



UNIVERSITAT  
POLITÈCNICA  
DE VALÈNCIA



Escola Tècnica  
Superior d'Enginyeria  
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica  
Universitat Politècnica de València

# Creación de una guía para el uso desde una perspectiva ética del Big Data en las empresas

Trabajo Fin de Grado

**Grado en Ingeniería Informática**

**Autor:** Francisco Javier Gil-Terrón Rodríguez

**Tutor:** Juan Vicente Oltra Gutiérrez

**Directora experimental:** María de los Ángeles Rodríguez  
Sánchez

Curso 2020-2021



# Resumen

---

Con la incesante expansión de las tecnologías, en la actualidad se están manifestando problemas relacionados con la privacidad y la seguridad de la información que exceden del ámbito técnico, penetrando cada vez más en la esfera de la ética y la moral. Estos problemas representan una amenaza real no solo para el sector empresarial y tecnológico, sino también para la sociedad. Las distintas organizaciones, poseedoras mayoritariamente de los datos, deberán reaccionar ante los cambios que se están produciendo en su entorno con el fin de alcanzar una cultura enfocada a prácticas que minimicen los riesgos asociados a esta situación.

Este trabajo de final de grado tratará de resolver dicha problemática, o al menos facilitar su resolución, mediante la creación de una guía para el uso ético de los datos en las empresas, y más específicamente del Big Data, basada fundamentalmente en recursos oficiales tales como guías de la Agencia española de protección de datos, normas ITIL, etc., dónde se contemplarán áreas como la privacidad o la transparencia.

Además, de forma complementaria se desarrollará una aplicación web con objeto de hacer más accesible su uso a las empresas, en la que se podrá acceder a la guía en su totalidad o a cualquiera de sus partes de forma independiente, y a través de un cuestionario la propia empresa podrá evaluar su situación actual con respecto al problema planteado y observar la evolución en el tiempo con evaluaciones anteriores.

**Palabras clave:** ética, big data, privacidad, guía, aplicación.

# Abstract

---

With the incessant expansion of technologies, the emerging problems related to privacy and information security reach beyond the technical sphere, increasingly penetrating the ethics and morals field. These problems mean a real threat not only to the business and technology sector, but also to society. The different organizations, which mainly own the data, must react to the changes that are taking place in their environment in order to achieve a culture focused on practices that minimize the risks associated to this situation.

This final degree project will try to solve those problems, or at least facilitate its resolution, by creating a guide for the ethical use of data in companies, more specifically Big Data, based mainly on official resources such as guides of the Spanish Data Protection Agency, ITIL standards, etc., where areas such as privacy or transparency will be considered.

In addition, a web application will be developed in order to make its use more accessible to companies, where the guide can be accessed in its entirety or any of its parts independently, and through a questionnaire the company will be able to assess its current situation regarding the problem raised and observe the evolution over time with previous evaluations.

**Keywords:** ethic, big data, privacy, guide, app.

## Resum

---

Amb la incessant expansió de les tecnologies, en l'actualitat s'estan manifestant problemes relacionats amb la privacitat i la seguretat de la informació que excedeixen de l'àmbit tècnic, penetrant cada vegada més en l'esfera de l'ètica i la moral. Aquests problemes representen una amenaça real no sols per al sector empresarial i tecnològic, sinó també per a la societat. Les diferents organitzacions, posseïdores majoritàriament de les dades, hauran de reaccionar davant els canvis que s'estan produint en el seu entorn amb la finalitat d'aconseguir una cultura enfocada a pràctiques que minimitzen els riscos associats a aquesta situació.

Aquest treball de final de grau tractarà de resoldre aquesta problemàtica, o almenys facilitar la seua resolució, mitjançant la creació d'una guia per a l'ús ètic de les dades en les empreses, i més específicament del Big Data, basada fonamentalment en recursos oficials com ara guies de l'Agència espanyola de protecció de dades, normes ITIL, etc., on es contemplaran àrees com la privacitat o la transparència.

A més, de manera complementària es desenvoluparà una aplicació web a fi de fer més accessible el seu ús a les empreses, en la qual es podrà accedir a la guia íntegrament o a qualsevol de les seues parts de manera independent, i a través d'un qüestionari la pròpia empresa podrà avaluar la seua situació actual respecte al problema plantejat i observar l'evolució en el temps amb avaluacions anteriors.

**Paraules clau:** ètica, big data, privacitat, guia, aplicació.

# Tabla de contenidos

---

<b>1. Introducción.....</b>	<b>9</b>
1.1 Motivación.....	10
1.2 Objetivos.....	11
1.3 Estructura.....	12
<b>2. Estado del arte.....</b>	<b>13</b>
2.1 Ética.....	13
2.2 Big Data.....	14
2.2.1 Origen.....	14
2.2.2 Características.....	15
2.2.3 Actualidad.....	16
2.3 Legislación.....	17
2.4 Crítica al estado del arte.....	17
<b>3. Análisis del problema.....</b>	<b>19</b>
3.1 Identificación y análisis de soluciones posibles.....	21
3.2 Solución propuesta.....	22
3.3 Análisis de requisitos.....	23
3.4 Plan de trabajo.....	28
3.5 Presupuesto.....	29
<b>4. Diseño de la solución.....</b>	<b>31</b>
4.1 Arquitectura del sistema.....	31
4.2 Diseño de la base de datos.....	32
4.3 Tecnología utilizada.....	37
4.3.1 Firebase.....	37
4.3.2 PyCharm.....	38
4.3.3 Python.....	38
4.3.4 Django.....	39
4.3.5 Bootstrap.....	40
4.3.6 JotForm.....	40
4.3.7 Chart.js.....	40
4.3.8 Docker.....	41
<b>5. Desarrollo de la solución propuesta.....</b>	<b>42</b>
5.1 Aplicación web.....	42
5.2 Página principal.....	45
5.3 Autodiagnóstico.....	47



<b>6. Implantación</b> .....	<b>53</b>
<b>7. Pruebas</b> .....	<b>54</b>
7.1    Innsai .....	54
7.2    Bankia Forward .....	54
7.3    Rpet Flake .....	55
<b>8. Conclusiones</b> .....	<b>56</b>
8.1    Relación del trabajo desarrollado con los estudios cursados .....	57
<b>9. Trabajos futuros</b> .....	<b>58</b>
<b>Bibliografía</b> .....	<b>59</b>
<b>Anexos</b> .....	<b>61</b>
Anexo I. Código del proyecto. ....	61
Anexo II. Guía para el uso desde una perspectiva ética del Big Data. ....	61

# Índice de ilustraciones

---

Ilustración 1 - UML. Fuente: Elaboración propia. ....	27
Ilustración 2 - Plan de trabajo. Fuente: Elaboración propia. ....	29
Ilustración 3 - Arquitectura del sistema. Fuente: Elaboración propia. ....	31
Ilustración 4 - Diseño conceptual. Fuente: Elaboración propia. ....	32
Ilustración 5 - Tabla usuarios. Fuente: Elaboración propia. ....	35
Ilustración 6 - Tabla documentos. Fuente: Elaboración propia. ....	35
Ilustración 7 - Tabla consulta. Fuente: Elaboración propia. ....	36
Ilustración 8 - Tabla evaluación. Fuente: Elaboración propia. ....	36
Ilustración 9 - Almacenamiento de documentos. Fuente: Elaboración propia. ....	37
Ilustración 10 - Página de presentación 1. Fuente: Elaboración propia. ....	42
Ilustración 11 - Página de presentación 2. Fuente: Elaboración propia. ....	43
Ilustración 12 - Página de presentación 3. Fuente: Elaboración propia. ....	44
Ilustración 13 - Página de presentación 4. Fuente: Elaboración propia. ....	44
Ilustración 14 - Inicio de sesión. Fuente: Elaboración propia. ....	45
Ilustración 15 - Inicio de sesión 2. Fuente: Elaboración principal. ....	45
Ilustración 16 - Aplicación 1. Fuente: Elaboración propia. ....	46
Ilustración 17 - Aplicación 2. Fuente: Elaboración propia. ....	46
Ilustración 18 - Aplicación 3. Fuente: Elaboración propia. ....	46
Ilustración 19 - Aplicación 4. Fuente: Elaboración propia. ....	47
Ilustración 20 - Autodiagnóstico 1. Fuente: Elaboración propia. ....	48
Ilustración 21 - Autodiagnóstico 2. Fuente: Elaboración propia. ....	49
Ilustración 22 - Autodiagnóstico 3. Fuente: Elaboración propia. ....	50
Ilustración 23 - Autodiagnóstico 4. Fuente: Elaboración propia. ....	51
Ilustración 24 - Autodiagnóstico 5. Fuente: Elaboración propia. ....	51
Ilustración 25 - Autodiagnóstico 6. Fuente: Elaboración propia. ....	52

# Índice de tablas

---

Tabla 1- Plantilla. Fuente :(García-Peñalvo, 2019) .....	23
Tabla 2 - Iniciar sesión. Fuente: Elaboración propia. ....	24
Tabla 3 – Registrar documento. Fuente: Elaboración propia. ....	25
Tabla 4 – Modificar documento. Fuente: Elaboración propia. ....	25
Tabla 5 – Eliminar documento. Fuente: Elaboración propia. ....	26
Tabla 6 - Realizar consulta. Fuente: Elaboración propia. ....	26
Tabla 7 - Realizar evaluación. Fuente: Elaboración propia.....	27
Tabla 8 - Presupuesto. Fuente: Elaboración propia.....	29
Tabla 9 - Presupuesto Software. Fuente: Elaboración propia. ....	30



# 1. Introducción

---

Desde que la humanidad hace uso de la razón, la ética, disciplina que estudia el comportamiento humano y tiene como propósito guiarlo hacia lo que es bueno mientras lo desvía de lo que es malo (Savater, 1991), se ha presentado como un concepto subjetivo y hasta *propio* de cada individuo, pues existen tantas concepciones del bien como personas. A pesar de esto, no implica que los valores en una sociedad diverjan en función del sujeto, al contrario, pues la ética ha servido como marco común estableciendo limitaciones morales y evolucionando a través de civilizaciones y generaciones, lo que dota de dinamismo a la ética en sí, pues ha ido transfigurándose a través del tiempo. En esencia, la ética ha vislumbrado lo que era correcto tanto sobre personas como instituciones estableciendo una especie de base normativa de carácter no obligatorio, en contraposición a lo que sugiere la legislación.

Pero este dinamismo que se ha comentado resulta en ocasiones insuficiente para adaptarse a otros campos que se desarrollen o progresen de manera acelerada, como es el caso de las nuevas tecnologías y por ende de Big Data, objeto del presente trabajo. En general, las sociedades se encuentran en entornos cambiantes, lo cual se acentúa en la actual denominada era de la información donde aparentemente se ha establecido un estándar de tendencia al cambio y que se muestra con más intensidad en el campo de las tecnologías. El progreso siempre ha portado consigo la aparición de conflictos éticos, y el dominio de los datos no es ni mucho menos una excepción como se desarrollará a continuación.

Entendemos Big Data como una constelación de fenómenos relacionados con la producción, consumo, recopilación y análisis de grandes colecciones de datos producidos por una gran variedad de fuentes, en gran número de formatos, en unas cantidades sin precedentes y que fluyen en red a gran velocidad, y cuya finalidad última es ser convenientemente analizados con propósito de extraer valor para empresas, gobiernos e instituciones públicas y privadas (Colmenarejo Fernández, 2017). En esta línea, no podemos hablar de Big Data sin hacer referencia a la generación masiva de datos y que los sistemas basados en esta tecnología han pasado de limitarse a la generación de información a tratarla de forma implícita, y han penetrado en nuestro entorno de una manera alarmante llegando a formar parte de la vida cotidiana en la esfera privada, como, por ejemplo, los tan populares asistentes de voz que se comercializan actualmente.

¿Pero dónde entra el sector empresarial en el asunto? Las empresas son los principales precursores del progreso y desarrollo tecnológico, y por supuesto del Big Data, por lo que gran cantidad de conflictos éticos han surgido dentro de estas. Este es el motivo de la existencia de los *Business ethics*, que constituyen el conjunto de principios morales orientados a la correcta toma de decisiones por parte de las organizaciones utilizando como fuente primordial la ética (incluso en el campo de Big Data).

Respecto a Big Data, las empresas se encuentran en la posición de sujeto moral desde el momento en que se convierten en las poseedoras de los datos, lo que conlleva en gran parte de los casos, su posterior uso, manejo y tratamiento. En este punto cabe destacar que se ha llegado a considerar que los sistemas de información son puramente objetivos

debido a su naturaleza automática y no dependencia humana, alcanzando un estado de inmunidad ética e imparcialidad idealizada. Este hecho llega a ser preocupante teniendo en cuenta que a raíz de la progresiva automatización "del todo", la toma de decisiones ha dejado de ser algo exclusivamente humano, lo que ha facilitado la elusión de responsabilidades morales. Aunque esa aparente neutralidad de los datos se ve comprometida si tomamos en consideración que estos provienen de una fuente concreta y se han obtenido mediante una metodología en particular, ambos, objeto de estudio ético. En este aspecto, pasamos de considerar dentro de lo moral no solo el uso que se le da a la tecnología, sino esta en sí misma.

De cualquier modo, si realmente Big Data trae consigo estos problemas que se han expuesto, ¿por qué la legislación no lo contempla? Recordando el desenfundado crecimiento de Big Data y de manera similar a como ocurría con la ética, existe un desfase temporal, en este caso debido a que el proceso de renovación de la jurisprudencia es lento y costoso. Bien es cierto que las instituciones intentan adaptar las leyes a la actualidad, como es el caso de la Ley Orgánica de Protección de Datos, la cual se encuentra en constante actualización, pero de nuevo, no es capaz de mantener un ritmo que vaya a la par con la innovación tecnológica, de hecho, este es un objetivo inalcanzable para la legislación. De esta manera, se permite la existencia de entornos en que la ley resulta demasiado laxa, o incluso inexistente.

En suma, en este trabajo de fin de grado se tratará de dar respuesta a esta problemática con la creación de una guía para el uso ético de Big Data en el entorno empresarial, complementada con una aplicación web para facilitar su utilización a los profesionales, permitiendo determinar la situación de la empresa en un momento dado con respecto a un marco ético adecuado, y evaluar la evolución en el tiempo.

## 1.1 Motivación

---

Una vez introducido qué es el Big data y cómo se relaciona con la ética y las empresas, se ahondará brevemente en las principales cuestiones éticas que atañen Big Data. Identidad, privacidad, propiedad y reputación son conceptos los cuales forman la estructura esencial para una ética aplicada a Big Data (Colmenarejo Fernández, 2017). Estos conceptos conforman el núcleo motivacional del proyecto, junto al establecimiento de un conjunto de principios éticos profesionales fundados en la justicia, como son la beneficencia y cumplimiento del trabajo y objetivos de forma profesional de acuerdo con un estándar, servicio del trabajo para un beneficio mayor más allá de este mismo, y autonomía y respeto hacia los individuos y su dignidad, intimidad y voluntad.

No obstante, la línea de acción que se propone ante la situación actual no es la de un posible endurecimiento o corrección de la ley, sino la concienciación y acercamiento de valores hacia la cultura organizativa, la cual determina el comportamiento general de las mismas empresas. Por este motivo, se busca el desarrollo de una cultura ética tanto por parte de las organizaciones como de los propios profesionales con el fin de enfocar la toma de decisiones hacia el bien común y alineado con los objetivos globales de la

Organización de las Naciones Unidas (ONU) de desarrollo sostenible. En suma, la motivación final se apoya en hacer de la revolución de los datos, uso de Big Data y de los sistemas y tecnologías actuales, mecanismos para lograr esos propósitos.

Por lo que respecta al objeto personal, está fundamentado en una serie de objetivos de aprendizaje ligados a la motivación profesional posterior. Como interesado en el dominio de los datos, y más en concreto en *Business Analytics*, los conocimientos sobre la presencia e importancia de la información en las empresas, así como de los principales instrumentos para su tratamiento como puede ser Python, constituyen una buena base con la que comenzar en este ámbito profesional. Asimismo, la elaboración de un trabajo basado en una perspectiva ética como es este, servirá de aliciente con el que orientar mi desarrollo profesional.

## 1.2 Objetivos

---

El objetivo principal del presente trabajo y del cual derivan el resto de los objetivos se cimenta en la mejora de la ética empresarial y el propósito de penetrar en la cultura, la cual supone el reflejo de la ética organizativa. Esta cultura influye y condiciona la toma de decisiones, y por supuesto el concepto que se tenga de privacidad y uso que se dará a los datos, así como de la consideración de los posibles efectos en los usuarios; razones que hacen de la cultura el centro de atención para alcanzar una ética adquirida basada en la transparencia y reflexión sobre futuras amenazas tecnológicas para la integridad ciudadana (ya sea por nuevo desarrollo o aplicación).

En otras palabras, el objetivo no es otro que proporcionar medios diligentes con el fin de guiar al sector tecnológico hacia las mejores prácticas en el terreno de responsabilidad y toma de decisiones. Así pues, se pueden identificar los siguientes objetivos específicos:

- Ofrecer una herramienta con la que la organización sea capaz de visualizar su situación actual en cuanto a uso ético de datos se refiere. Para ello se va a realizar una guía como documento principal de apoyo al profesional y una aplicación web complementaria.
- Establecer una base ética que permita el transcurso y crecimiento normal de la organización, desarrollando o ayudando a mejorar el marco ético de la compañía.
- Concienciar a la empresa de los riesgos a los que es vulnerable (intensidad y probabilidad de la vulnerabilidad), así como el posible daño que pudiera causar directa o indirectamente a través de estos.
- Mostrar que la utilización y tratamiento de Big Data e información suponen un problema real en la actualidad para las empresas.
- Proporcionar a la empresa un elemento de valía, no una norma a la que atenerse con el fin de permanecer dentro de la legislación.
- Incidir en la cultura organizativa de las empresas para la transmisión de valores a través de esta misma.

- Manifestar el valor generado por la ética para el conjunto de grupos de interés no limitados a la propia organización, con intereses más que simplemente monetarios.

## 1.3 Estructura

---

En los siguientes capítulos correspondientes al estado del arte y análisis del problema, se expondrá de manera sistemática toda la información y detalles en materia de ética, Big Data y la problemática que surge de la relación entre ambos con el fin de situar al lector en un contexto idóneo para el adecuado seguimiento y entendimiento de la solución propuesta. Asimismo, se hará un breve estudio a nivel tecnológico del contexto actual y evolución de las aplicaciones web y se mencionará superficialmente la legislación pertinente y se arrojará luz sobre qué aporta la propuesta actual en contraste con posibles alternativas o trabajos anteriores.

Por lo que respecta a los capítulos de diseño y desarrollo de la solución, se compondrán del conjunto de consideraciones previas al desarrollo de la aplicación a nivel tecnológico y el desarrollo en sí mismo de la misma aplicación y de la guía. Tras esto y una vez finalizada la solución, se llevará a cabo la implantación del proyecto y las posteriores pruebas para la validación de este.

Finalmente, en la sección de conclusiones se realizará un análisis de forma retrospectiva del conjunto del trabajo, evaluando las dificultades encontradas y el cumplimiento de los objetivos, tanto de aprendizaje como del proyecto en sí, y relacionándolo con los estudios cursados durante el grado.

Adicionalmente, en los anexos se añadirá la guía resultante final y aquellos fragmentos de código de la aplicación desarrollada que resultaran interesantes comentar a lo largo de la memoria.

## 2. Estado del arte

---

### 2.1 Ética

---

Los orígenes de la ética se remontan a los inicios de la propia reflexión filosófica, siendo la rama de esta misma que estudia la diferencia entre lo bueno y lo malo, lo correcto y lo incorrecto, y, en definitiva, la conducta humana en base al carácter y la costumbre (Savater, 1991). La ética ha estado presente en las sociedades desde su nacimiento, impulsando al motor de estas, las administraciones públicas, hacia lo que se presupone adecuado para el bien común, mientras evolucionaba con el paso del tiempo y de las civilizaciones. Pero, aunque se trate de un concepto subjetivo y dependiente del marco, tanto temporal como geográfico, en esencia se ha consolidado como fundamento de toda práctica y raíz de todos los actos humanos (Colmenarejo Fernández, 2017).

Esta filosofía moral de la que hablamos se puede dividir en distintas ramas, pero en el presente trabajo se indagará sobre la ética aplicada al profesionalismo, y más en concreto, en el dominio de los datos. De esta surge la noción de deontología, la cual contempla los principios morales adscritos a una profesión. En este sentido, y en relación con los datos, un ejemplo podría ser la ética de la información en el campo biomédico, donde la privacidad se presenta como objeto de estudio moral puesto que la confidencialidad es uno de los pilares de la relación entre profesionales y pacientes: Cuando se realiza divulgación o debates sobre problemas de salud puede ser muy útil ilustrarlo con casos reales, pero, al hacerlo, a veces no se contempla el riesgo que representa para la preservación de la confidencialidad del paciente (Fosas, 2019). Existen infinidad de éticas aplicadas como esta, e incluso incipientes en la actualidad, sin ir más lejos, la “ciber-ética” en internet, vinculada a relaciones sociales, libertad de expresión y censura en redes sociales; o la ética en manejo y utilización de datos en entornos digitales, motivación principal del trabajo.

Por lo que respecta a la ética contextualizada en el entorno empresarial y de Big Data, la aplicación de esta misma en la toma de decisiones está tomando una tendencia cuanto menos atractiva, y no solo para aquellos grupos de interés que se vean directamente afectados por esta (como puede ser el propietario de los datos), ya que cada vez más, se entiende la ética como generador de valor en una organización. De esta forma, nacen las denominadas empresas 3.0, que son aquellas que sitúan la ética como base de su cultura y ofrecen valor adicional, tanto a sí mismas como al resto de interesados, a través de su responsabilidad social corporativa. Comportamientos de esta índole han llegado a ser exaltados públicamente por gobiernos e instituciones públicas, y cada vez son más las fuentes de financiación que se ven atraídas por empresas con este valor añadido.



## 2.2 Big Data

---

En términos generales, se puede entender Big Data como el conjunto de datos que no pueden ser procesados por bases de datos estándar debido a su tamaño o complejidad, en otras palabras, cuando los mecanismos clásicos no permiten arreglar un problema de tratamiento, visualización o almacenaje, suele ser problema de Big Data (Zarza, 2017).

### 2.2.1 Origen

---

En la historia de los datos, se pueden observar una serie de hitos que encaminaron el desarrollo y nacimiento de Big Data, el cual ha ido evolucionando y ampliando su espectro desde entonces por las distintas generaciones.

El nacimiento del Big Data fue producto de las bases de datos relacionales de finales de la década de los 90 e inicios del siglo XXI, las cuales no eran capaces de manejar las cantidades de datos que se estaban generando.

Ante esta tesitura, el procesamiento masivo de datos originó un nuevo tipo de bases de datos, las bases de datos no relacionales. Fue entonces cuando se popularizó la técnica de *clustering* ante los volúmenes de datos, técnica basada en el concepto de escalabilidad mediante la interconexión de pequeñas máquinas para lograr una capacidad de cómputo global mayor, y que podría ser calificada como el paso hacia el auténtico Big Data, pues la inmensa mayoría de tecnología NoSQL está pensada para *clusters* y estos sistemas dieron lugar a lo que hoy conocemos como CPDs o centros de procesamiento de datos, lugar en que muchas grandes empresas alojan sus datos, de forma directa o a través de la nube mediante arrendamiento.

Pero la tecnología NoSQL no supuso la desaparición de las bases de datos relacionales, ni mucho menos, pues desde 2010 han seguido suponiendo más del 70% de las bases de datos mundiales. Con esto, se formó un entorno potente y flexible donde ambas opciones son válidas, aunque bien es cierto que la tendencia es que el uso de la tecnología relacional se está decrementando de manera progresiva con el tiempo.

De hecho, entre 2000 y 2010 empresas como Google abandonaron parcialmente el modelo relacional por el NoSQL para poder manejar, sin ir más lejos, la información asociada a la búsqueda de todas las páginas de Internet, decisión a la que se sumaron Amazon y Facebook entre otros. De nuevo, hay que recordar que no son excluyentes, todas estas empresas siguieron utilizando tecnología relacional con otros fines. También hay que destacar que el estandarizado uso de NoSQL en Big Data no implica que NoSQL sea utilizado únicamente en este ámbito, pues resulta de gran utilidad en otros campos.

