Universidad Politécnica
de Madrid

Escuela Técnica Superior de
Ingenieros Informáticos

Máster Universitario en Ingeniería Informática

Memoria de Final del Trabajo de Fin de Máster

Taxonomía de Fuentes de Información Aplicables para la Optimización de Indicadores de Seguridad

Autor(a): Caro Gómez, Celia

Co-Tutor(a): Benito Gómez, Mariano José

Tutor(a): Tovar Caro, Edmundo

Madrid, 16 de junio de 2020

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

Trabajo Fin de Máster
Máster Universitario en Ingeniería Informática
Título: Taxonomía de Fuentes de Información Aplicables Para la Optimización de Indicadores de Seguridad
Junio 2020

Autor(a): Celia Caro Gómez

Tutor(a):
Edmundo Tovar Caro
Departamento de Inteligencia Artificial
ETSI Informáticos
Universidad Politécnica de Madrid

Co-Tutor(a):
Mariano José Benito Gómez
CISO de GMV Soluciones Globales e Internet S.A.U

Agradecimientos

A ti, la vida porque…

Tengo la suerte, más que de ver, de poder mirar bien; de ganar siempre; de querer vivir mi presente y de contemplar lo más simple como un milagro; de perder la cordura, perderme en la cordura y salir sana y salva con aire desaliñado; de no usar brújulas pues mi corazón empuja mis pies; de querer matar de amor; de tenerme y conocerte; de ver mi día como un lienzo en blanco; de que la música me salve; de que mi boca abrace y mis abrazos besen; de ser agua y de sentir fuego; de salvar con las palabras y de viajar entre el polvo de páginas de libros viejos.

Mariano J. Benito, gracias por estar en mi presente, por ser fuente de conocimiento y poder aprender de ti curiosidades que matarían a gatos.

Familia, os amo y os juro que no os pierdo y perderé en ningún tiempo verbal.

Resumen

Este TFM es un trabajo de investigación que explora las posibilidades que ofrece disponer de una taxonomía de fuentes de información como herramienta útil para mejorar en costes del cálculo del cuadro de mandos de indicadores de la seguridad de la información. La motivación de este trabajo de investigación reside en el enfoque tan generalizado por parte de los estándares sobre la evaluación de la efectividad de los controles de la seguridad de la información y las fuentes de información empleadas. Esto hace que la mayoría de cuadros de mandos sean de lo más variopintos, alejados de la realidad y en algunas ocasiones un cuadro imposible de interpretar. Este trabajo pretende desacoplar el concepto de cuadro de mandos en diferentes niveles de abstracción con el fin de llegar a su nivel más bajo, las fuentes de información, y entender las piezas clave que harán que un cuadro de mandos sea sostenible y de calidad. Las fuentes de información son una parte fundamental, una pieza clave en el cálculo de indicadores.

El lector conocerá algunos de los diferentes estándares y marcos de trabajo más recurridos hoy en día por las empresas, verá que se ha realizado una búsqueda de recursos útiles a través de estos estándares para el desarrollo de este trabajo, cosa que ha sido una ardua tarea, y podrá encontrarse con todos los pasos seguidos para la construcción justificada de la taxonomía de fuentes de información hasta ahora utilizadas, la validación de esta con fuentes de información nuevas y la medición de su utilidad en un cuadro de mandos de un caso real a través de escenarios que han ido apareciendo mientras este trabajo de investigación se iba desarrollando. Por el enfoque investigador del trabajo, el lector debe ver el desarrollo de la taxonomía como una prueba que se puede convertir en una oportunidad de mejora, más que como una herramienta ya probada.

Por último, el lector podrá ver que este trabajo de investigación no acaba aquí y puede enlazarse con la recurrencia en diferentes áreas de la informática, como lo son la Inteligencia Artificial y Big Data, y tecnologías prometedoras de oportunidades de mejoras y transformaciones de procesos de negocio como los son el aseguramiento de la información y la tediosa tarea del cálculo del cuadro de mandos.

Abstract

This TFM is a research work that explores the potential usefulness of a taxonomy of information sources as a useful tool to improve the cost of calculating the information security scorecard. The motivation for this research work lies in the widespread approach of standards on evaluating the effectiveness of information security controls and the sources of information used. This makes the majority of scorecards very diverse, far from reality and sometimes an impossible to interpret picture. This work aims to decouple the concept of the scorecard at different levels of abstraction in order to reach its lowest level, the sources of information, and understand the key pieces that will make a scorecard sustainable and of quality. Information sources are a fundamental part, a key piece in the calculation of indicators.

The reader will know some of the different standards and frameworks most used by companies today, will see that a search for useful resources has been made through these standards for the development of this work, which has been an arduous task , and you will be able to find all the steps followed for the justified construction of the taxonomy of information sources used up to now, the validation of this with new information sources and the measurement of its usefulness in a scorecard of a real case through of scenarios that have appeared while this research work was developing. Please, take into account the research nature of this work: the reader should see it as a test that can become an opportunity for improvement, rather than an useful, field-tested tool.

Finally, the reader will be able to see that this research work does not end here and can be linked to recurrence in different areas of computing, such as Artificial Intelligence and Big Data, and promising technologies for opportunities for improvement and transformation of business processes such as the assurance of information and the tedious task of calculating the scorecard.

Tabla de contenidos

Listado de figuras

Listado de tablas

1 Introducción

La **Seguridad de la Información** es uno de los temas que últimamente está ganando más protagonismo y que a su vez provoca varios comederos de cabeza, en mayor o en menor medida, a muchas de las organizaciones que existen hoy en día. ¿Qué empresa no tiene actualmente equipamiento informático y de telecomunicaciones para almacenar, recuperar, transmitir y manipular sus datos? El **valor** de la mayoría de las empresas recae en sus datos y el valor de este **activo**, por ende se han de **proteger**. Pero, ¿qué supone la protección de estos datos? La protección de estos datos supone definir el alcance de protección de los activos de la organización de los que los activos de tipo dato dependen. ¿Cómo es definido este alcance? o mejor dicho, ¿cómo la organización discierne qué activos asegurar o no? Aquí, tiene cabida el término **amenaza**. Una amenaza es cualquier elemento o acción que pueda atentar contra la seguridad de la información y aparece si algún activo sufre de algún tipo de **vulnerabilidad**. Una amenaza puede provocar una cierta **degradación** en un activo con una cierta **frecuencia**. Los términos degradación y frecuencia están directamente relacionados con la amenaza y son los que permiten que tanto el **impacto** como el **riesgo** sean estimados. Existen muchos tipos de amenazas e incluso se podría decir que hay amenazas para cada tipo de activo, pero al igual que existen amenazas también existen las **salvaguardas**. Las salvaguardas son medidas **mitigadoras del riesgo**, **limitadoras del impacto** y **reductoras de la ocurrencia de amenazas**. Si las salvaguardas son seleccionadas y aplicadas correctamente, la organización alcanzará un determinado nivel o estado de seguridad mejor que sin su aplicación. Por lo tanto, con un buen **análisis y gestión de riesgos** una empresa asegura que los servicios y continuidad de su negocio no se vean afectados, de lo contrario, esta empresa, podría acarrear pérdidas de liquidez significativas o en el peor de los casos caer en la quiebra.

Un **SGSI** o **Sistema de Gestión de la Seguridad de la Información**, es un sistema que, no tiene porque pero, suele estar basado en un estándar o norma con el objetivo de asegurar que las organizaciones tienen implementadas ciertas características que se definen en el estándar para proteger la información de los stakeholders (*interesados* en español). Uno de los estándares más reconocidos, es la norma **ISO/IEC 27001:2013**, perteneciente a la familia de estándares **ISO/IEC 27000**, y junto a otros marcos teóricos de la seguridad de la información como **NIST 800-55 Rev.1:2008**, va a ser estudiado para el desarrollo del presente trabajo de investigación. ISO/IEC 27001:2013, tiene el objetivo de asegurar que **las empresas implementan todos los controles adecuados sobre las dimensiones**, principalmente entre otras, **de confidencialidad**, **integridad** y **disponibilidad**. Este estándar proporciona un marco de SGSI en el que implementar los principios de ***planificar***, ***hacer***, ***verificar*** y ***actuar***, además de ***los procesos de un sistema de gestión***. El análisis y gestión de riesgos es uno de los procesos más relevantes de un SGSI porque permite dar una representación cuantitativa del **nivel de riesgo** o alarma de la organización, identificar los activos más vulnerables y decidir cuáles salvaguardas aplicar para mitigar el riesgo y obtener así un valor de **riesgo residual** que demuestre que el estado de riesgo de la organización es el permisible. El nivel de riesgo deseado siempre será de riesgo cero.

Conocer el nivel de riesgo permite saber hacia dónde se debería enfocar la empresa o a qué activos debería prestar más atención, pero ¿cómo sabe la

empresa que alcanzará el nivel de riesgo residual? ¿Cómo sabe la empresa que lo está haciendo bien o tiene que mejorar?, ¿Ha elegido correctamente los controles a aplicar? ¿Está aplicando adecuadamente los controles de seguridad? ¿Cómo de eficaz es su SGSI? ¿Cuál es el estado de seguridad de la organización? ¿Con qué frecuencia debería la empresa representar y conocer este estado de seguridad? Todas las anteriores preguntas se responden si la empresa ***monitoriza, mide, analiza y evalúa la seguridad de su SGSI****,* por lo que es el momento de adentrar un poco más al lector en el tema que va a comprender todo el presente trabajo.

Muchas empresas que deciden implantar o tienen implantado un SGSI, cuando tienen que proceder a la medición de su seguridad, piden o dependen de los estándares de la industria. Manejar el riesgo, en términos de seguridad en este caso, de forma cuantitativa y fácil de entender es bastante complejo. Entonces, sea cual sea el marco teórico que las empresas elijan, este les proporciona un estándar con el cual comparar cómo lo están haciendo. De aquí, la recurrencia a **cuadros de mando** para medir sus prácticas de seguridad de la información. Un cuadro de mando es una herramienta de medición, en este caso de seguridad, y según el portal del Instituto Nacional de Ciberseguridad (INCIBE), ayuda a "*definir, diseñar e implementar indicadores y alarmas de seguridad tanto a nivel de gestión como de tecnología, por lo que se podrá verificar si se están cumpliendo los requisitos y objetivos de seguridad."* Llevar a cabo el cálculo de indicadores tiene varios problemas que ningún estándar tiene en cuenta, y es que suponen:

- **Coste en cálculo**: Esfuerzo presupuestario dedicado al cálculo de indicadores.
- **Precisión del indicador**: Capacidad del indicador de sostener una decisión.
- **Utilidad/Eficiencia del indicador**: El indicador permite tomar las mejores decisiones.
- **Representatividad**: Cómo se asegura tener todos los indicadores que se necesitan y que describan correctamente el estado de la organización.
- **Comparabilidad**: Cómo de comparable es el estado de seguridad de la empresa con estados de seguridad anteriormente calculados y con los de otras organizaciones.

Por dichas razones, las organizaciones toman sus propias decisiones respecto al cálculo de indicadores. Por otro lado, los cuadros de mando si bien son útiles como herramienta de evaluación también están listos para su mal uso. Dada la *Ley de Campbell, "El riesgo de corrupción de un indicador social es proporcional a la intensidad de su uso para la toma de decisiones."* Lo explica muy bien el catedrático emérito de la UCM en su artículo, Carabaña, Julio. (2019). "¿Ley de Campbell o duendes informáticos?". *El País.* En términos de cuadro de mando quiere decir que, cuanto más es usado cualquier indicador cuantitativo para la toma de decisiones, más sujeto estará a las presiones de corrupción y más apto será para distorsionar y corromper los procesos para los que está destinado a monitorear.

Por su propia naturaleza, los marcos teóricos o estándares han sido definidos de forma general para que las diferentes empresas se apoyen en ellos, pero no pueden definir casos particulares porque la complejidad de manejo en un solo

documento de todas las casuísticas que podrían llegar a darse para cada una de las inversiones en seguridad particulares de cada una de las empresas, sería muy alta. La postura de generalizar por parte de los estándares ha hecho que los estudios empíricos sobre la efectividad de los estándares o marcos de trabajo y los cuadros de mando sean limitados. Un cuadro de mando propio de una empresa puede diferir en gran medida de otro cuadro de mando propio de otra empresa y aun así, las dos empresas cumplir con el estándar. También, cada uno de estos cuadros de mando puede sufrir deficiencias y tener un cierto *grado de eficacia o efectividad* en mayor o menor medida con el estándar. Este *grado de eficacia o efectividad* se ve representado por la salida del cuadro de mando, con los *indicadores de seguridad.* Un indicador debe ser objetivo, sin embargo, la definición de la métrica de este tiene una naturaleza heterogénea que la hace subjetiva. Este peso recae en que la cantidad de atributos medibles relacionados con la seguridad de la información que puede poseer la empresa, puede ser abrumadora. No es del todo obvio saber cuáles atributos deben medirse. Con demasiados atributos o atributos incorrectos sería impracticable, costoso y contraproducente el medir, analizar e informar y aun así, atributos clave podrían seguir estando ocultos dentro de un gran volumen de información u omitirse por completo si no se implementan las medidas adecuadas. Este es un problema que, en el siguiente párrafo con una propuesta de solución, este trabajo tiene la intención de paliar. Se puede deducir entonces, que el marco de trabajo, al ser tan general, es útil para cualquier SGSI pero no se puede inferir que todos los cuadros de mando son igualmente efectivos con el estándar. Por ello, la creación de un cuadro de mando es una tarea compleja sobre la cual el personal responsable de la seguridad de la información de la organización tendrá que expresar su lado más creativo e ingenieril con el fin, y desde la honestidad, representar la realidad del estado de seguridad de su organización.

Se ha visto que el que la empresa tenga carta de libertad a la hora de crear su propio cuadro de mando puede desencadenar una situación falsa obteniendo un estado de seguridad de su SGSI inservible para la buena toma de decisiones y que esto al final recae en la cantidad masiva de atributos medibles que posee una organización que pueden ser inservibles o imprecisos y provenir de fuentes de información dispersas, heterogéneas y desagregadas, es decir, fuentes de información poco favorables para el posterior cálculo de indicadores. Va a ser objeto de este trabajo resolver este problema facilitando y agilizando la creación de cuadros de mando propios de las empresas a través de la creación de una **taxonomía de fuentes de información** existentes que ayude al cálculo de **indicadores de seguridad de la información** que se encontrarán representados en dichos cuadros de mando. Con una buena taxonomía de fuentes de información, se gana precisión en la posterior selección de indicadores de seguridad, en la elección del buen indicador, ya que al tener una fuente de información bien clasificada se aseguran ciertas características en los datos salientes de dicha fuente de información que serán combinados y empleados en las métricas de los indicadores. Además, la aplicación de la taxonomía consigue que, aunque la cantidad exagerada de atributos medibles que la empresa puede manejar siga existiendo, la empresa se asegura que, sea cual sea el conjunto de atributos medibles seleccionados, siempre sean válidos. El fin de este trabajo es que los indicadores que se utilizaban antes de la taxonomía, después de esta se calculen mejor y más rápido. Por ende, para realizar la taxonomía se estudiarán los marcos teóricos de la seguridad de la información más recurridos hoy en día en busca de

aspectos de apoyo para clasificar las fuentes de información existentes con la finalidad de facilitar el cálculo de indicadores de seguridad de la información.

