scope1 rdf:type owl:NamedIndividual ,
         :Scope ;
    rdfs:comment "Scope 1 (direct emissions) emissions are those from
activities owned or controlled by your organisation. Examples of Scope 1
emissions include emissions from combustion in owned or controlled
boilers, furnaces and vehicles; and emissions from chemical production in
owned or controlled process equipment." ;
    rdfs:label "scope 1" .

### https://w3id.org/ecfo#Scope2
:Scope2 rdf:type owl:NamedIndividual ,
         :Scope ;
    rdfs:comment "Scope 2 (energy indirect) emissions are those
released into the atmosphere that are associated with your consumption of
purchased electricity, heat, steam and cooling. These indirect emissions
are a consequence of your organisation's energy use, but occur at sources
you do not own or control." ;
    rdfs:label "scope 2" .

### https://w3id.org/ecfo#Scope3
:Scope3 rdf:type owl:NamedIndividual ,
         :Scope ;
    rdfs:comment "Scope 3 (other indirect) emissions are a
consequence of your actions that occur at sources you do not own or
control and are not classed as Scope 2 emissions. Examples of Scope 3
emissions are business travel by means not owned or controlled by your
organisation, waste disposal, materials or fuels your organisation
purchases. Deciding if emissions from a vehicle, office or factory that
you use are Scope 1 or Scope 3 may depend on how you define your
operational boundaries. Scope 3 emissions can be from activities that are
upstream or downstream of your organisation. More information on Scope 3
and other aspects of reporting can be found in the Greenhouse Gas
Protocol Corporate Standard." ;

```

```

rdfs:label "scope 3" .

### https://www.abdn.ac.uk#this
<https://www.abdn.ac.uk#this> rdf:type owl:NamedIndividual ,
                                foaf:Organization ;
                                foaf:homepage "https://www.abdn.ac.uk/" ;
                                foaf:name "University of Aberdeen" .

### https://www.open.ac.uk#this
<https://www.open.ac.uk#this> rdf:type owl:NamedIndividual ,
                                foaf:Organization ;
                                foaf:homepage "https://www.open.ac.uk" ;
                                foaf:name "The Open University" .

### https://www.ox.ac.uk#this
<https://www.ox.ac.uk#this> rdf:type owl:NamedIndividual ,
                                foaf:Organization ;
                                foaf:homepage "https://www.ox.ac.uk/" ;
                                foaf:name "University of Oxford" .

### https://www.upm.es#this
<https://www.upm.es#this> rdf:type owl:NamedIndividual ,
                                foaf:Organization ;
                                foaf:homepage "https://www.upm.es" ;
                                foaf:name "Universidad Politécnica de Madrid" .

#####
# Annotations
#####

<https://orcid.org/0000-0001-6634-3266> foaf:name "Iman Naja"@en ;
                                         org:memberOf

<https://www.open.ac.uk#this> .

<https://orcid.org/0000-0001-6993-0618> org:memberOf
<https://www.ox.ac.uk#this> ;
                                         foaf:name "Stefano Germano"@en .

```


```
<https://orcid.org/0000-0002-5477-287X> org:memberOf
<https://www.abdn.ac.uk#this> ;
                                foaf:name "Milan Markovic"@en .

<https://orcid.org/0000-0003-0454-7145> foaf:name "Daniel Garijo"@en ;
                                org:memberOf
<https://www.upm.es#this> .

### Generated by the OWL API (version 4.5.9.2019-02-01T07:24:44Z)
https://github.com/owlcs/owlapi
```

Para esta parte del código, el alumno ha tenido que introducir los conceptos de GWP con sus respectivas relaciones a la ontología.

Este documento esta firmado por



Firmante	CN=tfgm.fi.upm.es, OU=CCFI, O=ETS Ingenieros Informaticos - UPM, C=ES
Fecha/Hora	Tue Jan 14 19:14:19 CET 2025
Emisor del Certificado	EMAILADDRESS=camanager@etsiinf.upm.es, CN=CA ETS Ingenieros Informaticos, O=ETS Ingenieros Informaticos - UPM, C=ES
Numero de Serie	561
Metodo	urn:adobe.com:Adobe.PPKLite:adbe.pkcs7.sha1 (Adobe Signature)