A partir de 2010 se extendió el procesamiento *NRT* o cercano a tiempo real, que trataba principalmente problemas de conjuntos de datos que se incrementan de forma constante en el tiempo. En este momento ya se buscaban arquitecturas escalables y tolerantes a fallos. En esta etapa el término Big Data ya no es una idea desconocida ni alejada de la realidad, de ahí la especulación sobre la definición y significado de este, pues las

empresas e instituciones comienzan a cerciorarse de los volúmenes y velocidades alcanzados en la generación y tratamiento de datos.

La nueva generación que llega hasta la actualidad suele ir estrechamente vinculada a la aplicación de técnicas de inteligencia artificial, aprendizaje automático y extracción de conocimiento, y el auge de estas tecnologías, entre otras, hace que el resto se intensifique.

### 2.2.2 Características

---

Big Data se caracteriza principalmente por las conocidas “3 V” introducidas en el año 2001 por Doug Laney (Laney, 2001). Estas son Volumen, Velocidad y Variedad:

- **Volumen:** Quizá la más importante y que dota de nombre al propio concepto. Hace referencia a la gran cantidad de datos que se ven involucrados, pues la cantidad de información generada ha crecido de forma exponencial en las últimas décadas. En la actualidad, generamos datos constantemente incluso sin ser conscientes de ello, por ejemplo, llevando el móvil se están enviando señales sobre la ubicación, al usar internet permitimos publicidad personalizada, recomendaciones de compra... en definitiva, partimos de un paradigma en que cada operación o transacción genera un efecto archivable.
- **Velocidad:** Nace de la relación entre la necesidad de eficiencia en cuanto al manejo y tratamiento de datos cada vez generados más rápido, en otras palabras, se refiere tanto a la rapidez con la que se generan los datos, como de transmisión y almacenaje, así como la velocidad a la que deberían ser procesados para mantener el flujo. Hemos pasado de un ámbito en el que se primaba la generación de datos (pues resultaba una tarea costosa), al paradigma que vivimos actualmente, donde la información fluye a un ritmo desenfrenado. Un ejemplo de esto es que, al realizar una publicación en una red social, se espera que se haga de forma instantánea.
- **Variedad:** De fuentes de información y de la forma misma de los datos. Esta diversidad de formatos fue la dura arremetida contra la consistencia en que se basaba el modelo relacional, pues los sistemas clásicos no son capaces de escalar la cantidad de información, pero sobre todo la cuantía de tipos de datos que se suele encontrar en las bases de datos.

A estas “3 V” Doug Laney añadió posteriormente otras dos, más referentes al entorno, con el fin de adaptarlo al contexto actual. Estas fueron, Veracidad y Valor:

- **Veracidad,** relacionado con los inputs, determina el grado de certeza de estos mismos. Esta se encuentra fuertemente relacionado con el siguiente punto y con el objetivo de Big Data, pues de nada sirven los datos si no se puede garantizar que su contenido es correcto y fiable.
- **Valor,** en contraposición a la veracidad, pues se trata de una categoría de output del resultado de la aplicación de Big Data. En este punto cabe destacar que el objetivo final de Big Data no es la generación masiva y almacenamiento de datos, pues el valor reside en la información y conocimiento que se puede extraer a partir de estos. Este es el principal motivo por el que se ha intensificado la presencia de Big Data en las empresas, pues hoy en día las organizaciones



comparten la meta de aprovechar sus datos y obtener valor real a partir de estos mismos.

### 2.2.3 Actualidad

---

Desde que entendemos los datos, información y conocimiento como sistemas dinámicos no unidireccionales (se relacionan unos con otros), las tomas de decisiones se han basado en estos mismos. Así pues, hoy por hoy Big Data se ha convertido en una necesidad estratégica empresarial, y núcleo de los principales saltos tecnológicos de transformación digital, en definitiva, los datos suponen una innegable fuente de valor para las empresas, de hecho, gran parte de la expansión de Big Data ha sido promovida por fines de marketing. Pero estas decisiones trascienden más allá del sector de los negocios, estando los datos involucrados en cambios económicos, sociales y políticos. De esta manera, Big Data se está consolidando como fuente de innovación, abriendo nuevas fronteras para mejorar la calidad de vida, como predicciones de enfermedades, meteorológicas...

En la práctica, Hadoop y MongoDB son las dos tecnologías más importantes utilizadas en este ámbito, ambas basadas en modelo NoSQL. Como ya se había comentado anteriormente, el uso de tecnología NoSQL no implicaba el abandono total de la tecnología relacional, al contrario, son complementarias, pues en muchas ocasiones el uso de Big Data mediante NoSQL, manteniendo el sistema relacional ya presente en la empresa, aporta valor adicional, por ejemplo, agregando una capa de negocio. No obstante, se hará hincapié de forma superficial únicamente en aquellos sistemas enfocados expresamente en Big Data.

Hadoop utiliza almacenamiento mediante ficheros distribuidos y procesamiento mediante MapReduce, modelo que da soporte a computación paralela sobre grandes volúmenes de datos distribuidos entre los núcleos de un *cluster*. Además, al ser software libre y sumado a la posibilidad de ser utilizado en la nube, hacen que el uso de Big Data sea accesible para prácticamente cualquier empresa, lo que lo han hecho destacar como herramienta dentro del amplio abanico de posibilidades.

Por otro lado, MongoDB, también Software libre, es un sistema de bases de datos orientado a documentos y enfocado al almacenamiento y acceso masivo de datos, en contraposición a la opción anterior, la cual se focaliza en el procesamiento. Esta herramienta proporciona una integración rápida y sencilla de los datos, siendo ampliamente utilizada en procesos industriales.

A pesar de que existan instrumentos muy extendidos como los que se han comentado, la tendencia actual es que los ciclos de vida de estos sean particularmente acelerados en el ámbito de Big Data, causando que herramientas aparezcan y desaparezcan en periodos de tiempo pequeños.



## 2.3 Legislación

---

En cuanto la Norma se refiere en el contexto de las tecnologías y la información, en los últimos años han aparecido marcos legales en distintos ámbitos territoriales, sin ir más lejos, el Reglamento General de Protección de Datos o RGPD de marzo de 2018; así como numerosas guías oficiales de apoyo al profesional, como la guía de Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial o las Orientaciones Para Prestadores de servicios de *'cloud computing'*. Sin embargo, estos recursos se contemplan desde la perspectiva del cumplimiento normativo, siendo esta la premisa que prevalece en sus objetivos.

Por otro lado, en referencia a la ética, ciertas entidades tanto públicas como privadas han desarrollado herramientas con este enfoque, como La Gestión ética de los datos, por el Banco Interamericano de Desarrollo. No obstante, suelen tener un ámbito de aplicación demasiado específico, o, por el contrario, ser demasiado generalistas como para tener una aplicabilidad real que se materialice en las empresas.

Aunque estos proyectos no estén alineados totalmente con los objetivos que se persiguen en este trabajo, su estudio exhaustivo será de gran utilidad para establecer una base para la guía y definir las recomendaciones que se recogerán en ella.

## 2.4 Crítica al estado del arte

---

Tras un estudio de los trabajos directamente relacionados con el campo que se trata presentados con anterioridad en la escuela, se puede observar que son varias las guías que han sido elaboradas, de entre las cuales se podría citar guías de auditoría para el cumplimiento de normas ISO o distintas guías para la aplicación, control del cumplimiento o apoyo en la transición de la legislación. No obstante, dichos trabajos se apoyan directamente en la Ley y no se han desarrollado desde una visión ética, pues, aunque exista una fuerte vinculación entre ambas como ya se había comentado, el objetivo final que se persigue no es el mismo en función del enfoque.

Por otro lado, también se encuentran otros proyectos como el análisis y clasificación de códigos éticos vinculados a la informática, más alineado con la temática principal actual, aunque debido a su naturaleza, carecen del componente didáctico y orientado a ofrecer recomendaciones para las mejores prácticas que se busca actualmente.

Cabe destacar que gran parte de estos trabajos que son fundamentalmente teóricos, carecen casi por completo de cualquier componente tecnológico, dejando un vacío en el espacio de conocimiento que el alumno es capaz de alcanzar y limitando la posibilidad de evidenciar los conocimientos adquiridos durante el grado.

En cuanto a otros trabajos de final de grado de fuera de la Universidad, ha habido algunos que abordan el tema de las implicaciones éticas que tiene Big Data e entornos profesionales o cómo genera valor en las empresas, mientras que otros se enfocan en analizar la norma desde la perspectiva del derecho internacional privado.

## Creación de una guía para el uso desde una perspectiva ética del Big Data en las empresas

En cualquier caso, ninguno de estos trabajos ofrecía una aplicación para darle uso a la guía desde las empresas, y aquellos trabajos que sí disponían de una aplicación distaban demasiado del actual tema.

### 3. Análisis del problema

---

Con lo anteriormente expuesto, nos encontramos en la actualidad ante la tesis de una revolución basada en la información que trasciende al sector empresarial, llegando a la llamada era de los datos a formar parte del entorno tanto público como privado, más allá de la comunicación y el desarrollo tecnológico de las organizaciones e incorporando las tecnologías en la vida cotidiana de los ciudadanos. Frente a este nuevo paradigma, se presentan nuevos conflictos éticos, y aunque el ámbito de Big Data no sea el único campo que ha supuesto que el progreso abra nuevos dilemas morales (alimentos transgénicos, por ejemplo), este se caracteriza por el entorno dinámico en el que se encuentra, también afectado por la ética, y que somete a constante cambio el marco actual, desde el uso que se les da a los datos hasta las metodologías aplicadas para su obtención.

Los principales problemas que surgen, giran en torno a la identidad, privacidad, propiedad y reputación de los datos e individuos, como ya se había comentado con anterioridad, aunque no son ni mucho menos los únicos, pues se recuerda que el dinamismo que impera en este campo provoca la constante aparición de nuevos campos de estudio moral, y que todas las fases Big Data, recolección, procesamiento y difusión, pueden ser objeto de ausencia de ética o ilegal como espionaje o tratamiento no consentido, siendo la última etapa la más peligrosa de todas a causa de posibles terceros indebidos o no permitidos.

Como se expondrá a continuación, estos conceptos están estrechamente relacionados, y a raíz de estos mismos se puede observar uno de los grandes problemas provocados por la automatización global de la toma de decisiones: la evasión de responsabilidades.

Por lo que respecta a la identidad del sujeto, esta puede verse modificada en función del uso que haga de la tecnología en línea. La huella digital es una realidad hoy en día, siendo las redes sociales un reflejo de nuestra propia identidad, lo que puede resultar preocupante si se tiene en consideración que el usuario tiene la posesión parcial de los datos, en relación con la propiedad, pues es el administrador de la red quién decide finalmente qué hacer con esa información (por ejemplo, al borrar una cuenta de una red social, los datos se almacenan en lugar de ser eliminados, simplemente se ocultan al público). Esta digitalización de la identidad la convierte en un elemento fluido y fácilmente moldeable, es decir, pasa a ser un componente cambiante en la red. Por otro lado, el concepto de identidad, tanto a nivel institucional como de las organizaciones, está fuertemente ligado a la reputación del individuo, la cual, a su vez, también dependerá del uso que hagan de sus datos. Pero no es solo el usuario quién se expone a conflictos éticos relacionados con la identidad, pues al otro lado del consumidor, la identidad de la fuente que genera los datos permanece en el anonimato y se permite el desvanecimiento parcial de la responsabilidad que conlleva, pues generalmente desempeñan estas funciones máquinas y algoritmos.

Por otro lado, en materia de privacidad, la primera cuestión destacable reside en el conflicto entre privacidad y seguridad, pues a mayor nivel de seguridad requerido, mayor es el riesgo potencial de minorar la privacidad en el sentido de que se requiere control sobre sí mismo, de esta forma podrían verse de cierta manera como inversamente



proporcionales. Pero, al igual que la sociedad, la noción tanto de privacidad como de seguridad, ha ido evolucionando con el paso del tiempo, pues no son excepción en la incommensurable lista de elementos que se han visto influenciados por la revolución tecnológica. Así pues, mecanismos y herramientas utilizadas para garantizar la seguridad han cambiado mucho en las últimas décadas, convergiendo hacia la simplificación y accesibilidad de las tareas de observación, y provocando del mismo modo que la esfera privada del individuo se expanda y se entrelace con la red. De esta manera, los límites que establece la privacidad parecen haberse visto difuminados ante la marea de información, punto donde vuelve a aparecer la elusión de responsabilidades, pues son máquinas principalmente las que realizan estas tareas. Esto ocurre gracias al enorme incremento de la capacidad en tráfico y almacenamiento de datos, que se ha manifestado en tecnologías capaces incluso de tomar decisiones propias. Estas tecnologías presentan nuevos problemas de privacidad sobre qué información podrían compartir y con quién, de nuevo, sometidos a su toma de decisiones automática.

En lo que se refiere a propiedad, no es el usuario quien posee en la práctica los datos, si no que los cede de manera voluntaria, por lo que se puede llegar a interpretar que no es el usuario el verdadero propietario de su información, alejándose de la toma de decisiones final sobre sus datos y ofreciendo la posibilidad de decidir por parte de los administradores sobre el uso que se va a dar a la información. Esto se acentúa si tenemos en cuenta el uso de técnicas de replicación sobre los datos con fines de seguridad, lo que significa que esos datos podrían estar reflejados en varias copias y estas a su vez estar alojadas en distintos países, prácticas que ponen en duda quién posee realmente los datos o qué uso se les va a dar, abriendo más conflictos sobre la privacidad y facilitando el abuso de poder sobre la información. Aquí surge una dualidad de conflictos, pues la replicación supone una potencial transmisión de datos a terceros, y, por el contrario, la ausencia de esta se basa en la centralización en el almacenamiento de los datos, lo que aumenta el riesgo de posibles ataques o robos.

Por último, en cuanto a reputación, el principal problema surge cuando esta escapa de la red, formada enteramente por datos, sobrepasando la barrera de internet e influenciando en la reputación real del individuo u offline. Este hecho podría atentar directamente contra la dignidad, ya que la reputación conforma un elemento inherente a esta misma, y la utilización de datos puede suponer efectos muy negativos sobre la reputación del usuario si se vulnera la privacidad del susodicho, afectando también a su intimidad. Dichos acontecimientos ocurren con frecuencia desde que la exposición de datos personales está totalmente normalizada (en redes sociales, por ejemplo), sobre todo entre jóvenes, lo que abre la posibilidad de manipular sus datos independientemente de la intención, de manera que las decisiones sobre la información quedan fuera de su control.

Con todo esto, lo que se pretende exponer es que el manejo que se le da a los datos de las personas tiene implicaciones reales para estas mismas, siendo estas negativas e incluso nefastas en ocasiones. Las cuestiones anteriores se aúnan y extienden en el actual contexto caracterizado por la rápida difusión y fácil accesibilidad de la información, donde la generación de datos forma parte del marco rutinario del ciudadano a través de su teléfono móvil, asistentes de voz... recopilando información incluso de dominio íntimo, como puede ser su religión o ideología política. Además, los posibles fallos en los inputs pueden ser propagados rápidamente, casi imposibilitando su enmendación. En la

mayoría de los casos, el individuo no es siquiera consciente de que está generando estos datos, por lo que también desconoce el tratamiento o finalidad que se les está dando, dando lugar a situaciones en que ocurre sin su consentimiento explícito y donde se aprovecha la ignorancia del usuario a favor de los intereses de quien realiza estas prácticas.

Cabe recordar que este conjunto de actividades que se han mencionado, hoy por hoy las realizan casi exclusivamente máquinas, habiendo sido reemplazado el factor humano progresivamente. Conforme se ha estandarizado el uso de tecnologías y conocimientos técnicos avanzados como la inteligencia artificial, más dificultades se han presentado para el entendimiento de las responsabilidades éticas que hay detrás, y sumado a que la toma de decisiones ya no es un procedimiento únicamente humano en consecuencia a la automatización mediante tecnología y algoritmos, se constituyen las bases del paradigma actual, en el que se observa una frecuente evasión de responsabilidades y una habitual falta de consideración moral en la toma de decisiones referentes a Big Data.

Esto se traduce a un entorno donde las organizaciones poseen políticas de privacidad vagas con una casi inexistente transparencia en el análisis y manejo de datos, y el uso que se le da a Big Data es principalmente fomentar comportamientos consumistas influenciando al usuario mediante patrones de compra basados en su histórico de comportamiento, o ejercer presión hacia ideales concretos inclusive, sin tener en consideración la validez o veracidad de los mecanismos a través de los que se plantee. Pero después de todo, ignorar estas responsabilidades en el desarrollo de tecnología o innovación sería equivalente a hacerlo en su posterior uso, siendo así, la tecnología un componente social y no independiente.

En conclusión, las disrupciones tecnológicas centradas en datos nos obligan a replantearnos la manera en que entendemos elementos constitutivos de la realidad, cultura e idiosincrasia, teniendo consecuencias que exceden la escala de lo social. ¿Hasta qué punto es ético el resguardo y la explotación de datos sobre las personas? ¿Existen datos que no deberían obtenerse? ¿Cómo se garantizan estas limitaciones? (Zarza, 2017) En cualquier caso, la relación que surge entre ciudadanos y sistemas de información se presenta en un primer supuesto, indiferente a los objetivos de desarrollo sostenible. Con el progresivo crecimiento tecnológico, parece que se ha ido nublando la visión y objetivo a escala social de sostenibilidad y crecimiento hacia “lo bueno”, donde aparentemente los profesionales han estado más ocupados en atender la complicación técnica de la computación, que las implicaciones éticas de su propia actividad (Colmenarejo Fernández, 2017).

### 3.1 Identificación y análisis de soluciones posibles

---

Una vez desarrollado el problema que se plantea, la línea de actuación que sería más efectiva de manera inmediata consistiría en una solución mediante combinación de ética, ley y decisiones políticas. Sin embargo, este planteamiento escapa totalmente al alcance abordado. Con esto, sea cual fuere la solución, debería tratar de garantizar una evolución

legal que se adapte al cambio, y que permita a su vez que el desarrollo tecnológico logre un progreso adecuado en la empresa al tiempo que se van integrando las potenciales posibilidades de mejora. A continuación, se identificarán una serie de soluciones posibles, todas ellas desde una perspectiva viable y realista, con objeto de declarar que la elección de la solución propuesta no está fundamentada en su unicidad, sino en un conjunto de características que la hacen más atractiva que el resto.

La primera de estas alternativas corresponde al uso de fuentes de conocimiento como foros, blogs o redes de divulgación. El ámbito público de estos medios los convierte en una opción con una gran accesibilidad, además de que presenta unos costes ínfimos, tanto a nivel económico como temporal. En cambio, esta publicidad hace que los contenidos sean vulnerables a posibles sesgos lo que podría alterar completamente los resultados esperados, por no hablar de que este tipo de recursos difícilmente se incorporarán de manera formal en una organización. Es por esto por lo que las siguientes propuestas de solución se enfocarán desde el corazón de la problemática: las organizaciones.

En cuanto a la segunda opción consiste en la realización de talleres o cursos dentro de las empresas para la formación, capacitación y desarrollo del personal, práctica cada vez más extendida incluso en el campo de la ética, sobre todo por grandes corporaciones. Esta se presenta como la alternativa más costosa, pues pueden llegar a ser muy caros y requieren del compromiso y apoyo de la alta dirección y plantilla, en otras palabras, es necesaria de una adaptación previa para la empresa en cuestión, lo que incluye su toma en consideración en la estrategia corporativa. Pese a estas contras, si se realiza de manera adecuada puede suponer grandes incrementos en la motivación y rendimiento del personal, así como en su compromiso posterior.

Por lo que respecta a la tercera y última propuesta, la cual es la seleccionada de entre todas, se basa en la utilización de una guía enfocada a empresas para el uso ético de los datos y la información. Al ser un recurso que la entidad puede poseer como tal, tiene un mayor potencial de ser incluido directamente en la cultura y de ofrecer valor tanto a la empresa de forma interna como al exterior a través de las prácticas estudiadas en la guía. No obstante, el seguimiento de un simple documento puede resultar tedioso para algunos profesionales, por lo que, de forma complementaria, se incluirá la guía dentro de una aplicación web que hará el seguimiento de la guía más sencillo y atractivo, además de añadir ciertas funcionalidades.

## 3.2 Solución propuesta

---

Con el trabajo actual se plantea la creación de una herramienta tal que aúne de manera global una serie de competencias adquiridas, tanto técnicas como transversales, durante el periodo formativo del grado, de manera que las mismas queden de manifiesto al tiempo que se satisfacen los objetivos planteados.

De esta manera, la propuesta actual trata de ofrecer una guía fundamentada derechamente en el marco de la ética para el uso de los datos, concretamente Big Data, en las empresas. Considerando la privacidad como el principal problema en el área de

los datos y la imposibilidad de la ley para adaptarse al ritmo de crecimiento tecnológico, se pretende abordar la problemática desde su origen, el trabajo realizado por los profesionales del campo en cuestión, en lugar de elaborar una guía para el seguimiento de la legislación.

Así pues, se propone el desarrollo conjunto de una guía para el uso ético de datos, y más específicamente de Big Data, en el entorno empresarial, y de una aplicación donde se encuentre incluida dicha guía con el fin de facilitar su comprensión, hacerla más versátil y dotarla de usabilidad para lograr de un modo más sencillo que la organización en cuestión alcance las mejores prácticas en la materia y haciendo de la guía más que un mero documento, añadiendo valor sin olvidar la finalidad que se plantea y que comparte con el resto de las guías.

Como consecuencia de esta propuesta en concreto, el desarrollo vendrá definido por dos procedimientos claramente diferenciados: la guía, de naturaleza más teórica, y la aplicación, de carácter más técnico y práctico. Ambos serán precedidos por una etapa de investigación y serán desarrollados en paralelo durante la etapa de desarrollo de la solución propuesta. No obstante, la guía deberá ser finalizada con anterioridad para adaptar adecuadamente las funcionalidades de la aplicación. Finalmente, será evaluada y se realizarán pruebas de validación en la empresa Innsai con la colaboración directa de su CEO, quién supervisó personalmente las políticas de datos de la firma; además de trabajar con datos sensibles de grandes entidades bancarias entre otras empresas.

### 3.3 Análisis de requisitos

---

Siguiendo la metodología de Sommerville (Sommerville, 2005), se ha desarrollado el análisis de requisitos a partir de las funcionalidades. Se especificarán las características del sistema siguiendo la definición de análisis de requisitos del estándar IEEE de 1990, dónde se divide en requisitos funcionales, requisitos no funcionales y requisitos de datos. No obstante, en el presente TFG solo se abordarán los requisitos funcionales, es decir, aquellos que describen lo que el sistema debe hacer.

El conjunto de requisitos funcionales que exigirá el sistema la describirán en las tablas 2, 3, 4, 5, 6 y 7. Para la especificación de los requisitos se empleará la plantilla (García-Peñalvo, 2019) que se puede ver en la tabla 1. Para este caso en concreto, no serán necesarias todas las filas de la tabla, por lo que se seleccionarán los siguientes campos: Versión, Autores, Requisitos asociados, Descripción, Precondición, Secuencia normal y Comentarios.

Tabla 1- Plantilla. Fuente :(García-Peñalvo, 2019)

<b>CU &lt;id&gt;</b>	<b>&lt;Nombre descriptivo&gt;</b>
<b>Versión</b>	<nº de la versión actual>
<b>Autores</b>	<autor de la versión actual>



<b>Requisitos asociados</b>	<nombre del requisito>
<b>Descripción</b>	El sistema debe comportarse tal como se describe en el siguiente caso de uso {abstracto durante la realización de los siguientes casos de uso: <lista de casos de uso>, cuando <evento de activación> [o durante la realización de los siguientes casos de uso: <lista de casos de uso>]}
<b>Precondición</b>	<precondición del caso de uso>
<b>Secuencia normal</b>	P1 {El actor <actor>, El sistema} <acción/es realizada/s por actor/sistema> P2 Se realiza el caso de uso <caso de uso> P3 Si <condición>, {el actor <actor>, el sistema} <acción/es realizada/s por actor/sistema> P4 Si <condición>, se realiza el caso de uso <caso de uso>
<b>Poscondición</b>	<poscondición del caso de uso>
<b>Excepciones</b>	Pi Si <condición excepción>, {el actor <actor>, el sistema} <acción/es realizada/s por actor/sistema>, a continuación este caso de uso {continúa, queda sin efecto}
<b>Comentarios</b>	<comentarios adicionales del requisito>

Tabla 2 - Iniciar sesión. Fuente: Elaboración propia.

