A GUIDE TO OPENING AND SHARING DATA IN THE BUSINESS ENVIROMENT

Basics, recommendations and best practices

September 2019

COTEC
FUNDACIÓN COTEC PARA LA INNOVACIÓN

INDEX

TABLE INDEX

ILLUSTRATIONS INDEX

HOW TO USE THIS GUIDE

september 2019

FUNDACIÓN COTEC PARA LA INNOVACIÓN

1. HOW TO USE THIS GUIDE

The ***guide to open data and data sharing in the business world*** has two objectives: firstly, to explain the opportunities that opening up or sharing data offers to companies, and secondly, to help organisations wanting to embark on a process of this type. The guide is divided into different sections or steps to follow.

1. EXPERIMENTATION AND RAISING AWARENESS

Objective of this phase

Establish the potential of the company's data.

The company's baseline situation

This is a preliminary step before undertaking an open data or data sharing project.
This phase is aimed at companies that don't know the potential of the information they handle and want to identify it.

If the company has **internal data analysis expertise**

The company must involve the employees in an analysis of the existing data to determine its potential.
After identifying the potential of the data, the next step is to Define the Strategy.

If the company **does not have internal data analysis expertise**

The best course of action is to **seek external technical support** to advise the company on the value and potential of the data.

The guide includes different types of support:

- **Technology facilitators:** Under a collaboration arrangement, the company offers part of its data to a technology centre, university or specialised consultant to analyse and draw conclusions about their potential.
- **Datathon**: A competition where the company provides teams of experts with some of its data. The teams are set a specific time to experiment with the data and extract their value.
- **Marketplace / Industrial Platform**: Trusted online environments where companies share their data with other companies and with reusing agents.

Go to Launching an initial open data or data sharing experience.

Section **Experimentation and Raising Awareness**: page 32

- **More information:**
  Marketplaces and Industrial **Platforms in data sharing channels**, page 41.
  Relationship models or business for data exchange, page 53

Next phase

How to define a good open data or data sharing strategy

Optional preliminary step

Launching an initial open data or data sharing experience

Launching an initial open data or data sharing experience

If the organisation has no prior experience in data sharing or open data processes, it is a good idea to start with an experimentation process to ascertain the value of the data. This type of project is also useful for understanding the components of a data sharing or open data process before embarking on an initiative of this type.

In the first phase, several ways to start sharing or opening up data if the organisation has no prior experience have been described. In the case of Datathon and Marketplace, it should be remembered that this entails sharing or opening up data with third parties. Although there are other models, we will focus on these two formulas.

1. Definition of the objectives and action plan

- Establish the **goal** to achieve through this first experience. In the case of the datathon, it could be finding a solution to a specific corporate challenge. Regarding the marketplace, the goal could be to check the value of the data for a future process, which may involve monetisation, among others.
- Establish which **dataset** will be shared or opened.
- Decide which **relationship model** will be applied to the initiative. Because this is an initial experience, this itinerary focuses on the Datathon and Marketplace modalities.
- Find **expert support**: it is important for the organisation to be supported by technological expertise during the process.
    - **Datathon**: the initiative can be carried out through a collaborating agent or technological facilitator such as a technology centre, university or specialised consultant. Examples of datathons: ArcelorMittal and Suez, page 32 and 33.
    - **Marketplace**: the platform itself may have advisory services for organisations with no experience in the processes of sharing or opening up data.
    - **Legal advice**: it is essential to get legal advice to ensure that the data made available to third parties complies with the applicable law.

More information:
Relationship models, page. 18
Legal framework For different relationship model, page. 60
Examples of open data and data sharing, page 102
Law applicable to open data and data sharing, page. 70

2. Organisational and technical training

This phase includes aspects related to data management and technology infrastructure. Although this first experiment does not require a comprehensive action plan, it is a good idea to consider some aspects of management processes, training and data transformation. This information will be useful for understanding the necessary organisational and technical infrastructure changes that will be needed if it is decided to open up and share the data in the future.

Aspects to consider:

- Analyse the **format** of the datasets to share and whether they are usable by third parties.
- Establish the **extraction, transformation and sharing processes** to carry out, with expert support.
- Establish the **data collection, preparation and publication** processes to carry out to obtain quality datasets that can be re-used easily by third parties, also accompanied by expert support.
- Decide on the **open licences** to apply to the published data.
- Establish measures for **monitoring** the initiative to be able to measure the results achieved.

More information:
Extracting, transforming and sharing data, page. 40
Starting the dataset life cycle (Process for generating datasets), page. 46
Licenses, page 63
Monitoring the initiative, page. 51

---

3. Communication and raising awareness

After this first initiative it is important to collect the results of the experience, the benefits and lessons obtained, and to inform the rest of the organisation, particularly at an executive level. This makes it possible to share the opportunities of sharing and opening up data, and is also useful for understanding the decisions that must be taken at a strategic, methodological and technological level in a process of this type.

Next phase

How to define a good open data or data sharing strategy

---

2. HOW TO DEFINE A GOOD OPEN DATA OR DATA SHARING STRATEGY

Objective of this phase

Before starting a data or data sharing strategy, it is important to have a clear strategy that defines the objectives, the strategy to adopt and the decisions to take at an organisational level.

The company's baseline situation

The organisation has already identified the potential of the data and wants to start opening up or sharing its data with third parties.

Steps included in this phase

1. Baseline diagnosis

A first step which consists of getting a clear overview of the organisation's current situation vis-a-vis its data. To do this, it is advisable to identify key people in the organisation who can help to make this diagnosis. Questions to ask:

- Which units of the organisation collect, use and produce data?
- How are they managing those data?
- Is there data that is already opened up or shared? If so, how are they being published?
- Which datasets do you want to share?
- What is the current data flow process like in the organisation?

2. Defining the strategic objectives

The next step consists of determining **what you want to** achieve by sharing or opening up the data and the timeline for doing so. Several decisions need to be taken here:

- Clarify the objective: generate a new line of business, encourage innovation, solve a problem, monetise the data, etc.
- According to the objective, determine the **Relationship Model** (or business model) for sharing the data. This could be a Business to Business (B2B) model or a Business to Government (B2G) model, following different modalities (publication on an open data portal, monetisation, sharing on an industrial sector platform.
- Identify **internal expertise** to carry out the initiative. Will I need to get specialised external support?
- Will I need to get **legal support**?
- Define indicators associated with the objectives to be able to measure the results.
- How to visualise the flow of destination data.

Related information:
- Clarify the objective: **Why do companies share/open up their data?**, page 30
- **Relationship models,** page 53
- **Legal framework For different relationship models,** page 60
- **Principles guiding the Relationship Models,** page 78
- **Roles, skills, development and knowledge management,** page 66
- **Examples of open data and data sharing,** page 102

3. Define the action plan and implement the strategy

Establish the plan for achieving the objectives. The main aspects of this phase are:

- Secure the involvement of the people assigned to the initiative (internal or external).
- Establish measures to ensure the **viability** of the initiative through a cost-benefit analysis.
- Define the **data management** model that best fits the objective (This aspect is developed in the Technical training phase).
- Analyse data **formats** and whether they are usable by third parties. Identify whether there is any sensitive data.
- Establish which **licences** to use.
- Define and establish ETPD processes and identify the technology required (This aspect is developed in the Technical training phase).

Related information:
- **Data Management** structures, page 38
- **Financing models** and methods for quantifying the value of the data, page 57
- **Licences** for sharing or opening up data, page 63
- **Technical training** before launching the open data initiative, page 38
- Tools for **Monitoring** reuse of the data, page 51

4. Defining and maintaining the data inventory

Focused on establishing the aspects to consider regarding the datasets:

- Ensure **regulatory compliance** with regard to data protection and identify whether there is any **sensitive data**.
- Quantify the **economic value** of the datasets based on the financing model.

Related information:
- **Law** applicable to open data and data sharing, page 79
- Management of **sensitive data**, page. 49
- **Valuing datasets** and calculating tariffs, page 57

5. Defining the open or shared data policy and communicating it

Defining and announcing a policy at organisational level is essential because it promotes transparency in the organisation, helps to raise awareness of the benefits of opening up data and adds to the success of the initiative.

Section Define the Strategy: page 35

Next phase

How to steer the company towards open data - Technical Training

3. HOW TO STEER THE COMPANY TOWARDS SHARING OR OPENING DATA: TECHNICAL TRAINING

Objective of this phase

During this phase, decisions regarding technical training and the implementation of the data sharing or open data process are made. This phase includes aspects related to data management and the technology infrastructure to be used.

Baseline situation

The organisation has already identified the potential of its data and has made essential decisions regarding how it will carry out the project by defining the strategy. It has already decided the goal of the initiative and the relationship model to follow, the expert team, the data it will share and the financing model for the project.

Technical decisions to take during this phase

1. How will the data be managed

The organisation must decide upon the management structure or governance of the data. In the absence of this, it is recommended one be established before starting to share or open up the data. Theoretically, this could be:

- **Decentralised data management**:
  The data is managed by units or persons who are responsible for managing their own data.
- **Centralised/federated data management**:
  Data management is shared, but responsibility is centralised.
- **Fully centralised data management**:
  Standardisation is defined at the highest level, ensuring consistency throughout the organisation.

Shifting towards a data management structure has a significant impact on the organisation, so if there is no prior experience, it is advisable to seek support from expert data management agents.

More information: - Data Management page 38

2. Extract, Transform and Share Data Process ETSD

The ETLD process consists of the technical specifications of how data flows within an organisation, is transformed into publishable datasets and is eventually shared. This process must be defined at an organisational level.
It consists of three steps:

Extract: Data can be extracted from any type of source (internal or external). The extraction process will depend on the data management structure.

Transform: This should be transformed into lists of open datasets. This phase must be standardised by defining a data processing guide.

Share: This phase is unnecessary if the datasets are for internal use. It consists of identifying the channels where the data is shared or opened up (next section).

The specific process used to prepare each dataset, including quality aspects, inclusion of metadata and legal aspects, are addressed in phase Data life cycle. page 46

For more information about the ETLD process, see **Annex I – ETSD processing scenarios**, page 74

---

3. Data sharing channels

There are various channels through which to share or open up data. The choice of channel depends on the number of datasets to share and the specific needs of each organisation. To start with, it is a good idea to keep things simple. The main channels are:

- Downloads from the website
- Application Programming Interface (API)
- Data Portal (own or third party)
- Marketplace
- Industrial Platform
- Technical Facilitator

More information: **Data Sharing Channels**, page 41

---

4. Data Searches and other technical aspects

It is advisable to implement functions to enable data searches. Mechanisms of varying complexity can be used to share the datasets according to the distribution channel chosen.
Other technical decisions can also be taken regarding the selection of the domain and hosting and management of the data sharing channel.

More information: **Data Searches and other technical aspects** page 44

Next phase

Implementing the data life cycle: collect, prepare, publish and maintain.

4. IMPLEMENTING THE DATA LIFE CYCLE: COLLECT, PREPARE, PUBLISH AND MAINTAIN

Objective of this phase

Establish specific steps to follow to generate datasets with the data to prepare it for sharing or opening up. It includes data collection, preparation, publication and maintenance.

The company's baseline situation

The company has already established its strategy for opening up or sharing its data and has made organisational and technical decisions regarding:

- How the data will be managed.
- How the ETLD process will be addressed.
- Which data sharing channel will be used.

Steps included in this phase

1. Data collection

This phase consists of:

- **Mapping** and identifying the available data.
- **Prioritising** the most suitable data (because they are in line with the objective, they are of a higher quality,...).
- **Classifying** the datasets and **publishing** the categories.

More information: **Data Collection** phase, page 46

2. Processing the data

Once the data has been collected it must be prepared for opening, because it cannot be published in raw format. **This is a key phase, because if the data does not comply with the applicable law, it cannot be shared or made open**.

There are several aspects to consider:

- Determine the **quality of the data**: If they are complete, clean and accurate, if they are up to date and comply with established standards.
- **Prepare the data for publication**:
  Data should be published using the **linked data** method to enable interoperability. It must be associated with **metadata**, which is structured information that describes the dataset and makes it discoverable.
- **Legal preparation:** It must be checked that the data complies with the applicable law. In addition, each dataset must be protected by a license that establishes the conditions of use.
- **Sensitive data management:** If the datasets contain sensitive data, they need specific processing.
- **Final check** to ensure that all the above aspects of the datasets are in order (quality, interoperability and inclusion of metadata, licences...).

More information: **Data Preparation** phase, page 47

Related information:
- What is **Metadata**, page 51
- Methods for **determining dataset quality** (MELODA and Tim Berners-Lee's 5 Star Linked Data), page 50
- Technical training for carrying out the **Extract, Transform and Load Data process,** page 40
- **Licences for sharing or opening up data,** page 63
- **Law** applicable to open data and data sharing, page 79
- **Sensitive** Data Management process, page 49

3. Publishing the data

The data is published in distribution channels or established publication. The choice of channel is addressed in the earlier **Technical training phase.**

More information: **Data Sharing Channels**, page 40

4. Maintaining the data and metadata

Data and metadata can change in time and become obsolete. Hence the recommendation to define a process for maintaining open or shared datasets. There are several aspects to consider:

- Update the data regularly and include the date in the metadata.
- Check that the URI and URL are still active.
- Include means of contact so that reusing agents can submit feedback.

More information: **Maintaining data and metadata**, page 51

Section **Starting the life cycle of open or shared datasets,** page 42

Next phase

Measuring success: monitoring the sharing or open data initiative

5. MEASURING SUCCESS: MONITORING THE SHARING OR OPEN DATA INITIATIVE

Objective of this phase

Follow up the entire initiative to measure the level of success.

The company's baseline situation

The company has now implemented an open data project and must establish a method for following it up.

Main actions to take to monitor the initiative:

Ascertain the commitment of the organisations and reusing agents

The commitment of reusing agents is critical to the success of any open data or data sharing initiative.
Tim Berners-Lee's 5 Star Linked Data model is a useful tool for measuring the level of interaction offered by the data.

Measure the success of the initiative and the level of continuous improvement

There are several indicators for measuring the success of the initiative:

- Calculating the number of times that datasets are downloaded.
- Calculating the number of times that the data is accessed.
- Using qualitative measures to evaluate the usefulness of the datasets.

PIWK is a free, open source tool for qualitative analysis of data according to the needs of reusing agents.

Related information:
- **Methods for determining dataset quality,** page 50
- **Maintaining data and metadata,** page 51

Section **Monitoring the data sharing or open data initiative**, page 51

Relevant aspects of a data sharing or open data initiative

The guide includes several key questions to consider when defining and implementing a data sharing or open data initiative. The most relevant ones are:

- Relationship or business models for sharing or opening up data
- Financing models
- Legal framework and principles that guide the relationship or business models
- Licences for sharing or opening up data
- Roles, skills, development and knowledge management

Relationship models (or business models)

The organisation can share or open up its data following different relationship or business models according to the strategic objective of the initiative: opening up new lines of business, monetising the data, finding new collaborations, or others.

