GUÍA PARA LA APERTURA Y COMPARTICIÓN DE DATOS EN EL ENTORNO EMPRESARIAL

fundamentos, recomendaciones y buenas prácticas

septiembre 2019

FUNDACIÓN COTEC PARA LA INNOVACIÓN

ÍNDICE

CÓMO UTILIZAR ESTA GUÍA 5

1. EXPERIMENTACIÓN Y CONCIENCIACIÓN 7

Poniendo en marcha una 1ª experiencia en compartición o apertura de datos 8

2. CÓMO DEFINIR UNA BUENA ESTRATEGIA DE DATOS ABIERTOS O COMPARTIDOS 10

3. CÓMO ORIENTAR LA EMPRESA HACIA LA COMPARTICIÓN O APERTURA DE DATOS: PREPARACIÓN TÉCNICA 13

4. PONIENDO EN PRÁCTICA EL CICLO DE VIDA DE LOS DATOS: RECOGER, PREPARAR, PUBLICAR Y MANTENER 15

5. ¿CUÁL ES EL NIVEL DE ÉXITO? MONITORIZACIÓN DE LA INICIATIVA 17

1. INTRODUCCIÓN 20

1.1 ¿Por qué debería compartir o abrir los datos de mi empresa? 20

1.2 ¿Cuál es la diferencia entre la compartición y la apertura de datos? 21

1.3 ¿Por dónde empezar? 22

1.4 ¿Qué contiene esta guía y a quién está dirigida? 23

1.5 Aspectos relevantes de la apertura/compartición de datos 26

1.6 ¿Quién ha realizado esta guía? 26

2. LA APERTURA Y COMPARTICIÓN DE DATOS EN EL ÁMBITO EMPRESARIAL 28

3. ¿QUÉ PUEDE HACER UNA EMPRESA CON SUS DATOS? 30

4. FUNDAMENTOS TÉCNICOS DE LA APERTURA/COMPARTICIÓN DE DATOS EN EL ENTORNO EMPRESARIAL 32

4.1 Experimentación y concienciación 33

4.2 Definir una buena estrategia de datos abiertos o compartidos 36

4.3 Preparación técnica: 39

4.3.1 Gestión de Datos 40

4.3.2 Extraer, Transformar y Compartir Datos (ETCD) 42

4.3.3 Canales de Compartición de datos 43

4.3.4 Búsqueda de Datos 47

4.3.5 Otros aspectos técnicos 47

4.4 Poner en práctica el ciclo de vida de los datos: recoger, preparar, publicar y mantener 48

4.4.1 Recoger datos 48

4.4.2 Preparar datos 50

4.4.3 Publicar datos 53

4.4.4 Mantener datos, metadatos persistentes 54

4.5 Monitorización de la iniciativa, análisis del nivel de éxito y mejora continua 54

5. ASPECTOS RELEVANTES DE LA APERTURA/COMPARTICIÓN DE DATOS EN EL ENTORNO EMPRESARIAL 56

5.1 Modelos de Relación 56

5.1.1 B2B (Business to Business) 56

5.1.2 B2G (Business to Government) 60

5.2 Modelos de Financiación 60

5.3 Marco Legal 64

5.3.1 Acuerdos Contractuales en el modelo de relación B2B 64

5.3.2 Acuerdos contractuales en el modelo de relación B2G 66

5.3.3 Licencias para la Compartición/Apertura de Datos 68

5.4 Roles, habilidades, desarrollo y gestión del conocimiento 71

5.4.1 Roles 71

5.4.2 Habilidades 72

5.4.3 Desarrollo del conocimiento 72

5.4.4 Gestión del conocimiento 73

6. GLOSARIO EXPLICATIVO DE TÉRMINOS Y SIGLAS 74

REFERENCIAS Y BIBLIOGRAFÍA 76

ANEXO I: ESCENARIOS DEL PROCESO ETCD (EXTRACCIÓN Y TRANSFORMACIÓN) 78

ANEXO II: PRINCIPIOS QUE GUÍAN LOS MODELOS DE RELACIÓN 82

ANEXO III: LEGISLACIÓN APLICABLE 83

ANEXO IV: DESCRIPCIÓN DE ALGUNOS CONCEPTOS TÉCNICOS CLAVE EN EL ÁMBITO DE OPEN DATA 91

ANEXO V: SELECCIÓN NO EXHAUSTIVA DE HERRAMIENTAS Y APLICACIONES DE SOFTWARE LIBRE PARA LA IMPLEMENTACIÓN DE OPEN DATA 96

ANEXO VI:CUESTIONARIO SOBRE INICIATIVAS DE DATOS ABIERTOS EN EL SECTOR PRIVADO 98

ANEXO VII: DESCRIPCIÓN DE CASOS DE APERTURA Y COMPARTICIÓN DE DATOS 107

PRIMAFRIO 107

GRUPO EUSKALTEL 109

TELEFÓNICA 111

SUEZ 113

ITI 115

EDP 117

ARCELOR MITTAL 119

ÍNDICE DE TABLAS

Tabla 1 ArcelorMittal: Ejemplo Datatón 34
Tabla 2 SUEZ: Ejemplo Datatón 35
Tabla 3 5 estrella de Tim Bernes Lee 53
Tabla 4 BBVA VALORA: Ejemplo Datos Abiertos/Compartidos 56
Tabla 5 PRIMAFRIO: Ejemplo Datos Abiertos/Compartidos 57
Tabla 6 EUSKALTEL: Ejemplo Monetización 58
Tabla 7 ITI: Ejemplo Compartición de Datos Sectorial (Financiación Europea - H2020) 59
Tabla 8 Licencias para compartir/abrir datos 69
Tabla 9 Disposiciones para Licencias 70
Tabla 10 Principios Modelo de Negocio B2B 82
Tabla 11 Principios Modelo de Negocio B2G 82
Tabla 12 Aplicación del Reglamento 84
Tabla 13 Principios del Reglamento 84
Tabla 14 Fundamentos Jurídicos para el Tratamiento de Datos 85
Tabla 15 Obligaciones 86
Tabla 16 Solicitudes de los Ciudadanos 87
Tabla 17 Cumplimiento y Sanciones 88

ÍNDICE DE ILUSTRACIONES

Ilustración 1 Fundamentos Apertura/Compartición de Datos en Entorno Empresarial 32
Ilustración 2 Preparación previa a la puesta en marcha del ciclo de vida de datos abiertos/compartidos 40
Ilustración 3 Estructura Descentralizada 40
Ilustración 4 Estructura Centralizada 41
Ilustración 5 Estructura Completamente Centralizada 42
Ilustración 6 Proceso ETCD o ETDA 42
Ilustración 7 Canales de Compartición de Datos 43
Ilustración 8 Ciclo de Vida de Apertura/Compartición de Datos 48
Ilustración 9 Actividades de Recolección 49

CÓMO UTILIZAR ESTA GUÍA

septiembre 2019

COTEC
FUNDACIÓN COTEC PARA LA INNOVACIÓN

CÓMO UTILIZAR ESTA GUÍA

La *Guía para la apertura y compartición de datos en el entorno empresarial* tiene un doble objetivo: por un lado, acercar a las empresas las oportunidades que supone la apertura o compartición de los datos que tiene en su poder, y por otro, acompañar a las organizaciones que quieran iniciar un proceso de este tipo. La guía incluye diferentes apartados o pasos a seguir:

1. **EXPERIMENTACIÓN Y CONCIENCIACIÓN** 7

   Poniendo en marcha una 1ª experiencia en compartición o apertura de datos 8

2. **CÓMO DEFINIR UNA BUENA ESTRATEGIA DE DATOS ABIERTOS O COMPARTIDOS** 10

3. **CÓMO ORIENTAR LA EMPRESA HACIA LA COMPARTICIÓN O APERTURA DE DATOS: PREPARACIÓN TÉCNICA** 13

4. **PONIENDO EN PRÁCTICA EL CICLO DE VIDA DE LOS DATOS: RECOGER, PREPARAR, PUBLICAR Y MANTENER** 15

5. **¿CUÁL ES EL NIVEL DE ÉXITO? MONITORIZACIÓN DE LA INICIATIVA** 17

1. EXPERIMENTACIÓN Y CONCIENCIACIÓN

Objetivo de esta fase

Identificar el potencial de los datos de la empresa.

Situación de partida de la empresa

Es el paso previo a iniciar un proyecto de compartición o apertura de datos. Esta fase está enfocada a aquellas compañías que no conocen el potencial de la información que manejan y quieren identificarlo.

Si la empresa **dispone de conocimiento interno** experto en analítica de datos

La empresa debe involucrar a ese personal para realizar un análisis de los datos existentes e identificar su potencial.Una vez identificado el potencial de los datos, el siguiente paso es la **Definición de la Estrategia**.

Si la empresa **no dispone de conocimiento interno** experto en analítica de datos

Lo más adecuado es buscar **apoyo en personal técnico** externo que pueda asesorar a la empresa en cuanto al valor y potencial de sus datos.

En la guía se incluyen diferentes tipos de apoyo:

- **Facilitadores tecnológicos**: Bajo un formato de colaboración, la empresa ofrece una parte de sus datos para que un centro tecnológico, universidad o consultora especializada los analice y extraiga conclusiones sobre su potencial.
- **Datatón**: Concurso donde la empresa pone a disposición una parte de sus datos a equipos de participantes con perfiles expertos. Durante un tiempo determinado, los equipos experimentan con los datos para extraer su valor.
- **Marketplace / Plataforma Industrial**: Entornos online de confianza donde las empresas comparten sus datos con otras compañías y agentes reutilizadores de datos.

Ir a Poniendo en marcha una 1ª experiencia en compartición o apertura de datos.

Apartado Experimentación y Concienciación: pág. 33

Más información:
Marketplaces y Plataformas Industriales en **Canales de compartición de datos**, pág. 43.
Modelos de relación o negocio para el intercambio de datos, pág. 56.

Siguiente fase

Cómo definir una buena estrategia de datos abiertos o compartidos

Paso previo opcional

Poniendo en marcha una 1ª experiencia en compartición o apertura de datos

Poniendo en marcha una 1ª experiencia en compartición o apertura de datos

Si la organización no tiene experiencia previa en procesos de compartición o apertura de datos, es una buena opción empezar mediante un proceso de experimentación que permita conocer el valor de los datos. Un proyecto de este tipo servirá también para entender en qué consiste un proceso de compartición o apertura de datos, de cara a impulsar en el futuro una iniciativa de este tipo.

En la primera fase se han indicado varias modalidades para iniciarse en el ámbito de la compartición o apertura de datos si la organización no tiene experiencia previa. Es importante tener en cuenta que en los casos de Datatón y Marketplace, implican la compartición o apertura de datos con terceros. Si bien existen otros modelos, nos centraremos en estas dos fórmulas.

1. Definición de los objetivos y plan de acción

- Establecer el **objetivo** a conseguir mediante esta primera experiencia. En el caso del datatón, puede ser la búsqueda a una solución a un reto concreto de la empresa. En el caso del marketplace, el objetivo puede ser comprobar el valor de los datos para un futuro proceso que implique su monetización, entre otros.
- Establecer qué **conjunto de datos** se van a compartir o abrir.
- Decidir bajo qué **modelo de relación** se va a realizar la iniciativa. Siendo una primera experiencia, este itinerario se centra en las modalidades de Datatón o Markeplace.
- Búsqueda de **apoyo experto**: es importante contar con un conocimiento experto a nivel tecnológico que acompañe a la organización en el proceso.
    - **Datatón**: la iniciativa puede llevarse a cabo de la mano de un agente colaborador o facilitador tecnológico como un centro tecnológico, una universidad o una consultora especializada. Ejemplos de datatones: ArcelorMittal y Suez, págs. 33 y 34.
    - **Marketplace**: la propia plataforma puede contar con servicios que asesoran a las organizaciones sin experiencia previa en los procesos de compartición o apertura de datos.
- **Asesoría jurídica**: es fundamental contar con un asesoramiento en cuestiones legales para garantizar que los datos que se pondrán a disposición de terceros cumplen con la legislación en vigor.

Más información:
Modelos de relación, pág. 18
Marco legal para los diferentes modelos de relación, pág. 64
Ejemplos de casos de apertura y compartición de datos, pág. 107
Legislación aplicable a la apertura o compartición de los datos, pág. 83

2. Preparación organizativa y técnica

Esta fase incluye los aspectos relacionados con la gestión de los datos y la infraestructura tecnológica. Si bien en esta primera experimentación no se requiere aún iniciar un plan de acción completo, si es conveniente tener en cuenta algunos aspectos relacionados con los procesos de gestión, preparación y transformación de los datos. Esta información será útil para entender qué cambios organizativos y de infraestructura tecnológica son necesarios en el caso de llevar a cabo una futura iniciativa de compartición o apertura de datos completa.

Aspectos a tener en cuenta:

- Analizar en qué **formatos** están los datos que se van a compartir y si son utilizables por terceros.
- Establecer qué **procesos de Extracción**, **Transformación y Compartició**n deberían llevarse a cabo, con apoyo experto.
- Establecer qué **procesos de recogida**, **preparación y publicació**n deberían llevarse a cabo para obtener datasets de calidad que puedan ser reutilizados fácilmente por terceros, acompañados también por apoyo experto.
- Decidir bajo qué **licencias** se van a publicar los datos.
- Establecer medidas para **monitorizar** la iniciativa y poder obtener una valoración acerca de los resultados conseguidos.

Más información:
Extraer, Transformar y Compartir datos, pág. 42
Puesta en marcha del ciclo de vida de los datos (proceso para generar datasets), pág. 48
Licencias, pág. 68
Monitorización de la iniciativa, pág. 54

---

3. Comunicación y concienciación

Tras esta primera iniciativa, es importante recoger los resultados de la experiencia, los beneficios y aprendizajes obtenidos, y comunicarlos al resto de la organización, en especial a los niveles directivos. Esto permitirá dar a conocer las oportunidades que ofrece la compartición y apertura de datos, pero también servirá para conocer las decisiones que deben ser tomadas a nivel estratégico, metodológico y tecnológico en un proceso de este tipo.

Siguiente fase

Cómo definir una buena estrategia de datos abiertos o compartidos

---

2. CÓMO DEFINIR UNA BUENA ESTRATEGIA DE DATOS ABIERTOS O COMPARTIDOS

Objetivo de esta fase

Antes de empezar una iniciativa de compartición o apertura de datos, es importante tener una estrategia clara que defina los objetivos, la estrategia a seguir y las decisiones que deben tomarse a nivel organizativo.

Situación de partida de la empresa

La organización ya ha identificado el potencial de sus datos y quiere iniciar un proyecto de apertura o compartición de sus datos con terceros.

Pasos que incluye esta fase

1. Diagnóstico de partida

El primer paso consiste en tener una visión clara de la **situación actual** de la organización en relación a los datos. Para ello, es aconsejable identificar las personas clave que pueden ayudar a realizar este diagnóstico. Preguntas a realizar:

- ¿Qué unidades de la organización recogen, utilizan y/o producen datos?
- ¿Cómo se están gestionando estos datos?
- ¿Hay ya datos abiertos o compartidos? Si es así, ¿cómo se están publicando?
- ¿Qué datos se quieren compartir?
- ¿Cuál es el flujo de datos actual en la organización?

2. Definición de los objetivos estratégicos

El siguiente paso consiste en establecer **qué se quiere conseguir** a través de la iniciativa de compartición o apertura de datos y en qué plazos. Para ello, existen varias decisiones a tomar:

- Clarificar el **objetivo**: generar una nueva línea de negocio, promover la innovación, resolver un problema, monetizar los datos...
- A partir del objetivo, decidir el **Modelo de Relación** (o negocio) bajo el cual compartiremos los datos. Puede ser bajo un modelo B2B (*Business to Business*) o B2G (Business to Government) y siguiendo diferentes modalidades (publicación en portal de datos abiertos, monetización, compartición en plataforma industrial sectorial...)
- Identificar el **conocimiento interno** para llevar a cabo la iniciativa. ¿Es necesario buscar apoyo especializado externo?
- ¿Es necesario buscar **apoyo legal**?
- Definir medidas asociadas a los objetivos que permitan medir los resultados.
- Cómo visualizar el flujo de datos destino.

Información relacionada:
- Clarificar objetivo: **¿Por qué las compañías comparten/abren sus datos?**, pág. 31
- **Modelos de relación**, pág. 56
- **Marco legal** para los diferentes modelos de relación, pág. 64
- **Principios que guían los Modelos de Relación**, pág. 82
- **Roles, habilidades, desarrollo y gestión del conocimiento**, pág. 71
- **Casos de apertura y compartición de datos**, pág. 107

---

3. Definir el plan de acción e implantar la estrategia

Establecer el plan que permita alcanzar los objetivos marcados. Los aspectos principales de esta fase son:

- Asegurar la implicación de las personas involucradas en al iniciativa (internas o externas).
- Establecer medidas para asegurar la **rentabilidad** de la iniciativa, analizando el coste-beneficio.
- Definir cuál es el modelo de **gestión de los datos** que mejor encaja con los objetivos (Este aspecto se desarrolla en la fase Preparación Técnica).
- Analizar en qué **formatos** están los datos y si son utilizables por terceros. Identificar si hay datos sensibles.
- Establecer qué **licencias** se van a utilizar.
- Definir y establecer los procesos ETPD e identificar la tecnología necesaria (este aspecto se desarrolla en la fase Preparación Técnica)

Información relacionada:
- Estructuras de **Gestión de los Datos**, pág. 40
- **Modelos de financiación** y prácticas para cuantificar el valor de los datos, pág. 60
- **Licencias** para la compartición o apertura de datos, pág. 68
- **Preparación Técnica** previa a la puesta en marcha de la iniciativa de compartición de datos, pág. 39
- Herramientas para **Monitorizar la reutilización de los datos**, pág. 54

---

4. Definir y mantener el inventario de datos

Enfocada a establecer los aspectos a tener en cuenta en relación a los datasets:

- Garantizar el cumplimiento de las **normativas** en relación a la protección de datos e identificar si hay **datos sensibles**.
- Cuantificar el **valor económico** de los datasets en base al modelo de financiación.

Información relacionada:
- **Legislación** aplicable a la apertura o compartición de datos, pág. 83
- Gestión de los **datos sensibles**, pág. 52
- **Valoración de los datasets** y cálculo de tarifas, pág. 61

5. Definir la política de datos abiertos o compartidos y comunicarla

Definir una política a nivel organizacional y comunicarla es esencial porque aumenta la transparencia de la organización, contribuye a concienciar acerca de los beneficios de la apertura de datos y redunda en el éxito de la iniciativa.

Apartado Definición de la Estrategia: pág. 36

Siguiente fase

Cómo orientar la empresa hacia la apertura de datos – Preparación Técnica

3. CÓMO ORIENTAR LA EMPRESA HACIA LA COMPARTICIÓN O APERTURA DE DATOS: PREPARACIÓN TÉCNICA

Objetivo de esta fase

En esta fase se toman las decisiones ligadas a la preparación técnica y la implementación del proceso de compartición o apertura de datos. Esto incluye los aspectos relacionados con la gestión de los datos y la infraestructura tecnológica que se va a utilizar.

Situación de partida

La organización ya ha identificado el potencial de los datos de su empresa y ha tomado las decisiones esenciales acerca de cómo se realizará el proyecto mediante la definición de la estrategia. Ya ha decidido cuál es el objetivo de la iniciativa y el modelo de relación a seguir, el equipo especializado, los datos que se van a compartir y el modelo de financiación del proyecto.

Decisiones técnicas a tomar en esta fase

1. Cómo va a ser la gestión de los datos

La organización debe decidir cuál debe ser la estructura de gestión o gobernanza de los datos. Si aún no existe, es muy aconsejable establecer una antes de iniciar la compartición o apertura de datos. A nivel teórico, pueden ser:

- **Gestión de datos descentralizada**:
  La administración de los datos se realiza por unidades o personas que tienen la responsabilidad de gestionar sus propios datos.
- **Gestión de datos centralizada – federada**:
  La gestión de los datos está distribuida, pero la responsabilidad está centralizada.
- **Gestión de datos completamente centralizada**:
  La estandarización se define en el nivel más alto, garantizando la consistencia a lo largo de toda la organización.

Cambiar hacia una estructura de gestión de los datos tiene un impacto importante en la organización, así que si no hay una experiencia previa, es aconsejable conseguir apoyo por parte de agentes expertos en gestión de los datos.

Más información: Gestión de los Datos, pág. 40

2. Proceso ETCD (Extraer, Transformar y Compartir Datos)

El proceso ETCD consiste en las especificaciones técnicas de cómo los datos fluyen dentro de la organización, se transforman en datasets publicables y, eventualmente, se comparten. Es un proceso que debe definirse a nivel organizacional.

Consta de 3 pasos:

Extraer: Los datos se pueden extraer de cualquier tipo de fuente (interna o externa). El proceso de extracción dependerá de la estructura de gestión de los datos.

Transformar: Los datos deben transformarse en datasets listos para compartir. Es una fase que ha de estandarizarse mediante la definición de una guía de preparación de los datos.

Compartir: Esta fase no es necesaria si los datasets son para uso interno. Consiste en identificar los canales donde compartir o abrir los datos (siguiente apartado).

El proceso relacionado con la preparación específica de cada dataset, incluyendo los aspectos de calidad, inclusión de metadatos y aspectos legales, están recogidos en la fase **Ciclo de vida de los datos**, pág. 48

Para más información sobre el proceso ETCD, se puede consultar el **Anexo I – Escenarios del proceso ETCD**, pág. 78

---

3. Canales donde compartir los datos

Existen diferentes canales donde compartir o abrir los datos. La elección del canal dependerá del número de datasets a compartir y de las necesidades específicas de cada organización. Se recomienda empezar por las soluciones más sencillas. Los principales canales son:

- Descargas vía Web.
- API (Interfaz de Programación de Aplicaciones).
- Portal de Datos (propio o de terceros).
- Marketplace.
- Plataforma Industrial.
- Facilitador técnico.

Más información: **Canales de compartición de datos**, pág. 43

---

4. Búsqueda de Datos y otros aspectos técnicos

Es aconsejable implementar funcionalidades para facilitar la búsqueda de los datos. En función del canal escogido para compartirlos, se pueden utilizar mecanismos de mayor o menor complejidad.

También deben tomarse otras decisiones técnicas, relacionadas con la selección del dominio y el hosting y la administración del canal donde se comparten los datos.

Más información: **Búsqueda de Datos y otros aspectos técnicos**, pág. 47

Siguiente fase

Poniendo en práctica el ciclo de vida de los datos: recoger, preparar, publicar y mantener

4. PONIENDO EN PRÁCTICA EL CICLO DE VIDA DE LOS DATOS: RECOGER, PREPARAR, PUBLICAR Y MANTENER

Objetivo de esta fase

Establecer los pasos concretos a seguir para, partiendo de los datos, generar cada uno de los datasets y que estén listos para ser compartidos o abiertos. Incluye los procesos de recoger los datos, prepararlos, publicarlos y mantenerlos.

Situación de partida de la empresa

La empresa ya ha establecido su estrategia para la apertura o compartición de sus datos y ha tomado las decisiones organizativas y técnicas en relación a:

- Cómo se van a gestionar los datos.
- Cómo se realizará el proceso ETCD.
- Qué canal se va a utilizar para compartir los datos.

Pasos que incluye esta fase:

1. Recogida de datos

Esta fase consiste en:

- Mapear o identificar los datos disponibles.
- Priorizar aquellos que son más adecuados (porque encajan con el objetivo, porque tienen mayor calidad,...).
- Categorizar los datasets y publicar las categorías.

Más información: Fase de **Recogida de datos**, pág. 48

2. Preparación de los datos

Una vez los datos han sido recolectados, deben ser preparados para su compartición o apertura, puesto que no pueden publicarse en bruto. **Esta fase es clave, puesto que si los datos no cumplen con la legislación vigente, no se pueden compartir o abrir**.

Hay varios aspectos a tener en cuenta:

- Determinar la **calidad del dato**: Si están completos, si están limpios y son precisos, si están actualizados y si siguen unos estándares establecidos.

- **Prepararlos para su publicación**:
  Los datos deben ser publicados bajo el método ***linked* data**, o datos enlazados, para permitir su interoperabilidad. Es esencial que lleven asociados **metadatos**, esto es, información estructurada que describe el dataset y permite que sea encontrado.

- **Preparación legal**: Se debe comprobar que los datos respetan la legislación en vigor. Además, cada dataset debe tener asociada una licencia que establece sus condiciones de uso.

- **Gestión de los datos sensibles**: Si los datasets incluyen datos sensibles, necesitarán una preparación especial.

- **Chequeo final** para verificar la preparación de los datasets en todos los aspectos nombrados (calidad, interoperabilidad e inclusión de metadatos, licencias,...)

Más información: Fase **Preparar los datos**, pág. 50

Información relacionada:

- Qué son los **Metadatos**, pág. 53
- Métodos para **determinar la calidad del dataset** (MELODA 5, Estrellas de Tim Berners-Lee y FAIR), pág. 53
- Preparación técnica para llevar a cabo el proceso de **Extraer**, **Transformar y Compartir** los datos, pág. 42
- **Licencias para la compartición o apertura de datos**, pág. 68
- **Legislación** aplicable a la apertura o compartición de datos, pág. 83
- Proceso de **Gestión de Datos Sensibles**, pág. 52

3. Publicación de los datos

La publicación de datos se realiza en canales de compartición o publicación establecidos. La decisión acerca de qué canal elegir se recoge en la fase previa de **Preparación Técnica**.

Más información: **Canales de compartición de datos**, pág. 43

4. Mantener los datos y metadatos

Los datos y metadatos pueden cambiar con el tiempo y quedar obsoletos. Por ello, se recomienda definir un proceso de mantenimiento de los datos compartidos o abiertos. Algunos aspectos a tener en cuenta:

- Realizar actualizaciones periódicas e incluir la fecha en los metadatos.
- Comprobar que las URI y URL siguen activas.
- Incluir un canal de contacto para que los agentes reutilizadores puedan dar *feedback*.

Más información: **Mantener datos y metadatos**, pág. 54

Apartado **Puesta en práctica del ciclo de vida de los datos abiertos o compartidos**, pág. 44

Siguiente fase

Midiendo el éxito: monitorización de la iniciativa de compartición o apertura de datos

5. MIDIENDO EL ÉXITO: MONITORIZACIÓN DE LA INICIATIVA DE COMPARTICIÓN O APERTURA DE DATOS

Objetivo de esta fase

Realizar un seguimiento de la iniciativa en su totalidad para comprobar su nivel de éxito.

Situación de partida de la empresa

La empresa ya tiene en marcha un proyecto de compartición o apertura de datos y debe implementar un método para poder realizar un seguimiento.

Principales acciones a realizar para la monitorización de la iniciativa:

Contrastar el compromiso de las organizaciones y agentes reutilizadores

El compromiso de los agentes reutilizadores es esencial para el éxito de cualquier iniciativa de compartición o apertura de datos.
El modelo 5 Estrellas de Tim Davies es una herramienta útil para analizar el nivel de interacción que ofrecen los datos.

Medir el éxito de la iniciativa y el nivel de mejora continua

Existen varias medidas para evaluar el desempeño de la iniciativa:

- Contabilizar el número de descargas de los datasets
- Contabilizar el número de accesos al dato.
- Utilizar métricas cualitativas para evaluar si los datasets son útiles.

PIWIK es una herramienta libre para analizar de forma cualitativa la preparación del dato en función de las necesidades de los agentes reutilizadores.

Información relacionada:
- **Métodos para determinar la calidad del dataset**, pág. 53
- **Mantener datos y metadatos**, pág. 54

Apartado **Monitorización de la iniciativa de compartición o apertura de datos**, pág. 54

Aspectos relevantes en una iniciativa de compartición o apertura de datos

La guía incluye diferentes cuestiones clave a tener en cuenta durante el proceso de definición y posterior puesta en marcha de una iniciativa de compartición o apertura de datos. Los más relevantes son:

- Modelos de relación o negocio para la compartición o apertura de los datos.
- Modelos de financiación.
- Marco legal y principios que guían los modelos de relación o negocio.
- Licencias para la compartición o apertura de datos.
- Roles, habilidades, desarrollo y gestión del conocimiento.

Modelos de Relación (o negocio)

La organización puede compartir o abrir sus datos siguiendo diferentes modelos de relación o de negocio en función del objetivo estratégico de la iniciativa: apertura de nuevas líneas de negocio, monetización de los datos, búsqueda de nuevas colaboraciones, u otros.

Se establecen dos modelos, B2B (*Business to Business*) o B2G (*Business to Government*) y diferentes modalidades. Algunas de ellas son:

- **Datos Abiertos/Compartidos**: Los datos son puestos a disposición de un rango abierto de agentes reutilizadores.
  Ejemplos:
  BBVA Valora, pág. 56
  Primafrio, pág. 107

- **Monetización de datos** (B2B o B2G): Enfoque unilateral en el que las empresas tienen ingresos adicionales gracias a los datos que comparten.
  Ejemplo: Euskaltel, pág. 109

- **Marketplace** (B2B): Proveedores de datos y agentes reutilizadores se reúnen en una plataforma online a través de intermediarios de confianza.

- **Plataforma Industrial** (B2B): Los datos se comparten en un entorno cerrado y seguro, entre un grupo restringido de empresas.
  Ejemplo: ITI, pág. 115

- **Facilitadores Tecnológicos** (B2B): Habilitadores técnicos que actúan como tercera parte ofreciendo una solución para el intercambio de datos o un servicio de consultoría.

- **Donación de datos** (B2G): Entendida como una forma de responsabilidad social corporativa.
  Ejemplo: Telefónica – *Big data for social good*, pág. 111

- **Acuerdos de colaboración** (B2G): Acuerdos entre organismos públicos y empresas para el intercambio mutuo de datos.

Información sobre los **Modelos de Relación**, pág. 56
Ejemplos de casos de apertura y compartición de datos, pág. 107

Modelos de Financiación

La organización debe ser capaz de cuantificar el valor económico de la información y los costes asociados a la iniciativa de compartición o apertura de datos, tanto en infraestructura como en personal.

Para ello, son destacables varias prácticas a tener en cuenta:

- **Asegurar la rentabilidad de la iniciativa**, analizando el coste/beneficio e identificando los impactos tangibles e intangibles, como pueden ser los ingresos por nuevos productos o servicios, el ahorro gracias a nuevos procesos o la mejora en la imagen de marca de la empresa.

- Contar con un **método para establecer el valor de los datasets** y establecer tarifas en caso de monetización de los datos. Existen diferentes criterios a tener en cuenta que pueden utilizarse para cuantificar el valor de los datasets y calcular tarifas.

Información sobre **Modelos de Financiación**, pág. 60

Marco legal de los acuerdos contractuales

Normalmente, la compartición de datos se implementa a partir de un acuerdo contractual establecido previamente. En la preparación de este tipo de acuerdos se deben tener en cuenta diferentes aspectos en función del modelo de relación.

En **modelos B2B**:
Algunas consideraciones a la hora de preparar los acuerdos:

- ¿Qué datos estarán disponibles?
- ¿Quién podrá acceder y reutilizar los datos?
- ¿Qué podrá hacer el agente reutilizador con los datos?
- ¿Qué medios técnicos se utilizarán para acceder a los datos?
- ¿Qué datos es necesario proteger?
- ¿Cuál será la duración del contrato?

En **modelos B2G**:
Algunas consideraciones específicas a tener en cuenta a la hora de preparar los acuerdos con entidades públicas:

- ¿Cuál es el propósito de interés público?
- ¿Cuáles son los objetivos y limitaciones relacionadas con la compartición de datos?
- ¿Cuáles son las condiciones para la implementación?
- ¿Cuál es la compensación por la compartición de datos?

Información sobre:
- **Marco Legal**, pág. 64
- **Principios que guían los modelos de relación**, pág. 82

Licencias para la compartición o apertura de datos

Todos los datos compartidos o abiertos deben incluir información explícita sobre sus condiciones de uso. Existen diferentes tipos de licencias que permiten establecer los requerimientos, limitaciones de uso y ámbitos de aplicación del acuerdo.

Información sobre **Licencias para la compartición o apertura de datos**, pág. 68

Roles, habilidades, desarrollo y gestión del talento

Roles

Dentro de una iniciativa de compartición o apertura de datos, existen diferentes roles necesarios para su puesta en marcha y posterior mantenimiento, incluyendo tanto perfiles estratégicos como técnicos. Los principales son:

- **Sponsor** o patrocinador/a: responsable de establecer la estrategia de la organización y supervisar su implementación.
- **Gestor/a de datos**: responsable de implementar la estrategia de compartición de datos.
- **Implementador/a**: responsable de la implementación de los requisitos técnicos.
- **Propietario/a de los datos**: responsable de decidir sobre el acceso y uso de cada base de datos.

Habilidades y desarrollo de conocimiento

La organización debe asegurar que el personal tiene la formación suficiente y el soporte adecuado para llevar a cabo sus responsabilidades en relación a la gestión de los datos. Si es necesario, la organización deberá ofrecer formación que permita desarrollar estas habilidades.

Información sobre **Roles, habilidades, desarrollo y gestión del talento**, pág. 71

1. INTRODUCCIÓN

Los avances en las tecnologías digitales han hecho que actualmente todo tipo de organizaciones generen de manera permanente y exponencial grandes cantidades de datos. Se trata de información procedente tanto de los sistemas internos de la organización como de su interacción con usuarios, proveedores y clientes. En el caso de las empresas, además, la expansión de las tecnologías como el Internet de las Cosas (IoT) y la Inteligencia Artificial (IA) aplicada a los procesos productivos hacen que buena parte de esos datos se generen de manera automática.

La recopilación y análisis de los datos de forma interna implica importantes beneficios para las organizaciones, como la optimización de los procesos y la mejora en la toma de decisiones, entre otros aspectos. Pero la compartición o apertura de estos datos a terceros también ofrece grandes oportunidades para el crecimiento económico, la competitividad y la innovación.

En el caso de las Administraciones Públicas, la apertura de datos tiene ya un largo recorrido y ha demostrado claramente los beneficios que aporta: según la Comisión Europea, el mercado directo de los datos abiertos generará cerca de 325.000 millones de euros en el periodo 2016-2020, así como 100.000[1] empleos[1]. En el caso del sector privado, la compartición y apertura de datos es una práctica menos extendida, aunque cada vez más compañías están viendo sus ventajas. Esta guía tiene como objetivo revisar las ventajas de compartir datos en entornos empresariales, así como recoger una serie de recomendaciones y buenas prácticas para potenciar la reutilización de datos privados.