2 Estado del Arte

Uno de los estándares o marcos de trabajo más reconocidos hoy en día y sobre el que el presente trabajo se va a respaldar es ISO/IEC 27001:2013. Este estándar pertenece a la serie de normas de la familia ISO/IEC 27000. Esta familia reúne una serie de normas relacionadas con la seguridad de la información. En este caso el estándar ISO/IEC 27001:2013, se compone de un conjunto de requisitos contra los cuales se puede auditar el SGSI de una organización. También existe la norma **ISO/IEC 27002:2013** que proporciona, con un nivel de detalle mayor que ISO/IEC 27001:2013, un código de buenas prácticas que si son implantadas correctamente en una organización podrá conseguir la certificación en la norma ISO/IEC 27001:2013. Pero si lo que se busca en estas normas es un apoyo sobre cómo monitorizar y medir un SGSI estas dos normas no ayudan de mucho porque describen qué se debe hacer pero no el cómo hacerlo. Por otro lado, de la familia ISO/IEC 27000, existe la norma **ISO/IE 27004:2016** que explica que "*proporciona orientación sobre la forma de evaluar el desempeño de la norma ISO/IEC 27001:2013"* y si se buscan otros marcos fuera de la *ISO (International Organization for Standardization),* existe la *NIST (National Institute of Standards and Technology)* y su estándar ***NIST 800-55 Rev.1:2008***. Según la *NIST,* este estándar "*proporciona orientación sobre cómo una organización, mediante el uso de métricas, identifica la idoneidad de los controles, políticas y procedimientos de seguridad, proporciona un enfoque para ayudar a decidir dónde invertir en recursos de protección de seguridad adicionales o identificar y evaluar controles no productivos y explica el proceso métrico de desarrollo e implementación y cómo también se puede utilizar para justificar adecuadamente las inversiones en control de seguridad.*" Además, destaca que "*los resultados de un programa métrico eficaz pueden proporcionar datos útiles para dirigir la asignación de recursos de seguridad de la información y deberían simplificar la preparación de informes relacionados con el rendimiento.*"

Tanto el estándar *ISO/IEC 27004:2016* como el estándar *NIST 800-55 Rev.1:2008* han sido examinados en busca de información que favorezca el desarrollo de una taxonomía de fuentes de información, pudiendo comentar lo siguiente de cada uno de los estándares:

- **ISO/IEC 27004:2016:** La *Ilustración 1* representa la traza de requisitos de la norma ISO/IEC 27001:2013 con las cláusulas más detalladas de la norma ISO/IEC 27004:2016.

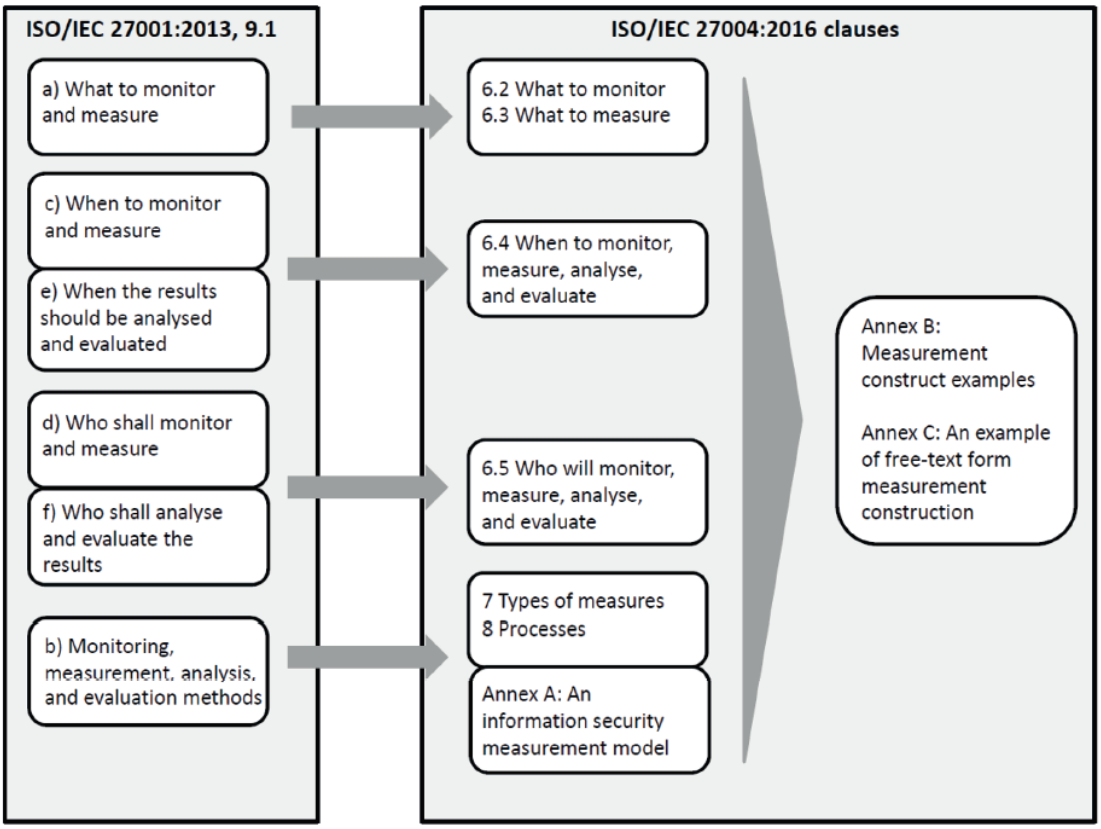


Ilustración 1. Mapping to ISO/IEC 27001:2013, 9.1 requirements

ISO/IEC 27004:2016 describe que las organizaciones deben tener en cuenta que hay una serie de cláusulas en ISO/IEC 27001:2013 que requieren explícitamente que se determine la efectividad de alguna actividad y que para hacer esto la organización debe definir primero una necesidad de información apropiada y una medida, o medidas, para satisfacerla. También, nos informa de los procesos y actividades de nuestro SGSI que podemos medir y a través de las cláusulas del capítulo 6 describe qué monitorear, qué medir, analizar y evaluar y con qué frecuencia hacerlo. En el capítulo 7 se afirma que el desempeño de las actividades planificadas y la efectividad de sus resultados se pueden medir mediante dos tipos de medidas, las medidas de desempeño, que son las medidas que demuestran el progreso de implementación de procesos del SGSI, procedimientos asociados y controles de seguridad específicos y las medidas de efectividad, que señalan la medida en que las actividades planificadas se han realizado y los resultados previstos han sido alcanzados. En este capítulo también se ponen como ejemplos orientativos, algunos datos que pueden considerarse útiles para estos dos tipos de medidas. Por último, a través de todo el capítulo 8, se describe cada una de las fases del proceso de monitorización, medición, análisis y evaluación del SGSI. Entre todas las fases que componen este proceso una de las fases que has sido examinada más en detalle es la fase de *8.3 Creación y mantenimiento de medidas,* pero solo orienta al lector, a través de ejemplos de datos salientes de fuentes de información que pueden recogerse para apoyar al cálculo de medidas de seguridad.

En conclusión sobre esta norma, podemos decir que proporciona un marco de trabajo para el proceso de monitorización, medición, análisis y evaluación de los SGSI pero no describe explícitamente los atributos intrínsecos que deberían de caracterizar las fuentes de información e indicadores o incluso, problemas existentes de las fuentes de información, que es lo que se intenta encontrar.

- **NIST 800-55 Rev.1:2008:** De este estándar los capítulos que se encuentran más interesantes para su estudio en este trabajo son el capítulo 5 de Proceso de Desarrollo de Medidas y el capítulo 6 de Implementación de Medidas de Seguridad de la Información. El capítulo 5 se puede resumir en la Ilustración 2.

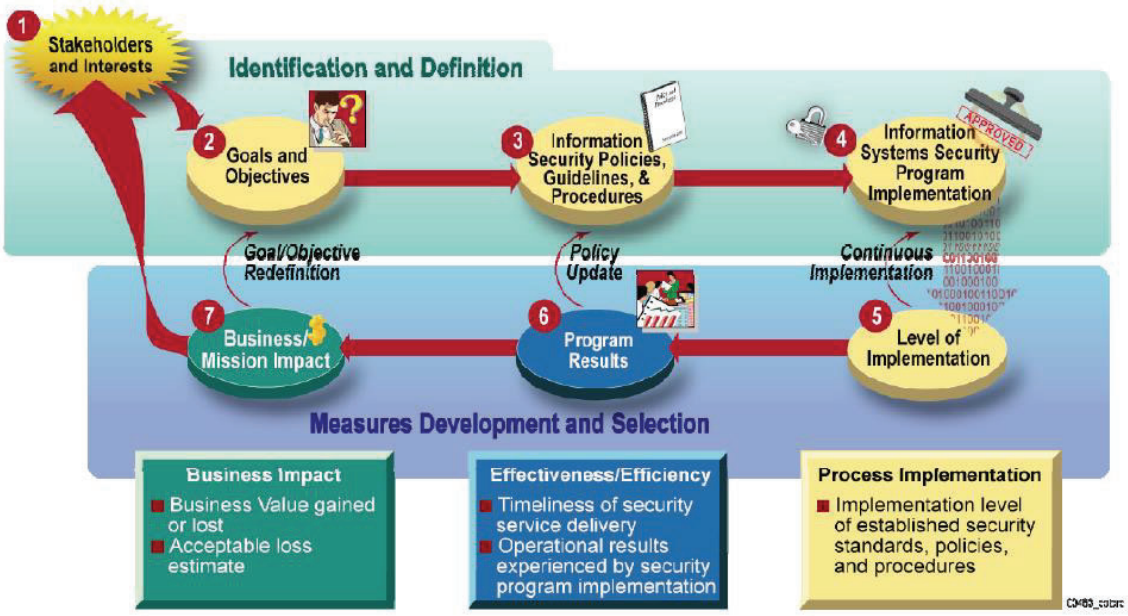


Ilustración 2. Information Security Measures Development Process

La anterior ilustración describe muy bien de forma gráfica el proceso de desarrollo de medidas de seguridad de la información. La primera parte de identificación y definición se compone de las fases 1, 2 y 3 que describen respectivamente que cada stakeholder requiere de medidas customizadas que representen el desempeño de la seguridad de la organización dentro de su área de responsabilidad, que se han de identificar y documentar las metas y objetivos del desempeño de la seguridad del sistema de información que guiarán la implementación de los controles de seguridad, como los controles de seguridad que deberían ser implementados constituyen las políticas y procedimientos de la organización que definen la línea base para el SGSI y que cualquier medida existente y repositorio de datos que pueda usarse para derivar datos de medidas debe ser revisada. La segunda parte de desarrollo y selección de medidas que implica desarrollar medidas que sigan la implementación del proceso, la eficiencia/efectividad y el impacto de la misión, se compone de las fases 5, 6 y 7 que describen cómo el desarrollo de medidas y la medición del desempeño de la seguridad de un control de seguridad en específico o grupo de controles, ayuda a la mejora continua de los SGSI y que el proceso conecta las actividades de seguridad de la información con las metas estratégicas de la organización a través del desarrollo y uso de medidas del desempeño. Este enfoque asume que las organizaciones tienen múltiples metas estratégicas y que

cada una de esas metas requiere de inputs de múltiples medidas. Sin embargo, en este capítulo 5 solo se comenta que los datos se pueden obtener de fuentes de información existentes y repositorios de datos exponiendo varios ejemplos y tal como *ISO/IEC 27004:2016* en su *Anexo B* ofrece ejemplos de construcción de medidas mapeadas con los controles de *ISO/IEC 27001:2013,* la *NIST 800-55 Rev.1:2016* ofrece al lector una serie de plantillas de desarrollo de medidas.

Continuando con el capítulo 6 de Implementación de Medidas de Seguridad de la Información que se resume en la Ilustración 3.

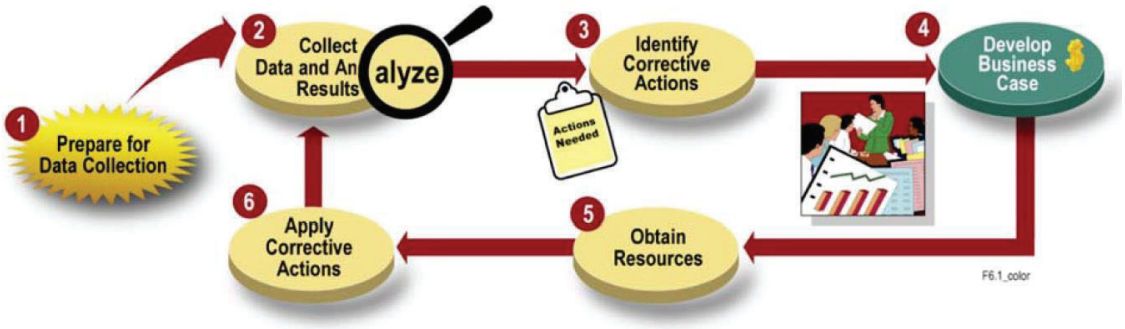


Ilustración 3. Information Security Measurament Program Implementation Process.

La fase de este Proceso de Implementación de Medidas de Seguridad de la Información que podría aportar alguna información más interesante para la realización de la taxonomía de fuentes de información es la fase 2 de Colección de Datos y Análisis de los resultados, primordialmente en la recolección de datos, ya que estos son outputs de las fuentes de información existentes. Sin embargo, no se ha encontrado nada respecto a fuentes de datos y simplemente se dice que la recolección de datos de medida debe seguir los procesos definidos en el Plan de Implementación de Medidas definido en la fase 1.

Como se ha dicho anteriormente, el objetivo final del presente trabajo es conseguir una taxonomía de fuentes de información a la que recurrir, pero ¿dónde se puede encontrar una guía para poder desarrollar esta taxonomía? No existen estándares o marcos de trabajo que describan cómo se tiene que desarrollar una taxonomía de fuentes de información o incluso qué atributos debería de tener ésta. Esto, también supone un inconveniente por lo que hará más laborioso el desarrollo de este trabajo. La calidad de las fuentes de información afecta al cuadro de mando y las fuentes de información tanto internas como externas a las que puede recurrir un responsable del SGSI para la recolección de datos son de lo más variadas y esto puede ser tanto un punto a favor como en contra. Es un punto a favor porque proporcionan información desde diversas perspectivas o contextos y es un punto en contra porque muchas de las fuentes de información que son accesibles son imprecisas, incompletas, heterogéneas, no filtrables, proporcionan datos que puede que no estén normalizados e incluso no destacan valores atípicos, y esto queda a interpretación de un responsable del SGSI, que pueden afectar al resultado e interpretación de los indicadores. Por lo tanto, la acción de seleccionar una fuente de información entre otras, es algo compleja. Por desgracia, este un

problema al que se enfrentan la mayoría de las empresas que deciden controlar su seguridad mediante un cuadro de mando. Como se puede ver en la Ilustración 4 del Modelo de Medición de la Información, la construcción del indicador, sobre el que posteriormente se realizarán interpretaciones, depende de combinaciones de medidas en las que cada una de estas es natural de las diferentes fuentes de información existentes.