There are two models, Business to Business (B2B) or Business to Government (B2G) and different modalities. Among these are:

- **Open/Shared Data:** The data is made available through an open range of reusing agents.
  Examples:
  BBVA Valora, page 53
  Primafrio, page 102

- **Monetising the data** (B2B or B2G): Unilateral focus on obtaining additional income for businesses from the shared data.
  Example: Euskaltel, pág. 104

- **Marketplace** (B2B): Data providers and reusing agents come together on an online platform through trusted intermediaries.

- **Industrial Platform** (B2B): Data is shared in a closed, secure environment, among a restricted group of companies.
  Example: ITI, page 110

- **Technology Facilitators** (B2B): Technology facilitators who act as third parties, offering data exchange solutions or consultancy services.

- **Data donation** (B2G): This is a type of corporate social responsibility.
  Example: Telefónica – *Big data for social good,* page 106

- **Collaboration agreements** (B2G): Agreements between public organisations and companies to exchange data.

Information about **Relationship Models**, page 53
Examples of **open data and data sharing**, page 102

Financing models

The organisation must be capable of quantifying the economic value of the information and the costs associated with the data sharing or open data initiative in terms of infrastructure and personnel.

To do this, there are several important practices to take into consideration:

- **Ensure the viability of the initiative**, by means of a cost/benefit analysis and identifying tangible and intangible impact, such as income from new products and services and savings from new processes or improvements to the company's brand image.

- Having a **method to establish the value of the datasets and establish tariffs** in case of data monetisation. There are several criteria to consider that can be used to quantify the value of the datasets and calculate tariffs.

Information about **Financing Models**, page 57

Legal framework of contractual agreements

Normally, data sharing is based on a prior contractual agreement. Several aspects must be considered when preparing these agreements according to the type of relationship model.

In **B2B models:**
Some aspects to consider when drafting agreements

- Which data will be available?
- Who can access and reuse the data?
- What can reusing agents do with the data?
- What technical measures will be used to access the data?
- Which data require protection?
- What is the duration of the agreement?

In **B2G models**:
Specific aspects to consider when drafting agreements with government agencies:

- What is the public interest purpose?
- What are the objectives and limitations of data sharing?
- What are the conditions for implementation?
- What is the compensation for sharing the data?

Information about:
- **Legal Framework,** page 60
- **Principles guiding the Relationship Models,** page 78

Licences for sharing or opening up data

All shared and open data must include explicit information about the conditions of use. There are different types of licences that make it possible to establish the requirements, limitations on use and areas of application of the agreement.

Information about **Licences for sharing or opening up data**, page 63

Roles, skills, development and talent management

Roles

As part of the initiative to share or open up data, there are different roles necessary to implement and maintain them, including both strategic and technical experts. The main ones are:

- **Sponsor**: responsible for establishing a strategy for the organisation and supervising its implementation.
- **Data manager**: responsible for implementing the data sharing strategy.
- **Implementer**: responsible for implementing technical requirements.
- **Data owner**: responsible for deciding who has access and can use each database.

Skills and knowledge development

The organisation must ensure that staff are suitably trained and have support to fulfil their data management responsibilities. If necessary, the organisation should offer training to develop these skills.

Information about **Roles, skills, development and talent management,** page 66

---

1. INTRODUCTION

Current advances in digital technologies have led all types of organisations to permanently and exponentially generate large amounts of data. This information comes both from the organisation's internal systems and from its interaction with users, suppliers and customers. Additionally, in the case of companies, the expansion of technologies such as the Internet of Things (IoT) and Artificial Intelligence (AI) applied to production processes mean that much of this data is generated automatically.

The internal collection and analysis of data involves important benefits for organisations, such as process optimisation and improved decision making, among other aspects. However, the sharing or opening of these data to third parties also offers great opportunities for economic growth, competitiveness and innovation.

In the case of Public Administrations, the opening of data has come a long way and it has clearly demonstrated the benefits it entails: according to the European Commission, the direct market for open data will generate approximately 325,000 million euros in the period 2016-2020, as well as 100,000 jobs[1]. In the case of the private sector, the sharing and opening of data is less widespread, although more and more companies are realising its advantages. This guide aims to review the benefits of sharing data in business environments, as well as collect a series of recommendations and best practices to enhance the re-use of private data.

[1]European Commission (2015): Creating value through Open Data

1.1 Why should I share or open my company's data?

The main benefits of sharing and/or opening data are the opportunities to generate new lines of business within the company and improve internal efficiency. There are also other reasons to share or open data, either in B2B environments (with other companies) or B2G (with public administrations and universities or research organisations). Some of these reasons can be the monetisation of data as a way to generate additional benefits, the opening of collaborations with other organisations, the strategic positioning of the company or the support for innovation.

> *Success story: BBVA Valora*
>
> *The BBVA Valora project has created a new online service based on data, which offers useful information before buying or renting a home and allows to manage all the information related to mortgages and household expenses in a single place.*

Data alone do not generate value; it is necessary to foster an ecosystem that can transform data into useful information and new business opportunities

1.2 What is the difference between data sharing and opening?

The opening or sharing of data in the business environment means making the data generated internally by the company available for re-use by third parties, such as other companies or public bodies. However, there are substantial differences between the two concepts:

> *Success story: Primafrio*
>
> *Company specialised in land transport of refrigerators. The company offers its customers free data related to the location and positioning of its fleet. As a result, it has managed to optimise the tracking of orders and developed greater speed in invoicing processes.*

Opening of data: It consists of offering data in an "open" way, through licences that allow to re-use the data so that anyone can access them, use them and share them under minimum restrictions. These limitations include compliance with data protection legislation and regulations for the re-use of open data[2]. Open Data also entails displaying such data by using open standards and formats, which facilitate their re-use.

Data sharing: It is a process through which a company makes its data available to other companies who are neither competitors in the market where the company operates, nor subcontractors, and are interested in using the data for their own business purposes[3].

In data sharing, provider companies have greater control over the data they make available to third parties. These sharing initiatives can take different forms, with unilateral or collaborative models, and data can be offered for fees, in exchange for the provision of services or free of charge.

> *Success story: ITI*
>
> *The Instituto Tecnológico de Informática [Technological Institute of Computer Science] participates in the European project Transforming Transport. The initiative consists of an online industrial platform where participating companies share their data in the field of transport and logistics in order to achieve greater efficiency in logistics processes.*

[2] Royal Decree 1495/2011 on the re-use of public sector information, current Spanish legislation. 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, new European legislation.

[3] European Commission (2016): Study on data sharing between companies in Europe

Since it is a concept that is still not widely used in the business field, it is important to bear in mind that:

- Initiating a data sharing process does not necessarily mean that the company will offer access to all its datasets. The company must decide what data it will share based on its business strategy.

- Sharing does not mean that data is offered for free. There are several business models that include pay-per-access.

- Data providers are the ones who decide with whom they will share their data, on what terms, under what conditions of use and the time frame in which they may be used.

1.3 Where to start?

As it was already mentioned, the opening and sharing of data in the business environment offers significant opportunities, both for large companies with previous experience in this field, and for SMEs who are not used to handling large amounts of data. In addition, with the expansion of the so-called Industry 4.0, many companies without previous experience in Big Data management have started to generate large volumes of data, and their interest in seeking new opportunities to enhance their value has increased.

> *Success story: ArcelorMittal*
>
> *In 2018, this steel company, through its R&D centre in the Basque Country, started up a datathon aimed at solving an existing problem in their steelworks. The datathon was in collaboration with Tecnalia. The company made part of its data available for a limited time so that groups made up of researchers could come up with innovative solutions.*

It is important to bear in mind that the processes of data opening or sharing involve changes both at the level of technical infrastructures and in the organisational processes—and, in short, in the company's culture. Therefore, it is advisable for those companies who want to take this road but do not have previous experience to start experimenting through small projects, opening part of their data to facilitators or technology platforms in order to evaluate its potential. This guide offers different alternatives to carry out these processes.

> *Success story: Euskaltel*
>
> *The company has developed a Big Data analysis tool based on data from connections to telephone antennas. As a result, the company obtains reports and offers them to other companies and public administrations.*

1.4 What are the contents of this guide and who is its target audience?

The Guide to Opening and Sharing Data in the Business Environment aims to be a tool to be used by all those companies interested in opening or sharing their data, whether or not they have previous experience in this field. It is therefore an informative, non-academic document that aims to bring the potential of Open Data closer to the private and business sectors.

> *Success story: SUEZ*
>
> *For a weekend, the company made its data available to 20 teams, in order to address various water-related challenges by working collaboratively. The results helped the company design new products and services.*

There are already numerous examples of companies that have launched initiatives to share their data with third parties, as well as startups that have found a new business niche in the field of data sharing. These examples show the potential of these processes for both data providers and re-users.

> *Success story: EDP*
>
> *The company organises online datathons where it presents challenges to solve, making its datasets available as raw material to develop new solutions. It also offers its data to institutions for research projects. Mode: Datathon and B2G (universities and technology centres).*

These pages include a series of recommendations and best practices, and the guide describes the steps to follow to start the process of opening or sharing data:

a) Experimentation and understanding

b) How to define a good open or shared data strategy

c) Technical preparation, how to steer the company towards the opening or sharing of data

d) Implementing the data life cycle: collecting, preparing, sharing or publishing and maintaining

e) Monitoring the initiative, analysis of the level of success and continuous improvement

a) Experimentation and understanding. Its purpose is to know the potential of an organisation's data. It is the recommended early stage for an organisation that is not aware of the potential of its data.

If the organisation does not have internal knowledge in data analytics (Big Data), the most suitable action is to have the support of external staff who have this knowledge and can advise on the value and potential of the data handled. External support can come from various sources, such as **Technology Facilitators**, organising open **Datathons** or joining **industrial marketplaces/platforms** that offer this type of services.

b) How to define a good open or shared data strategy. The aim is to define, establish and review the organisation's open/shared data policy and strategy based on the organisation's actual initial situation. Such a strategy influences both the sharing/opening and re-use of third-party data and is an organisational transformation process that should be part of a broader digitisation strategy. Some best practices in this regard are:

- *Knowing the initial situation of the organisation regarding open/shared data.*
- *Knowing the target situation of the organisation regarding open/shared data.*
- *Defining and implementing the appropriate strategy to achieve the objectives.* Derive and establish the improvement plan for data sharing/opening; Define the appropriate data governance model; Define the organisational processes for deriving and sharing/publishing datasets and for monitoring their re-use by third parties and, if necessary, deploy the change management process associated with it.
- *Defining and maintaining the organisation's data inventory.* In the inventory, include both internal datasets and re-used datasets from other organisations.
- *Defining the open/shared data policy, disclosing it and making it public.*

c) Technical preparation, how to steer the company towards the opening or sharing of data. The aim is to help the organisation make the right decisions regarding data management and associated organisational changes, as well as regarding the technological infrastructure to be put in place to extract and process the data prior to opening/sharing, all according to the organisation's goals and strategy. Some relevant aspects to consider:

- *Data management.* Covers all the organisation's datasets (those generated internally, as well as third party datasets). There are different theoretical types of data management (decentralised data management, centralised data management and fully decentralised data management—the latter is ideal for data sharing/opening). Each organisation should adopt the most appropriate data management type according to its specific situation (reality and goals of the data targeting initiative).
- *Extract, Transform and Share Data (ETSD).* It involves defining the technical specifications to implement a system/platform that enables the flow and the transformation of data that the organisation needs to finalise the sharing or

opening of data. The ETSD process, along with the implemented technology, is the central process that will be executed repeatedly to obtain the datasets for internal use or to share with/open to third parties.

- *Data sharing channels*. These are the different technical means available for the distribution of the organisation's data to third parties[4]: Web, API, Data Portal, Marketplace, Industrial Platform or Technical Facilitators. An organisation can distribute data through more than one channel, change the distribution channels over time according to the purpose of the sharing/opening of data, etc.

- *Data search*. There are BASIC (e.g. CMS), ADVANCED (using SPARQL) or universal dataset-specific search engines (e.g. https:/toolbox.google.com/datasearch). The use of one specific tool may depend on the data sharing/opening strategy.

- *Other technical aspects to be considered* in a sharing/opening initiative: Where will the data be stored? Where will they be shared? Who will manage the distribution channels? What service will the organisation offer, etc.

d) Implementing the data life cycle: collecting, preparing, sharing or publishing and maintaining. It describes the steps to follow to generate each dataset to share/publish. The relevant aspects are addressed, such as data quality, legal preparation, management of sensitive data, etc.

- *Data collection*. For organisations without experience in sharing/opening of data, we recommend to follow a process to identify and prioritise the data to be shared/published.

    1. *Data preparation*. Legal data preparation for sharing/opening. This aspect is key for sharing/publishing. If the data do not comply with current regulations and do not have the appropriate licences, they cannot be shared.

    2. *Data quality*:

    - Content: Are the data complete? Are the data clean? Are the data accurate enough for the relevant purpose? We suggest verifying empty fields, erroneous values, double entries, sensitive private information, etc.

    - Promptness: How often are they updated? Are they real time data?

    - Consistency in the presentation of data in order to facilitate its re-use by third parties.

    - Data quality can be measured by its level of compliance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) by including appropriate metadata and persistent identifiers for identification and re-use.

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

3. Determining the *publication format of the dataset* and facilitating access to the dataset. To this effect, it will be necessary to add metadata, use standard ontologies, etc.

4. Defining a process for processing sensitive data to be shared/published.

- *Data publishing*. Done through the identified data sharing channels.

- *Maintaining data and metadata*. Define a process to keep them up to date and identify improvement actions in the data life cycle based on feedback from re-users.

e) Monitoring the initiative, analysis of the level of success and continuous improvement. The aim is to audit both the life cycle of the dataset from its sharing/publication (data traceability and re-use in different scenarios), as well as to monitor the data sharing/opening initiative of the organisation as a whole (from policy and strategy, established processes, etc.). It is recommended to base improvement decisions on qualitative and quantitative metrics and, also, to contrast the engagement of the re-users by using formal methods such as that of Tim Davies[5]).

1.5 Relevant aspects of the opening and sharing of datas

For an open/shared data initiative to be sustainable, these **relevant aspects** must, at least, be taken into account:

- It has to be based on a viable **relationship or business model**.

- It must have an associated **financing model** that is profitable.

- The existing **legal framework** must be rigorously respected.

- The organisational culture needs to be changed and **qualified staff needs to be available to manage and share/open the data**.

Most marketplaces or industrial platforms offer technological solutions and services to organisations, mainly small and medium, to address these relevant issues in a simple way. That is to say, they offer their clients the possibility of sharing technology platforms and the common investment costs of this type of initiatives.