<b>CU 1</b>	<b>Iniciar sesión</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Registrar documento (CU 2), Modificar documento (CU 3), Eliminar documento (CU 4), Realizar consulta (CU 5), Realizar evaluación (CU 6)
<b>Descripción</b>	El sistema debe controlar qué usuarios acceden al aplicativo.
<b>Precondición</b>	-
<b>Secuencia normal</b>	P1 el usuario inicia sesión mediante unas credenciales P2 si estas son válidas accede a la aplicación P3 si no lo fueran no tendría acceso al sistema
<b>Poscondición</b>	-
<b>Excepciones</b>	-
<b>Comentarios</b>	-



Tabla 3 – Registrar documento. Fuente: Elaboración propia.

<b>CU 2</b>	<b>Registrar documento</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Iniciar sesión (CU 1), Realizar consulta (CU 5)
<b>Descripción</b>	El sistema debe permitir a los administradores del sistema añadir nuevos documentos para que los usuarios puedan acceder a ellos.
<b>Precondición</b>	Tener una sesión iniciada. Ser administrador del sistema.
<b>Secuencia normal</b>	P1 el usuario asigna un archivo nuevo P2 proporciona la información adicional necesaria
<b>Poscondición</b>	-
<b>Excepciones</b>	-
<b>Comentarios</b>	-

Tabla 4 – Modificar documento. Fuente: Elaboración propia.

<b>CU 3</b>	<b>Modificar documento</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Iniciar sesión (CU 1), Realizar consulta (CU 5)
<b>Descripción</b>	El sistema debe permitir a los administradores del sistema modificar los recursos ya registrados con anterioridad.
<b>Precondición</b>	Tener una sesión iniciada. Ser administrador del sistema.
<b>Secuencia normal</b>	P1 el usuario selecciona el documento a modificar P2 determina los elementos que serán modificados P3 proporciona la información que desee actualizar
<b>Poscondición</b>	-
<b>Excepciones</b>	-
<b>Comentarios</b>	-

Tabla 5 – Eliminar documento. Fuente: Elaboración propia.

<b>CU 4</b>	<b>Eliminar documento</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Iniciar sesión (CU 1), Realizar consulta (CU 5)
<b>Descripción</b>	El sistema debe permitir a los administradores del sistema la eliminación de recursos registrados con anterioridad.
<b>Precondición</b>	Tener una sesión iniciada. Ser administrador del sistema.
<b>Secuencia normal</b>	P1 el usuario selecciona el documento a eliminar P2 el documento será removido del sistema
<b>Poscondición</b>	-
<b>Excepciones</b>	-
<b>Comentarios</b>	-

Tabla 6 - Realizar consulta. Fuente: Elaboración propia.

<b>CU 5</b>	<b>Realizar consulta</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Iniciar sesión (CU 1)
<b>Descripción</b>	El sistema debe permitir el acceso a diferentes documentos a través de consultas, permitiendo la descarga del archivo por parte del usuario.
<b>Precondición</b>	Tener una sesión iniciada.
<b>Secuencia normal</b>	P1 el usuario realiza una consulta a un documento en concreto P2 el usuario obtiene el resultado de la consulta
<b>Poscondición</b>	-
<b>Excepciones</b>	-

Tabla 7- Realizar evaluación. Fuente: Elaboración propia.

<b>CU 6</b>	<b>Realizar evaluación</b>
<b>Versión</b>	1
<b>Autores</b>	Francisco Javier Gil-Terrón Rodríguez
<b>Requisitos asociados</b>	Iniciar sesión (CU 1), Realizar consulta (CU 5)
<b>Descripción</b>	El sistema debe ser capaz evaluar al usuario mediante una serie de preguntas y mostrar de manera clara los resultados, recomendando el acceso a unos documentos u otros en consecuencia de dichos resultados.
<b>Precondición</b>	Tener una sesión iniciada.
<b>Secuencia normal</b>	P1 el usuario accederá a una nueva evaluación P2 responderá a un cuestionario predefinido P3 visualizará los resultados de su evaluación P4 en función de las respuestas el sistema le recomendará realizar consultas específicas
<b>Comentarios</b>	La visualización gráfica de los resultados representa un factor importante.

Para finalizar el análisis de requisitos se presenta un diagrama UML (Orallo, 2002) donde se agruparán los distintos casos de uso, de manera que quede plasmado visualmente la forma en que los sujetos harán uso del software mediante la ilustración 1.

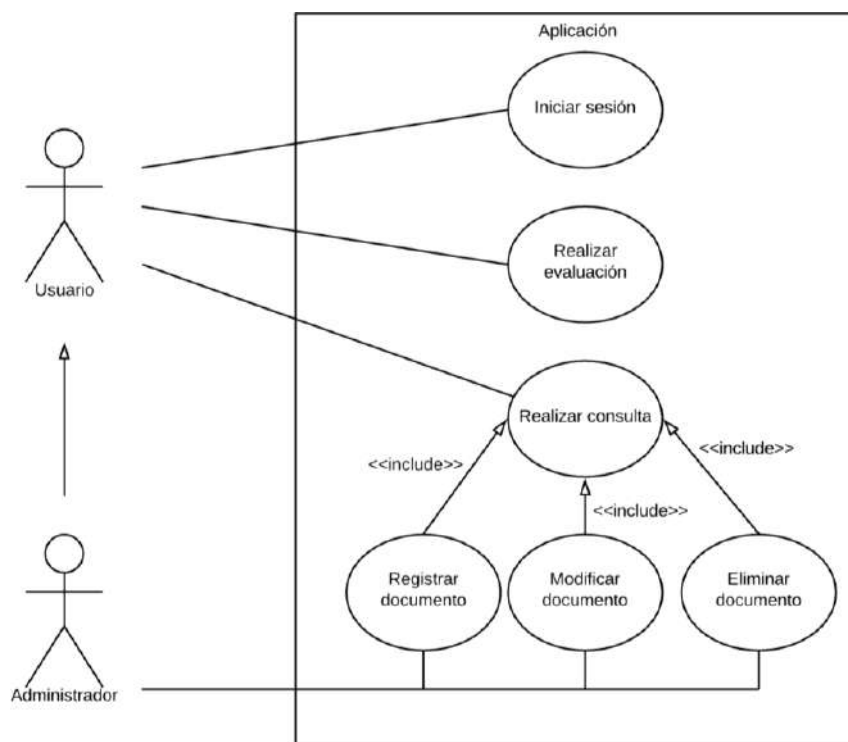


Ilustración 1 - UML. Fuente: Elaboración propia.

La operatividad del sistema está basada, principalmente, en la descarga de documentos por parte del usuario, en este caso la guía o las secciones en que esta se divide, las cuales serán descargables de manera independiente. Adicionalmente, el usuario podrá acceder a recursos oficiales en los que se ha basado la guía desarrollada en este trabajo, además de tener a su disposición un sistema de autodiagnóstico donde podrá realizar evaluaciones sobre su situación actual con respecto al marco ético propuesto, y acceder a ellas posteriormente, por lo que el sistema deberá proporcionar un visionado adecuado de esta información. Para ello, todo usuario deberá iniciar sesión, requisito necesario para acceder al sistema.

### 3.4 Plan de trabajo

---

El total de trabajo requerido para este TFG se estima en un horizonte temporal de unas 250 horas. El plan de trabajo inicial se distribuye de igual forma que la estructura recomendada por la facultad para la realización de los trabajos de fin de grado. Siguiendo esta línea de acción, la planificación temporal se ha elaborado con Microsoft Project y se distribuirá de la siguiente forma:

- Introducción: 30 horas
- Estado del arte: 40 horas
- Análisis del problema: 40 horas
- Diseño de la solución: 30 horas
- Desarrollo de la solución propuesta: 80 horas
- Implantación y pruebas: 15 horas
- Conclusiones y trabajos futuros: 15 horas

Se puede visualizar la disposición temporal de las tareas a través del siguiente diagrama de Gantt de la ilustración 2, donde se muestra en color azul el trabajo realizado por el alumno, y en naranja las tareas de supervisión y revisión de los tutores, cuya finalización marca los hitos del trabajo, da por validada cada sección y da comienzo a la siguiente.

Para la planificación de las jornadas de trabajo se asumen semanas de cinco días laborables. El comienzo del proyecto coincide con el inicio del mes de noviembre, donde se dedicará aproximadamente una hora y media cada día, asignando 20 días para la introducción y diseño de la solución, y 30 para el estado del arte y análisis del problema. Esta primera fase se centra en la redacción de la memoria, la profundización teórica en la materia y la investigación para dar soporte al proyecto en conjunto.

A partir de ese punto, dará comienzo el desarrollo de la solución, etapa en la que se incrementará el trabajo diario a dos horas durante un total de 40 días. Por último, durante el último mes, se retomará la dinámica de una hora y media para finalizar el trabajo con la implantación, pruebas, conclusiones y trabajos futuros.

Con un plan de trabajo tan dilatado en el tiempo con una duración de ocho meses en total, se pretende amortiguar los posibles imprevistos y llegar correctamente al plazo

planeado para la convocatoria de julio. Pese a esto, se comentarán desviaciones ocurridas en capítulos posteriores en caso de haberlas.

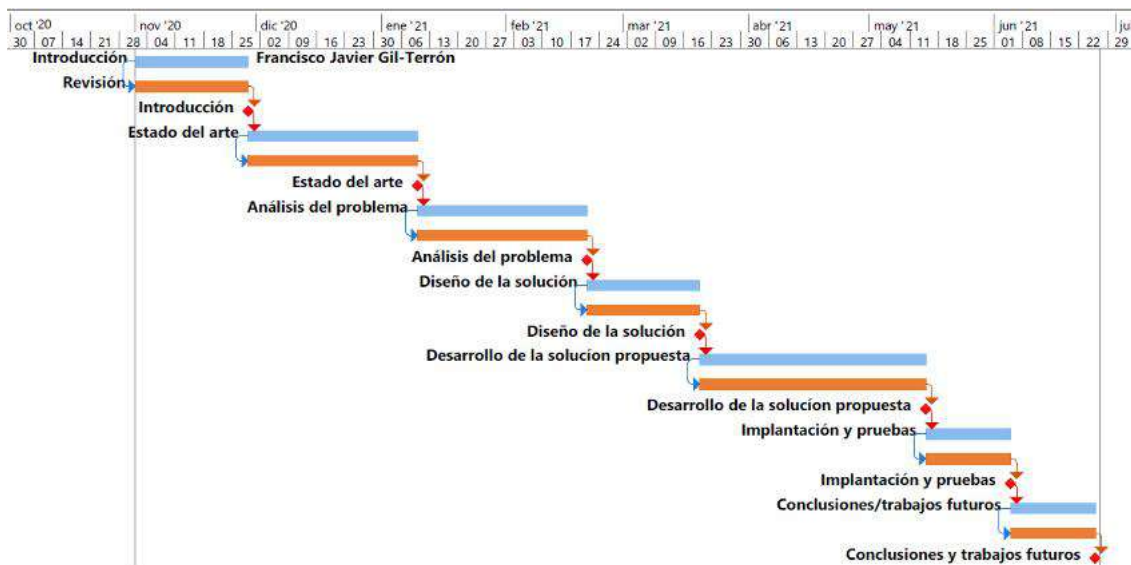


Ilustración 2 - Plan de trabajo. Fuente: Elaboración propia.

### 3.5 Presupuesto

En este apartado se tratará de estimar los costes referentes a los recursos empleados durante la elaboración completa del TFG, tanto humanos como técnicos, lo que incluye software y hardware. Para el apartado personal, tras un breve estudio de mercado se ha determinado un salario medio de 12.8 €/hora para un desarrollador software. Como se puede observar en la tabla 4 el coste total del proyecto se estima en 3.609€.

Tabla 8 - Presupuesto. Fuente: Elaboración propia.

Recurso	Observaciones	Coste
<b>Personal</b>	Atendiendo al plan de trabajo anterior (total de 250 horas) y estimando un precio de 12.8 €/hora.	<b>3200€</b>
<b>Hardware</b>	Amortización correspondiente a la utilización del equipo hardware estimando una vida útil de 4 años con un valor inicial de 900€.	<b>150€</b>
<b>Software</b>	Gastos incurridos en la utilización de todo elemento software.	<b>259€</b>
<b>Total</b>	<b>Personal + software + hardware</b>	<b>3609€</b>

La totalidad de las herramientas software empleadas se muestra en la tabla 5:

Tabla 9 - Presupuesto Software. Fuente: Elaboración propia.

<b>Software</b>	<b>Versión</b>	<b>Utilidad</b>	<b>Coste</b>
Windows 10	Versión Pro-64 bits.	Sistema operativo sobre el que se ha trabajado.	259€
Microsoft Word	Licencia académica.	Redacción y estructuración de la memoria.	0€
Microsoft Project	Licencia académica.	Elaboración del plan del trabajo y diagrama de Gantt.	0€
Mendeley	Licencia académica.	Generador de bibliografía y referencias.	0€
Overleaf	Versión personal sin coste.	Herramienta de edición LaTeX.	0€
Lucidchart	Versión gratuita.	Elaboración de esquemas, UML...	0€
Firebase	Plan Spark sin coste.	Base de datos y almacenamiento.	0€
Python	Versión 3.8 de código abierto.	Lenguaje base para el desarrollo de la aplicación.	0€
PyCharm	Versión Community gratuita.	Entorno de desarrollo de soporte Python.	0€
Django	Versión 3.0.8 de código abierto.	Framework de desarrollo web.	0€
Bootstrap	Versión 4.6 de código abierto.	Biblioteca de diseño web.	0€
JotForm	Versión inicial gratuita.	Creación de formularios en línea.	0€
Chart.js	Versión 2.0 de código abierto.	Generación de gráficas dinámicas.	0€
Docker	Versión 20.10.7 de código abierto.	Implantación y puesta en producción de la aplicación web.	0€
Adobe InDesign	Versión 16.1 bajo licencia.	Maquetación de la guía.	29,99€/mes

## 4. Diseño de la solución

---

En este capítulo se profundizará exclusivamente en el diseño de la aplicación web debido a la ausencia de tecnología en la parte correspondiente a la guía, la cual será mostrada en los anexos.

### 4.1 Arquitectura del sistema

---

A continuación, en la ilustración 3 se mostrará de forma esquematizada los elementos mayoritarios que componen la arquitectura software elegida para el sistema. Cabe destacar que las decisiones tomadas en materia de selección de tecnologías están principalmente motivadas por objetivos de aprendizaje, aprovechando la versatilidad de estas para adaptarlas a las necesidades y requisitos del proyecto actual. Así pues, la arquitectura se resume como muestra la siguiente imagen:

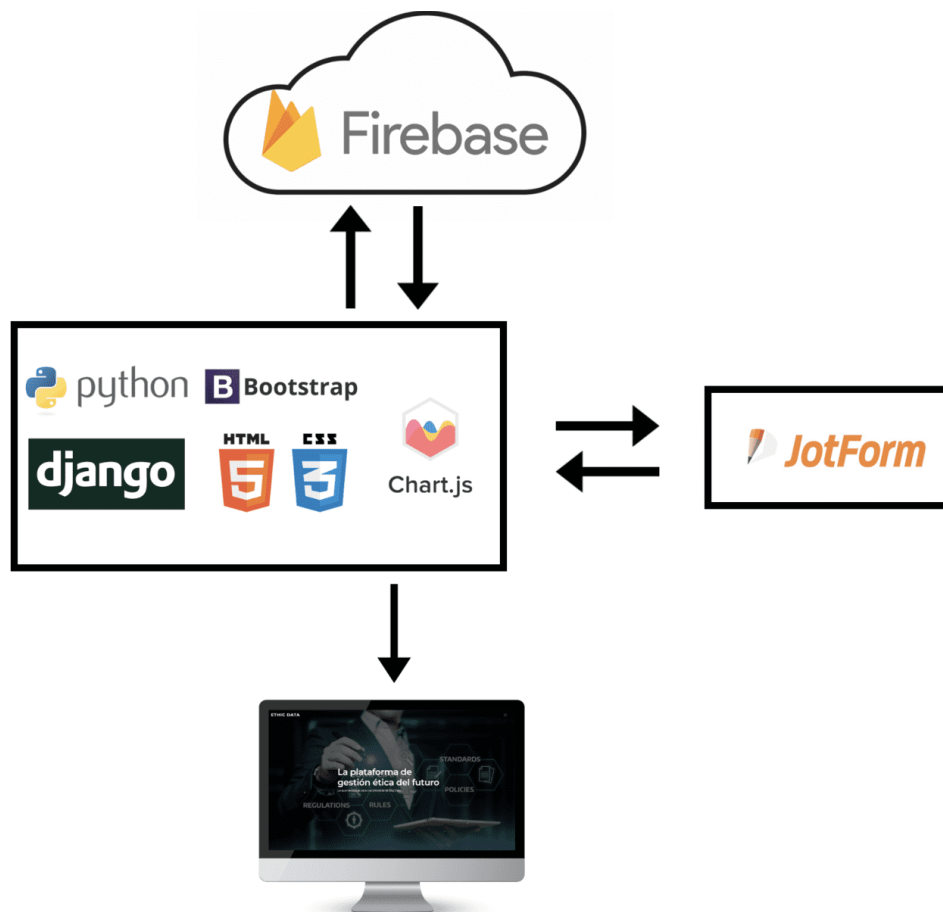


Ilustración 3 - Arquitectura del sistema. Fuente: Elaboración propia.

En cuanto al desarrollo de la parte del servidor o *back-end*, se empleará principalmente el *framework* de desarrollo web Django, basado en lenguaje Python. En esta sección es donde se encontrará el grosor del código y el núcleo funcional de la aplicación.

En referencia al almacenamiento, se dispondrá de una base de datos de tipo no relacional o NoSQL, concretamente la ofrecida por la entidad Google: Firebase. En este punto se distinguen dos componentes de almacenamiento. El primero consiste en una base de datos actualizada a tiempo real o *Realtime Database* proporcionada por Firebase donde se alojará la información de naturaleza dinámica tal como los datos de los usuarios y sus credenciales. El segundo se trata de un sistema de almacenamiento estático en la nube, *Cloud Storage* de Firebase, donde se ubicará la guía y los documentos necesarios. Ambas se encontrarán íntegramente conectadas al núcleo operacional de la aplicación gracias a la conexión Firebase.

Por último, en lo que respecta al *front-end* o lado del cliente, será desarrollado en HTML 5 y CSS 3 aplicando el marco de trabajo Bootstrap con objeto de perfeccionar el apartado estético y mejorar la interacción persona-computador. Adicionalmente, y en un segundo plano, se empleará la librería Chart.js para la visualización de gráficos, y se hará uso de la API de JotForm para

## 4.2 Diseño de la base de datos

El diseño de la base de datos para la aplicación se tratará de desarrollar siguiendo una metodología (Elmasri, 2007) que garantice que se tienen en cuenta todos los requisitos funcionales contemplados en el capítulo de análisis de requisitos. Dicho diseño se dividirá en las etapas de diseño conceptual, diseño lógico y diseño físico.

### Diseño conceptual

El diseño conceptual debe reflejar el flujo de la información de la aplicación a través de las distintas funcionalidades y cómo se relaciona con los usuarios, y es independiente a su posterior implementación. El diseño conceptual se hará siguiendo el modelo UML (Orallo, 2002), y representándolo mediante un diagrama de clases tal como se puede ver en la ilustración 4.

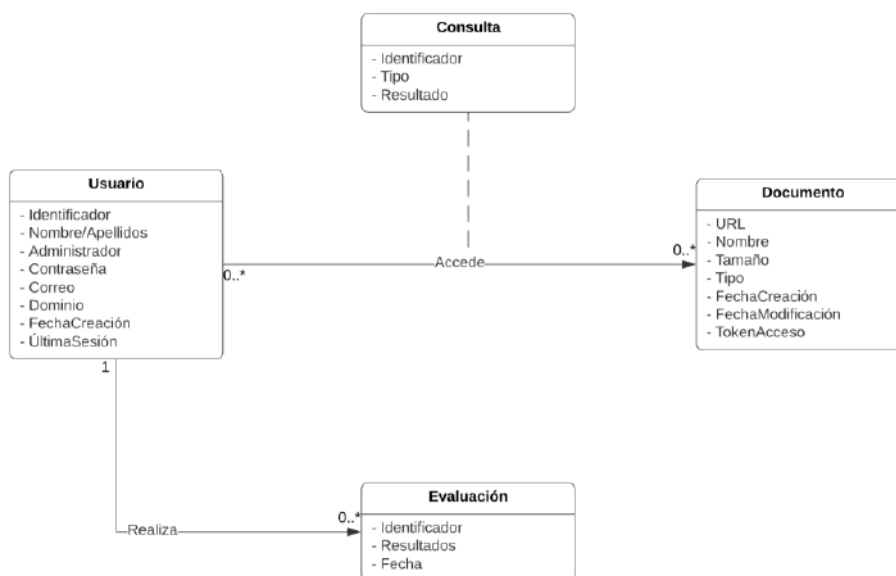


Ilustración 4 - Diseño conceptual. Fuente: Elaboración propia.



Los usuarios, como ya se había mencionado, se encontrarán en *Realtime Database*, siendo necesaria la información referente a su nombre completo y su correo electrónico. Asimismo, se requerirá de una contraseña para su acceso y se podrán distinguir usuarios de tipo estándar o administrador. El resto de los elementos pertenecientes al usuario (identificador y fechas de creación y última sesión) son generados, registrados y actualizados de manera automática por Firebase, siendo el identificador la clave primaria del usuario.

En cuanto a la clase documento, se refiere a la guía en cuestión, pero no como un único documento, ya que se añadirán también las subsecciones de la guía como documentos independientes, teniendo por un lado la guía al completo, y por otro, todas las partes de las que se compone. En cualquier caso, únicamente se deberá proporcionar el archivo y un nombre, y a partir de estos elementos el sistema de Firebase completa de forma automática el resto de los campos. La URL, dirección concreta donde se encuentra almacenado el archivo en Cloud Storage, puede ser modificada por un administrador si así lo desea, donde también es interesante comentar la existencia de un *token* o clave de acceso, la cual permitirá restringir el acceso o manipulación de los archivos. De nuevo, las fechas y tamaño de archivo son información inherente a este sistema, así como el tipo de archivo (formato del archivo, por ejemplo, PDF).

De la interacción entre estos dos agentes, usuario y documento, surge la clase consulta, pues cabe tener en cuenta que el objetivo de la aplicación no es otro que favorecer la experiencia del usuario en torno al uso de la guía. Su clave primaria constará de un identificador propio para poder diferenciarla de las distintas consultas que podrá realizar un mismo usuario. En general, el tipo de consultas estará predefinido siendo la mayoría de únicamente lectura de la guía, aunque se permitirá a los administradores la modificación de esta. De cualquier manera, el resultado de la consulta será almacenado por el sistema.

Por último, se encuentra la clase evaluación, donde se almacenarán los resultados obtenidos por un usuario en un formulario de autodiagnóstico o autoevaluación como una cadena de números enteros en formato de texto. Al igual que sucedía con la clase documento, únicamente tendremos que facilitar este campo y Firebase completará el resto: identificador, y fecha de creación, que permitirá estudiar la evolución de los diagnósticos en el tiempo.