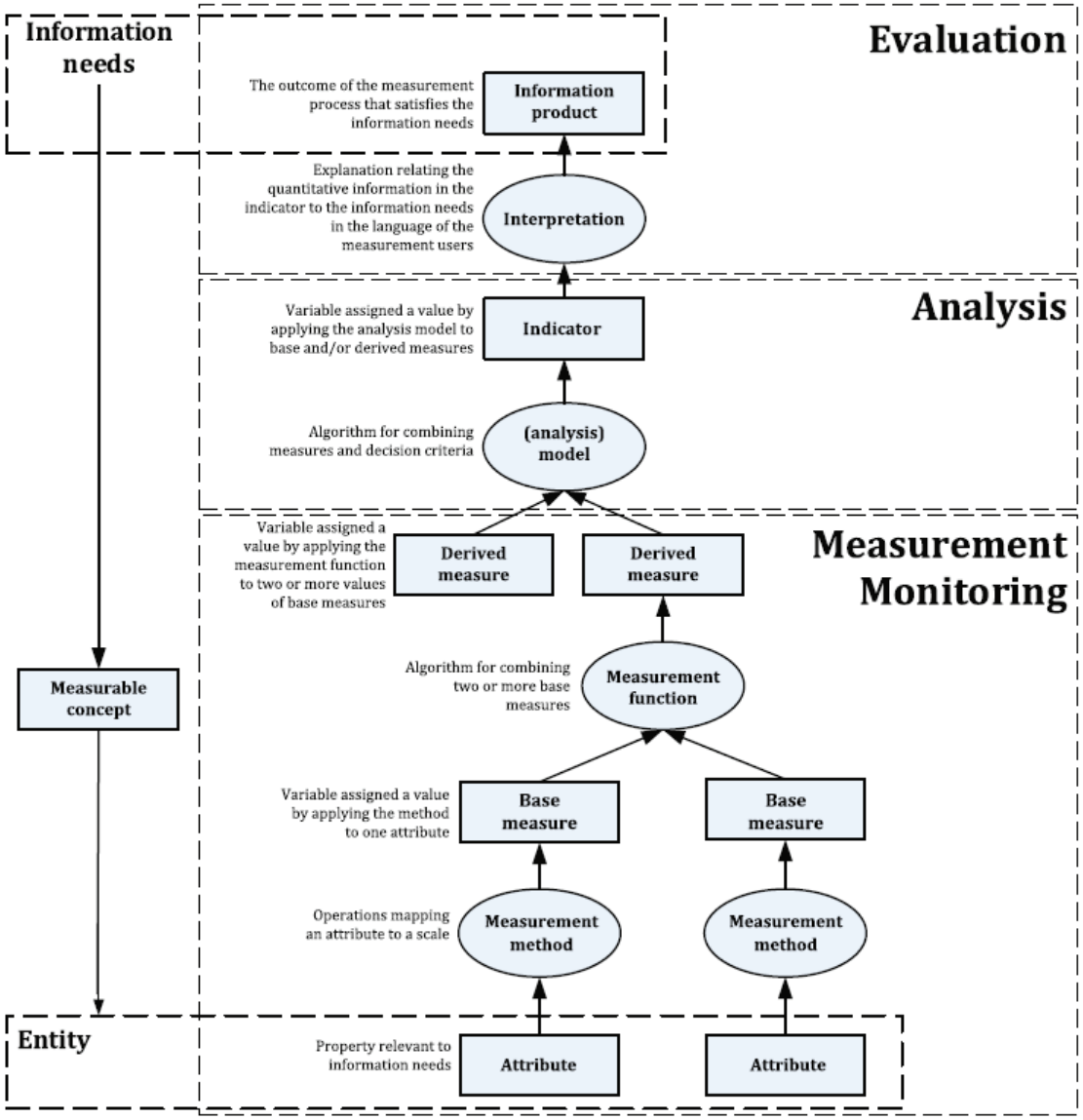


Ilustración 4. Key relationships in the measurement information model.

Entonces, ¿qué se puede hacer? Si al final lo que se quiere son indicadores de calidad, *¿por qué no clasificar la fuente de información a través de la utilidad que tenga para el cálculo del indicador?* Mediante el desarrollo de este trabajo se construirá una taxonomía completa de fuentes de información mediante la comparación entre fuentes de información existentes y ponderación de sus atributos, siempre teniendo en cuenta como criterio lo útiles que son para el cálculo, gestión y eficacia de indicadores, que se verán en el siguiente capítulo. *¿Cuáles son los atributos de las fuentes de información disponibles que más ayudarán en el posterior cálculo, gestión y eficacia de indicadores?* El presente trabajo pretende responder a esta pregunta.

2.1.1 La fuente de información como parte fundamental en el cálculo de indicadores

Antes de iniciar con la tarea de identificación de atributos, merece la pena aclarar, aunque pueda ser obvia, la diferencia entre un indicador y una fuente de información. Sin olvidar el contexto de la Seguridad de la Información, un *indicador* es un valor generado a partir de una combinación de medidas, una métrica, que tiene el fin de representar el estado o eficacia del desempeño del SGSI implantado. Pero no todo valor resultante de una métrica es un indicador. Esto quiere decir, que cada indicador tiene su propia métrica definida y que no le es útil cualquiera. Además, la métrica para un indicador puede evolucionar debido a la aparición de cambios en esta que mejoren los resultados anteriores haciéndolo más preciso o realista si cabe. Con un conjunto de indicadores, la organización puede interpretar con una mayor facilidad si se están alcanzando los objetivos planificados o no y a partir de ahí tomar determinadas decisiones.

Por otro lado, las *fuentes de información* son el *origen de los datos* utilizados en las distintas métricas empleadas. ¿Significa esto que los atributos que tiene que tener un indicador son los mismos que tiene que tener la fuente de información para calcular tal indicador? Con estas definiciones los conceptos de indicador, métrica y dato están demasiado acoplados tal que confunde en la definición y atribución de los atributos que deberían caracterizar a todo indicador y por otro lado los atributos que deberían caracterizar a las fuentes de información. Es importante que estos atributos sean atribuidos correctamente porque pueden perjudicar a la efectividad de la taxonomía. Distinguir si los atributos de una fuente de información y un indicador deberían ser los mismos resulta dificultoso. Por dicha razón y con el objetivo de inferir la taxonomía de fuentes de información, a continuación, la definición de indicador será desglosada, o mejor dicho desacoplada, con el fin de encontrar las partes fundamentales de las que depende su cálculo y destacando el papel de la fuente de información en este, a través del siguiente esquema:

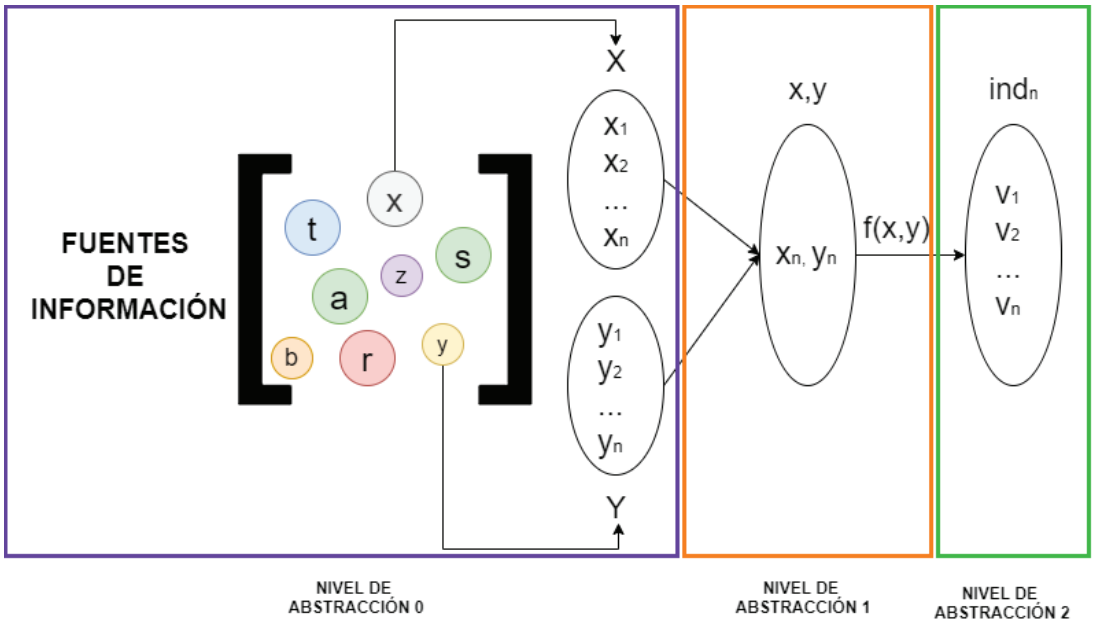


Ilustración 5. Niveles de abstracción contenidos en el cálculo de un indicador.

Las características de los indicadores las relacionan con el acrónimo ***SMART***, para decir que un indicador debe de ser e**s**pecífico, **m**edible, **a**lcanzable,

relevante y disponible en el **t**iempo requerido y además Wayne Eckerson profesional reconocido en el campo del análisis y el business intelligence, añade a parte de otras características, que deben estar alineados con los objetivos estratégicos de la empresa. Estas características están meditadas y estudiadas para la finalidad que tiene que tener un indicador, pero no tiene sentido trasladarlas como atributos de las fuentes de información. Sin embargo, de forma explícita el esquema anterior demuestra la relación de utilidad de cada uno de los niveles de abstracción con el cálculo del indicador. **Esta relación de utilidad puede usarse como criterio para desarrollar la taxonomía de fuentes de información.** Este trabajo se centra en el nivel de abstracción 0, las fuentes de información, que son la base para el cálculo de los indicadores y son de las que a partir de ahora se centrará este trabajo de investigación.

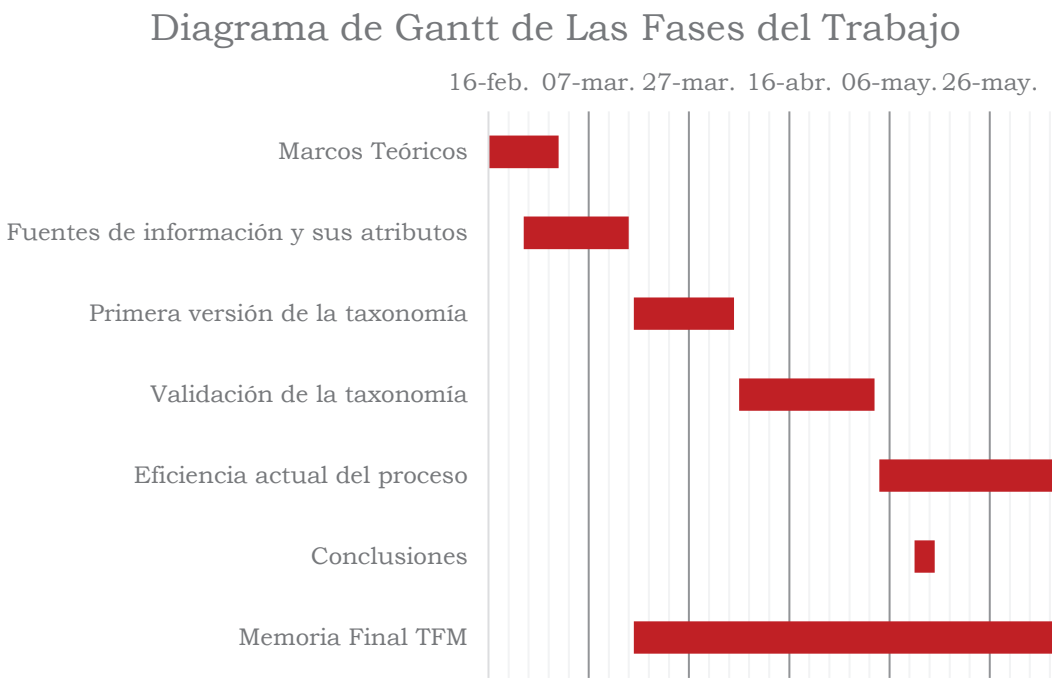
3 Desarrollo

Se describe en este capítulo las actividades desarrolladas en este TFM para la creación de una taxonomía de fuentes de información.

3.1 Fases del trabajo

El trabajo se ha desarrollado en las fases que se describen :

FASE	*DESCRIPCIÓN*
[FA][1]	Estado del arte de la medición de seguridad: **marcos teóricos** de trabajo.
[FA][2]	**Fuentes de información y sus atributos. Criterios de utilidad aplicables a los atributos. Identificación de atributos** exigibles de indicadores.
[FA][3]	**Primera versión de la taxonomía de fuentes de información**. Ponderación de importancia de los atributos.
[FA][4]	**Validación de la taxonomía.** Utilización de fuentes de información nuevas no recurridas para la identificación de los atributos de la taxonomía. **Posibles mejoras.**
[FA][5]	Selección de **indicadores** de seguridad **reales** para aplicación de Taxonomía. **Eficiencia** actual del **proceso**. Aplicación de taxonomía para **cálculo** de indicadores de seguridad con fuentes de información propuestas por la taxonomía. Valoración de **nuevos indicadores**. Comparación de escenario. Identificación de acciones de mejora.



3.2 Fuentes de información y sus atributos

Como primera tarea para el desarrollo de la taxonomía, se identificarán atributos de las fuentes de información de interés para la creación de la taxonomía, a través del análisis de fuentes de información y la aplicación de criterios de su utilidad para el cálculo de indicadores.

Para ello, se utilizarán diversas fuentes de información utilizadas para el cálculo de indicadores de seguridad. Del análisis de estas fuentes:

1. Se identificarán los atributos que las pueden caracterizar en base a criterios de utilidad.
2. Se optimizará esta lista de atributos para su posterior ponderación.

En todo este proceso, se tendrá siempre presente la orientación de la taxonomía como herramienta para determinar la utilidad de estas fuentes de información para el cálculo, gestión y eficacia de indicadores de seguridad.

3.2.1 Criterios de utilidad aplicables a los atributos

Disponer de una taxonomía de fuentes de información es de utilidad para un Responsable de Seguridad para seleccionar con rapidez posibles fuentes de información. Pero se trata también de que la fuente sea de calidad. Para conseguir seleccionar una fuente de calidad, primero hay que **definir un concepto de calidad propio**. Y es que este concepto de calidad tiene que construirse a base de un conjunto de criterios que apoyen la utilidad o fin que tendrán las fuentes de información mejor puntuadas por la taxonomía. ¿Cómo se pueden conocer esos criterios? Se propone extraerlos a partir del criterio de utilidad de la taxonomía, el cálculo de indicadores. Estos criterios tendrán un principio base, siempre tendrán que favorecer al cálculo de indicadores.