1.6 Who made this guide?

This guide was written as part of the activities carried out by the Working Group of the Fundación Cotec de Datos Abiertos [Cotec Open Data Foundation], which ran from September 2018 to June 2019. The development of this Working Group was coordinated by Cristina Oyón and Amaia Martínez, of SPRI's Department of Strategic Initiatives, the Basque Business Development Agency.

Additional collaborators were: Ana Ayerbe, Elixabete Ostolaza and Izaskun Santamaría of the technological corporation Tecnalia; Diego Moñux and David Gago

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

of the consultancy firm SILO Science & Innovation Link Office; Alberto Abella as Cotec's expert in open data; Irene Lopez de Vallejo of Ocean Protocol; Eva Méndez of Alianza 4 Universidades (Universidad Carlos III from Madrid) as expert in Open Science and research data; and Álvaro Ramos of Clarke & Modet as advisor on legal issues.

The guide compiles the contributions made by the entities that participated in the Cotec Working Group, collected during the 4 working sessions held and through the questionnaire and interviews conducted, as well as contributions from other entities that have shared their experiences (Arcelor Mittal and Euskaltel).

Designated members:

1. **La Rioja Economic Development Agency *(ADER, Agencia de Desarrollo Económico de la Rioja)*:** Almudena Larrea, Head of Statistics and Justification of Funds; and Javier Solozábal, Head of the Planning and Monitoring Programme.
2. **Alberto Abellá**: Expert of the 100 de Cotec group of experts.
3. **Alianza 4 Universidades**: Prof. Eva Méndez, Assistant Vice President for Open Science at Universidad Carlos III from Madrid.
4. **Banco de Santander**: Lara de Mesa, Director of Public Policy.
5. **Bankia**: Ignacio Cea, Corporate Director of Innovation.
6. **BBVA**: Iván Moreno, CTO of New Digital Business; and Ana Isabel Segovia, Digital Regulation, Supervisors, Regulation and Compliance Department.
7. **Clarke Modet**: Álvaro Ramos, Data Protection Delegate.
8. **Department of Economy, Knowledge, Enterprise and University *(CECEU, Consejería de Economía, Conocimiento, Empresa y Universidad)* of the Andalusian Government**: Manuel Ortigosa, General Secretary for Enterprise, Innovation and Entrepreneurship.
9. **Ecoembalajes España**: Manuel de Arcocha, Director of Technology; and Daniel Moradiellos, Strategic Office.
10. **EDP Spain**: Luis Manuel Santos, Head of Innovation, Directorate of Environment, Sustainability, Innovation and Quality; and Ernesto Barroso, Head of Emerging Companies.
11. **F. Iniciativas**: Enrique Blas, Partnership Manager.
12. **Vodafone Foundation**: Luz Olivas, Project manager in the Directorate of Public Administration; and Carlos Becker, Head of the Public Administration Segment.
13. **Indra (Business Consulting Minsait)**: Francisco Javier Ruiz, Partner; and David Fernández, Head of CyberSecurity Business.
14. **Valencian Institute of Business Competitiveness (*IVACE, Instituto Valenciano de Competitividad Empresarial*)**: Daniel Sáez, Technology Transfer Director – ITI; and Francisco Valverde, R&D Projects Manager – ITI (Instituto Tecnológico de Informática).
15. **Orange**: Juan Vallés, Big Data Product Owner.
16. **Primafrío**: Adrián Valverde, Head of R&D&I management.
17. **Suez Advanced Solutions**: Catalina Balseiro, Director of Innovation.
18. **Tecnalia**: Ana Ayerbe, Director of ICT Business Area; Izaskun Santamaría, Researcher; and Elixabete Ostolaza, Senior Consulting.
19. **Telefónica**: Richard Benjamins, Data & AI Ambassador; and Estanislao Fernández, Project Manager at Telefónica I+D.

This project has received technical support from the Department of Studies and Knowledge Management of the Cotec Foundation.

2. OPENING AND SHARING DATA IN THE BUSINESS ENVIRONMENT

Data plays a fundamental role in the era of digital transformation. It is the "new oil" that politicians, technicians, scientists and strategists refer to. Many governments, as well as most of the large private organisations whose business involves the intensive use of data (e.g.: banking, insurance, telephone operators, etc.), are aware of its potential and impact, both in improving the internal performance of organisations and in economy and society in general. A wide range of possibilities opens up with the sharing and/or opening of an organisation's data with third parties—both for the organisation itself and for third parties.

Currently, **innovation based on data** is considered one of the key facilitators for the growth of countries and the creation of new jobs in Europe. The importance of data produced and processed online, the continuous growth of data generated by objects connected to the Internet of Things (IoT), the growing availability of Big Data analysis tools and the availability of an increasing number of Artificial Intelligence (AI) applications are key technical facilitators in the digital transformation. The non-competitive nature of data itself makes it possible for the same data to be used as the basis for a whole range of products and services or new production methods. This leads to the conclusion that it may be beneficial for companies to share more data with others, so that the source data can be exploited to the maximum degree.

However, data alone do not generate value, and it is necessary to foster an ecosystem[6] that provides the context for transforming data into information[7], and information into knowledge.

Since the beginning of the data economy, the public sector has been an important contributor, both in defining and adapting the legal frameworks, regulations and recommendations to facilitate and promote this economy, and for its role in encouraging the implementation of initiatives for the opening and sharing of data in both public and private organisations. The private sector falls behind when it comes to this process of data opening and sharing, although some interest is beginning to be aroused among companies and some successful initiatives have begun to be implemented. We should not forget that new technologies have turned many organisations that traditionally did not work intensively with data into "data-intensive" organisations and therefore ideal candidates for data sharing/opening initiatives.

- When the organisations that share/open the data are private organisations, there is talk of B2B (Business to Business) and/or B2G (Business to Government) sharing/opening of data. The first refers to the sharing/opening of data between private organisations; the second, to the sharing/opening of data between private organisations and the Public Administration. A private organisation may have data shared with other private organisations (B2B), with Public Administrations (B2G), or with both, depending on its open data policy and strategy. In this sense, the final report of the European Commission, which includes a study on data sharing practices between companies in Europe[8], describes 16 successful examples of data sharing in private companies in different sectors, which can be an inspiration for the sharing of business data. The report also draws the following conclusions: Around 40% of the companies participating in the study publish and/or re-use data.

- A considerable proportion of companies that do not yet share/publish their data expect to do so in the next 5 years.

- All companies that share/publish and/or re-use data consider that sharing/opening has helped them improve internal efficiency and/or improved their business opportunities.

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

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

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

3. WHAT CAN A COMPANY DO WITH ITS DATA?

As mentioned before, many organisations that begin to manage large volumes of data are not traditionally "data intensive" organisations and do not know their potential and/or do not know where to start, as they do not have specialised resources to exploit that data. This is the case of many companies in the so-called industry 4.0. In these cases, it is suggested to start with experimentation by opening the data to technology facilitators, who analyse how to value the data and, after this experience, decide on the next steps to take.

In the case of companies of a certain size that have a large amount of data (e.g. banks, telecommunications operators, electricity companies, etc.), and that also have specialised resources to exploit them, the data can be used not only for internal improvement, but also to make them more profitable by opting for business diversification beyond tadsadsahe core activities of the company through the exploitation and re-use of these data.

The European Commission's final report, which we have already cited[9], identifies two possible approaches to categorise initiatives related to data sharing/opening:

- **Top-down approach,** which involves identifying data that are useful for solving an internal problem or a global or local challenge. In these cases, the companies that own the data are proactively sought out so they can share/open the data. Most success cases of data sharing between companies fit this type.

- **Bottom-up approach,** where the opening/sharing of data is done with third parties, in order to explore the potential for their re-use. They are usually companies that feed on customer data to develop products and services that are not their core business (cited as success cases are GEO, Orange and the Mastercard foundation). These types of initiatives may or may not be monetised. These cases can be considered, following Chesbrough[10] , as experiences of Open Innovation, which is understood as the opening of the innovation process to the external knowledge of the organisation. Open innovation is a paradigm that assumes that companies can and should use external ideas, as well as internal ones. The boundaries between a company and its ecosystem will become increasingly invisible, and innovations will be easily transferred into and out of the organisation.

From the case study analysis of the European Commission's report on data sharing in enterprises, some patterns of best practices can also be drawn:

- Large companies without experience in opening of data and without the appropriate technological infrastructure for its exploitation rely on technology facilitators and/or platforms to start the process. The scope of the support can range from the outsourcing of the entire implementation of the initiative to the punctual outsourcing of resources and/or services (ORANGE success case in which it outsources the cloud to AMAZON to store/share massive data and to process it before the opening; success case of DAWEX, an intersectoral or cross-sectional platform aimed at all types of private companies).

- Companies without previous experience in opening of data, but who are aware of its value, can start by allowing startups, universities and research centres to use the data, so that they can explore the potentialities of data by applying Big Data techniques, artificial intelligence, etc. (ORANGE success case).

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

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

- From the point of view of data opening, there are mature companies that allow research centres and universities working on medium- to long-term challenges to access their data (such is the case of the non-profit foundation the Mastercard Center for Inclusive Growth, which does not charge research centres that work on solving existing problems in developing countries—looking for patterns of disease in the third world to identify the causes, etc.—for the re-use of data).

Why do companies share/open up their data?

- **To generate new lines of business and/or new companies based on the re-use of their data,** which enhance the exponential increase in the volume of data they manage, catapulted by the massive use of smart components—the Internet of Things (IoT)—and smartphones. The cases of BBVA, Telefónica, Orange and GEO are examples of this.

- **To solve problems that organisations cannot solve on their own**. An example is the case of the URBAN Institute which, using data from different companies (for example, the Mastercard Foundation), has been able to accurately locate the pockets of poverty in the different geographical areas of the United States.

- **To help other companies develop innovative and higher value-added solutions that they are unable to develop on their own.** This is true for both sectoral and cross-sectional data sharing/opening. An example of this are the publications that GEO sells to energy companies that use its smart meters, which contain the results of the analysis of the collected data. GEO has signed agreements with those electricity companies that allow it to analyse the data from their meters for certain purposes.

- **Companies that publish open data are often companies that have previous experience in data sharing.** The publishing of open data is a further step in making the most of the company's data and requires a major effort on the part of the company to ensure all the requirements for open data (technical openness, legality, etc.).

- **To comply with national regulations regarding transparency and accountability.** Normally, this aspect mainly affects public administrations, although there are more and more cases, such as ENEDIS, in which a strategy of opening of data is chosen in order to respond to national legal obligations.

There are various reasons that lead an organisation towards the opening/sharing of data, but they all share a common component: innovation and value generation from data. Sharing/opening of data can open the possibility for organisations to rethink their products and services more effectively to position themselves better. However, investments in technology and expertise are required to use the data **effectively**. The deployment of adequate data governance in the organisation will be a key aspect to achieve maximum effectiveness of the initiative.

4. TECHNICAL BASIS FOR OPENING/SHARING DATA IN THE BUSINESS ENVIRONMENT

The following are a series of technical recommendations and best practices for any organisation wishing to start a data sharing/opening initiative[11]. Therefore, this section describes each of the **stages and processes** that an organisation should follow regarding a data sharing and/or opening initiative:

1. Experimentation and understanding.
2. Defining a good open and shared data strategy.
3. Technical preparation.
4. Implementing the data life cycle: collecting, preparing, publishing and maintaining.
5. Monitoring the initiative, analysis of the level of success and continuous improvement.

The process of opening up and sharing data in the business environment must also be supported by an in-depth analysis of what we will call **supporting information** related to:

- Definition of the Business Model and Financing.
- Compliance with the Legal Framework.
- Technological Infrastructure.
- Development of internal skills (roles, skills and knowledge).

[11] Annexes IV and V can help to better understand and implement the technical complexity of data opening or sharing by companies. Annex IV contains some important concepts (linked data, metadata, etc.), while Annex V contains a list of open source software and applications that can be easily worked with in the field of Open Data and the knowledge of which can help companies starting with these practices.

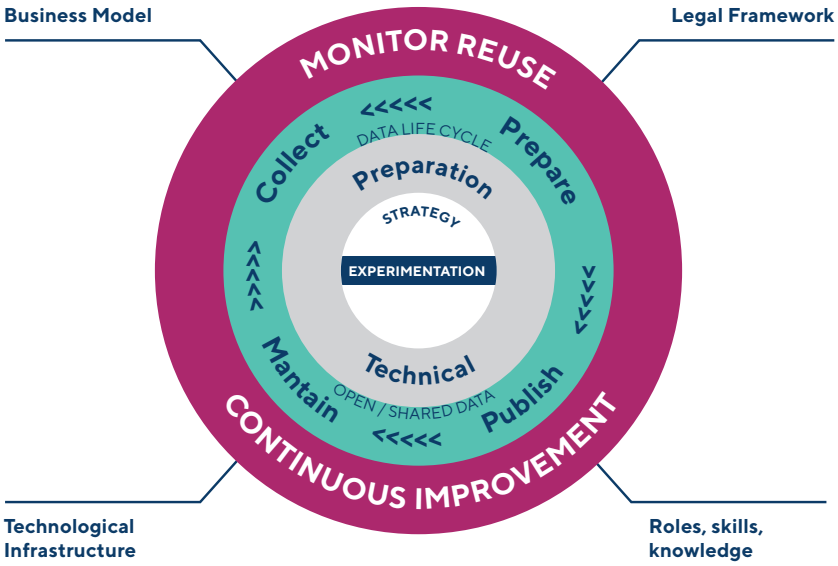


Illustration 1- Basics of the Opening/Sharing of Data in a Business Environment

Illustration 1 graphically represents the life cycle of open/shared data. Each of the stages and processes that the life cycle entails is detailed below.

4.1 Experimentation and understanding

Experimentation is the initial stage that an organisation faces when it is not aware of the potential of the information it handles. An organisation whose core business is focused on other areas of action (such as industrial companies) and that is far removed from the opening/sharing of data must start by knowing the potential of its data.

We can find organisations that are going through the following situations:

- **The organisation has internal expertise in data analytics.** In this situation, it may be enough to involve these personnel in the process of analysing existing data, to help management identify the potential that exists in the data managed and make decisions about the business model with which they should start.

  The organisation does not have internal knowledge in data analytics. In this situation, the most appropriate course of action is to rely on external personnel who have technical knowledge in data analysis and who can **advise on the value and potential of the data handled**. External support can come from various sources:

  - **Technology facilitators**: Technological centres, technical universities, consulting firms specialised in data analysis, specialised startups, etc., with which collaboration agreements can be made for mutual benefit. On the one hand, the organisation that transfers its data obtains an analysis of the existing potential from the data it manages; and on the other hand, the organisation that carries out the analysis obtains either remuneration in exchange for the analysis carried out, practical knowledge on the analysis of real data or testing algorithms for the analysis of developed data. Normally, the agreements are made with potential technological partners with whom a future relationship can be maintained. It should be borne in mind that collaboration agreements between both parties are not always associated with financial remuneration in this first stage of experimentation.