[1] Comisión Europea (2015): Creating value through Open Data

1.1 ¿Por qué debería compartir o abrir los datos de mi empresa?

Los principales beneficios que aporta la compartición y/o apertura de datos son la oportunidad de generar nuevas líneas de negocio dentro de la empresa y mejorar la eficiencia interna. Existen además otras razones para compartir o abrir los datos, ya sea en entornos B2B (con otras empresas) o B2G (con administraciones públicas y universidades u organismos de investigación). Estos pueden ser la monetización de los datos como vía para generar beneficios adicionales, la apertura de colaboraciones con otras organizaciones, el posicionamiento estratégico de la compañía o el apoyo a la innovación.

> *Caso de éxito: BBVA Valora*
>
> *El proyecto BBVA Valora ha creado un nuevo servicio online basado en datos que ofrece información útil antes de comprar o alquilar una vivienda y permite gestionar en un único lugar toda la información relacionada con la hipoteca y los gastos del hogar.*

Los datos por sí solos no generan valor, y es necesario fomentar un ecosistema que permita transformar datos en información útil y nuevas oportunidades de negocio.

1.2 ¿Cuál es la diferencia entre la compartición y la apertura de datos?

La apertura o compartición de datos en el ámbito empresarial significa la puesta a disposición de los datos generados de forma interna por la compañía para que puedan reutilizarlos terceros, como otras empresas u organismos públicos. Sin embargo, existen diferencias sustanciales entre ambos conceptos:

> *Caso de éxito: Primafrio*
>
> *Empresa especializada en el transporte frigorífico terrestre. La compañía ofrece a sus clientes de manera gratuita datos relativos a la localización y el posicionamiento de su flota. Como resultado, ha conseguido optimizar el tracking de los pedidos y una mayor agilidad en los procesos de facturación.*

Apertura de datos: Consiste en ofrecer los datos de forma "abierta", mediante licencias que permiten la reutilización de los datos para que cualquiera pueda acceder a ellos, utilizarlos y compartirlos bajo unas restricciones mínimas. Estas limitaciones incluyen el cumplimiento de la legislación en materia de protección de datos y el reglamento relativo a la reutilización de datos abiertos[2]. Asimismo los datos abiertos (*Open Data*, en inglés) implican la exposición de dichos datos utilizando estándares y formatos abiertos, que facilitan su reutilización.

Compartición de datos: Consiste en un proceso a través del cual una empresa pone sus datos a disposición de otras, que no son ni competidoras en el mercado donde la empresa opera, ni subcontratistas, y están interesadas en usar los datos para sus propios propósitos de negocio[3].

En la compartición de datos, las empresas proveedoras tienen un mayor control sobre los datos que ponen a disposición de terceros. Estas iniciativas de compartición pueden adoptar diferentes formas, con modelos unilaterales o colaborativos y donde los datos se ofrecen a cambio de tarifas, la provisión de servicios o de manera gratuita.

> *Caso de éxito: ITI*
>
> *El Instituto Tecnológico de Informática, participa en el proyecto europeo Transforming Transport. La iniciativa consiste en una plataforma industrial online donde las empresas participantes comparten sus datos en el ámbito del transporte y la logística, con el objetivo de lograr una mayor eficiencia de los procesos logísticos.*

[2] Real Decreto 1495/2011 sobre la reutilización de la información del sector público, en la legislación vigente española. Directive (EU) 2019/1024 of the European Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information, en la nueva legislación europea.

[3] Comisión Europea (2016): Study on data sharing between companies in Europe

Puesto que es un concepto aún poco extendido en el ámbito empresarial, es importante tener en cuenta que:

- Iniciar un proceso para compartir datos no implica necesariamente que la empresa ofrezca el acceso a todos sus *datasets*. Qué datos comparte es una decisión que debe tomar la compañía en función de su estrategia de negocio.

- La compartición no significa que se ofrezcan los datos de manera gratuita. Existen diversos modelos de negocio que incluyen el pago por acceso.

- Las empresas proveedoras de los datos son las que deciden con quién compartirán sus datos, en qué términos, con qué condiciones de uso y el marco temporal en el que podrán ser utilizados.

1.3 ¿Por dónde empezar?

Como ya se ha apuntado, la apertura y la compartición de datos en el entorno empresarial ofrecen importantes oportunidades, tanto para grandes compañías con experiencia previa en este ámbito, como para pymes no acostumbradas a manejar grandes cantidades de datos. Además, con la expansión de la llamada Industria 4.0, muchas empresas sin experiencia previa en la gestión de *Big Data* han empezado a generar grandes volúmenes de datos y ha aumentado su interés en buscar nuevas oportunidades para ponerlos en valor.

> *Caso de éxito: ArcelorMittal*
>
> *La empresa siderúrgica, a través de su centro I+D en Euskadi, y Tecnalia pusieron en marcha en 2018 un datatón dirigido a resolver un problema existente sus las acerías. La compañía puso a disposición una parte de sus datos durante un tiempo limitado para que grupos formados por personas investigadoras dieran con soluciones innovadoras.*

Es importante tener en cuenta que los procesos de apertura o compartición de datos implican un cambio tanto a nivel de infraestructuras técnicas como en los procesos organizacionales y, en definitiva, en la cultura de la empresa. Por ello, para aquellas compañías sin experiencia previa que quieran iniciar este camino, es aconsejable empezar experimentando mediante pequeños proyectos, abriendo una parte de sus datos a facilitadores o plataformas tecnológicas para valorizar su potencial. En esta guía se ofrecen diferentes alternativas para llevar a cabo estos procesos.

> *Caso de éxito: Euskaltel*
>
> *A partir de los datos procedentes de las conexiones a las antenas de telefonía, la compañía ha desarrollado una herramienta de análisis de Big Data. Como resultado, obtiene informes que ofrece a empresas y administraciones públicas.*

1.4 ¿Qué contiene esta guía y a quién está dirigida?

La *Guía para la apertura y compartición de datos en el entorno empresarial* pretende ser una herramienta para acompañar a todas aquellas compañías interesadas en abrir o compartir sus datos, tengan o no experiencia previa en este ámbito. Se trata pues de un documento informativo, no académico, que pretende acercar la potencialidad del *Open data* al sector privado y empresarial.

> *Caso de éxito: SUEZ*
>
> *Durante un fin de semana, la empresa puso sus datos a disposición de 20 equipos para responder a varios retos relacionados con el agua, trabajando de forma colaborativa. Los resultados ayudaron a la empresa a diseñar nuevos productos y servicios.*

Existen ya numerosos ejemplos de compañías que han puesto en marcha iniciativas para compartir sus datos con terceros, así como *startups* que han encontrado un nuevo nicho de negocio en el ámbito de la compartición de datos. Dichos ejemplos muestran el potencial que ofrecen estos procesos, tanto para las empresas proveedoras de datos como para las reutilizadoras.

> *Caso de éxito: EDP*
>
> *La compañía lleva a cabo datatones online donde lanza retos a resolver, poniendo a disposición sus datasets como materia prima para desarrollar nuevas soluciones. También ofrece sus datos a instituciones para proyectos de investigación. Modalidad: Datatón y B2G (universidades y centros tecnológicos).*

A lo largo de estas páginas, se incluyen una serie de recomendaciones y buenas prácticas y se describen los pasos a seguir para iniciar un proceso de apertura o compartición de datos:

a) Experimentación y concienciación.

b) Cómo definir una buena estrategia de datos abiertos o compartidos.

c) Preparación técnica, cómo orientar la empresa hacia la compartición o apertura de datos.

d) Poner en práctica el ciclo de vida de los datos: recoger, preparar, compartir o publicar y mantener.

e) Monitorización de la iniciativa, análisis del nivel de éxito y mejora continua.

a) Experimentación y concienciación. Tiene como propósito conocer el potencial de los datos de una organización. Es la fase inicial recomendada para una organización que no es consciente del potencial de sus datos.

Si la organización no dispone de conocimiento interno en analítica de datos *(Big Data)* lo más apropiado es contar con el apoyo de personal externo que disponga de este conocimiento y le asesore sobre el valor y potencial de los datos manejados. El apoyo externo puede provenir de diversas fuentes como **Facilitadores Tecnológicos**, organización de **Datatones** abiertos o adhesión a **marketplaces/ plataformas industriales** que ofrecen este tipo de servicios.

b) Cómo definir una buena estrategia de datos abiertos o compartidos. El propósito es definir, establecer y revisar la política y estrategia de la organización respecto a los datos abiertos/compartidos partiendo de la situación real inicial de la organización. Dicha estrategia afecta tanto a la compartición/apertura como a la reutilización de datos de terceros y es un proceso de transformación organizacional que debiera ser parte de una estrategia más amplia de digitalización. Algunas buenas prácticas en este sentido son:

- *Conocer la situación de partida de la organización respecto a los datos abiertos/compartidos.*
- *Conocer la situación objetivo/destino de la organización respecto a los datos abiertos/compartidos.*
- *Definir e implantar la estrategia adecuada para alcanzar los objetivos.* Derivar y establecer el plan de mejora para la compartición/apertura de datos; Definir el modelo de gobernanza de datos adecuado; Definir los procesos organizacionales para derivar y compartir/publicar los *datasets* y para hacer seguimiento de la reutilización de los mismos por terceros y, si fuera necesario, desplegar el proceso de gestión del cambio asociado.
- *Definir y mantener el inventario de datos de la organización.* Incluir en el inventario tanto los *datasets* internos como los *datasets* provenientes de otras organizaciones que se reutilizan.
- *Definir la política de datos abiertos/compartidos, comunicarla y hacerla pública.*

c) Preparación técnica, cómo orientar la empresa hacia la compartición o apertura de datos. El propósito es ayudar a la organización a tomar las decisiones adecuadas sobre la gestión de datos y cambios organizacionales asociados, así como sobre la infraestructura tecnológica de la que se debe dotar para extraer y tratar los datos antes de su apertura/compartición según sea la ambición y estrategia de la organización. Algunos aspectos relevantes que considerar:

- *Gestión de datos*. Abarca todos los *datasets* de la organización (los generados internamente, así como *dataset* de terceros). Existen distintos tipos teóricos de gestión de datos (gestión de datos descentralizada, gestión de datos centralizada y gestión de datos completamente descentralizada – esta última es la ideal para la compartición/apertura de datos) y cada organización debe adoptar la gestión de datos más apropiada según su casuística (realidad y ambición de la iniciativa de orientación al dato).

- *Extraer, Transformar y Compartir datos (ETCD)*. Consiste en definir las especificaciones técnicas para implantar un sistema/plataforma que permita el flujo de datos y la transformación de datos a llevar a cabo en la organización para que estos datos estén listos para compartir o abrir. El proceso de ETCD con la tecnología implementada es el proceso central que se ejecutará de manera recurrente para la obtención de los *datasets* de uso interno o para compartir/abrir a terceros.

- *Canales de compartición de datos*. Son los distintos mecanismos técnicos habilitados para la distribución de los datos de la organización a terceros[4]: Web, API, Portal de Datos, Marketplace, Plataforma Industrial o Facilitadores técnicos. Una organización puede distribuir datos a través de más de un canal, variar los canales de distribución en el tiempo según el propósito de la compartición/apertura de datos, etc.

- *Búsqueda de datos*. Existen mecanismos de búsqueda BÁSICOS (ej. CMS), AVANZADOS (usando SPARQL) o buscadores universales específicos de *datasets* (ej. https:/toolbox.google.com/datasearch). Dependiendo de la estrategia de compartición/apertura de datos convendrá utilizar unos u otros.

- *Otros aspectos técnicos que considerar* en una iniciativa de compartición/apertura: ¿dónde almacenar los datos?; ¿dónde compartirlos?; ¿quién administrará los canales de distribución?; ¿qué servicio ofrecerá la organización?, etc.

[4] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0125&rid=2

d) Poner en práctica el ciclo de vida de los datos: recoger, preparar, compartir o publicar y mantener. Describe los pasos a seguir para la generación de cada dataset a compartir/publicar. Se abordan los aspectos relevantes delo: calidad del dato, preparación legal, gestión de datos sensibles, etc.

- *Recogida de datos*. Para las organizaciones sin experiencia en la compartición/apertura de datos se recomienda seguir un proceso para la identificación y priorización de los datos a compartir/publicar.

    1. *Preparación de datos*. Preparación legal del dato para la compartición/apertura. Este aspecto es clave para la compartición/publicación. Si el dato no cumple con la normativa vigente, y tiene las licencias adecuadas, no se puede compartir.

    2. *Calidad del dato*:

    - Contenido: ¿El dato está completo?; ¿Están los datos limpios?; ¿Son los datos suficientemente precisos para el propósito en cuestión? Se sugiere verificar campos vacíos, valores erróneos, dobles entradas, información privada sensible; etc.

    - Prontitud: ¿con qué frecuencia se actualizan?; ¿Son datos en tiempo real?

    - Consistencia en la presentación de datos para facilitar su reutilización por terceros.

    - La calidad del dato se puede medir por su nivel de cumplimiento de los principios FAIR (*Findable, Accessible, Interoperable, Reusable)* al incluir los metadatos adecuados y los identificadores persistentes para su identificación y reutilización.

3. Determinar el *formato de publicación del dataset* y facilitar el acceso al *dataset*. Para ello será necesario añadir metadatos, utilizar ontologías estándar, etc.

4. Definir un proceso para el *tratamiento de datos sensibles* a compartir/publicar.

- *Publicación del dato*. A través de los canales de compartición de datos identificados.

- *Mantener datos y metadatos*. Definir un proceso para mantenerlos al día e identificar acciones de mejora en el ciclo de vida del dato basadas en el *feedback* de los reutilizadores.

e) Monitorización de la iniciativa, análisis de nivel de éxito y mejora continua. El propósito es auditar tanto el ciclo de vida del *dataset* desde su compartición/publicación (trazabilidad del dato y reutilización en los distintos escenarios), como hacer el seguimiento de la iniciativa de compartición/apertura de datos de la organización en su conjunto (desde la política y estrategia, procesos establecidos, etc.). Se recomienda basar las decisiones de mejora en métricas cualitativas y cuantitativas y contrastar además el compromiso de los reutilizadores utilizando métodos formales como el de Tim Davies[5]).

1.5 Aspectos relevantes de la apertura/compartición de datos

Para que una iniciativa de datos compartidos/abiertos sea sostenible hay que tener en cuenta, al menos, estos **aspectos relevantes**:

- Tiene que sustentarse en un **modelo de relación o negocio** viable.

- Tiene que tener un **modelo de financiación** asociado que sea rentable.

- Se debe respetar escrupulosamente **el marco legal** existente.

- Se necesita cambiar la cultura organizacional y disponer de **personal cualificado para la gestión y compartición/apertura de los datos**.

La mayoría de marketplaces o plataformas industriales ofrecen soluciones tecnológicas y servicios a las organizaciones, principalmente pequeñas y medianas, para abordar estos aspectos relevantes de manera sencilla. Es decir, ofrecen a sus clientes la posibilidad de compartir plataformas tecnológicas y gastos de inversión comunes de este tipo de iniciativas.

1.6 ¿Quién ha realizado esta guía?

La redacción de esta guía se enmarca en la actividad del Grupo de Trabajo de la Fundación Cotec de Datos Abiertos, que se ha desarrollado entre septiembre de 2018 y junio de 2019. El desarrollo de este Grupo de Trabajo ha sido coordinado por Cristina Oyón y Amaia Martínez del Departamento de Iniciativas Estratégicas de SPRI, la Agencia Vasca de Desarrollo Empresarial del Gobierno Vasco.

[5] Time Davies' five stars Open Data engagement, 2012. https://www.timdavies.org.uk/2012/01/21/5-stars-of-open-data-engagement/

En la redacción de la guía han colaborado además Ana Ayerbe, Elixabete Ostolaza e Izaskun Santamaría de la corporación tecnológica Tecnalia, Diego Moñux y David Gago de la consultora Science & Innovation Link Office SILO, Alberto Abella como experto de Cotec en datos abiertos, Irene Lopez de Vallejo de Ocean Protocol, Eva Méndez de la Alianza 4 Universidades (Universidad Carlos III de Madrid) como experta en Open Science y datos de investigación, y Álvaro Ramos de Clarke & Modet en el asesoramiento en temas legales.

La guía recoge las contribuciones de las entidades participantes en el Grupo de Trabajo de Cotec , recogidas durante las 4 sesiones de trabajo mantenidas y a través del cuestionario y entrevistas realizadas, además de contribuciones de otras entidades que han contribuido con sus experiencias (Arcelor Mittal y Euskaltel). Miembros designados:

1. Agencia de Desarrollo Económico de la Rioja (ADER): Almudena Larrea, Responsable de la Ud. de Estadística y Justificación de Fondos; y Javier Solozábal, Responsable del Programa de Planificación y Seguimiento.
2. Alberto Abellá: Experto de los 100 de Cotec.
3. Alianza 4 Universidades: Prof. Eva Méndez, Vicerrectora Adjunta para Open Science de UC3M.
4. Banco de Santander: Lara de Mesa, Directora de Public Policy.
5. Bankia: Ignacio Cea, Director Corporativo de Innovación.
6. BBVA: Iván Moreno, CTO de New Digital Business; y Ana Isabel Segovia, Digital Regulation, Supervisors, Regulation and Compliance Department.
7. Clarke Modet: Álvaro Ramos, Delegado de Protección de Datos.
8. Consejería de Economía, Conocimiento, Empresa y Universidad (CECEU) Junta de Andalucía: Manuel Ortigosa, Secretario General de Empresa, Innovación y Emprendimiento.
9. Ecoembalajes España: Manuel de Arcocha, Director de Tecnología; y Daniel Moradiellos, Oficina Estratégica.
10. EDP España: Luis Manuel Santos, Jefe de Innovación, Dirección de Ambiente, Sostenibilidad, Innovación y Calidad; y Ernesto Barroso, Responsable de Empresas Emergentes.
11. F. Iniciativas: Enrique Blas, Partnership Manager.
12. Fundación Vodafone: Luz Olivas, Project manager en la Dirección de Administración Pública; y Carlos Becker, Head del Segmento de Administraciones Públicas.
13. Indra (Business Consulting Minsait): Francisco Javier Ruiz, Partner; y David Fernández, Head of CyberSecurity Business.
14. Instituto Valenciano de Competitividad Empresarial (IVACE): Daniel Sáez, Director de Transferencia de Tecnología – ITI; y Francisco Valverde, Jefe de Proyectos de I+D – ITI (Instituto Tecnológico de Informática).
15. Orange: Juan Vallés, Big Data Product Owner.
16. Primafrío: Adrián Valverde, Responsable de gestión de I+D+i.
17. Suez Advanced Solutions: Catalina Balseiro, Directora de Innovación.
18. Tecnalia: Ana Ayerbe, Dir. Área Negocio ICT; Izaskun Santamaría, Investigadora; y Elixabete Ostolaza, Senior Consulting.
19. Telefónica: Richard Benjamins, Data & AI Ambassador; y Estanislao Fernández, Jefe de proyecto en Telefónica I+D.

[6] Componentes del Grupo de Trabajo de la Fundación COTEC: Fundación COTEC (Adelaida Sacristán y Pilar Caro), SPRI (Cristina Oyón, Amaia Martínez, Javier Diéguez e Iratxe Martín), Agencia de Desarrollo Económico de La Rioja ADER (Almudena Pérez), Alianza 4 Universidades (Eva Méndez), BBVA (Iván Moreno y Ana Isabel Segovia), Clarke Modet (Álvaro Ramos) , Telefónica (Richard Benjamins), Ecoembes (Daniel Moradiellos y Manuel de Arcocha), Indra (Francisco Javier Ruiz y David Fernández), Fundación Vodafone (Luz Olivas), Instituto Valenciano de Competitividad Empresarial IVACE (Daniel Sáez y Francisco Valverde), Junta de Andalucía - Empresa y Comercio- (Manuel Ortigosa), Primafrio (Adrián Valverde), Suez Advanced Solutions (Catalina Balseiro) y EDP (Luis Manuel Santos y Ernesto Barroso).

2. LA APERTURA Y COMPARTICIÓN DE DATOS EN EL ÁMBITO EMPRESARIAL

En la era de la transformación digital, los datos juegan un papel fundamental, son el *"new oil"* como aluden políticos, técnicos, científicos y estrategas. Los distintos gobiernos y la mayoría de las grandes organizaciones privadas, cuyo negocio implica el uso intensivo de datos (ej.: banca, seguros, operadores telefónicos, etc.) son conscientes de su potencial e impacto, tanto en la mejora del rendimiento interno de las organizaciones, como en la economía y en la sociedad en general. Con la compartición y/o apertura de los datos que dispone una organización con terceros se abre un amplio abanico de posibilidades tanto para la propia organización, como para terceros.

Actualmente la **innovación basada en datos** se considera uno de los facilitadores clave para el crecimiento de los países y la generación de nuevos empleos en Europa. La importancia del dato producido y procesado *online*, el continuo crecimiento de datos generados por objetos conectados a Internet de las cosas (IoT), la disponibilidad creciente de herramientas de análisis de *Big Data* y la disponibilidad de cada vez mayor número de aplicaciones de Inteligencia Artificial (IA) son facilitadores técnicos clave en la transformación digital. La naturaleza no competitiva del dato en sí, hace posible que el mismo dato se utilice como base para toda una gama de productos y servicios o nuevos métodos de producción. Esto nos lleva a la conclusión de que puede ser beneficioso para las compañías compartir más datos con otras, de tal forma que, los datos de origen se puedan explotar al máximo.

"Sin embargo, los datos por sí solos no generan valor, y lo que se necesita es fomentar un ecosistema[7] que provea el contexto que permita transformar datos en información[8], y la información en conocimiento.

El sector público ha sido desde el comienzo de la economía de los datos un importante contribuidor, tanto a la hora de definir y adaptar los marcos legales, normativas y recomendaciones para facilitar e impulsar dicha economía, como por su papel para incentivar la puesta en marcha de iniciativas de apertura y compartición de datos en las organizaciones tanto públicas como privadas. El sector privado se encuentra por detrás en este proceso de apertura y compartición de datos, aunque empieza a despertarse cierto interés entre las empresas y han comenzado algunas iniciativas de éxito. No se debe olvidar que las nuevas tecnologías han convertido a muchas organizaciones que tradicionalmente no trabajaban de forma intensiva con datos, en organizaciones *"data-intensive"* y por tanto, candidatas ideales para iniciativas de compartición/apertura de datos.

- Cuando las organizaciones que comparten/abren los datos son organizaciones privadas se habla de compartición/apertura de datos B2B (*Business to Business*) y/o B2G (*Business to Government*), refiriéndose en el primer caso a la compartición/apertura de datos entre organizaciones privadas y el segundo a la compartición/apertura de datos entre los organismos privados y la Administración Pública. Una organización privada puede tener datos compartidos con otras organizaciones privadas (B2B), con las Administraciones Públicas (B2G), o con ambas, según sea su política y estrategia de datos abiertos. En este sentido, en el informe final de la Comisión Europea donde se recoge un estudio sobre las prácticas de compartición de datos

[7] https://datos.gob.es/sites/default/files/actualizacion_informe_iniciativa_sectorial_0.pdf

[8] A very short introduction: Luciano Floridi, Oxford press, Published: 25 February 2010, ISBN: 9780199551378

entre empresas en Europa,[9] se describen 16 casos de éxito de compartición de datos en empresas privadas de distintos sectores, de los que se pueden obtener fuentes de inspiración para la compartición de datos empresariales. En dicho informe, además se llega a las siguientes conclusiones: Alrededor del 40% de las compañías que participaron en el estudio publican y/o reutilizan datos.

- Una proporción considerable de compañías que aún no comparten/publican sus datos espera hacerlo en los próximos 5 años.

- Todas las compañías que comparten/publican y/o reutilizan datos consideran que la compartición/apertura les ha ayudado a mejorar la eficiencia interna y/o ha mejorado sus oportunidades de negocio.

[9] https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

3. ¿QUÉ PUEDE HACER UNA EMPRESA CON SUS DATOS?

Como se ha mencionado, muchas organizaciones que empiezan a manejar grandes volúmenes de datos no son tradicionalmente organizaciones "intensivas en datos" y desconocen las potencialidades de éstos y/o no saben por dónde empezar, al no contar con recursos especializados para valorizar esos datos. Este es el caso de muchas compañías de la denominada industria 4.0. En estos casos se sugiere empezar mediante la experimentación abriendo los datos a facilitadores tecnológicos que analicen cómo valorizar los datos para, tras esta experiencia, decidir los siguientes pasos a dar.

En el caso de empresas de una cierta dimensión que cuentan con gran cantidad de datos, (ej.: bancos, operadoras de telecomunicaciones, compañías eléctricas, etc.), y que además cuentan con recursos especializados para valorizarlos, los datos pueden ser utilizados no sólo para la mejora interna, sino también para sacarles más rendimiento optando por la diversificación del negocio más allá de las actividades centrales de la empresa mediante la explotación y reutilización de esos datos.

El informe final de la Comisión Europea, que ya hemos citado[10], señala dos enfoques posibles para tipificar las iniciativas relacionadas con la compartición/apertura de datos:

- **Enfoque *Top-Down*,** que implica la identificación de datos que sean útiles para la resolución de un problema interno o reto global o local. En estos casos se busca proactivamente a las compañías que poseen los datos para que los compartan/abran. La mayoría de los casos de éxito de compartición de datos entre empresas se ajustan a este tipo.

- **Enfoque *Bottom-up*:** donde la apertura/compartición de datos se realiza con terceros para explorar la potencialidad de la reutilización de estos. Normalmente son compañías que se nutren de datos de sus clientes para desarrollar productos y servicios que no son su negocio principal (se citan como casos de éxito, GEO, Orange y la fundación de Mastercard). Este tipo de iniciativas pueden ser monetizadas o no. Estos casos pueden considerarse, siguiendo a Chesbrough[11], experiencias de Innovación abierta, entendida como la apertura del proceso de innovación al conocimiento externo de la organización. La innovación abierta es un paradigma que asume que las empresas pueden y debe usar ideas externas, así como ideas internas. Los límites entre una empresa y su ecosistema serán cada vez más invisibles, y las innovaciones podrán ser fácilmente transferidas hacia adentro y hacia afuera de la organización.

Del análisis de los casos de estudio del informe de la Comisión Europea sobre la compartición de datos en empresas se pueden extraer también algunos patrones de buenas prácticas:

- Las grandes compañías sin experiencia en la apertura de datos y sin la infraestructura tecnológica adecuada para su explotación se apoyan en facilitadores y/o plataformas tecnológicas para iniciar el camino. El alcance del apoyo puede ir desde la subcontratación de toda la puesta en marcha de la iniciativa a la subcontratación puntual de recursos y/o servicios (caso de éxito de ORANGE en el que subcontrata la nube a AMAZON para almacenar/compartir datos masivos y para el tratamiento de estos antes de su apertura; caso de éxito de DAWEX, plataforma intersectorial o transversal orientada a todo tipo de empresas privadas).

[10] https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

[11] https://www.bbvaopenmind.com/articulos/articuloinnovacion-abierta-innovar-con-exito-en-el-siglo-xxi

- Las compañías sin experiencia previa en apertura de datos, pero conscientes del valor de estos, se pueden iniciar permitiendo el uso de los datos a start-ups, universidades y centros de investigación para que éstas exploren las potencialidades de los datos aplicando técnicas de Big-Data, inteligencia artificial, etc. (caso de éxito de ORANGE).

- Hay compañías maduras desde el punto de vista de la apertura de datos que permiten el acceso de sus datos a centros de investigación y universidades que trabajan en retos de medio-largo plazo (es el caso de la fundación sin ánimo de lucro "*Center for Inclusive Growth*" de Mastercard, que no cobra por la reutilización de datos a los centros de investigación que trabajan en resolver problemas existentes en países en vías de desarrollo, por ejemplo, buscando patrones de una enfermedad en el tercer mundo para identificar las causas, etc.).

¿Por qué las compañías comparten/abren sus datos?

- **Para generar nuevas líneas de negocio y/o nuevas empresas basadas en la reutilización de sus datos,** que pongan en valor el exponencial incremento del volumen de datos que gestionan, catapultado por el uso masivo de componentes *smart* -el Internet de las cosas (IoT)- y de los *smartphones*. Los casos de BBVA, Telefónica, Orange y GEO son ejemplos de ello.

- Para **resolver problemas que las organizaciones no pueden resolver por sí solas**. Un ejemplo es el caso del Instituto URBAN que partiendo de datos de distintas compañías (por ejemplo, la fundación Mastercard) ha podido localizar con precisión las bolsas de pobreza de las distintas zonas geográficas de Estados Unidos.

- **Para contribuir a que otras empresas desarrollen soluciones innovadoras y de mayor valor añadido que por sí solas no son capaces de desarrollar.** Esto es válido tanto en lo relativo a la compartición/apertura de datos sectoriales como transversales. Un ejemplo son las publicaciones que vende GEO a las compañías energéticas que utilizan sus contadores inteligentes (*smart meters)* con los resultados de los análisis de los datos recogidos. GEO tiene firmados acuerdos con las eléctricas que le permiten analizar los datos de sus contadores para determinados propósitos.

- **Las compañías que publican datos abiertos son frecuentemente compañías que vienen de una experiencia previa en la compartición de datos.** La publicación de datos abiertos es un paso más para sacar máximo provecho a los datos de la compañía y requiere un esfuerzo importante por parte de ésta para garantizar todos los requisitos exigibles a los datos abiertos (apertura técnica, legal, etc.).

- **Para cumplir con normativas nacionales respecto a transparencia y rendición de cuentas.** Normalmente este aspecto afecta principalmente a administraciones públicas, aunque existen cada vez más casos, como el de ENEDIS, en el que se opta por una estrategia de apertura de datos para dar respuesta a las obligaciones legales nacionales a este respecto.

Los motivos que guían a una organización hacia la apertura/compartición de datos son diversos, pero todas comparten un componente común: la innovación y generación de valor a partir de los datos. La compartición/apertura de datos puede abrir a las organizaciones la posibilidad de replantearse sus productos y servicios de manera más efectiva para situarse en una mejor posición competitiva. Sin embargo, se requieren inversiones en tecnología y experiencia para utilizar los datos de manera **efectiva**. El despliegue de una gobernanza adecuada de los datos en la organización será un aspecto clave para lograr la máxima eficacia de la iniciativa.

4. FUNDAMENTOS TÉCNICOS DE LA APERTURA/COMPARTICIÓN DE DATOS EN EL ENTORNO EMPRESARIAL

A continuación, se detallan una serie de recomendaciones técnicas y buenas prácticas dirigidas a cualquier organización que desee comenzar una iniciativa de compartición/apertura de datos[12]. Así pues en este apartado se describen cada una de las **fases y procesos** que debería seguir de una iniciativa de compartición y/o apertura de datos:

1. Experimentación y concienciación.
2. Definir una buena estrategia de datos abiertos y compartidos.
3. Preparación técnica.
4. Poner en práctica el ciclo de vida de los datos: recoger, preparar, publicar y mantener.
5. Monitorización de la iniciativa, análisis del nivel de éxito y mejora continua.

El proceso de apertura y compartición de datos en el entorno empresarial ha de apoyarse además en un análisis en profundidad de la que llamaremos **información de soporte** relativa a la:

- Definición del Modelo de Negocio y Financiación.
- Cumplimiento del Marco Legal.
- Infraestructura Tecnológica.
- Desarrollo de habilidades internas (roles, habilidades y conocimiento).

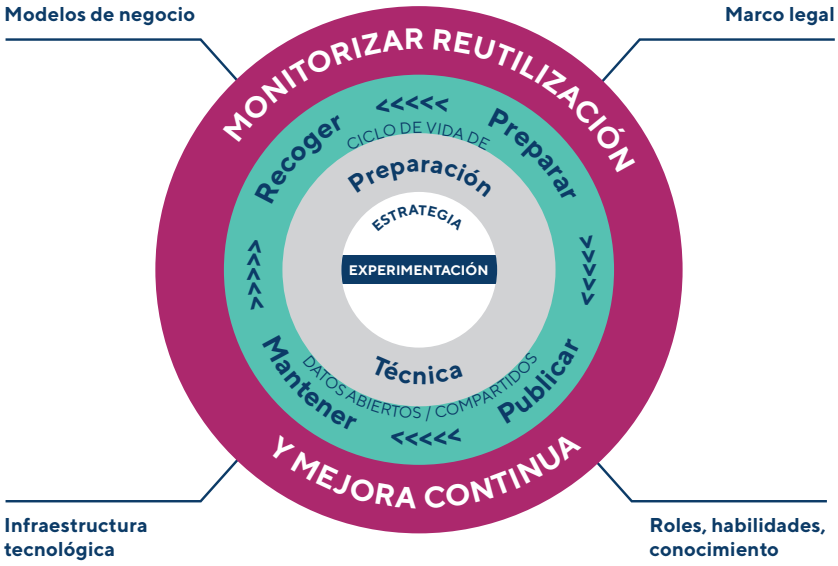


Ilustración 1- Fundamentos Apertura/Compartición de Datos en Entorno Empresarial

[12] Los anexos IV y V, pueden ayudar a un mejor entendimiento y puesta en práctica de la complejidad técnica que puede tener la apertura de datos o su compartición por parte de las empresas. El anexo IV recoge algunos conceptos importantes (linked data, metadatos, etc.), mientras que el anexo V recoge un listado de software y aplicaciones de software libre con las que se puede trabajar fácilmente en el ámbito de Open Data y cuyo conocimiento puede ayudar a las empresas que comiencen con estas prácticas.

La ilustración 1 representa gráficamente el ciclo de vida de los datos abiertos/compartidos, se detalla a continuación cada una de las fases y procesos implícitos en dicho ciclo de vida.

4.1 Experimentación y concienciación

La **experimentación** es la fase inicial a la que se enfrenta una organización cuando no se es consciente del potencial de la información que maneja. Una organización alejada de la apertura/compartición de datos cuyo negocio principal está centrado en otros ámbitos de actuación, como pueden ser las empresas industriales, debe comenzar por conocer el potencial de sus datos.