## Diseño lógico

El diseño lógico se desarrolla a partir del diseño conceptual contemplando las estructuras de datos que se emplearán y evitando la posible redundancia de los datos. El diseño lógico resultante sería:

```
USUARIO (Identificador: char(20), Nombre_apellidos: char(50),
Administrador: integer(1), Contraseña: char(20), Correo:
char(50), Dominio: char(50), FechaCreación: date,
ÚltimaSesión:date )
CP: {Identificador}
VNN: {Nombre_apellidos, Administrador, Contraseña, Dominio,
```



FechaCreación, ÚltimaSesión}  
Único: {Correo}

DOCUMENTO (URL: char(150), Nombre: char(50), Tamaño: integer(20), Tipo: char(20), FechaCreación: date, FechaModificación: date, TokenAcceso: char(20))  
CP: {URL}  
VNN: {Nombre, Tamaño, Tipo, FechaCreación, FechaModificación}  
Único: {TokenAcceso}

CONSULTA (Identificador:char(20), Tipo: char(20), Resultado: char(100), Identificador\_usuario: char(20), URL: char(150))  
CP: {Identificador}  
VNN: {Tipo, Resultado, Identificador\_usuario, URL}  
Caj: {Identificador\_usuario} -> USUARIO (Identificador)  
Caj: {URL} -> DOCUMENTO (URL)

EVALUACIÓN (Identificador: char(20), Resultados: char(20), Fecha: date, Identificador\_usuario: char(20))  
CP: {Identificador}  
VNN: {Resultados, Fecha, Identificador\_usuario}  
Caj: {Identificador\_usuario} -> USUARIO (Identificador)

## Diseño físico

El diseño físico, al igual que sucedía con el diseño lógico, se realiza a partir de su predecesor, en este caso el diseño lógico. El diseño físico define la forma en que la base de datos es implementada físicamente, así como las estructuras de almacenamiento, a partir de dónde se obtendrán las tablas de la base de datos. En este proyecto se va a emplear una base de datos de tipo no relacional, en concreto Firebase Realtime Database de Google, que posee una estructura similar al formato JSON. Seguidamente se aportará un ejemplo de cada tabla en las ilustraciones 5, 6, 7 y 8:

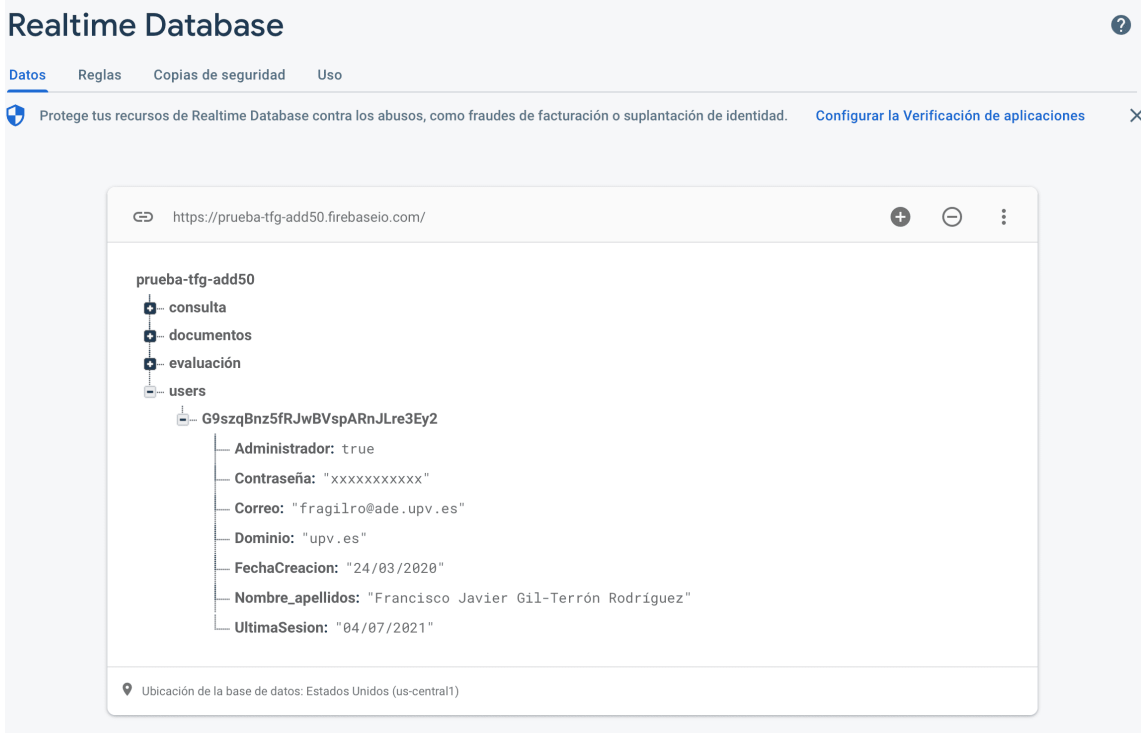


Ilustración 5 - Tabla usuarios. Fuente: Elaboración propia.

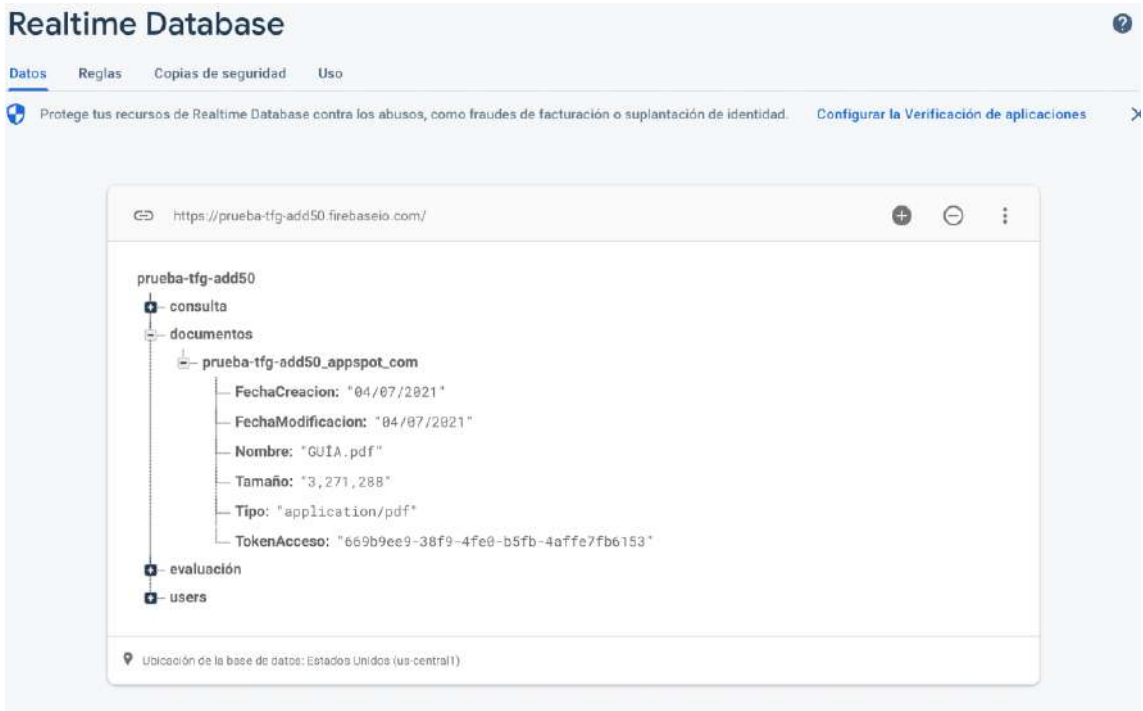


Ilustración 6 - Tabla documentos. Fuente: Elaboración propia.

# Creación de una guía para el uso desde una perspectiva ética del Big Data en las empresas

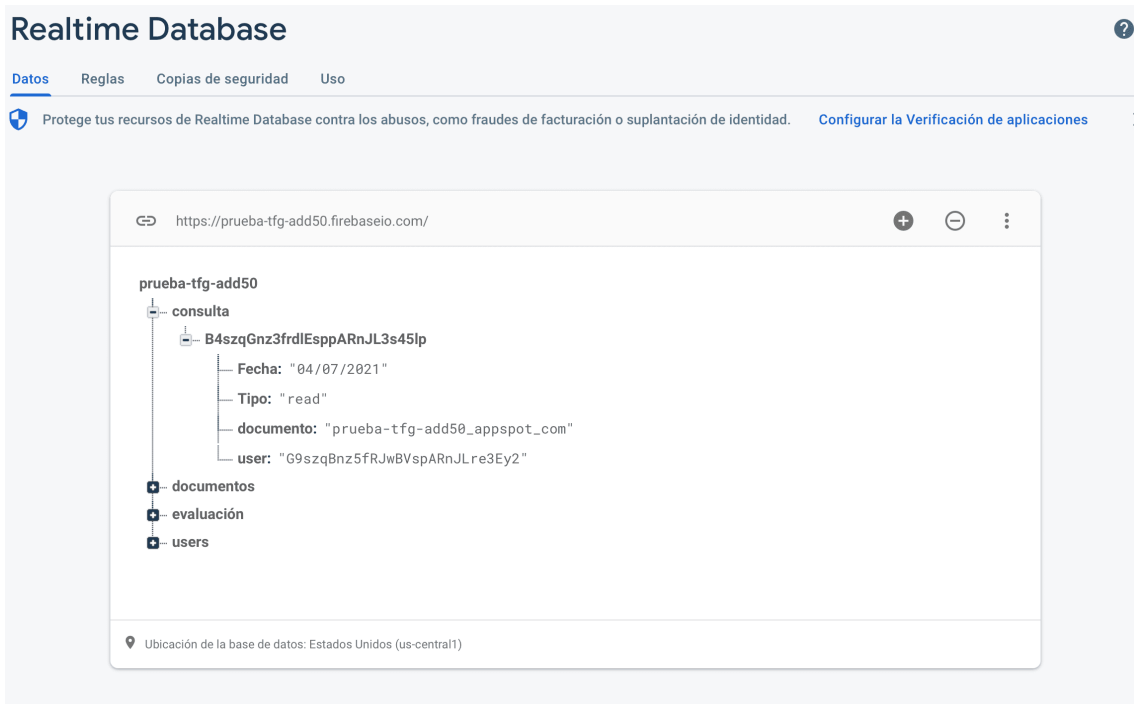


Ilustración 7 - Tabla consulta. Fuente: Elaboración propia.

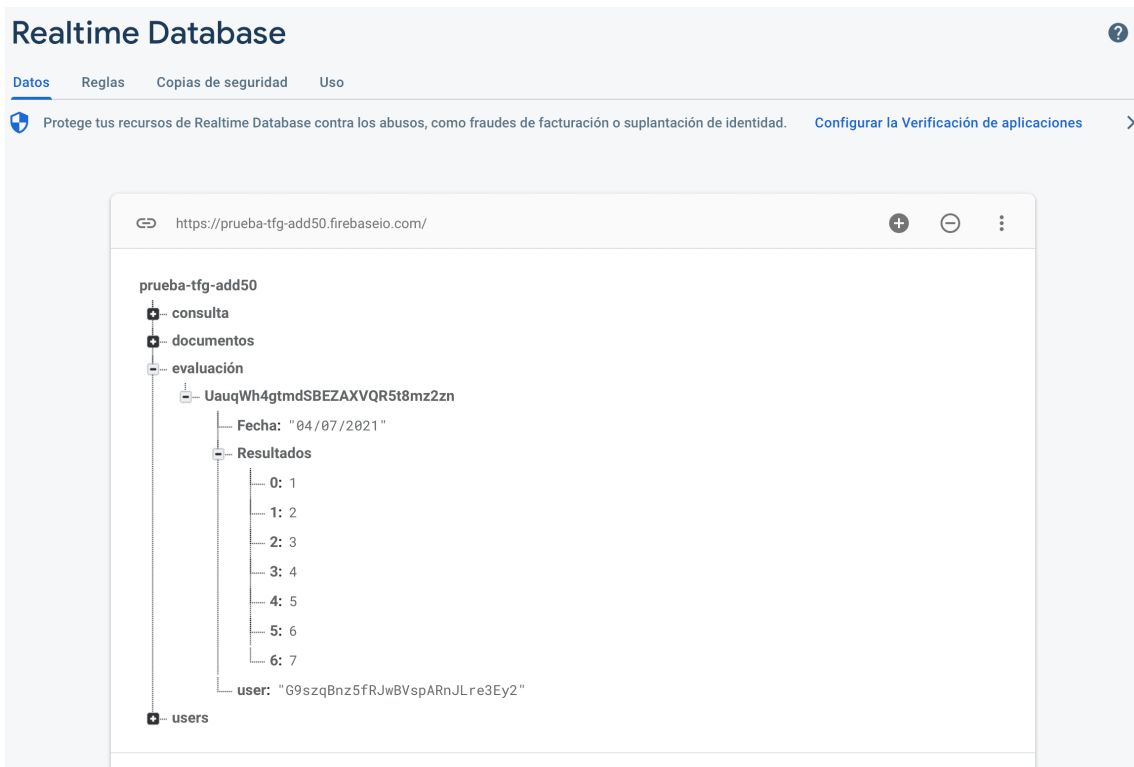


Ilustración 8 - Tabla evaluación. Fuente: Elaboración propia.

Por último, el almacenamiento de documentos, como se había comentado, se hará a través de *Firebase Cloud Storage*, en la ilustración 9 se ve un ejemplo de documento registrado por un administrador.



Ilustración 9 - Almacenamiento de documentos. Fuente: Elaboración propia.

## 4.3 Tecnología utilizada

---

Para el desarrollo de la solución se emplearán las siguientes tecnologías:

### 4.3.1 Firebase

---

Firebase es una plataforma desarrollada por Google destinada al desarrollo de aplicaciones web, aunque en el proyecto actual se utilizará únicamente para la gestión de usuarios mediante su sistema de autenticación y para el almacenamiento.

El coste de las autenticaciones es gratuito a menos que se opte por la autenticación vía telefónica, caso en que será gratuito para las 10.000 primeras de cada mes, y a partir de esa cifra cada autenticación tendrá un coste de 0.06 USD en España, aunque dados los requisitos actuales será más que suficiente el límite gratuito. Para el almacenamiento se hará uso de dos servicios de esta plataforma: *Realtime Database* y *Cloud Storage*.

El primero consiste en una base de datos NoSQL dinámica alojada en la nube que sincroniza a tiempo real todos los datos en formato JSON con el resto de los usuarios. Además, permite el uso de reglas personalizables para el control de acceso y establecimiento de permisos de los usuarios. La versión gratuita ofrece cien conexiones simultáneas, 1 GB almacenado y 10 GB descargados por mes, ambos ampliables por 5 USD/GB y 1 USD/GB respectivamente. Recordando que en *Realtime Database* será utilizado para gestionar la información de los usuarios, los márgenes sin coste también serán idóneos.

En cuanto a *Cloud Storage*, se trata de un sistema de almacenamiento estático basado en sencillas operaciones de carga y descarga pensado inicialmente para archivos multimedia, aunque acepta todo tipo de archivos. Una vez más no será necesario recurrir al método de pago uso, pues de base permite almacenar 5 GB, descargar 1 GB por día y realizar 20.000 y 50.000 operaciones de carga y descarga respectivamente y en este caso en particular únicamente iba a ser utilizado para almacenar la guía u otros documentos adicionales complementarios. De igual manera, se podrían superar esos límites a unos precios de 0.026 USD/GB almacenado, 0.12 USD/GB descargado, 0.05 USD/10.000 operaciones de carga y 0.004 USD/10.000 operaciones de carga.

### 4.3.2 PyCharm

---

PyCharm es un entorno de desarrollo específico para el uso de lenguaje Python, aunque también permite otros como HTML. Pertenece a la firma JetBrains, conocidos en la industria por su amplio repertorio de entornos de desarrollo adaptados a distintos lenguajes de programación, como PhpStorm para PHP o RubyMine para Ruby.

La plataforma cuenta con una versión estándar sin coste y una profesional, la cual añade soporte remoto para máquinas virtuales o contenedores, soporte directo a bases de datos SQL y otros lenguajes como JavaScript o ventajas adicionales en el uso de *frameworks* como Flask o Django; todo ello a un precio de 199 €. Para el presente trabajo se empleará únicamente, dentro de este entorno, Python-Django, HTML/CSS y JavaScript, por lo que basta con la versión estándar.

La totalidad del código se desarrollará dentro de este entorno, y las librerías requeridas, que se comentarán a continuación, se instalarán directamente dentro del mismo. Adicionalmente también será necesaria la incorporación de la librería Pyrebase para la conexión con Firebase, paquete de código abierto que se encuentran integrado en PyCharm y que no se detallará posteriormente ya que su única funcionalidad es esta última.

### 4.3.3 Python

---

En la década de los 90, nace y se impone el término minería de datos para el proceso de extracción de conocimiento a partir de grandes cantidades de datos, técnica que obtuvo un amplio desarrollo gracias a Big Data, dando lugar a lenguajes específicos para el manejo de datos, como R en 1995, o Python en 1991. Ambos se caracterizan principalmente por el gran número de librerías de las que disponen, pero, aunque estén adaptados para trabajar con grandes volúmenes de datos, cabe destacar que no lo están a la variedad y velocidad de estos mismos. Sin embargo, Python tiene librerías desarrolladas específicamente para el uso de Big Data, motivo que le hace destacar en este ámbito (García Ruíz & Sarasa Cabezuelo, 2018).

Se trata de un lenguaje de programación de código abierto, alto nivel, portable, interpretado (no debe compilarse antes de ejecutarse), multiparadigma, multiplataforma, y el tipo de datos es dinámico en tiempo de ejecución.

Pero lo que le ha permitido distinguirse y ganarse un hueco en el sector de los datos, siendo el que más ha crecido en los últimos años, es principalmente la enorme cantidad de librerías que tiene a su disposición el usuario relacionadas con el tratamiento de datos, como ya se ha comentado, sumado a que es de código libre mientras que por el contrario las licencias de herramientas disponibles en el mercado tienen en su gran mayoría un elevado coste. Pero no solo esto es lo que ha formado una amplia comunidad que respalda Python, pues además es adaptable a muchos otros tipos de aplicaciones y se basa en una sintaxis sencilla y fácil de comprender, lo que le dota de una accesibilidad de la que muchos lenguajes carecen.

En cambio, esta herramienta también posee unas carencias que lo hacen inadecuado para ciertas prácticas. Sin ir más lejos, se trata de un lenguaje que no tiene precisamente un amplio soporte de servidores debido a la complejidad de su hosting y resulta bastante ineficiente en materia de computación paralela. Incluso parte de la comunidad que lo emplea, critica la pobre documentación pública de la que se dispone.

Pese a esto, Python ha logrado imperar en el ámbito de Big Data gracias a su simplicidad y facilidad para el aprendizaje, lo que permite un alto rendimiento para profesionales poco experimentados, sumado a la gran variedad de posibilidades que ofrece para la visualización de datos y su compatibilidad con otras herramientas populares dentro del sector, como es el caso de Hadoop.

#### 4.3.4 Django

---

Django es un *framework* de desarrollo web de código abierto creado por The Django Software Foundation. Se trata de un marco de trabajo alto nivel y escrito en Python que persigue el objetivo de permitir al desarrollador centrarse en el núcleo de su aplicación, sin dejar de lado un entorno seguro, rápido y altamente escalable.

Gracias a la gran cantidad de servicios pre-construidos que dispone, como sistema de registro e inicio de sesión de usuarios o conexión a distintas bases de datos, lo convierten en un marco de trabajo muy versátil que resulta igual de válido tanto para pequeñas aplicaciones como para proyectos a gran escala. Algunos ejemplos de estos podrían ser Instagram o Pinterest.

Cuenta con un gran apoyo por parte de su comunidad, y una considerable cantidad de paquetes desarrollados por esta, que junto a la documentación tan completa de que dispone, hacen de este marco el ideal con el que iniciarse en el ámbito del desarrollo web, pero disponiendo de una profundidad más que suficiente para ser de igual utilidad a desarrolladores más experimentados.

Los proyectos de Django se basan en la creación de aplicaciones, módulos independientes entre ellos con distintas funcionalidades que pueden ser transversales a distintos proyectos, aunque en este caso únicamente será necesaria la creación de uno de estos módulos. Además, cuenta con varios mecanismos del campo de la ciberseguridad, como protección ante inyecciones SQL, pero este ámbito escapa al alcance del trabajo.



### 4.3.5 Bootstrap

---

Las distintas vistas de la aplicación web serán implementadas mediante HTML y diseñadas con CSS y la biblioteca Bootstrap. Bootstrap es de las librerías de código abierto para el diseño de páginas web más populares que existen en la actualidad. Pertenece a Twitter y tal es su extensión que su uso podrían considerarse casi un estándar en la industria. Trabaja con HTML, CSS y JavaScript y conforma un marco de desarrollo adaptable a cualquier dispositivo y casi a cualquier navegador web que sigue la filosofía “*mobile-first*”, es decir, que está pensada para desarrollar desde un inicio la vista que se tendría desde un dispositivo móvil hacia el resto de los dispositivos para hacerlo totalmente adaptable.

Esta librería tiene distintas opciones de inclusión, de forma total o parcial según las necesidades del proyecto, además de permitir su instalación de manera sencilla mediante una red de distribución de contenidos o CDN (opción escogida en este caso) o bien mediante la descarga de los archivos CSS y JavaScript compilados en el caso de desear modificarlos para una mayor personalización.

Al igual que Django, cuenta con una excelente documentación repleta de ejemplos y casos de uso más frecuentes que facilitan un rápido aprendizaje y una implementación sencilla con poco código.

### 4.3.6 JotForm

---

JotForm es un generador y configurador de formularios en línea orientado a acelerar procesos que requieran información del usuario, como gestión de pagos o encuestas dinámicas.

Tiene múltiples opciones de pago que van desde una versión gratuita hasta 79 dólares al mes. Las cotas respectivamente son de 5 formularios a ilimitados, de 100 a 100.000 envíos al mes, de 100 MB a 1 TB de almacenamiento y de 1.000 a ilimitadas vistas al mes de los formularios.

JotForm dispone integraciones directas con herramientas como PayPal, Dropbox o Zoom, y aunque no tenga con Django, se hará uso de su API para que los formularios aparezcan dentro de la aplicación web y poder almacenar los resultados en la base de datos.

### 4.3.7 Chart.js

---

Chart.js es una librería de código abierto que funciona sobre JavaScript para la creación de gráficas y visualización de información. Consta de ocho tipos de gráficas: de líneas, de barras, de radar, de tarta, de área polar, de burbujas, de dispersión y de área o superficie, todas ellas combinables entre sí.



En este caso, se empleará únicamente la gráfica de tipo radar para la visualización de los datos referentes a las autoevaluaciones o autodiagnósticos realizados por los usuarios.

Como sucedía con Bootstrap, también dispone de diferentes vías de instalación, y de nuevo, se empleará en este caso la opción en red distribución de contenidos por su simplicidad.

### 4.3.8 Docker

---

Esta tecnología se empleará para la implantación y puesta en producción de la aplicación, no para el desarrollo como tal, por lo que en su respectivo capítulo se tratará con más detalle.

Docker se basa en la distribución de aplicaciones y servicios en entornos estandarizados y aislados llamados contenedores, los cuales contendrán únicamente lo necesario para que funcione correctamente la aplicación, por lo que son ligeros y minimizan el riesgo de errores producidos a causa de dependencias. Gracias al uso de estos contenedores, se puede desplegar aplicaciones de forma rápida, consistente y segura, además de hacerlas más escalables.

Cada contenedor es una instancia de una imagen, plantilla que contiene las instrucciones para crear un contenedor específico. Existe una gran cantidad de estas imágenes en el repositorio público, desde imágenes con la información para construir un contenedor con un servidor de bases de datos hasta servidores proxy, siendo a su vez editables para crear un entorno personalizado en un caso dado.

## 5. Desarrollo de la solución propuesta

---

Como se había citado con anterioridad, la guía se encontrará adjuntada en los anexos, por lo que este capítulo se centrará en mostrar el resultado del desarrollo de la aplicación complementaria. Cabe citar que todas las imágenes empleadas en la aplicación son libres de derechos de autor, y que no se han empleado plantillas para el diseño de las páginas, a excepción del formulario de evaluación como se comentará en su respectiva sección, habiendo sido todas estas desarrolladas de primera mano con la ayuda de la documentación de Bootstrap.

### 5.1 Aplicación web

---

Lo primero con lo que se encontrará el usuario en la aplicación es con una página de presentación, tal como se muestra en la ilustración 10 y siguientes. Esta página corresponde a la página de inicio de la web, siendo la única a la que se puede tener acceso sin haber iniciado sesión previamente.

El objetivo de esta página es valorizar la guía desde la visión de las empresas, ya que al fin y al cabo es el público objetivo al que se trata de alcanzar. Por ello, ha sido diseñada desde un enfoque de producto con el objeto de hacerla más atractiva para las organizaciones.




Ilustración 10 - Página de presentación 1. Fuente: Elaboración propia.

En ella se vislumbran superficialmente los contenidos de la guía, reflejado en la ilustración 11, acompañados de una serie de indicadores y metas que conforman una llamada a la acción para acceder a la aplicación, tal como aparece en las ilustraciones 12 y 13.

**SEGURIDAD**

**Crea un entorno seguro para tu información y la de tus usuarios**


Cientos de empresas en España han sufrido brechas de seguridad que han alterado o incluso llegado a destruir la organización.



**COMPETITIVIDAD**

**Genera ventajas competitivas sostenibles gracias a la gobernanza de los datos**


Una conducta ética no tiene por qué mermar tus objetivos, al contrario, es una fuente de valor para tu empresa



**LIDERAZGO**

**Conviértete en una empresa ética líder y contribuye al cambio**

Puedes llegar a ser un modelo de referencia e incrementar la confianza en tu organización.