    In this situation, it is common for organisations to make available a reduced subset of the data they use as a sample to technology partners. This sample must be sufficiently representative to be able to draw conclusions from it. The technology partner will support the organisation in extracting the most representative sample if necessary.

    Collaboration agreements at this early stage will include confidentiality agreements that give guarantees to the data-providing organisation.

Company name:	**ArcelorMittal**
Web address (company or initiative):	https://spain.arcelormittal.com/
Description of the initiative:	▪ Optimise the ferrous performance of each type of scrap by referencing real data from the melting process of the steel mill and taking into account a set of factors that affect performance. ▪ Predict the final copper content in liquid steel at a point in time for each type of scrap mixed.
Shared/open data type:	▪ Data from the use of the scrap blocks generated in situ (up to 16), from the melting process of the steelworks and from the results: from 07/2017 to 04/2018, 4,000 consecutive samples.
Reasons for the opening/sharing of data:	▪ Exploring new analytical solutions for the interrelationship of the production process. ▪ Supporting open and collaborative innovation
Monetisation of business:	▪ No direct monetisation.

Table 1 – ArcelorMittal: Datatón Example

- ***Datathon Example*:** In this case, support involves making a selected set of data available to third parties in an open environment and for a limited period of time, so that participants can apply algorithms in order to try to get the most value out of the data. These activities are usually led by organisations oriented towards innovation in the ecosystem (local, regional, national or international) and with demonstrated access to Artificial Intelligence and Data Science communities, which can use and experiment with these data. In this type of initiatives, it is advisable to encourage the participation of organisations (research centres, universities, companies, etc.), as well as to publish the conditions related to the use of results obtained in the datathon. An interesting example of this type of initiative is the APPS FOR EUROPE project, which organised twenty local and global competitions in 10 European countries between 2013-2015 to identify viable businesses related to the creation of apps based on open data[12].

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

Company name:	**SUEZ**
Web address (company or initiative):	https://www.suez.es/es-es http://www.hackath2on.es
Description of the initiative:	▪ Over a weekend, 20 teams met to tackle, from a collaborative perspective, three water-related challenges, specifically: 1) excellence in customer experience; 2) promotion of the social value of water; and 3) new services for the administration. The initiative has been useful to the company, since carrying it out helped the company to design new products and services, and it encouraged the launch of other open data initiatives in the immediate future.
Shared/open data type:	▪ Transaction data (customers, suppliers) ▪ Internal operational data (quality, production, etc.)
Reasons for the opening/sharing of data:	▪ Diversifying the company's offer of products and/or services through better knowledge of the market and the client. ▪ Creating new business models. ▪ Social purpose, public common good.

Table 2 – SUEZ: Datathon Example

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

- **Marketplace/Industrial Platform**: In this case, the experience involves making available to third parties a sample of the data handled by the organisation, in an open environment such as a marketplace or an industrial platform, in order to identify the potential of the information and data, following the rules set by the sharing platform used. Agents who will explore/analyse data will be required to sign a data confidentiality agreement. Sometimes, the marketplace itself has services to explore the potential of the data. An example of this is the use of the Datary platform[13].

The experimentation exercise should help the organisation's management to make decisions as to whether the data opening/sharing initiative can be aligned with its business strategy, to what extent, through what business model and even what technology infrastructure to use. However, in this previous stage, conclusions reached will be incipient and will require further refined analysis, which will be carried out during the strategy definition stage.

Whether the organisation chooses to carry out a datathon or to make its data available to third parties in a marketplace, it is essential to seek prior legal advice to ensure that the data to be made available to third parties complies with the legislation in force.

Once the experimentation activity has been carried out, we suggest carrying out activities to promote **understanding** and raise awareness of the practice of open data in the organisation. In the first place, these activities should be aimed at the management level, as it is essential to know what the opening/sharing of data means in order to make appropriate decisions to define a strategy, policy and methodology aligned with the needs of each organisation.

4.2. Defining a good open or shared data strategy

The aim of this process is to define, establish and review the organisation's open/shared data policy and strategy based on the organisation's actual initial situation. Such a strategy influences both the sharing/opening and re-use of third-party data and is an organisational transformation process that should be part of a broader digitisation strategy. The following are some practices that can be considered at this stage:

- **Knowing the initial situation of the organisation regarding open/shared data**.
    - Identify the role of each unit/department/division of the organisation: Which units collect data, use data, produce data? What type of data do they collect/use/produce? In what format?
    - Identify the representatives or key people from different departments of the organisation who allow to perform a diagnosis of the starting situation regarding the data.
    - Are the data centralised or distributed? Is data management centralised or decentralised? If data is generated in multiple units/departments/divisions, and everyone has their own way of working and their own policies, the change may be much more complex.
    - Is there open/shared data? What data? Is there a data publication policy? What data will be shared and with whom? Which will be closed and available only to the organisation?
    - In this stage of institutional knowledge of the data, it may also be convenient to visualise the current data flow through a graph.

Depending on the scope and goals of the data sharing/opening initiative, the baseline diagnosis can be done in a more or less formal way, using the Open Data Institute (ODI) open data maturity model as a reference[14]. This model makes it possible to evaluate the maturity of an organisation regarding the openness of data in terms of operational and strategic activities. The model provides guidance on possible areas for improvement and helps organisations to compare themselves in order to highlight their respective strengths and weaknesses and adopt best practices and processes.

At this stage, it is advisable to look for examples of benefits that can be derived from publishing certain data in open format and to avoid discussions about whether or not to share/open up data. It is equally important that all members of the organisation are involved in identifying opportunities for opening up/sharing data. Different people in the organisation may see value in some data that others do not, because they observe a specific context where the data add value.

- **Knowing the target situation of the organisation regarding open/shared data**.
    - Define strategic objectives: What should be achieved? In what time frame? Where can the organisation be in 2-5 years?
    - What is the B2B or B2G relationship model that best suits you: data monetisation, Marketplace, industrial platform, etc.?

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

- Should there be a technical team specialised in data analysis? What internal knowledge does the organisation have? Should there be agreements with technology partners (technology centres, universities, specialised consultancy firms, etc.) who can provide their experience?

- Should the organisation have a sales team for this line of business? Should the organisation establish collaboration agreements with third parties whose mission is to act as intermediaries between the organisation that provides the data and the end user of the data?

- Is legal support necessary to help license shared/published data?

- Define measures associated with the objectives (in relation to scope, time, number of deliveries, etc.) that allow to monitor the objectives.

- Visualise the flow of destination data.

- **Defining and implementing the appropriate strategy and action plan to achieve the objectives**.

  - Develop an action plan to achieve the objectives and manage the plan as a project.

  - Ensure the involvement of all those participating in the initiative, whether internal or external, where necessary.

  - Define the appropriate data governance model for the organisation:

    - How does the organisation manage its data (both its own and third-party data)?
    - Is there sensitive data? Is there confidential information?
    - Is it necessary to anonymise and aggregate data before publication?
    - Is the data in formats that third parties can easily use?
    - Have responsibilities been assigned within the organisation for the execution of the various processes related to data sharing/opening, including data pricing and payment methods?

See section Data Management and Governance.

  - What type of licences will be used? *(see section regarding licences)*.

  - Define the financing model *(see section Financing Models).*

  - Define and establish processes for extracting, transforming and publishing datasets, and progressively identify and implement the technology needed to prepare the data and to share/publish them as high quality datasets *(see technical preparation and data preparation)*.

  - Define and establish processes to monitor the re-use of shared/open data and improve established processes for sharing.

  - Remember that an organisation's evolution towards an adequate data management structure may require the deployment of a change management process.

- **Defining and maintaining the organisation's data inventory**. When it comes to the sharing/opening of data, a mature organisation treats data as organisational assets. The catalogue must contain internal datasets and the re-used datasets belonging to other organisations (in the case it re-uses data from third parties).

    - Possible attributes on published/shared datasets:

        - Risks associated with the publication of the dataset (does it or does it not contain personal data? Can such data be considered sensitive?) It must ensure compliance with applicable regulations. The most relevant element is compliance with the GDPR. Publishing of personal data will require the consent of the data subject. In addition, it will be necessary to notify the identity of those who receive personal information at the time of the request for consent, as well as the other information to be provided to the data subject whose data would be processed and published. In an Open Data strategy, the risk of not complying with these requirements makes it necessary to anonymise the information to be published. With the anonymisation of the information (making it so that the owner of the data cannot be identified), the data can be published, since it will not be of a personal nature.

        - It should be borne in mind that the company may have established a trade secret policy. The Trade Secrets Directive [Directiva de Secreto Empresarial] and the new Trade Secrets Act [Ley de Secretos Empresariales] establish the requirements under which certain information that gives a competitive advantage to a company can be considered a Trade Secret. Therefore, companies must take into account whether the information they are going to publish can be a violation of the Trade Secret regulations, whether it is a Trade Secret belonging to the company itself or if the secret is licensed by a third party.

        - Value of the datasets (determined by the organisation's process of valuation).

        - Details of the financial investment associated with the publication of the dataset.

        - Planning the publication of the dataset.

        - Key metrics collected during the monitoring and continuous improvement process; etc.

        - Possible attributes on re-used datasets: rights to use the dataset; costs associated with re-use (for business data); risks associated with a price increase (business data); risks that the dataset is not maintained or not available; how the dataset is re-used by the organisation; etc.

- **Defining the open/shared data policy and disclosing it**.
  Defining a solid policy at the organisational level increases the transparency of the organisation and ensures the good use of your data. Disclosing the open data policy helps to promote awareness of the benefits of open data, and this results in an increase in open/shared data and data quality.
  - What is part of the policy?
  - What data will be published? What will be shared? Has it been previously anonymised?
  - Under what conditions? When? How often?
  - What is the expected impact? And the benefits?
  - In any case, before the organisation defines a policy and a strategy for the sharing/opening of data, it is important to identify the benefits it can bring and the purpose of sharing/publishing data; to give examples of both internal and external benefits of sharing/opening, etc. That is, to develop the culture of openness/sharing of data in the organisation.

4.3. Technical preparation[15]

The opening/sharing of data in a private company can have a significant impact in terms of technical elements and data governance. The organisation has to be equipped to efficiently and sustainably support the data life cycle, from data collection to publishing and maintenance.

The aim of the preparation stage, prior to the implementation of the open/shared data life cycle, is to help the organisation make the right decisions regarding data management and associated organisational changes, as well as regarding the technological infrastructure to be put in place to extract and process the data prior to opening/sharing, all according to the organisation's goals and strategy. The guide also addresses some aspects of data governance.

It then details the aspects to be considered at the organisational level prior to the implementation of the life cycle of data opening/sharing:

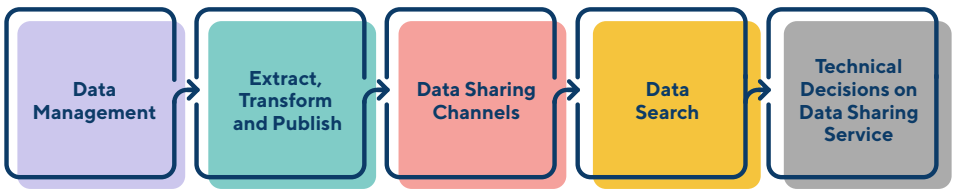


Illustration 2- Preparation prior to the start of the life cycle of open/shared data

4.3.1. Data Management

Data management must cover all the organisation's datasets that are internally generated, whether they are to be shared or not. Organisations may decide to share only certain types of data. In any case, the company should have a management policy that takes this possibility into account. If applicable, third party

[15] The contents of this section were inspired by section "3. Technical preparation and implementation" of https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf (illustrations 3, 4, 5 and 6 are from that document).

datasets will also be considered. An organisation oriented towards the sharing/opening of data will treat third party shared data as if it were its own asset.

Possible data management structures are detailed below. These structures are theoretical and must be adapted to the caseload of each organisation.

- **Decentralised Data Management (DDM):** The management of data is distributed across units or people who are responsible for managing their own data. This situation is an example of short-term data governance and is only recommended for organisations where data sharing is an infrequent activity and which handle very small amounts of data (e.g. a maximum of 10 datasets that are updated a maximum of 4 times per year):

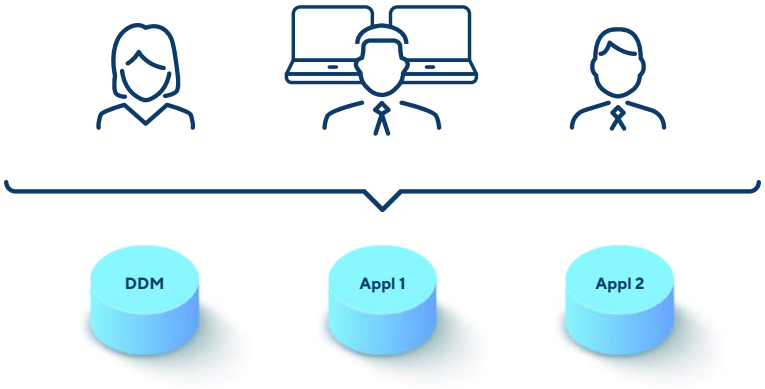


Illustration 3- Decentralised Structure

This type of management is often seen in organisations that are not traditionally data-intensive and who begin to manage large volumes of data (e.g. Industry 4.0). These organisations will need to improve the management and governance of their data as a step prior to its sharing/opening.

- **Centralised Data Management (CDM):** In this case, data management is still distributed across units, but there is a clear allocation of centralised responsibility. There is more coordination in this scenario and, therefore, there is more consistency between the different units that manage data. This is a better approach for organisations handling large volumes of data. It is a first step towards fully centralised data management.

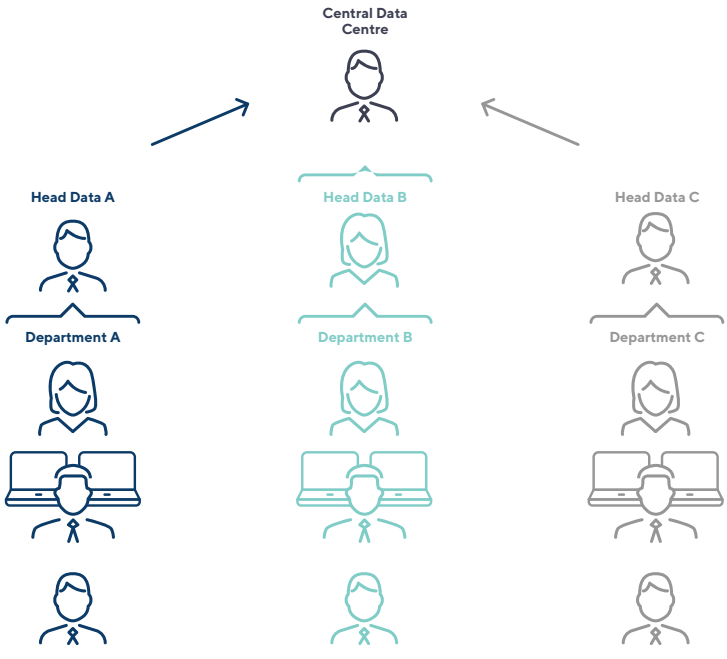


Illustration 4- Centralised Structure

- **Fully Centralised Data Management (FCDM):** In this case, standardisation is defined at the highest level and consistency is guaranteed throughout the organisation. This scenario minimises execution errors during the data life cycle, reducing long-term costs. Furthermore, this governance structure is more sustainable, even when the structure of the organisation changes over time. In large corporations with a great wealth and diversity of data, centralisation may be limited to certain similar data silos, in which case we would not refer to a FCDM, but a CDM.