Los criterios de selección y ponderación de los atributos de las fuentes se pueden definir a partir de las características que definen a un indicador. Sin

embargo, algunas de estas características son naturales de las fuentes de información, por ejemplo:

- Si se habla de la propiedad de **específico**, un indicador tiene que ser particular de un control. Si el fin es el cálculo de indicadores de seguridad, las fuentes que se usen tienen que ser específicas para el cálculo de indicadores de seguridad. Por lo tanto, **este sería un criterio de selección de fuente**, no un criterio de ponderación, es decir la fuente quedaría inmediatamente descartada o no.
- La información de **una fuente de información** siempre o la mayoría de las veces **se puede cuantificar o categorizar**, es decir harían a un indicador cumplir la propiedad de **medible**. Por lo tanto, este no sería un criterio a considerar a la hora de seleccionar y ponderar los atributos, ya que hace descartar o no la fuente pero no la destaca o valora ante las demás. Esto **también ocurre si se habla de realismo** de la fuente de información.
- Si se habla de **alcance**, esta propiedad no tiene tanto que ver con la forma de calcular el indicador, es decir, con su fuente, sino con el umbral de consecución con el que el responsable del cálculo se sienta satisfecho. Por lo tanto, este no sería un criterio a considerar a la hora de seleccionar y ponderar los atributos, ya que hace ni descartar o no la fuente ni la destaca o valora ante las demás.
- Por último, se tiene la propiedad de **disponible en el tiempo**. Esta es la propiedad del indicador que puede verse más favorecida según la selección de la fuente de información. Por ello **será un criterio de ponderación a considerar.**

Como se ha visto, a través de las propiedades de los indicadores, han sido encontrados criterios aplicables que ayudarán en la selección de los atributos de las fuentes de información y ponderación de los valores de estos y con ello, se favorecerá al cálculo de indicadores. Se ha de dejar claro, que los indicadores per se tienen sus propiedades y que en ningún momento se dice que estos criterios de ponderación y selección, que han sido extraídos, formen parte de la definición de un indicador.

Una vez se tienen estos criterios generales, se pretende desglosarlos en aspectos, para poder trazarlos con cada uno de los atributos que compondrán a la taxonomía. Esto permitirá tener siempre conectada la taxonomía con los criterios. ¿Qué utilidad tiene esta metodología seguida?

1. Asegurar que los atributos tienen relación con los criterios que interesan para favorecer el cálculo de indicadores.
2. Una vez realizada la ponderación de los atributos, resulta mucho más fácil extraer conclusiones e identificar mejoras sobre el cuadro de mandos utilizado hasta el momento, ya que, se podrán ver apoyadas en los diferentes aspectos.

El cálculo de indicadores es una tarea operacional y la posibilidad de automatización de esas operaciones simplifica las mismas, por dicha razón, se ha podido extraer de la propiedad del indicador de *disponible en el tiempo*, **el criterio de automatización** que será un criterio del tipo ponderable. Por otro lado, aunque también se ha identificado, a través de la propiedad *específico* del indicador, **el criterio de selección,** este criterio será del tipo selectivo, con lo que los atributos relacionados con este criterio no ponderarán en la taxonomía pero tendrán la función o rol de descartar o no la fuente de información.

Para el **criterio de automatización**, solo se ponderarán los atributos en base a la relevancia que tienen a la hora de que una fuente de información se pueda automatizar y así favorecer a la disponibilidad en el tiempo del indicador, agilizando su cálculo. **Este criterio se puede relacionar con tres aspectos que pueden cambiar su grado de cumplimiento** y se definen tal que:

- **El aspecto tecnológico:** La automatización de una fuente de información dependerá de la naturaleza de la fuente y de las tecnologías que puedan aplicarse en su automatización.
- **El aspecto de coste de utilización:** La automatización de una fuente de información será posible siempre que sea eficiente en términos de coste temporal y monetario.
- **El aspecto de dominio/control:** La automatización de una fuente de información que tiene un mayor alcance y profundidad de controles, será más recomendable que una fuente de información para un control en específico. Además, una fuente de información volátil ayudaría a tener un control más preciso que una fuente menos volátil, al generar un mayor volumen de información.

Estos aspectos estarán presentes en mayor o en menor grado, pero presentes, tanto en la taxonomía de fuentes de información a través de los atributos identificados y se verá que también en el cuadro de mandos utilizado como escenario en este proyecto para validar la taxonomía. Más adelante, en el *capítulo 3.3*, se verá que ese mayor o menor grado, es decir la calibración ideal de las ponderaciones de los atributos, será la que se busque para que la salida de la taxonomía sea la ideal para el cuadro de mandos utilizado. Para ello, se ajustarán las ponderaciones de los atributos a los pesos reales o aproximados del cuadro de mandos. Antes de esto, en el siguiente subcapítulo, se identificarán los atributos empleados en la taxonomía.

3.2.2 Identificación de atributos exigibles a las fuentes de información

El presente capítulo se centrará exclusivamente en la identificación de los atributos exigidos a las fuentes de información seleccionadas, dejando para otros capítulos posteriores las distintas ponderaciones de estos.

Como indica la **Ilustración 5**, las fuentes de información son de lo más variadas, muchas reúnen cantidades ingentes de datos y la mayoría de las veces supone toda una aventura encontrar la información que se busca, y no por el hecho de que no se sepa la fuente en la que hay que buscar sino por el hecho de que el estándar ISO 27002 tiene 114 controles, de 15 ámbitos muy distintos y habitualmente independientes en las organizaciones. El responsable de seguridad debe implantar controles y obtener información en ámbitos tan dispersos como la organización empresarial (controles del capítulo 6 de ISO 27002), la gestión de personal/talento (controles del capítulo 7), la seguridad física (capítulo 11), departamento legal (capítulo 18), departamento de compras (capítulo 15) y gestión financiera de la empresa (capítulo 9 de ISO 27001). Y por supuesto, cada uno tiene sus sistemas y formas de trabajo. Forzosamente, es una tarea que requiere de un esfuerzo prolongado, sostenido y distribuido.

Como fuentes de información han sido reunidas unas 57 fuentes de distinta naturaleza: Entre ellas se han incluido logs de servicio TI, normativas, políticas,

manuales, mails, conversaciones cara a cara, BBDD de herramientas de gestión, archivos de configuración, reportes de servicios e indicadores de ataque. De ellos, se identificarán los atributos que formarán la taxonomía de fuentes de información, que serán excluyentes entre sí, objetivos y compartidos por todas las fuentes de información consideradas.

Esta tarea se trata de conseguir un conjunto de atributos mínimo con las máximas capacidades de evaluación posibles sobre las fuentes de información. Se pretende que este conjunto de atributos no sea muy extenso, ya que, se quiere evitar que la tarea de aplicar la taxonomía a una fuente de información sea una tarea laboriosa. Es decir, si se tienen unas 57 fuentes de información y el conjunto de atributos tiene una cardinalidad de 25 atributos, el responsable de realizar la tarea tendía que atribuir 1425 valores en total y como se ha dicho a lo largo de la lectura, que el coste temporal sea el mínimo posible es primordial. Por lo que se busca, una cardinalidad aproximada a 10 atributos.

¿Cómo se puede saber si los atributos seleccionados son los mejores que se pueden seleccionar? Aquí se hará una primera identificación, más adelante se verá cómo seleccionar los atributos más óptimos. Sin embargo, esta es una pregunta difícil de responder con exactitud porque a priori se puede tener una ligera idea pero, al no tener la taxonomía aplicada y validada no se puede saber si los resultados han mejorado o no. Sin embargo, no es necesario que la taxonomía se componga de un conjunto de atributos ideales o inmejorables, sino que sean lo bastante buenos. Lo importante es que los resultados después de utilizar la taxonomía para el cálculo de indicadores, mejoren, y esto lo provocarán las nuevas fuentes seleccionadas en la salida de la taxonomía y gracias a la calidad de esta. Se busca una buena base de atributos, las oportunidades de mejora siempre están ahí esperando.

Este conjunto de atributos se ha ido extrayendo de las necesidades o requisitos impuestos de forma implícita a las fuentes de información hasta ahora utilizadas para el cálculo del cuadro de mandos. Estos requisitos, sin importar su orden, son los siguientes:

1. Como se ha comentado en el capítulo 1 un CMS o cuadro de mandos se suele trazar con un estándar en este caso de la Seguridad de la Información, *ISO 27001*, por lo que, es de esperar que se muestre *interés por fuentes de información que estén relacionadas con la Seguridad de la Información.*
2. Si los indicadores tienen que mostrar la efectividad de los controles, *las fuentes tienen que ser fiables.*
3. Las fuentes de información se aplican a aspectos de gestión o controles específicos.
4. Son preferibles fuentes de información que informan de varios controles de seguridad a fuentes específicas de un control.
5. Son preferibles fuentes que proporcionen la información localizada en sí mismas y no distribuida en diferentes fuentes.
6. Las fuentes de información siempre han de ser accesibles por el responsable que las precisa.
7. *Cualquier característica de una fuente de información*, ya sea, su formato, capacidad de generación de grandes volúmenes de información, metodologías de consulta y de extracción de la información o incluso la disponibilidad del sistema al que pertenecen, *que afecte a su automatización.*

Una vez se han definido los siete requisitos anteriores, han ido apareciendo con mayor facilidad los atributos de forma que, se ha conseguido el siguiente conjunto inicial de 15 atributos, tal que:

ATRIBUTO	**DEFINICIÓN**	**ROL**	**VALORES**
Relación con la SI	El contenido de la fuente está relacionado con algún control de la seguridad de la información.	Selecciona	Sí/No
Categoría	Naturaleza o tipo de la fuente de información.	Selecciona	Persona/Audiovisuales/ Textuales/ Sistemas TI **Persona:** Contacto directo de forma oral o escrita con la fuente. **Audiovisuales:** Audio, video e imágenes. **Textuales:** Documentos físicos, .doc, txt, .pdf, .xls, entrevistas, informes, cuestionarios, encuestas, etc. También se incluyen fuentes como aplicaciones que tienen la funcionalidad/capacidad de extraer su información en un archivo de tipo informe. **Sistemas TI:** Archivos .log, .conf, etc. Son archivos que se extraen de un sistema TI. Suelen ser registros, debugs o logs, configuraciones, etc. Son archivos que no tienden a seguir una estructura documental.
Presentación	Formato en el que se presenta la información de la fuente.	Pondera	Documento físico/Documento electrónico/ Oral/Mail/

ATRIBUTO	DEFINICIÓN	ROL	VALORES
			BBDD/Archivo log/Arhivo conf.
Fiabilidad	Necesidad de contrastar la información de la fuente porque el origen de la fuente no es fiable.	Selecciona	Sí/No
Aplicación	Es el ámbito de aplicación de la fuente dentro de la SI. El ámbito de aplicación está relacionado con aspectos de gestión de la seguridad de la información o con controles específicos de algún dominio de la seguridad de la información. Todos ellos medibles. (Gestión de riesgos, gestión de incidentes, gestión de activos.	Selecciona	Aplicación de controles/ Datos de gestión de la SI
Especificidad	Relacionado con la finalidad medible de la fuente de información. ¿Qué ayuda a medir la fuente de información la aplicabilidad del control, el rendimiento/efectividad del control o la gestión del SGSI?	Selecciona	Gestión / Rendimiento / Aplicabilidad
Multicontrol	¿La fuente de información es específica de un control u ofrece información de un número mayor de controles de seguridad de la información?	Pondera	Sí/No

ATRIBUTO	DEFINICIÓN	ROL	VALORES
	La fuente de información almacena información relacionada con diferentes dominios o temas de la seguridad de la información. A mayor número de controles, la fuente de información favorece el coste temporal de cálculo del CMS.		
Esfuerzo de acceso a la información	¿Cuánto cuesta en tiempo llegar a encontrar lo que se busca?	Pondera	Los tiempos de respuesta y búsqueda son elevados; Los tiempos de respuesta y búsqueda se compensan; Los tiempos de respuesta y búsqueda son mínimos.
Disponibilidad	Capacidad del sistema al que pertenece la fuente de estar operable/disponible de forma continua 24/7.	Pondera	Inmediata / Ninguna/ No inmediata
Volatilidad	Capacidad de la fuente de información de generar grandes volúmenes de información.	Pondera	Alta/Media/Baja **Alta:** Diariamente o mensualmente. **Media:** Más de 2 veces al año. **Baja:** En el año una vez o en años.
Tipo de búsqueda	Metodología de consulta en la fuente de información para realizar la búsqueda de los datos.	Pondera	Manual/Script/Manual con filtros
Extracción de la información	Metodología de extracción de información en la fuente de información.	Pondera	Consulta DDBB/Exportando a otro formatos/Script

ATRIBUTO	DEFINICIÓN	ROL	VALORES
Localización	La información puede encontrarse en una misma fuente de información o la información está distribuida en distintas fuentes de información.	Pondera	Centralizada/Distribuida
Accesibilidad	Dependiendo del nivel de clasificación asignado a la información de la fuente, se requerirá de permisos de acceso a solicitar previamente bajo justificación lo que perjudica al esfuerzo de acceso a la información. En el caso de que la fuente de información no sea accesible no se podrá utilizar.	Pondera	Sí/No/Aplica control de acceso/ Requiere autorización previa
Metadatos	Los datos de los datos. Pueden servir para contrastar información como las fechas de creación, tamaño, extensión, etc.	Pondera	Sí/ No

Tabla 1. Primera versión de los atributos de la taxonomía

Aun así, este conjunto de atributos no es el conjunto final propuesto para la taxonomía. Sobre esta lista de atributos, se han ido tomando algunas decisiones que han acabado en descarte, desgloses y unificaciones de atributos quedando el conjunto de atributos con una cardinalidad de 11, 9 atributos ponderables y 2 atributos de selección.

A través de varios análisis, estas son las conclusiones más relevantes que se extrajeron:

1. Se detectó que varios atributos con rol de selección eran prescindibles en la taxonomía, ya que, sus posibles valores no tenían utilidad para la taxonomía porque no había ninguna casuística que hiciera descartar a la fuente de información. Estos atributos eran *Aplicación* y *Especificidad.* Sin embargo, aunque atributos como *Relación con la SI* y *Fiabilidad* tengan un rol de selecciona, estos eran imprescindibles, es decir, que si una fuente de información no los cumple esta quedaría inmediatamente descartada.