**ANTICIPACIÓN**

**Anticípate a las nuevas tecnologías y a los dilemas morales que conllevan**

Té entorno está cambiando, no esperes a que sea demasiado tarde para adaptarte a él.




Ilustración 11 - Página de presentación 2. Fuente: Elaboración propia.

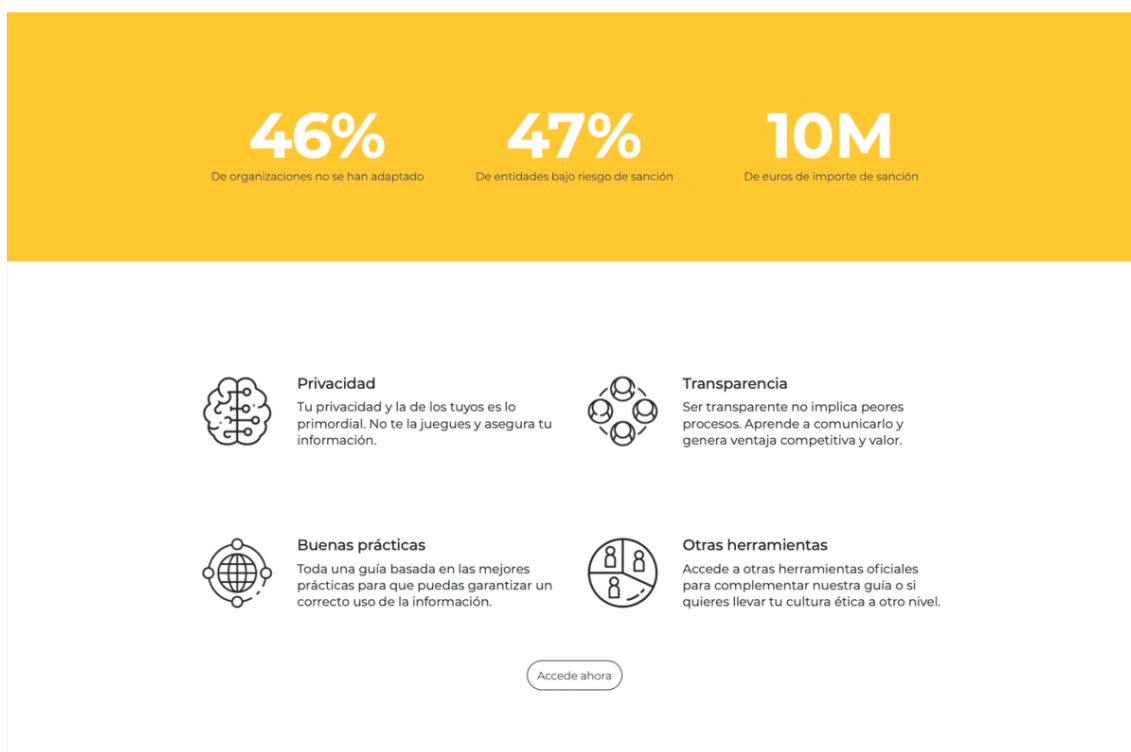


Ilustración 12 - Página de presentación 3. Fuente: Elaboración propia.



Ilustración 13 - Página de presentación 4. Fuente: Elaboración propia.

Interactuando con cualquiera de los botones de llamada a la acción presentes en esta página, se accederá al formulario de inicio de sesión para redireccionar a la aplicación principal en el caso de que las credenciales sean válidas. Dicha página se puede ver en la ilustración 14.

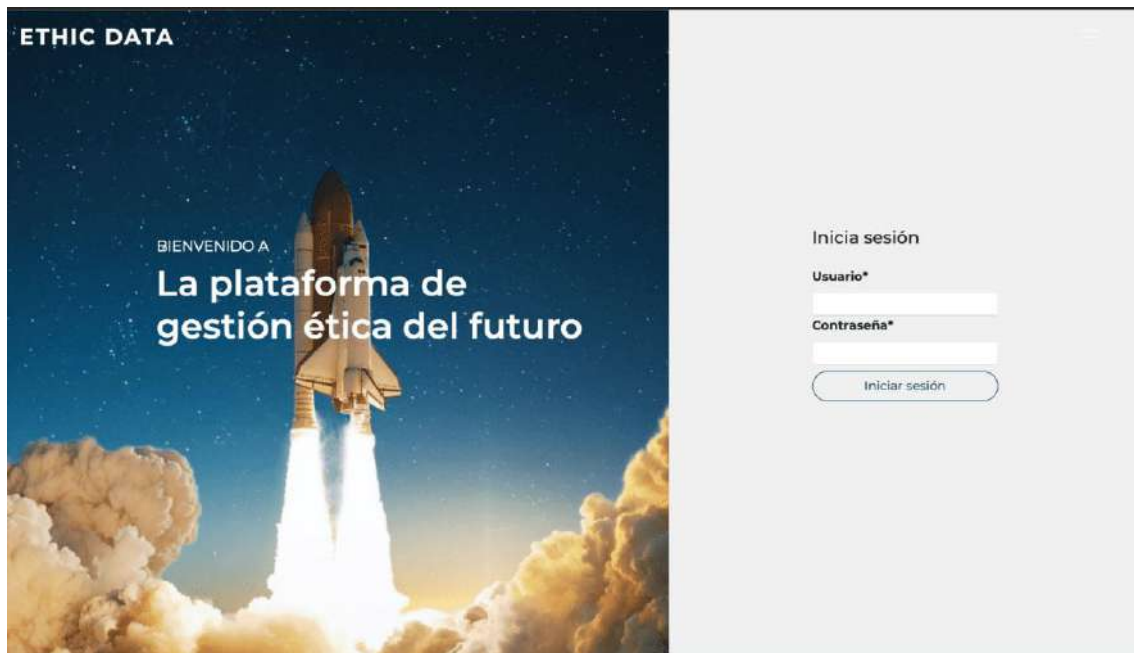


Ilustración 14 - Inicio de sesión. Fuente: Elaboración propia.

La gestión de las credenciales la realizará Firebase a través de Firebase Authentication. En la ilustración 15 aparece un ejemplo de usuario registrado correctamente.



Ilustración 15 - Inicio de sesión 2. Fuente: Elaboración principal.

## 5.2 Página principal

En cuanto a la aplicación en sí misma, se ha dividido en cuatro secciones:

La primera, correspondiente a las ilustraciones 16 y 17, y considerando el fragmento de presentación de la página como parte de esta, incluye un índice navegable que desplazará la página hasta la sección seleccionada, y enlaces de descarga de la guía completa.





Ilustración 16 - Aplicación 1. Fuente: Elaboración propia.



Ilustración 17 - Aplicación 2. Fuente: Elaboración propia.

A continuación, se le da la posibilidad al usuario de Descargar los distintos capítulos de la guía de manera individual, para aquellos que estuvieran interesados en una parte específica o quisieran ahondar en un campo determinado. Esta sección se visualiza en la ilustración 18.

### Por secciones



Ilustración 18 - Aplicación 3. Fuente: Elaboración propia.

La siguiente sección, corresponde al acceso al panel de autoevaluación o autodiagnóstico, el cual se comentará en profundidad en su respectiva subsección.

Por último, la cuarta parte de la página ofrece en un carrusel el acceso a otras guías. Pone en valor los contenidos de la guía relacionándola con guías oficiales, ya que todos los recursos que aparecen en esta sección han sido estudiados para desarrollar la guía, además de complementarla cubriendo algunas áreas de conocimiento cercanas que escapan al alcance de la guía. Estas dos últimas se muestran en la ilustración 19.

## Autodiagnóstico



EVALÚA TU POSICIÓN ACTUAL

**Responde a este cuestionario y recibe un informe sobre tu nivel de adecuación a un marco ético.**

Te indicamos aquellas áreas en las que necesitas mejorar y los recursos para hacerlo.

Podrás visualizar un diagnóstico en base a tus respuestas, y observar tu evolución con autodiagnósticos anteriores.

[Acceder](#)

## Otras guías



Guía sobre el uso de las cookies

Guía práctica para las evaluaciones de impacto...

Guía práctica de análisis de riesgos...

Ilustración 19 - Aplicación 4. Fuente: Elaboración propia.

## 5.3 Autodiagnóstico

El panel de autodiagnóstico o autoevaluación permite analizar la posición actual de la empresa de acuerdo con un marco ético a través de la evaluación, y estudiar el progreso en el tiempo a través del contraste entre evaluaciones.

Inicialmente, la primera vez que el usuario acceda a este panel, únicamente verá la opción de realizar una evaluación, tal como se ve en la ilustración 20. A continuación, se le indicará que rellene un cuestionario.



Ilustración 20 - Autodiagnóstico 1. Fuente: Elaboración propia.

El cuestionario propuesto para el diagnóstico se evaluará mediante una escala de Likert con valores del uno al seis, donde cabe destacar que no se está empleando el estándar de una escala de cinco para evitar la neutralidad a la hora de responder las preguntas. Así pues, se compondrá de las siguientes cuestiones:

- Los datos que se recogen persiguen fines determinados, explícitos y legítimos y se tratan únicamente de acuerdo con estos fines y no otros
- Los datos son exactos/completos y están actualizados, y se han implantado mecanismos para rectificar/completar rápidamente cuando esto no suceda, llegando a eliminarlos en caso de no ser esto posible
- No se mantiene la información por más tiempo del necesario para el tratamiento, suprimiendo los datos cuando se haya completado la finalidad, se haya retirado el consentimiento o se haya producido un uso ilegítimo
- Se tiene el consentimiento para cada una de las posibles finalidades que recibe la información, se ofrece información con carácter previo, y es demostrable la obtención de este
- El consentimiento se obtiene de forma explícita, clara, sencilla, inteligible y se permite su retirada con la misma facilidad en cualquier momento
- Es demostrable que los datos anonimizados no permiten identificar a los individuos, y en caso contrario, se cancelará la información
- Se facilita a los interesados la información relativa al tratamiento de manera clara, sencilla e inteligible y por medios accesibles, entre ellas, el interés legítimo, finalidad y destinatarios de transferencias en caso de haberlos, y el plazo de conservación



- Se facilita, de manera gratuita, el ejercicio de sus derechos al interesado, como el derecho al acceso, rectificación, limitación u oposición; además de informar acerca de la existencia de estos derechos
- La información de contacto del responsable del tratamiento, y del Delegado en caso de haberlo, es accesible y está a disposición de los usuarios
- En caso de producirse decisiones automatizadas, elaboración de perfiles o de realizar tratamientos con finalidad distinta a la inicialmente propuesta, se informa a los interesados de manera adecuada
- Se informa de la fuente de la que proceden los datos, en caso de no recogerse del propio interesado
- Se realizan tratamientos de información que se basen en la toma de decisiones enteramente automatizada sin supervisión humana
- Se toman decisiones automatizadas, incluyendo perfilado, en base a datos especiales, teniendo en consideración los riesgos que conlleva esta práctica
- Se toman medidas para proteger la confidencialidad de los datos, evitando su pérdida o daño, y se tiene el control y registro sobre los movimientos y actividades que se realizan sobre ellos
- Los procesos de tratamiento, así como los fines de este y etapas del ciclo de vida de los datos están claramente definidos, registrados y reflejan la realidad
- Se tiene en cuenta el ámbito y contexto de aplicación del tratamiento para evaluar los posibles riesgos para los derechos y libertades de los interesados, y se aplican y actualizan medidas para contrarrestarlos.
- El personal encargado del tratamiento tiene el nivel de formación adecuado para desempeñar correctamente sus responsabilidades de acuerdo con un marco ético de gestión de la información
- Se trata de extrapolar, de manera activa, el desarrollo de una cultura ética internamente a otras áreas más allá del departamento tecnológico

Finalmente, el formulario será visualizado de la siguiente manera indicada en la ilustración 21:

The screenshot shows a survey question with a Likert scale. The question text is: "Los datos que se recogen persiguen fines determinados, explícitos y legítimos y se tratan únicamente de acuerdo con estos fines y no otros\*". The scale has six options: "Muy en desacuerdo", "En desacuerdo", "Ligeramente desacuerdo", "Ligeramente de acuerdo", "De acuerdo", and "Muy de acuerdo". The first option "Muy en desacuerdo" is selected, indicated by a radio button and a grey bar below it. A "SEGUIR →" button is at the bottom right. A progress indicator "1/18" is on the left.

Ilustración 21 - Autodiagnóstico 2. Fuente: Elaboración propia.

Una vez completado el cuestionario se enviará al usuario a la página de resultados, cuya primera parte exhibe la ilustración 22. Aquí se hará un resumen de las áreas que están siendo evaluadas y se podrá visualizar mediante un gráfico de tipo radar de la situación actual de la empresa con respecto a un marco ético idóneo y contrastarla con los valores medios del sector en otras organizaciones que también traten con Big Data.

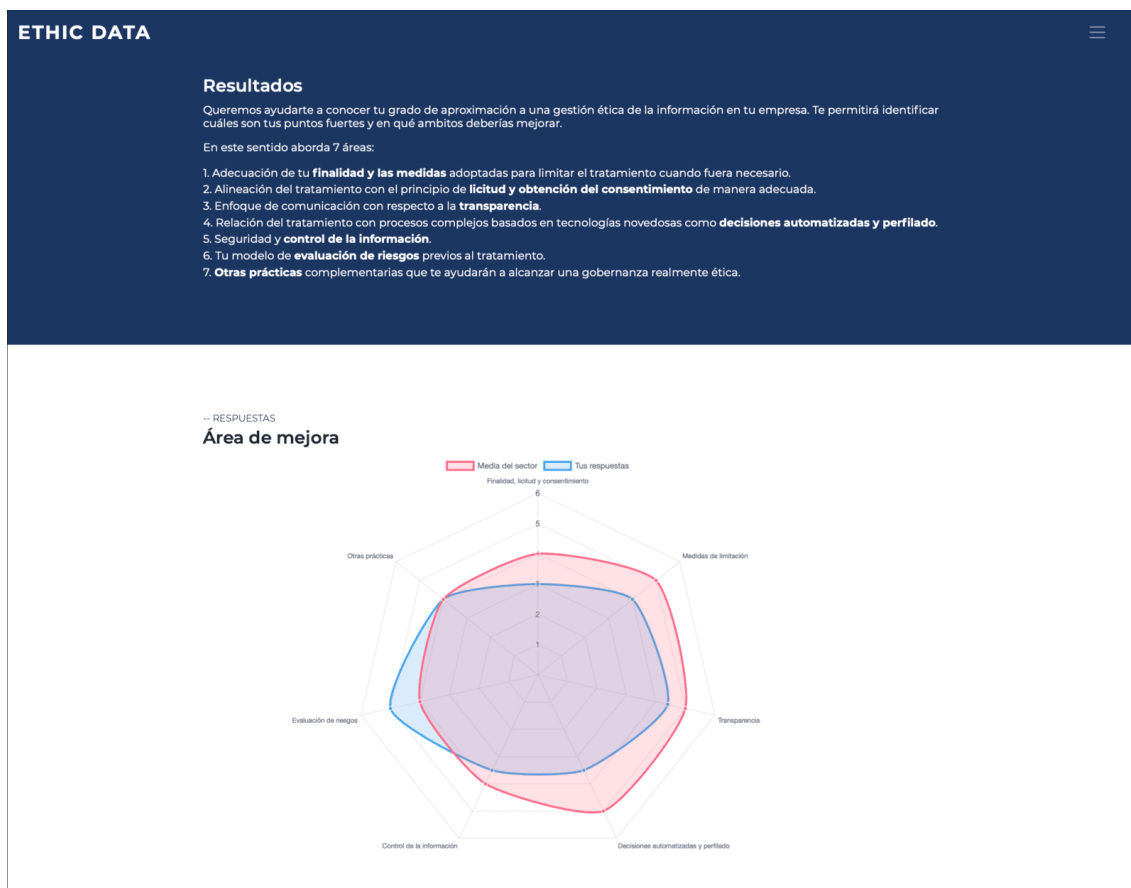


Ilustración 22 - Autodiagnóstico 3. Fuente: Elaboración propia.

Además, en función de las respuestas, se recomendará concretamente unos capítulos de la guía u otros (o la guía completa en su defecto) dependiendo de las carencias detectadas en la evaluación. Estas funcionalidades se pueden ver en las ilustraciones 23 y 24.

-- DIAGNÓSTICO

### Nivel de adecuación a un marco ético

#### Finalidad, licitud y consentimiento

La finalidad y licitud que sigue tu tratamiento, así como la forma en que obtienes el consentimiento **no están alineados con un enfoque ético**. Puesto que este es el primer punto que se plantea en cualquier tratamiento de datos, quizás **deberías replantear el uso de la información** que se da en tu empresa.

#### Medidas de limitación

Los mecanismos empleados para garantizar la privacidad de los interesados, tales como anonimización o minimización **son insuficientes**. Este es un punto clave para minimizar las posibles consecuencias negativas dadas por el uso de la información, por lo que **deberías destinar más recursos a este campo**.

#### Transparencia

El plan de comunicación que sigue tu entidad **no es suficientemente transparente**. Para poder generar confianza en los grupos de interés debes **enfocar de otra manera la manera en que comunicas**.

#### Decisiones automatizadas y perfilado

Tus procesos de toma de decisiones automatizadas y/o perfilado **carecen de las medidas necesarias**. Debido a la naturaleza de estos tratamientos deberías **incrementar los esfuerzos en asegurar** que se realizan de forma ética.

#### Control

El control que llevas sobre los datos en tu organización **no es el adecuado**. **Debes ser consciente de la relevancia** que tiene la vigilancia de la información y actuar en consecuencia.

#### Evaluación de riesgos

Tu grado de conocimiento de los posibles riesgos a los que te expones con tu tratamiento de datos es **suficiente**.

#### En conjunto

Parca que en general la gobernanza de los datos en tu empresa **dista de seguir un modelo ético**. Te recomendamos que estudies la propuesta que te ofrecemos en el siguiente apartado, dónde destacamos tus puntos débiles, y que analices nuestra guía al completo al margen de estos.

Ilustración 23 - Autodiagnóstico 4. Fuente: Elaboración propia.

-- NUESTRA RECOMENDACIÓN

### Creemos que esto es lo que necesitas



[Volver al panel](#)

Ilustración 24 - Autodiagnóstico 5. Fuente: Elaboración propia.

Tras haber realizado al menos una evaluación, el usuario tendrá total acceso a diagnósticos anteriores con el fin de estudiar la evolución en el tiempo y constatar los progresos realizados. En este caso, la ilustración 25 ejemplificará la situación posterior a la realización de una evaluación.



Ilustración 25 - Autodiagnóstico 6. Fuente: Elaboración propia.

## 6. Implantación

---

Para la implantación de la solución se ha montado la aplicación en una estructura de docker, que podrá construirse directamente en el servidor mediante conexión `docker-machine`, que permite conectarse a docker de un host remoto a través de la dirección IP y credenciales *Secure Shell* (ssh) del servidor.

El servidor que se ha utilizado se encuentra alojado en la plataforma DigitalOcean y ha sido proporcionado por la empresa Innsai, así como el dominio prestado para la aplicación web (el cuál no se citará por petición de la empresa). El servidor tiene un sistema operativo de base Linux, aunque resulta irrelevante para este caso ya que únicamente se necesitará docker instalado en el servidor.

Se han utilizado los siguientes contenedores de Docker, que se vincularán a través de `docker-compose`:

- **Aplicación:** En este contenedor se ejecutará la aplicación Django. Emplea una imagen personalizada, partiendo de la pública de Python a la que se le instalarán adicionalmente los paquetes requeridos para su funcionamiento, tales como Django o Pyrebase.
- **Proxy:** Este contenedor actuará de proxy o intermediario entre el cliente y la aplicación, ofreciendo los archivos estáticos al cliente tales como archivos JavaScript o multimedia. Se empleará la imagen pública de Nginx sin necesidad de modificarla.
- **Certificado:** Este contenedor se empleará para la emisión de certificados por la entidad Certbot para poder establecer una conexión segura al dominio a través del protocolo https. Para ello, una vez más, sólo será necesaria la imagen pública sin necesidad de editarla, en este caso la imagen será la de Certbot.

Ya que la base de datos que se emplea en la aplicación se encuentra en la nube (Firebase Realtime Database), no será necesario un contenedor para esta, pero en caso de utilizar una base de datos como MySQL o PostgreSQL sería necesario añadir un contenedor con su respectiva imagen.

Una vez conectado a docker del servidor, únicamente será necesario construir los docker mediante la directiva `docker-compose build`, seguido de `docker-compose up` para su puesta en marcha.

Al momento de lanzarse, ya será accesible desde cualquier navegador y rastreable por los motores de búsqueda, aunque cabe destacar que el posicionamiento SEO escapa al alcance del TFG.



## 7. Pruebas

---

En esta sección se mostrarán las pruebas de validación realizadas. Debido a la naturaleza del proyecto, resulta complicado evaluar si ha tenido un impacto real en la cultura organizacional, o si aquellos resultados obtenidos derivan de un cambio puntual. No obstante, sí que se han podido materializar en las empresas prácticas específicas derivadas de la guía.

Se ha tenido la oportunidad de aplicar la guía, mediante la aplicación web desarrollada, en las siguientes tres empresas:

### 7.1 Innsai

---

Innsai es una consultoría de innovación estratégica e inteligencia de mercado que trabaja con grandes cantidades de información en sus procesos de investigación de mercado, incluyendo datos sensibles. Esta fue la primera entidad donde se pudo aplicar por primera vez las recomendaciones recogidas en la guía.

Ya que las investigaciones que se llevan a cabo en Innsai se desarrollan durante largos periodos de tiempo, en ocasiones superiores al año, no era difícil que la finalidad inicial se viera modificada a lo largo del proceso, o que una vez se hubiera logrado parcialmente aparecieran finalidades adicionales a partir de la inicial. Este hecho da pie a que parte de la información obtenida no fuera necesaria en fases tardías de la investigación o que no llegara a emplearse nunca.

Con la ayuda de la guía y siguiendo el principio de minimización de la información, se consiguió alinear la cantidad de información recogida con los objetivos finales, prescindiendo de aquellos ya recogidos una vez cumplida su finalidad y limitando la recolección a aquellos datos que fueran realmente necesarios para conseguirlos.

### 7.2 Bankia Forward

---

Bankia Forward es una iniciativa que nace desde Innsai con la colaboración de Bankia cuya finalidad es ayudar a las empresas a entender esos cambios en los mercados, anticipando cómo serán los modelos de negocio del futuro y generando hoy, nuevas fuentes de ventaja competitiva para la empresa.

Este proyecto contaba con información de la gran entidad bancaria que le pone nombre, que, aun habiendo sido anonimizada, se comprobó que era posible la re-identificación de ciertos individuos a partir de sus datos. Tras la detección de este riesgo, se logró que se incrementara la robustez del proceso de anonimización con el fin de asegurar la privacidad de los ciudadanos cuya información se encontraba entre los ficheros disponibles.

## 7.3 Rpet Flake

---

Esta empresa representa uno de los más grandes centros de reciclaje de plásticos a nivel nacional y aspira a ser un referente dentro de este ámbito en el contexto europeo. Pese a su actividad principal, la firma trabaja con grandes cantidades de datos y realiza/recibe numerosas transferencias de información.

En este caso, la empresa pudo realizar pruebas en la aplicación complementaria, validando sus funcionalidades, sobre todo la herramienta de autodiagnóstico, donde puedo comprobar que se encontraba dentro de un rango aceptable en cuanto a gestión ética de la información, aunque cabe destacar que podría mejorar en el ámbito de la transparencia.



## 8. Conclusiones

---

Este trabajo ha ofrecido la posibilidad de integrar distintas disciplinas del terreno de la tecnología, tanto a nivel práctico con una aplicación web, como a nivel teórico con una guía para el uso ético de Big Data en las empresas. A su vez, se ha tenido la oportunidad de poder aplicarlo directamente en empresas, materializando resultados contrastables con los objetivos planteados, tal como se expuso en el capítulo de pruebas.