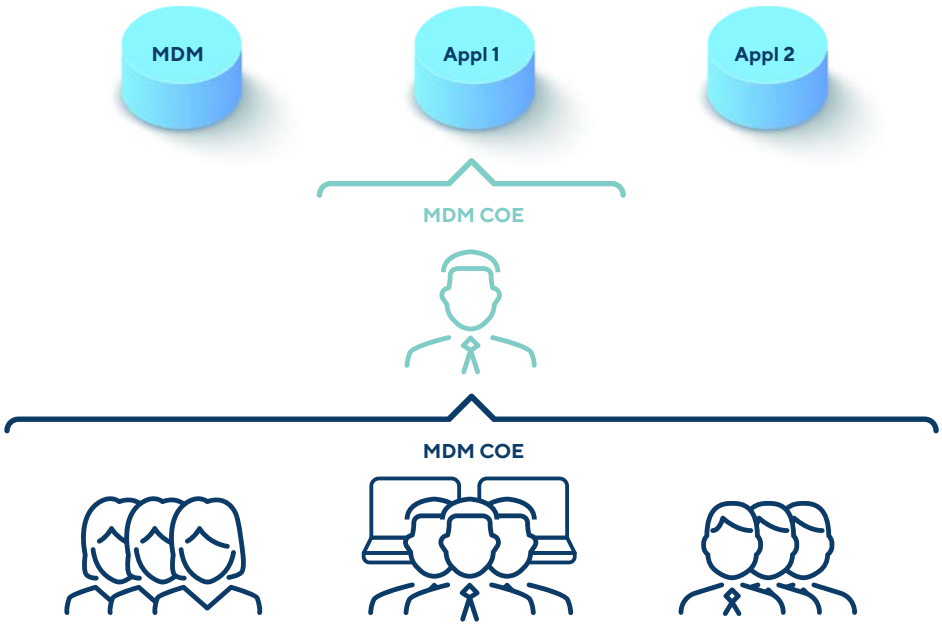


Illustration 5- Fully Centralised Structure

4.3.2. Extract, Transform and Share Data (ETSD)

The ETSD process refers to the technical specifications of how data flow and data transformation is to be carried out in the organisation so that this data is ready to be shared or opened up (published as open data). This is regardless of whether the data is finally shared and/or published as open data or used only internally. This process is defined at the organisational level and is complemented with the preparation of each dataset, as it is addressed in section 4.4.

The ETSD process, along with the implemented technology, is the central process that will be executed repeatedly to obtain datasets (either for internal use or to be shared/opened up).

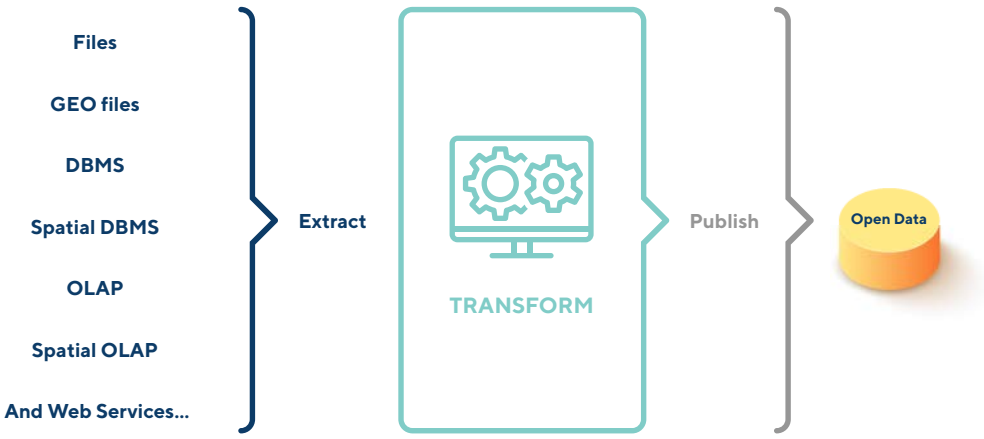


Illustration 6- ETSD Process

The three steps of the ETSD process are:

1. **Extract**: Data is extracted from any type of source, including new data generated from existing data and data from internal or external sources (documents, databases and/or applications or systems). The extraction process is conditioned by the organisation's data management. Depending on the organisation's strategic objectives or goals, there are many options open regarding ETSD processes. (see possible scenarios in Annex 1).

2. **Transform**: It consists of transforming the starting data into data that is ready to be shared (internally or externally), or published in the open or as Linked Data. The organisation must define a guide for data preparation (e.g. by dataset types, whether or not it contains sensitive data, geo-localised data, real-time data) and a policy that refers to these guides. (See possible scenarios in Annex 1).

3. **Share/Publish (optional – not necessary if datasets are for internal use only)**: Identify channels for sharing or publishing datasets. In addition, the following steps have to be completed. These are part of the process of the Life Cycle of Open/Shared Data (See section 4.4. Implementing the life cycle of open/shared data):
   - Choose the licence model.
   - Define the metadata associated with the datasets.
   - Open data: Publish the dataset (preferably automatically).
   - Offer the possibility of converting data for certain platforms and develop APIs (if necessary).
   - Establish the life cycle, indicating the details of the organisation's contact person.
   - Ensure periodic sharing/publication of datasets

4.3.3 Data Sharing Channels

Depending on the policy and strategy defined by the organisation with respect to data sharing/opening, there are different technical mechanisms for distributing information. Data sharing channels are the different technical mechanisms for distributing the organisation's data to third parties[16]. The organisation should identify those data sharing channels that best meet its needs:

Illustration 7 – Data Sharing Channels

[16] https://eur-lex.europa.eu/legal-content/EN/TXT/

Web Downloads

This is the easiest way to share data and it is the most widely used channel in those cases where the organisation chooses to share small amounts of datasets. The data can be published on the organisation's website, in a private section with controlled access to the data links.

The software solution used could be any Content Management System (CMS), such as Drupal, Wordpress, Django or others, possibly combined with an internal database.

API

An API (Application Programming Interface) is a more advanced and technical way of connecting different systems or applications to each other. In the context of data sharing, an API is a software that provides a system with direct access to data from another system, giving access to the use of data catalogues and their functionalities. Without interference from end-user interfaces, such as web portals or web pages, a third-party application can load the data via a request protocol. It is particularly useful if the data must be up to date, if third parties must have direct access and if the application using the data needs direct access to the database without any interference.

The European Commission encourages[17] companies across Europe to consider the use of open, standardised and well-documented APIs for data sharing. There are also multiple technical solutions that can be used, such as CKAN, Socrata, Open Data Soft and DKAN (See Annex V).

[17] European Commission (2017): "Building a European Data Economy", Brussels, COM (2017) 9 final

Data Portal

A more advanced way of publishing datasets is through a data portal, as is usual for the public data from governmental bodies. In the business environment, this solution is particularly useful if the organisation has large amounts of datasets that they want to publish, and which must also be updated at certain periods. Portals are web-based solutions that normally have their own domain name.

Another approach would be to publish in third party data portals. This is a low cost solution and is very applicable in some situations. Publishing on existing data portals facilitates access to the data, as deployment increases and, therefore, data may reach more interested entities. In this case, two compatible options can be considered: 1) the data is loaded directly into the portal and downloaded directly; or 2) the portal automatically redirects access to the organisation's website containing the data, and the download is made from the organisation's own website. In the case where the data is entered in open portals, it will probably be necessary to implement additional configurations, such as metadata formats, use of terminology specific to the disciplinary scope of the data or other technical standards that promote interoperability and re-use.

Marketplace

These are trusted intermediaries that bring together data providers and users in order to exchange data on a secure online platform. These companies derive revenue from the data transactions that take place on the platform. The basic functions of a Marketplace are:

- Enable matches between the potential data provider and the data consumer. This may include specific configurations in which the potential provider and potential consumer may remain anonymous in the first part of the data transfer preparation setup, as the intention to provide or purchase may reveal trade information that is already secret (future business strategies).
- Actual transfer of data and agreed remuneration. The Marketplace creates a space of trust by ensuring that the object of the negotiation will not be altered during the course of negotiations.
- Certify that the transaction has actually happened.
- Additional services, such as model contractual clauses or anonymisation services, may also be provided if personal or confidential data is exchanged.

The role of this type of intermediary normally ends once the data has been transferred, unless the interacting group wants to continue collaborating and sharing data and sees the possibility of growing by adding more participants.

Most Marketplaces are centralised, although technologies such as DLTs (Distributed Ledger Technologies) and the generation of decentralised algorithms (e.g. OceanProtocol[18] are enabling the creation of decentralised Marketplaces.

[18] https://oceanprotocol.com

Industrial Platform

Industrial data platforms are created to foster a collaborative approach to data exchange between companies. More specifically, a restricted group of users voluntarily join such platforms to mutually benefit from shared data. Data is generally exchanged free of charge, but fees for additional services may be considered in a closed, exclusive and secure environment. The goal of these platforms is to improve the performance and productivity of platform members. Industrial platforms are typically used by large organisations.

Technical Facilitator

Although they may resemble industrial data platforms or marketplaces, these organisations can be described as technical enablers for the exchange of data between companies. These are organisations that develop their own web and cloud-based solutions to allow data exchange between a group of users or business partners. These companies act as a third party in the data exchange process by providing the enabling technical solution. Data users who participate in data sharing through these solutions can exchange data within a particular community in a flexible manner. In addition, they make use of a tool that already exists and that can be customised, instead of using something new.

Unlike industrial data platforms or data markets, their revenue comes from the configuration, implementation and maintenance of their solutions.

This type of intermediary offers additional functionalities that allow the data provider to control how the data is used. It includes ways of tracking and tracing the use of data, for example, by recording all actions of access and data processing,

using blockchain technology, or developing digital watermarks. The intermediary can also develop requests for self-regulation within the user community of the data space or platform, including, for example, penalties for data users who violate established data transfer agreements.

Sharing/opening of real-time data streams

All of the sharing channels mentioned above address the sharing/opening of static datasets. In addition, most Marketplace and Industrial Platforms also offer companies their own solutions and services to share large volumes of data (e.g. data streams from Industry 4.0 sensors) that are updated in real time. The technology facilitator provides the companies or the end customer with the platforms, whether open or proprietary, that enable such solutions and services. Some examples of platforms for real-time data stream analysis: Apache Flink (https://flink.apache.org/), Spart streaming (https://spark.apache.org/), IBM Streams, Software AG's Streaming Analytics, Azure Stream Analytics, Amazon Kinesis Data Streams, etc.

'Data sovereignty' is when an organisation's data is stored outside its country and is subject to the law of the country in which it resides. The main concern regarding data sovereignty is maintaining the privacy regulations of each country and protecting customers from subpoenas on their data. **EU rules on personal data protection set limits for organisations that transmit personal data created in Europe when interacting with countries whose data protection legislation is considered inadequate by the EU**[19]. Some countries require confidential data to be stored on government-supervised servers or in authorised, third-party data centres. Monitoring and complying with data sovereignty requirements raises a difficult and persistent issue for global companies. Major cloud service providers (CSPs), such as AWS, Azure, Google, Oracle and IBM, have national data centres, which is the first requirement to comply with data sovereignty regulations. These providers offer authentication, encryption, and security services, as well as a set of management tools designed to help customers comply with local regulations. However, to implement the architecture, processes, and security policies needed to meet these standards, global organisations must understand country-specific data regulations and work with cloud service providers and the management tools they offer.

International Data SPACE (IDS)[20]. It is an association made up of more than 90 companies and institutions from over 18 countries. Its initial purpose was to develop the concept of "data sovereignty". It has defined a framework for the creation of a virtual data space that allows the exchange of data between organisations, creating value around the data in a trustworthy environment. This virtual data space uses common standards and governance models to facilitate secure data exchange and to easily link data in business ecosystems within a framework of mutual trust. It provides a foundation to create and use smart services and innovative business processes, while ensuring the digital sovereignty of data owners.

[19] https://eugdpr.org
[20] https://www.internationaldataspaces.org

4.3.4. Data Search

The level of complexity of the search mechanism to be used depends on the volume of data to be shared and the data sharing channel. For example, if data is shared through the organisation's website, the basic search mechanisms provided by content management systems (CMS) will be perfectly valid. Other data sharing channels (e.g. open data portal, platforms, etc.) have their own search engines to access their datasets or third-party datasets, and these operate by

harvesting or metadata extraction techniques (see Annex IV for additional information).

For advanced searches of datasets that meet the desired technical requirements of linked data technologies (Linked Data) or better, Linked Open Data (LOD), there is the option of using SPARQL for queries. SPARQL is an advanced search language that should not be used if the datasets do not comply with Linked Data requirements (datasets with metadata and data described according to a standard or recognised ontology) or are encoded in standard languages such as RDF (Research Description Framework).

There are also specific dataset and universal search engines such as Google Data Search, which was launched in September 2018, initially for searching open research data, but which also aggregates other types of Open Data on the web. There are also large global data aggregators such as Fireware.

4.3.5 Other technical aspects

Other technical decisions related to the provision and maintenance of the opening/sharing data service that must be considered are:

- **Domain selection:** The organisation may choose to use its own website to make the data available or use a separate website with its own domain name (e.g. data.institution.es).
- **Host selection:** The organisation must determine if the data will be stored and made available on its own servers or third-party servers will be used.
- **Selecting functionalities**: For example, the organisation should identify which databases will be shared, if a payment module will be required, how much space will be needed on the server, the speed of access required, etc.
- **Management of the sharing channel**: There must be a person in charge of managing the sharing channel. In addition, the organisation must define the degree of availability, security and monitoring, among other aspects, of the sharing channel.
- **Service maintenance**: A service maintenance manager should also be appointed to ensure the availability, functionality and performance of the data sharing service.

4.4. Implementing the data life cycle: collecting, preparing, publishing and maintaining

The implementation of the life cycle of open/shared data[21] is a process that describes the steps to follow to produce the datasets (each dataset) that will be shared/published. It is a process that must be included in the organisation's policy and involves the following sub-processes:

Illustration 8 – Life Cycle of the Opening/Sharing of Data

4.4.1 Collecting data

The purpose is to identify which datasets are going to be shared/published, prioritise them, register them in the organisation's dataset catalogue to share/publish (if it exists), categorise them and determine how to promote/publish the datasets. It is based on data already available in the organisation, either in a centralised or distributed way. Data collection can be specific (seeking a quick pay-off) or it can have a more detailed data management approach (based on preparation processes).