2. El atributo *Categoría,* que en un principio tenía un rol de selecciona pasó a ser ponderador, ya que se ha considerado que cada uno de los valores que puede tomar tiene cierto peso en la taxonomía. Son preferibles fuentes que favorezcan la automatización a fuentes que vayan en contra de esta.
3. Para los atributos *Disponibilidad, Localización* y *Accesibilidad*, se detectaron las siguientes situaciones:

    a. El atributo *Disponibilidad,* que fue definido como "Capacidad del servicio, en el que se encuentra la fuente de información, de estar operable de forma continua 24/7.", se pudo decir por experiencia que casi todas las fuentes están disponibles y por lo tanto este no es un atributo determinante en la puntuación de la fuente de información. Sin embargo, se verá al final de este capítulo, que en cierta forma la disponibilidad, aunque con un matiz distinto a esta definición, estará presente en la taxonomía.
    b. El atributo *Localización*, que fue definido como "Dónde se encuentra la información que se busca, centralizada en un sistema o distribuida en varios sistemas.", no es útil para la taxonomía, tal y como está definido porque a la hora de valorar la fuente en la taxonomía, no se sabe a priori si la información que se va a buscar está centralizada en esa fuente o distribuida en varias y más si la fuente fuera multicontrol.
    c. Para el atributo *Accesibilidad*, que fue definido como "Dependiendo del nivel de clasificación asignado a la información de la fuente, se requerirá de permisos de acceso a solicitar previamente bajo justificación. En el caso de que la fuente de información no sea accesible no se podrá utilizar.", a priori podría ser un atributo interesante tal y como está definido, sin embargo la mayoría de la fuentes suelen estar accesibles con lo que este atributo no sería determinante o no tendría un peso relevante en nuestra taxonomía como para discernir entre una fuente de información u otra. Además, una fuente de información con accesibilidad dificultosa no se suele solicitar al momento del cálculo del cuadro de mandos, sino que se solicita previamente. Se entiende que todas las fuentes que se necesitan para el cálculo del cuadro de mandos se encuentran accesibles en ese momento.
    d. Para el atributo *Multicontrol*, definido como "La fuente de información almacena información relacionada con diferentes dominios o temas de la seguridad de la información. A mayor número de controles, la fuente de información favorece el coste temporal de cálculo del CMS.", al contrario que los casos anteriores, se planteó la siguiente situación:

        Hasta ahora, como se había definido el atributo, las fuentes eran multicontroles o específicas, pero si se plantea lo siguiente: ¿Qué conviene más una fuente multiusos/multicontrol o una fuente específica de un control que tenga un mayor grado de control sobre este? A priori, se prefiere **un menor coste al tener una fuente multicontrol que un mayor coste con una fuente específica con un mayor grado de control**, a no ser que esta fuente pueda justificar por sus propias características que por su mayor grado de control sobre un control, el uso de esa sea preferible. Una vez hecha esta reflexión, se decidió incluir dentro de la definición el

grado de control y el coste de la fuente, añadiendo más valores al atributo *Multicontrol* y pasándose a llamar, *Multicontrol/Grado de Control/Coste*. Sin embargo, al tener cada uno de los conceptos *Multicontrol/Grado de Control/Coste* concentrados en un atributo y con varios valores tal que: Multicontrol/Alto/Alto (70%), Multicontrol/Alto/Bajo (100%), Multicontrol/Bajo/Alto (10%), Multicontrol/Bajo/Bajo (50%), Específico/Alto/Alto (60%), Específico/Alto/Bajo (90%), Específico/Bajo/Alto (5%) y Específico/Bajo/Bajo (40%), se detectó que en el peso del concepto Multicontrol en el atributo, una fuente específica en cualesquiera de los casos tiene un peso siempre de un 10% menos a una fuente multicontrol. Si se considera que una fuente es una fuente multicontrol si maneja información de al menos 2 controles, esto demuestra que al final el valor multicontrol tiende a aportar lo mismo para una fuente que maneja 5 controles que para una fuente que maneja 15. Por lo que se decide separar este atributo en tres, el atributo *Multicontrol*, el atributo *Grado de Control* y el atributo *Coste de la fuente*. Además, para el atributo *Multicontrol* se decide definir valores diferentes a los Sí/No o Multicontrol/Específico utilizados hasta ahora.

Los cambios, eliminaciones, adiciones y redefiniciones de los atributos quedan plasmados en la siguiente tabla:

Atributo	**Eliminado**	**Redefinido**	**Seleccionado**	**Comentario**
Relación con la SI			X	Descarta
Categoría			X	Pondera
Presentación			X	Pondera
Fiabilidad			X	Descarta
Aplicación	X			
Especificidad	X			
Multicontrol		X	X	Pondera
Grado de control			X	Adición y pondera
Coste de la fuente			X	Adición y pondera
Esfuerzo de acceso a la información		X	X	Pondera
Disponibilidad	X			
Volatilidad			X	Pondera
Tipo de búsqueda			X	Pondera

Extracción de la información		X	X	Pondera
Localización	X			
Accesibilidad	X			
Metadatos	X			

Tabla 2. Modificaciones en el conjunto de atributos.

Después de todo este análisis, ahora sí, se obtuvieron los 11 atributos finales.

Como se ve, todos estos atributos son perfectamente trazables con los criterios de automatización y selección definidos en el capítulo 3.2.1.

He aquí la siguiente ilustración que lo resume:

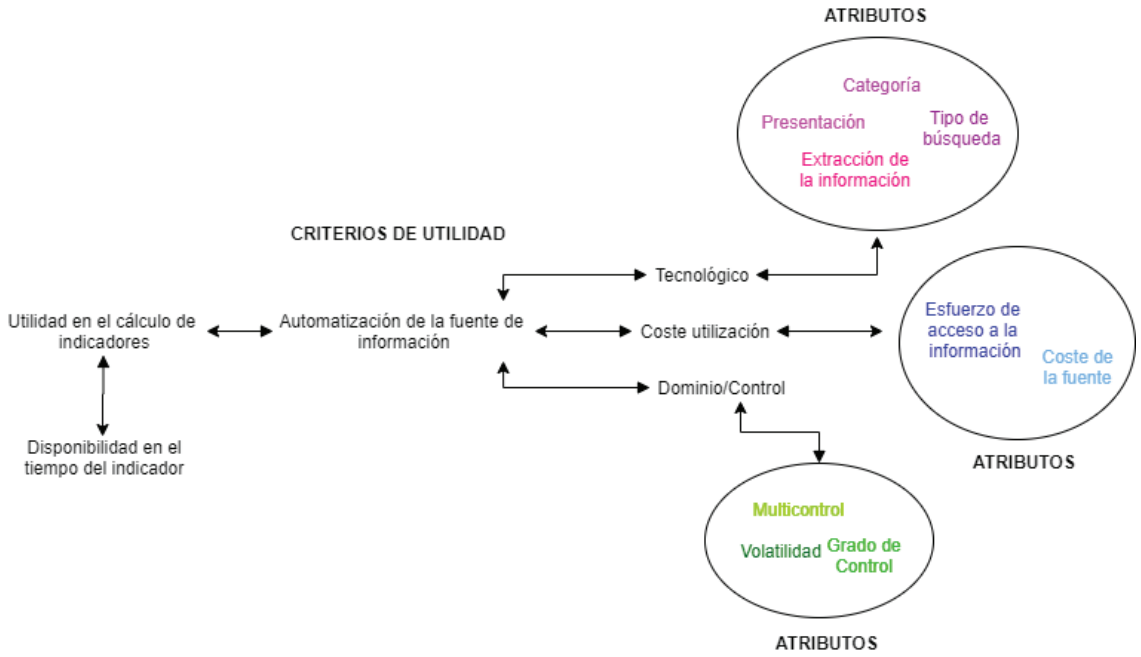


Ilustración 12. Traza de criterios y atributos

Una vez ha sido obtenida la primera versión de los atributos de la taxonomía, el siguiente paso será ponderarlos, tal que quede elaborada la primera versión de la taxonomía de fuentes de información.

3.3 Primera Versión de la Taxonomía de Fuentes de Información

El objetivo de este capítulo consiste en la valoración de la importancia relativa de los atributos seleccionados en el capítulo anterior para determinar su relevancia a la hora de seleccionar fuentes de información para el cálculo de indicadores de seguridad, favoreciendo la selección de fuentes potencialmente más útiles. Así, se realizará:

1. Una primera ponderación de la relevancia de los atributos identificados, que permita seleccionar unas u otras fuentes para el cálculo de indicadores, y
2. los criterios aplicados en dicha ponderación.

3.3.1 Criterios de Ponderación de los atributos

La valoración de la importancia relativa de los atributos identificados se apoya en una ponderación del peso que tiene cada uno de ellos, frente a los demás, para crear un índice que permita medir y comparar las distintas fuentes entre sí.

Dada la complejidad de la tarea, se ha optado por realizar esta valoración en dos pasos. Que utilizarán las 57 fuentes de información identificadas como base para la taxonomía de forma distinta:

- **El conjunto de entrenamiento,** que se basará en las fuentes de información utilizadas actualmente para el cálculo de la mayoría de los indicadores de seguridad. Con ellas se calibrarán las ponderaciones de los atributos hasta obtener una versión candidata a ser final.
- **El conjunto de test** que se utilizará para validar la taxonomía. Estas fuentes son tanto fuentes poco usadas, como fuentes de información candidatas, futuras o que podrían considerarse.

De la utilización de ambas fuentes se obtendrá la taxonomía propuesta, que se mostrará en el capítulo 0.

Pero, ¿cómo se pueden descubrir los pesos reales o aproximados de estos atributos? En esta fase, se **deberá descubrir cuál es esta ponderación numérica implícita de los atributos en el cuadro de mandos** mediante la calibración justificada de ponderaciones. Para ello, a partir del cuadro de mandos utilizado para este trabajo, se llevó a cabo un breve análisis de cuáles eran las fuentes más relevantes, extrayendo los porcentajes de utilización de una determinada fuente de información para calcular un indicador. De aquí salieron tres grupos de fuentes, A, B y C. Las fuentes del grupo A, eran las fuentes que, combinadas, permitían calcular el 65% de los indicadores de seguridad existentes. Las fuentes del grupo B, eran las fuentes que combinadas con las del grupo A, permitían calcular el 25% de los indicadores restantes. Por último, las fuentes del grupo C eran las fuentes que, combinadas, `permitían calcular un 10% de los indicadores actuales.

Por ello, se decide utilizar como fuentes del conjunto de entrenamiento las A y B, y se deciden utilizar como conjunto de test las C, junto con otras fuentes adicionales que se puedan considerar.

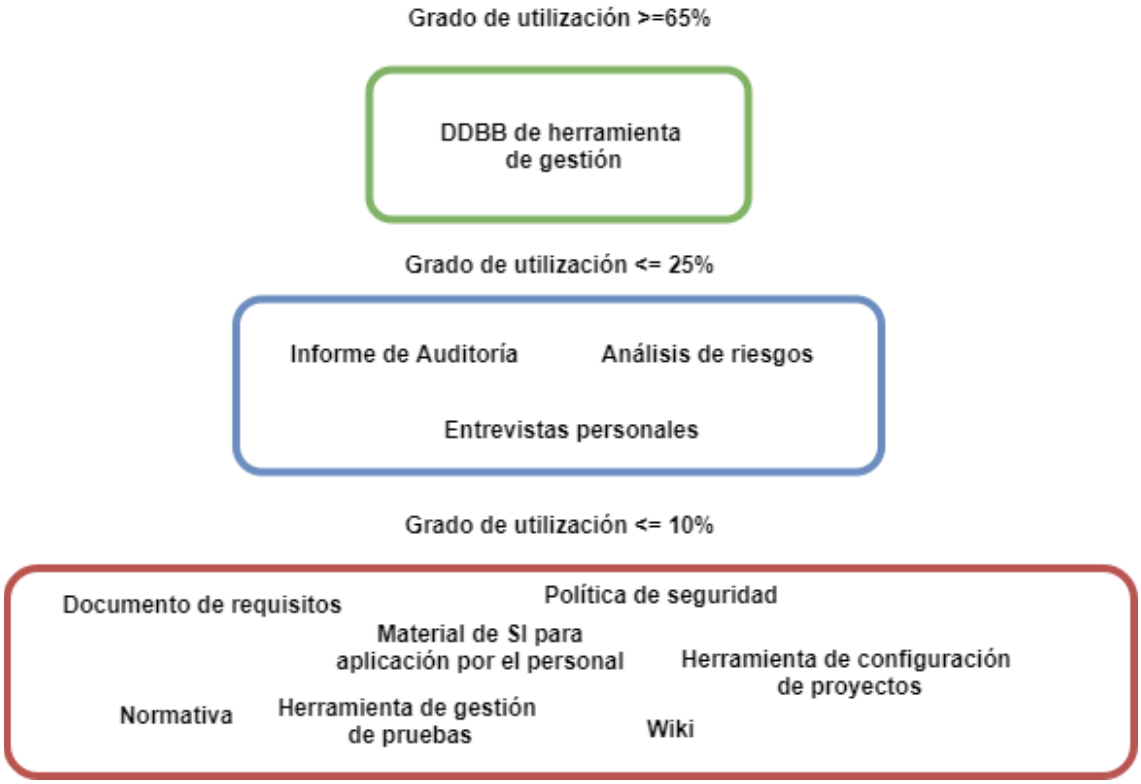


Ilustración 6. Grado de utilización de las fuentes de información en el CMS

3.3.2 Escenario de Ponderación de los atributos. Prioridades

Con esto se supuso que aquellas fuentes de información que se utilizaban con más frecuencia pues probablemente eran las más relevantes, mientras que aquellas que apenas se usaban podrían ser imprescindibles. En función de esto, se propusieron cuatro escenarios de aplicabilidad para dar con las ponderaciones implícitas en el cuadro de mandos:

1. Un escenario de cuadro de mando, llamado **E0**, dónde se valoraron prácticamente todos los atributos de forma **equitativa**. La ponderación de este escenario fue P0.
2. Un escenario de cuadro de mando, llamado **E1**, dónde los atributos más valorados fueron los relacionados con la facilidad de la **automatización** de la fuente. La ponderación de este escenario fue P1.
3. Un escenario de cuadro de mando, llamado **E2**, dónde los atributos más valorados fueron los relacionados con el la **precisión de medición** de la seguridad. La ponderación de este escenario fue P2.
4. Un escenario de cuadro de mando, llamado **E3**, dónde los atributos más valorados fueron los relacionados con el **grado de control de la fuente**. La ponderación de este escenario fue P3.

A continuación, se muestran cada una de las ponderaciones de los escenarios descritos anteriormente. Se ha de comentar también que cada uno de los escenarios se fue retroalimentando de la salida de los escenarios anteriores.

Para E0 se tiene lo siguiente:

ATRIBUTO	**PONDERACIÓN P0**
Categoría	112%
Multicontrol	111%
Grado de Control	111%
Coste de la Fuente	111%
Esfuerzo de Acceso a la Información	111%
Volatilidad	111%
Tipo de búsqueda	111%
Extracción de la información	111%
Presentación	111%

Tabla 3. Ponderación del escenario E0

FUENTE DE INFORMACIÓN	**RANKING PARA P0**
Normativa	6,72
Material de SI para aplicación por el personal	7,78
Wiki	7,44
DDBB de herramienta de gestión	8,66
Análisis de riesgos	5,94
Documento de requisitos	6,16
Política de seguridad	6,16
Entrevistas personales	3,60
Informe de Auditoría	6,72
Herramienta de gestión de pruebas	9,44
Herramienta de configuración de proyectos	8,89

Tabla 4. Ranking de fuentes de información de E0.