En esta línea, y en referencia a los objetivos inicialmente propuestos, gracias a las pruebas llevadas a cabo se pudo comprobar que se alcanzaron satisfactoriamente los objetivos que proponían:

- Ofrecer una herramienta con la que la organización sea capaz de visualizar su situación actual en cuanto a uso ético de datos se refiere.
- Concienciar a la empresa de los riesgos a los que es vulnerable (intensidad y probabilidad de la vulnerabilidad), así como el posible daño que pudiera causar directa o indirectamente a través de estos.
- Mostrar que la utilización y tratamiento de Big Data e información suponen un problema real en la actualidad para las empresas.
- Proporcionar a la empresa un elemento de valía, no una norma a la que atenerse con el fin de permanecer dentro de la Ley.

No obstante, por lo que respecta al resto de objetivos, aunque se manifestaran indicios a partir de los que se pudiera deducir que se habían alcanzado, estos no eran suficientemente consistentes como para afirmarlo. Esto es debido, como se citó con anterioridad, en parte a que son difícilmente evaluables, sumado a que quizás fueran demasiado ambiciosos. Dichos objetivos son los siguientes:

- Establecer una base ética que permita el transcurso y crecimiento normal de la organización, desarrollando o ayudando a mejorar el marco ético de la compañía.
- Incidir en la cultura organizativa de las empresas para la transmisión de valores a través de esta misma.
- Manifestar el valor generado por la ética para el conjunto de grupos de interés, no limitados a la propia organización, con intereses más que simplemente monetarios.

En cualquier caso, puede concluirse con que se ha alcanzado una mejora de la ética empresarial, en mayor o menor medida en función de la entidad, que era la base de la motivación del trabajo.

En cuanto a los problemas que se han encontrado, el más relevante está relacionado con el uso de Firebase, ya que Django no dispone de integración directa con esta base de datos, por lo que tanto la conexión como las peticiones se debían gestionar de una manera más compleja. Esto se podría haber evitado empleando una base de datos como MySQL o PostgreSQL, pero gracias a la utilización de Firebase se ha podido aprender sobre la tecnología no relacional de bases de datos.



Otro problema al que se le dedicó una considerable cantidad de tiempo era la forma en que se iban a representar las gráficas de los resultados de cada autoevaluación, pero se solucionó gracias a la documentación de chart.js que ofrecía suficiente información como para aventajarlo.

Por otro lado, aunque no fuera un problema como tal, cabe destacar que gran parte del esfuerzo empleado en la parte de la guía, y por tanto del trabajo, fue destinado a la revisión bibliográfica de recursos necesarios para desarrollarla.

Con todo esto, el desarrollo de este TFG me ha permitido aprender en profundidad sobre las normativas que rigen el sector de las tecnologías de la información, sin dejar de lado el apartado práctico, ya que también se ha aprendido a construir una aplicación web entera, lo que me permitió acceder a mi actual puesto de trabajo desarrollando en Django.

## 8.1 Relación del trabajo desarrollado con los estudios cursados

---

En continuidad con lo aprendido, los conocimientos adquiridos mediante el desarrollo de este proyecto, en su inmensa mayoría eran complementarios a los estudios cursados durante el grado, ampliando lo que ofrecían varias asignaturas de este.

El ejemplo más claro es el de la asignatura de deontología y profesionalismo, ya que la propuesta de una guía para el uso ético de la información presenta una fuerte vinculación con los contenidos de esta.

En lo que respecta al diseño de la página, recordando que no se han recurrido a plantillas y se han diseñado y desarrollado enteramente las distintas páginas, se han podido aplicar los conocimientos adquiridos en la asignatura de interfaces persona computador.

Durante el desarrollo de la base de datos, se ha ampliado el aprendizaje ofrecido por las asignaturas bases de datos y sistemas de información, y diseño y gestión de bases de datos, profundizando en la tecnología de tipo de no relacional complementando la relacional ya aprendida durante el grado.

A su vez, para el desarrollo de la aplicación web, se partía de la base obtenida en las asignaturas de Introducción a la informática y a la programación, programación y lenguajes, tecnologías y paradigmas de la programación, ampliándola con la inclusión del lenguaje Python; mientras que para el esquema de casos de uso se utilizó lo aprendido en ingeniería del software.

Finalmente, y en otro orden de ideas, también han podido desarrollarse competencias transversales como la responsabilidad ética y profesional, conocimiento de problemas contemporáneos y pensamiento crítico; además de que el TFG estuviera alineado con los objetivos de desarrollo sostenible propuestos por las Naciones Unidas, tal como producción y consumo responsables, paz, justicia e instituciones sólidas y trabajo decente y crecimiento económico.

## 9. Trabajos futuros

---

En cuanto a posibles ampliaciones o mejoras que no se han podido realizar debido al límite de tiempo, se planteó el desarrollo de un *chatbot* para la aplicación web con el que el usuario pudiera interactuar, por ejemplo, preguntándole directamente por alguna sección o contenidos de la guía. Otra alternativa que se propuso fue la de ofrecer la guía en formato móvil a través de la aplicación Telegram, pero tampoco se pudo desarrollar por el tiempo disponible.

En esta línea, a pesar de que el número de empresas donde se ha podido poner a prueba el resultado del TFG es bastante aceptable, y aunque no se trate como tal de un elemento inacabado del trabajo, quisiera finalizar la memoria destacando que me hubiera gustado realizar pruebas en más organizaciones, ya que el observar que realmente se estaban materializando resultados en las empresas resultaba cuanto menos motivador.

En cualquier caso, se espera que este trabajo sirva de referente para futuros TFGs y que motive a otros alumnos a profundizar en áreas de estudio más teóricas sin abandonar un desarrollo técnico de la solución, de manera que puedan demostrar la adquisición de distintas competencias durante el grado.

# Bibliografía

---

- Caballero Roldán, R., & Martín Martín, E. (2015). *Las bases de Big Data* (Primera). Catarata. [http://encore.fama.us.es/iii/encore/record/C\\_\\_Rb2691210\\_\\_Sinternet de las cosas\\_\\_P0%2C9\\_\\_Orightresult\\_\\_U\\_\\_X3?lang=spi&suite=cobalt](http://encore.fama.us.es/iii/encore/record/C__Rb2691210__Sinternet de las cosas__P0%2C9__Orightresult__U__X3?lang=spi&suite=cobalt)
- Colmenarejo Fernández, R. (2017). *Una ética para Big data: introducción a la gestión ética de datos masivos* (Primera). UOC.
- Diène, B., Rodrigues, J. J. P. C., Diallo, O., Ndoye, E. H. M., & Korotaev, V. V. (2020). Data management techniques for Internet of Things. *Mechanical Systems and Signal Processing*, 138. <https://doi.org/10.1016/j.ymssp.2019.106564>
- Docker Inc. (2020). *Docker Documentation* (<https://docs.docker.com/>). Docker Docs. <https://docs.docker.com/get-started/overview/%0Ahttps://docs.docker.com/> [Consulta: 20 de junio de 2021]
- Downie, N. (2015). *Chart.js/ Open source HTML5 Charts for your website*. Chart. Js. <https://www.chartjs.org/> [Consulta 10 de junio de 2021]
- Elmasri, R., & Navathe, S. B. (2007). *Fundamentos de sistemas de bases de datos*. Pearson Educación.
- Fosas, M. R. i. (2019). Ética Y Redes Sociales. *FMC Formacion Medica Continuada En Atencion Primaria*, 26(9), 498–502. <https://doi.org/10.1016/j.fmc.2019.05.011>
- F. J. García-Peñalvo, A. García-Holgado y A. Vázquez-Ingelmo (2019) "Requisitos," *Recursos docentes de la asignatura Ingeniería de Software I. Grado en Ingeniería Informática. Curso 2019-2020*, Eds., Salamanca, España: Grupo GRIAL, Universidad de Salamanca.
- García Ruíz, Y., & Sarasa Cabezuelo, A. (2018). *Big Data. Análisis de datos con Python* (Primera). Garceta.
- Google. *Documentación / Firebase*. from <https://firebase.google.com/docs?hl=es> [Consulta: 14 de junio de 2021]
- He, K., Hermanns, H., Wu, H., & Chen, Y. (2020). Connection models for the Internet-of-Things. *Frontiers of Computer Science*, 14(3). <https://doi.org/10.1007/s11704-018-7395-3>
- JotForm. *Constructor & Creador de Formularios en Línea Gratis | JotForm*. <https://www.jotform.com/es/> [Consulta: 14 de junio de 2021]
- Kapur, V. (2015). The data revolution. In *Hydrocarbon Processing* (First, Vols. 2015-Febru). Sage. <https://doi.org/10.29085/9781783304752.021>
- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*.



*Application Delivery Strategies*, 949(February 2001), 4.

Marz, N., & Warren, J. (2015). *Big Data: Principles and best practices of scalable realtime data systems* (First). Manning.

Orallo, E. H. (2002). El Lenguaje Unificado de Modelado (UML). Manuales Formativos ACTA, (26).

Otto, M. (2020). *Bootstrap · The most popular HTML, CSS, and JS framework in the world*. Bootstrap Website. <https://getbootstrap.com/>

Savater, F. (1991). *Ética para Amador* (Ariel (ed.)).

Sommerville, I. (2005). *Ingeniería del software Ingeniería del software Séptima edición*. 651. <https://ulagos.files.wordpress.com/2010/07/ian-sommerville-ingenieria-de-software-7-ed.pdf>

The Django Foundation (2012). *Django Documentation*. <https://docs.djangoproject.com/en/3.2/> [Consulta: 3 de junio de 2021]

Zarza, J. J.-L. M. G. (2017). *La Ingeniería Del Big Data: Como Trabajar Con Datos* (Primera). UOC. <https://www.casadellibro.com/libro-la-ingenieria-del-big-data-como-trabajar-con-datos/9788491800033/6028797>

# Anexos

---

## Anexo I. Código del proyecto.

---

A continuación, se indicará el enlace del código del proyecto publicado en el repositorio de GitHub.

<https://github.com/fragilro/TFG.git>

## Anexo II. Guía para el uso desde una perspectiva ética del Big Data.

---

A partir de la siguiente página, se adjuntará la guía para el uso del Big Data desde una perspectiva ética en las empresas desarrollada durante el trabajo de final de grado.



**Guía para el  
uso desde una  
perspectiva  
ética del  
Big Data**

Francisco Javier Gil-Terrón Rodríguez  
Universitat Politècnica de València



# Índice

Introducción	5
1. ¿A quién va dirigida la guía? – Alcance	7
2. Terminología y definiciones	9
3. Consideraciones generales	13
4. Buenas prácticas	17
4.1 Privacidad	18
Consentimiento	19
Finalidad y Licitud	20
Minimización y adecuación de la información	21
Anonimización	22
4.2 Transparencia	24
Qué comunicar	25
Cómo comunicar	26
Recomendaciones	27
4.3 Decisiones automatizadas y perfilado	28
4.4 Control	30
4.5 Transferencias de información	32
4.6 EIPDs	33
4.7 Otras buenas prácticas	35
5. Conclusiones	37
6. Referencias	39





# Introducción

En la actual revolución tecnológica, todo tipo de entidades, tanto públicas como privadas han hecho de la tecnología una pieza clave en sus modelos de negocio o diseños estructurales, hasta el punto de muchas de estas verse forzadas a usarla para conservar al menos un impacto mínimo en su sector. Esta realidad trae consigo una gran intensificación en el tráfico y uso de los datos, así como una progresiva automatización de los procesos de obtención, tratamiento, almacenamiento y eliminación de los datos, dando lugar al fenómeno del Big Data. Esto conlleva una serie de conflictos éticos y morales desde el inicio del ciclo de vida de la información, involucrando el consentimiento explícito del usuario, hasta su tratamiento.

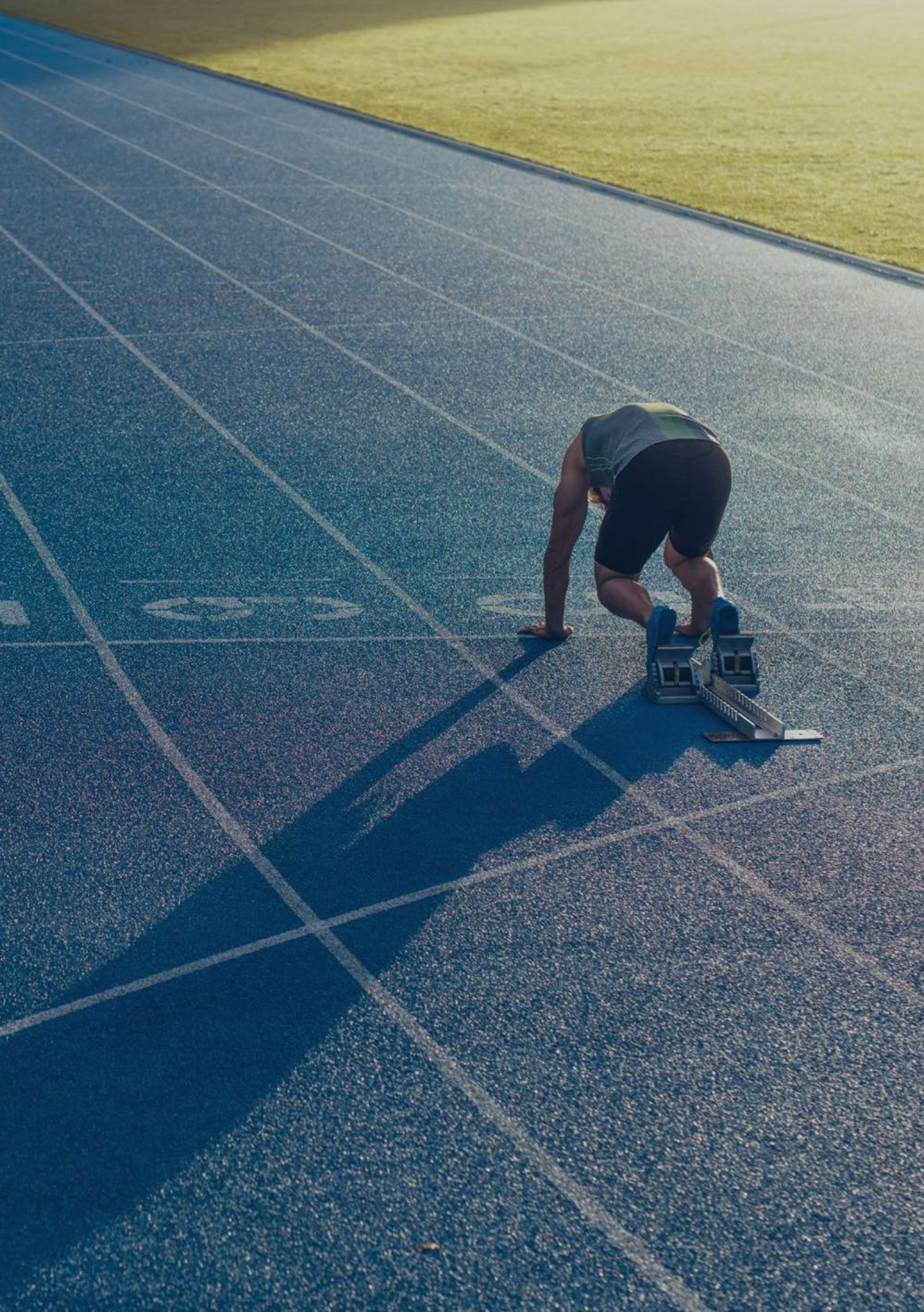
La utilización del Big Data ha servido para mejorar productos, prestación de servicios, eficiencia productiva, etc.; Pero, por otro lado, acarrea grandes consecuencias sobre la privacidad y seguridad de la información, hecho cada vez más presente en la sociedad y que preocupa no solo a nivel institucional, sino que la ciudadanía se hace consciente de ello paulatinamente. A su vez ha permitido beneficiarse a las organizaciones a través de, por ejemplo, la promoción de la transparencia mediante la publicación de sus cifras.

Ante este nuevo paradigma donde la privacidad y protección de la información tiene una mayúscula importancia, lo que en el pasado podría haberse considerado como iniciativas aisladas tras el interés propio ilustrado de una entidad, hoy en día se presentan como acciones que generan valor para todas las partes interesadas a nivel empresarial y social. A pesar de esto y teniendo en cuenta la creciente necesidad de actuar con transparencia e integridad en un mundo gobernado por los datos, no se ha establecido un estándar donde se integren privacidad, uso de la información y transparencia como ejes estratégicos en las empresas desde una perspectiva ética.

Pese a esta inexistencia de un estándar homologado a la hora de gestionar la información desde un enfoque ético, son numerosos los marcos de referencia que se han desarrollado con el fin de alcanzar las mejores prácticas en la materia. Estos marcos van desde enfoques generalistas hasta específicos sectoriales, pasando por el análisis de inteligencia artificial, algoritmos y cómo se relacionan con el Big Data, aunque bien es cierto que las inconsistencias entre estos mismos resultan demasiado significativas como para establecer un estándar global y en ocasiones la distancia entre cumplimiento normativo y uso ético puede llegar a ser considerable y

En esta guía se tratará de ofrecer un marco de referencia orientado hacia las mejores prácticas desde su diseño en cuanto a uso ético y responsable de los datos se refiere, y más específicamente, del Big Data, a la par que se advierte de los riesgos que afrontan las empresas al mantener políticas que no respeten un uso y tratamiento de datos adecuado. De esta manera, el objetivo final que se persigue es una utilización responsable de los datos por parte de las empresas para alcanzar un nivel de confianza fundamentado en una gestión adecuada de la información durante todo el ciclo de vida de esta, y el fomento de una cultura responsable comprometida con los derechos y libertades de la ciudadanía, sin renunciar a proporcionar mejores servicios, pues que un proyecto de Big Data sea ético no implica una pérdida de valor.







# 1. ¿A quién va dirigida la guía? – Alcance

El presente documento pretende servir de herramienta de soporte para el profesional o entidad que decide sobre la finalidad, contenido, uso del tratamiento de la información y, por tanto, de la existencia o no de esta, sobre quienes recaerá la responsabilidad última de dichos procesos.

De esta manera, la guía contiene recomendaciones y orientaciones acerca de mejores prácticas en materia de privacidad, transparencia y gestión responsable de la información. Cabe destacar que la guía únicamente se abordará desde una perspectiva más bien generalista para el uso ético del Big Data desde cualquier enfoque, sin llegar a especificar casos basados en sectores determinados (como podrían ser la automoción o la salud) ni resultar demasiado abstracto como para tener aplicabilidad real. En caso de tratar de alcanzar un propósito más específico, resultaría necesaria la utilización complementaria de otros recursos o herramientas externas.

Por otra parte, dadas las múltiples complejidades que plantea el uso de los datos, las recomendaciones aquí recogidas no pretenden ofrecer una solución uniforme para el cumplimiento de la ley, sino que deben servir de guía para que las entidades afectadas reflexionen y adopten decisiones sobre la solución más adecuada a sus intereses y modelo de negocio, sin llegar a ser necesario un estudio individual pormenorizado de la entidad, más allá de los tipos de tratamientos que se realizan, los sistemas informáticos o los sistemas de gestión documental para obtener un resultado correcto.

7

En definitiva, se pretende concienciar a las organizaciones que se enfrentan, no solo a un marco legal cambiante, sino a uno ético debido, en primer lugar, a la implicación directa de sus acciones a través del uso y tratamiento de datos de los ciudadanos, y en segundo, a que la extensión y profundidad de los valores éticos de las personas depende en gran medida del entorno cultural en el que se desarrollan, percibiendo valores como la privacidad y la intimidad de los individuos de formas distintas pese a que se hayan ido acordando internacionalmente unos principios aceptados de forma global.

Así pues, el actual procedimiento va dirigido, en primer lugar, a responsables y encargados de tratamiento de la información y a Delegados de Protección de Datos (en adelante aludiremos a ésta figura pro su acrónimo extendido: DPD), quienes deben ser conscientes de que el empleo de soluciones basadas en Big Data les sitúa en un escenario en el que el tratamiento, por sus características y, en particular, por la tecnología en cuestión, puede conllevar un alto nivel de riesgo. Con el fin de aplicar correctamente las orientaciones presentadas y adoptar medidas en el tiempo que permitan encontrarse en condiciones, no solo de cumplir la ley, sino de permanecer dentro de un marco ético óptimo, deberán estar claramente identificados los roles y responsabilidades de los implicados.



## 2. Terminología y definiciones

### ./ Dato

Consiste en la información obtenida del individuo, y será personal en el caso de pertenecer a personas físicas identificadas o identificables, es decir, que se pueda determinar la identidad del ciudadano a partir de esta. No obstante, La definición de datos personales no solo involucra información directa del usuario, sino que comprende cualquier elemento con el que se pudiera inferir la identidad de un individuo determinado (añadir todos los campos sensibles posibles), incluyendo identificadores de dispositivos electrónicos aún sin estar vinculados a través de un nombre.

9

### ./ Datos especiales

Aquellos datos que identifiquen de algún modo el origen étnico o racial, las opiniones políticas, las convicciones religiosas o filosóficas, y la afiliación a un sindicato. Es información especial los datos biométricos dirigidos a identificar de manera unívoca a una persona y los datos relativos a la salud o a la vida sexual y/o la orientación sexual de una persona. Se ha establecido la prohibición de registrar este tipo de datos salvo los de carácter biométrico en ciertas excepciones.

## ./ Ciclo de vida de los datos

Aunque existan distintas versiones para definir el ciclo de vida de la información, el estándar es dividirlo en las siguientes etapas:

**1. Obtención:** Proceso donde se recogen los datos, para su almacenamiento y posterior uso, de manera automática o manual a través de la compra, cesión, transferencia, extracción desde otros sistemas o generando de primera mano esa información. Algunos de los métodos de obtención más frecuentes pueden ser formularios, muestreos, encuestas o mediante tecnologías punteras como sensores.

**2. Almacenamiento:** Mantenimiento de la información para su posterior acceso. Suele estar directamente relacionado con la categorización de la información para poder clasificarla. Debe llevarse a cabo con la protección necesaria tanto a nivel físico de los equipos como para evitar posibles fallos humanos o accesos no permitidos.

**3. Tratamiento:** Conjunto de operaciones de explotación con el fin de obtener información adicional útil para la toma de decisiones, de manera automatizada o manual, realizados sobre los datos personales. Incluye planteamientos de hipótesis y extracción de conclusiones, así como la visualización de los propios datos.

**4. Transferencia:** Traspaso o comunicación de la información a un tercero (en cualquiera de sus formas, tanto a nivel interno como externo) recogiendo la conexión, consulta, publicación, comunicación u otras formas de acceso a los datos.

**5. Eliminación:** Proceso de destrucción de la información almacenada con la finalidad de que no pueda ser recuperada de nuevo de los sistemas donde estuviera soportada.

10

## ./ Tratamiento

Aunque el tratamiento haya sido considerado como una parte del ciclo de vida de los datos, cabe profundizar un poco más en este término debido a su relevancia en la gobernanza de la información.

De esta manera, se considera tratamiento a cualquier operación o conjunto de operaciones realizadas sobre datos personales o conjuntos de datos personales, con el objetivo de alcanzar una finalidad determinada, ya sea por procedimientos automatizados o no, como la recogida, registro, organización, estructuración, conservación, adaptación o modificación, extracción, consulta, utilización, comunicación por transmisión, difusión o cualquier otra forma de habilitación de acceso, cotejo o interconexión, limitación, supresión o destrucción.

Además, Se considerará que existe tratamiento de datos personales cuando el usuario esté identificado por un nombre o dirección de email que lo identifique o cuando se utilicen identificadores únicos que permitan distinguir unos usuarios de otros y realizar un seguimiento individualizado de los mismos.

## ./ Big Data

Partiendo del concepto de dato anteriormente expuesto, el Big Data o macrodatos componen los conjuntos de datos o de estructuras de información que, debido a su volumen, velocidad y forma resulta complicada su gestión, en cualquier parte de su ciclo de vida, con tecnologías tradicionales, como bases de datos relacionales. En otras palabras, cuando, a causa de emplear mecanismos clásicos de gestión de la información, se imposibilita alguna etapa del ciclo de vida de los datos a causa de la cantidad o variedad de estos, se trata de un problema de macrodatos.

Los términos macrodato y Big Data serán usados en adelante indistintamente, siendo el primero el preferido por instituciones públicas, y el segundo el más frecuente en cuanto a extensión de uso entre profesionales del sector, aplicando uno u otro en función de la fuente de apoyo.

## ./ Responsable

Persona física o jurídica, autoridad, servicio u otro organismo que, solo o junto con otros, determine los fines y medios del tratamiento, ya sea en forma de profesionales que contribuyen dentro de sus Organizaciones o bien como Encargados.