For organisations without experience in data sharing/opening, the main activities to be carried out are:

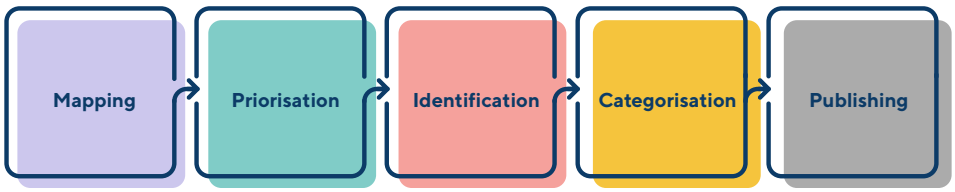


Illustration 9 - Collection Activities

1. **Mapping the data available in the organisation for the purpose or strategic objective in question**: Identify data that can be shared with other organisations and data to be published as open data. For this task, it is important to seek help from those responsible for the data.

2. **Prioritisation of datasets**: Criteria for prioritising datasets:

    - Are they suitable for the purpose in question?

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

- Can they be shared/opened up as they are (legally, politically and organisationally)?
- Are they of the required quality or do they need to go through a treatment process before the sharing/opening?
- Is it necessary to "clean" them, anonymise them, and improve their quality and formatting?
- Is it a dataset with a high value theme?

It is recommended to start first with the datasets that bring us benefits with the minimum effort (quick wins). In other words: those datasets for which the answers to all the above questions are positive.

3. **Establishing the process for the identification and prioritisation of datasets**: Define and establish the process for identifying, modifying, creating and eliminating the organisation's datasets.

4. **Categorising datasets:** Consists of grouping of datasets by themes or categories.

   The categories must be action oriented. When classifying/assigning datasets to categories, the dataset must be analysed from the perspective of the re-users. Categories allow datasets to be structured logically, but excessive categorisation should be avoided.

   There are no examples of categorisation of datasets for private organisations, but we can use the existing ones for governmental organisations, such as the categories of the European[22] open data portal and the categories of the Spanish data portal (at state level), as a reference[23]. The existing dataset categories and the number of datasets per category help the organisation to prioritise the datasets.

5. **Publishing available data categories and datasets by category**: The classification of datasets into categories can be done automatically through the metadata associated with the datasets. For example, in the European data portal, the categorisation of datasets[24] is done automatically through the extraction of metadata (metadata harvesting) and based on the DCAT standard.

4.4.2 Preparing data

The goal is to prepare the data (each dataset) to be openly shared or published. Identified baseline data, most of the time, must be processed before publication.

The processing or preparation of the data covers the following aspects (not necessarily sequentially):

- Data quality.
- Data preparation.
- Adding metadata.
- Legal preparation.
- Management of sensitive data.

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

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

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

The legal feasibility of the opening/sharing of data is key to launching the dataset preparation process. That is to say, if the data does not comply with the legislation in force, it cannot be shared/opened up, so it would be pointless to start the preparation process.

1. **Data quality**: Data quality covers aspects such as:

- **Content**: The usefulness of the data is conditioned by its quality. Main aspects that determine the quality of the data and what can be done to improve it:

    - Is the data complete? Define a header with a unique description in the metadata. Even if the data changes, do not change the structure. It must have a version number and, if the content changes, a new version must be assigned to make it easier to track the changes; it also must describe the content of the information: what the data is, where it comes from, for what purpose it is to be shared or published, and what is its status (e.g. draft/validated/final).

    - Is the data clean? Check empty fields; do default values and fictitious data make sense? Check erroneous values, double entries, sensitive private information; are there any violations of legal requirements

    - Is the data accurate? Is the data accurate enough for the specific purpose (data sharing) or the potential one (open data)? Does accuracy affect its reliability? Was the choice of time intervals described? Does the data need to go through an aggregation or disaggregation process?

Data quality can be measured by its level of compliance with the FAIR principles (Findable, Accessible, Interoperable, Reusable) by including appropriate metadata and persistent identifiers for identification and re-use.

- **Promptness**: Address the notion of time as it applies to data. Data changes over time, so it is important to establish a process to update it. This aspect is closely related to data maintenance.

- **Consistency**: It is essential to keep consistency in the presentation of data. It is important to use standards to share/publish datasets and to be consistent in sharing/publishing data of the same quality. For example: If datasets are published annually and different attributes are used to define the data each year, the handling/effort/cost required by the re-user will be greater and often beyond the capacity of many organisation.

2. **Data preparation for technical sharing/opening**. Two conditions required to share an organisation's data with other people or organisations are the interoperability of the data and the information that gives us access to the data itself, especially in the case of open data. However, these conditions are not enough. The first depends on the format(s) in which the dataset is shared/published, which determines the ease with which third parties or individuals will be able to make use of the data with minimum effort and cost. The second, which is essential mainly for the opening of data, has to do with the ease with which third parties understand the content of datasets and is linked to the use of metadata.

3. **Add persistent metadata and identifiers.**
   In the case of data shared on platforms, and especially in the case of open data,

it is important to ensure access to the data. This means that the data should be easily found by the interested people or organisations. This concept is called discoverability. However, more and more people are now talking about FAIR data. Although "FAIRness" is an emerging requirement applicable to research data, it is also applicable to any data set. It is not for nothing that data from the business sector can be used for research or even arise from research originating in the private sphere. In any case, for datasets to be FAIR, it is essential that they have associated metadata and persistent identifiers. Metadata is data about data, structured information that describes, explains, locates or, in any case, facilitates the recovery, use or management of data. It allows to understand the nature and structure of the dataset, know its origin, and know the terms with which it can be used. On the other hand, persistent identifiers allow a dataset to always have the same URI and therefore guarantee access to it.

4. **Legal preparation for sharing/opening.**
   As we have already said—and will explain in more detail in the following section—, the data must respect the legislation in force and each shared or open published dataset must have an associated licence, which must be respected by both the publisher and the re-users of the data. These licences must comply with all current legislation. The data publisher is responsible for ensuring this compliance (see section 5.3.3.).

5. **Sensitive Data Management Process**
   Datasets that include sensitive data may require special preparation. This special preparation may involve:

   1. Conducting risk assessments and impact analyses before sharing datasets with sensitive information.

   2. Defining risk mitigation mechanisms:

      - Defining and deploying data anonymisation policies and processes.
      - Identifying sensitive information, e.g. personal health data, racial origin, etc.
      - Preparing subsets of data to eliminate sensitive fields.
      - Data aggregation.
      - Seeking the consent of those whose personal data may be disclosed. In this case, consent should be presented as a positive action by the data subject, clearly providing information related to the recipients and purposes. In addition, consent is revocable, which may result in the subject revoking his or her consent—this means that the data which had previously been published must be deleted.
      - Properly defining the data environment: related datasets, people who have access to the datasets, people who might make malicious use of the dataset, etc.

   3. The organisation is responsible for complying with the current legislation applicable depending on the data handled in each case (see annex on legislation.

   4. The organisation is also responsible for creating an appropriate data management culture. Personnel must be trained in the new policies, processes, tools and technologies deployed and related to data management.

   5. It is necessary to define legal procedures for contingencies in case of

incidents related to the sharing/opening of sensitive data.

An organisation initiating the data sharing/opening process does not normally have a legal inventory of the ownership of its data (this inventory would identify which data is sensitive and which is not). However, identifying the organisation's sensitive data is a necessary condition to apply this process.

6. Checking the preparation of the data:

- Check whether the datasets are legally open.
- *Check the quality of the dataset.*
- *Check the updating frequency of the datasets.*
- Check that the datasets comply with established standards.
- Check that the datasets have the appropriate metadata and are persistently identified.
- Check whether the metadata is described as linked data.
- Check whether the datasets are technically open.
- Check the licences of each dataset and the information regarding the origin of these licences.

Some of the most popular methods/models to determine the quality of the dataset (the information associated with the dataset that facilitates its re-use and accessibility):

- Tim Berners-Lee's 5-stars[25] are a simple way of checking the technical quality of datasets by virtue of the publication format:

☆	Make the dataset available on the web (in any format) under an open licence.
☆☆	Make it available as structured data (e.g. Excel instead of a table image).
☆☆☆	Do not use proprietary formats (e.g. CSV instead of Excel.
☆☆☆☆	Use URIs to identify elements, so that people can point to elements through URIs.
☆☆☆☆☆	Link data to other data to provide context (linked data).

Table 3 – Tim Bernes-Lee's 5 star.

- MELODA[26] is a simple evaluation method that allows to evaluate the quality of datasets (it is applied to each dataset) based on the associated metadata.

- FAIR Data maturity model[27] is a developing model born in the Research Data Alliance (RDA) community. It tries to create metrics and indicators to measure the findability, accessibility, interoperability and re-usability (FAIR) of the datasets; it initially applies to the research data, but as we have already mentioned, the FAIR principles are applicable to any dataset. In fact, there is a whole web dimension linked to the data that has begun to be called the Internet of FAIR Data and Services (IFDS).

These methods/models, as well as others that are applied in specific contexts, allow the organisation that shares/publishes data to have an indicative idea of

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

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

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

the value of its data. However, what will really give value to the data will be the success when it is re-used.

4.4.3 Publishing data

Data publishing is done on established sharing/publishing channels. Some publication channels have been mentioned in previous sections, so it is worth mentioning here that the channel to be used depends on the number of datasets to be shared. It is recommended to start with the simplest solutions for the expected number of datasets.

4.4.4 Maintaining data, persistent metadata

Both data and metadata change over time. We recommend establishing a process to keep shared/open data and metadata up to date. Aspects to take into account for this process:

When updating a dataset, indicate the date of updating in the metadata.

Depending on the publication format, the URIs and URLs associated with the datasets have to be reviewed periodically to check that they are still in force. The WWW is dynamic and URIs and URLs may change.

Provide a contact, either through metadata or by incorporating a contact channel for re-users. Your opinion is essential to improve the quality of the shared/open data.

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

4.5 Monitoring the initiative, analysis of the level of success and continuous improvement

When datasets are shared or published following the established life cycle, it is important to follow the data sharing/opening initiative in its entirety, from the review of policy and strategy to the definition and implementation of the different processes. Auditing the data's life is essential, specially determining the future traceability of the data (from its origin) and its re-use in different scenarios, both commercial and non-commercial. In any initiative, there are two key aspects to review:

- **Contrasting the commitment of the re-users,** since their commitment is essential for the success of any open data sharing initiative. Tim Davies[28] also establishes a 5-star model (as Tim Berners-Lee did to measure data quality based on its coding format), in this case, to analyse the commitment of the re-users:
  - **1 star**: Be demand driven.
  - **2 stars**: Contextualise the data.
  - **3 stars**: Establish conversation channels for the data.
  - **4 stars**: Build capacity, skills and networks.
  - **5 stars**: Collaborate on data as a common resource.
- **Measuring the success of the initiative and continuous improvement**: Some basic measures that can be used are:
  - **Number of downloads and visualisations of the datasets**. However, this indicator does not give information about the usefulness of the dataset.

- **Number of accesses**. It indicates if the system can manage the request, if the data has been downloaded, if the system can respond to the requests, if there are consequences for other systems when there are requests, etc.

- **Qualitative metrics of dataset users**: Is the dataset useful for the defined purpose? Is the dataset of quality (clean, clearly described, adequate granularity level)?

The data provider can influence the latter indicator by adapting the data preparation process better to the needs of the re-users. There are free software tools for this type of analysis (e.g. **PIWIK**[29]). Some open data portals, platforms, Marketplace, etc. also offer software solutions with this type of metrics.

[29] https://piwik.pro/

5. RELEVANT ASPECTS OF THE OPENING/SHARING OF DATA IN THE BUSINESS ENVIRONMENT

5.1. Relationship Models

5.1.1. B2B (*Business to Business*)

Business models related to data exchange can differ considerably. They depend, to a large extent, on the type of data shared and the strategic commercial interest of both parties. The model can range from an open data approach to exclusive data associations where data is shared with a single party:

- **Open/Shared Data Approach**: Data is made available to an open range of re-users with little or no data re-use restrictions and with little or no remuneration. This data sharing approach takes place when the data provider has a strong interest in re-using its data, such as those service providers who wish to use an ecosystem of third-party application developers to reach end customers.

Name of the initiative:	**BBVA VALORA**
Web address (company or initiative):	https://www.bbva.es/general/banca-online/bbva-valora/index.jsp
Description of the initiative	BBVA Valora is an online tool available through the BBVA app, and it is especially useful for two main types of activities. On the one hand, it helps to discover the estimated value of the desired home, either for purchase or rent, based on the large number of available analysis functions that integrate data, such as the evolution of land prices, the provision of public and social facilities in the neighbourhood where the home is located, the existing housing stock or the average price of homes for sale or rent. On the other hand, and for those people who are BBVA customers, the tool allows all the expenses derived from the home to be grouped together, as well as all the information related to it, in order to help them make the best decisions when it comes to renovating the home or monitoring invoices.
Reasons for the opening/sharing of data	▪ Diversifying the company's offer of products and/or services through better knowledge of the client ▪ Social purpose, public good
Monetisation of business	No direct monetisation. The data is offered free of charge to all interested people (in the case of the valuation of the desired home) and to the bank's customers (grouping of expenses derived from the home).
Shared/open data type	▪ Location/positioning data. ▪ Transaction data (customers)

Table 4 – BBVA VALORA: Open/Shared Data Example

Company name:	**Primafrio**
Web address (company or initiative):	https://primafrio.com/
Description of the initiative:	Real-time data is shared with the customer. Parameters include truck number plate, position, driver, etc. as well as the company's billing and accounting information..
Shared/open data type:	▪ Real-time truck location/positioning; driver's name; etc. ▪ Transactions (customers, suppliers).
Reasons for the opening/sharing of data:	▪ Improving process efficiency ▪ Solving an internal management problem ▪ Supporting open and collaborative innovation ▪ Pioneering these processes (strategic positioning)
Sectors that re-use data:	▪ Distribution/Logistics ▪ ICT services, including app/software developers ▪ Automotive and transport
Monetisation of business:	No direct monetisation. Data is shared free of charge and is available only to PRIMAFRIO customers..

Table 5 – PRIMAFRIO: Open/Shared Data Example

- **Monetisation of Data**: A one-sided approach in which companies derive additional revenue from data they share with other companies. Data monetisation can be achieved through the provision of new services based on shared data.