De este escenario E0, se detectaron las siguientes incongruencias u observaciones:

- Se observó que el atributo Categoría podría estar sobrevalorado porque aunque cada uno de sus posibles valores definen la vía, más buena o no, de automatización de la fuente, cuando el responsable del cálculo del cuadro de mandos realiza esta tarea tiene en cuenta otros factores por encima de la Categoría de la fuente. Por ejemplo, podría preferir una fuente con un alcance de controles mayor aunque esta fuente sea mediante una entrevista personal. El atributo Categoría es un fiel candidato a reducir bruscamente su ponderación.
- Se identificó que la fuente *Wiki* tenía una puntuación mayor que la fuente *Informe de auditoría*, cuando un informe de auditoría realmente tiene mayor provecho, ya que, informa de la aplicabilidad de los controles en la organización.

- Se identificó que la fuente *Análisis de riesgos* quedaba infravalorada por fuentes como *Documento de requisitos* o *Política de seguridad*. Por supuesto, un análisis de riesgos aporta mucha más información como por ejemplo el inventario de activos, la madurez de los controles de seguridad y valores de impacto/riesgo.
- Se identificó que la fuente *Wiki* al igual que la fuente *Material de SI para aplicación por el personal* tenían una puntuación bastante alta en comparación con la fuente *DDBB de herramienta de gestión* que es la fuente más recurrida en el cálculo del cuadro de mandos.

Para E1, se tiene lo siguiente:

ATRIBUTO	**PONDERACIÓN P1**
Categoría	20%
Multicontrol	80%
Grado de Control	75%
Coste de la Fuente	75%
Esfuerzo de Acceso a la Información	50%
Volatilidad	100%
Tipo de búsqueda	200%
Extracción de la información	200%
Presentación	200%

Tabla 5. Ponderación del escenario E1

FUENTE DE INFORMACIÓN	**RANKING PARA P1**
Normativa	6,69
Material de SI para aplicación por el personal	8,02
Wiki	7,81
DDBB de herramienta de gestión	8,72
Análisis de riesgos	5,26
Documento de requisitos	6,29
Política de seguridad	6,29
Entrevistas personales	3,34
Informe de Auditoría	6,69
Herramienta de gestión de pruebas	9,50
Herramienta de configuración de proyectos	9,10

Tabla 6. Ranking de fuentes de información de E1.

De este escenario E1, se detectaron las siguientes incongruencias u observaciones:

- Se identificó que se mantenían las mismas incongruencias que surgieron en E0, incluso en algunas ocasiones sobrevalorando fuentes sin sentido.

De este escenario, no se extrajeron conclusiones positivas tal que mereciese la pena continuar con su evolución en el estudio, y aun así era preferible el escenario E0.

Para E2, se tiene lo siguiente:

ATRIBUTO	**PONDERACIÓN P2**
Categoría	20%
Multicontrol	200%
Grado de Control	200%
Coste de la Fuente	200%
Esfuerzo de Acceso a la Información	50%
Volatilidad	100%
Tipo de búsqueda	80%
Extracción de la información	75%
Presentación	75%

Tabla 7. Ponderación del escenario E2

FUENTE DE INFORMACIÓN	**RANKING PARA P2**
Normativa	6,50
Material de SI para aplicación por el personal	7,05
Wiki	6,45
DDBB de herramienta de gestión	8,35
Análisis de riesgos	5,96
Documento de requisitos	5,50
Política de seguridad	5,50
Entrevistas personales	4,14
Informe de Auditoría	6,50
Herramienta de gestión de pruebas	9,50
Herramienta de configuración de proyectos	8,62

Tabla 8. Ranking de fuentes de información de E2.

De este escenario E2, se detectaron las siguientes incongruencias u observaciones:

- Se observó positivamente que la puntuación de la fuente *DDBB de herramienta de gestión* tenía una diferencia mayor con la fuente *Wiki* y la fuente *Material de SI para aplicación por el personal*, al contrario que en el escenario E0.

- Se volvió a identificar que el que la fuente *Informe de Auditoría* se valore prácticamente igual que la fuente *Wiki* o la fuente *Normativa* o un manual no tiene mucho sentido, ya que, un informe de auditoría es muy útil en un cuadro de mandos al retroalimentarlo de la aplicabilidad de los controles cuando estos no se pueden medir con facilidad.

Con estos resultados, se decidió mantener esta mayor relevancia de los atributos relacionados con el alcance y profundidad de los controles, pero reconociendo que el atributo *Grado de Control* tendrá un peso mayor, porque para el cálculo del cuadro de mandos (incluso para la mejora de indicadores) es preferible que la fuente ofrezca un detalle o profundidad del control considerable. Sin embargo, también se decidió calibrar los pesos de los atributos relacionados con la facilidad de automatización repartiendo el peso que se les había sido asignado hasta ahora entre ellos. Entre estos, se decidió que el atributo *Tipo de Búsqueda* debía ser el mejor valorado, ya que, se consideró que determina si la fuente es automatizable realmente o no.

Para E3, se tiene lo siguiente:

ATRIBUTO	**PONDERACIÓN P3**
Categoría	20%
Multicontrol	150%
Grado de Control	300%
Coste de la Fuente	150%
Esfuerzo de Acceso a la Información	50%
Volatilidad	100%
Tipo de búsqueda	130%
Extracción de la información	50%
Presentación	50%

Tabla 9. Ponderación del escenario E3

FUENTE DE INFORMACIÓN	**RANKING PARA P3**
Normativa	5,71
Material de SI para aplicación por el personal	6,85
Wiki	5,65
DDBB de herramienta de gestión	8,65
Análisis de riesgos	6,62
Documento de requisitos	4,96
Política de seguridad	4,96
Entrevistas personales	4,15
Informe de Auditoría	7,21
Herramienta de gestión de pruebas	9,50
Herramienta de configuración de proyectos	8,75

Tabla 10. Ranking de fuentes de información de E3.

De este escenario E3, se concluyó lo siguiente:

- Se afirmó que el atributo Categoría estaba sobrevalorado como se propuso en E0.
- Se consiguió que la fuente *Wiki* tuviera una puntuación mucho menor que la fuente *Informe de auditoría*.
- Se consiguió que la fuente *Análisis de riesgos* no quedara infravalorada por fuentes como *Documento de requisitos* o *Política de seguridad*.
- Se consiguió que la fuente *Wiki* al igual que la fuente *Material de SI para aplicación por el personal* tuvieran una gran diferencia en comparación con la fuente *DDBB de herramienta de gestión* que es la fuente más recurrida en el cálculo del cuadro de mandos.

Este escenario consiguió resolver todas las incongruencias y observaciones encontradas en los escenarios anteriores hasta ese momento. Por lo que, esta ponderación es la ponderación final a ser validada.

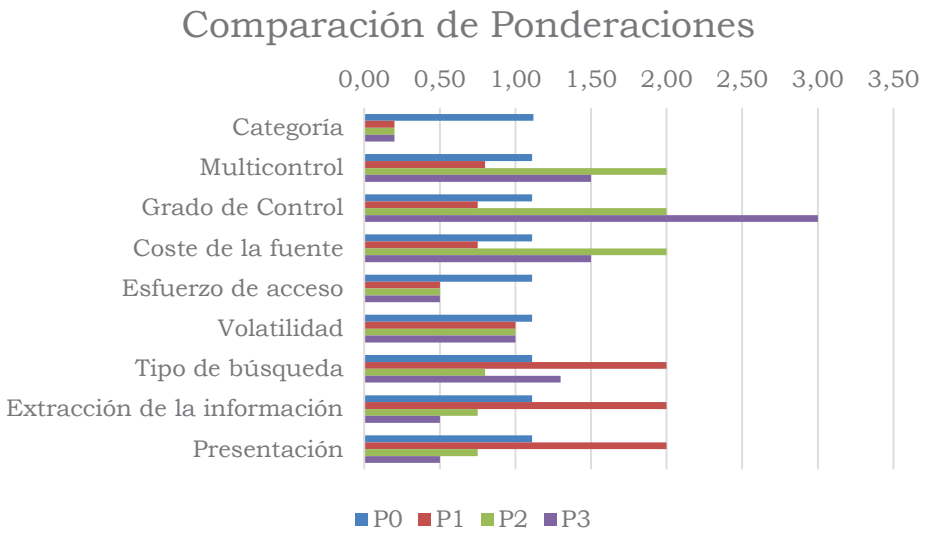


Ilustración 7. Comparación de ponderaciones

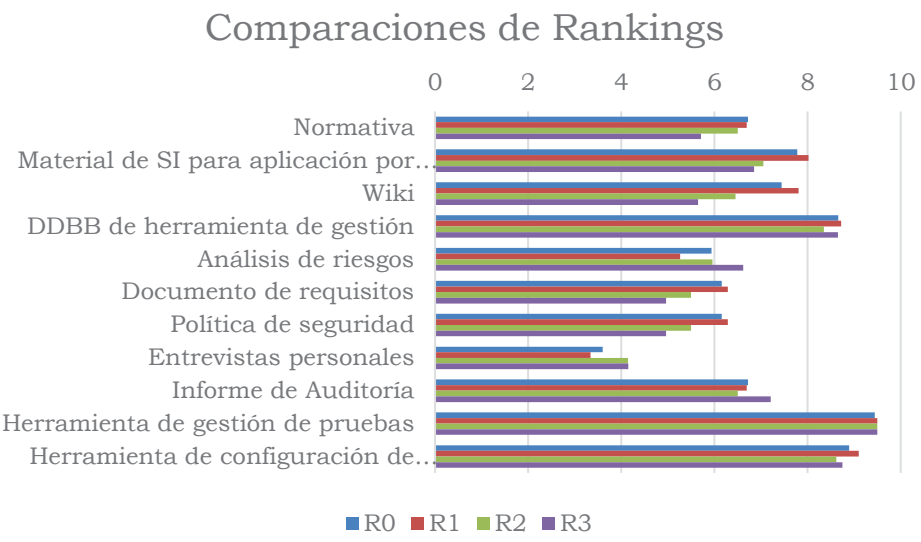


Ilustración 8. Comparación de rankings

3.3.3 Ponderación seleccionada para la taxonomía. Fase 1

Una vez finalizadas las calibraciones para obtener la ponderación final para la posterior validación de la taxonomía mediante el conjunto de fuentes de test, se consiguió una ponderación representativa del cuadro de mandos utilizado. Aunque el lector, no conozca ni pueda conocer el cuadro de mandos, calculado por el responsable de la SI de esta tarea, por motivos de confidencialidad, gracias a los breves análisis y criterios de calibración descritos en las diferentes ponderaciones mostradas, es posible hacerse una imagen de las preferencias que tiene el cuadro de mandos sobre las fuentes de información. Como se pretendía al principio de este capítulo, se quería descubrir la ponderación de aspectos (a través de los atributos) intrínseca en el cuadro de mandos y, aunque aún tiene que ser validada la taxonomía con esta ponderación final obtenida, se ha conseguido descubrir una ponderación coherente con la realidad y trasladarla a la taxonomía de fuentes de información. Por lo que, se puede decir a partir del escenario E3, que en el cuadro de mandos utilizado tienen mayor peso las fuentes de información que se caracterizan por proporcionar información extendida de varios controles de seguridad y a un medio/bajo coste consiguiendo con esto un mayor control del estado de seguridad de la organización y de las que se pueden aprovechar las características que las hacen automatizables.

Se pretende que esta deducción local en base a un ejemplo particular de fuentes de información sea extrapolable a un caso general. ¿Resistirá la taxonomía a la adición de nuevas fuentes de información?

3.4 Validación de la taxonomía

El objetivo de este capítulo consiste en la validación de la primera versión de la taxonomía de fuentes de información.

Para ello, se utilizarán las fuentes de información pertenecientes al conjunto de test no utilizadas para el cálculo de indicadores de seguridad como grupo de control. Se llevarán a cabo las siguientes tareas:

1. Detección de incongruencias en la salida de la taxonomía con las nuevas fuentes.
2. Si se precisan, se aplicarán mejoras en la ponderación que se valida, y
3. se justificaran los criterios aplicados en dicha ponderación.

Debido a la cantidad elevada de fuentes de información que se incluyen en el conjunto de test, no se podrán citar todas las fuentes. Sin embargo, se aportará en este trabajo una muestra de estas, que contiene aquellas que han sido más relevantes para la identificación de mejoras en el ajuste de la ponderación de la taxonomía de fuentes de información.

Aplicando la ponderación de las fuentes de la **Tabla 9**, se obtuvo que estas fuentes tenían el valor de utilidad que se indica en la siguiente tabla:

FUENTE DE INFORMACIÓN	**RANKING DE FUENTES NUEVAS**
Lectoras de acceso físico	7,75
Log de control de acceso a directorio	8,80
Reporte de personal de vigilancia	6,20
Reporte de antivirus	8,42
Monitorización de red	8,42
Informe de altas y bajas de usuarios	6,87
Contenido de directorio	2,825

Tabla 11. Ranking de fuentes nuevas con P3.

Después de aplicar la taxonomía con P3 a las nuevas fuentes de información se observó que muchas de las fuentes que ofrecían información sobre varios controles de SI, no se desmarcaban tanto de fuentes específicas. La diferencia entre estas fuentes era de 1 o 2 décimas a favor de las fuentes multicontrol. Esta diferencia se considera importante porque en la realidad del cálculo del cuadro de mandos siempre es preferible usar una fuente capaz de informar de una gran variedad de controles, ya que, esto hace ganar en tiempo de cálculo.

Esta nueva observación produce una nueva reponderación que pretende aumentar esta diferencia, tal que:

ATRIBUTO	**PONDERACIÓN DE MEJORA**
Categoría	20%
Multicontrol	250%
Grado de Control	300%

ATRIBUTO	**PONDERACIÓN DE MEJORA**
Coste de la Fuente	50%
Esfuerzo de Acceso a la Información	50%
Volatilidad	100%
Tipo de búsqueda	130%
Extracción de la información	50%
Presentación	50%

Tabla 12. Ponderación de mejora.

FUENTE DE INFORMACIÓN	**RANKING DE FUENTES NUEVAS**
Lectoras de acceso físico	7,75
Log de control de acceso a directorio	8,50
Reporte de personal de vigilancia	7,20
Reporte de antivirus	8,12
Monitorización de red	8,12
Informe de altas y bajas de usuarios	6,87
Contenido de directorio	3,32

Tabla 13. Muestra del ranking de fuentes de información nuevas.

Con la aplicación de la ponderación P3, se han metido bloques de fuentes nuevas y han salido fuentes con puntuaciones bastante correctas exceptuando fuentes relacionadas con la observación identificada en este capítulo. Esto ha sido porque la calibración ha pecado un poco de infravalorar el que una fuente maneje información de varios controles. Con la nueva ponderación de mejora aplicada, dando mayor importancia al atributo *Multicontrol*, se ha visto aumentada la diferencia entre fuentes multicontrol y fuentes específicas favoreciendo las primeras.