Podemos encontrar organizaciones que se encuentran en las siguientes situaciones:

- **La organización dispone de conocimiento interno experto en analítica de datos**. En esta situación puede ser suficiente con involucrar a este personal en el proceso de análisis de los datos existentes, para ayudar a la dirección a identificar el potencial que existe en los datos manejados y a tomar decisiones sobre el modelo de negocio con el que arrancar.

  La organización no dispone de conocimiento interno en analítica de datos. En esta situación lo más apropiado es apoyarse en personal externo que disponga de conocimiento técnico en analítica de datos y que pueda **asesorar en cuanto al valor y potencial de los datos manejados**. El apoyo externo puede provenir de diversas fuentes:

  - **Facilitadores tecnológicos**: Centros tecnológicos, universidades técnicas, consultoras especializadas en análisis de datos, start-ups especializadas, etc., con las que se extienden acuerdos de colaboración en los que existe un beneficio mutuo, ya que por una parte la organización que cede sus datos obtiene un análisis del potencial existente a partir de los datos que gestiona, y por otra parte la organización que realiza el análisis obtiene bien una remuneración a cambio del análisis realizado, bien conocimiento práctico sobre análisis de datos reales o bien testear algoritmos de analítica de datos desarrollados. Normalmente, los acuerdos se extienden con potenciales socios tecnológicos con los que posteriormente se podrá mantener una relación futura. Hay que tener en cuenta que los acuerdos de colaboración entre ambas partes no siempre van asociados a remuneración económica en esta primera fase de experimentación.

    En esta situación es habitual que las organizaciones pongan a disposición de los socios tecnológicos un subconjunto reducido de los datos que manejan a modo de muestra. Esta muestra debe ser lo suficientemente representativa para poder obtener conclusiones a partir de la misma. El socio tecnológico apoyará a la organización a extraer la muestra más representativa en caso de ser necesario.

    Los acuerdos de colaboración en este primer momento incluirán acuerdos de confidencialidad que den garantías a la organización proveedora de los datos.

Nombre de la empresa:	**ArcelorMittal**
Dirección web (empresa o iniciativa):	https://spain.arcelormittal.com/
Descripción de la iniciativa:	▪ Optimizar el rendimiento férrico de cada tipo de chatarra tomando como referencia datos reales del proceso de fusión de acería y teniendo en cuenta un conjunto de factores que influyen en el rendimiento. ▪ Predecir el contenido de cobre final en acero líquido en un momento puntual en el tiempo para cada tipo de chatarra que se mezcla.
Tipo de datos compartidos/abiertos:	▪ Datos procedentes del uso de los bloques de chatarra de origen (hasta 16), del proceso de fusión de acería y de los resultados: desde 07/2017 a 04/2018, 4.000 muestras consecutivas.
Razones para la apertura/compartición de datos:	▪ Explorar nuevas soluciones analíticas para la correlación del proceso de producción. ▪ Apoyar la innovación abierta y colaborativa.
Monetización del Negocio:	▪ No hay monetización directa.

Tabla 1 – ArcelorMittal: Ejemplo Datatón

[13] https://cordis.europa.eu/project/rcn/191761/factsheet/es

- ***Datatones* abiertos:** El apoyo en este caso implica poner a disposición de terceros un conjunto seleccionado de datos en un entorno abierto y por un periodo limitado, para que los participantes puedan aplicar algoritmia con objeto de intentar sacar el mayor valor a los datos. Estas actividades suelen ser facilitadas por organismos orientados a la innovación del ecosistema (local, regional, nacional o internacional) y acceso demostrado a las comunidades de Inteligencia Artificial y *Data Science* que pueden consumir y experimentar con estos datos. En este tipo de iniciativas es recomendable incentivar la participación de las organizaciones (centros de investigación, universidades, empresas, etc.), así como publicar las condiciones relacionadas con la explotación de los resultados obtenidos en el *datatón*. Un ejemplo interesante de este tipo de iniciativas es el proyecto APPS FOR EUROPE, que organizó veinte competiciones locales y globales en 10 países europeos entre 2013-2015 para identificar negocios viables con la creación de apps basadas en datos abiertos[13].

Nombre de la empresa:	**SUEZ**
Dirección web (empresa o iniciativa):	https://www.suez.es/es-es http://www.hackath2on.es
Descripción de la iniciativa:	▪ Durante un fin de semana, 20 equipos se reunieron para abordar, desde una óptica colaborativa, tres retos relacionados con el agua, en concreto: 1) la excelencia en la experiencia de cliente; 2) la promoción del valor social del agua; y 3) nuevos servicios a la administración. La iniciativa ha sido de utilidad para la compañía, ya que su celebración ha ayudado a la empresa a diseñar nuevos productos y servicios, y ha animado la puesta en funcionamiento de otras iniciativas de datos abiertos en el futuro inmediato.
Tipo de datos compartidos/abiertos:	▪ Datos de transacciones (clientes, proveedores). ▪ Datos internos operativos (calidad, producción, etc.).
Razones para la apertura/compartición de datos:	▪ Diversificar la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del mercado y del cliente. ▪ Crear nuevos modelos de negocio. ▪ Propósito social, de bien común público.

Tabla 2 – SUEZ: Ejemplo Datatón

[14] https://consulting.datary.io/casos-de-uso

- ***Marketplace*/Plataforma Industrial**: En este caso la experiencia consiste en poner a disposición de terceros en un entorno abierto como puede ser un *marketplace* o plataforma industrial una muestra de los datos que maneja la organización con el objetivo de identificar igualmente la potencialidad de la información y de los datos, siguiendo las normas marcadas por la plataforma de compartición que se utilice. A los agentes que van a explorar/analizar datos se les exigirá firmar un acuerdo de confidencialidad sobre los mismos. En ocasiones, el propio *marketplace* dispone de servicios para explorar la potencialidad de los datos. Un ejemplo de ello son los casos de uso de la plataforma Datary[14].

El ejercicio de experimentación debe ayudar a la dirección de la organización a tomar decisiones en cuanto a si la iniciativa de apertura/compartición de datos puede estar alineada con su estrategia de negocio, en qué medida, a través de qué modelo de negocio e incluso qué infraestructura tecnológica utilizar. Aunque en esta fase previa, las conclusiones a las que se lleguen serán incipientes y requerirán de un análisis posterior más refinado, que se realizará durante la fase de definición de la estrategia.

Tanto si la organización opta por organizar un *datatón* o por poner sus datos disponibles a terceros en un *marketplace*, es fundamental que previamente cuente con asesoría jurídica que le garantice que los datos que se van a poner a disposición de terceros cumplen la legislación en vigor.

Una vez realizada la actividad de experimentación se sugiere realizar actividades de **concienciación** y sensibilización a la práctica de los datos abiertos en la organización. En un primer lugar, dirigidas a los niveles directivos, ya que es imprescindible conocer lo que supone la apertura/compartición de datos para poder tomar decisiones adecuadas orientadas a definir una estrategia, política y metodología alineadas a las necesidades de cada organización.

4.2. Definir una buena estrategia de datos abiertos o compartidos

El propósito de este proceso es establecer y revisar la política y estrategia de la organización respecto a los datos abiertos/compartidos partiendo de la situación real inicial de la organización. Dicha estrategia afecta tanto a la compartición/apertura, como a la reutilización de datos de terceros y es un proceso de transformación organizacional que debería ser parte de una estrategia más amplia de digitalización. A continuación, se detallan algunas prácticas que se pueden tener en cuenta en esta fase:

- **Conocer la situación de partida de la organización respecto a los datos abiertos/compartidos**.
    - Identificar qué hace cada unidad/departamento/división de la organización: ¿Qué unidades recogen datos, utilizan datos, producen datos? ¿Qué tipo de datos recogen/utilizan/producen? ¿En qué formato?
    - Identificar los representantes o personas clave de diferentes departamentos de la organización que permitan hacer un diagnóstico de la situación de partida respecto a los datos.
    - ¿Están los datos centralizados o distribuidos? ¿La gestión de los datos es de forma centralizada o descentralizada? Si los datos se generan en múltiples unidades/departamentos/divisiones, y cada cual tiene su manera de trabajar y unas políticas, el cambio puede resultar mucho más complejo.
    - ¿Hay datos abiertos/compartidos? ¿Qué datos? ¿Hay una política de publicación de los datos? ¿Qué datos se compartirán y con quiénes? ¿Cuáles serán cerrados y disponibles solo para la organización?
    - En esta fase de conocimiento institucional de los datos, podría ser también conveniente visualizar el flujo de datos actual a través de un gráfico.

Dependiendo del alcance y la ambición de la iniciativa de compartición/apertura de datos, el diagnóstico de partida se puede realizar de manera más o menos formal, tomando como referencia el modelo de madurez de datos abiertos de ODI (Open Data Institute)[15]. Este modelo permite evaluar la madurez de una organización con respecto a la apertura de datos en cuanto a actividades operativas y estratégicas se refiere. El modelo proporciona orientación sobre posibles áreas de mejora y ayuda a las organizaciones a compararse entre sí, con el objetivo de resaltar sus respectivas fortalezas y debilidades, y adoptar las mejores prácticas y procesos.

En esta fase conviene evitar las discusiones sobre si conviene o no compartir/abrir datos y buscar ejemplos de beneficios que puede reportar publicar ciertos datos en abierto. Es importante igualmente que todos los miembros de la organización se involucren en la detección de oportunidades de apertura/compartición de datos. Distintas personas de la organización pueden ver valor en ciertos

[15] https://theodi.org/article/open-data-maturity-model-2/

datos que otras no ven, porque tienen un contexto especifico donde los datos aportan valor.

- **Conocer la situación objetivo/destino de la organización respecto a los datos abiertos/compartidos**.
    - Definir los objetivos estratégicos: ¿Qué se quiere conseguir? ¿En qué plazo? ¿Dónde puede estar la organización en 2 - 5 años?
    - ¿Cuál es el modelo de relación B2B o B2G que mejor nos encaja?: monetización de datos, *Marketplace*, plataforma industrial?, etc.
    - ¿Interesa disponer de un equipo técnico especializado en analítica de datos? ¿De qué conocimiento interno dispone la organización? ¿Interesa extender acuerdos con socios tecnológicos (centros tecnológicos, universidades, consultoras especializadas, etc.) que aporten esta experiencia?
    - ¿Interesa disponer de un equipo comercializador de esa línea de negocio? ¿Interesa extender acuerdos de colaboración con terceros cuya misión sea hacer de intermediario entre la organización que provee los datos y el usuario final de los mismos?
    - ¿Es necesario apoyo legal para ayudar a licenciar los datos compartidos/publicados?
    - Definir medidas asociadas a los objetivos (en relación con el alcance, tiempo, número de entregas, etc.) que permitan hacer el seguimiento de los objetivos.
    - Visualizar el flujo de datos destino.

- **Definir e implantar la estrategia y plan de acción adecuados para alcanzar los objetivos**.
    - Desarrollar un plan de acción para alcanzar los objetivos y gestionar el plan como un proyecto.
    - Asegurar la implicación de todos aquellos involucrados en la iniciativa, ya sean internos o externos cuando sea necesario.
    - Definir el modelo de gobernanza de datos adecuado para la organización:
        - ¿Cómo gestiona la organización sus datos (propios y de terceros).
        - ¿Hay datos sensibles? ¿Información confidencial?
        - ¿Es necesario anonimizar y agregar datos antes de su publicación?
        - ¿Están los datos en formatos fácilmente utilizables por terceros?
        - Se han asignado responsabilidades en la organización para la ejecución de los distintos procesos relacionados con la compartición/apertura de datos, incluido el precio del dato y formas de pago?

Ver apartado Gestión y gobernanza de Datos.

- ¿Qué tipo de licencias se van a utilizar? (*ver apartado relacionado con licencias*).

- Definir el modelo de financiación (*ver apartado Modelos de financiación*).

- Definir y establecer los procesos para extraer, transformar y publicar los conjuntos de datos e, identificar e implantar progresivamente la tecnología necesaria para preparar los datos y para compartirlos/publicarlos como *datasets* de calidad alta (ver preparación técnica y preparación de datos).

- Definir y establecer los procesos para monitorizar la reutilización de los datos compartidos/abiertos y mejorar los procesos establecidos para la compartición.

- Tener en cuenta que la evolución en una organización hacia una estructura de gestión de datos adecuada puede requerir el despliegue de un proceso de gestión del cambio.

- **Definir y mantener el inventario de datos de la organización**. Una organización madura en la compartición/apertura de datos trata los datos como activos organizacionales. El catálogo debe contener *datasets* internos y los *datasets* de otras organizaciones que reutiliza (en caso de reutilizar datos a su vez de terceros).

  - Posibles atributos sobre los *datasets* publicados/compartidos:

    - Riesgos asociados a la publicación del *dataset* (¿contiene o no datos personales? ¿dichos datos pueden considerarse sensibles?). Se debe garantizar el cumplimiento de las normativas aplicables. La más relevante es el cumplimiento del RGPD. La publicación de datos de carácter personal requerirá el consentimiento del titular de los datos. Además, sería necesario comunicar la identidad de los destinatarios de la información de carácter personal en el momento de la solicitud del consentimiento, además del resto de información que se debe facilitar al titular cuyos datos serían tratado y publicados. En una estrategia de *Open Data,* el riesgo de no cumplir con estos requisitos provoca que resulte necesario anonimizar la información que va a ser publicada. Con la anonimización de la información (que no se pueda identificar al titular de los datos) se podrán publicar los datos, ya que estos no serán de carácter personal.

    - Se debe tener en cuenta que la empresa puede haber establecido una política de secretos empresariales. La Directiva de Secreto Empresarial y la Nueva Ley de Secretos Empresariales establecen los requisitos para que determinada información que suponga una ventaja competitiva para una empresa pueda ser considerada Secreto Empresarial. Por tanto, las empresas deben tener en cuenta si la información que van a publicar puede quedar afectada por la vulneración del Secreto Empresarial, tanto por si se trata de un Secreto Empresarial propio de la empresa como si dicho secreto viene licenciado por un tercero.

    - Valor de los *datasets* (determinado por el proceso de valorización de la organización).

- Detalles de la inversión financiera asociada a la publicación del *dataset*.
- Planificación de la publicación del *dataset*.
- Métricas claves recogidas durante el proceso de monitorización y mejora continua; etc.
- Posibles atributos sobre los *dataset* reutilizados: derechos para el uso del *dataset;* costes asociados a la reutilización (para datos comerciales); riesgos asociados con una subida de precio (datos comerciales), riesgos de que el *dataset* no se mantengo o no esté disponible; cómo se reutiliza el *dataset* en la organización; etc.

- **Definir la política de datos abiertos/compartidos y comunicarla**.
  Definir una política sólida a nivel organizacional aumenta la trasparencia de la organización y asegura el buen uso de tus datos. Comunicar la política de datos abiertos contribuye a concienciar sobre los beneficios de los datos abiertos y ello redunda en un aumento de los datos abiertos/compartidos y en la calidad de éstos.
  - ¿Qué es parte de la política?
  - ¿Qué datos se publicarán?; Cuáles se compartirán?; ¿Se han anonimizado previamente?
  - ¿En qué condiciones?; ¿Cuándo?; ¿Con qué frecuencia?
  - ¿Cuál es el impacto esperado? ¿Y los beneficios?
  - En cualquier caso, antes de que una organización defina una política y estrategia para la compartición/apertura de datos es importante que identifique los beneficios que ésta le puede reportar, cual es el propósito de compartir/publicar datos, dar a conocer ejemplos de beneficios tanto internos como externos de la compartición/apertura, etc. Es decir, trabajar en la organización la cultura de la apertura/compartición de datos.

4.3 Preparación técnica[16]

La apertura/compartición de datos en la empresa privada puede tener un impacto relevante en términos técnicos y de gobierno de datos. La organización tiene que prepararse para poder soportar de manera eficiente y sostenible el ciclo de vida de los datos, desde su recogida hasta su publicación y mantenimiento.

El propósito de la fase de preparación, previa a la puesta en marcha del ciclo de vida de datos abiertos/compartidos, es ayudar a la organización a tomar las decisiones adecuadas sobre la gestión de datos y los cambios organizacionales que conlleva, así sobre la infraestructura tecnológica de la que se debe dotar para extraer y tratar los datos antes de su apertura/compartición según sea la ambición y estrategia de la organización. En la guía se abordan también algunos aspectos relacionados con la gobernabilidad del dato.

A continuación, se detalla los aspectos a considerar a nivel organizativo previos a la puesta en marcha del ciclo de vida de apertura/compartición de datos:

[16] El contenido de este apartado se ha inspirado en el apartado "3. Technical preparation and implementation" de https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf (las ilustraciones 3, 4, 5 y 6 provienen de dicho documento).

Ilustración 2- Preparación previa a la puesta en marcha del ciclo de vida de datos abiertos/compartidos.

4.3.1 Gestión de Datos

La gestión de los datos debe abarcar a todos los *datasets* de la organización, los generados internamente, tanto si se van a compartir como si no. Las organizaciones pueden decidir compartir ciertas clases de datos y otras no. En cualquier caso, se deberá tener una política de gestión que incluya esta posibilidad. También se tendrán en cuenta los *datasets* procedentes de terceras partes en el caso de que aplique. Una organización orientada a la compartición/apertura de datos tratará los datos compartidos de terceros como si fuera su propio activo.

A continuación, se detallan posibles estructuras de gestión de datos. Estas estructuras son teóricas y se deben adaptar a la casuística de cada organización.

- **Gestión de Datos Descentralizada (GDD):** La administración de los datos está distribuida en unidades o personas que son las que tienen la responsabilidad de gestionar sus propios datos. Esta situación, es un ejemplo de gobernanza de datos a corto plazo y sólo se recomienda para aquellas organizaciones en las que la compartición de datos sea una actividad ocasional y manejan cantidades muy pequeñas de datos (ej. un máximo de 10 *datasets* que se actualizan un máximo de 4 veces por año):

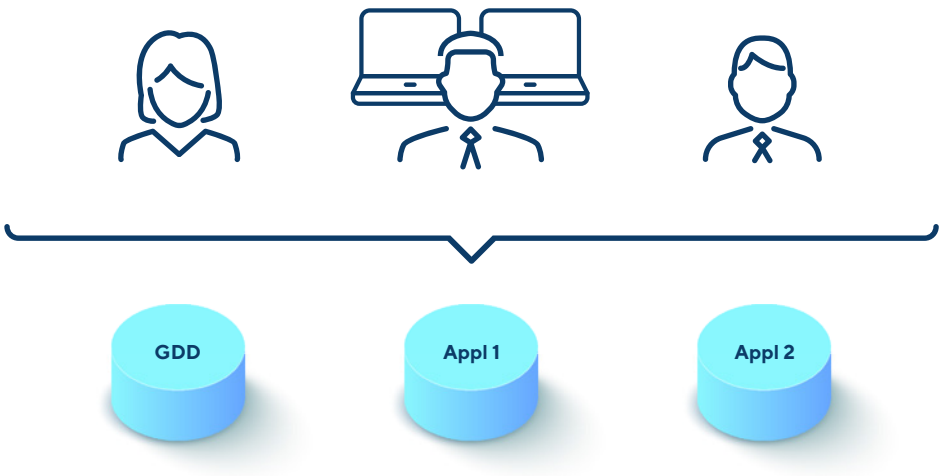


Ilustración 3- Estructura Descentralizada.

Este tipo de gestión es típica en organizaciones tradicionalmente no intensivas en datos que empiezan a gestionar grandes volúmenes de datos (ej. Industria 4.0). Estas organizaciones deberán mejorar la gestión y gobernanza de sus datos como paso previo a su compartición/apertura.

- **Gestión de Datos Centralizada (GDC):** En este caso, la gestión de datos todavía está distribuida en las distintas unidades, pero existe una clara asignación de responsabilidad centralizada. En esta situación existe más coordinación y por lo tanto más consistencia entre las distintas unidades que gestionan datos. Este es un enfoque más acertado para organizaciones que tratan con grandes volúmenes de datos. Es un primer paso hacia la gestión de datos completamente centralizada.

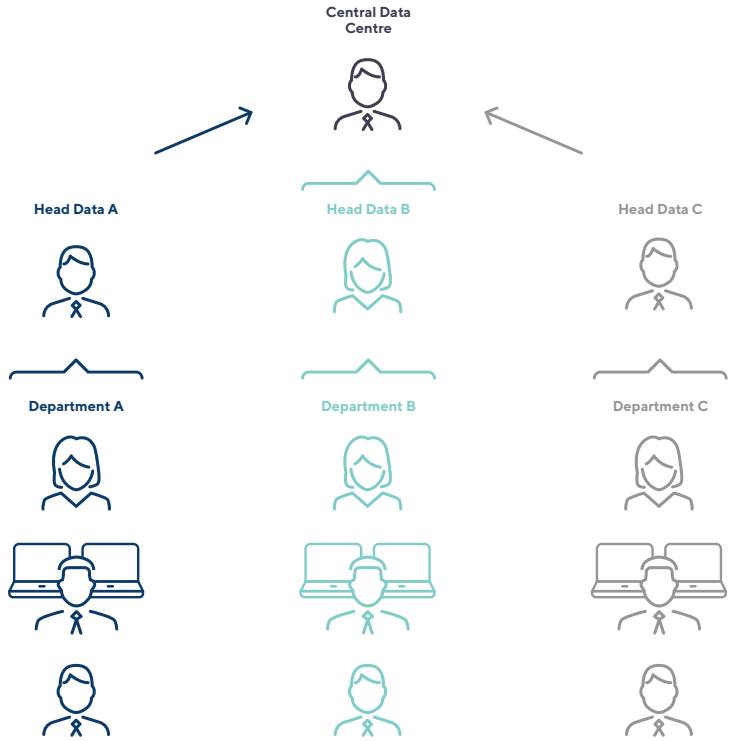


Ilustración 4- Estructura Centralizada

- **Gestión de Datos Completamente Centralizada (GDCC):** En este caso, la estandarización se define en el nivel más alto y la consistencia se garantiza en toda la organización. En este caso, se minimizan los errores de ejecución durante el ciclo de vida de los datos, reduciendo costes a largo plazo. Además, esta estructura de gobierno es más sostenible incluso cuando la estructura de la organización cambia a lo largo del tiempo. En grandes corporaciones con una gran riqueza y diversidad de datos la centralización puede limitarse a ciertos silos semejantes de datos en cuyo caso no hablaríamos de una GDCC, sino de una GDC.

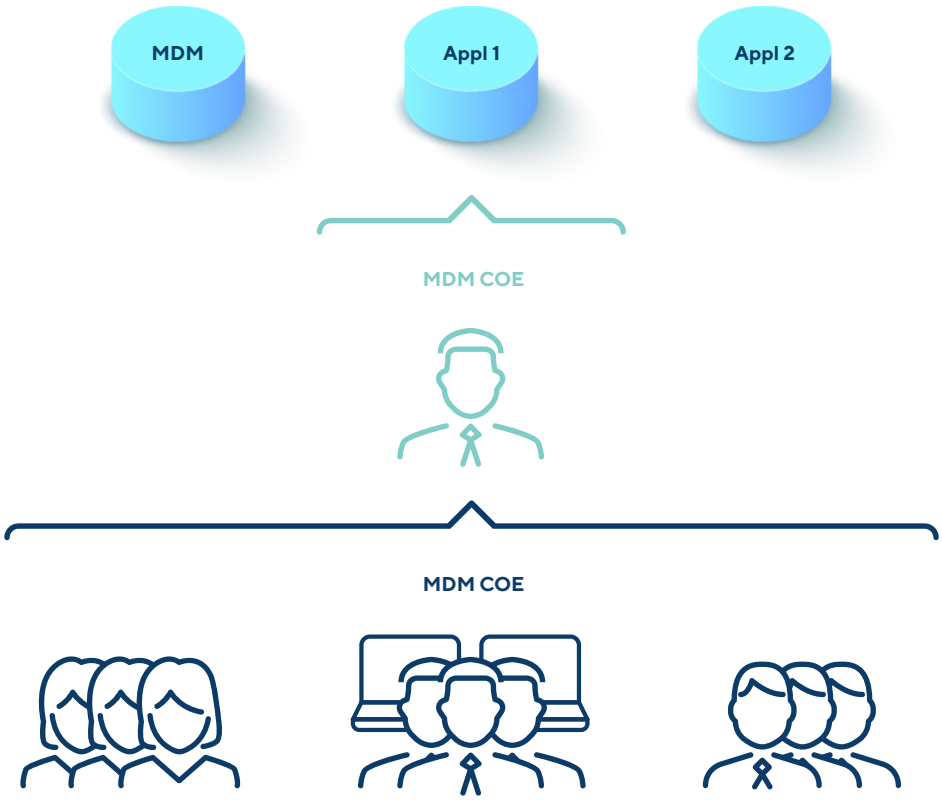


Ilustración 5- Estructura Completamente Centralizada

4.3.2 Extraer, Transformar y Compartir Datos (ETCD)

El proceso de ETCD/ETCA son las especificaciones técnicas de cómo es el flujo de datos y la transformación de datos a llevar a cabo en la organización para que estos datos estén listos para compartir o abrir (publicarlos como datos abiertos). Independientemente de si finalmente se comparten y/o publican como datos abiertos o se usan solo internamente. Este proceso se define a nivel organizacional y se complementa con la preparación de cada *dataset* que se aborda en el apartado 4.4.

El proceso de ETCD/ETCA con la tecnología implementada es el proceso central que se ejecutará de manera recurrente para la obtención de *datasets* (bien de uso interno o a compartir/abrir).

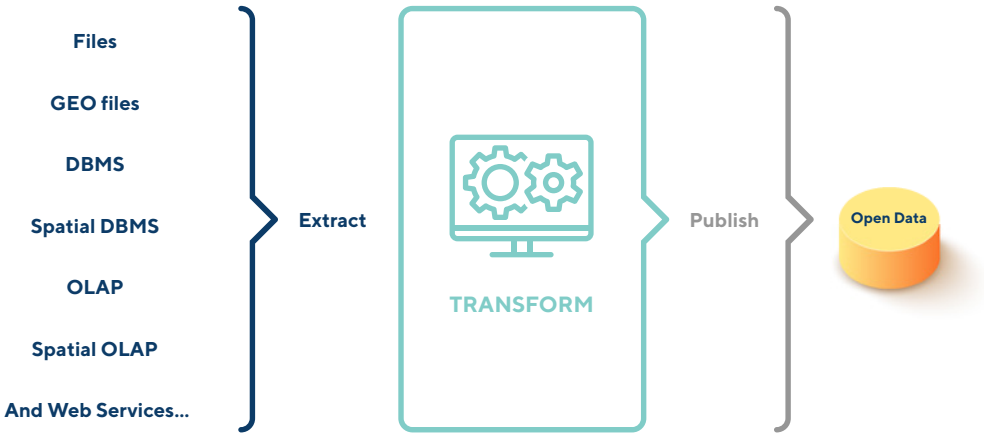


Ilustración 6- Proceso ETCD o ETDA

Los tres pasos de los procesos ETCD/ETDA son:

1. **Extraer**: los datos se extraen de cualquier tipo de fuente, incluye nuevos datos generados a partir de los existentes y datos de fuentes internas o externas (documentos, Bases de Datos y/o aplicaciones o sistemas). El proceso de extracción está condicionado por la gestión de datos de la organización. Dependiendo de los objetivos estratégicos o de la ambición de la organización se abren múltiples escenarios para los procesos de ETCD y ETDA. (ver posibles escenarios en Anexo 1).

2. **Transformar**: Consiste en transformar los datos de partida en datos listos para compartir (interna o externamente) o publicar en abierto, o como datos enlazados (*Linked Data)*. La organización tiene que definir una guía para la preparación de los datos (por ejemplo: por tipos de *dataset,* según si contienen o no datos sensibles, datos geolocalizados, datos en tiempo real, y una política que haga referencia a dichas guías. (Ver posibles escenarios en Anexo 1).

3. **Compartir/Publicar (optativa – no es necesaria si los *dataset* son solo para uso interno)**: Identificar los canales para compartir o publicar los *datasets*. Adicionalmente se tienen que ejecutar los siguientes pasos, que forman parte del proceso Ciclo de Vida de los datos abiertos/compartidos (Ver apartado 4.4. Poner en la práctica el Ciclo de Vida de los datos abiertos/compartidos):

- Elegir el modelo de licencia.
- Definir los metadatos asociados a los *datasets*.
- Datos abiertos: Publicar el *dataset* (preferiblemente de manera automática).
- Ofrecer la posibilidad de convertir los datos para determinadas plataformas y, desarrollar APIs (en caso necesario).
- Establecer el ciclo de vida indicando los detalles de la persona de contacto de la organización.
- Asegurar la compartición/publicación periódica de los *datasets*.

4.3.3 Canales de Compartición de datos

Dependiendo de la política y estrategia definidas en la organización relacionadas con la compartición/apertura de datos, existen distintos mecanismos técnicos de distribución de la información. Se denomina canales de compartición de datos a los distintos mecanismos técnicos habilitados para la distribución de los datos de la organización a terceros[17]. La organización deberá identificar aquellos canales de compartición de datos que más se ajusten a sus necesidades:

Ilustración 7 – Canales de Compartición de Datos

[17] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0125&rid=2

Descargas Web

Esta es la forma más sencilla de compartir datos y es el canal más utilizado en los casos en los que la organización decide compartir pequeñas cantidades de conjuntos de datos. Los datos pueden estar publicados en el sitio web de la organización, en una parte privada con acceso controlado a los enlaces de los datos.

La solución de software utilizada podría ser cualquier Sistema de gestión de contenido (CMS), como Drupal, Wordpress, Django, u otros posiblemente en combinación con una base de datos interna.

API

Una API (Interfaz de Programación de Aplicaciones) es una forma más avanzada y técnica de conectar distintos sistemas o aplicaciones entre sí. En el contexto de la compartición de datos, una API es un *software* que proporciona un sistema con acceso directo a los datos de otro sistema, otorgando acceso al uso de los catálogos de datos y sus funcionalidades. Sin la interferencia de las interfaces de usuario final, como portales o páginas web, una aplicación de terceros puede cargar los datos mediante un protocolo de solicitud. Es particularmente útil si los datos deben estar actualizados, los terceros deben acceder directamente a ellos y la aplicación que los usa necesita acceso directo a los datos en la base de datos sin ninguna interferencia.

La Comisión Europea anima[18] a las empresas de toda Europa a considerar el uso de APIs abiertas, estandarizadas y bien documentadas para la compartición de datos. Asimismo existen múltiples soluciones técnicas que pueden utilizarse como CKAN, Sócrata, Open Data Soft y DKAN (Ver Anexo V).

[18] Comisión Europea (2017): "La construcción de una economía de los datos europea", Bruselas, COM (2017) 9 final.

Portal de Datos

Una forma más avanzada de publicar conjuntos de datos es a través de un portal de datos, como es habitual en los datos púbicos de las Administraciones. En el entorno empresarial esta solución es particularmente útil si la organización dispone de grandes cantidades de conjuntos de datos que desea publicar y que además deben mantenerse actualizados con cierta periodicidad. Los portales son soluciones web al uso que normalmente tienen su propio nombre de dominio.

Otro enfoque sería la publicación en portales de datos de terceros. Esta es una solución de bajo costo y es muy aplicable en algunas situaciones. Al publicar los datos en los portales de datos existentes, se facilita el acceso a los mismos ya que se incrementa el despliegue pudiendo llegar a ser conocidos por más entidades interesados en los datos. En este caso, se pueden tener en cuenta dos opciones compatibles: 1) los datos se encuentran cargados directamente en el portal y se descargan de forma directa o bien 2) el portal redirige el acceso automáticamente a la web de la organización donde se encuentran los datos y la descarga se realiza desde la propia web de la organización. En el caso en el que se decida introducir los datos en portales abiertos, seguramente se requiera implementar configuraciones adicionales, como formatos de metadatos, uso de vocabularios específicos del ámbito disciplinar de los datos u otros estándares técnicos que permitan fomentar su interoperabilidad y reutilización.

Marketplace

Se trata de intermediarios de confianza que reúnen a proveedores y usuarios de datos para intercambiar datos en una plataforma en línea segura. Estas empresas obtienen ingresos de las transacciones de datos que ocurren en la plataforma. Las funciones básicas de un *Marketplace* son:

- Propiciar coincidencias entre el posible proveedor de datos y el consumidor de datos. Esto puede incluir configuraciones específicas en las que el proveedor potencial y el consumidor potencial pueden permanecer en el anonimato en la primera parte del establecimiento de la preparación de transferencia de datos, ya que la intención de proveer o comprar puede revelar información comercial ya secreta (estrategias comerciales futuras).

- Transferencia real de los datos y la compensación acordada. El *Marketplace* genera un espacio de confianza garantizando que el objeto de la negociación no será alterado durante el curso de las negociaciones.

- Certificar que la transacción realmente ha sucedido.

- Adicionalmente se pueden proporcionar servicios adicionales, como cláusulas contractuales modelo o servicios de anonimización si se intercambian datos personales o confidenciales.

El rol de este tipo de intermediario termina normalmente una vez que los datos han sido transferidos, a menos que el grupo que ha interactuado quiera seguir colaborando y compartiendo datos y vea la posibilidad de crecer con más participantes.

La mayoría de los *Marketplace* son centralizados, aunque tecnologías como DLTs (*Distributed Ledger Technologies*) y la creación de algoritmos de carácter descentralizado (ej. OceanProtocol[19] están permitiendo la creación de *Marketplaces* descentralizados.

[19] https://oceanprotocol.com

Plataforma Industrial

Las plataformas de datos industriales están creadas para fomentar un enfoque de colaboración para el intercambio de datos entre empresas. Más concretamente, un grupo restringido de usuarios se une voluntariamente a dichas plataformas para beneficiarse mutuamente de los datos compartidos. Por lo general, los datos se intercambian de forma gratuita, pero se puede considerar la aplicación de tarifas por la prestación de servicios adicionales, en un entorno cerrado, exclusivo y seguro. Estas plataformas tienen como objetivo mejorar el rendimiento y la productividad de los miembros de la plataforma. Las plataformas industriales son utilizadas normalmente por grandes organizaciones.