Name of the initiative:	**EUSKALTEL**
Description of the initiative:	EUSKALTEL offers its customers (from small shops to companies and public administrations) reports based on completely anonymised and aggregated data. The aim is for Euskaltel's client companies to be able to advance their digital transformation.
Reasons for the opening/ sharing of data	▪ Diversifying the company's offer of products and/or services through better knowledge of the client ▪ Creating new business models
Sectors that re-use data:	▪ Transport ▪ Public sector ▪ Retail/commerce
Shared/open data type	▪ Location/positioning data.
Monetisation of business:	The price of the service depends on the value given to the client. Criteria to determine the value of the data: ▪ Level of data enrichment. ▪ Geographical area. ▪ Time period of the study.

Table 6 – EUSKALTEL: Monetisation Example

- ***Marketplace***: Trusted intermediaries that bring together data providers and users in order to exchange data on a secure online platform. These companies derive revenue from the data transactions that take place on the platform. This approach may be appropriate when:
  - There are risks associated with the unlawful use of data.
  - The data provider needs guarantees regarding the trustfulness of the re-users.
  - The data provider has technical mechanisms to prevent or identify unlawful uses of the data.
- **Industrial Platform**: Collaborative and strategic approach to exchanging data between a restricted group of companies. Data is shared in a closed and secure environment in order to encourage the development of new products or services, or to improve their efficiency. This solution makes it possible to offer value-added services. Therefore, it provides a more complete solution: it encourages data association, has control mechanisms regarding the use of the data, the model contracts made available to the platform reduce the costs of drafting framework agreements, etc. Platform expenses are borne by the companies that share data.

Company name:	**ITI**
Web address (company or initiative):	https://www.iti.es/
Description of the initiative:	Participation in the Transforming Transport project (https://data.transformingtransport.eu/), still in force, for the creation of a pilot initiative in collaboration with several national and international companies. The project as a whole has involved several data opening pilots in the transport and logistics sector. The sharing took place between a small number of companies.
Shared/open data type:	▪ Varies depending on the pilot (localisation, transactions, etc.).
Reasons for the opening/ sharing of data:	▪ Supporting open and collaborative innovation ▪ Contributing to the resolution of a local or global challenge
Sectors that re-use data:	▪ Distribution/Logistics ▪ Transport

Table 7 – ITI: Sectoral Data Sharing Example (European Funding - H2020).

- **Technical Facilitators,** with two possible approaches:
  - They allow the exchange of data through a technical solution. Revenues are obtained from setting up, using and/or maintaining the solution. Although they may resemble industrial data platforms or marketplaces, these companies can be described as technical enablers for the exchange of data between companies. These companies act as a third party in the data exchange process by providing the enabling technical solution. Unlike industrial data platforms or data markets, their revenue

comes from the configuration, implementation and maintenance of their solutions. Data users who participate in B2B data exchange through these solutions can exchange data within a particular community in a flexible manner. In addition, they make use of a tool that already exists and that can be customised, instead of using something new. Technical facilitators charge for the provision of the services they offer.

- They provide consulting services in data analytics from information obtained from third parties. Usually, they are organisations that have the necessary technical knowledge in data analytics and centre their business on providing such services. The end clients of these organisations can be both public administrations and final companies interested in the conclusions obtained after applying data analysis.

There are possible variations and combinations of these models, which can be adapted, in each case, to the existing business need. The term "data sharing" is used to describe all possible forms and models that support B2B data access or transfer.

5.1.2. B2G (Business to Government)

The provision and re-use of data in dealings with public administrations can take many forms, in both legal and technical terms. This section describes some of them in more detail:

- **Data Donation:** It can be considered a form of corporate social responsibility. The data donation program is supported by a dedicated team that supports any potential party interested in using the data.

- **Awards:** Launching competitions that encourage individual participation and participation of companies specialising in data analytics to find solutions to challenges presented by a specific public interest, e.g. H2020: *Horizon Prize Big Data Technologies*[30].

- **Collaboration Agreements with Public Administrations:** Public sector bodies may sign agreements with private companies, including mutual exchange of data, in accordance with the PSI Directive 22[31] concerning public sector information shared with the private sector. This can also benefit private companies, as they will be able to obtain information from the correlation of public and private sector data.

- **Intermediaries**: In cases where there is no prior relationship of trust between a company and a public sector body, an intermediary entity may be asked to obtain the knowledge needed for the public interest and to communicate it to the private company.

- **Civic Data Exchange**: Individuals may be encouraged to authorise public sector bodies to process their personal data by a private company. It should be noted that, in this case, the authorities should also comply with data protection legislation. Processing must be carried out in accordance with the applicable law (e.g. consent must be in accordance with Article 6(1)(a) or the performance of a task performed in the public interest in accordance with Article 6(1)(e)[32]). This model is more likely to work in situations where there is a sufficiently strong link between the citizen and the public sector body in question (e.g. the municipality where they live) or where the public interest purpose is particularly compelling from a public point of view.

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

[31] Directive 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).

[32] Where authorities rely on Article 6(1) of the General Regulations of the Data Protection Act ("processing is necessary for the performance of a task performed in the public interest"), such a legal basis must be imposed by European or Member State law. Furthermore, in the case of civic data exchange, data subjects shall have to be adequately informed on the topic, including the right to withdraw consent and any other possible processing of their personal data that might be carried out by the authorities.

5.2 Financing Models

An organisation involved in a data opening/sharing initiative must be able to quantify the economic value of the information it manages and the associated costs of carrying out the initiative, in terms of infrastructure and personnel costs.

Some practices to consider are detailed below.

1. Ensuring the initiative's cost-effectiveness

A cost-benefit analysis should be carried out, identifying and monitoring both the costs and benefits of an innovative open data publication initiative or a restricted data sharing one, depending on the chosen business model. It is a question of estimating the ROI of the innovation. Innovation has two types of impacts on the organisation: tangible and intangible.

Tangible impacts are relatively easy to calculate, following the recommendations of the Oslo Manual[33]:

- Revenue from new products/services sold in the last three years
- Cost savings from new processes in the last three years
- Economic benefit from productivity improvements in the last three years

Additionally, it should be considered that an innovation project may generate new intangible goods that will be transferred to the company's accounting assets: patents, trademarks or use models, etc. These assets are relatively easy to measure and there are advanced financial methodologies that allow to do it. In this way, we will be able to obtain values from trademarks or patents.

Such a business innovation project also affects other intangible assets:

- The brand image of a company
- Its positioning
- Environmental impact
- Intellectual capital that already existed before
- Attractiveness of a company as an employer

When monitoring the costs of the initiative, the costs associated with the start-up of new projects, as well as the maintenance costs of existing projects, must be taken into account.

The organisation shall also differentiate between costs related to data governance and those specifically associated with the technical process of data publication. In this way:

- an incorrect allocation of funding for new investments in both areas will be avoided.
- the efficiency of the data opening/sharing initiative will be improved by reducing cost overruns in both areas.

2. Valuation of datasets

The organisation shall have a quantification method to establish the value of

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

its datasets. Ideally, it should carry out a prior analysis in order to evaluate the existing interest or demand and possible models for the use of the information, although this may not be an easy task. At least, it is suggested to maintain, as far as possible, close contact with the re-users in order to be able to analyse the existing interest in the data of the organisation and to be able to make decisions in a more contrasted way.

The absence of a data valuation method should not hinder the data opening/sharing initiative. However, it is recommended to have one, as it helps to establish that datasets are higher priorities and therefore, it makes it easier to define the initiative's strategy (where to focus investments, data publication sequence, target market, etc.).

In order to assess a dataset, certain criteria related to the re-use of the dataset can be taken into account, such as:

- Availability of the data (information distributed in different sources, etc.).
- Format of the dataset.
- Type of information in the dataset.
- Singularity of the dataset.
- Need for geolocation (information classified by city, country, etc.).
- Time space in which the data will be accessed (days, weeks, months, etc.).
- Volume of shared data.
- Final use of the information (type of user, future reuses, benefits obtained by the final user, etc.).

It should be noted that the improvement in the re-use of a dataset does not necessarily have to be directly related to a higher value of the dataset. The final value of the dataset will be higher or lower depending on the re-user's use of the dataset.

3. Calculating Prices[34]

The following are some cost categories that can be taken into account when establishing prices in an opening/sharing data initiative:

- Costs related to the creation of data (necessary technology infrastructure):
  - Production: data and metadata generation, quality control, coding, etc.
  - Compilation: data collection and classification.
  - Anonymisation: deletion, obfuscation and impoverishment of databases.

- Distribution-related costs::
  - Infrastructure: development, maintenance of physical and logical equipment, software, etc.
  - Consultation: telephone and e-mail conversations with re-users, customer service costs, etc.
  - Other

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

- Data governance costs:
  - Human team.
  - Management of the initiative (monitoring and control of the process).
  - Awareness raising and training activities (if necessary).
  - Other.

Costs should be reviewed and adjusted periodically. An annual review is recommended.

Price calculation could therefore be based on:

- Pre-identified costs (governance costs and technical costs of the initiative).
- Valuation of datasets.
- Percentage adjustment, taking into account the business risks of the initiative (if any).
- Profit margin.

Some of the example companies from the ***Study on data sharing between companies in Europe***[35] , such as **ORANGE**, have reported using the following criteria to determine the value and prices of the data:

- Complexity of the project (visualisation options, data crossing from different Databases, etc.)
- Geographical coverage of the data (municipal, provincial, state, etc.)
- Data access time (day, week, month, etc.)

In the case of **GEO**, different criteria are applied to calculate the value of the datasets depending on each specific situation. However, the determining criterion in all cases is related to the expected benefit for the data re-user, which ultimately depends on the application of the data in each case.

In the case of **Michelin**, the price of the data is related to the cost reduction that organisations obtain through the use of their services (e.g. the EFFITRAILER service).

In the case of **API-AGRO**, the value of the datasets is estimated based on the technical costs of sharing the data across the platform. API-AGRO does not take into account the re-user's potential use of its datasets. A similar case is that of **ENEDIS**. This organisation makes its datasets available to third parties through an Open Data Platform. Only in certain specific cases does it choose to charge a small fee for access to its data. In these cases, the value of the datasets includes the cost of personnel arising from the effort required to share the datasets with quality guarantees.

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

5.3 Legal Framework

5.3.1 Contractual agreements in the B2B relationship model

Data sharing is normally implemented on the basis of contractual agreements. Data monetisation agreements are not only bilateral in nature, but they can also be executed by multiple parties.

The definition of contractual terms for the use of data or licensing agreements requires special attention, since they must comply with existing legislation, specifically with any legislation that prevents the exchange of data or is subject to specific conditions to ensure the strategic interests of each party.

Initiatives to develop model contract terms for different types of data exchange agreements and for some sectors or types of data exchange are currently under way. The European Commission plans to compile best practices, existing model contract terms and model checklists[36] through a Data Exchange Support Centre that will be operational in early 2019.

The following are some considerations that may assist companies in the preparation and/or negotiation of data sharing agreements:

1. What data will be available?

- Identify the data you want to share as concretely and accurately as possible (e.g. R&D data, customer data, diagnostic data), including the levels of updates expected in the future. When the mechanisms that make data analysis possible (e.g. methods, models, algorithms, etc.) are shared together with the datasets, they should be described.
- What quality levels can be guaranteed for the data now and over time? Shared data should be of good quality, i.e. accurate, reliable and, where necessary, up to date. Make sure there is no missing, duplicate, unstructured data. Specify the source/origin of the data and how it was collected/constructed. A mechanism to report data errors can be set up.
- Is data shared over a dataset or data stream?
- Ensure compliance with legal obligations that may prohibit data access or transfer. Ensure compliance with any rights others may have over the data. Verify intellectual and industrial property rights.
- Ensure compliance with data protection legislation. Among others, verify that there is a legal basis for the processing of personal data in line with the General Data Protection Regulation.

[36] See 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. **Who can access and re-use the data?**

- Ensure that the contract defines in a transparent, clear and understandable way who has the right of access, the right to re-use and the right to distribute the data and under what conditions. Specify if and how the data can be licensed for re-use. Be clear when explaining the licensing conditions for data re-use and distribution. Sublicensing (the original licensee is given the legal authority to grant an additional licence to others) should also be considered: 1) it must either be specifically excluded, or 2) the conditions under which it is permitted and for what type of data must be specified.

- The right of access and re-use of data does not have to be unlimited. The agreement may, for instance, limit the right of access: for example, only to members of specific professional groups or linked to certain data use purposes (e.g. for commercial use).

3. **What can the re-user do with the data?**

- During contract negotiations, the re-user should be as open and clear as possible about how the data will be used. This will ensure transparency and increase the trust of the data provider.

- Specify the exact use that can be made of the data, including rights over those uses derived from the data (analysis).

- Establish rules for subsequent non-disclosure to third parties.

4. **Establish the technical means for access and/or exchange of data, including**

- Access frequency and number of downloads.

- Security requirements.

- Service levels for support.

5. **What data do I need to protect and how do I protect it?**

- Ensure that the appropriate measures to protect your data have been implemented. The measures should be applied to data exchange transactions and data storage, as data can be subject to theft or malicious use. Data can also be accidentally released, for example through human error or due to a technical problem. Data may also be subject to unauthorised access or disclosure or it may be lost.

- Ensure the protection of trade secrets, sensitive trade information, licences, patents, intellectual property rights. Neither party shall aim to recover confidential information from the other party as a result of data sharing.

6. **Include liability rules for the provision of erroneous data, interruptions in data transmission, poor quality interpretive work, or for destruction/loss or alteration of data (if illegal or accidental) that may cause damage.**

7. **Establish the rights of both parties to conduct audits on the fulfilment of mutual obligations.**

8. **What is the expected duration of the contract? What are the rights when it comes to the termination of the contract? What notice should you give to your partners?**

9. **Agree on applicable legislation and enforcement mechanisms in the event of non-compliance.**

5.3.2 Contractual agreements in the B2G relationship model[37]

The following are some considerations that may assist private companies and public agencies in data sharing initiatives:

1. Public agencies should identify a **public interest purpose**, the private sector data needed for the purpose in question, and the level of granularity needed. Businesses, on the other hand, may also reflect that their data can contribute to a public interest purpose.

2. Parties should **identify the internal objectives and constraints** they face related to data sharing:

- Investment in expanding knowledge in data governance (awareness needs and team training, if needed).
- Organisations that have corporate departments responsible for data sharing will find that sharing data with public agencies (or also with private companies) is less costly in terms of data governance, infrastructure, legality, and legal issues.
- Companies and public sector bodies should ensure compliance with existing legislation in terms of GDPR data protection (ensure legality of processing, consent, proper use of anonymisation techniques, confidentiality and data protection from its design and by default, etc.). and ePrivacy (proposed regulation has not been approved at the date of publication of this guide).
- Public sector agencies should conduct a careful analysis of potential data sources and determine the limitations of a specific data provider.

3. Parties should choose the **technical and practical types of data sharing** that are best suited to their internal and data governance processes:

- Public bodies should safeguard the legitimate business interests of the private company (e.g. confidential business information, trade secrets) and ensure security of access to the data. Private sector data transferred to a public sector body should be treated as confidential data. This needs to be explicitly stated and adequate measures must be in place to ensure the security of the network and information systems.
- Public agencies may need to expand their technical capacities and their personnel in order to take advantage of the possibilities provided by private sector data.