Que haya aparecido solo una observación, lleva a pensar que la taxonomía ha cogido forma y es capaz de clasificar las fuentes de información asegurándole al responsable del cálculo del cuadro de mandos que la fuente que vaya a elegir para su cálculo es entre todas la mejor que puede elegir. Para ello, en el siguiente capítulo se pondrá a prueba la taxonomía de forma que se aplicará en el cálculo del cuadro de mandos.

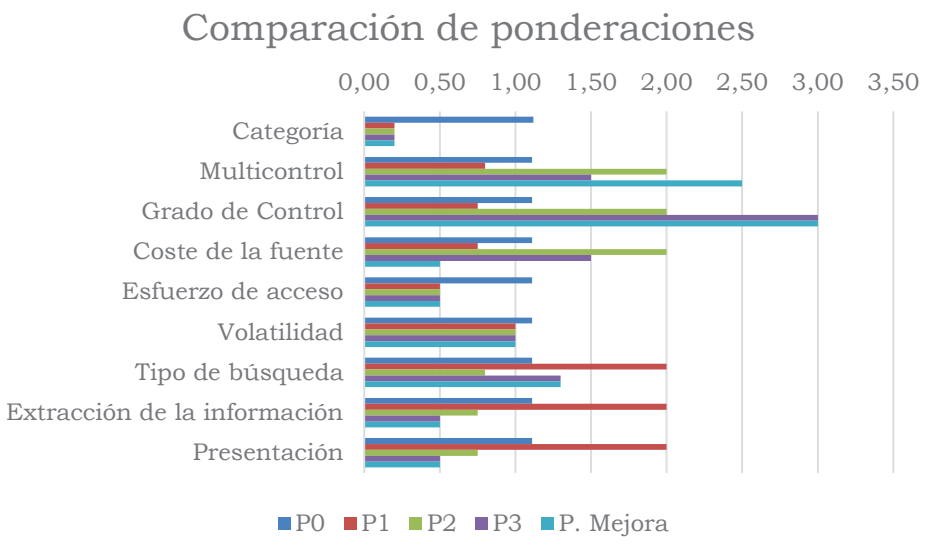


Tabla 14. Comparaciones de ponderaciones anteriores con la ponderación de mejora.

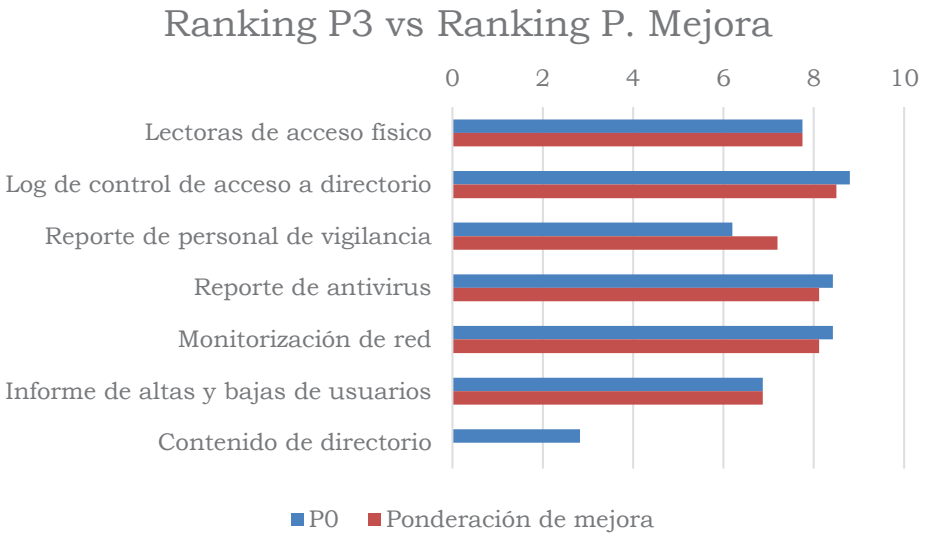


Tabla 15. Comparación ranking P3 con ranking de la ponderación de mejora.

En este conjunto mostrado de fuentes de información interesaba que la fuente *Reporte de personal de vigilancia*, fuente multicontrol¸ se desmarcara un poco más como pretendía la mejora y se ha conseguido.

3.5 Selección de indicadores de seguridad reales para su aplicación en la taxonomía. Eficiencia del proceso

El objetivo de este capítulo consiste en la medición de la eficiencia del proceso del cálculo de indicadores reales con fuentes de información actuales y con fuentes prometedoras con la intención de detectar mejoras en el cuadro de mandos gracias a la taxonomía que se ha elaborado durante todo este trabajo.

Para ello, se llevarán a cabo las siguientes tareas:

1. Medición del coste de cálculo y aporte de los indicadores con fuentes actuales e identificación de oportunidades de mejora en los indicadores mediante la utilización de fuentes propuestas por la taxonomía.
2. Optimización de los indicadores de SI y medición del coste cálculo y aporte de los indicadores con las nuevas fuentes propuestas por la taxonomía.

Antes de continuar con el desarrollo de este capítulo, se ha de comentar que no se mostrarán todos los indicadores que forman el cuadro de mandos sino un conjunto pequeño de estos que servirá para comprobar si la aplicación de la taxonomía ha sido efectiva o no.

3.5.1 Medición del coste de cálculo y aporte de los indicadores con fuentes actuales

Para llevar a cabo esta tarea se ha realizado una medición completa del esfuerzo de cálculo actual de un cuadro de mandos de indicadores de seguridad con las fuentes actuales, de forma que por cada indicador calculado se ha medido el tiempo de cálculo del mismo y se ha valorado el aporte de valor del indicador en los términos que lo requiere el capítulo 9.1 de ISO 27001 y la filosofía de mejora continua de la norma.

La medición del tiempo de cálculo permitirá comparar el coste de obtención del indicador actual, frente al coste de obtención de ese indicador u otro equivalente basándonos en nuevas fuentes.

Se describe el aporte de valor del indicador como necesario para conocer el beneficio que obtiene la organización al obtenerlo.

De esta forma, obtenidos coste y beneficio, puede medirse la eficiencia del indicador en el marco de un proceso de medición de la seguridad. Y pueden identificarse mejoras en el mismo, utilizando una nueva fuente prometedora este aporte de valor podría ser diferente pudiendo despertar mayor interés para la toma de decisiones y pudiendo modificar su criticidad para la toma de decisiones a crítica. Se recuerda aquella frase cuando se hablaba de los cuadros de mando en la Introducción de este trabajo, que decía lo siguiente *"los cuadros de mando si bien son útiles como herramienta de evaluación también están listos para su mal uso"*.

3.5.2 Identificación y valoración de Fuentes de información candidatas

Una vez, se tuvieron calculados estos datos, se pasó al cálculo de la variable de, *Posibilidad de Mejora*. Este cálculo ayudó a identificar los indicadores que por

orden de prioridad estaban “pidiendo auxilio”. Esta variable de *Posibilidad de Mejora*, dependía del producto de dos datos, tal que:

$$Posibilidad\ de\ Mejora = Impacto\ en\ caso\ de\ fallo * Tiempo\ de\ cálculo\ del\ indicador$$

$$Impacto\ en\ caso\ de\ fallo\ \epsilon\ \{1, 5, 10\}$$

Los valores de 1, 5 y 10 de la variable de *Impacto de fallo* se justificaron con la repercusión que podían tener sobre los objetivos de seguridad definidos en la organización en caso de que los indicadores fallasen en su cálculo.

De entre los indicadores de seguridadcalculados, se han escogido los siguientes indicadores con posibilidad de mejora:

INDICADOR	**INDICADOR ACTUAL**	**Fuente de Información**
1.Profundidad de testeo en seguridad de aplicaciones internas	$\frac{Grado\ testeo_{aplicación}interna}{Total\ Aplicaciones\ testadas}$	Informe de auditoría de calidad de proyectos
2.% Acciones correctivas cerradas	$\frac{\#\ AC\ cerradas}{\#AC\ totales}$	DDBB de herramienta de gestión
3.grado de formación actualizada	1-(#personas sin formación en últimos 6 meses)/todo el personal	Informes de Desarrollo profesional
4. Antigüedad del análisis de riesgos	Verificación manual	Análisis de riesgos

Tabla 16. Muestra de indicadores con posibilidades de mejora.

En el ejercicio realizado de medición de tiempo de cálculo de indicadores de seguridad, los valores obtenidos para estos indicadores fueron los de la **7**. Estos valores se utilizarán como referencia para calcular las posibilidades de mejoras de las fuentes alternativas.

ID	**TIEMPO**	**IMPACTO**	**POSIBILIDAD DE MEJORA**
1	111 min	5	555
2	30,14 min	5	150,7
3	6,38 min	10	63,8
4	0,5 min	5	2,5

Tabla 17. Mediciones de la muestra de indicadores con posibilidades de mejora.

Una vez identificados los indicadores con posibilidades de mejora, se propusieron las siguientes fuentes nuevas tal que:

INDICADOR	**FUENTE ACTUAL**	**FUENTE NUEVA**
1. Profundidad de testeo en seguridad de aplicaciones internas	Persona	Extracción directa de información desde Herramienta de gestión de pruebas
2.% Acciones correctivas cerradas	DDBB de herramienta de gestión (manual)	DDBB de herramienta de gestión (Script)

INDICADOR	**FUENTE ACTUAL**	**FUENTE NUEVA**
3. Grado de formación actualizada	Persona	Consulta directa a directorio de empleados
4.Antigüedad del análisis de riesgos	Análisis de riesgos	DDBB de herramienta de gestión (Manual)

18. Selección de fuentes nuevas para indicadores

Para valorar cada una de las fuentes, se ha realizado una estimación del tiempo que se prevé que se obtendría en caso de utilización de esta fuente.

ID	**TIEMPO**	**MEJORA**
1	0 min	0 min
2	2,5 min	27,6 min
3	0,2 min	6,2 min
4	0,3 min	0,2 min

Tabla 19. Mediciones de los indicadores con fuentes nuevas

¿Alguna de las fuentes nuevas prometió una mejora en el cuadro de mandos, concretamente, en el cálculo del indicador correspondiente? Durante esta tarea siempre se tuvo en mente la idea de que el futuro cuadro de mandos tendría que seguir *los principios de mínimos costes de cálculo y mayor precisión o control del indicador sobre la representación de la efectividad de un control de seguridad de la información implantado en la organización*. A partir de esto, parece ser que se pudieron encontrar varias mejoras en los indicadores dirigidas a cumplir estos principios mediante fuentes nuevas y a través de tres escenarios diferentes:

1. **La fuente nueva es efectiva.** Este escenario demuestra que la aplicación de la fuente prometedora por sí sola, en el cálculo del indicador, ofrece una mejora considerable. Este es el escenario más deseado.
2. **La fuente nueva es opcional.** Este escenario demuestra que la aplicación de la fuente prometedora por sí sola, en el cálculo del indicador, ofrece una mejora mínima pudiendo ser opcional su utilización.
3. **La fuente nueva requiere de automatización.** Este escenario demuestra que la aplicación de la fuente prometedora por sí sola, en el cálculo del indicador, no es suficiente y necesita ir acompañada de herramientas que sepan explotarla. Este escenario conlleva estimar el coste de transición para descubrir si es conveniente o no utilizar la nueva fuente automatizada.

Como se ha visto en la última tabla,

- Utilizando la nueva fuente para el indicador *4.* se da el caso del escenario 2. La mejora en tiempo ha sido de 2 décimas de minuto. ¿Merece la pena el cambio de fuente de información para obtener una mejora de 20 segundos de diferencia en el nuevo escenario? Teniendo en cuenta que

en todo el cuadro de mandos en unos 5 indicadores han ocurrido casos parecidos, el ahorro en tiempo de cálculo sería de unos 2,5 minutos. Esto no supone una mejora considerable por lo que aquí el cambio de fuente podría ser opcional.

- Para el indicador *2.* se da el caso del escenario 1. La mejora en tiempo ha sido de 27,64 minutos, por lo que es evidente que la fuente ofrecida por la taxonomía supera en calidad a la fuente que se había utilizado hasta ahora. Al igual ocurre con el indicador 3.

Estos dos casos comentados hasta el momento, no suponen ningún coste adicional, sino que simplemente con un cambio directo de fuente basta para que el cuadro de mandos se beneficie.

Por otro lado, con el indicador 1 , se experimenta el escenario 3. Para este indicador no existen más fuentes de información prometedoras que las que la taxonomía ofrece y esto supone que la información que se necesita para calcularlo o se obtiene de la fuente utilizada hasta el momento o se utiliza esa única fuente prometedora. El utilizar en este caso la nueva fuente supone que para extraer la información de esta se necesitan herramientas de extracción que sepan explotarla. Esto quiere decir que, al no tener estas herramientas de explotación disponibles, ya no se tiene una aplicación directa de la fuente en el cuadro de mandos sino que hay un coste de transición de por medio que será el que decida si merece la pena o no realizar esta transición. Por supuesto, que si estas herramientas estuvieran ya disponibles, la mejora estaría garantizada. Para comprender la envergadura de estas herramientas y que el coste de transición no es algo que se tiene que pasar por encima sino que es un punto muy importante a mirar, se va a presentar de forma breve el concepto de *ETL* y en su continuación se estimará este aparecido coste de transición.

Una ETL viene de las palabras, Extraer (*extract*), Transformar (*transform*) y Cargar (*load*), y para entenderla es necesario conocer el funcionamiento y algunas claves de cada una de las etapas, así como comprender las medidas de seguridad que se tendrían de tomar en cada una de estas fases para evitar las consecuencias negativas que un mal proceso de extracción, transformación o carga de datos puede producir.

Los procesos ETL son aquellos que se encargan de:

- **Traer** los datos de una o múltiples fuentes de información.
- **Reformatear** los datos para adecuarlos a un formato uniforme y común.
- **Limpiar** los datos para evitar problemas de calidad de los datos.
- **Analizar** los datos.
- **Cargar** los datos en su lugar destino.
- **Reutilizar** los datos en cualquier otro momento.

El primer proceso, es el **proceso de extracción** y tiene el objetivo de **traer los datos** de origen, **ya sea de una única fuente o múltiples fuentes de información**. Se deben de analizar y realizar procesos de calidad sobre los datos para ver si se satisfacen todas las restricciones de calidad que son impuestas a estos. El objetivo de esta fase de extracción es preparar los datos para la siguiente fase. Se debe recordar que los datos pueden tener diferentes formatos y estructuras y que existen diferentes tipos de datos, como los *datos estructurados (formato tabla)*, los *datos semiestructurados (formato JSON)* y los *datos no estructurados (párrafo).* Aunque no es objeto de este trabajo adentrar en estos detalles. Volviendo al proceso, una cosa muy importante a tener en cuenta es que el proceso de extracción debe causar un impacto mínimo en el sistema, ya que si se están extrayendo muchos datos, este proceso puede

ralentizar el sistema generando pérdidas. Por ende, muchas veces este proceso se planifica en horas en las que el sistema no tiene mucha demanda o carga de trabajo.