Facilitador Técnico

Aunque pueden parecerse a las plataformas de datos industriales o los *marketplace*, estas organizaciones pueden entenderse como habilitadores técnicos para el intercambio de datos entre empresas. Se trata de organizaciones que desarrollan sus propias soluciones web y basadas en la nube, para permitir el intercambio de datos entre un grupo de usuarios o socios comerciales. Estas compañías actúan como un tercero en el proceso de intercambio de datos al proporcionar la solución técnica habilitadora. Los usuarios de datos que participan en la compartición de datos a través de estas soluciones tienen la oportunidad de

intercambiar datos dentro de una comunidad particular de manera ágil. Además, hacen uso de una herramienta que ya existe y que se puede personalizar, en lugar de utilizar algo nuevo.

Contrariamente a las plataformas de datos industriales o los mercados de datos, sus ingresos se logran a través de la configuración, implementación y mantenimiento de sus soluciones.

Este tipo de intermediario ofrece funcionalidades adicionales que permiten al proveedor de datos controlar el uso que se hace de los datos. Incluye formas de seguimiento y rastreo del uso de datos realizado, por ejemplo, el registro de todas las acciones de acceso y procesamiento de datos, utilizando la tecnología *blockchain*, o desarrollando formas de marca de agua digital. El intermediario también puede desarrollar instancias de autorregulación dentro de la comunidad de usuarios del espacio de datos o plataforma, incluyendo por ejemplo sanciones para los usuarios de datos que violen los acuerdos de transferencia de datos establecidos.

Compartición/apertura de flujos de datos en tiempo real

Todos los canales de compartición arriba mencionados abordan la compartición/apertura de *datasets* de carácter estático. Adicionalmente la mayoría de *Marketplace* y Plataformas industriales ofrecen también sus propias soluciones y servicios a las empresas para compartir grandes volúmenes de datos (ej. flujos de datos provenientes de sensores de la Industria 4.0) y actualizados en tiempo real. El facilitador tecnológico proporciona a las primeras o al cliente final las plataformas, tanto de software abierto o privativas, que posibilitan dichas soluciones y servicios. Algunos ejemplos de plataformas para análisis de flujos de datos en tiempo real: Apache Flink (https://flink.apache.org/), Spart streaming (https://spark.apache.org/), IBM Steams, Software AG's Steaming Analytics, Azure Stream Analytics, Amazon Kinesis Data Steams, etc.

Cuando los datos de una organización se almacenan fuera de su país y están sujetos a la legislación del país en el que residen, es lo que se denomina "soberanía de los datos". La principal preocupación de la soberanía de los datos es mantener la normativa sobre protección de la intimidad de cada país y proteger a los clientes de citaciones legales sobre sus datos. La **normativa de la UE sobre protección de datos personales restringe las organizaciones de transmisión de datos personales creados en Europa a países cuya legislación en materia de protección de datos es considerada inadecuada por la UE[20]**. Algunos países exigen que los datos confidenciales se almacenen en servidores supervisados por el Gobierno o en centros de datos autorizados, de terceros. El seguimiento y el cumplimiento de los requisitos de la soberanía de datos plantean una cuestión difícil y persistente para las empresas globales. Los principales proveedores de servicios en la nube (PSC), como AWS, Azure, Google, Oracle e IBM, tienen centros de datos nacionales, el primer requisito para cumplir los reglamentos sobre la soberanía de datos. Estos proveedores ofrecen servicios de autenticación, codificación y seguridad, más un conjunto de herramientas de gestión destinadas a ayudar a los clientes a cumplir la normativa local. No obstante, las organizaciones mundiales deben comprender las normativas de datos de cada país específico y trabajar con los proveedores de servicios en la nube y los instrumentos de gestión que les ofrece cada uno de ellos para aplicar la arquitectura, los procesos y las políticas de seguridad necesarias para cumplir estas normas.

[20] https://eugdpr.org

International Data SPACE (IDS)[21]. Es una asociación integrada por más de 90 compañías e instituciones de más de 18 países planteada inicialmente para desarrollar el concepto de la "soberanía del dato". Han definido un marco para la creación de un espacio de datos virtual que permita el intercambio de datos entre organizaciones, creando valor alrededor de los datos, en un entorno de confianza. Este espacio de datos virtual utiliza estándares y modelos de gobernanza comunes para facilitar un intercambio seguro de datos y poder relacionar fácilmente los datos en ecosistemas de negocio en un marco de confianza mutua. Proporciona una base para crear y utilizar servicios inteligentes y procesos de negocio innovadores, al mismo tiempo que asegura la soberanía digital de los propietarios de los datos.

4.3.4 Búsqueda de Datos

Dependiendo del volumen de datos que se desee compartir y del canal de compartición de datos utilizado se puede utilizar un mecanismo de búsqueda de mayor o menor complejidad. Por ejemplo, si los datos se comparten a través de la web de la organización, los mecanismos de búsqueda básicos que facilitan los sistemas de gestión de contenido (CMS) son perfectamente válidos. Otros canales de compartición de datos (ej. portal de datos abiertos, plataformas, etc.) tienen sus propios buscadores para acceder a sus *datasets* o a los *datasets* de terceros referenciados mediante técnicas de *harvesting* o extracción de metadatos (ver información adicional en Anexo IV).

Para búsquedas avanzadas de *datasets* que cumplan con los requisitos técnicos deseables de las tecnologías de datos enlazados *(Linked Data)* o mejor, datos abiertos enlazados *(LOD, Linked Open Data)*, existe la opción de usar SPARQL para las consultas. SPARQL es un lenguaje de búsqueda avanzada que no tiene sentido utilizar si los *datasets* no cumplen los requisitos de *Linked Data* (*datasets* con metadatos y datos descritos según una ontología estándar o reconocida) o están codificados en lenguajes estándar como RDF *(Research Description Framework)*.

Existen también buscadores específicos de *datasets* y universales como **Google Data Search** que se lanzó en septiembre 2018, inicialmente para la búsqueda de datos abiertos de investigación, pero que también agrega otros tipos de Open Data en la web. También existen grandes agregadores de datos a nivel global como **Fireware**.

4.3.5 Otros aspectos técnicos

Otras decisiones técnicas relacionadas con la provisión y mantenimiento del servicio de apertura/compartición de datos que deben ser tomadas en cuenta son:

- **Selección del dominio:** La organización puede optar por usar su propio sitio web para hacer que los datos estén disponibles o a través de un sitio web separado con su propio nombre de dominio (ej. data.institucion.es).
- **Selección del *Host*:** La organización debe determinar si los datos se almacenarán y estarán disponibles en sus propios servidores o si utilizará servidores de terceros.
- **Selección de funcionalidades**: Se deberá identificar por ejemplo que bases de datos serán compartidas, si se va a requerir disponer de un módulo de pago, espacio necesario en el servidor, velocidad de acceso requerida, etc.

[21] https://www.internationaldataspaces.org

- **Administración del canal de compartición**: Se debe asignar un responsable de administrar el canal de compartición. Adicionalmente se debe definir el grado de disponibilidad, seguridad y seguimiento, entre otros aspectos, del canal de compartición.

- **Mantenimiento del servicio**: Asimismo debería nombrarse un responsable del mantenimiento del servicio, de forma que se garantice la disponibilidad, la funcionalidad y el rendimiento del servicio de compartición de datos.

4.4. Poner en práctica el ciclo de vida de los datos: recoger, preparar, publicar y mantener

La puesta en marcha del ciclo de vida de datos abiertos/compartidos[22] es un proceso que describe los pasos a seguir para generar los *datasets* (cada *dataset*) a compartir/publicar. Es un proceso que se debe incluir en la política de la organización y que engloba los siguientes subprocesos:

Ilustración 8 – Ciclo de Vida de Apertura/Compartición de Datos

4.4.1 Recoger datos

El propósito es identificar qué *datasets* vamos a compartir/publicar, priorizarlos, registrarlos en el catálogo de *datasets* a compartir/publicar de la organización en caso de existir, categorizarlos y determinar cómo promocionar/publicar los *datasets*. Se parte de datos ya disponibles en la organización, de manera centralizada o distribuida. La recogida de datos puede ser puntual (búsqueda de un beneficio rápido) o tener un enfoque de gestión de datos más minucioso (basado en procesos de preparación).

[22] https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

Para organizaciones sin experiencia en la compartición/apertura de datos, las principales actividades a realizar son:

Ilustración 9 -Actividades de Recolección

1. **Mapear los datos disponibles en la organización para el propósito u objetivo estratégico en cuestión**: Identificar los datos que se pueden compartir con otras organizaciones y los datos a publicar como datos abiertos. Debe pedirse ayuda a los responsables de los datos para la tarea.

2. **Priorización de los *datasets***: Criterios para priorizar los *datasets*:

   - ¿Son adecuados para el propósito en cuestión?
   - ¿Se pueden compartir/abrir tal cual están (legalmente, políticamente y, organizacionalmente)?
   - ¿Tienen la calidad requerida o es necesario que pasen por un proceso de tratamiento antes de su compartición/apertura?
   - ¿Es necesario "limpiarlos", anonimizarlos, mejorar la calidad y el formato?
   - ¿Es un *dataset* con una temática de alto valor?

Se recomienda empezar primero por los *datasets* que nos reporten beneficios con el mínimo esfuerzo (*quick wins*). Es decir, aquéllos para los que las respuestas a todas las cuestiones que hemos formulado arriba, son positivas.

3. **Establecer el proceso para la identificación y priorización de los *datasets***: Definir y establecer el proceso para la identificación, modificación, creación y eliminación de los *datasets* de la organización.

4. **Categorizar los *datasets*:** Consiste en la agrupación de *datasets* por temáticas o categorías.

   Las categorías deben estar orientadas a la acción. A la hora de clasificar/asignar los *datasets* a categorías, se debe analizar el *dataset* desde la perspectiva de los reutilizadores. Las categorías permiten estructurar de manera lógica los *datasets,* pero debe evitarse la categorización excesiva.

   No existen ejemplos de categorización de *datasets* para organizaciones privadas, pero puede utilizarse como referencia las existentes para organizaciones gubernamentales como las categorías del portal europeo[23] de datos abiertos, y las categorías del portal de datos español, a nivel estatal[24]. Las categorías de *dataset* existentes y el número de *datasets* por categoría ayudan a la organización a priorizar los *datasets*.

[23] https://joinup.ec.europa.eu/solution/dcat-application-profile-data-portals-europe

[24] https://datos.gob.es/es/doc-tags/nti-risp

5. **Publicar las categorías de datos y los *datasets* disponibles por categoría**: La clasificación de los *datasets* en categorías se puede hacer de manera automática a través de los metadatos asociados a los *datasets*. Por ejemplo en el portal de datos europeos, la categorización de los *datasets*[25] se realiza de manera automática basándose en el estándar DCAT a través de la extracción de metadatos (*metadata harvesting)*.

4.4.2 Preparar datos

El propósito es preparar los datos (cada *dataset*) para compartirlos o publicarlos en abierto. Los datos de partida identificados, la mayoría de las veces, hay que tratarlos antes de su publicación.

El tratamiento o preparación de los datos abarca los siguientes aspectos (a ejecutar no necesariamente de manera secuencial):

- Calidad del dato.
- Preparación del dato.
- Añadiendo metadatos.
- Preparación legal.
- Gestión de datos sensibles.

La viabilidad legal de la apertura/compartición del dato es clave para poner en marcha el proceso de preparación del *dataset*. Es decir, si los datos no cumplen con la legislación vigente, no se pueden compartir/abrir por lo que no tiene sentido iniciar el proceso de preparación.

1. **Calidad del dato**: La calidad del dato abarca aspectos como:

- **Contenido**: La utilidad del dato está condicionada por su calidad. Principales aspectos que determinan la calidad del dato y qué hacer para mejorarlos:
    - ¿Es el dato completo? Describir una cabecera con una descripción única en los metadatos. No cambiar la estructura, aunque cambien los datos; tiene que tener un número de versión y, si el contenido cambia, se debe asignar una nueva versión para facilitar la traza de los cambios; describir el contenido de la información: sobre qué es el dato, de donde procede y con qué propósito se quiere compartir o publicar; asignarle un estado (ej. borrador/validado/final).
    - ¿Están los datos limpios? Verificar campos vacíos; ¿tienen sentido los valores por defecto y los datos ficticios?; valores erróneos; dobles entradas; información privada sensible; ¿se viola algún requisito legal?
    - ¿Son los datos precisos? ¿Es el dato suficientemente preciso para el propósito concreto (compartición de datos) o potencial (datos abiertos) ?; ¿Afecta la precisión a su fiabilidad? ¿Se han descrito la elección de los intervalos de tiempo? ¿Necesitan los datos pasar por un proceso de agregación o desagregación?

[25] http://www.europeandataportal.eu

La calidad del dato se puede medir por su nivel de cumplimiento de los principios FAIR (*Findable, Accessible, Interoperable, Reusable)* al incluir los metadatos adecuados y los identificadores persistentes para su identificación y reutilización.

- **Prontitud**: Abordar la noción del tiempo en los datos. Los datos cambian con el tiempo, es importante establecer un proceso para la actualización del dato. Este aspecto está muy relacionado con el mantenimiento de los datos.

- **Consistencia**: La consistencia en la presentación de los datos es un aspecto fundamental. Es importante utilizar estándares para compartir/publicar *datasets* y ser consistente en compartir/publicar datos de la misma calidad. Por ejemplo: Si los *datasets* se publican anualmente y, cada año, los datos se definen con distintos atributos, la manipulación/esfuerzo/coste requerido por el reutilizador es mayor y, frecuentemente, inasumible por muchas organizaciones.

2. **Preparación de los datos para la compartición/apertura técnica**. Para compartir los datos de una organización con otras personas u organizaciones la interoperabilidad del dato, así como la información sobre el dato que nos facilita el acceso al mismo son condiciones necesarias, en especial en el caso de datos abiertos, aunque no suficientes. La primera depende del/los formatos en los que se comparte/publica el *dataset*, que determinan la facilidad con la que terceras organizaciones o individuos podrán hacer uso de esos datos con el mínimo esfuerzo y coste. La segunda, fundamental principalmente para la apertura de datos, tiene que ver con la facilidad con la que terceras partes entienden el contenido de los *datasets* y está ligada al uso de metadatos.

3. **Añadir metadatos e identificadores persistentes.**
   En los datos compartidos en plataformas y, principalmente, en el caso de datos abiertos, es importante asegurar el acceso a los datos. Es decir, que los datos puedan ser fácilmente encontrados por las personas u organizaciones interesadas. A este concepto se llama en inglés "*discoverability*". Sin embargo, cada vez más se habla más de datos FAIR. Aunque la "FAIRness" es un requisito surgido y aplicable a los datos de investigación, es también aplicable a cualquier conjunto de datos. No en vano, los datos del sector empresarial pueden utilizarse para la investigación o incluso surgir de investigaciones originadas en el ámbito privado. En cualquier caso, para que los conjuntos de datos sean FAIR es esencial que éstos tengan asociados metadatos e identificadores persistentes. Los metadatos son datos sobre los datos, , información estructurada que describe, explica, localiza o en cualquier caso facilita la recuperación, uso o gestión de los datos, y sirven para entender la naturaleza y estructura del *dataset*, conocer su origen, así como los términos en los que puede ser usado. Por su parte, los identificadores persistentes permiten que un conjunto de datos tenga siempre el mismo URI y por tanto garantizar su acceso.

4. **Preparación legal para la compartición/apertura.**
   Como ya hemos dicho y se explicará más ampliamente en el apartado siguiente, los datos tienen que respetar la legislación en vigor y cada *dataset* compartido o publicado en abierto tiene que tener asociada una licencia que tiene que respetar tanto el publicador como los reutilizadores de los datos. Estas licencias tienen que cumplir con toda la legislación vigente y es responsabilidad del que publica los datos asegurarse de ello (*ver apartado de 5.3.3.*).

5. Proceso de Gestión de Datos Sensibles

Los *datasets* que incluyen datos sensibles pueden requerir una preparación especial que incluya:

1. Realizar evaluaciones de riesgo y análisis de impacto antes de compartir conjuntos de datos con información sensible.

2. Definir mecanismos de mitigación de los riesgos:

   - Definición y despliegue de políticas y procesos de anonimización de datos.

   - Identificación de la información sensible, por ejemplo, datos personales de salud, origen racial, etc.

   - Preparación de subconjuntos de datos para eliminar campos sensibles.

   - Agregación de los datos.

   - Búsqueda del consentimiento de las personas cuyos datos personales puedan ser divulgados. El consentimiento en este caso deberá ser una acción en positivo del titular de los datos facilitándole claramente la información de los destinatarios y finalidades. Además, el consentimiento es revocable por lo que puede darse la circunstancia de que el titular revoque su consentimiento y se tenga que eliminar dicho dato, que previamente había sido publicado.

   - Definir de forma adecuada el entorno de los datos: conjuntos de datos relacionados, personas que tienen acceso a los conjuntos de datos, personas que podrían hacer un uso malintencionado del conjunto de datos, etc.

3. Es responsabilidad de la organización cumplir con la legislación vigente aplicable en cada caso en función de los datos manejados (*ver anexo sobre legislación*).

4. Es responsabilidad de la organización también, crear una cultura de gestión del dato apropiada. Se debe formar al personal en los nuevas políticas, procesos, herramientas y tecnologías desplegadas y relacionadas con la gestión de los datos.

5. Es necesario definir procedimientos legales para contingencias en caso de incidencias en la compartición/apertura de datos sensibles.

Una organización que se inicia en el proceso de compartición/apertura de datos normalmente no tiene un inventario legal sobre la propiedad de sus datos en el que se identifican qué datos son sensibles y cuáles no. Sin embargo, el tener identificados los datos sensibles de la organización es condición necesaria para aplicar este proceso.

6. Verificar la preparación de los datos:

- Verificar si los *datasets* son legalmente abiertos.
- Verificar la calidad del *dataset*.
- Verificar la frecuencia de actualización de los *datasets*.
- Verificar si los *datasets* se ajustan a estándares establecidos.
- Verificar si los *dataset* tienen los metadatos adecuados y están identificados de forma persistente
- Verificar si los metadatos están descritos como *linked data*.
- Verificar si los *datasets* son técnicamente abiertos.
- Verificar las licencias de cada *dataset* e información sobre la procedencia de dichas licencias.

Algunos de los métodos/modelos más populares para determinar la calidad del dataset (la información asociada al dataset que facilita la reutilización y accesibilidad al mismo):

- Las 5-estrellas de Tim Berners-Lee[26] son una forma sencilla de verificación de la calidad técnica de los conjuntos de datos en virtud del formato de publicación:

☆	Poner el *dataset* disponible en la web (en cualquier formato) bajo una licencia abierta.
☆☆	Ponerlo disponible como datos estructurados (ej. Excel en lugar de la imagen de una tabla).
☆☆☆	No utilizar formatos propietarios (ej. CSV en lugar de Excel).
☆☆☆☆	Usar URIs para identificar cosas, de manera que la gente pueda apuntar a las cosas a través de las URIs.
☆☆☆☆☆	Enlaza los datos a otros datos para proveerlos de contexto (*linked data*).

Tabla 3 – 5 estrella de Tim Bernes Lee.

- MELODA[27] es un método de evaluación sencillo que permite evaluar la calidad de los *datasets* (se aplica a cada *dataset*) en base a los metadatos asociados.

- FAIR Data maturity model[28] es un modelo en desarrollo, surgido en la comunidad de la *Research Data Alliance* (RDA) que trata de crear métricas e indicadores para medir el nivel de encontrabilidad, accesibilidad, interoperabilidad y posibilidad de reutilización (FAIR) de los conjuntos de datos, inicialmente de los datos de investigación, pero como ya hemos mencionado, los principios FAIR son aplicables a cualquier *dataset*. De hecho, existe toda una dimensión en la web vinculada a los datos que se ha empezado a denominar *Internet of FAIR Data and Services* (IFDS).

Estos métodos/modelos y otros que se aplican en contextos concretos, permiten a la organización que comparte/publica datos tener una idea orientativa del valor de sus datos. Sin embargo, lo que realmente dará valor o no al dato será el éxito en su reutilización.

4.4.3 Publicar datos

La publicación de datos se realiza en los canales de compartición/publicación establecidos. Algunos canales de publicación han sido vistos en apartados anteriores, por lo que aquí cabe mencionar que el canal a utilizar depende del número de *datasets* a compartir y se recomienda empezar por las soluciones más sencillas para el número de *datasets* previsto.

[26] http://opendatahandbook.org/glossary/en/terms/five-stars-of-open-data

[27] http://www.meloda.org/wp-content/uploads/2017/03/Meloda4.11.pdf

[28] https://www.rd-alliance.org/groups/fair-data-maturity-model-wg

4.4.4 Mantener datos, metadatos persistentes

Tanto los datos como los metadatos cambian en el tiempo. Se recomienda definir un proceso para mantener los datos y metadatos compartidos/abiertos actualizados. Aspectos a tener en cuenta en el proceso:

Cuando se actualiza un *dataset*, indicar la fecha de actualización en el metadato. Dependiendo del formato de publicación, los URIs y URLs asociados a los *datasets* tienen que revisarse periódicamente para comprobar que siguen en vigor. La WWW es dinámica y los URIs y URLs pueden cambiar.

Proveer un contacto, bien a través de los metadatos o bien incorporando un canal de contacto para los re-utilizadores. Su opinión es fundamental para mejorar la calidad de los datos compartidos/abiertos.

4.5 Monitorización de la iniciativa, análisis del nivel de éxito y mejora continua

Cuando los *datasets* se comparten o publican siguiendo el ciclo de vida establecido es importante hacer un seguimiento de la iniciativa de compartición/apertura de datos en su totalidad, desde la revisión de la política y estrategia hasta la definición e implementación de los distintos procesos. Auditar la vida del dato es esencial, en particular determinar en un futuro la trazabilidad del dato (desde su origen) y su reutilización en diferentes escenarios, tanto comerciales como no comerciales. Hay dos aspectos clave a revisar en cualquier iniciativa:

- **Contrastar el compromiso de los reutilizadores** ya que el compromiso de los reutilizadores es esencial para el éxito de cualquier iniciativa de compartición o datos abiertos. Tim Davies[29] define también un modelo de 5 estrellas (como hacía Tim Berners Lee para medir la calidad de los datos basándose en su formato de codificación), en este caso para analizar el compromiso de los reutilizadores:
  - **1-estrella**: Responder a la demanda.
  - **2-estrellas**: Poner el dato en contexto.
  - **3-estrellas**: Establecer canales de conversación sobre el dato.
  - **4-estrellas**: Construir capacidad, habilidades y red.
  - **5-estrellas**: Colaborar sobre el dato como recurso común.

- **Medir el éxito de la iniciativa y mejora continua**: Algunas medidas básicas que pueden utilizarse son:
  - **Número de descargas y visualizaciones del *datasets***. Sin embargo, es un indicador que no da información sobre la utilidad del *dataset*.
  - **Número de accesos al dato**. Indica si el sistema puede gestionar la petición, si se han descargado los datos, si el sistema puede responder a las peticiones, si hay consecuencias para otros sistemas cuando hay peticiones, etc.
  - **Métricas cualitativas de los usuarios del *dataset***: ¿Es útil el *dataset* para el propósito definido? ¿Es el *dataset* de calidad (limpio, claramente descrito, nivel de granularidad adecuado)?

[29] Time Davies' five stars Open Data engagement, 2012. https://www.timdavies.org.uk/2012/01/21/5-stars-of-open-data-engagement/

El proveedor de datos puede influir sobre este último indicador, adaptando el proceso de preparación del dato mejor a las necesidades de los reutilizadores. Existen herramientas de software libre para este tipo de análisis (ej. **PIWIK**[30]) Algunos portales de open data, plataformas, *Marketplace*, etc. también ofrecen soluciones software con este tipo de métricas.

[30] https://piwik.pro/eu

5. ASPECTOS RELEVANTES DE LA APERTURA/ COMPARTICIÓN DE DATOS EN EL ENTORNO EMPRESARIAL

5.1. Modelos de Relación

5.1.1. B2B (*Business to Business*)

Los modelos de negocio relacionados con el intercambio de datos pueden diferir considerablemente y dependen en gran medida del tipo de datos compartidos y del interés comercial estratégico de ambas partes. Se puede plantear desde un enfoque de datos abiertos hasta asociaciones de datos exclusivas en las que se comparte con una sola parte:

- **Enfoque Datos Abiertos/Compartidos**: Los datos son puestos a disposición de un rango abierto de reutilizadores con pocas o ninguna restricción de reutilización de los datos y con una remuneración nula o muy limitada. Este enfoque de compartición de datos se da en aquellas ocasiones en las que el proveedor de los datos tiene un gran interés en la reutilización de sus datos, por ejemplo, aquellos proveedores de servicios que desean utilizar un ecosistema de desarrolladores de aplicaciones de terceros para llegar a clientes finales.

Nombre de la iniciativa:	**BBVA VALORA**
Dirección web (empresa o iniciativa):	https://www.bbva.es/general/banca-online/bbva-valora/index.jsp
Descripción de la iniciativa:	BBVA Valora es una herramienta online disponible a través de la *app* BBVA que resulta especialmente útil para dos grandes tipos de actividades. Por un lado, ayuda a descubrir el valor estimado de la casa deseada, ya sea para comprarla o alquilarla, a partir de la gran cantidad de funcionalidades de análisis disponibles que integran datos como la evolución del precio del suelo, la dotación de equipamientos públicos y sociales del barrio donde se ubica la vivienda, el parque de viviendas existente o el precio medio de las viviendas en venta o en alquiler. Por otro lado, y para aquellas personas que son clientes de BBVA, la herramienta permite agrupar todos los gastos derivados de la vivienda, así como toda la información relacionada con ella para ayudar a tomar las mejores decisiones si se quieren realizar reformas o controlar las facturas.
Razones para la apertura /compartición de datos	▪ Diversificar la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del cliente ▪ Propósito social, de bien público
Monetización del negocio:	No hay monetización directa. Los datos se ofrecen gratuitamente a todas las personas interesadas (en el caso de la valoración de la casa deseada) y a los clientes del banco (agrupación de los gastos derivados de la vivienda).
Tipo de datos compartidos/abiertos	▪ Datos de localización/posicionamiento. ▪ Datos de transacciones (clientes).

Tabla 4 – BBVA VALORA: Ejemplo Datos Abiertos/Compartidos

Nombre de la empresa:	**Primafrio**
Dirección web (empresa o iniciativa):	https://primafrio.com/
Descripción de la iniciativa:	Se comparten con el cliente datos en tiempo real de parámetros como matrícula del camión, posición, conductor, etc. así como información de facturación y contabilidad de la empresa.
Tipo de datos compartidos/abiertos:	▪ Localización/Posicionamiento del camión en tiempo real; nombre del conductor; etc. ▪ Transacciones (clientes, proveedores).
Razones para la apertura/compartición de datos:	▪ Mejorar la eficiencia de los procesos. ▪ Resolver un problema de gestión interno. ▪ Apoyar la innovación abierta y colaborativa. ▪ Ser pionero en estos procesos (posicionamiento estratégico).
Sectores que re-utilizan los datos:	▪ Distribución/logística. ▪ Servicios TIC, incluyendo desarrolladores app/software. ▪ Automoción y transporte.
Monetización del negocio:	No hay monetización directa. Los datos se comparten de manera gratuita y están disponibles solo para clientes de PRIMAFRIO.

Tabla 5 – PRIMAFRIO: Ejemplo Datos Abiertos/Compartidos

Monetización de los Datos: Enfoque unilateral según el cual las empresas obtienen ingresos adicionales de los datos que comparten con otras empresas. La monetización de los datos puede obtenerse a través de la prestación de nuevos servicios basados en los datos compartidos.

Nombre de la iniciativa:	**EUSKALTEL**
Descripción de la iniciativa:	EUSKALTEL pone a disposición de sus clientes, desde pequeños comercios, empresas o administraciones públicas, informes basados en datos completamente anonimizados y agregados. El objetivo es que las empresas clientes de Euskaltel puedan avanzar en su transformación digital.
Razones para la apertura/ compartición de datos	▪ Diversificar la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del cliente. ▪ Crear nuevos modelos de negocio.
Sectores que re-utilizan los datos:	▪ Transporte. ▪ Sector Público. ▪ Retail/comercio.
Tipo de datos compartidos/abiertos	▪ Datos de localización/posicionamiento.
Monetización del negocio:	La tarifa del servicio depende del valor aportado al cliente. Criterios para determinar el valor de los datos: ▪ Nivel de enriquecimiento de los datos. ▪ Zona geográfica. ▪ Período de tiempo del estudio.

Tabla 6 – EUSKALTEL: Ejemplo Monetización

- ***Marketplace***: Intermediarios de confianza que reúnen a proveedores de datos y usuarios de datos para intercambiar datos en una plataforma en línea segura. Estas empresas obtienen ingresos de las transacciones de datos que ocurren en la plataforma. Este enfoque puede ser adecuado cuando:
  - Existen riesgos relacionados con el uso ilícito de los datos.
  - El proveedor de datos necesita garantía de confianza en los reutilizadores.
  - El proveedor de datos dispone de mecanismos técnicos para prevenir o identificar usos ilícitos de los datos.
- **Plataforma Industrial**: Enfoque colaborativo y estratégico para intercambiar datos entre un grupo restringido de empresas. Los datos se comparten en un entorno cerrado y seguro con el objetivo de fomentar el desarrollo de nuevos productos o servicios, o para mejorar su eficiencia. Esta solución permite ofrecer servicios de valor agregado y, por lo tanto, proporciona una solución más completa: fomenta la asociación de datos, dispone de mecanismos de control sobre el uso que se hace de los datos, los contratos modelos puestos a disposición de la plataforma reducen costos de redacción de acuerdos marco, etc. Los gastos de la plataforma se costean entre las empresas que comparten datos.

Nombre de la empresa:	**ITI**
Dirección web (empresa o iniciativa):	https://www.iti.es/
Descripción de la iniciativa:	Participación en el proyecto *Transforming Transport* (https://data.transformingtransport.eu/), aún vigente, en la elaboración de un piloto en colaboración con varias empresas nacionales e internacionales. El proyecto en su conjunto ha supuesto la realización de varios pilotos de apertura de datos en el sector del transporte y la logística. La compartición se ha realizado entre un número reducido de empresas.
Tipo de datos compartidos/abiertos:	▪ Varían en función del piloto (los hay de localización, transacciones, etc.).
Razones para la apertura/ compartición de datos:	▪ Apoyar la innovación abierta y colaborativa. ▪ Contribuir a la resolución de un reto local o global.
Sectores que re-utilizan los datos:	▪ Distribución/logística. ▪ Transporte.

Tabla 7 – ITI: Ejemplo Compartición de Datos Sectorial (Financiación Europea - H2020).

- **Facilitadores Técnicos**, con dos posibles aproximaciones:
  - Permiten el intercambio de datos a través de una solución técnica. Los ingresos se obtienen de configurar, usar y/o mantener la solución. Aunque pueden parecerse a las plataformas de datos industriales o los mercados de datos, estas empresas pueden entenderse como habilitadores técnicos para el intercambio de datos entre empresas. Estas compañías actúan como un tercero en el proceso de intercambio de datos al proporcionar la solución técnica habilitadora. Contrariamente a las plataformas de datos industriales o los mercados de datos, sus ingresos se logran a través de la configuración, implementación y mantenimiento de sus soluciones. Los usuarios de datos que participan en el intercambio de datos B2B a través de estas soluciones tienen la oportunidad de intercambiar datos dentro de una comunidad particular de manera ágil. Además, hacen uso de una herramienta que ya existe y que se puede personalizar, en lugar de utilizar algo nuevo. Los facilitadores técnicos cobran por la provisión de los servicios que ofrecen.
  - Proveen de servicios de consultoría en analítica de datos a partir de información obtenida de terceros. Suelen ser organizaciones que disponen del conocimiento técnico necesario en analítica de datos y que centran su negocio en la provisión de ese tipo de servicios. Los clientes finales de estas organizaciones pueden ser tanto administraciones públicas como empresas finales interesadas en las conclusiones obtenidas tras la aplicación de la analítica de datos.

Las variaciones y combinaciones de estos modelos son factibles adaptándose en cada caso a la necesidad empresarial existente. El término "compartición de datos" se utiliza para describir todas las formas y modelos posibles que respaldan el acceso o transferencia de datos B2B.

5.1.2 B2G (*Business to Government*)

La provisión y la reutilización de datos en las relaciones con administraciones públicas pueden tomar muchas formas tanto en términos legales como técnicos. Esta sección describe algunos de ellos con más detalle:

- **Donación de Datos**: Puede considerarse como una forma de responsabilidad social corporativa. El programa de donación de datos está respaldado por un equipo dedicado que apoya a cualquier parte potencial interesada en usar los datos.

- **Premios**: Puesta en marcha de concursos que fomentan la participación individual y de compañías especializadas en analítica de datos para encontrar solución a retos de un interés público particular, ej. H2020: *Horizon Prize Big Data Technologies*[31]

- **Acuerdos de Colaboración con Administraciones Públicas**: Los organismos del sector público pueden celebrar acuerdos con empresas privadas, que incluyen el intercambio mutuo de datos, de conformidad con la Directiva PSI22[32] en lo que respecta a la información del sector público compartida con el sector privado. Esto puede traer beneficios también para la empresa privada, ya que podrá obtener información de la correlación de los datos del sector público y privado.

- **Intermediarios**: En los casos en que no existe una relación previa de confianza entre una empresa y un organismo del sector público, se puede encargar a una entidad intermediaria para que obtenga los conocimientos necesarios para el interés público y los transmita a la empresa privada.

- **Intercambio de Datos Cívico**: Se puede alentar a las personas a autorizar a los organismos del sector público para el procesamiento de sus datos personales por parte una empresa privada. Cabe destacar que, en este caso, las autoridades también deberían cumplir con la legislación de protección de datos. El procesamiento debe realizarse de acuerdo con la legalidad vigente (ej. el consentimiento debe estar conforme al Artículo 6 (1) (a) o el desempeño de una tarea realizada en interés público según el Artículo 6 (1) (e)[33]). Este modelo es más probable que funcione en situaciones donde existe un vínculo suficientemente fuerte entre el ciudadano y el organismo del sector público en cuestión (ej. el municipio donde viven) o cuando el propósito de interés público es particularmente convincente desde un punto de vista público.