## ./ Encargado

Persona física o jurídica, autoridad, servicio u otro organismo que trate datos personales por cuenta del responsable del tratamiento, así como quienes desempeñan las tareas de asesorar a los Responsables de Tratamientos.

## ./ Delegado de Protección de Datos

Persona física o jurídica, empleado en plantilla o mediante contrato de servicio, que informa y asesora al Responsable, al Encargado y a otros empleados sobre las obligaciones del RGPD y supervisa su cumplimiento, cooperando y actuando como punto de contacto con las Autoridades de Control.





### 3. Consideraciones generales

Aunque la visión ofrecida en este documento no sea desde el punto de la norma sino del de la ética, aparecen una serie de elementos mínimos comunes a ambas, que se cumplirán de manera de implícita reflejando su estrecha vinculación, y conforman los fundamentos de los marcos legales y éticos que se conocen actualmente en torno a la protección de datos.

Estos principios básicos que se recogen en la mayoría de los documentos de este ámbito y tienen una perspectiva global, son formulados por entidades como la OCDE (Organización para el Desarrollo y la Cooperación Económica) o la Asamblea General de las Naciones Unidas. Algunos de estos pueden ser los principios de Licitud, Seguridad o Transparencia.

Posteriormente a estas directrices, el RGPD (Reglamento General de Protección de Datos) surgió como respuesta al creciente uso intensivo de datos, aportando nuevas herramientas para facilitar las garantías de privacidad, y alertando sobre los nuevos riesgos consecuentes, convirtiéndose, desde su entrada en vigor en 2018, en el primer marco que exige cierta proactividad por parte de las empresas.

Pero, ya que la esfera de lo moral es más amplia que la de la norma, esta última puede llegar a mostrar una extraordinaria flexibilidad a la hora de garantizar su cumplimiento, dando lugar a un margen en el que se puede actuar siguiendo, o contrariando, un modelo que se considere ético. De esta manera, no es suficiente que las entidades se ciñan a lo estipulado en el RGPD, pues deja demasiadas incógnitas y cuenta con un margen de maniobra demasiado amplio, sobre todo si se trabaja con tecnologías punteras a las que la Ley es incapaz de seguir en velocidad de desarrollo. En otras palabras, se parte del compromiso en el cumplimiento de la normativa, bajo la premisa de que este no es suficiente para alcanzar una gestión ética de los datos.

Por otro lado, ante la constante evolución tecnológica y los procesos de transformación digital que sufren las actividades de tratamiento de datos personales, el modelo tradicional para afrontar las medidas que garantizan la protección de los datos personales ha virado hacia un nuevo modelo más dinámico, **una nueva visión**, donde el foco no se centra en las amenazas que se ciernen sobre la compañía, sino en las amenazas sobre los derechos y libertades de los interesados y en la gestión continua de

los riesgos potenciales asociados al tratamiento. En definitiva, un enfoque ético en la gestión de la información, donde la ética aparece como un generador de valor para los diferentes grupos de interés.

Con esto, nace el concepto de Responsabilidad Proactiva, el cual está presente en aquellas prácticas que la Ley no establece de forma específica o la forma de implementarlas, siendo estas de cumplimiento voluntario bajo la iniciativa de la organización. La gestión del riesgo para los derechos y libertades, por ejemplo, es una actividad continua y forma parte del concepto de Responsabilidad Proactiva. Gran parte de las orientaciones aquí ofrecidas giran en torno a esta idea, pues es la proactividad lo que diferenciará una cultura de la información realmente ética de una que únicamente busca ceñirse a lo establecido en la legislación. Para poder garantizar la responsabilidad proactiva, es fundamental definir correctamente las actividades de tratamiento desde su diseño.

14

Al mismo tiempo, debe tenerse presente que la utilización de tecnologías nuevas, entre ellas el Big Data, suele conllevar de manera generalizada un mayor nivel de riesgo para los derechos y libertades de las personas físicas al margen de su naturaleza, alcance, contexto o finalidad. En este caso específico, el de los macrodatos, tiene implicaciones importantes en relación con su privacidad y, por tanto, en la protección de datos, pues son varios los estudios que han comprobado que vulneraciones en la privacidad de los ciudadanos pueden acarrear grandes consecuencias en la reputación personal e imagen pública, afectando, entre otros, a su libertad de expresión e incluso presunción de inocencia. Ciudadanos, políticos y organizaciones de consumidores, han puesto de manifiesto la necesidad de que se garantice la efectiva tutela de la privacidad con el fin de obtener la confianza de los usuarios.

De esta manera, se persigue que las diferentes entidades adopten fórmulas de información y consentimiento que reúnan los requisitos de transparencia y control, manifestando así el interés de la propia industria por implantar un sistema transparente en este ámbito, a la par que reflexionan sobre la licitud de sus procesos e implicaciones que los tratamientos de datos que llevan a cabo tienen sobre los interesados. Estas implicaciones, a su vez, exigen una valoración del mejor modo de incorporar las garantías contempladas en la normativa, modulándolas para adaptarlas a un modelo que resulte lo más ético posible en función de las características específicas de la compañía.

Pese a que, como se ha venido mencionando, esta guía únicamente recoge recomendaciones, la omisión total o parcial de estas puede no sólo comprometer seriamente la confianza de los grupos de interés en las empresas, sino que puede desembocar en serias crisis organizativas incluso en aquellos casos que estuvieran estrictamente ceñidas a la Ley, hecho que se intensifica si se tratan de datos personales y/o sensibles los que maneja la entidad en cuestión.



## 4. Buenas prácticas

Aplicar medidas apropiadas a fin de garantizar y poder demostrar que los procesos que involucran tratamientos de información, incluyendo Big Data, llevados a cabo en una organización se encuentran dentro de un marco ético adecuado y garantizan la protección de la información del interesado exige una actitud consciente, diligente y proactiva por parte de esta.

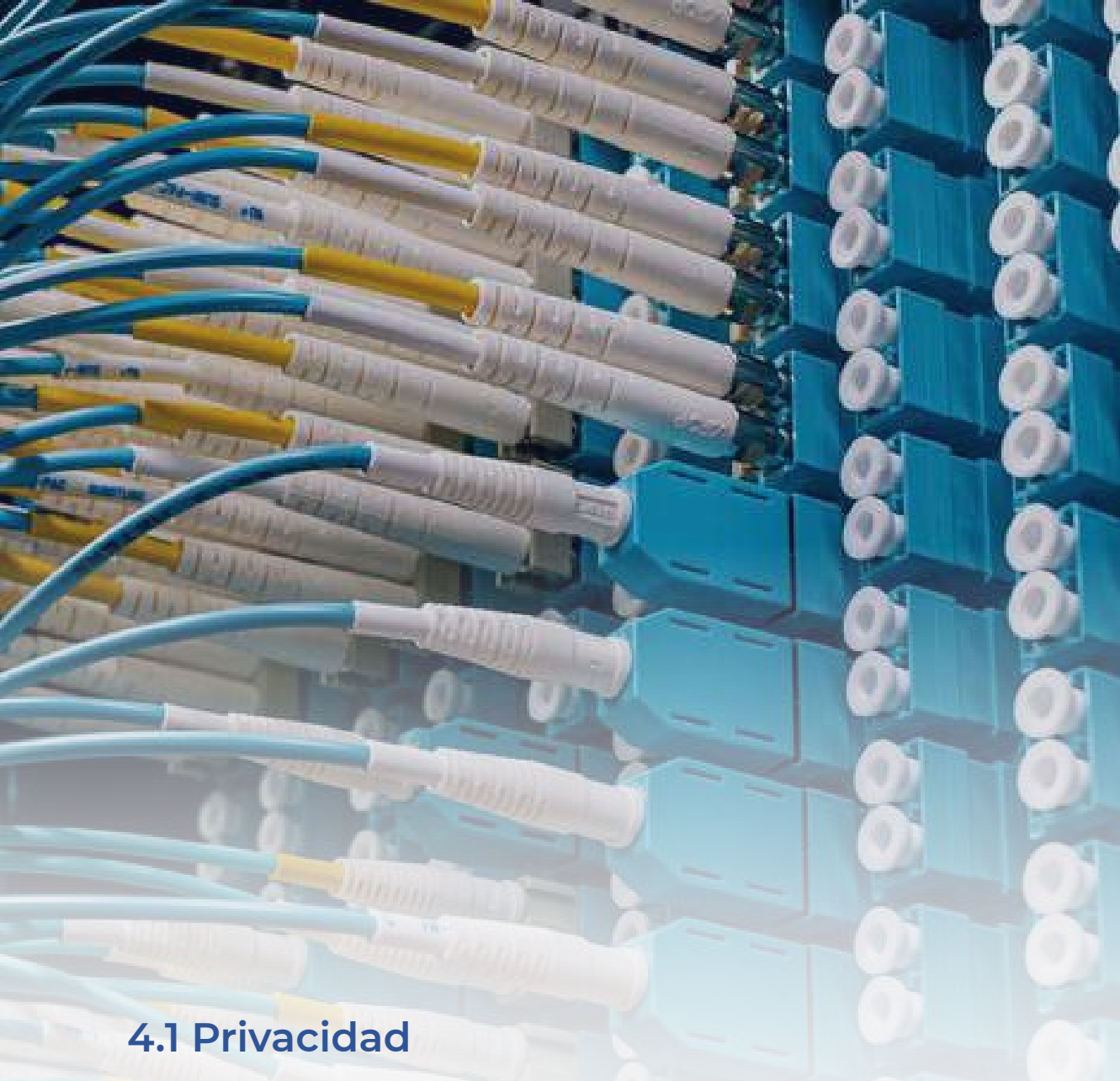
Esto se traduce a un abanico de requerimientos asociados a la protección de la información con el foco central en la integridad, disponibilidad y confidencialidad de los datos y requiere que las entidades analicen qué datos tratan, con qué finalidades lo hacen y qué tipo de operaciones de tratamiento llevan a cabo y las medidas dirigidas a garantizar su cumplimiento deben tener en cuenta la naturaleza, el ámbito, el contexto y los fines del tratamiento, así como el riesgo que conlleva para los derechos y libertades de las personas físicas.

Algunos principios pueden resultar obvios para los profesionales con sólo mencionarlos, como puede ser que los datos deben ser tratados de forma lícita, leal y transparente. No obstante, la forma de conseguirlo o de garantizar estos principios puede no ser tan obvia.

Por este motivo, la formación del capital humano responsable de la gestión de los datos aparece como un punto clave para garantizar un correcto uso de los datos y evitar que las empresas queden expuestas a los potenciales riesgos a los que se enfrentan, pues es el responsable quien debe aplicar las medidas oportunas y eficaces y ha de poder demostrar la conformidad de las actividades de tratamiento y la eficacia de estas.

En esta línea, las recomendaciones que se presentan a continuación tratan de mostrar medidas específicas que, aun garantizando su seguridad, deben ser evaluadas por los responsables con el fin de detectar las áreas más susceptibles de mejora y aplicando, según lo considerado por estos, aquellas que resulten pertinentes con el fin de identificar las amenazas y riesgos a los que está expuesta la actividad, evaluarlos, gestionarlos y minimizarlos.





## 4.1 Privacidad

La información es un recurso fundamental en la sociedad actual para la toma de decisiones en todos los aspectos y sectores de esta, donde el tratamiento y la explotación de grandes volúmenes de información puede ofrecer múltiples beneficios a la sociedad siempre que se mantenga el respeto por los derechos de las personas, su privacidad y la protección de sus datos personales.

A pesar de que la privacidad constituye un derecho de la ciudadanía en torno al cual se han desarrollado normas jurídicas, sin ir más lejos la RGPD (Reglamento General de Protección de datos), sigue presentando un margen de maniobra moral donde las empresas pueden optar por ceñirse a la norma o ir más allá y adoptar una política que persiga un uso realmente ético de la información.

La privacidad debe posicionarse como eje fundamental del diseño y planificación de proyectos que involucren Big Data y debe tenerse en cuenta en todas las fases del ciclo de vida de la información.

## ./ Consentimiento

Bajo el principio de privacidad, el consentimiento debe prevalecer en todos los procesos llevados a cabo que involucren los datos del interesado y es condición indispensable para obtener y tratar los datos dada su relevancia legal y moral. Las mejores prácticas relacionadas con el consentimiento empiezan antes del tratamiento de los datos: en su obtención y, por tanto, se debe tener muy presente en dicho proceso.

El consentimiento se toma como base para la gestión de datos personales del individuo para la inmensa mayoría de situaciones y se presenta como eje fundamental en procesos de gran impacto tecnológico como aplicaciones de modelos predictivos, permitiendo mantener la confianza de los grupos de interés.

Para que dicho consentimiento sea válido será necesario que haya sido otorgado desde la neutralidad de forma libre e informada y siempre de manera inequívoca, es decir, que se haya prestado mediante una manifestación del interesado o mediante una clara acción afirmativa.

El consentimiento, además de inequívoco, ha de ser explícito en aquellas situaciones que involucren tratamiento de datos sensibles, adopción de decisiones automatizadas o transferencias internacionales, por lo que, en el caso de los macrodatos, prácticamente siempre se requerirá que se cumpla esta dualidad.

Algunas recomendaciones relacionadas con la solicitud del consentimiento pueden ser:

- **Separar** la aceptación de los términos o condiciones de la aceptación de la política de privacidad con el fin de clarificar las posibles dudas de los interesados y asegurar que la información aportada es siempre suficiente.

- **Evitar** la obtención de consentimientos por omisión y establecer mecanismos capaces de asegurar que el usuario ha accedido de manera inequívoca a ceder su información.

- **Actuar** con transparencia incluso en aquellos casos específicos que la Ley contemple que no sea necesario solicitar el consentimiento y proporcionar información de manera activa sobre dicha situación.

- **Renovar** el consentimiento de los interesados en intervalos apropiados y tantas veces como cambios de finalidad hubiera, siendo estos justificados y compatibles con la inicial.

- **Asegurar** que la edad de los interesados es superior a 14 años, y en caso contrario, solicitar el consentimiento de los padres o tutor legal del mismo.

Además, los interesados deben ser informados de sus derechos y de cómo poder ejercerlos. Entre estos destaca el Derecho al olvido por ser el que mayor repercusión puede llegar a tener en ambas partes, pues permite retirar el consentimiento tras haberlo otorgado en ciertos casos, obligando a las entidades a la eliminación de sus datos.

Ya que la información debe dirigirse directamente al usuario para que pueda expresar su consentimiento o su rechazo, el consentimiento tiene un estrecho vínculo con la transparencia, que se tratará con mayor profundidad en secciones posteriores.



## ./ Finalidad y Licitud

De manera generalizada, será necesario que se informe sobre las finalidades para las que se tratarán los datos. No obstante, y a pesar de que la finalidad debe estar claramente definida antes de obtener el consentimiento pertinente, el tener el consentimiento del usuario no garantiza que la finalidad sea realmente lícita, lo que se agrava en aquellos casos que se contemplen distintas finalidades para los datos.

Resulta fundamental establecer hasta qué punto una actividad de tratamiento, por sus características, tipo de datos a los que se refiere o tipo de operaciones que ejecuta, puede causar algún daño a los interesados. Para ello se requiere clarificar cuáles son las finalidades del tratamiento, para qué se usarán y quién se beneficiará o se verá afectado por ello, lo que se acentúa a la hora de trabajar con datos personales. El principio de Responsabilidad Proactiva tiene un fuerte vínculo con el de licitud, pues de este dependerá la adecuación o no de las medidas tomadas por los responsables.

20 Siguiendo el principio de licitud, el tratamiento, y a su vez, la manera de obtener los datos debe tener una finalidad determinada, explícita y un interés legítimo y responsable. Además, los datos no podrán ser tratados ulteriormente de manera incompatible con dichos fines. Tanto los modelos incluidos en un tratamiento, como el tratamiento en sí, han de tener el propósito de dar respuesta a una necesidad real de la empresa y la industria, y no deben crear y/o reforzar sesgos. Estas soluciones trascienden el ámbito experimental y se van a someter a las leyes del mercado, un mercado regulado y que está obligado a cumplir con unas normas y estándares de calidad.

Es preciso ser vigilantes tanto sobre la legitimidad ética de los tratamientos como de los efectos inesperados de estos, pues debe considerarse y cuantificarse el posible impacto colateral de dichos tratamientos en un entorno social, más allá de las limitaciones concebidas inicialmente de propósito, de duración en el tiempo y de extensión.

En otras palabras, el tratamiento debe ser objeto de ponderación, realizando un análisis de si este, en la forma en que se plantea, es equilibrado porque se deriven más beneficios y ventajas concretas para el interés general y la sociedad en su conjunto, que perjuicios, entendidos estos como los riesgos sobre los derechos y libertades de los sujetos cuyos datos son objeto de tratamiento.

En definitiva, intervienen múltiples elementos técnicos, pero también humanos, de forma directa o colateral, que hay que tener en cuenta para determinar las implicaciones que tiene un tratamiento y su finalidad. De hecho, es el factor humano el que determinará si la finalidad es realmente ética, pues los objetivos del propio proyecto deberán estar alineados con este enfoque para alcanzar una gestión de la información adecuada.

## ./ Minimización y adecuación de la información

Tradicionalmente, a la hora de conseguir información, las organizaciones trataban de adquirir la máxima cantidad posible de datos con el fin de emplearlos en posibles futuros tratamientos y almacenar aquella información que ya hubiera sido utilizado o incluso que quedara sin tratar. Estas prácticas ponen en peligro tanto la protección de los usuarios como la integridad de la empresa, pues se exponen a enormes riesgos de privacidad.

Los procesos de minimización y de adecuación de la información surgen para dar respuesta a este problema protegiendo la privacidad de los interesados sin interferir en las finalidades de tratamiento, dando un mayor paso a la calidad frente a la cantidad de datos para que un proyecto tenga éxito desde un punto de vista ético.

La minimización defiende que la información que se vaya a recolectar debe ser limitada, tanto en el tiempo como en cantidad, a la que vaya a ser estrictamente necesaria de acuerdo con los fines determinados del tratamiento, los cuales, a su vez, deben ser explícitos y legítimos como se citó en su respectiva sección. La minimización requiere de un compromiso activo por parte de los responsables, pues deben asegurar que no existen datos que no se prevén utilizar o recopilar sin utilidad para la finalidad de las actividades de tratamiento. Para aplicarlo correctamente se deberán establecer límites claros identificando los datos que ya se poseen y los que se desean, con el fin de determinar si es necesario obtener más datos, o si fuera necesario incrementar la calidad de estos a la hora de obtenerlos modificando la forma en que se hace.

21

A pesar de que la utilización de Big Data suele venir dada por necesidad debido a la naturaleza de la información de que se dispone (volumen, variedad, etc.), en ocasiones no es así, siendo el empleo de esta tecnología una decisión arbitraria, lo que va en contra del principio de minimización y dificulta un gobierno responsable de los datos. En este tipo de situaciones, el objeto del tratamiento puede ser alcanzado utilizando otro tipo de solución con un margen de rendimiento aceptable y un nivel de riesgo menor.

Por otro lado, por lo que respecta a la adecuación, los datos deben ser adecuados, exactos, pertinentes y estar siempre actualizados bajo revisión periódica, ya que la desactualización de los datos puede perjudicar a los individuos o a terceros. Además, únicamente se podrá identificar a los usuarios a través de sus datos por el tiempo que sea necesario para cumplir los fines del tratamiento, y en aquellos casos que no se pudieran conseguir los fines mediante el uso de anonimización (se tratará con más detalle en la siguiente sección).

Por otro lado, se deberán adoptar todas las medidas razonables para que los datos se supriman o rectifiquen sin dilación si estos son inexactos con respecto a los finales para los que se tratan.

## ./Anonimización

La anonimización es el proceso de tratamiento de datos personales de manera tal que sea imposible vincularlos con la persona a quien se hubiese podido identificar a través de aquellos, mecanismo que tiene como primer objetivo la protección de la privacidad y la confidencialidad de los interesados. La seudonimización, por su parte, es el tratamiento de datos personales de manera tal que no se los pueda atribuir a un individuo determinado sin utilizar información adicional. En este caso se reemplaza el dato identificador (nombre, identificador único, etc.) por un código.

En este contexto, el desarrollo de técnicas de anonimización de datos personales adquiere un valor especial como una fórmula que puede garantizar el avance de la sociedad de la información, donde permanecer en el anonimato absoluto resulta imposible, sin menoscabar el respeto a la protección de datos en el desarrollo de estudios e investigaciones de interés social, científico y económico, pues con esta práctica se puede incrementar considerablemente la protección de la privacidad de los datos de los ciudadanos sin influir en los resultados del tratamiento, ya que un análisis masivo de los datos o macrodatos que puedan derivar de los datos anonimizados no debería diferir del análisis que pudiera obtenerse si hubiera sido realizado con datos no anonimizados.

22

Se debe garantizar que estos procesos se realizan de manera adecuada, pues no es posible garantizar que no se pueda realizar una posterior re-identificación de las personas. Este caso sucedería cuando el diseño de la anonimización no ha sido lo suficientemente robusto (cifrado incompleto, incorrecto...), por lo que, aún diseñándose pensando en su irreversibilidad, será necesario sustentar la fortaleza de la anonimización en medidas de evaluación de impacto, organizativas, de seguridad de la información, tecnológicas y, en definitiva, cualquier medida que sirva tanto para atenuar los riesgos de reidentificación de las personas, como para paliar las consecuencias de que éstos se materialicen.

Someter la información a un proceso de anonimización resulta útil tanto para una posible distribución o transmisión de los datos (detallado en su respectiva sección), como para asegurar que no se emplearán en usos incorrectos o distintos a los expresados en el momento de su recolección, además de resultar también útil **para reforzar la seguridad de la información**, aportando mayores garantías en los tratamientos y en la conservación de datos personales. Si se trabaja con datos que serán publicados parcial o totalmente, será necesario especial hincapié en la anonimización, pues la reidentificación sería posible mediante cruces de datos con otras fuentes.

Recordando y destacando la relevancia del factor humano, una de las claves para garantizar la privacidad de los interesados es la formación e información que se facilite al personal involucrado en el proceso de anonimización y en la explotación de la información anonimizada, donde todo el personal con acceso a los datos debe ser consciente acerca de sus obligaciones.

En aquellos casos que la información no es susceptible de anonimización, es decir, datos no personales, la Ley es considerablemente más permisiva en cuanto a su tratamiento se refiere. Pese a esto, no hay que olvidar los principios anteriormente citados, manteniéndolos como base de cualquier proceso.



## 4.2 Transparencia

El concepto de transparencia en si misma, se fundamenta en la idea de que los individuos cuyos datos son objeto de tratamiento puedan tener, en su caso, un conocimiento efectivo de su información y de los tratamientos que la organización realiza sobre ellos. En otras palabras, los responsables deben informar a los interesados, en virtud del principio de transparencia, acerca de las circunstancias y condiciones del tratamiento de datos a efectuar, así como de los derechos que les asisten, en especial el derecho fundamental a la protección de datos de carácter personal, y facilitar el ejercicio de estos derechos con sus respectivos procedimientos.

La transparencia se ha convertido en una exigencia estandarizada en este nuevo contexto que recae directamente en las empresas y los responsables de la información, y que debe presidir las relaciones entre las partes, especialmente en los casos en que el proveedor de servicios ocupa una posición preeminente sobre los clientes. Esta, de nuevo, debe estar presente en todas las partes del ciclo de vida de los datos (recolección, almacenamiento, tratamiento...) accesos de terceros y en decisiones en base a esta información.

Puesto que abarca un amplio espectro, la opacidad de la información podría provocar una crisis de confianza en los grupos de interés, riesgo que se acentúa si se tiene en consideración que los sistemas informáticos representan un campo de conocimiento a menudo incomprensible para la ciudadanía. Por el contrario, un uso adecuado de la transparencia puede suponer una fuente de confianza para los grupos de interés y de valor para la empresa en consecuencia.

## ./ Qué comunicar

Para que un proveedor pueda ser considerado transparente debe facilitar información específica sobre los ámbitos que se tratan a continuación, en particular sobre los mecanismos que garantizan el cumplimiento de las obligaciones derivadas de la normativa de protección de datos.

En referencia a la información del individuo y su tratamiento:

- La existencia del fichero o tratamiento, su finalidad y destinatarios, en caso de haberlos.
- El plazo o periodo y los criterios de conservación de la información.
- La existencia de decisiones automatizadas y/o elaboración de perfiles.
- La previsión de transferencias a Terceros y los destinatarios.
- El origen o procedencia y las categorías de los datos, en el caso de que los datos no se obtengan del propio interesado.