El segundo proceso, es el **proceso de transformación**. En este se deben aplicar funciones o reglas sobre los datos extraídos. Estas funciones o reglas de negocio tienen que ser:

- Declarativas
- Independientes entre sí
- Claras
- Útiles para el objetivo deseado

En esta fase, los datos se podrán seleccionar, traducir a códigos, obtener nuevos valores, unir con otros datos de otras fuentes de información, generar nueva información, etc.

Finalmente, **el proceso de carga**, se encarga de llevar los datos al sistema destino. En esta fase se tiene que tener en cuenta si los datos tienen que ser sobrescritos, se deben duplicar o se quiere añadir nueva información. En muchos casos, se ha de tener presente que puede bastar con resumir los datos y almacenar un promedio de una magnitud considerada. También, sean de aplicar todas las restricciones de calidad definidas:

- Unicidad
- Campos obligatorios
- Rangos
- Integridad

El proceso de carga se puede realizar de dos formas:

- **Acumulación Simple:** consiste en realizar un resumen de las últimas transacciones realizadas en un periodo seleccionado aplicando un promedio o un sumatorio y transportarlo hacia el sistema destino o *data warehouse*, ya sea con una frecuencia periódica o no.
- **Rolling:** es recomendable cuando se quiere almacenar con diferentes niveles de granularidad (totales diarios, semanales, etc.) o diferentes jerarquías dentro de unas determinadas dimensiones de la base de datos (acciones correctivas cerradas por periodo o por determinado etiquetado que las clasifique, etc.)

Sea cual sea la forma elegida y teniendo en cuenta que estas operaciones se realizan sobre el sistema destino, se tienen que aplicar todas las restricciones de calidad que se hayan definido sobre el sistema destino. Si están bien definidas, la calidad de los datos en el proceso ETL se garantiza.

¿Sería una ETL una buena oportunidad de mejora para el cuadro de mandos? No hay una respuesta única, sino que varía entre indicadores. Una ETL es una buena optimización para el cálculo del cuadro de mandos, si todas las ETLs capaces de explotar la fuentes están implementadas y se tienen disponibles o incluso si estas son reutilizables. Como se ha visto no se tienen ninguno de estos casos, por lo que se tiene que estimar el coste de esta transición. Para facilitar el entendimiento del fin que se busca con la estimación del coste de transición, se expondrá un ejemplo sencillo previo a la estimación del indicador 1. Por ejemplo, si se tuviese una fuente actual con un coste de tiempo de 2 horas, para estimar el coste de transición se tendrá en cuenta, el coste fijo de cuánto cuesta desarrollar la ETL y cuánto cuesta su utilización. Si el desarrollo de la ETL tiene un coste de 20 minutos más 5 minutos de explotación de la primera

extracción, serían 25 minutos frente a 2 horas de utilización de la fuente actual. Esto quiere decir que la ETL debería de ser desarrollada porque desde la primera vez que se utiliza se le está ganando tiempo y dinero. Ahora, si implementar la ETL lleva 1 semana, es decir 40 horas, más 5 minutos de explotación de la primera extracción son 40,083 horas frente a 2 horas de utilización de la fuente actual. Se están perdiendo 38,083 horas. ¿Qué pasaría por ejemplo con la extracción de datos número 20? En la extracción número 20 para la fuente actual se habrían empleado 40 horas de extracción de datos mientras que utilizando la ETL se habrían empleado 40 horas de desarrollo más 20 veces 5 minutos son 1,66 horas, lo que daría un total de 41,66 horas. Es decir, el punto en el que las horas empleadas en desarrollar y espera de la extracción de datos compensa con el seguir o no con la fuente actual ocurre en la extracción de datos número 21. La rentabilidad de esta acción quedará supeditada a la frecuencia de extracción de indicadores: sería muy eficaz en caso de extracción de información en tiempo real, mientras que será menos eficaz en caso de indicadores que se computan semanal, mensual o anualmente. En todo caso, una ETL se rentabiliza mejor si puede amortizarse en menos tiempo. Debe considerarse también en este caso que la evolución de la seguridad en las compañías y la necesaria adaptación al cambio de la organización, sus objetivos, las circunstancias, la evolución tecnológica, … puede derivar en cambios en los controles de seguridad aplicados y en las fuentes de información disponibles. Si esto ocurre durante el tiempo de amortización, la rentabilidad del cambio de fuente puede quedar en entredicho. La clave está en las veces que se tiene que recurrir a la herramienta para la extracción de los datos. Este número de veces es el que hará que el tiempo en amortizar los costes de la ETL sea permisible o no. Se busca más calidad en menos tiempo por lo que esta fuente que necesita ir acompañada de una ETL no sería una buena opción y es preferible continuar las mediciones con la fuente actual.

Ahora sí, ¿merecerá la pena la implementación de ETLs con un tiempo medio de implementación de 40 horas y tiempo medio de extracción de los datos de 5 minutos para el indicador 1?

INDICADOR	**TIEMPO EXTRACCIÓN CON FUENTE ACTUAL**	**TIEMPO DE PRIMERA EXTRACCIÓN**	**# EXTRACCIÓN DE COMPENSACIÓN O BREAK EVEN**
1. Profundidad de testeo en seguridad de aplicaciones internas	1,85 horas (111min)	0,083 horas (5 min)	Extracción número 22

Tabla 20. Estimación coste transición para el indicador 1.

Como muestran las estimaciones, no es conveniente la utilización de ETLs para este indicador. Se tendría que esperar 6 iteraciones del cálculo del cuadro de mandos para amortizar el tiempo y dinero empleados. Si se dieran casos en los que el tiempo de amortización fuera mucho menor, merecería la pena estudiar las estimaciones más en detalle de forma que los tiempos de amortización se vieran reducidos y variables como el número de cálculos del cuadro de mandos, que durante todas las estimaciones se ha mantenido en 4, si se tiene una ETL para la fuente de información con mayor grado de utilización, el número de cálculos del cuadro de mandos podría ser mayor a 4 y con ello se podría reducir el tiempo de amortización a tales puntos que la ETL desarrollada supusiera una optimización para el cálculo del cuadro de mandos. Al final, todas estas

variables dependen de situaciones, necesidades y decisiones que se tienen que tomar en la organización y que no tienen cabida en el presente trabajo.

Con cada uno de los escenarios encontrados se ha demostrado que la taxonomía es útil y no solo con sustituciones directas de nuevas fuentes de información mejores en el cuadro de mandos sino también con propuestas de fuentes de información prometedoras, que aunque puedan conllevar costes de transición y convenga o no utilizarlas (cosa que hay que estudiar) según los resultados, no quita que la propuesta de fuente no sea buena de por sí que es la función de la taxonomía, exponer un ranking de clasificación de las fuentes de información.

3.6 Líneas futuras

Como líneas futuras se ha querido comentar el papel de la Inteligencia Artificial y Big Data en el mejor aprovechamiento de las fuentes de información.

Como se sabe Big Data son datos que exceden la capacidad de procesamiento de los sistemas de bases de datos convencionales. Los datos son demasiado grandes, se mueven demasiado rápido o no se ajustan a las estructuras de las arquitecturas de las bases de datos existentes. Para obtener valor de estos datos, se debe elegir una forma alternativa de procesarlos. Big Data trata de abordar nuevos desafíos donde técnicas como la inferencia estadística, las bases de datos tradicionales y procedimientos estándar del business intelligence se quedan cortos. Big Data alcanza situaciones en las que se están produciendo datos demasiado rápido (velocidad), en una cantidad extremadamente grande (volumen) y de muchas fuentes heterogéneas (variedad). Podría ser interesante aprovechar el poder del Big Data y la Inteligencia Artificial, concretamente el área de esta llamado Machine Learning, para optimizar el cuadro de mandos de seguridad de la organización, de forma que sean aprovechadas en su totalidad cada una de las fases que se trabajan en estos campos; el *procesamiento de los datos* para trabajar con datos que provengan de fuentes de información heterogéneas de forma que se almacenen sus datos y estos se puedan gestionar y operar; el *análisis de los datos* de forma que a través de la extracción de conocimiento se ayude a la creación y redefinición de indicadores de seguridad haciendo que estos sean más precisos y así más representativos de la efectividad de los controles de seguridad que se apliquen y de la realidad del estado de seguridad de la organización; y por último la *visualización* de forma que la información saliente de las fases anteriores se represente visualmente ayudando al entendimiento y mejor comprensión del estado de la seguridad y favoreciendo y apoyando así la toma de decisiones en la organización. De hecho, actualmente existen los llamados robots RPA o Automatización de Procesos Robóticos implementados de forma personalizada y que tienen la extraordinaria función de liberar a los profesionales de tareas tediosas y repetitivas, lo que ayuda a aumentar la eficiencia operativa. El uso de estos robots RPA para gestionar cualquier proceso de negocio, en este caso el de aseguramiento de la información, transformaría y optimizaría el flujo de trabajo del departamento además de reducir el índice de errores humanos y los costes asociados. Por lo tanto, la fusión de la IA con los robots RPA permitirían automatizar procesos cognitivos y en este caso en el cuadro de mandos de la SI ganando en tiempo de cálculo de indicadores, control de la efectividad de los controles de seguridad y facilidad en la toma de decisiones.

4 Resultados y conclusiones

Una vez construida la taxonomía de fuentes de información a través de la identificación de atributos exigibles a las fuentes, en base a los criterios definidos aplicables a los atributos, y mejorarla y validarla mediante las fuentes de información nuevas se ha obtenido una lista de fuentes de información prometedoras correctamente puntuadas. Esta taxonomía se caracteriza por ser sostenida y de calidad debido a la cantidad de iteraciones de análisis realizadas en la selección de sus atributos y las calibraciones llevadas a cabo en las diferentes ponderaciones de estos sin olvidar nunca los principios de *mínimos costes de cálculo y mayor precisión o control del indicador sobre la representación de la efectividad de un control de seguridad de la información implantado en la organización*.

Una vez medida la eficiencia del proceso mediante el caso real del cálculo de un cuadro de mandos de una organización, se puede decir que la taxonomía ha sido útil en mayor o menor grado en cada uno de los escenarios encontrados en el cálculo de indicadores de seguridad. El escenario en el que la taxonomía es muy útil es el escenario que se basa en la sustitución directa de una fuente actual utilizada para el cálculo de un indicador por una fuente prometedora propuesta por la taxonomía. Este es el escenario más deseable porque no supone ningún coste adicional y mejora considerablemente los tiempos de cálculo del cuadro de mandos. También se han dado escenarios en los que la mejora en tiempos ha sido mínima por lo que la sustitución de la fuente hasta el momento utilizada por la fuente prometedora sería opcional. Los demás escenarios en los que la fuente de información prometedora tiene que ir acompañada de ETLs demuestran que la taxonomía tiene su utilidad si estas ETLs ya están desarrolladas y se pueden utilizar mejorando así los costes cálculo del cuadro de mandos, sin embargo, si estas ETLs no están implementadas, la taxonomía, aunque que dé por buena la fuente de información, simplemente estaría informando de que la fuente sería útil si estuviera implementada aunque no se pueda utilizar debido a los costes de transición y el tiempo de amortización. Por lo tanto, se concluye con este trabajo de investigación que la taxonomía funciona.

Algunas conclusiones parciales sobre aspectos relevantes durante la ejecución de este trabajo de investigación han sido la buena extracción y selección de atributos objetivos y exigibles a todas las fuentes de información con la dificultad de evitar características o aspectos no útiles para el fin de la taxonomía; la orientación favorable de la taxonomía hacia un cuadro de mandos basado en controles del estándar de certificación ISO 27001:2013; la capacidad de la taxonomía de dar espacio a la adición de nuevas fuentes de información y la sorpresa de alguna que otra fuente muy prometedora por la taxonomía y por la que el responsable del cálculo de indicadores no apostaba demasiado, resultando en una mejora en tiempo importante y demostrando la calidad de la taxonomía.

Como opinión personal, la realización de este trabajo de investigación ha sido muy provechosa, ya que, me ha permitido como responsable con poca experiencia del cálculo de indicadores de seguridad de la información conocer con mayor profundidad las virtudes y defectos del cuadro de mandos sobre el que he estado trabajando, ganando en visión de juego de este y con ello facilitar la detección de oportunidades de mejora.

Considero que esta es y ha sido una gran oportunidad para exponer lo aprendido en mi experiencia en las prácticas dentro de un SGSI y en la asignatura del MUII de Evaluación y Aseguramiento de Sistemas de Información, ya que, la seguridad de la información no tiene apenas protagonismo en los planes de estudio de grado y máster definidos, o por lo menos desde mi experiencia en todos los planes de estudio en los que he tenido la oportunidad de aplicarme. Además, ha sido una gran ayuda para continuar mejorando en mi trabajo.

5 Bibliografía

[1] AENOR ISO/IEC 27001:2013. Disponible: https://www.aenor.com/normas-y-libros/buscador-de-normas/ISO?c=054534

[2] AENOR ISO/IEC 27002:2013. Disponible: https://www.aenor.com/normas-y-libros/buscador-de-normas/ISO?c=054533

[3] AENOR ISO/IEC 27004:2016. Disponible: https://www.aenor.com/normas-y-libros/buscador-de-normas/ISO?c=064120

[4] Rana Khudhair, Abbas Ahmed. (December 2016). "Overview of Security Metrics". Disponible: http://www.sciencepublishinggroup.com/journal/paperinfo?journalid=237&doi=10.11648/j.se.20160404.11

[5] Instituto Nacional de Ciberseguridad. Disponible: https://www.incibe.es/protege-tu-empresa/catalogo-de-ciberseguridad/listado-soluciones/cuadro-mando-seguridad-0

[6] Carabaña, Julio. (15 nov 2019). "¿Ley de Campbell o duendes informáticos?". Disponible: https://elpais.com/sociedad/2019/11/15/actualidad/1573840634_649464.html

[7] Waldron, Kathryn. (October 2019). "Resources For Measuring Cybersecurity – A partial annotated bibliography". *RStreet.* Disponible: https://www.rstreet.org/wp-content/uploads/2019/10/Final-Cyberbibliography-2019.pdf

[8] NIST 800-55. Disponible: https://www.nist.gov/publications/performance-measurement-guide-information-security

[9] PowerData, blog (6 de junio de 2017). "El valor de la gestión de datos, ¿Qué son los procesos ETL?". Disponible: https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/qu-son-los-procesos-etl

[10] Ramírez, Vicente. (17 diciembre 2019). "GMV apuesta por la solución RPA de Automatización Anywhere". Disponible: https://cybersecuritynews.es/gmv-apuesta-por-la-solucion-rpa-de-automation-anywhere/

[11] Eckerson, Wayne. (July 17, 2016). "Tem Characteristics of a Good KPI." Disponible: https://www.bpmpartners.com/2016/07/17/characteristics-of-a-good-kpi/