5.2 Modelos de Financiación

Una organización involucrada en una iniciativa de apertura/compartición de datos debe ser capaz de cuantificar el valor económico de la información que maneja y los costes asociados para llevar a cabo la iniciativa, en términos de infraestructura y coste de personal.

A continuación, se detallan algunas prácticas a considerar:

1. Asegurar la rentabilidad de la iniciativa

Se deberá realizar un análisis coste/beneficio, identificando y controlando tanto los costos como los beneficios de plantear una iniciativa innovadora de publicación abierta de los datos o bien una compartición restringida de los mismos dependiendo del modelo de negocio escogido. Se trata de estimar el ROI de la innovación. La innovación tiene dos tipos de impactos en la organización: tangibles e intangibles.

[31] http://ec.europa.eu/research/horizonprize/index.cfm?prize=bigdata

[32] Directiva 2003/98/EC of the European Parliament and the Council on the re-use of public sector information (OJ L 345, 31.12.2003, p. 90).

[33] En el caso de que las autoridades se basen en el artículo 6(1) del Reglamento General de la Ley de Protección de Datos ("el tratamiento es necesario para el cumplimiento de una misión realizada en interés público "), tal base legal debe ser impuesta por las leyes europeas o del Estado Miembro. Además, en el caso del intercambio de datos cívico, los interesados tendrán que estar adecuadamente informados, incluyendo el derecho de retirar el consentimiento y cualquier otro posible tratamiento de sus datos personales que pudieran realizar las autoridades.

Los tangibles son relativamente fáciles de calcular, siguiendo las recomendaciones del Manual de Oslo[34]:

- Ingresos por nuevos productos/servicios vendidos en los últimos tres años.
- Ahorros de coste por nuevos procesos en los últimos tres años.
- Beneficio económico por mejoras de productividad en los últimos tres años.

Adicionalmente se debe considerar que un proyecto de innovación puede generar nuevos intangibles que pasan al activo contable de la empresa: patentes, marcas o modelos de uso, etc. Estos activos son relativamente sencillos de medir y existen metodologías financieras avanzadas que lo permiten. Así podremos obtener valores de marcas o de una patente.

Un proyecto de innovación empresarial de estas características también afecta a otros activos intangibles:

- En la imagen de marca de una empresa.
- En su posicionamiento.
- En el impacto medioambiental.
- En el capital intelectual que ya existía anteriormente.
- En el atractivo de una empresa como empleador.

A la hora de controlar los costes de la iniciativa, se deben tener en cuenta los costes asociados con la puesta en marcha de los nuevos proyectos, así como los costes de mantenimiento de los ya existentes.

La organización, además, deberá diferenciar los costes relacionados con la gobernanza de los datos y aquellos asociados específicamente con el proceso técnico de publicación de los datos. De esta forma:

- Se evitará atribuir incorrectamente financiación para nuevas inversiones en ambos ámbitos.
- Mejorará la eficiencia de la iniciativa de apertura/compartición de datos reduciendo sobrecostes en ambos ámbitos.

2. Valoración de los *datasets*

La organización deberá disponer de un método de cuantificación para establecer el valor de sus *datasets*. Lo ideal sería realizar un análisis previo para evaluar el interés o demanda existente y los posibles modelos de uso de la información, aunque esto puede no resultar una tarea sencilla. Al menos se sugiere mantener, en la medida de lo posible, contacto cercano con los reutilizadores con el objetivo de poder analizar el interés existente en los datos de la organización y poder tomar decisiones de una manera más contrastada.

La falta de un método de valoración de datos no debe obstaculizar la iniciativa de apertura/compartición de datos, aunque se recomienda disponer de uno ya que ayuda a establecer que *datasets* son más prioritarios y por lo tanto a orientar mejor la estrategia de la iniciativa (dónde enfocar las inversiones, secuencia de publicación de los datos, mercado objetivo, etc.).

[34] http://www.itq.edu.mx/convocatorias/manualdeoslo.pdf

Para valorizar un *dataset* se pueden tener en cuenta criterios relacionados con la reutilización del *dataset,* tales como:

- Disponibilidad de los datos (información distribuida en distintas fuentes, etc.).
- Formato del *dataset*.
- Tipología de la información del *dataset.*
- Singularidad del *dataset.*
- Necesidad de geolocalización (información clasificada por ciudad, país, etc.).
- Espacio temporal en el que se tendrá acceso a los datos (días, semanas, meses, etc.).
- Volumen de datos compartidos.
- Uso final que se realizará de la información (tipo de usuario, reutilizaciones futuras, beneficios obtenidos por el usuario final, etc.)

Cabe destacar que la mejora en la reutilización de un *dataset* no necesariamente tiene que estar directamente relacionada con un mayor valor del *dataset*. Dependiendo del uso que el reutilizador haga del *dataset*, el valor final del mismo será mayor o menor.

3. Cálculo de Tarifas[35]

A continuación, se detallan algunas categorías de costes que pueden tenerse en cuenta a la hora de establecer las tarifas en una iniciativa de apertura/compartición de datos:

- Costes relacionados con la creación de datos (infraestructura tecnológica necesaria):
  - Producción: generación de datos y metadatos, control de calidad, codificación, etc.
  - Recopilación: recogida y clasificación de los datos.
  - Anonimización: supresión, ofuscación y empobrecimiento de bases de datos.
- Costes relacionados con la distribución:
  - Infraestructura: desarrollo, mantenimiento del equipo físico y lógico, soportes, etc.
  - Consulta: conversaciones telefónicas y por correo electrónico con los reutilizadores, costes del servicio al cliente, etc.
  - Otros.

[35] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:52014XC0724(01)&from=EN

- Costes de Gobernanza de los datos.
    - Equipo humano.
    - Gestión de la iniciativa (monitorización y control del proceso).
    - Actividades de concienciación y formación (si fueran necesarias).
    - Otros.

Los costes deberán ser revisados y ajustados con cierta periodicidad. Se recomienda una revisión anual.

El cálculo de las tarifas podría basarse por lo tanto en:

- Los costes identificados previamente (costes de gobernanza y costes técnicos de la iniciativa).
- La valorización de los *datasets.*
- Porcentaje de ajuste considerando los riesgos empresariales de la iniciativa (si hubiera).
- Margen de beneficio.

Algunas de las empresas ejemplo del estudio ***Study on data sharing between companies in Europe***[36], como por ejemplo **ORANGE**, han informado del uso de los siguientes criterios para determinar el valor y tarifas de los datos:

- Complejidad del proyecto (opciones de visualización, cruce de datos de distintas Bases de Datos, etc.)
- Cobertura geográfica de los datos (municipal, provincial, estatal, etc.)
- Tiempo de acceso a los datos (día, semana, mes, etc.)

En el caso de **GEO**, se aplican diferentes criterios para el cálculo del valor de los *datasets* dependiendo de cada situación específica. Aunque el criterio determinante en todos los casos está relacionado con el beneficio esperado por parte del reutilizador de los datos, que depende en última instancia de la aplicación de los datos en cada caso.

En el caso de **Michelin**, el precio de los datos está relacionado con la reducción de costes que las organizaciones obtienen a través del uso de sus servicios (ej. el servicio EFFITRAILER).

En el caso de **API-AGRO**, el valor de los *datasets* se estima en base a los costes técnicos derivados de la compartición de los datos a través de la plataforma. API-AGRO no tiene en cuenta el potencial uso que el reutilizador haga de sus *datasets*. Un caso similar es el de **ENEDIS**, esta organización pone a disposición de terceros sus *datasets* a través de una plataforma abierta "Open Data Platform". Sólo en algunos casos puntuales opta por el cobro de una pequeña tarifa de acceso a sus datos. En estos casos, el valor de los *datasets* incluye el coste de personal derivado del esfuerzo requerido para compartir los *datasets* con garantías de calidad.

[36] https://publications.europa.eu/en/publication-detail/-/publication/8b8776ff-4834-11e8-be1d-01aa75ed71a1/language-en

5.3 Marco Legal

5.3.1 Acuerdos Contractuales en el modelo de relación B2B

La compartición de datos normalmente se implementa sobre la base de acuerdos contractuales. Los acuerdos de monetización de datos no solo pueden ser de naturaleza bilateral, sino que también pueden celebrarse entre múltiples partes.

La definición de los términos contractuales para el uso de datos o los acuerdos de licencia requieren una atención especial para cumplir con la legislación existente, en particular aquella legislación que impida el intercambio de datos o que esté sujeta a condiciones específicas para garantizar los intereses estratégicos de cada parte.

Actualmente existen iniciativas para desarrollar los términos de contrato modelo para diferentes tipos de acuerdos de intercambio de datos y para algunos sectores o tipos de intercambio de datos. La Comisión Europea, a través de un Centro de Apoyo para el intercambio de datos que estará operativo a principios de 2019, planea recopilar las mejores prácticas, los términos de contrato modelo existentes y listas de verificación tipo[37].

A continuación, se detallan algunas consideraciones que pueden ayudar a las empresas en la preparación y/o negociación de los acuerdos de compartición de datos:

1. ¿Qué datos estarán disponibles?

- Identificar los datos que desea compartir de la manera más concreta y precisa posible (ej. datos de I + D, datos de clientes, datos de diagnóstico), incluidos los niveles de actualizaciones que se esperan en el futuro. Cuando los mecanismos que hacen posible el análisis de los datos (ej. métodos, modelos, algoritmos, etc.) se compartan junto con los conjuntos de datos, deben describirse.

- ¿Qué niveles de calidad se pueden asegurar para los datos actualmente y a lo largo del tiempo? Los datos compartidos deben ser de buena calidad, es decir, precisos, confiables y, cuando sea necesario, actualizados. Asegúrese de que no falten datos, que estén duplicados, no estructurados. Especifique la fuente/origen de los datos y cómo se recopilaron/construyeron. Se puede configurar un mecanismo para informar de errores relacionados con los datos.

- ¿Se comparten los datos sobre un conjunto de datos o un flujo de datos?

- Asegurar el cumplimiento de las obligaciones legales que pueden prohibir el acceso o la transferencia de los datos. Asegurar el cumplimiento de los derechos que otros puedan tener sobre los datos. Verificar los derechos de propiedad intelectual e industrial.

- Asegurar el cumplimiento de la legislación de protección de datos. Entre otros, verificar que existe una base legal para el procesamiento de datos personales en línea con el Reglamento General de Protección de Datos.

[37] Cf. Annex to the Commission implementing decision on the adoption of the work programme for 2018 and on the financing of Connecting Europe Facility (CEF) - Telecommunications Sector, p. 42.

2. ¿Quién puede acceder y reutilizar los datos?

- Asegurar que el contrato defina de manera transparente, clara y comprensible quién tiene derecho de acceso, derecho de reutilización y derecho de distribución de datos y en qué condiciones. Especificar si y cómo los datos pueden tener licencia para su reutilización. Sea claro al explicar las condiciones de las licencias para la reutilización y distribución de datos. La sublicencia (se concede al licenciatario original la autoridad legal de otorgar una licencia adicional a otros) también debe ser considerada: 1) bien debe ser excluida específicamente, o 2) se deben especificar las condiciones bajo las cuales se permite y para qué tipo de datos.

- El derecho de acceso y reutilización de los datos no tiene que ser ilimitado. El acuerdo puede, por ejemplo, limitar el derecho de acceso: por ejemplo, solo a miembros de grupos profesionales específicos o vincularlo a ciertos propósitos de uso de datos (ej. para un uso comercial).

3. ¿Qué puede hacer el reutilizador con los datos?

- Durante las negociaciones del contrato, el reutilizador debe ser lo más abierto y claro posible acerca de la forma en que se utilizarán los datos. Esto asegurará la transparencia y aumentará la confianza del proveedor de los datos.

- Especificar el uso exacto que se puede hacer de los datos, incluidos los derechos sobre los derivados de los datos (análisis).

- Definir reglas de no divulgación posterior a terceros.

4. Definir los medios técnicos para el acceso y/o intercambio de datos, incluyendo

- Frecuencia de acceso y número de descargas.

- Requisitos de seguridad.

- Niveles de servicio para el soporte.

5. ¿Qué datos necesito proteger y cómo los protejo?

- Asegurar que las medidas adecuadas para proteger sus datos han sido implantadas. Las medidas deben aplicarse a las transacciones de intercambio de datos y al almacenamiento de datos, ya que los datos pueden ser objeto de robo o uso malintencionado. Los datos también pueden ser liberados accidentalmente, por ejemplo, a través de un error humano o debido a un problema técnico. Los datos también pueden estar sujetos a acceso o divulgación no autorizados o pueden perderse.

- Asegurar la protección de secretos comerciales, información comercial sensible, licencias, patentes, derechos de propiedad intelectual. Ninguna de las partes tendrá como objetivo recuperar información confidencial de la otra parte como resultado de la compartición de datos.

6. Incluir normativa de **responsabilidad por la provisión** de datos erróneos, interrupciones en la transmisión de datos, trabajo interpretativo de baja calidad, o por destrucción/pérdida o alteración de datos (si es ilegal o accidental) que puedan causar daños.

7. Definir los derechos de ambas partes para realizar **auditorías sobre el cumplimiento de las obligaciones mutuas.**

8. **¿Cuál es la duración prevista del contrato? ¿Qué derechos para rescindir el contrato? ¿Qué aviso dar a sus socios?**

9. **Acordar la legislación aplicable y los mecanismos de aplicación ante incumplimientos.**

5.3.2 Acuerdos contractuales en el modelo de relación B2G[38]

A continuación, se detallan algunas consideraciones que pueden ayudar a las empresas privadas y organismos públicos en las iniciativas de compartición de datos:

1. Los organismos públicos deben **identificar un propósito de interés público**, los datos del sector privado necesarios para el propósito en cuestión y el nivel de granularidad necesario. Las empresas, por su parte, también pueden reflejar que sus datos pueden contribuir a un propósito de interés público.

2. Las partes deberían **identificar los objetivos internos y limitaciones** a las que se enfrentan relacionadas con la compartición de datos:

- Inversión en ampliar el conocimiento en gobernanza de datos (necesidades de concienciación y formación del equipo, si fueran necesarias).

- Las organizaciones que disponen de departamentos corporativos responsables de la compartición de datos encontrarán que compartir datos con organismos públicos (o igualmente con empresa privada) es menos costoso en términos de gobernanza de datos, infraestructura, legalidad y aspectos legales.

- Las empresas y los organismos del sector público deben garantizar el cumplimiento de la legislación vigente en términos de protección de datos GDPR (garantizar la legalidad del procesamiento, el consentimiento, el uso adecuado de técnicas de anonimización, la confidencialidad y protección de datos desde el diseño y por defecto, etc.). y de privacidad ePrivacy (propuesta de reglamento no aprobada a fecha de publicación de la presente guía).

- Los organismos del sector público deben realizar un análisis cuidadoso de las posibles fuentes de datos y determinar las limitaciones de un proveedor de datos específico.

3. Las partes deben elegir las **modalidades técnicas y prácticas de compartición de datos** que mejor se adapten a sus procesos internos y de gobernanza de datos:

- Los organismos públicos deben salvaguardar los intereses comerciales legítimos de la empresa privada (ej. información comercial confidencial, secretos comerciales) y garantizar la seguridad de acceso a los datos. Los datos del sector privado que se transfieren a un organismo del sector público deben tratarse como datos confidenciales. Es necesario explicitarlo y deben existir medidas adecuadas para garantizar la seguridad de la red y los sistemas de información.

[38] https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52018SC0125&rid=2

- Los organismos públicos podrían necesitar ampliar sus capacidades técnicas y de personal para aprovechar las posibilidades que brindan datos del sector privado.

4. El contrato debe **incluir las condiciones para la implementación**, las limitaciones de tiempo y los conjuntos de datos que serán utilizados:

- Los organismos públicos deben garantizar que su solicitud de datos privados cumpla con el principio de proporcionalidad y sea necesaria para lograr el objetivo de interés público definido. El acuerdo debe especificar que después de que se haya logrado el propósito, o bien se haya alcanzado el límite de la duración, los datos transmitidos se borrarán. El uso de los mismos datos para un propósito diferente debe estar sujeto a un acuerdo de colaboración nuevo o modificado.

- Las partes deben definir las condiciones a nivel operacional para la transferencia de datos: formato de los datos y metadatos, calidad, granularidad y duración del acceso y modo de acceso.

- Las partes deben determinar la compensación por la compartición de datos. A este respecto, existen diferentes opciones:

    - Limitar la compensación a una recuperación proporcional de los costos incurridos en la producción, preservación y difusión de los datos.

    - Limitar la compensación a los costos relacionados con la difusión de los datos, considerando que los costos de producción y la preservación de los datos ya han sido cubiertos por otras fuentes de ingresos.

      La elección de una opción u otra podría vincularse con el objetivo de interés público definido.

- Con el fin de permitir que los organismos públicos realicen la evaluación de calidad de los datos, las empresas que proveen los datos deben ofrecer su apoyo. Las empresas permitirán la evaluación de la calidad de sus datos para los fines declarados, incluso en los casos en los que se requiera realizar una auditoría o verificación formal.

5. Las partes deben acordar el **uso de principios comunes** para el seguimiento de la implementación del contrato:

- Se puede acordar un código de conducta o usar las reglas éticas existentes, (como las definidas en el "*European Statistics Code of Practice*"[39]), definir un comité de coordinación o nombrar un auditor independiente para supervisar el uso de los datos.

- Los organismos públicos deben implementar las salvaguardas necesarias que eviten el uso indebido de los datos, para otros fines distintos a los definidos en el contrato.

6. Incluir **normativa de responsabilidad** por la provisión de datos erróneos, interrupciones en la transmisión de datos, trabajo interpretativo de baja calidad, o por destrucción/pérdida o alteración de datos (si es ilegal o accidental) que puedan causar daños.

[39] http://ec.europa.eu/eurostat/web/products-manuals-and-guidelines/-/KS-32-11-955

7. Acordar la **legislación aplicable** y los mecanismos de aplicación ante incumplimientos. Cualquier parte debe tener la libertad de rescindir el contrato cuando exista un riesgo legal o técnico en relación con el tratamiento o el uso de los datos compartidos.

8. Los organismos públicos deben **difundir los resultados de la colaboración y garantizar mecanismos de retroalimentación** pública, cuando sea necesario o relevante, sin comprometer la confidencialidad de los datos del sector privado.

5.3.3 Licencias para la Compartición/Apertura de Datos

Como se ha comentado anteriormente, es fundamental que cualquier dato compartido/abierto incluya información explícita sobre las condiciones de uso con el fin de establecer cuáles son las condiciones de reutilización de esos datos y en qué contexto se pueden reutilizar y cuando no.

Condiciones de Uso Simples (apertura de datos)

Los datos están plenamente disponibles para su reutilización bajo una serie de condiciones mínimas, para las cuales se puede tomar como referencia el Real Decreto 1495/2011, de 24 de octubre[40]:

- Queda prohibido desnaturalizar el sentido de los datos.
- La fuente de los datos debe ser citada.
- La fecha de la última actualización debe mencionarse.
- No se podrá hacer alusión a que los titulares de los datos participan, patrocinan o apoyan la reutilización de los datos.
- Los metadatos deben ser mantenidos y no alterados, así como las condiciones de reutilización aplicables.

Licenciamiento predefinido

En este otro caso se indicará que los datos están disponibles para su reutilización bajo las condiciones establecidas por alguna de las licencias compatibles con los principios de los datos abiertos. Estas licencias podrán a lo sumo establecer restricciones relacionadas con la atribución de la fuente original y la obligatoriedad de compartir cualquier trabajo realizado con los datos bajo las mismas condiciones. A continuación, se detallan algunas licencias de referencia[41]:

[40] https://www.boe.es/diario_boe/txt.php?id=BOE-A-2011-17560

[41] https://help.data.world/hc/en-us/articles/115006114287-Common-license-types-for-datasets

Licencias de dominio público, nos permiten compartir los datos sin ningún tipo de restricción	▪ Creative Commons Zero (CC Zero)[42] ▪ Public Domain Dedication and License (PDDL 1.0)[43]
Licencias que únicamente obligan a atribuir la fuente original de los datos	▪ **Creative Commons-Reconocimiento** (CC BY 4.0)[44] ▪ **Open Data Commons Attribution License 1.0** (ODC-By 1.0)[45]
Licencias que requieren atribución de la fuente original y compartir los resultados bajo las mismas condiciones	▪ ***Creative Commons*-Reconocimiento Compartir Igual** (CC BY-SA 4.0)[46] ▪ ***Open Data Commons Open Database License*** (ODbL 1.0)[47] ▪ ***Community Data License Agreement – Permissive,*** (CDLA-Permissive-1.0)[48] ▪ ***Community Data License Agreement – Sharing*, Version 1.0** (CDLA Sharing)[49] ▪ ***Creative Commons Attribution-NonCommercial 4.0 International*** (CC BY-NC)[50] ▪ ***Creative Commons Attribution-NoDerivatives 4.0 International*** (CC BY-ND)[51] ▪ ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International*** (CC BY-NC-SA)[52] ▪ ***Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International*** (CC BY-NC-ND)[53]

Tabla 8 – Licencias para compartir/abrir datos.

La Comisión Europea dispone de un asistente de licencias[54] que proporciona una descripción de las licencias cómo aplicarlas y cómo combinarlas (si fuera necesario).

Las condiciones de uso o licencia de los datos deben quedar claramente definidas para evitar problemas tales como:

- No conocer cuáles son las condiciones aplicables de reutilización por no tener ninguna indicación sobre ello.
- Continuar con la aplicación de derechos por defecto (ej. relacionados con los derechos de actor o de propiedad de las bases de datos) por no haber renunciado a ellos previamente.

[42] https://creativecommons.org/publicdomain/zero/1.0/deed.es

[43] https://opendatacommons.org/licenses/pddl/index.html

[44] https://creativecommons.org/licenses/by/4.0/deed.es

[45] https://opendatacommons.org/licenses/by/summary/

[46] https://creativecommons.org/licenses/by-sa/4.0/deed.es

[47] https://opendatacommons.org/licenses/odbl/summary/

[48] https://cdla.io/permissive-1-0/

[49] https://cdla.io/sharing-1-0/

[50] https://creativecommons.org/licenses/by-nc/4.0/

[51] https://creativecommons.org/licenses/by-nd/4.0/

[52] https://creativecommons.org/licenses/by-nc-sa/4.0/

[53] https://creativecommons.org/licenses/by-nc-nd/4.0/

[54] https://www.europeandataportal.eu/en/content/show-license

Disposiciones recomendadas para las licencias

Ámbito de Aplicación	Esta disposición debe definir el ámbito geográfico y temporal de los derechos cubiertos por el acuerdo de licencia, los tipos de derechos concedidos y la gama de reutilizaciones permitidas. Es aconsejable indicar explícitamente los derechos no cubiertos por la licencia y definir de manera amplia los tipos de derecho concedidos (derechos de autor, derechos sobre las bases de datos y derechos afines, etc.).
Atribución	Los requisitos de atribución deben quedar claramente definidos. Al menos se debe indicar la procedencia de los datos compartidos: ▪ Declaración que indique la procedencia de los datos. ▪ Un vínculo a la información relativa a la licencia.
Exenciones	En caso de que se ofrezcan conjuntos de datos reutilizables en conjunción con otros conjuntos no reutilizables), es conveniente indicar explícitamente qué conjuntos de datos no están cubiertos por la licencia.
Definiciones	Es aconsejable definir los principales términos de la licencia (licenciante, utilización, información, licenciatario, etc.) de manera concisa y en lenguaje lo más sencillo posible, y ajustándose a los de la Directiva y la legislación nacional de transposición.
Exclusión de Responsabilidad	Esta disposición debe ser utilizada (en la medida en que lo permita la legislación aplicable) para llamar la atención sobre el hecho de que el licenciante facilita la información «tal cual» y no asume ninguna responsabilidad sobre su exactitud o exhaustividad. Se debe indicar además el caso en el que el licenciante no esté en condiciones de garantizar un suministro y un acceso sostenibles a la información en cuestión.
Consecuencias por incumplimiento	Se deben especificar las consecuencias del incumplimiento de las condiciones de la licencia, en particular si incluyen la revocación automática e inmediata de los derechos del reutilizador.
Compatibilidad y versiones	Indicar otras licencias con las que es compatible la licencia, es decir, que puede reutilizarse conjuntamente la información procedente de diversas fuentes con arreglo a diferentes licencias compatibles siempre que se respete cualquiera de las licencias. Es importante mantener un sistema claro de gestión de versiones y fechas de las licencias y remitirse a él para indicar las actualizaciones.

Tabla 9 – Disposiciones para Licencias.

5.4 Roles, habilidades, desarrollo y gestión del conocimiento

Este apartado incluye actividades de soporte relacionadas con las nuevas responsabilidades[55] asociadas a la gobernanza de datos, así como a las habilidades y conocimientos[56] necesarios a llevar a cabo en una organización que desea compartir/publicar datos.

5.4.1. Roles

Existen varios actores involucrados a lo largo del ciclo de vida de una iniciativa de compartición de datos en una organización. Dependiendo del carácter de la actividad a desarrollar se requerirán de perfiles de estrategia, técnicos, etc. Típicamente se puede encontrar los siguientes perfiles en una iniciativa de compartición de datos:

- **Sponsor**: Es el patrocinador de la iniciativa de compartición de datos en la organización. El patrocinador validará el enfoque general, supervisará la implementación de la iniciativa de compartición de datos y, en última instancia, será el responsable de establecer la estrategia en la organización.

  Este perfil no tiene responsabilidad técnica sobre los datos, pero es quien dota a la organización de los recursos económicos, técnicos y humanos necesarios para que la iniciativa sea exitosa.

  El patrocinador tiene la visión clara de organización y es quién establece las políticas, la visión y misión y en última instancia la estrategia de la compartición de datos. Su compromiso hacia la compartición de datos debe ser claro y estable durante todo el proceso.

- **Gestor de Datos**: Es el responsable de la iniciativa de compartición de datos en la organización. Es la persona responsable de definir la estrategia de compartición de datos y es el responsable de todo el proceso. Este perfil agrupa el conocimiento de todo lo que concierne a la iniciativa desde los beneficios, las barreras, aspectos organizativos, técnicos, etc. El gestor de datos[57] tiene la responsabilidad de implementar la estrategia de compartición de datos con los recursos económicos, técnicos y humanos que el sponsor pone a su disposición. Conocer la legalidad vigente en torno a la compartición de datos y cumplirla es responsabilidad también de este perfil.

  Debe ser capaz de identificar situaciones de resistencia internas y de haberlas, debe saber cómo gestionarlas adecuadamente.

  Adicionalmente, este rol es el responsable de garantizar la calidad de los datos, realizando chequeos frecuentes sobre los mismos, garantizando que se cumplen los estándares de calidad definidos. En el caso de identificar no conformidades, es responsabilidad del Propietario de Datos solventarlas.

- **Implementador**: El implementador es generalmente responsable de la implementación de los requisitos técnicos. Por lo tanto, es necesario conocer las normas técnicas, las herramientas específicas y los requisitos organizativos básicos. El implementador puede ser un recurso interno o externo asignado por el director de datos. Ambos perfiles deben trabajar de manera conjunta en la iniciativa de compartición de datos.

[55] https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

[56] https://theodi.org/article/open-data-maturity-model/

[57] https://medium.com/data-stewards-network/the-three-goals-and-five-functions-of-data-stewards-60242449f378

- **Propietario de Datos:** Cada campo de datos en cada base de datos de la organización debe ser propiedad de un propietario de los datos, quien tiene la autoridad para decidir en última instancia sobre el acceso y el uso de los datos.

  Los roles y responsabilidades deben estar alineados con la gobernanza de datos en vigor en la organización, siendo ésta a su vez uno de los pilares fundamentales para el despliegue de la estrategia de datos junto con los procesos. Sin embargo, dependiendo de la madurez organizacional, tamaño de la organización y modelo de gestión de datos (centralizado vs. descentralizado), éstos roles o responsabilidades se asumirán o bien de manera puntual y no exclusiva o, con roles y responsabilidades asumidos por perfiles especializados en la temática.

5.4.2 Habilidades

Una organización madura deberá asegurar que el personal tiene la formación suficiente y el soporte adecuado para llevar a cabo sus responsabilidades respecto a la gestión de datos y debe ayudar a desarrollar habilidades en todos los niveles organizativos, dando soporte tanto a las necesidades operacionales como a las estratégicas.

Una organización orientada a la compartición/apertura de datos debe comprender los beneficios de la compartición/apertura y la transparencia y aplicar esos principios adecuadamente. Como apoyo al crecimiento de esta cultura en la organización, se debe garantizar que las personas dispongan de los conocimientos y experiencia necesarias en distintos ámbitos:

- Conocimiento compartido del valor de los datos y su aplicación en la organización.

- Conocimiento operativo para la gobernanza de los datos y la implementación de los canales de compartición de los datos.

- Conocimiento estratégico a nivel de dirección para definir objetivos futuros relacionados con la compartición de datos.

La organización apoyará a las personas a desarrollar las habilidades necesarias para poder poner en marcha la estrategia de compartición de datos. Si fuera necesario, facilitará a las personas la formación necesaria en aspectos técnicos, operativos, contractuales, de gobernanza de los datos, etc.

El acceso a la información también es un componente importante en el proceso de generar una cultura orientada a los datos. La organización deberá asegurar que todos los estándares organizacionales, conjuntos de datos y cualquier documentación relevante está accesible a quién lo necesite.

5.4.3 Desarrollo del conocimiento

La organización debe asegurar que el personal tenga la capacitación y el apoyo necesarios para cumplir con las responsabilidades individuales relacionadas con la compartición de datos. Este es probablemente el factor clave de éxito para concienciar a las personas sobre los beneficios de una buena gobernanza de datos y su aplicación en la estrategia de la organización.

Es relevante que las personas estén formadas para asegurar que disponen de los conocimientos necesarios para:

- Entender los riesgos y beneficios relacionados con la compartición de datos y como afectan a sus responsabilidades.
- Entender la política y estrategia de compartición de datos en la organización.
- Tener la capacidad de aplicar el nivel adecuado de gobernanza de datos en los proyectos.
- Identificar la formación necesaria en cualquier ámbito, por ejemplo, formación técnica, de normativa, estándares, etc.

5.4.4 Gestión del conocimiento

La gestión del conocimiento es un proceso imprescindible para cualquier organización. Desde la perspectiva de la compartición de datos, la gestión del conocimiento es relevante por dos motivos:

1. Las personas necesitan localizar y utilizar la documentación relacionada con los estándares y políticas de compartición de datos. Esta información también es probablemente relevante para los reutilizadores de los datos. La política de datos definirá el compromiso de la organización respecto a la compartición/apertura de datos y cómo enganchar a los reutilizadores.
2. En segundo lugar, tanto los usuarios internos como los externos a la organización deben tener acceso a la documentación necesaria que dé soporte al uso de datos compartidos. Para que los datos conserven su valor a largo del tiempo, las personas que los usan deben saber cómo están recopilados, cómo se procesan y quién los posee.

6. GLOSARIO EXPLICATIVO DE TÉRMINOS Y SIGLAS

API (*Application Programing Interface,* Interfaz de Programación de Aplicaciones). Conjunto de reglas (código) y especificaciones que las aplicaciones pueden utilizar para comunicarse entre ellas.

Anonimización. Proceso de convertir los datos en una forma en la que no se pueda identificar a individuos concretos.

Apps. Aplicación o programa que puede ser instalada en dispositivos móviles y ordenadores para que el usuario realice distintos tipos de tareas, de forma dinámica y específica.

B2B (*Business to Business*). Modelos de negocio en los que las transacciones de bienes o la prestación de servicios se producen entre dos empresas.

B2G (*Business to Government*). Empresas que venden u ofrecen sus servicios a las instituciones del gobierno.

Big Data. Cualquier cantidad voluminosa de datos estructurados, semiestructurados y no estructurados que tienen el potencial de ser extraídos para obtener información.

Blockchain. Estructura de datos en la que la información contenida se agrupa en conjuntos (bloques) a los que se le añade metainformaci**ón** relativa a otro bloque de la cadena anterior, de manera que, gracias a técnicas criptográficas, la información contenida en un bloque solo puede ser repudiada o editada modificando todos los bloques posteriores.

Business Intelligence. Habilidad para transformar los datos en información, y la información en conocimiento, para apoyar los procesos de toma de decisión en las organizaciones.

CKAN. Solución software que hace los datos accesibles y reutilizables al proveer herramientas para publicar, compartir, encontrar y usar los datos (incluyendo almacenamiento de datos y provisión de APIs de datos robustas).

CMS (Sistema de Gestión de Contenidos). Software para administrar y gestionar contenidos de una web con facilidad.

Data Mining. (Minería de datos). Proceso de extracción de información de grandes bases de datos, información que revela inteligencia del negocio, a través de tendencias y correlaciones. De esta forma se facilita la generación de predicciones que ayuden a resolver problemas del negocio, pudiendo proporcionar una ventaja competitiva.

Data Science. Campo interdisciplinario que involucra métodos científicos, procesos y sistemas para extraer conocimiento o un mejor entendimiento de datos en sus diferentes formas, ya sea estructurados o no.

Dataset (Conjunto de Datos). Colección de datos, habitualmente tabulada.

Datatón. Jornada o reunión donde se demuestran los beneficios que tiene la apertura y análisis de datos y de grandes volúmenes de información. Se concibe como un maratón de datos donde se convocan distintos agentes interesados (técnicos, ciudadanos, investigadores) y se pone a su disposición distintos conjuntos de datos.

Datawarehouse. Base de datos corporativa que se caracteriza por integrar y depurar información de una o más fuentes.

DKAN. Plataforma de código abierto basada en Drupal que permite a las Administraciones Públicas publicar datos, proporcionar visualizaciones e historias de datos y crear paneles internos de control analítico.

DLTs (*Distributed Ledger Technologies*). Sistema en el que los registros o los datos son compartidos, accesibles y usados por todos los miembros de una red.

FAIR (*Findable, Accessible, Interoperable and Reusable*). Principios (no estándares) definidos inicialmente por la comunidad de datos de investigación, para hacer que éstos sean encontrables, accesibles, interoperables y reutilizables. Estos principios son aplicables a cualquier tipo de *dataset* e implican la calidad de los datos para su reutilización.

IA (Inteligencia Artificial, también AI, *Artificial Intelligence,* por sus siglas en inglés). Simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos.