Sobre los derechos de los interesados, los cuales les facultan para modificar sus datos y limitar el tratamiento, portabilidad o evitar la toma de decisiones individualizada en base a sus datos:

25

Informar sobre el poder de toma de decisiones que mantiene el interesado y de sus garantías exigibles:

- La posibilidad de ejercitar los derechos de acceso, rectificación, cancelación, oposición y olvido.
- La base jurídica o legitimación para el tratamiento.
- El derecho a presentar una reclamación ante las Autoridades de Control.

Y, por último, en cuanto a la forma de contactar con la entidad responsable:

- La identidad y datos de contacto del responsable del tratamiento, así como los datos de contacto del DPD, en su caso.

Por otro lado, también deberán comunicarse, si se dieran, aquellas violaciones de seguridad que ocasionen la destrucción, pérdida o alteración accidental o ilícita de datos personales transmitidos, conservados o tratados de otra forma, o accesos no autorizados a dichos datos. El uso de técnicas de cifrado o seudonimización en este tipo de situaciones reduce considerablemente tanto los posibles riesgos como las posibilidades de que estos se materialicen



## ./ Cómo comunicar

Para que un plan de comunicación sea adecuado y genere confianza, la información debe proporcionarse, en todo momento, de forma concisa, transparente e inteligible evitando el uso de frases que induzcan a confusión o desvirtúen la claridad del mensaje. Además, debe ser sucinta para evitar el cansancio informativo y debe utilizar un lenguaje claro y sencillo, de forma que pueda ser entendida por un usuario medio a la vez que se utilizan canales visibles y fácilmente accesibles (generalmente serán correo postal, mensajería electrónica y/o notificaciones emergentes en servicios y aplicaciones).

Ya que los procedimientos de recogida de información pueden ser muy variados, los modos de informar a los interesados deben adaptarse a las circunstancias de cada uno de los medios empleados para la recopilación o registro de los datos. Si los datos han sido obtenidos directamente del interesado, la información facilitará previamente a la recogida o registro de estos. Si por el contrario fuera por cesión legítima o accesibles desde una fuente pública, el responsable deberá informar dentro de un plazo razonable.

26

Para conseguir que las personas interesadas obtengan la información más relevante de forma rápida y simplificada, pero sin que ello suponga ningún menoscabo de los principios de licitud, lealtad y transparencia, se recomienda ofrecer la información de forma escalonada en dos niveles o capas: presentar una información básica en el primer nivel, de forma resumida, en el mismo momento y en el mismo medio en que se recojan los datos, y remitir a la información adicional en un segundo nivel, donde se presentarán detalladamente el resto de las informaciones, en un medio más adecuado para su presentación y comprensión.

El nombramiento de un DPD es una de las mejores medidas que puede adoptar el responsable para orientar la implementación de las políticas de transparencia, en particular, para gestionar un canal de información a los interesados.

## ./ Recomendaciones

En ocasiones, la opacidad puede ser provocada con la intención de mantener la propiedad intelectual del algoritmo, por la dificultad de entendimiento que genera en los individuos el algoritmo en sí aun siendo este explicable, o bien por las dificultades en explicar el propio sistema, lo que sucede frecuentemente en casos de uso de aprendizaje automático.

Cabe destacar que la transparencia no se limita al deber de publicar información acerca de derechos y procesos determinados, pues requiere además implementar medidas que permitan rendir cuentas sobre las decisiones que toman, sus motivaciones y los impactos esperados.

Así pues, es la transparencia en la planificación, en la toma de decisiones y en la evaluación del impacto es lo que articula una cultura comprometida con la gestión responsable de la información y los datos, identificando sus impactos tanto positivos como negativos, y que debe estar presente más allá de la esfera tecnológica de una empresa calando en la estrategia corporativa.

Otra buena práctica, cada vez más presente y demandada, consiste en someter los sistemas y algoritmos de forma periódica y pública a procesos de autoevaluación, auditoría y monitorización del impacto ético de los algoritmos, productos y/o servicios que se desarrollan. Aquellos que fueran menos relevantes podrían realizarse de manera interna, aunque lo más recomendable sería recurrir a proveedores independientes.



## 4.3 Decisiones automatizadas y perfilado

28

A menudo, la utilización de una tecnología novedosa, como es el caso del Big Data, trae consigo la utilización de otras, por ejemplo, la Inteligencia Artificial. La Inteligencia Artificial se ha convertido en un componente más de los tratamientos de datos realizados por los responsables y que, en muchos casos, aparece en forma de soluciones desarrolladas por terceros.

Esta genera muchas dudas entre los usuarios, investigadores, especialistas, autoridades y la industria con relación a aspectos de cumplimiento normativo, respeto a los derechos de los interesados y seguridad jurídica de todos los intervinientes. Un tratamiento que tome decisiones automatizadas usando la inteligencia artificial puede afectar a personas físicas, como por ejemplo un sistema de autenticación de usuarios. Es por ello por lo que los ciudadanos tienen el derecho a no ser sometidos a procesos de decisión puramente automatizados si se considera que hay riesgos significativos.

A su vez, en la toma de decisiones, la Inteligencia Artificial puede adoptar dos roles: El de una ayuda para el proceso de decisión, proporcionando una inferencia o perfil sobre un sujeto o una situación, para que un ser humano tome la decisión final; o el de toma y ejecución de la decisión. En el primero de los casos, ante un proceso de perfilado, ya que consiste en la clasificación de información sensible en base a ciertos parámetros de manera automática, se presenta un nuevo problema: la discriminación algorítmica.

Esta consiste en el trato diferente de la información de un individuo y perjudicial para el mismo a causa de categorizaciones arbitrarias capaz de proliferar en sistemas de información debido al uso de inteligencia artificial y aprendizaje auto-

mático. Este reflejo de discriminaciones reales producidas en entornos de datos puede ser provocado por el sesgo de las decisiones de aquellos que recolectaron la información, o bien por inconsistencias o incapacidades intrínsecas de los datos a la hora de representar grupos sociales y conduciendo a decisiones algorítmicas erróneas. Por ejemplo, que el sistema difiera en la decisión dependiendo de la raza o etnia.

Estos efectos se pueden reducir teniendo en consideración los riesgos y actuando en consecuencia a la hora de programar los algoritmos con la intención de corregir posibles sesgos que extrapolan los sistemas para predecir el futuro a partir de dinámicas producidas en el pasado. Además, el responsable del tratamiento ha de utilizar procedimientos adecuados para la elaboración de perfiles que garanticen que los datos vinculados con el interesado son exactos, por lo que el proceso de adecuación y la calidad de la información reciben una mayor importancia en relación con el perfilado, especialmente cuando el tratamiento está basado en información biométrica, como reconocimiento facial, huellas dactilares, voz, etc.

Cualquier proceso de toma de decisión, humano o automático, trae, per se, un sesgo implícito que resulta particularmente grave cuando deriva a discriminaciones, **pero** cuando es el algoritmo el que toma enteramente la decisión existe otro tipo de sesgo que puede ser aún más preocupante y que es el del sesgo en la interpretación de los resultados, sesgo humano que consiste en aceptar, sin espíritu crítico, los resultados como ciertos e inamovibles, asumiendo un “principio de autoridad” derivado de las expectativas creadas por los sistemas.

Para evitar estos sesgos, o al menos minimizar su expansión, es necesario analizar la solución desde su diseño, el marco del tratamiento en el que se integra y las relaciones de dicho tratamiento con el entorno cultural, contextual y en los aspectos que se derivan de la interconexión masiva de componentes en la sociedad de la información. Es por ello por lo que es recomendable que los procesos donde intervengan los datos no sean de aplicación enteramente automática, sino que conformen un elemento más de identificación y tipificación los posibles riesgos con el fin alcanzar modelos que permitan una utilización ética de los datos y manteniendo la responsabilidad de la toma de decisiones en el capital humano.

29

La transparencia algorítmica ha surgido como respuesta directa a la discriminación algorítmica con el objetivo de comunicar la ausencia de estos sesgos, y la voluntad de evitarlos por parte de las empresas, otorgándoles un incremento de confianza de los grupos de interés. De esta manera, cuando la elaboración de perfiles implique la toma de decisiones automatizadas será necesario mantener en la esfera de lo público la lógica utilizada, así como la importancia y las consecuencias previstas de dicho tratamiento. Siguiendo esta tendencia, se recomienda la inclusión de la transparencia en el diseño de sistemas de decisión automática.

En esta línea, también han surgido otros derechos como el derecho a la explicabilidad, a través del cual las empresas que manejan procesos de tomas de decisiones automatizadas deberán explicar la manera en que las toman, la lógica aplicada, la importancia y las consecuencias previstas, sobre todo en caso de procesar datos sensibles; o el derecho a la compensación, que tiene el fin de revertir o corregir aquella información errónea que perjudica al individuo por vulneraciones de su privacidad. En este último caso, es aconsejable contar con mecanismos para resarcir a aquellos individuos que pudieran ser afectados de manera negativa.





## 4.4 Control

El control de la seguridad de la información es uno de los ámbitos donde es más común cometer errores ya que, generalmente, estas tareas suelen ser manuales o presentar algún tipo de factor humano y gran parte de los errores cometidos en seguridad de la información son a causa de factores humanos, llegando a ser improbable que no se produzcan errores cuando los procesos se realizan únicamente de forma manual.

Por ello, los datos deben ser tratados de tal manera que se garantice una seguridad adecuada mediante la aplicación de medidas de control apropiadas contra riesgos naturales y humanos, capaces de asegurar la integridad y confidencialidad de la información. A su vez, la Norma exige a las empresas una serie de procedimientos en materia de control y seguridad, pero resulta aconsejable tratar de someter estos procesos a una automatización parcial, en aquellas situaciones que fuera posible, en pro de la privacidad para asegurarla, por ejemplo, en el caso de mantener un registro de actividades de tratamiento.

Otro asunto que requiere de especial atención por parte de los responsables y encargados es el de determinar, en función de los riesgos detectados en cada caso en particular, las medidas de seguridad técnicas y organizativas apropiadas aplicables a los tratamientos que realizan, a través de las que pueden acreditar diligencia. Adherirse a códigos de conducta o certificarse dentro de los esquemas previstos por el RGPD puede ayudar en gran medida en este aspecto.

Entre las medidas que los responsables deben tomar se pueden encontrar

- Documentación de la totalidad de las operaciones y actividades realizadas sobre cada conjunto estructurado de datos.
- Evitar accesos no autorizados o ilícitos a la información y mantener monitorizados a los usuarios.
- Realizar copias de seguridad y controles de almacenamiento para evitar posibles pérdidas de datos.
- Cifrar y/o seudonimizar la información.

Designar a un DPD, como se había comentado en el capítulo de transparencia, es una medida que también supone un beneficio en cuanto al control y seguridad de la información se refiere. Los responsables habrán de elegir únicamente encargados que ofrezcan garantías suficientes para aplicar dichas medidas técnicas y organizativas apropiadas, de manera que el tratamiento sea conforme con los requisitos del Reglamento y con el marco ético al que aspire la entidad.





## 4.5 Transferencias de información

En cuanto a transferencias de información se refiere, son de suma importancia la transparencia y la anonimización de la información. Una vez más, bajo el principio de transparencia, se debe informar a los interesados de la realización del proceso, siendo estrictamente obligatorio en el caso de transferencias internacionales.

Las garantías sobre la protección que recibirán los datos en su destino las debe ofrecer el exportador, que podrá ser tanto un responsable como un encargado de tratamiento. La transferencia solo se realizará una vez se hayan ofrecido garantías adecuadas sobre la protección que los datos recibirán en su destino y que se encontrarán dentro de un marco ético aceptable, y se realizarán únicamente a territorios o sectores específicos sobre los que la Comisión haya adoptado una decisión reconociendo que ofrecen un nivel de protección adecuado,

Además, es importante establecer mecanismos para permitir que las contrataciones que se realicen en este contexto de transferencias se gestionan con fluidez, asegurando al mismo tiempo que el cliente responsable tiene información suficiente sobre los contratistas, o potenciales contratistas, y mantiene la capacidad de tomar decisiones; especialmente cuando existan flujos transfronterizos de datos a terceros países que conlleven una transferencia internacional de datos.



## 4.6 EIPDs

En la gran mayoría de escenarios que involucren el uso de macrodatos, los responsables se encontrarán en la obligación de realizar un EIPD o evaluación de impacto de protección de datos con carácter previo a la puesta en marcha de los tratamientos.

Esto es debido al riesgo significativo para los derechos y libertades de las personas físicas que conlleva un tratamiento a gran escala, pues como se ha venido diciendo, el uso de nuevas tecnologías intensifica los riesgos a la hora del tratamiento y, en el caso de los macrodatos, estos riesgos se acentúan ya que su tratamiento suele ir vinculado a la evaluación sistemática y exhaustiva de aspectos personales de personas físicas basados en un tratamiento automatizado, como la elaboración de perfiles o procesos de toma de decisiones automatizados.

Puesto que el desarrollo en profundidad de una EIPD escapa al alcance de este documento, se hace saber que la Agencia española de protección de datos tiene a disposición de los responsables una Guía práctica para las Evaluaciones de Impacto en la Protección de los datos sujetas al RGPD. No obstante, se expondrá brevemente en qué consiste y cómo debe realizarse.

Una evaluación de impacto, en esencia consiste en un ejercicio teórico que requiere su puesta en práctica de forma íntegra para garantizar los derechos y las libertades de los interesados siguiendo un modelo proactivo. Tiene la finalidad de determinar, de manera anticipada, el nivel de riesgo que entraña un tratamiento, con el objetivo de establecer las medidas de control más adecuadas para reducir el mismo hasta un nivel considerado aceptable.

Para poder gestionar los riesgos a los que están expuestas las actividades de tratamiento, el responsable, asesorado en su caso por el DPD y con carácter previo a la realización del tratamiento, deberá clarificar la descripción prevista del ciclo de vida de los datos, e identificar y evaluar los riesgos, así como la estimación de la necesidad y proporcionalidad del tratamiento, de acuerdo con la finalidad que se persiga, para desarrollar una respuesta ante estos que se materialice en mecanismos, medidas de seguridad y garantías en pro de la protección de la información.

Una vez finalizado, el resultado de la evaluación debe influir, necesariamente, en la decisión de realizar, o no, finalmente el tratamiento de dato. Asimismo, debe considerarse como un proceso de mejora continua que se realizará iterativamente cuando se produzcan cambios en el planteamiento de los tratamientos realizados por la organización.





## 4.7 Otras buenas prácticas

Ya que el incesante cambio al que están sometidos los diferentes sectores provoca a su vez la constante innovación por parte de las empresas, cada vez son más las empresas que monitorizan otras buenas prácticas realizadas por entidades similares dentro del mismo sector aplicables directamente a este. Puesto que la industria aparentemente está avanzando en dirección a una gestión ética, apunta a que la creación de observatorios de buenas prácticas se materializará tarde o temprano.

Una de las mejoras prácticas que puede realizar una empresa es tratar de extrapolar al resto de equipos o entidad en conjunto el diseño de proyectos y cultura ética más allá del área tecnológica. Esta es la manera más accesible de perpetuar una gestión adecuada, aunque su dificultad dependerá completamente de la capacidad de la empresa de cambiar su cultura desde sus raíces. En esta línea, contar con un comité ético o la realización de pruebas o proyectos piloto a menor escala puede ayudar a perfeccionar una metodología ética en otros ámbitos que no fueran el tecnológico.

Consiguientemente y, aunque venga implícito en los objetivos de la guía, cabe destacar una vez más la importancia de la formación del personal para poder hacer efectiva una cultura ética, pero no sólo de los responsables directamente implicados. Todo el personal debe ser consciente del impacto social, moral y legal que conlleva su labor, ya que con equipos multidisciplinares se alcanza una visión de conjunto más amplia con la que conseguir más fácil y eficazmente los objetivos éticos.



## 5. Conclusiones

La gestión ética de los datos supone una preocupación relevante en la actualidad, con vistas a incrementarse de cara al futuro. Siguiendo un marco ético basado en buenas prácticas desde el diseño, de manera que se contribuya a reducir los riesgos asociados a la equidad, privacidad y transparencia de la información, las empresas están contribuyendo a construir un entorno fundamentado en la confianza ciudadana, generando valor y beneficiando tanto a sí mismas como a la sociedad.

La tecnología en general, y en particular las tecnologías basadas en Big Data, puede tener un efecto multiplicador de las carencias éticas que ya están presentes en la sociedad o aquellas que se arrastran desde el pasado y que están registradas en datos históricos. El empleo de estrategias de transparencia, gestión del riesgo y mecanismos de auditoría y certificación no sólo permitirán el cumplimiento de lo establecido en el RGPD, sino que mejorarán la confianza de los usuarios en los productos y servicios. Es más, las aplicaciones de Big Data pueden suponer un gran apoyo para proteger la privacidad e implementar mecanismos que aseguren la protección de datos.

En esta guía se han recogido aspectos que deben ser tenidos en cuenta para asegurar la protección de los datos desde el diseño y por defecto, además de ser de utilidad para verificar el nivel de compromiso ético de cualquier organización con el fin de garantizar la licitud del tratamiento o mejorar aspectos relacionados, por ejemplo, con la transparencia y la información que se facilita a los interesados. Estas medidas reflejan muy directamente el enfoque de responsabilidad proactiva, pensando en términos de protección de datos desde el mismo momento en que se diseña un tratamiento, un producto o servicio que implica el tratamiento de datos personales.

```
    .trigger("themes:update")
    .on("themes:update", function(a) {
        var $ = this;
        $ .close-full-overlay("
        review"), render: function()
        .navigate(c.router.base
        .sel.addClass("iframe-read
        .removeClass("iframe-ready"
        .trigger("preview:close"), thi
        this.sel.toggleClass("coll
        view-device", c), this.toogle
        attr("aria-pressed", !0)}, k
        ("disabled") || (wp.updates.
        ))))}, c.view.Themes=wp.Bar
        .currentTheme(), this.lister
        nt.text(c.collection.length
        .text("There are 1 theme(s) for
```

## 6. Referencias

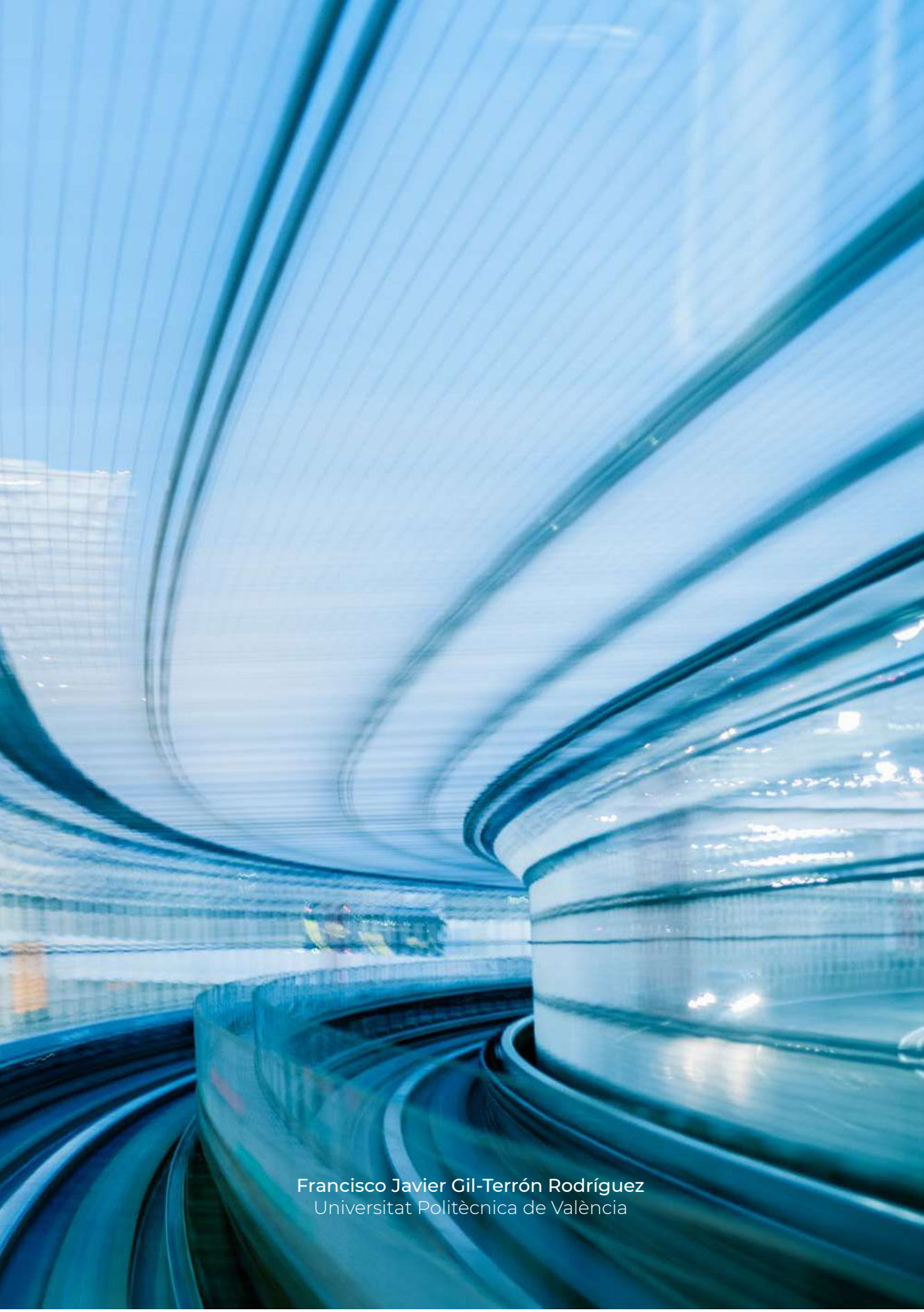
- Agencia española de Protección de Datos (2018). Adecuación a la normativa de protección de datos “a coste cero.”
- Agencia española de Protección de Datos (2018a). Adecuación al RGPD de tratamientos que incorporan Inteligencia Artificial.
- Agencia española de Protección de Datos (2019). EL DEBER DE INFORMAR Y OTRAS MEDIDAS DE RESPONSABILIDAD PROACTIVA EN APPS PARA DISPOSITIVOS MÓVILES.
- Agencia española de Protección de Datos (2018b). Guía del Reglamento General de Protección de Datos para Responsables de Tratamiento.
- Agencia española de Protección de Datos (2018c). Guía para el cumplimiento del deber de informar.
- Agencia española de Protección de Datos (2018d). GUIA PRÁCTICA de Análisis de riesgos en los tratamientos de datos personales sujetos al RGPD.
- Agencia española de Protección de Datos (2018e). Guía práctica para las Evaluaciones de Impacto en la Protección de los datos sujetas al RGPD.
- Agencia española de Protección de Datos (2020). Guía sobre el uso de las cookies.
- Buenadicha, C., Clavell, G., Hermosilla, M. P., Loewe, D., & Pombo, C. (2019). La gestión ética de los datos. Banco Interamericano de Desarrollo.
- Agencia española de Protección de Datos (2018f). LEY ORGÁNICA 3/2018, DE 5 DE DICIEMBRE, DE PROTECCIÓN DE DATOS PERSONALES Y GARANTÍA DE LOS DERECHOS DIGITALES NOVEDADES PARA LOS CIUDADANOS.
- Boletín Oficial del Estado. (2018). Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales. BOE (Boletín Oficial Del Estado).
- Agencia española de Protección de Datos (2018g). Listado de cumplimiento normativo.
- Agencia española de Protección de Datos (2016). LISTAS DE TIPOS DE

## TRATAMIENTOS DE DATOS QUE REQUIEREN EVALUACIÓN DE IMPACTO RELATIVA A PROTECCIÓN DE DATOS (art 35.4).

- Agencia española de Protección de Datos (2018h). ORIENTACIONES PARA PRESTADORES DE SERVICIOS DE 'CLOUD COMPUTING.'
- Agencia española de Protección de Datos (2016a). Orientaciones y garantías en los procedimientos de anonimización de datos personales.
- Agencia española de Protección de Datos (2018i). Protección de Datos: Guía para el Ciudadano.
- Parlamento Europeo. (2016). REGLAMENTO (UE) 2016/679 DEL PARLAMENTO EUROPEO Y DEL CONSEJO de 27 de abril de 2016 relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga l. Diario Oficial de La Unión Europea.







Francisco Javier Gil-Terrón Rodríguez  
Universitat Politècnica de València