Industria 4.0. Fabricación informatizada con todos los procesos interconectados por Internet de las Cosas (IoT).

IoT (Internet de las Cosas). Sistema de dispositivos conectados entre sí a través de una red con capacidad de transmitir información sin necesidad de la interacción humana.

Linked Data. Método de publicación estructurada de los datos (Ver Anexo IV).

Machine-Readable. Datos en un formato tal que puede ser procesado fácilmente por una computadora.

Marketplace. El sitio en Internet donde se llevan a cabo interacciones comerciales entre diferentes empresas.

Metadata. Datos que describen otros datos. Describen el contenido, calidad, condiciones, historia, disponibilidad y otros aspectos descriptivos, estructurales o administrativos de los datos.

ODS (*Operational Data Store*). Contenedor datos activos, diseñado para integrar datos de múltiples fuentes con los que realizar operaciones adicionales en los propios datos.

OLAP. Solución para agilizar la consulta de grandes cantidades de datos en el contexto del *Business Intelligence*.

OLTP (*Online Transaction Processing*). Herramienta para el procesamiento, administración y mantenimiento diario de transacciones generadas por una compañía para garantizar altos niveles de disponibilidad, seguridad y confiabilidad.

PIWIK (Matomo). Herramienta de analítica web perteneciente a una comunidad *OpenSource* cuyas características le permiten competir con la mayoría de las opciones disponibles en el mercado.

Reutilizador. Persona o entidad que hace un uso de los datos abiertos o compartidos por un tercero. El uso de los datos puede ser igual o diferente a aquel para el que fueron concebidos.

ROI (*Return of Investment,* Retorno de la Inversión). Índice financiero que mide y compara el beneficio o la utilidad obtenida en relación a la inversión realizada.

RGPD (Reglamento General de Protección de Datos, también GDPR, por sus siglas en inglés). Es el reglamento europeo relativo a la protección de las personas físicas en lo que respecta al tratamiento de sus datos personales y a la libre circulación de estos datos.

Smart Meter. Medidor de consumibles (como por ejemplo gas, agua o electricidad) el cual es capaz de transmitir y recibir información. Es capaz de comunicarse con otros dispositivos utilizando una red de área local.

SPARQL (*Protocol and RDF Query Language*). Lenguaje estandarizado para la consulta de grafos RDF, normalizado por el RDF *Data Access Working Group (DAWG) del World Wide Web Consortium (W3C)*.

Utilities. Compañías que ofrecen servicios públicos de consumibles como electricidad, gas, agua, etc.

REFERENCIAS Y BIBLIOGRAFÍA

Abella, A., Ortiz-de-Urbina-Criado, M., De Pablos-Heredero, C. (2014a). "Melodа, métrica para evaluar la reutilización de datos abiertos". *El profesional de la información*, Nov-Dec 2014, v. 23, n. 6.

Abella, A.; Ortiz-de-Urbina-Criado, M.; De-Pablos-Heredero, C. (2014b): Melodа, a metric to assess open data reuse, *El profesional de la información*, 23/6, noviembre-diciembre, 582-588.

Berends, J.; Carrara, W. y H. Vollers (2017). Barriers in working with Open Data, Informe Analítico nº 5- Portal de Datos Europeo, Comisión Europea, Bruselas.

British Ecological Society (2018): *Guide to Data Management in Ecology and Evolution*, Londres.

Comisión Europea (2014): *Directrices sobre las licencias normalizadas recomendadas, los conjuntos de datos y el cobro por la reutilización de los documentos*, Comunicación de la Comisión 2014/C 240/01, Diario Oficial de la Unión Europea.

Comisión Europea (2015): *Creating Value through Open Data, Portal de Datos Europeo*, Comisión Europea, Bruselas.

Comisión Europea (2016): *Study on data sharing between companies in Europe, Informe Final*, Comisión Europea, Bruselas.

Comisión Europea (2017): *Open Data Goldbook for Data Managers and Data Holders-Practical guidebook for organisations wanting to publish Open Data*, Portal de Datos Europeo, Comisión Europea, Bruselas.

Comisión Europea (2017): *recommendations for Open Data Portals: from setup to sustainaibility,* Portal de Datos Europeo, Comisión Europea, Bruselas.

Comisión Europea (2018): *Open Data Goldbook for Data Managers and Data Holders-Practical guidebook for organisations wanting to publish Open Data*, Portal de Datos Europeo, Comisión Europea, Bruselas.

Comisión Europea (2018): *Directrices para el intercambio de datos del sector privado en la economía de datos europea*, Comisión Europea, Bruselas.

Chesbrough, H. (2003). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business School Press.

Gobierno de España, Red.es (22 de octubre de 2014): *Decálogo del Reutilizador de Datos Abiertos del Sector Público,* [entrada en noticias]. Recuperado de https://datos.gob.es/es/noticia/decalogo-del-reutilizador-de-datos-abiertos-del-sector-publico.

Red.es (21 de diciembre de 2017): *Guía para publicar datos abiertos de manera rápida y sencilla (con CKAN)* [entrada en noticias]. Recuperado de https://datos.gob.es/es/documentacion/guia-para-publicar-datos-abiertos-de-manera-rapida-y-sencilla-con-ckan.

European Open Data Portal (2018). *Open Data Goldbook for Data Managers and Data Holders Practical guidebook for organisations wanting to publish Open Data.* European Open Data Portal, January 2018. https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

Federación Española de Municipios y Provincias (2017): *Datos Abiertos-Guía estratégica para su puesta en marcha – Conjunto de datos mínimos a publicar*, Wolters Kluwer España.

Gobierno de España; Red.es (2016): *Guía metodológica para planes open data sectoriales- agosto 2016*. Madrid: Ministerio de Industria, Energía y Turismo, Ministerio de Hacienda y AAPP y Red.es. Recuperado de: https://datos.gob.es/sites/default/files/actualizacion_informe_iniciativa_sectorial_0.pdf

Hapuarhachchi, T.; Balley, B.; Dodds, L. y A. Fletcher (2016): *Shareable by default-creating resilient data ecosystems*, Open Data Institute- Thomson Reuters.

Lewis, J. (2017). *2018 Planning Guide for Data Analytics*, Gartner.

Open Data Institute (2015): *A guide to the Open Data Maturity Model- Assessing your open data.*

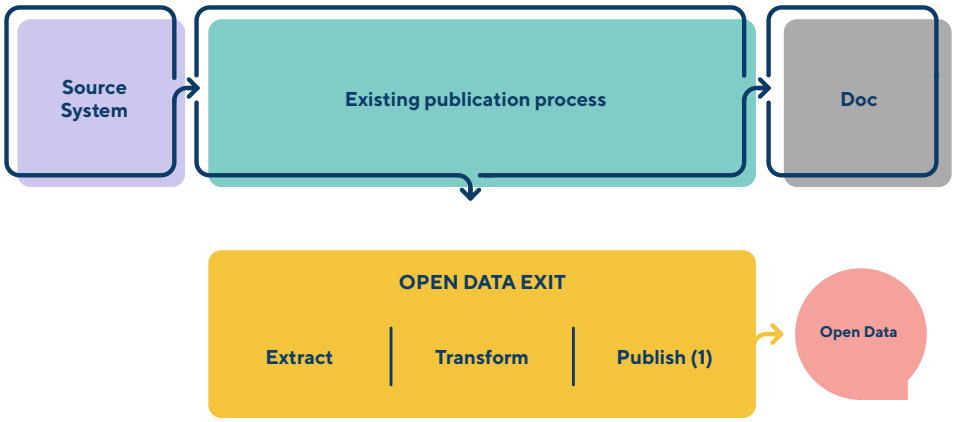
Secretaría de Estado de Telecomunicaciones y para la Sociedad de la Información (2009): *Guía Aporta sobre reutilización de la información del sector público-Proyecto Aporta.*

Servicio de Datos del Reino Unido (2015): *Depositing shareable survey data*.

Williams, C. (5 de abril de 2018): The future of Data Portals, [entrada en blog]. Recuperado de https://medium.com/vizzuality-blog/the-future-of-data-portals-1a9a20b20164.

ANEXO I: ESCENARIOS DEL PROCESO ETCD (EXTRACCIÓN Y TRANSFORMACIÓN)[58]

ESCENARIO 1
Organización que publica datos en informes periódicos



ETP-Process with existing publication

Una organización (ej. Informes de Gartner) **publica informes** (documentos en Doc/Pdf/ etc.) que contienen datos que provienen de distintas fuentes (internas o externas) listas para publicar. Los datos recogidos para la publicación de los informes pueden transformarse en un flujo de datos actualizados compartidos por terceros y/o abiertos:

- Datos existentes en publicaciones anteriores: se adaptará el proceso de publicación existente para permitir crear un flujo de datos actualizados para compartir/publicar con las frecuencias estipuladas.
- Nuevos informes: se añadirán entradas de datos adicionales al proceso de compartición/publicación de los nuevos flujos de datos.
- Datos Abiertos: Los datos se deben someten al proceso de publicación de datos abiertos (ej. definir los metadatos, publicar el *dataset* o flujo de datos de manera preferiblemente automática, elegir el modelo de licencia, transformar el *dataset* usando diferentes formatos, asegurar actualización periódica de los datos, etc.).

ESCENARIO 2
La organización publica datos en distintos formatos

Descripción: La organización ya publica *datasets* o flujos de datos en distintos formatos a través de sus canales de difusión (ej. la web). Estos datos no siempre cumplen los requerimientos de los datos compartidos ni de datos abiertos. Sin embargo, con muy poco esfuerzo, se pueden adaptar los datos para ser compartidos o publicarlos como datos abiertos.

Extracción
Aislar y filtrar los datos en un *dataset* uniforme. Podemos también ofrecer vistas filtradas del *dataset* (menos campos, anonimizados, agrupados) según se trate de un proceso de compartición o publicación de datos en abierto.

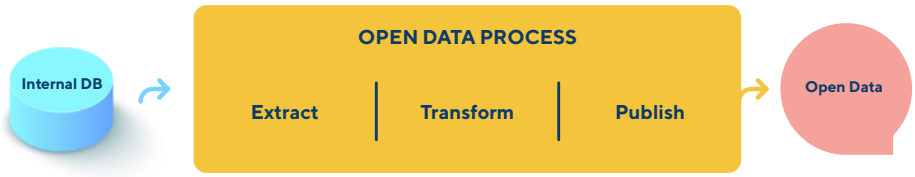
[58] Los escenarios descritos en este anexo son adaptaciones y, en algún caso los mismos, del Appendix 3 del documento https://www.european-dataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

Datos compartidos/abiertos: Revisar los procesos existentes para asegurarse que cumplen todos los requisitos legales para poder compartir con terceros o publicarlos como datos abiertos. Los segundos más restrictivos que los primeros.

Transformar
Aplicar el proceso de preparación para transformar los datos publicados en datos listos para compartir/abrir o *linked data*.

Para la publicación de los datos en abierto, éstos deben de someterse al proceso de publicación de la organización, al igual que en el escenario 1.

ESCENARIO 3
Los datos a publicar/compartir están en base de datos



The ETP-Process starting from an existing database

Descripción: Extraer datos de una Base de Datos y transformarla en un flujo de datos a compartir con terceros o publicar en abierto. Este escenario es distinto de los anteriores y asume que la base de datos es estándar (ej. Oracle, SQL server) y la organización es propietaria de esta o, en cualquier caso, puede aplicar las técnicas de extracción y transformación, sin ningún impedimento por parte de terceros, a las tablas gestionadas por la base de datos.

Este es un caso típico que requiere aplicar técnicas de extracción y transformación.

¿Por qué es necesario extraer datos a un fichero? Porque los datos de la base de datos (datos en origen) están estructurados para que resulte óptimo su acceso para las aplicaciones operativas de la organización y no para ser compartidos por terceros o publicarlos como datos abiertos. Esta estructura se llama OLTP (on-line transaction processing) y es de naturaleza relacional (los datos están ligados a través de relaciones (claves) que permiten relacionar los datos entre distintas tablas.

Si los datos están en bases de datos distintas, es preferible crear un **mapa lógico de datos** (*dataset* destino) que se abstraiga de las relaciones físicas existentes en las bases de datos de origen y aplicar las transformaciones oportunas (ej. sustituir identificadores por valores, hacer que las referencias sean consistentes – 0 = hombre; 1=Mujer -, etc.) hasta obtener el *dataset* correspondiente al mapa lógico de datos. En cualquier caso, muy raramente se pueden compartir o abrir los datos tal cual están en la base de datos de partida.

Si además los datos están en distintos formatos fuente (ej. bases de datos, ficheros planos, etc.), el escenario sería el siguiente (escenario 4).

Extraer

La mayoría de los sistemas de bases de datos (ej. ORACLE, SQL Server) disponen de técnicas estándar para leer de tablas y escribir en ficheros planos. Por ejemplo:

- Oracle: vía EXPORT tool.
- Microsfot SQL Server: via mysqldump tool.
- ProgressSQL: vía SQL Dump procedure.

Sin embargo, cuando los datos de partida cambian con frecuencia, una buena alternativa es escribir programas que lean los datos directamente del Sistema de gestión de datos (DBMS) vía ODBC – JDBC. ODBC (o SQL directo) permite programar extracciones más complejas.

Finalmente, existe también la posibilidad de utilizar herramientas de ETC(ETL).

Desde un punto de vista práctico, en caso de que sea necesario hacer transformaciones complejas, es recomendable almacenar los datos en una base de datos intermedia, temporal o permanente, de la que se obtendrá el flujo de datos final (a compartir o publicar en abierto). En este caso, hay que tener en cuenta las dos dimensiones del flujo de datos:

- Exportar todo el contenido de la base de datos para que sea compartido o publicado en abierto.
- Descargar las modificaciones/delta con respecto a la versión previa y combinarlas con el flujo de datos.

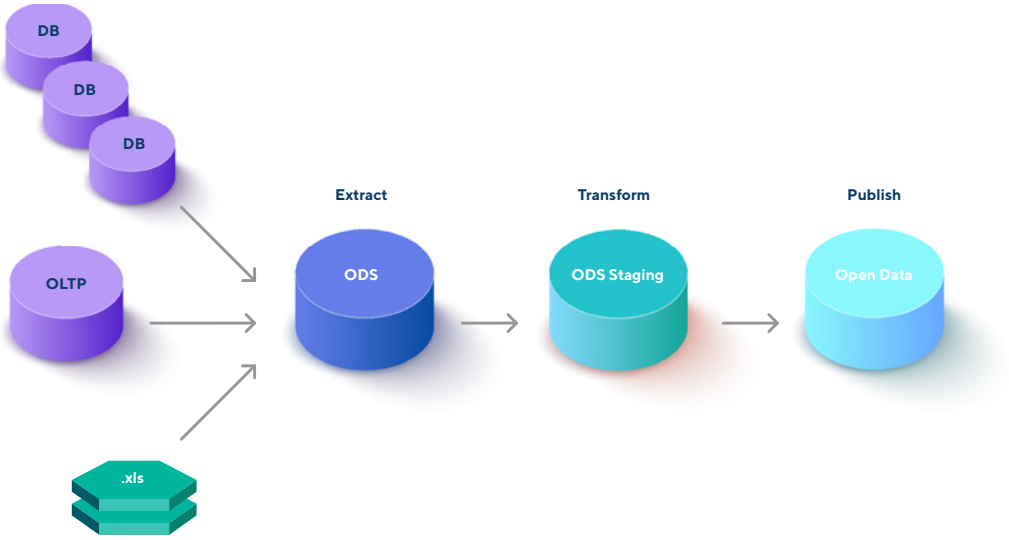
Transformar

En esta fase se abordan, como en el caso del entorno del datawarehouse, la revisión de la calidad de los datos. Por ejemplo, utilizar los mismos nombres de campos para los mismos datos, transformar las abreviaturas en nombres completos, guardar los datos de manera consistente, etc. Los procesos de anonimización, agrupación de datos para conseguir el mismo nivel de granularidad, etc. se realizan en este paso.

> *Recomendación: en caso de que los datos cambien con mucha rapidez, es preferible poner los datos a disposición de terceros lo antes posible en lugar de perder tiempo en transformaciones que no son estrictamente necesarios para el propósito en cuestión.*

Para la publicación de los datos en abierto, éstos deben de someterse al proceso de publicación de la organización, al igual que en el escenario 1.

ESCENARIO 4
Organizaciones con datos consolidados extraídos de distintas fuentes



Descripción: En este escenario se asume que la organización dispone de la experiencia y herramientas de ETC(ETL) necesarias. Todas las técnicas de extracción y transformación que se aplican para el datawarehouse o Business Intelligence son válidas para los procesos de ETCD o ETDA.

En un proceso de ETC(ETL) los datos de origen se extraen de distintas fuentes y se almacenan de manera consistente en un el ODS – Operational Data Store -. Los datos están además disponibles al más bajo nivel de granularidad. Los datos del ODS se procesan en distintas fases que requieren almacenar los datos en tablas intermedias. Posteriormente se agregan los datos y se refinan hasta que sean consistentes antes de guardarlos en el datawarehouse. En el caso de datos compartidos o datos a publicar en abierto, los pasos y tablas intermedias de transformación pueden ser otros.

Procesos ETCD o ETDA no es necesario extraer los datos de manera separada, podemos partir directamente del ODS. En el proceso de transformación, los flujos de datos pueden provenir de los procesos de ETC. En algunos casos se necesitará modificar alguna fase del ETL o sustituir una fase por otra. Se recomienda reutilizar la mayor parte del proceso ETC. No es necesario ninguna herramienta adicional para estos procesos.

Para la publicación de los datos en abierto, éstos deben de someterse al proceso de publicación de la organización, al igual que en el escenario 1.

ANEXO II: PRINCIPIOS QUE GUÍAN LOS MODELOS DE RELACIÓN

Modelos de negocio B2B

Transparencia	Los acuerdos contractuales deben identificar de manera transparente y clara (i) las personas o entidades que tendrán acceso a los datos que genera el producto o servicio, el tipo y detalle de los datos; y (ii) el propósito de uso de dichos datos.
Protección de los Intereses Comerciales	Los acuerdos contractuales deben proteger tanto los intereses comerciales como la información de titularidad de los datos y usuarios de estos.
Portabilidad	Las empresas que ofrecen un producto o servicio que genera datos como subproducto deben permitir la máxima portabilidad posible de sus datos.
Valor Compartido	Los acuerdos contractuales deben reconocer que, cuando los datos se generan como un producto derivado del uso de un producto o servicio, varias partes han contribuido a la creación de los datos.
Competencia Leal	Los acuerdos contractuales deben garantizar una competencia leal al intercambiar datos comerciales sensibles.

Tabla 10 – Principios Modelo de Negocio B2B

Modelos de negocio B2G

Proporcionalidad de Uso	Las solicitudes de provisión de datos del sector privado en condiciones preferenciales para su reutilización deben justificarse por un interés público claro y demostrable. La solicitud de datos del sector privado debe ser adecuada y relevante para el propósito de interés público. El costo y el esfuerzo requeridos para la provisión y la reutilización de los datos del sector privado deben ser razonables en comparación con los beneficios públicos previstos.
Propósito Limitado	El uso de los datos del sector privado debe estar claramente limitado para uno o varios propósitos que se especifiquen lo más claramente posible en los contratos de colaboración entre empresas y gobiernos. La empresa del sector privado debe recibir garantías específicas de que los datos obtenidos no se utilizarán para procedimientos administrativos o judiciales no relacionados.
Protección del Secreto Comercial	Se debe garantizar que se respeten los intereses legítimos, en particular la protección de los secretos comerciales. La provisión de datos de empresa a gobierno debe permitir a las empresas continuar monetizando los datos en cuestión.
Condiciones de Reutilización	Los acuerdos contractuales deben tratar de ser mutuamente beneficiosos al tiempo que reconozcan el objetivo de interés público al otorgar al organismo del sector público un trato preferencial sobre otros clientes. El nivel de compensación acordado podría estar vinculado al propósito de interés público esperado. Los acuerdos de colaboración de datos entre empresa y gobierno que involucran a los mismos organismos públicos que realizan las mismas funciones, deben tratarse de manera no discriminatoria.

Apoyo Razonable	Los organismos públicos que se benefician de la provisión de datos del sector privado deben ofrecer un apoyo razonable y proporcionado a las empresas que suministran los datos para ayudar por ejemplo a evaluar la calidad de los datos para los fines declarados, incluyendo la posibilidad de auditar o verificar los datos. No se debe exigir a las empresas que inviertan en mejorar la calidad de los datos
Transparencia y Participación Social	La colaboración entre empresas y organismos públicos debe ser transparente. Las mejores prácticas de colaboración entre empresas y gobiernos deben ponerse a disposición del público siempre y cuando no comprometan la confidencialidad de los datos.

Tabla 11 – Principios Modelo de Negocio B2G

ANEXO III: LEGISLACIÓN APLICABLE

GDPR (Reglamento General de Protección de Datos)

El Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo[59], el nuevo Reglamento general de protección de datos («RGPD») regula el tratamiento que realizan personas, empresas u organizaciones de los datos personales relacionados con personas en la Unión Europea (UE).

Este reglamento se actualiza posteriormente en la versión corregida del 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 la Directiva 95/46/CE (Reglamento general de protección de datos) (DO L 119 de 4.5.2016)[60]

A partir del 25 de mayo de 2018, con la entrada en vigor del Reglamento general de protección de datos, hay un único conjunto de normas de protección de datos para todas las empresas que operan en la Unión Europea (UE), con independencia de dónde tengan su sede.

Las normas más estrictas en materia de protección de datos implican que:

- Las personas tienen más control sobre sus datos personales.
- Las empresas se benefician de igualdad de condiciones.

[59] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679&from=EN

[60] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32016R0679R(02)&from=ES

[61] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations_es

A continuación, se detallan algunos aspectos que una organización debe tener en cuenta para el cumplimiento de la normativa[61]:

APLICACIÓN DEL REGLAMENTO	
1.	Tratamiento de datos personales como parte de las actividades de una de sus sucursales establecidas en la Unión Europea (UE), independientemente del lugar donde sean tratados los datos.
2.	La organización está establecida fuera de la UE y ofrece productos o servicios (de pago o gratuitos) u observa el comportamiento de las personas en la UE.
3.	No depende del tamaño de la empresa, sino de la naturaleza de sus actividades.
4.	Las normas se aplican únicamente a los datos de personas, no rigen los datos sobre empresas ni ninguna otra persona jurídica.

Tabla 12 – Aplicación del Reglamento

PRINCIPIOS DEL REGLAMENTO	
1.	El tipo y la cantidad de datos personales que pueden tratar dependen de las razones del tratamiento (razón jurídica usada) y lo que se quiera hacer con ellos: ▪ los datos personales deben tratarse de forma lícita y transparente ▪ se deben tener fines específicos para el tratamiento de los datos e indicar dichos fines a las personas al recopilar sus datos personales ▪ solo se deben recopilar y tratar los datos personales que sean necesarios para cumplir esa finalidad ▪ se debe garantizar que los datos personales sean exactos y estén actualizados ▪ se debe garantizar que los datos personales no se conserven más tiempo del necesario para los fines para los que fueron recopilados ▪ se deben establecer garantías técnicas y organizativas apropiadas que garanticen la seguridad de los datos personales.
2.	Principio de la limitación de la finalidad, se debe saber para qué fin se quiere tratar los datos personales e informar a las personas de las que se traten datos.
3.	Se pueden utilizar los datos para otros fines distintos de los inicialmente definidos, aunque sólo en los casos en los que se hayan recopilado datos basándose en un interés legítimo, un contrato o intereses vitales, y sólo después de comprobar que el nuevo fin sea compatible con el fin original: ▪ la relación entre el fin inicial y el nuevo o futuro fin, ▪ el contexto en que se recopilaron los datos (¿cuál es la relación entre la empresa y la persona?), ▪ el tipo y la naturaleza de los datos (¿son sensibles?), ▪ las posibles consecuencias del tratamiento ulterior (¿cómo afectará a la persona?), ▪ la existencia de garantías adecuadas (como cifrado o seudonimización).
4.	Los datos personales deben ser tratados únicamente cuando no sea razonablemente posible realizar el tratamiento de otra forma.
5.	Los datos deben conservarse durante el menor tiempo posible.

[61] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rules-business-and-organisations_es

6.	En el momento de recopilar los datos, se debe decir claramente a las personas de las que se estén recopilando datos, por lo menos lo siguiente: ▪ quiénes son (su información de contacto y la de su delegado de protección de datos, si lo tienen), ▪ para qué utilizarán sus datos personales (fines), ▪ las categorías de datos personales de que se trate, ▪ la justificación jurídica del tratamiento de los datos, ▪ durante cuánto tiempo conservarán los datos, ▪ quién más puede recibirlos, ▪ si los datos personales se transferirán a un destinatario de fuera de la UE, ▪ que tienen derecho a una copia de los datos (derecho de acceso a los datos personales) y otros derechos básicos del ámbito de la protección de datos (véase la lista completa de derechos[62]), ▪ su derecho a presentar una reclamación ante una autoridad de protección de datos, ▪ su derecho a retirar el consentimiento en todo momento, ▪ en su caso, la existencia de decisiones automatizadas y la lógica aplicada, además de las consecuencias que conlleva. Véase toda la lista de información que se debe proporcionar[63].

Tabla 13 – Principios del Reglamento

FUNDAMENTOS JURÍDICOS PARA EL TRATAMIENTO DE DATOS	
1.	Únicamente se pueden tratar datos en alguna de las circunstancias siguientes: ▪ cuando tengan el consentimiento de las personas en cuestión, ▪ cuando tengan una obligación contractual (un contrato entre ustedes y el cliente), ▪ para cumplir con una obligación legal (expuesta en la legislación de la Unión Europea o nacional), ▪ cuando el tratamiento sea necesario para el cumplimiento de una misión realizada en interés público (expuesto en la legislación de la Unión Europea o nacional), ▪ para proteger los intereses vitales del interesado, ▪ para satisfacer los intereses legítimos de su organización, pero solo tras haber comprobado que los derechos y libertades fundamentales de la persona cuyos datos estén tratando no se vean significativamente afectados.
2.	Cuando se necesite obtener consentimiento para tratar los datos personales, para que dicho consentimiento tenga validez debe cumplir las siguientes condiciones: ▪ darse libremente, ▪ ser informado, ▪ darse para un fin específico, ▪ establecerse claramente todos los motivos para el tratamiento, ▪ ser explícito y haberse dado a través de un acto positivo (como una casilla de selección electrónica que el interesado debe marcar explícitamente en línea o la firma en un formulario), ▪ utilizar un lenguaje claro y sencillo y ser fácilmente visible, ▪ explicarse y ofrecerse la oportunidad de retirar el consentimiento (por ejemplo, un enlace para darse de baja al final del correo de un boletín informativo electrónico).
3.	Debe ser tan fácil dar el consentimiento como retirarlo.
4.	Se permite cierta flexibilidad en relación con el nivel de especificación y de detalle del consentimiento en el contexto de la investigación científica.

Tabla 14 – Fundamentos Jurídicos para el Tratamiento de Datos.

[62] https://ec.europa.eu/info/law/law-topic/data-protection/reform/rights-citizens/my-rights_es

[63] https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX:32016R0679#d1e2313-1-1

OBLIGACIONES

1. Roles y Responsabilidades:

- El responsable del tratamiento determina los fines y los medios relacionados con el tratamiento de los datos personales, de modo que, si decide por qué y cómo deberán tratarse los datos personales, usted es el responsable del tratamiento.
- Corresponsable del tratamiento cuando, junto con una o más organizaciones, determine conjuntamente por qué y cómo deberán tratarse los datos personales. Los corresponsables del tratamiento deben concertar un acuerdo en el que se establezcan sus respectivas responsabilidades de cumplimiento con las normas del Reglamento General de Protección de Datos.
- El encargado del tratamiento trata los datos personales únicamente por cuenta del responsable del tratamiento. El encargado del tratamiento de los datos suele ser un tercero externo a la empresa; sin embargo, en el caso de los grupos de empresas, una de ellas puede actuar como encargada del tratamiento para otra.

2. Otra persona (física o jurídica o cualquier otro organismo) puede tratar datos personales por usted siempre que exista un contrato u otro acto jurídico.

3. El Reglamento general de protección de datos (RGPD) se basa en el enfoque basado en los riesgos, en otras palabras, se anima a las organizaciones que tratan datos personales a que apliquen las medidas de protección que sean acordes con el nivel de riesgo de sus actividades de tratamiento de datos.

4. Una violación de la seguridad de los datos se produce cuando los datos de los que ustedes son responsables sufren un incidente de seguridad que da lugar a la violación de la confidencialidad, disponibilidad o integridad de los datos.

5. Se exige una EIPD (Evaluación de Impacto relativa a la Protección de Datos) cuando el tratamiento pueda entrañar un alto riesgo para los derechos y libertades de las personas, por lo menos en los tres casos siguientes:

- la evaluación sistemática y exhaustiva de aspectos personales de una persona, incluida la elaboración de perfiles,
- el tratamiento a gran escala de datos sensibles,
- la observación sistemática a gran escala de una zona pública

6. Se deberá nombrar un DPD (Delegado de Protección de Datos), tanto si es un responsable como un encargado del tratamiento, si sus actividades principales guardan relación con el tratamiento de datos sensibles a gran escala o si sus actividades principales guardan relación con la observación habitual y sistemática a gran escala de personas.

7. El DPD puede ser un miembro del personal de su organización o puede contratar uno externo en el marco de un contrato de servicio. Sus funciones son las siguientes:

- informar y asesorar al responsable o encargado del tratamiento, así como a sus empleados, de sus obligaciones en virtud del Reglamento de protección de datos,
- controlar el cumplimiento por parte de la organización de toda la legislación relacionada con la protección de datos, incluidas las auditorías, las actividades de concienciación, así como la formación del personal que participe en las operaciones del tratamiento,
- proporcionar asesoramiento cuando se haya llevado a cabo una evaluación de impacto relativa a la protección de datos y controlar su cumplimiento,
- ejercer como punto de contacto para las solicitudes de los interesados relativas al tratamiento de sus datos personales y el ejercicio de sus derechos,
- cooperar y ejercer de punto de contacto con las autoridades de protección de datos sobre cuestiones relacionadas con el tratamiento.

8.	La protección que ofrece el Reglamento general de protección de datos (RGPD) viaja con los datos, lo cual significa que las normas que protegen los datos personales seguirán aplicándose independientemente de dónde vayan a parar los datos. Esto también se aplica cuando los datos se transfieren a un Estado no perteneciente a la Unión Europea (UE).
9.	El principio de responsabilidad proactiva es una piedra angular del Reglamento general de protección de datos (RGPD). De acuerdo con el RGPD, toda empresa u organización es responsable del cumplimiento de todos los principios de protección de datos, así como de demostrar dicho cumplimiento.

Tabla 15 – Obligaciones.

SOLICITUDES DE LOS CIUDADANOS	
1.	Las personas pueden ponerse en contacto con el responsable con el fin de ejercer sus derechos en virtud del Reglamento general de protección de datos (derecho de acceso, rectificación, supresión, oposición, portabilidad y limitación al tratamiento.
2.	Cuando alguien solicita acceder a sus datos personales, usted deberá hacer lo siguiente: ▪ confirmar si está tratando o no los datos personales del interesado, ▪ proporcionar una copia de los datos personales que posee de esa persona, ▪ proporcionar información sobre el tratamiento (como los fines, las categorías de datos personales, los destinatarios, etc.).
3.	El Reglamento general de protección de datos otorga a las personas el derecho a pedir que se borren sus datos y las organizaciones tienen la obligación de hacerlo, salvo en los casos siguientes: ▪ los datos personales que ustedes poseen son necesarios para ejercer el derecho de libertad de expresión, ▪ cuando una obligación legal implica que deben conservar los datos, ▪ por motivos de interés público (como salud pública, fines de investigación científica o histórica).
4.	Las personas tienen derecho a oponerse al tratamiento de los datos personales si exponen las razones específicas que les afectan.
5.	Las personas tienen el derecho a recibir de su organización los datos personales que le facilitaron en un formato estructurado y de lectura mecánica y a que se transmitan a otra empresa u organización (derecho a la portabilidad de los datos).

Tabla 16 – Solicitudes de los Ciudadanos.

CUMPLIMIENTO Y SANCIONES
1. Una de las funciones de la autoridad de protección de datos (APD) es publicar asesoramiento experto sobre cuestiones relativas a la protección de datos
2. El CEPD (Comité Europea de Protección de Datos)[64] es un organismo de la Unión Europea (UE) responsable de la aplicación del Reglamento general de protección de datos (RGPD) a partir del 25 de mayo de 2018. Está compuesto por el director de cada autoridad de protección de datos (APD) y el Supervisor Europeo de Protección de Datos o sus representantes. La Comisión Europea participa en las reuniones del CEPD sin derecho a voto. La secretaría del CEPD estará a cargo del Supervisor Europeo de Protección de Datos.
3. Si se realiza el tratamiento de datos en distintos países, la APD competente (que será la autoridad principal en sus relaciones con otras APD implicadas en la UE) es la APD del país de la UE en el que tenga su establecimiento principal. Se identifica su administración central en la UE como el lugar donde se toman las decisiones sobre los fines y los medios del tratamiento de los datos personales y donde se encuentra el poder para aplicar estas decisiones. Si no cuenta con una administración central en la UE, el establecimiento principal será el lugar donde se realice el ejercicio real y efectivo de las actividades de gestión que determinan las principales decisiones, tales como los fines y los medios del tratamiento.
4. El Reglamento general de protección de datos (RGPD) ofrece distintas herramientas a las autoridades de protección de datos (APD) en caso de incumplimiento de las normas de protección de datos: ▪ posible infracción: por ejemplo, se puede lanzar una advertencia, ▪ infracción: las posibilidades incluyen un apercibimiento, una prohibición temporal o definitiva del tratamiento y una multa de hasta 20 millones de euros o un 4 % del volumen de negocio total anual mundial.
5. Las personas pueden reclamar una indemnización si una organización no ha respetado el Reglamento de protección de datos (RGPD) y han sufrido daños materiales (como pérdidas económicas) o daños inmateriales (como ansiedad o pérdida de reputación).

Tabla 17 – Cumplimiento y Sanciones.

e-privacy

La propuesta[65] de Directiva sobre privacidad y comunicaciones electrónicas garantiza la protección de los derechos y libertades fundamentales, en particular el respeto de la vida privada, la confidencialidad de las comunicaciones y la protección de los datos personales en el sector de las comunicaciones electrónicas. También garantiza la libre circulación de datos, equipos y servicios de comunicaciones electrónicas en la Unión. Incorpora en el Derecho derivado de la Unión el derecho fundamental al respeto de la vida privada en el sector de las comunicaciones, establecido en el artículo 7 de la Carta de los Derechos Fundamentales de la Unión Europea[66].

La legislación sobre privacidad debe adaptarse para estar alineada con la nueva normativa europea de protección de datos (GDPR).

La propuesta para la nueva normativa e-privacy alineada con la GDPR de la Comisión Europea incluye[67] aspectos tales como:

- **Nuevos actores**: las reglas de privacidad en el futuro también se aplicarán a los nuevos actores que brindan servicios de comunicaciones electrónicas como WhatsApp, Facebook Messenger y Skype. Esto asegurará que estos servicios populares garanticen el mismo nivel de confidencialidad de las comunicaciones que los operadores de telecomunicaciones tradicionales.

- **Normas más estrictas**: todas las personas y empresas de la UE disfrutarán del mismo nivel de protección de sus comunicaciones electrónicas a través de este reglamento. Las empresas también se beneficiarán de un único conjunto de normas en toda la UE.

[64] https://edpb.europa.eu/edpb_es

[65] En el momento de publicación de la presente guía, se trata de una propuesta de Directiva, aún no aprobada

[66] https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52017PC0010 TXT/?uri=-CELEX:32016R0679#d1e2313-1-1

[67] https://ec.europa.eu/digital-single-market/en/proposal-eprivacy-regulation

- **Contenido y metadatos de comunicaciones**: se garantiza la privacidad de los metadatos y el contenido de las comunicaciones, por ejemplo: hora de una llamada y localización. Los metadatos tienen un alto componente de privacidad y deben anonimizarse o eliminarse si los usuarios no han dado su consentimiento, a menos que los datos sean necesarios para la facturación.

- **Nuevas oportunidades de negocio**: una vez que se otorga el consentimiento para los datos de las comunicaciones (contenido y / o metadatos) que se procesarán, los operadores de telecomunicaciones tradicionales tendrán más oportunidades para proporcionar servicios adicionales y desarrollar sus negocios.

- **Reglas más simples para las cookies**: la nueva regla será más fácil de usar ya que la configuración del navegador proporcionará una forma fácil de aceptar o rechazar el seguimiento de cookies y otros identificadores. La propuesta también establece que no se necesita consentimiento para que las cookies intrusivas no privadas mejoren la experiencia de Internet (por ejemplo, para recordar el historial del carrito de compras) o las cookies que utiliza un sitio web para contar el número de visitantes.

- **Protección contra el spam**: esta propuesta prohíbe las comunicaciones electrónicas no solicitadas por correo electrónico, SMS y máquinas de llamadas automáticas. Según la legislación nacional, las personas estarán protegidas de forma predeterminada para no recibir llamadas de marketing. Las personas que llamen con fines de marketing deberán mostrar su número de teléfono o usar un prefijo especial que indique que se trata de una llamada de marketing.

- **Aplicación más efectiva**: la aplicación de las normas de confidencialidad será responsabilidad de las autoridades de protección de datos, que ya están a cargo de las normas en virtud de lo que establece la RGPD.

Directiva sobre Bases de Datos

La Directiva sobre Bases de Datos[68] estipula que los creadores de bases de datos tienen el derecho de controlar el acceso y reutilizar sus contenidos, pero solo en aquellos casos en los que la suponga una gran inversión económica. En mayo del 2017, la Comisión Europea lanzó una consulta pública[69] para evaluar el impacto de la Directiva en los usuarios e identificar los ajustes necesarios para mejorar su aplicabilidad.

Los resultados preliminares de esta consulta pública se dieron a conocer en octubre de 2017[70]. La consulta recibió un total de 113 respuestas, de las cuales 18 provinieron de empresas. Sobre la base de las respuestas recibidas, la mayoría de los encuestados consideraron que los objetivos originales de la Directiva siguen vigentes hoy en día, aunque no existe una opinión unificada sobre el grado en que la Directiva ha logrado alcanzarlos.

La mayoría de los encuestados opinaron que el derecho "sui generis" establecido por la Directiva protege suficientemente las inversiones realizadas en la creación y el mantenimiento de las bases de datos, pero las opiniones están divididas en cuanto a si el alcance actual de este derecho es lo suficientemente amplio hoy en día. También hay una divergencia de opiniones con respecto a la capacidad de la Directiva para lograr un equilibrio entre los derechos e intereses de los propietarios y usuarios de la Base de Datos, y no existe consenso sobre qué enfoque debe prevalecer.

[68] http://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31996L0009

[69] https://ec.europa.eu/info/consultations/public-consultation-database-directive-application-and-impact-0_en

[70] https://ec.europa.eu/digital-single-market/en/news/summary-report-public-consultation-legal-protection-databases

Directiva PSI y Open Data

Existe una Directiva Europea específica para la reutilización de datos en el ámbito del sector público que, si bien inicialmente no es el centro de esta guía, es conveniente tenerla en cuenta para aquellos casos en los que la compartición de datos se realice con organismos públicos, sobre todo, a partir de su última versión de 20 de junio de 2019, donde pasa a llamarse "directiva de Open Data y PSI"[71]

La Directiva sobre la reutilización de la información del sector público proporciona un marco legal común para un mercado europeo de datos gubernamentales (información del sector público). Se construye en torno a dos pilares fundamentales: la transparencia y la competencia leal.

Esta Directiva sobre la reutilización de la información del sector público, también conocida, hasta junio de este año como la "Directiva PSI" (Directiva 2003/98 / EC[72]) entró en vigor el 31 de diciembre de 2003. Fue revisada por la Directiva 2013/37 / EU[73], que entró en vigor el 17 de julio de 2013. Su última versión se denomina: *Directiva (UE) 2019/1024 del Parlamento y del Consejo de 20 de junio de 2019 relativa a los datos abiertos y la reutilización de la información del sector público.* Esta revisión exhaustiva de la Directiva de 2013, recoge también los resultados de la consulta pública que realizó la Comisión Europea a finales de 2017 y entró en vigor el 26 de junio de 2019. La Directiva PSI se centra en los aspectos económicos de la reutilización de la información en lugar de en el acceso a la información por parte de los ciudadanos. Anima a los Estados miembros a poner la mayor cantidad de información disponible para su reutilización. Asimismo en la última revisión de la directiva se anima además a los Estados miembros a ir más allá de los requisitos mínimos establecidos la directiva aplicando los requisitos de ésta *a los documentos que obran en poder de empresas públicas relacionados con actividades que, en virtud del artículo 34 de la Directiva 2014/25/UE del Parlamento Europeo y del Consejo, estén sometidas directamente a la competencia*. Los Estados miembros también pueden decidir aplicar los requisitos de la Directiva de Open Data a las empresas privadas, en particular, a las que presten servicios de interés general.

En julio de 2014, la Comisión publicó directrices para ayudar a los Estados miembros a transponer las normas revisadas e indicar las mejores prácticas para la reutilización de la información del sector público en cada país. En España se crea la Ley 37/2007 sobre reutilización de la información del sector público del 16/11/2007, BOE n° 276 of 17/11/2007 n°19814, p. 47160 – 47165 modificada posteriormente por la Ley 18/2015, de 9 de julio, BOE no. 164/2015 of 10 July 2015, p. 57436[74], y esperamos que pronto se vuelva a modificar en función de la nueva Directiva 1024/2019.

[71] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32019L1024&from=EN

[72] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32003L0098&from=en

[73] https://eur-lex.europa.eu/legal-content/ES/TXT/PDF/?uri=CELEX:32013L0037&from=FR

[74] https://ec.europa.eu/digital-single-market/en/news/implementation-psi-directive-spain

ANEXO IV: DESCRIPCIÓN DE ALGUNOS CONCEPTOS TÉCNICOS CLAVE EN EL ÁMBITO DE OPEN DATA

En este apartado se describen algunos conceptos relevantes relacionados con los datos abiertos/compartidos, así como técnicas de datos abiertos necesarias para la apertura y la compartición de datos siguiendo las buenas prácticas ya existentes en el ámbito público.

Linked data

Linked data es una tecnología que conlleva un conjunto de principios de diseño para compartir y enlazar datos en Internet *o en la Web de Datos,* y supone una evolución de la idea de Web Semántica vinculada a los datos y a los metadatos. *Linked data* implica no sólo la publicación normalizada de conjuntos de datos, sino su interconexión o enlazado a través de grafos. Los datos se etiquetan con descriptores (*uniform descriptors*), que convierten el dato en información dinámica e interpretable, independientemente del idioma en el que esté o el tipo de fichero que lo contiene. Estas tecnologías y los estándares asociados (RDF, JSON-LD, etc.) mejoran la interoperabilidad y facilita que terceras partes encuentren el dato o *dataset* (discoverability).

Linked data utiliza varias técnicas, entre ellas RDF, vocabularios y URIs. Muchos catálogos de portales de datos abiertos que publican datos abiertos conectados (linked open data) utilizan un vocabulario predefinido.

En el ámbito de *linked data* es crucial el modelado de los datos donde se trata de reflejar la naturaleza y valor de los datos y establecer relaciones semánticamente pertinentes con otros conjuntos de datos. El modelado debe conseguir representar el área o dominio temático u organizacional definido y para ello es necesario un lenguaje de modelado específico, un vocabulario estandarizado, o incluso el diseño local de un vocabulario u ontología propia. Dado que las relaciones semánticas se pueden establecer entre *datasets*, el modelado de los metadatos descriptivos de conjuntos de datos adquiere una especial relevancia y es muy importante para establecer la interoperabilidad global. El proceso de asignación de licencias es otro aspecto muy relevante y depende de los condicionantes de publicación de los datos que tenga la organización, que pueden requerir, desde licencias apenas restrictivas o completamente abiertas *(Linked Open Data* o LOD), a la publicación de datos no abiertos *(Linked Closed Data).*

Tal y como los define Tim Berners-Lee en su esquema incremental de cinco estrellas que vimos en el apartado 4.4.2, si realmente se pretende llegar a una reutilización y explotación eficiente de los datos empresariales el objetivo debería ser cumplir con los principios de *linked data* o de LOD más allá de la mera publicación de los conjuntos de datos en formatos propietarios no preparados ni pensados para la reutilización.

Más información sobre *linked data:* https://www.w3.org/standards/semanticweb/data, También se recomienda el libro electrónico surgido el proyecto EUCLID, *Using linked data effectibly*

Metadatos, estándares de metadatos y publicación de catálogos de metadatos

Los metadatos son datos, datos sobre los datos, información estructurada que describe, explica, localiza o en cualquier caso facilita la recuperación, uso o gestión de la información.

Los metadatos tienen que ser entendibles tanto por las personas como por las máquinas *(machine-readable).* Los metadatos son además esenciales para la contextualización del dato, así como para enlazar/mapear usuarios y software con los datos disponibles en internet.

Los metadatos ayudan, entre otras cosas, a:

- Descubrir (encontrar) el *dataset*
- Entender la naturaleza y estructura del *dataset*
- Conocer el origen del *dataset* y bajo qué términos puede ser usado.

Y sobre todo, ayudan a hacer útiles los datos y a su calidad. Cuanto mejores son los metadatos (precisos, adecuados, completos, etc.) mejores serán los datos y sus posibilidades de recuperación y reutilización.

Estándares de metadatos

Los estándares son esenciales para la contextualización de los metadatos. Proveen una estructura común para los metadatos, facilitan la interoperabilidad entre sistemas, así como la accesibilidad y la posibilidad de encontrar (*discoverability*) los *datasets*.

Podemos clasificar los metadatos, los vocabularios o los vocabularios de metadatos, mezclando diferentes niveles y criterios (sistemas, prácticas, modelos, *schemas,* elementos, usos y tendencias). De los múltiples criterios a la hora de tipificar los metadatos, algunos de ellos son aplicables a nivel de registro de metadatos *(record)* y otros a nivel de *schema*. Vamos a analizar ahora sólo los que son aplicables a nivel schema o formato.

- Según el **propósito o campo de aplicación de los metadatos**, distinguimos: metadatos de propósito general (como el DC) y metadatos para propósitos específicos (como metadatos orientados a un dominio, p. ej. ISO 19115/FGDC para describir la información geoespacial).

- Según el **uso** que se le da a los metadatos. Este es el criterio quizás más vago pero más importante para clasificar tipos de metadatos, puesto que el principal objetivo de los metadatos, como ya hemos dicho, es hacer útiles los datos. Así, cualquier tipo de aplicación de los metadatos, cualquier utilidad que se les atribuya, puede incluirse en esta tipificación. Según la aplicación o funcionalidad de los metadatos, podemos deducir dos categorías fundamentales: metadatos para la recuperación de información y metadatos para la utilización del recurso.

- Atendiendo al **nivel de normalización.** Aunque todas las iniciativas de metadatos, y todos los vocabularios nacen con la vocación de ser estándares, podemos clasificar los modelos, *schemas*, etc. de metadatos, también según el nivel de normalización que alcancen. De forma genérica podemos distinguir:

    - Estándares de metadatos *de jure,* usualmente reconocidos a nivel nacional o internacional, y son modelos de metadatos o infraestructuras para la creación de vocabularios que han adquirido una aprobación formal gracias al proceso de normalización llevado a cabo por una institución normalizadora. Por ejemplo el caso del Dublin Core que es una norma ISO 15836-2008 o la ISO 19115-2003 para la descripción de la información geoespacial, etc.

- Estándares de metadatos *de facto,* que normalmente son *Public Access Specifications* o PAS, que son aquellos que se encargan de uno o varios aspectos de una política de metadatos (elementos, valores, buenas prácticas, etc.) que han conseguido un cierto crédito o posición dominante y cuya consulta es abierta para todo el mundo, porque se difunden públicamente.

Generalmente, antes de que un estándar llegue a ser norma *de facto* o *de jure* nos gusta decir (Méndez & van Holland, 2014) que las iniciativas de metadatos son estándares "por seducción", lo que implica que una comunidad, un proyecto o un dominio científico adopta un modelo o un estándar porque es útil, simple, sencillo, extensible, etc. Además establecer niveles de normalización para nuestros metadatos y vocabularios es crucial en cualquier proyecto de **creación y mantenimiento** de un servicio/sistema de compartición o apertura de datos.

Algunos de los estándares más usados para la apertura de datos en el sector públicos son:

- **DC** o DCMI – *Dublin Core/ Dublin Core Metadata Initiative*[75]: Estándar de metadatos de propósito general, desarrollado desde 1995, en un principio para describir los contenidos de las páginas web, y posteriormente en el ámbito de las bibliotecas digitales y en los repositorios de *Open Access,* entre otros múltiples dominios, para fomentar la interoperabilidad. El formato Dublin Core simple es un estándar ISO 15836 donde se recoge y explica la semántica y el alcance de 15 elementos de este modelo de metadatos, sin entrar en detalles de codificación o sintaxis.

- **DCAT** – *Data CATegory Vocabulary*[76]*:* DCAT es un vocabulario RDF destinado a facilitar la interoperabilidad entre catálogos de datos publicados en la Web, entre ellos, los portales de datos de información del sector público (PSI). Es un estándar de metadatos realizado en el contexto del Consorcio de la Web, y utilizado para los *datasets* de portales de datos publicados en la Web. Está basado en la filosofía del Dublin Core pero con el propósito específico de describir conjuntos de datos e información del sector público. Actualmente se está trabajando en la extensión de DCAT para otros tipos de datos en la web a través de distintos perfiles de aplicación de metadatos u otros vocabularios. Si se utiliza DCAT para describir *datasets* de distintos catálogos de datos, los responsables de su publicación (por ejemplo empresas del sector privado) aumentan las posibilidades de acceso e interoperabilidad entre múltiples catálogos de metadatos.

- **CERIF** – *Common European Research Information Format*[77] es un estándar de metadatos para compartir datos sobre la investigación en Europa, vinculado a los CRIS *(Current Research Information Systems)* de las universidades y otros organismos de investigación.

- **Inspire**: es un estándar europeo utilizado para la descripción de información y datos geoespaciales.

Se recomienda publicar los metadatos en formatos *machine-readable* y usar términos estándar para definir los metadatos . Adicionalmente, describir las características generales del *dataset* con información sobre parámetros locales, licencias, origen y calidad del *dataset*. Seleccionar formatos estándar de metadatos garantiza su interoperabilidad y reutilización. El estándar o schema de metadatos relacionados Se recomienda usar el estándar DCAT-AP.

[75] https://www.dublincore.org/

[76] https://www.w3.org/TR/vocab-dcat/

[77] http://www.dcc.ac.uk/resources/metadata-standards/cerif-common-european-research-information-format

Publicación de metadatos.

Aunque los metadatos están ligados a los *datasets* que describen, puede ser útil tener los metadatos asociados a un *dataset* en múltiples sitios. Por ejemplo, los portales de datos abiertos de la Unión Europea comparten los metadatos de los portales de datos abiertos de los países miembros. En algún caso, se comparten solo un subconjunto de los metadatos asociados a los *datasets*, los correspondientes al subconjunto del *dataset* que se quiere compartir. En otros, además de la totalidad o parte del metadato, se comparte el propio *dataset*. Esta práctica de datos abiertos con datos públicos, sería posible extrapolarla a los datos privados o de carácter empresarial. Ya que la compartición de metadatos en múltiples sitios mejora sustancialmente la disponibilidad de los *datasets*. Esta compartición se puede hacer de manera manual o automáticamente a través de los portales de datos abiertos a través de la extracción o *metadata harvesting*.

Consultar https://www.europeandataportal.eu/en/providing-data/how-to-be-harvested-by-us para saber cómo federar los *datasets* en el portal de datos abiertos europeo.

Se recomienda publicar siempre que sea posible los metadatos como *linked data*[78]. Mejora la facilidad de acceso al dato y la interoperabilidad. Muchos portales incorporan una solución software que permite crear fácilmente los metadatos asociados a los *datasets*.

Otros Vocabularios de metadatos.

Los metadatos en general describen diferentes atributos o propiedades de los datos y otros objetos digitales, dándoles significado, contexto y organización de forma normalizada. La primera tipificación de los metadatos apareció hace 20 años, y estudia los metadatos analizando las diferentes funcionalidades que soportan los elementos de metadatos existentes. Sin embargo, el reconocimiento de los muchos usos y dimensiones que tienen los metadatos nos lleva a la construcción de una tipología muy amplia. La mayor parte de las tipologías tradicionales se refiere al uso funcional y a los objetivos de los metadatos más que a las cualidades inherentes de los ***schemas*** de metadatos, los elementos de metadatos o a cualquier valor de dichos elementos ***(schemes).***Información adicional y formación sobre metadatos en:

- http://w3c.github.io/dwbp/bp.html#metadata
- https://www.europeandataportal.eu/en/resources/elearning
- https://theodi.org/guides/marking-up-your-dataset-with-dcat

Recopilando: buenas prácticas de metadatos para la descripción de conjuntos de datos:

- Publicar los metadatos junto a los datos.
- Usar formatos *machine-readable* para los metadatos.
- Usar términos estándares para definir los metadatos.
- Describir las características generales del *dataset*.
- Proveer información sobre parámetros locales (fecha, duración, valor, tipos, lenguaje).
- Proveer información sobre la licencia asociada al *dataset*.
- Proveer información sobre el origen del *dataset*.
- Proveer información acerca de la calidad del *dataset*.

[78] http://www.w3.org/TR/dwbp/#-metadata

Ontologías
RDF (Resource Description Framework)

RDF (*Resource Description Framework* - Marco de descripción de recursos) es un lenguaje estándar de codificación de datos que permite describir e intercambiar recursos en la web mediante metadatos, no sólo documentos, sino todo tipo de entidades y objetos digitales que tengan presencia en la web, incluso referidos a personas, organizaciones, productos, servicios, etc. RDF es una herramienta fundamental para hacer de la Web una infraestructura global para compartir y reutilizar datos y documentos entre diferentes tipos de usuarios y sistemas. RDF es un estándar para la codificación normalizada de datos y metadatos en la Web. Es el lenguaje clave para la web semántica y *linked open data.*

RDF proporciona un método general y flexible para descomponer cualquier conjunto de datos en piezas pequeñas llamadas tripletas, con algunas reglas de la semántica para relacionar esas piezas. Cada una de estas tripletas son sentencias o declaraciones sobre el conjunto de datos, o cualquier otra entidad que se quiera describir de forma normalizada. RDF se basa en la idea de que las cosas que se describen tienen **propiedades** que a su vez tienen **valores** y que los recursos se pueden describir mediante sentencias como la del ejemplo, que especifican cuáles son esas propiedades y valores. La terminología empleada para denominar esas partes de la **sentencia** o **declaración** (*statement*) son: sujeto, predicado y objeto:

- **sujeto**: identifica el recurso o *dataset* sobre el que versa la declaración
- **predicado**: identifica la propiedad o característica del sujeto que se expresa mediante esta declaración (ej. Idioma de los datos, fecha de creación, etc.).
- **objeto**: identifica el valor de la propiedad a la que se refiere el predicado.

Más información sobre RDF: *http://www.linkeddatatools.com/introducing-rdf*

URI (Uniform Resource Identifier / IRI (International Resource Identifier)

Los **identificadores uniformes de recursos** (URI) son un tipo de identificación persistente de recursos o *datasets.* Un URI hace referencia bien a un texto o literal ej. URN (*Uniform Resource Name*), o a una localización, URL (*Uniform Resource Locator*). En el ámbito de *linked data* y sus lenguajes de codificación (RDF, JSON-LD, etc.) a través de tripletas y grafos, los URI sirven para definir cada uno de los elementos de una declaración o tripleta. , para identificar el sujeto de la declaración (el *dataset)* que permite su localización y acceso en la red. No obstante, muchos de los elementos que pueden formar parte de una declaración en RDF no tienen por qué tener una ubicación en la red, por lo que RDF propone el uso de **IRI (International Resource Identifier)** (denominación que sustituye a la de URI (Uniform Resource Identifier) en la nueva versión de RDF 1.1). Los IRI son un tipo especial de URL que permite referirse a cualquier cosa que se necesite mencionar en una expresión RDF, incluyendo los conjuntos de datos, sobre todo si estos sólo son compartidos, pero no expresamente abiertos. RDF utiliza los IRIs como base de su mecanismo de identificación de sujetos, predicados y objetos.

Más información sobre URIs en el ámbito de la reutilización de *datasets https://joinup.ec.europa.eu/sites/default/files/c0/7d/10/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf*

ANEXO V: SELECCIÓN NO EXHAUSTIVA DE HERRAMIENTAS Y APLICACIONES DE SOFTWARE LIBRE PARA LA IMPLEMENTACIÓN DE OPEN DATA

Atlas (Apache): https://atlas.apache.org
ATLAS es una herramienta de software libre de APACHE que facilita la integración del ecosistema de datos de una organización utilizando metadatos compatibles con Hadoop.

APACHE ATLAS ofrece una gestión de metadatos abiertos (*open metadata*) y permite a las organizaciones construir su catálogo de *datasets*, clasificarlos y gobernar los *datasets*. Incluye también funcionalidades que facilitan la colaboración con investigadores de datos (ej. expertos de big-data) y los responsables de los datos en la organización.

CKAN : http://ckan.org
Es una herramienta desarrollada por Open Knowledge que se ha convertido en un estándar para los portales de datos abiertos. Es una aplicación de software libre disponible de manera gratuita soportada por una comunidad que la mejora continuamente. El sistema de catálogo CKAN incluye funcionalidades como *harvesting* (recolección de metadatos), publicación y auditoría de datos además de disponer de un sistema de almacenamiento de datos integrado propio. CKAN necesita un CMS para funcionar. Se recomienda usarlo con Drupal, Wordpress, and Django.

El portal de datos europeos y muchos otros portales gubernamentales conocidos utilizan CKAN.

Datatank: http://thedatatank.com
Es una herramienta de software libre, como CKAN, Drupal o Elastic Search, que permite transformar *datasets* en formato http API.

DKAN: http://demo.getdkan.com
Es también un sistema de catálogo de software libre desarrollado por Drupal. Está recomendado por los portales de datos abiertos de USA. Se complementa bien con CKAN y tiene un CMS integrado basado en Drupal, aunque se puede integrar fácilmente con otros CMS.

ODI Certificate Tool: https://certificates.theodi.org
ODI *(Open Data Institute)* ha desarrollado una herramienta de autoevaluación que permite identificar la madurez de los *datasets* respecto a su apertura técnica y legal. Evalúa también los pasos que se deben dar/dados para que los *datasets* se puedan encontrar por terceros más fácilmente. La autoevaluación es una fuente de información relevante para identificar acciones de mejora en la iniciativa de compartición/apertura de datos en una organización. Esa aplicación ayuda también a la institución que la utilice para la certificación con el modelo de madurez de datos abiertos.

OPENREFINE : http://openrefine.org
Es una herramienta desarrollada por Google que permite limpiar y analizar *datasets*.

OPEN DATA SOFT: https://www.opendatasoft.com
Plataforma de uso compartido de datos que los equipos utilizan para acceder, reutilizar y compartir datos que hacen crecer el negocio.

SNORQL: http://snorql.nextprot.org
Esta herramienta es útil para aprender a hacer *queries* en SPARQL.

SÓCRATA: https://socrata.com
Es una plataforma API segura que permite a las organizaciones gubernamentales modernizar las infraestructuras de datos de sus organizaciones. Impulsa un proceso repetible mediante el cual el personal puede tomar datos de varios sistemas de origen, automatizar su flujo a la nube, organizarlo de manera uniforme y luego convertirlo en un servicio siempre activo para la reutilización de datos aguas abajo.

SpagoBI. Dashboards I (Fi-ware): http://www.stratebi.com/spagobi
Es una plataforma Business Intelligence Open Source que cubre y satisface todos los requisitos de BI, tanto en términos de análisis y de gestión de datos, administración y seguridad.

SpagoBI para la parte analítica ofrece soluciones para la presentación de informes, análisis multidimensional (OLAP), minería de datos (Data Mining), cuadros de mando (Dashboard) y consultas ad-hoc. Tiene además módulos propios para la gestión de procesos de colaboración y análisis de geo-referencia.

SpagoBI es una plataforma de integración (y no una plataforma de producto), ya que no se construye y se cierra en torno a un conjunto predefinido de herramientas. Tiene una estructura modular en la cual todos los módulos se relacionan con el núcleo del sistema, lo que garantiza la armonía de la plataforma junto con su capacidad evolutiva.

VIRTUOSO: https://dbpedia.org/sparql
Es una de las herramientas más conocidas para ejecutar *queries* en SPARQL.

ANEXO VI: CUESTIONARIO SOBRE INICIATIVAS DE DATOS ABIERTOS EN EL SECTOR PRIVADO

Justificación y objetivos

El Grupo de Datos Abiertos de COTEC coordinado por SPRI tiene como objetivo fundamental contribuir al fomento de la apertura de datos como recursos de valor por parte de las empresas privadas. Para animar estos procesos de apertura de datos, y como punto de partida, se ha determinado elaborar una *Guía para la apertura y compartición de datos en el entorno empresarial,* que en un buen número de casos no se acometen porque no hay una conciencia clara de los beneficios que pueden acarrear y/o porque se desconocen los protocolos y pasos intermedios que hay que recorrer.

En este sentido, las empresas que ya se han embarcado en este tipo de iniciativas disponen de una experiencia previa que puede animar a otras empresas a recorrer este camino. El ejemplo de esas otras empresas (recogido en la *Guía* a modo de casos de estudio), así como el conocimiento de las razones reales o barreras que impiden a las empresas involucrarse en este tipo de iniciativas, de los factores que ayudarían a cambiar ese enfoque y de los planes de la empresa con respecto a los datos abiertos en el futuro cercano, son cuestiones capitales para impulsar estos procesos de apertura.

Así, para conocer de primera mano toda esta valiosa información e integrarla en la *Guía para la apertura y compartición de datos en el entorno empresarial*, solicitamos vuestra colaboración para contestar este cuestionario[79], que es válido tanto si ha realizado proyectos de datos abiertos como si no lo ha hecho. El cuestionario especifica claramente qué preguntas han de contestarse en cada caso, si bien éste se ha construido con lógica de unidad, con el objetivo de que su lectura completa sirva de aprendizaje para todo tipo de empresas. Las Administraciones Públicas o asociaciones también pueden aportar experiencias que hayan tenido o que conozcan de empresas de su entorno.

La *Guía para la apertura y compartición de datos en el entorno empresarial* va a permitir compartir la terminología y procesos para una apertura de datos exitosa, así como apoyar la definición de una estrategia de datos abiertos, siendo su característica básica la flexibilidad, dado que se dirigirá a cualquier sector de actividad y a cualquier tamaño de empresa. La *Guía* describe los principales patrones de compartición/apertura de datos así como las principales razones que llevan a las empresas a compartir/abrir sus datos, aborda también el marco metodológico de apertura/compartición de datos, se centra en la descripción de los diferentes modelos de apertura de datos, la necesaria consideración del marco legal a la hora de abordar cualquier iniciativa de este tipo, así como la consideración detallada de los pasos que es necesario dar para abordar proyectos de apertura (estrategia). Finalmente, incluirá las principales recomendaciones de diferentes empresas que han realizado apertura de datos y que han sido recogidas por medio de este cuestionario y entrevistas.

¡Muchas gracias por la colaboración!

[78] Elaboración propia - Basado en: https://publications.europa.eu/en/publication-detail/-/publication/004fcf02-49c7-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-69800191

1. INFORMACIÓN PREVIA

Nombre de la empresa/organización

Nombre y apellidos de la persona que responde este cuestionario

---

Posición en la empresa

INFORMACIÓN SOBRE LA INICIATIVA DE APERTURA DE DATOS

2. Título y breve descripción de la iniciativa de datos abiertos *(objetivos, justificación de la oportunidad). Por favor incluya también la dirección de la/s página/s web si existe información on line de esa iniciativa*:

2.1. ¿Cuándo se puso en marcha la iniciativa?

2.2. ¿Cuál ha sido su duración?

2.3. ¿Sigue todavía activa? ☐SI ☐NO

2.4. ¿Ha dado pie esta iniciativa a otras iniciativas posteriores de apertura de datos? ☐SI ☐NO

3.a ¿Qué tipo de datos ha abierto?

☐ Datos de localización/posicionamiento.

☐ Datos de transacciones (clientes, proveedores).

☐ Datos internos operativos (calidad, producción, ...).

☐ Datos en tiempo real (o casi a tiempo real).

☐ No sé.

☐ Otros (*señalar cuál*):

3.b. SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. En caso de **involucrarse en una iniciativa de apertura de datos en el futuro**, ¿qué tipo de datos ve más probable que abra su empresa? (*Pueden marcarse varios*)

☐ Datos de localización/posicionamiento.

☐ Datos de transacciones (clientes, proveedores).

☐ Datos internos operativos (calidad, producción, ...).

☐ Datos en tiempo real (o casi a tiempo real).

☐ No sé.

☐ Otros (*señalar cuál*):

_____

4. ¿Cuáles han sido los principales usuarios de los datos abiertos de tu organización? *(elección múltiple)*

☐ Manufacturas.

☐ Distribución/logística.

☐ Servicios TIC, incluyendo desarrolladores app/software.

☐ Sanidad.

☐ Automoción y transporte.

☐ Servicios financieros y seguros.

☐ Turismo y hostelería.

☐ Administraciones Públicas.

☐ Educación.

☐ Construcción.

☐ Investigación.

☐ Medios de comunicación y publicidad.

☐ Construcción.

☐ Utilities.

☐ Sector primario.

☐ Otros:

_____

5. Relación de los usuarios de datos con su empresa

☐ Clientes.

☐ Proveedores.

☐ Otras empresas dentro de un grupo empresarial.

☐ Asociación sectorial.

☐ Otros (*señalar cuál*):

_____

6. ¿Tiene constancia del uso principal que estos usuarios han dado a sus datos?

☐ SI

¿Puede decir cuál?

_____

☐ NO

7.a La iniciativa en datos abiertos en tu organización puede catalogarse como *(señalar varios si se han realizado)*:

☐ Monetización de los Datos: Obtención de ingresos a través de los datos generados por la organización.

☐ Marketplace de Datos: Intermediarios de confianza que reúnen a proveedores de datos y usuarios de los datos, facilitando una plataforma segura de intercambio.

☐ Plataforma de Datos Industrial: Enfoque colaborativo y estratégico para intercambiar datos entre un grupo restringido de empresas.

☐ Datos Abiertos: Apertura de datos de forma gratuita para fomentar el desarrollo de nuevos productos y/o servicios.

☐ Modelos B2G (*business to government*): donación de datos, premios, etc.

☐ Facilitadores Técnicos: Permiten el intercambio de datos a través de una solución técnica y cuyos ingresos se obtienen de configurar, usar y/o mantener la solución.

7.b. SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. En caso de **involucrarse en una iniciativa de apertura de datos**, ¿en qué tipo de iniciativa cree más probable que centre sus esfuerzos? (*Marcar un máximo de dos*)

☐ Monetización de los Datos: Obtención de ingresos a través de los datos generados por la organización.

☐ Marketplace de Datos: Intermediarios de confianza que reúnen a proveedores de datos y usuarios de los datos, facilitando una plataforma segura de intercambio.

☐ Plataforma de Datos Industrial: Enfoque colaborativo y estratégico para intercambiar datos entre un grupo restringido de empresas.

☐ Datos abiertos: Apertura de datos de forma gratuita para fomentar el desarrollo de nuevos productos y/o servicios.

☐ Modelos B2G (*business to government*): donación de datos, premios, etc.

☐ Facilitadores Técnicos: Permiten el intercambio de datos a través de una solución técnica. Los ingresos se obtienen de configurar, usar y/o mantener la solución.

RAZONES PARA LA APERTURA DE DATOS

8. ¿Por qué **razón principal** la empresa decidió llevar a cabo la iniciativa de apertura de datos? *(señale las tres más importantes)*

☐ Mejorar la eficiencia de los procesos internos.

☐ Diversificar la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del mercado y del cliente.

☐ Resolver un problema de gestión interno concreto.

☐ Apoyar la innovación abierta y colaborativa.

☐ Crear nuevos modelos de negocio.

☐ Ser pionero en estos procesos (posicionamiento estratégico).

☐ Mejorar la transparencia/rendición de cuentas de la empresa.

☐ Propósito social, de bien común público.

☐ Contribuir a la resolución de un reto local o global.

☐ Otros (*señalar cuál*):

_____

9. Una vez implementada la iniciativa, ¿el resultado obtenido **se ha correspondido con las expectativas** planteadas?

☐ Ha estado por encima de las expectativas.

☐ Ha estado en consonancia con las expectativas.

☐ Ha estado por debajo de las expectativas.

10. Si ha estado por encima o por debajo de las expectativas, señale la razón principal para justificar su respuesta *(respuesta abierta)*

OBSTÁCULOS PARA LA APERTURA DE DATOS

11.a ¿Cuáles fueron los **principales obstáculos** a los que tuvo que hacer frente? (*señale los tres más importantes).*

☐ Falta de conocimiento de los beneficios asociados a la iniciativa y/o del valor de los datos.

☐ Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial).

☐ Desconocimiento de los pasos intermedios/protocolos que deben tenerse en cuenta.

☐ Riesgos en el ámbito de la ciberseguridad.

☐ Falta de habilidades adecuadas en la empresa para gestionar iniciativas de este tipo.

☐ Filtración de información sensible a empresas competidoras.

☐ Alto coste de la iniciativa (en la preparación de los datos, en la extracción y tratamientos de datos, etc.).

☐ Problemas de tipo técnico/falta de interoperabilidad o estandarización de los datos.

☐ Falta de acuerdos contractuales apropiados que garanticen el uso adecuado de los datos, la propiedad, etc.

☐ Falta de interés en los datos de la organización por parte de terceros.

☐ Otros (*señalar cuál*):

11.b SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. ¿Por qué razón la empresa no se ha involucrado hasta la fecha en iniciativas de apertura de datos? (*como máximo señale cuatro*).

☐ Falta de conocimiento de los beneficios asociados a la iniciativa y/o del valor de los datos.

☐ Cuestiones de privacidad.

☐ Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial).

☐ Desconocimiento de los pasos intermedios/protocolos que deben tenerse en cuenta.

☐ Riesgos en el ámbito de la ciberseguridad.

☐ Falta de habilidades adecuadas en la empresa para gestionar iniciativas de este tipo.

☐ Filtración de información sensible a empresas competidoras.

☐ Problemas de tipo técnico/falta de interoperabilidad o estandarización de los datos.

☐ Falta de acuerdos contractuales apropiados que garanticen el uso adecuado de los datos, la propiedad, etc.

☐ Falta de interés en los datos de la organización por parte de terceros.

☐ Costes económicos de la apertura de datos (ejemplo: costes de hacer disponibles los datos en el formato deseado, costes de infraestructura relacionados con la recogida de datos, costes de curación de los datos, etc.)

☐ Costes reputacionales en caso de mal uso de los datos abiertos.

☐ Otros (*señalar cuál*):

_____

FACTORES DE ÉXITO PARA UNA CORRECTA APERTURA DE DATOS

12.a. ¿Cuáles son los **factores de éxito** en las iniciativas de apertura de datos? (*señale los dos más importantes).*

☐ Disponer de un marco legal claro en relación con la protección de datos personales y la propiedad intelectual.

☐ Realizar una inversión inicial acertada o disponer de financiación de soporte externo.

☐ Simplicidad y facilidad de uso de los datos (acceso, interoperabilidad, etc.).

☐ Entendimiento de la demanda de datos (necesidades, casos de uso, usuarios, etc.).

☐ Disponer de personal con las cualificaciones y competencias necesarias.

☐ Crear acuerdos de colaboración con empresas con conocimientos complementarios (win-win).

☐ Ser pionero (por tipo de empresa, sector, tipo de proyecto, etc.).

☐ La creación de un marco de confianza, garantizando la seguridad y control de los datos.

☐ El intercambio de experiencias y conocimientos en la compartición de datos.

☐ Otros (s*eñalar cuál*):

12.b. SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. ¿Qué **haría más probable que su empresa se embarcase en iniciativas de apertura de datos**? (*por favor marque un máximo de tres*).

☐ Mayor certeza de cómo abrir los datos desde un punto de vista contractual.

☐ Mayor claridad legal acerca de lo que se puede hacer y lo que no se puede hacer con los datos abiertos.

☐ Mayor control sobre el uso de los datos una vez que éstos se han abierto.

☐ Mayor disponibilidad de cualificaciones técnicas para aprovechar el valor de los datos.

☐ Mayores niveles de interoperabilidad/infraestructura para facilitar la apertura de los datos.

☐ Una mayor compensación económica por la apertura de esos datos.

☐ Otros (s*eñalar cuál*):

IMPACTO DE LA APERTURA DE DATOS Y PROSPECTIVA

13. ¿Cuál es el grado de importancia cuantitativa de las actividades de apertura de datos en la actividad de la empresa? (porcentaje de datos abiertos mensualmente sobre la totalidad de los datos generados por la empresa) *(marcar sólo una):*

☐ Menos de un 1%

☐ Hasta un 5%

☐ Hasta un 10%

☐ Hasta un 25%

☐ Hasta un 50%

☐ Si es más de un 50%, indicar el porcentaje

---

14 ¿Cuál es el grado de importancia cualitativa de las actividades de apertura de datos en la actividad de la empresa? *(marcar sólo una):*

☐ La apertura de datos es una de las principales actividades de mi empresa (ejemplo: la empresa vende sensores para el control de la temperatura, y obtiene la mayor parte de la facturación de la puesta en valor de los datos obtenidos por estos sensores, no de la venta de sensores).

☐ La apertura de datos es una actividad secundaria de mi empresa (ejemplo: la empresa produce aspiradoras que recogen información de las superficies con objeto de optimizar la limpieza. La empresa obtiene sus mayores ingresos de la venta de robots, pero también -de manera secundaria- de la venta de los datos recogidos a otras empresas interesadas, obteniendo así un ingreso adicional).

☐ La apertura de datos es una actividad secundaria actualmente, pero es probable que sea una actividad principal en los próximos cinco años.

☐ La apertura de datos se circunscribe a la iniciativa descrita aquí y no hay intención de que aumente su estatus en la actividad de la empresa próximamente.

☐ La apertura de datos se circunscribe a la iniciativa descrita aquí, pero está previsto llevar a cabo otras iniciativas próximamente.

15. SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. ¿Hasta qué punto **considera probable** que su empresa active la apertura de datos en los próximos cinco años? (*señale sólo uno).*

☐ Altamente improbable.

☐ Improbable.

☐ Probable.

☐ Altamente probable.

☐ No lo sé

RECOMENDACIONES PARA LA APERTURA DE DATOS

16. Desde la experiencia en el mundo de los datos abiertos, ¿**por dónde debería empezar una empresa** que quiera embarcarse en un proyecto de este tipo? ¿Alguna **otra recomendación** final?

17. SOLO para las EMPRESAS QUE NO han realizado iniciativas de apertura de datos. ¿Qué **tipo de asesoramiento necesitaría** para poder comenzar una iniciativa de apertura de datos? *(pregunta abierta).*

ANEXO VII: DESCRIPCIÓN DE CASOS DE APERTURA Y COMPARTICIÓN DE DATOS

PRIMAFRIO

La apertura de datos es un elemento de carácter estratégico en la empresa que todavía tiene un carácter relativamente secundario y exploratorio, pero que se apunta como principal en los próximos cinco años.

Es por ello, que el Grupo Primafrio, empresa con más de cincuenta años de experiencia especializada en el transporte por carretera y en el transporte refrigerado, ha introducido los datos abiertos como un elemento de valor en sus operaciones, en concreto en los ámbitos de localización y posicionamiento (datos en tiempo real de parámetros que el cliente puede necesitar, como la matrícula del camión, la posición, el conductor, etc.) y relacionados con transacciones realizadas con ellos (facturación o contabilidad en su propia cuenta), que se gestionan a través de plataformas que se han creado internamente ad hoc.

Existe constancia expresa del uso que los clientes han dado a estos datos, relacionado con la optimización del tracking de los pedidos y también la capacidad de adelantar la facturación.

Objetivo:
Esta iniciativa ha buscado fundamentalmente la mejora en la eficiencia de los procesos internos de la compañía, al mismo tiempo que resolver un problema de gestión (tracking de pedidos) posicionando estratégicamente a la entidad, pero ha trascendido esos límites para configurarse como una herramienta que busca también incardinarse en procesos de innovación abierta, en la medida que van dirigidos al cliente y existe voluntad clara para, a través de la escucha activa de sus necesidades, poder embarcarle en dinámicas de co-innovación. Aunque no se ha llevado a cabo una evaluación formal de la iniciativa, de acuerdo con sus responsables ha estado en consonancia con las expectativas depositadas.
A futuro la empresa ya está embarcada en nuevos desarrollos y plataformas tecnológicas que beneficiarán la cadena de valor del sector transporte.

Riesgos:
En todo caso, la iniciativa no ha estado exenta de dificultades y riesgos asociados. Entre ellos, los más relevantes (que son extrapolables a buena parte de proyectos de apertura de datos) son los riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con la LGPD o el secreto industrial), riesgos en el ámbito de la ciberseguridad y un relativo alto coste de la iniciativa. Relacionado con ello, los factores de éxito más críticos tienen que ver con la disposición de un marco legal claro, transparente, para alejar esos riesgos de tipo regulatorio/legislativo y de confianza para garantizar la seguridad y el control de los datos. Igualmente, disponer de personal adecuado y utilizar marcos de uso amigables se antojan como claves.

Recomendaciones:
La principal recomendación para empresas que se han decidido a embarcarse en iniciativas de datos abiertos es tener bien definido y bajo relación contractual el usuario final para evitar cualquier tipo de incidencia, dada la importancia de los datos a compartir. También lo es tener una buena estructura y soporte que permita que se pueda transferir el dato de interés con el cliente final.

Tipología de colaboración y modalidad de datos abiertos:	• Datos abiertos/compartidos • Datos: - de localización/posicionamiento - de transacciones (clientes, proveedores) - en tiempo real
Sectores de los principales clientes de Primafrio	• Distribución/logística. • Servicios TIC, incluyendo desarrolladores app/software. • Automoción y transporte.
Monetización del negocio: **Razón/es de la apertura**	No hay monetización directa: son datos que se ofrecen como servicio de valor añadido al cliente, ya que ofrecer un mejor servicio al cliente impacta en optimizar su fidelidad a medio plazo y, por tanto, incrementa en última instancia la capacidad de monetización. • Mejorar la eficiencia de los procesos internos. • Resolver un problema concreto de gestión interna. • Ser pionero en estos procesos (posicionamiento estratégico). • Apoyar la innovación abierta y colaborativa. • Disponer de un marco legal claro en relación con la protección de datos personales y la propiedad intelectual. • Simplicidad y facilidad de uso de los datos (acceso, interoperabilidad, etc.)
Factores de éxito:	• Disponer de personal con las cualificaciones y competencias necesarias. • Ser pionero (por tipología de empresa, sector, proyecto, etc.). • Creación de un marco de confianza, garantizando la seguridad y control de los datos.
Obstáculos:	• Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial). • Riesgos en el ámbito de la ciberseguridad. • Alto coste de la iniciativa (en la preparación de los datos, en la extracción y tratamiento de datos, etc.).

GRUPO EUSKALTEL

La iniciativa del grupo Euskaltel para abordar el ámbito de la compartición de datos surgió en el año 2017, momento en el cual se empezó a estudiar las posibilidades que podrían surgir a partir del análisis de la información contenida en sus CDR (Call Detail Records) que recogen información sobre las conexiones de terminales móviles a las antenas.

El estudio de la potencialidad de esta información concluyó que se pueden generar indicadores muy interesantes para diferentes negocios e incluso para el sector público sobre el comportamiento de los clientes del grupo y generar indicadores valiosos para la mejora sus prestaciones.

Debido a esto se decidió utilizar una solución *Big Data* desarrollada dentro del grupo, Mileva, que aporta soluciones de almacenamiento de datos *Big Data*, visualización y análisis avanzado de datos a través de sus componentes. La herramienta Mileva también tiene la capacidad de juntar diferentes fuentes de datos, tanto públicas o de fuentes privadas que permitirían enriquecer esta información, adecuándola a cada uno de los intereses individuales que pueden surgir de diferentes negocios o gobiernos.

La información que se genera a partir de los datos de origen está siempre debidamente anonimizada y agregada de manera que cumpliendo la presente regulación de protección de datos no permita que puedan existir casos de re-identificación de ningún usuario. Esto unido a que la solución *Big Data* ha sido creada dentro del equipo de *Big Data* del propio grupo, genera que los datos no salgan de los sistemas de la compañía.

Para la algoritmia en función del caso se va a buscar la ayuda de socios tecnológicos y proveedores externos especializados en cada uno de los ámbitos de los cuales se va a generar una propuesta de información que aporte valor al negocio. En este caso los encargados de realizar los pertinentes análisis serán socios tecnológicos del grupo, donde todos los elementos externos que entren en contacto con la información seguirán los requisitos de cumplimiento de la RGPD para garantizar la privacidad, confidencialidad y el correcto uso de los datos del grupo Euskaltel. Los acuerdos serán específicos para cada proyecto.

El cliente final sólo podrá tener acceso a informes finales bajo las condiciones que haya decidido (rango temporal, zona geográfica) pero en ningún momento tendrá acceso a los datos en crudo utilizados para la generación del informe que han contratado, por lo que la información bruta de nuestros clientes no queda expuesta en ningún momento.

Tipo de colaboración:	• B2B. • B2G (AAPP del estado).
Sectores de los principales clientes:	• Transporte. • Sector Público. • Retail/comercio.
Monetización del negocio:	La tarifa del servicio depende del valor aportado al cliente. Criterios para determinar el valor de los datos: • Nivel de enriquecimiento de los datos. • Zona geográfica. • Período de tiempo del estudio.
Principales Obstáculos:	• Adecuación de la infraestructura propia para el caso particular. • Convencer a los clientes potenciales del valor de los datos a compartir para su negocio.

TELEFÓNICA

TELEFÓNICA es una empresa con amplia experiencia en el ámbito de la gestión de datos. A modo de ejemplo, y como prueba de ello, en el año 2018 ha sido reconocida por el Informe FORRESTER (*Wave Leaders 2018. Specialized Insights. Service Providers*) como proveedor líder en *insights* (información que se extrae de grandes cantidades de datos anonimizados y agregados). El baluarte de esta posición de referencia de Telefónica es LUCA, la unidad de negocio especialista en *Big Data*, que ofrece servicios que combinan *business insights*, herramientas e infraestructura y consultoría y analítica adaptados a sus clientes. LUCA ofrece soluciones y servicios para apoyar a las empresas en su recorrido hacia la transformación digital y desde su fase inicial.

Como proveedor de datos, LUCA comparte principalmente insights generados de los datos que se recogen de la red móvil tales como movimientos de multitud, datos sobre localización, etc. garantizando en todo momento la privacidad del cliente y cumpliendo estrictamente con la legislación existente sobre protección de datos. En menos de tres años, desde LUCA se han implementado varios centenares de proyectos (alrededor un 30% en el Sector Público), cubriendo realidades y sectores muy diversos, como turismo, transporte, comercio, servicios financieros, publicidad y medios y también transversales (como ahorro energético o gestión de flota).

TELEFÓNICA no dispone como tal de *datasets* en abierto, aunque sí hay disposición de abrir datos a demanda cuando el propósito de su utilización es social. De vez en cuando comparte datos para retos en hackatons.

De todas las iniciativas que se han llevado con datos, destaca el protagonismo de "***Big Data for Social Good***" por su compromiso social y su carácter abierto. En esta iniciativa se hace uso de datos internos de TELEFÓNICA junto con otros externos para devolver el valor del dato a la sociedad, contribuyendo así a los Objetivos de Desarrollo Sostenible fijados por la ONU para 2030. Dentro de esta iniciativa destacan proyectos como el estudio de los patrones de movilidad en comunidades vulnerables a episodios climatológicos extremos, el proyecto OPAL (de Open Algorithms) para apoyar al Departamento Nacional de Planificación en Colombia en la explotación de los datos del sector privado de forma que tengan un impacto social escalable y económicamente sostenible, priorizando la seguridad del dato, o la predicción de la difusión de enfermedades como el Dengue y Zyka en Colombia y Brasil a través de ONU Global Pulse. En https://luca-d3.com/es/big-data-social-good-es/index.html se analiza con mayor detalle el contenido de estos proyectos.

Tipo de colaboración:	• B2B (principalmente clientes de TELEFONICA). • B2G (AAPP del Estado y donación de datos en caso de *Big Data for Social Good*).
Sectores de los principales clientes de *LUCA:*	• Publicidad y Medios. • Servicios Financieros. • Comercio al por menor (retail). • Turismo. • Transporte. • Gobierno.
Servicios cross-sector	• Optimización de comunicaciones. • Ahorro Energético. • Gestión de flota. • Identidad móvil. • Márketing y publicidad.
Monetización del negocio:	Cobra por *insight* según el valor que le aporta al cliente. Oferta soluciones tecnológicas y servicios de consultoría para el establecimiento de una estrategia data-driven en las empresas clientes.
Principal obstáculo:	Hacer entender al mercado el verdadero potencial de los datos.
Factores de éxito:	La estrecha relación con sus clientes que les permite conocer sus necesidades y ofrecerles las soluciones en compartición de datos que mejor les encajan (modularidad). Los datos de la red móvil son un activo diferencial que otras empresas no tienen.

SUEZ

Suez, empresa multinacional que opera en la industria de la distribución y tratamiento del agua, así como la industria del tratamiento de residuos, desarrolló en el año 2017 una interesante iniciativa de apertura de datos gestada a partir de la celebración de un *hackathon*, que se denominó Hackath20n y con información disponible en http://www.hackath2on.es/.

Se trató de un evento en el que, durante un fin de semana, 20 equipos se reunieron para abordar, desde una óptica colaborativa, tres retos relacionados con el agua, en concreto: 1) la excelencia en la experiencia de cliente; 2) la promoción del valor social del agua; y 3) nuevos servicios a la administración.

La iniciativa ha sido de utilidad para la compañía, ya que su celebración ha ayudado a la empresa a diseñar nuevos productos y servicios, y ha animado la puesta en funcionamiento de otras iniciativas de datos abiertos en el futuro inmediato. Ha estado, por tanto, en consonancia con las expectativas.

En cuanto a las razones principales que explicaron la celebración del hackhaton, se apuntan como fundamentales la diversificación de la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del mercado y del cliente y la creación de nuevos modelos de negocio, a lo que se une también un propósito de carácter social, de bien común público.

Los obstáculos principales en iniciativas de datos abiertos tienen que ver, de acuerdo con la compañía, con riesgos de tipo regulatorio/legislativo, así como con la falta de acuerdos contractuales apropiados que garanticen el uso adecuado o la propiedad de los datos. En línea con esto último, uno de los factores principales de éxito tiene que ver con la disposición de un marco legal apropiado en relación con la protección de datos personales y la propiedad intelectual, además de las cuestiones de acceso e interoperabilidad (simplicidad y facilidad en el uso de los datos) y el entendimiento de la demanda de datos (necesidades, casos de uso, usuarios, etc.).

En cuanto a las recomendaciones, se pone el acento en la necesidad de marcarse objetivos cuando se lleva a cabo una iniciativa de apertura de datos con objeto de asegurar su éxito. Igualmente, resulta muy relevante poner los recursos necesarios en el ámbito de la seguridad y la protección de datos personales y sensibles desde el primer momento, ya que no tener esto en cuenta puede suponer tener que abortar la iniciativa sin que pueda llevar a término.

Tipo de colaboración y tipo de datos abiertos	• Datos Abiertos: Apertura de datos de forma gratuita para fomentar el desarrollo de nuevos productos y/o servicios). • Datos: - Datos de transacciones (clientes, proveedores). - Datos internos operativos (calidad, producción, ...). • Diversificar la oferta de productos y/o servicios de la empresa a través de un mejor conocimiento del mercado y del cliente. • Crear nuevos modelos de negocio.
Razón de la apertura de da	• Propósito social, de bien común público.
Factores de éxito	• Disponer de un marco legal claro en relación con la protección de datos personales y la propiedad intelectual. • Simplicidad y facilidad de uso de los datos (acceso, interoperabilidad, etc.). • Entendimiento de la demanda de datos (necesidades, casos de uso, usuarios, etc.).
Obstáculos	• Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial). • Falta de acuerdos contractuales apropiados que garanticen el uso adecuado de los datos, la propiedad, etc.

ITI

El Instituto Tecnológico de Informática (ITI), es un Centro Tecnológico de Investigación, Desarrollo e Innovación en TIC radicado en la Comunidad Valenciana y con actividad desde el año 1994. Aunque ITI no es una institución orientada a la generación de datos, su participación en un proyecto H2020 sobre datos abiertos da lugar a una serie de enseñanzas y recomendaciones que deben ponerse en valor.

Se trata del proyecto *Transforming Transport* (https://data.transformingtransport.eu/), aún vigente, en el que ITI ha participado en la elaboración de un piloto en colaboración con varias empresas nacionales e internacionales. El proyecto en su conjunto ha supuesto la realización de varios pilotos de apertura de datos en el ámbito de los sectores del transporte y la logística. Desde esta perspectiva, los distintos pilotos creados pueden adscribirse dentro de la categoría de plataforma de datos industrial, esto es, enfoque colaborativo y estratégico para intercambiar datos entre un grupo restringido de empresas.

Las dos razones fundamentales que han apoyado la implementación del proyecto es la contribución a la resolución de un reto (local o global), para lo cual se han creado los pilotos, así como apoyar la innovación abierta. En cuanto a las principales barreras, se citan tanto los riesgos de tipo regulatorio/legislativo (especialmente los que tienen que ver con el secreto industrial), la filtración de información sensible a empresas competidoras y cuestiones de tipo económico (alto coste de la iniciativa) y técnico (falta de interoperabilidad o de estandarización de los datos). De hecho, una de las razones de que los resultados del proyecto hayan estado por debajo de las expectativas tiene que ver con el hecho de que la mayoría de las empresas participantes realmente no han abierto datos, ya que eran públicos antes de la implementación del proyecto. Sólo en casos muy contados, y mediante acuerdos muy concretos y previa firma de cláusulas de confidencialidad se han abierto datos.

En coherencia con las barreras económicas y técnicas se citan como factores de éxito la realización de una inversión inicial acertada (o disponer de financiación de soporte externo) y la simplicidad y facilidad del uso de los datos, en términos, por ejemplo, de acceso e interoperabilidad. También la creación de un marco de confianza para garantizar la seguridad de los datos, lo que vuelve a poner de manifiesto la relevancia de los riesgos regulatorios, así como disponer de personal con cualificaciones y competencias necesarias.

En clave prospectiva, ITI plantea la realización de otras iniciativas de apertura de datos en el futuro, y en cuanto a las principales recomendaciones, se insiste en la necesidad de evaluar el coste de mantener una infraestructura de datos abiertos, que resultan costosas especialmente si los datos se consumen de forma constante y por parte de múltiples clientes. Otro reto interesante es la necesaria anonimización de datos sensibles del negocio que sea compatible con preservar el valor del dato.

Tipo de colaboración y tipo de datos abiertos	• Plataforma de datos industrial. • Datos: - Varían en función del piloto (los hay de localización, transacciones...), pero fundamentalmente la descripción de las fuentes de datos utilizadas y cómo acceder a las mismas. Existen datos de las categorías antes mencionadas de forma privada. La mayoría de los datos "abiertos" ya eran públicos con anterioridad.
Sectores de aplicación	• Distribución/logística. • Transporte.
Razón/es de la apertura	• Contribuir a la resolución de un reto local o global. • Apoyar la innovación abierta y colaborativa.
Factores de éxito	• Realizar una inversión inicial acertada o disponer de financiación de soporte externo. • Simplicidad y facilidad de uso de los datos (acceso, interoperabilidad, etc.). • Disponer de personal con las cualificaciones y competencias necesarias. • La creación de un marco de confianza, garantizando la seguridad y control de los datos.
Obstáculos	• Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial). • Alto coste de la iniciativa (en la preparación de los datos, en la extracción y tratamientos de datos, etc.). • Filtración de información sensible a empresas competidoras. • Problemas de tipo técnico/falta de interoperabilidad o estandarización de los datos.

EDP

La experiencia en apertura de datos por parte de EDP nace del profundo convencimiento acerca del papel que éstos asumen como herramienta de valor y de promoción de la innovación para la empresa. Su experiencia en datos abiertos se incardina fundamentalmente en la propuesta de lanzamiento de retos (*challenges*), vigente desde 2018, cuyo abordaje se hace a partir de la apertura de *datasets*, que son accesibles con un simple *click* (a través del link: opendata.edp.com), y de una gestión innovadora de los mismos por parte de personas y equipos que a ello se aventuran a modo de *"hackatones" on line*. Se trata, por tanto, de iniciativas de innovación abierta que usan los *datasets* como materia prima para abordar esos retos, que son premiados con recompensas (pueden ser monetarias o testimoniales). A modo de ejemplo, dos de estos retos son los siguientes:

- Detección de fallos en turbinas de viento. Se trata de un reto dotado de una recompensa de 5000 $. Ha implicado determinar, a partir de modelos *Big Data*, cuándo un equipo de estas características tiene más probabilidades de fallo y, en consecuencia, permite optimizar las tareas de mantenimiento de las turbinas.

- Predicción de energía solar. En este caso la recompensa fue de carácter testimonial (el reconocimiento de trabajar por hacer un mundo más sostenible). Basado en el proyecto SunLab de EDP (un laboratorio que tiene por objeto medir el desempeño de diferentes tecnologías fotovoltaicas que están instaladas en diferentes configuraciones y entornos), se trató de construir un algoritmo que sea capaz de predecir la producción de energía solar para los primeros siete días de 2018 a partir de un *dataset* de variables meteorológicas.

Siguiendo también este trasfondo de innovación abierta, hay una segunda modalidad en la que no existe como tal un reto que hay que abordar, sino que se trata de promocionar todos aquellos trabajos (tesis, publicaciones, creación de nuevos algoritmos, etc.), que han visto la luz gracias a la apertura de los *datasets,* promoviéndose de esta forma la colaboración con instituciones académicas, fundamentalmente Universidades. Un ejemplo es la Tesis Doctoral sobre distribución angular de la radiación solar (Universidad de Ciencias de Lisboa), que utiliza *datasets* del proyecto SunLab en su sede de Faro (Portugal).

Como resumen, las principales razones para la apertura de datos de EDP son apoyar la innovación abierta, a la vez que contribuir a la resolución de un reto local o global y a la mejora de la eficiencia de los procesos internos.

Por su parte, uno de los principales obstáculos tiene que ver con la falta de conocimiento del verdadero potencial (tanto interno como externo) de los datos y de los beneficios que pueden generar estas iniciativas. De hecho, uno de los principales factores de éxito es involucrar a agentes de todo tipo en el uso y la creación de valor con los datos, generándose de esta forma mecanismos de co-creación muy dinámicos y fructíferos. Otros obstáculos que también resultan relevantes son los riesgos de tipo regulatorio y la falta de habilidades adecuadas en la empresa para gestionar este tipo de iniciativas.

Tipo de colaboración y tipo de datos abiertos:	• Datos abiertos: apertura de datos de forma gratuita para fomentar el desarrollo de nuevos productos y/o servicios. • Se trata de datos internos operativos (calidad, producción...). En breve se pretende crear también datos en tiempo real derivados de algunos de sus activos.

Usuarios principales:	• Los datos se han utilizado para propósitos de I+D. Por tanto, se ha dado un uso desde una triple perspectiva:
	  • Servicios TIC, incluyendo desarrolladores app/ software.
	  • Educación.
	  • Investigación.

Razón/es de la apertura:	• Mejorar la eficiencia de los procesos internos. • Apoyar la innovación abierta y colaborativa. • Contribuir a la resolución de un reto global o local.
Principal/es obstáculo/s:	• Falta de conocimiento de los beneficios asociados a la iniciativa y/o del valor de los datos. • Riesgos de tipo regulatorio/legislativo (por ejemplo, colisión con LGPD o secreto industrial). • Falta de habilidades adecuadas en la empresa para gestionar iniciativas de este tipo.
Factor/es de éxito	• Entendimiento de la demanda de datos (necesidades, casos de uso, usuarios, etc.). • El intercambio de experiencias y conocimientos en la compartición de datos.

ARCELOR MITTAL

Arcelor Mittal, conjuntamente con TECNALIA, organizó los días 24 y 25 de septiembre de 2018 un datatón orientado a la búsqueda de soluciones "útiles y creativas" para resolver un problema real con las acerías a partir del análisis de datos. De forma más concreta, el datatón tuvo un objetivo doble:

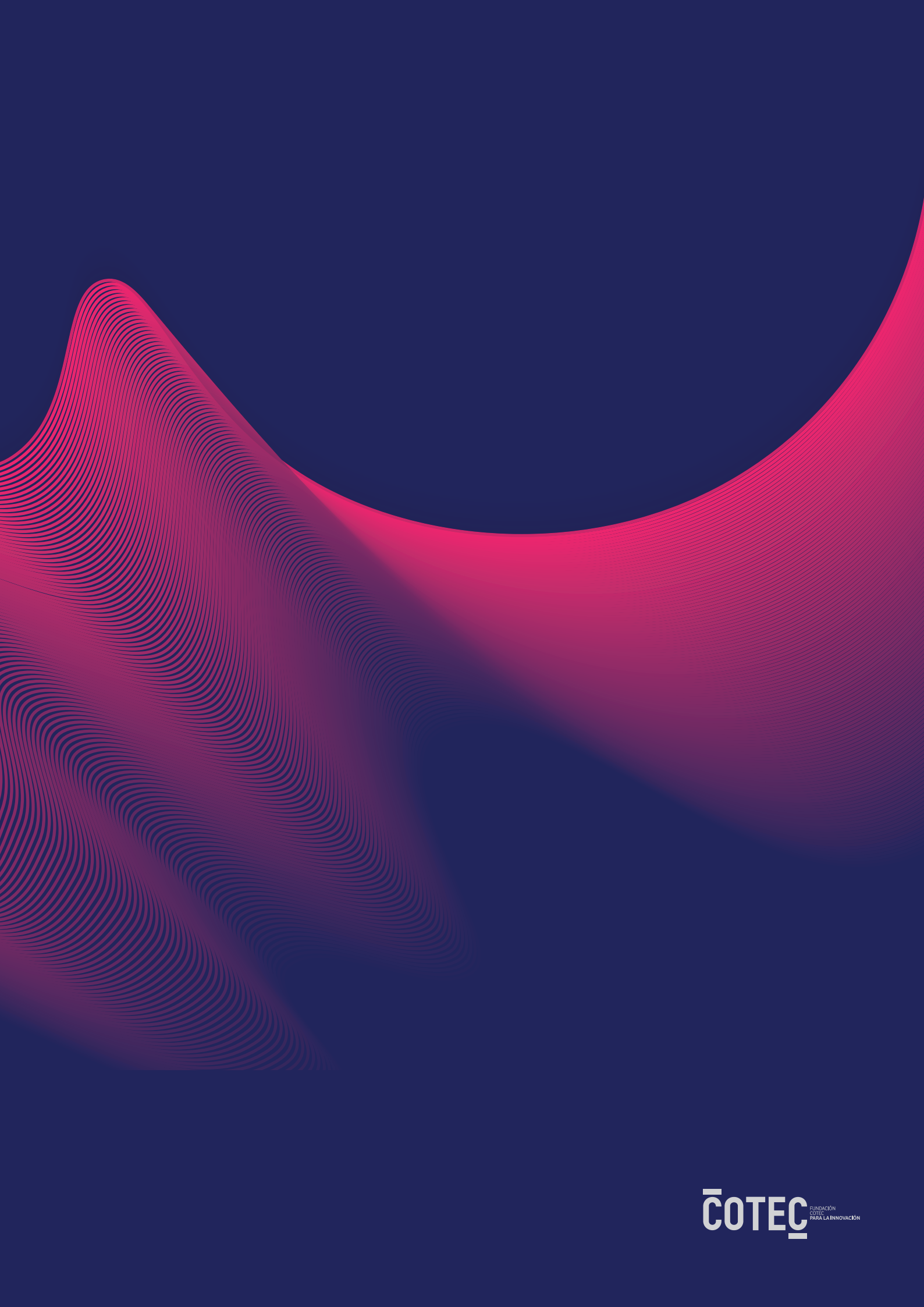
- Optimizar el rendimiento férrico de cada tipo de chatarra tomando como referencia datos reales del proceso de fusión de acería y teniendo en cuenta un conjunto de factores que influyen en el rendimiento.

- Predecir el contenido de cobre final en acero líquido en un momento puntual en el tiempo para cada tipo de chatarra que se mezcla.

La iniciativa contó con más de 20 participantes, entre ellos investigadores de Tecnalia, la Universidad del País Vasco y el BCAM (Basque Center for Applied Mathematics). Las propuestas fueron analizadas por un jurado compuesto por investigadores de Arcelor Mittal y TECNALIA, que valoraron aspectos como la calidad técnica de la misma, su novedad y aportación al "estado del arte", su aplicabilidad industrial, creatividad e innovación. La recompensa para la mejor propuesta fue el desarrollo de una prueba de concepto para una eventual implementación posterior.

Esta iniciativa de innovación abierta ha sido valorada muy positivamente tanto por ArcelorMittal ya que le ha permitido abordar un amplio conjunto de soluciones innovadoras con impacto en la mejora de sus procesos.

Tipología de colaboración y modalidad de datos abiertos:	• Datos procedentes del uso de los bloques de chatarra de origen (hasta 16), del proceso de fusión de acería y de los resultados: desde 07/2017 a 04/2018, 4.000 muestras consecutivas.
Sectores de los principales clientes	• Distribución/logística. • Servicios TIC, incluyendo desarrolladores app/software. • Automoción y transporte.
Monetización del negocio **Razón/es de la apertura**	• No hay monetización directa. • Explorar nuevas soluciones analíticas para la correlación del proceso de producción. • Apoyar la innovación abierta y colaborativa.


COTEC
FUNDACIÓN
COTEC
PARA LA INNOVACIÓN