4. The contract should **include conditions for implementation**, time constraints and the datasets that will be used:

- Public bodies should ensure that their request for private data complies with the principle of proportionality and is necessary to achieve the defined public interest objective. The agreement should specify that the transmitted data will be erased after the purpose has been achieved or when the time limit has been reached. The use of the same data for a different purpose should be subject to a new or modified collaboration agreement.
- The parties should define the conditions for data transfer at the operational level: data format and metadata, quality, granularity and duration of access, and mode of access.

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

- The parties should determine the compensation for data sharing. In this respect, there are different option:
  - Limit compensation to a proportionate recovery of the costs incurred in the production, preservation and dissemination of data.
  - Limit compensation to costs associated with the dissemination of data, considering that the costs of producing and preserving the data have already been covered by other sources of revenue.

    The choice of one option or another could be linked to the defined public interest objective.
- In order to enable public bodies to carry out data quality assessment, companies providing the data should offer their support. Companies will allow the quality assessment of their data for the stated purposes, even in cases where a formal audit or verification is required.

5. The parties should agree on the **use of common principles** for monitoring the implementation of the contract:

- A code of conduct can be agreed or existing ethical rules (such as those defined in the European Statistics Code of Practice[38]) can be used. A coordination committee can also be established or an independent auditor can be appointed to oversee the use of the data.
- Public bodies must implement the necessary protections to prevent misuse of the data for purposes other than those defined in the contract.

6. Include **liability rules** for the provision of erroneous data, interruptions in data transmission, poor quality interpretive work, or for destruction/loss or alteration of data (if illegal or accidental) that may cause damage.

7. Agree on **applicable legislation** and enforcement mechanisms in the event of non-compliance. Either party should be free to terminate the contract in the case there is a legal or technical risk relating to the processing or use of shared data.

8. Public bodies should disseminate the results of collaboration and guarantee the presence of public feedback mechanisms, where necessary or relevant, without compromising the confidentiality of private sector data.

5.3.3 Licences for Data Sharing/Opening

As mentioned above, it is essential that any shared/open data includes explicit information on the conditions of use; this is to establish the conditions for the re-use of the data and the context in which they can and cannot be re-used.

Simple Conditions of Use (opening of data)

The data is fully available for re-use under a series of minimum conditions, for which Royal Decree 1495/2011 of 24 October can be taken as a reference[39]:

- It is forbidden to distort the meaning of the data.
- The source of the data must be cited.

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

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

- The date of the last update should be mentioned.
- No reference may be made to data subjects participating in, sponsoring or supporting the re-use of the data.
- Metadata must be maintained and not altered, as well as the applicable conditions for re-use.

Predefined licensing

In this other case, it will be indicated that the data is available for re-use under the conditions established by any of the licences compatible with the principles of open data. At most, these licences may establish restrictions related to the attribution of the original source and the obligation to share any work done with the data under the same conditions. Some reference licences are listed below[40]:

Public domain licences, which allow us to share data without any restriction.	▪ **Creative Commons Zero** (CC Zero)[41] ▪ **Public Domain Dedication and License** (PDDL 1.0)[42]
Licences that only require attribution of the original source of the data	▪ **Creative Commons-Attribution** (CC BY 4.0)[43] ▪ **Open Data Commons Attribution License 1.0** (ODC-By 1.0)[44]
Licences that require attribution of the original source and sharing of results under the same conditions	▪ ***Creative Commons-Attribution-ShareAlike*** (CC BY-SA 4.0)[45] ▪ ***Open Data Commons Open Database License*** (ODbL 1.0)[46] ▪ ***Community Data License Agreement – Permissive,*** (CDLA-Permissive-1.0)[47] ▪ ***Community Data License Agreement – Sharing,*** **Version 1.0** (CDLA Sharing)[48] ▪ ***Creative Commons Attribution-NonCommercial 4.0 International*** (CC BY-NC)[49] ▪ ***Creative Commons Attribution-NoDerivatives 4.0 International*** (CC BY-ND)[50] ▪ ***Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International*** (CC BY-NC-SA)[51] ▪ ***Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International*** (CC BY-NC-ND)[52]

Table 8 – Licences to share/open up data.

The European Commission has a Licensing Assistant[53] who provides a description of the licences, how to apply them and how to combine them (if necessary).

The conditions of use or license of the data must be clearly defined to avoid problems such as:

- Not knowing what the applicable conditions for re-use are because there are no indications about it.
- Continuing applying default rights (e.g. those related to author's rights or ownership of databases) for not having previously waived them.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

Recommended Licensing Provisions

Scope of Application	This provision should define the geographical and temporal scope of the rights covered by the licence agreement, the types of rights granted and the range of permitted re-uses. It is advisable to indicate explicitly the rights not covered by the licence and to define broadly the types of rights granted (copyright, database rights and related rights, etc.).
Attribution	The attribution requirements must be clearly defined. At least the source of the shared data should be indicated: ▪ Declaration indicating the source of the data ▪ A link to the licence information
Exemptions	In case re-usable data sets are offered in conjunction with other non-reusable data sets, it is appropriate to explicitly indicate which data sets are not covered by the licence.
Definitions	It is advisable to define the main terms of the licence (licensor, use, information, licensee, etc.) concisely and in as simple a language as possible, and in line with those of the Directive and those incorporated into national legislation.
Exclusion of Liability	This provision should be used (to the extent permitted by applicable law) to draw attention to the fact that the licensor provides the information "as is" and assumes no responsibility for its accuracy or completeness. It should also be noted where the licensor is not in a position to ensure sustainable supply and access to the information in question.
Consequences of non-compliance	The consequences of non-compliance with licence conditions should be specified, especially if they include the automatic and immediate revocation of the re-user's rights.
Compatibility and versions	IIndicate other licences that are compatible with the licence, i.e. information from different sources can be re-used together under different compatible licences as long as any of the licences are respected. It is important to maintain a clear system for managing versions and dates of licences and to refer to it for updates.

Table 9 – Licensing Provisions.

5.4 Roles, skills, knowledge development and management

This section includes support activities related to the new responsibilities[54] associated with data governance, as well as the skills and knowledge[55] that an organisation that wishes to share/publish data needs to perform.

5.4.1. Roles

There are several actors involved throughout the life cycle of a data sharing initiative in an organisation. Depending on the nature of the activity to be developed, strategy profiles, technicians, etc. will be required. Typically, the following profiles can be found in a data sharing initiative:

- **Sponsor**: It is the sponsor of the data sharing initiative in the organisation. The sponsor will validate the overall approach, oversee the implementation of the data sharing initiative, and ultimately be responsible for establishing the strategy in the organisation.

  This role has no technical responsibility over the data, but provides the organisation with the financial, technical, and human resources necessary for the initiative to be successful.

  The sponsor has a clear organisational vision and is the one who sets the policies, vision and mission and, ultimately, the data sharing strategy. The sponsor's commitment to data sharing must be clear and stable throughout the process.

- **Data Manager**: The data manager is responsible for the data sharing initiative in the organisation. It is the person responsible for defining the data sharing strategy and is responsible for the entire process. This role groups the knowledge of everything concerning the initiative, from benefits, obstacles, organisational aspects, technical aspects, etc. The data manager[56] is responsible for implementing the strategy of sharing data with the economic, technical and human resources that the sponsor makes available. This role also must be aware of the current legislation on data sharing and is responsible for its compliance.

  The data manager must be able to identify internal resistance situations and, if there are any, they must know how to manage them properly.

  Additionally, this role is responsible for ensuring the quality of the data, performing frequent checks and ensuring that the defined quality standards are met. In the case of identifying non-conformities, the Data Owner is responsible for resolving the situation.

- **Implementer**: The implementer is generally responsible for the implementation of the technical requirements. Therefore, the role requires knowledge of the technical standards, specific tools and basic organisational requirements. The implementer can be an internal or external resource assigned by the data manager. Both roles should work together on the data sharing initiative.

- **Data Owner**: Each data field in each database of the organisation must be owned by a data owner, who has the authority to ultimately decide on the access and use of the data.

  Roles and responsibilities should be aligned with the data governance in place in the organisation, which in turn is one of the fundamental bases for deploying the data strategy along with the processes. However, depending on the organisation's organisational maturity, its size and its data manage-

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

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

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

ment model (centralised vs. decentralised), these roles or responsibilities will be assumed either in a timely and non-exclusive manner or with roles and responsibilities assumed by specialised players in the subject matter.

5.4.2 Skills

A mature organisation should ensure that the personnel has sufficient training and adequate support to carry out data management responsibilities. It should help develop skills at all levels of the organisation, supporting both operational and strategic needs.

An organisation focused on data sharing/opening must understand the benefits of sharing/opening and transparency, and apply those principles appropriately. To support the growth of this culture in the organisation, it must ensure that people have the necessary knowledge and experience in different fields.

- Shared knowledge of the value of data and its application in the organisation.
- Operational knowledge for data governance and implementation of data sharing channels.
- Strategic knowledge at management level to define future objectives related to data sharing.

The organisation will support people when it comes to developing the necessary skills to implement the data sharing strategy. If necessary, it will provide people with the necessary training in technical, operational, contractual and data governance aspects, among others.

Access to information is also an important component in the process of generating a data-oriented culture. The organisation shall ensure that all organisational standards, datasets and any relevant documentation are accessible to whomever needs them.

5.4.3 Knowledge development

The organisation should ensure that the personnel has the necessary training and support to fulfil individual responsibilities related to data sharing. This is probably the key success factor in raising people's awareness of the benefits of good data governance and its application to the organisation's strategy.
It is important that people are trained to ensure that they have the knowledge they need to:

- Understand the risks and benefits associated with data sharing and how they affect their responsibilities.
- Understand the organisation's data sharing policy and strategy.
- Have the ability to apply the appropriate level of data governance to projects.
- Identify necessary training in any area, e.g. technical training, regulatory training, standards, etc.

5.4.4 Knowledge management

Knowledge management is an essential process for any organisation. From the perspective of data sharing, knowledge management is relevant for two reasons:

1. People need to locate and use documentation related to data sharing standards and policies. This information is also likely to be relevant to data re-users. The data policy will define the organisation's commitment to data sharing/opening and how to engage re-users.

2. Second, both internal and external users must have access to the documentation necessary to support the use of shared data. For data to retain its value over time, the people who use it must know how it is collected, how it is processed, and who owns it.

6. EXPLANATORY GLOSSARY OF TERMS AND ACRONYMS

API (Application Programming Interface). A set of rules (code) and specifications that applications can use to communicate with each other.

Anonymisation. The process of converting data into a form in which specific individuals cannot be identified.

Apps. Application or program that can be installed on mobile devices and computers for the user to perform different types of tasks, dynamically and specifically.

B2B (Business to Business). Business models in which the transactions of goods or provision of services takes place between two companies.

B2G (Business to Government). Companies that sell or offer their services to government institutions.

Big Data. Any voluminous amount of structured, semi-structured and unstructured data that has the potential to be extracted to obtain information.

Blockchain. Data structure in which the information it contains is grouped into sets (blocks) to which meta-information relative to another block of the previous chain is added. Thus, thanks to cryptographic techniques, the information contained in a block can only be repudiated or edited by modifying all subsequent blocks.

Business Intelligence. Ability to transform data into information, and information into knowledge, to support decision-making processes in organisations.

CKAN. Software solution that makes data accessible and re-usable by providing tools for publishing, sharing, finding and using data (including data storage and provision of robust data APIs).

CMS (Content Management System). Software to facilitate the handling and managing of web content.

Data Mining. (Mining of data). Process of extracting information from large databases (information that reveals business intelligence) through trends and correlations. This allows to generate predictions that help solve business problems, and is able to provide a competitive advantage.

Data Science. Interdisciplinary field that involves scientific methods, processes and systems to extract knowledge or gain a better understanding of data in its different forms, whether structured or not.

Dataset Data collection, usually tabulated.

Datathon. Event or meeting that demonstrates the benefits of opening and analysing data and large volumes of information. It is conceived as a data marathon where different interested agents (technicians, citizens, researchers) are brought together and different sets of data are made available to them.

Data warehouse. Corporate database characterised by the integration and cleaning of information from one or more sources.

DKAN. Open source platform based on Drupal that allows Public Administra-

tions to publish data, provide visualisations and data histories and create internal analytical control panels.

DLTs (Distributed Ledger Technologies). A system in which records or data are shared, made accessible, and used by all members of a network.

FAIR (Findable, Accessible, Interoperable and Reusable). (Non-standard) principles initially defined by the research data community to make data findable, accessible, interoperable and re-usable. These principles are applicable to any type of dataset and involve the quality of the data for re-use.

IA (**Artificial Intelligence)**. Simulation of human intelligence processes by machines, especially computer systems.

Industry 4.0. Computerised manufacturing with all processes interconnected by Internet of Things (IoT).

IoT (Internet of Things). System of devices connected to each other through a network capable of transmitting information without the need for human interaction.

Linked Data. Structured data publication method (See Annex IV).

Machine-Readable. Data in such a format that it can be easily processed by a computer.

Marketplace. Internet site where business interactions take place between different companies.

Metadata. Data that describes other data. Describes the content, quality, conditions, history, availability, and other descriptive, structural, or administrative aspects of the data.

ODS (Operational Data Store). Active data container designed to integrate data from multiple sources with which to perform additional operations on the data itself.

OLAP. Solution used to speed up the consultation of large amounts of data in the context of Business Intelligence.

OLTP (Online Transaction Processing). Tool for the processing, management and daily maintenance of transactions generated by a company to guarantee high levels of availability, security and reliability.

PIWIK (Matomo). Web analytics tool belonging to an OpenSource community whose characteristics allow it to compete with most of the options available in the market.

Re-user. A person or entity that makes use of data opened or shared by a third party. The use of the data can be the same or different from the one for which they were conceived.

ROI (Return of Investment). A financial index that measures and compares the benefit or utility obtained in relation to the investment made.

GDPR (General Data Protection Regulation). This is the European regulation on the protection of individuals with regard to the processing of their personal data and the free movement of such data.

Smart Meter. Meter of consumables (such as gas, water or electricity), capable of transmitting and receiving information. It is able to communicate with other devices using a local area network.

SPARQL (Protocol and RDF Query Language). Standardised language for consulting RDF graphs, normalised by the RDF Data Access Working Group (DAWG) of the World Wide Web Consortium (W3C).

Utilities. Companies that offer public services of consumables such as electricity, gas, water, etc.

REFERENCES AND BIBLIOGRAPHY

Abella, A., Ortiz-de-Urbina-Criado, M., De Pablos-Heredero, C. (2014a). "Melo-da, métrica para evaluar la reutilización de datos abiertos". *El profesional de la información*, Nov-Dec 2014, v. 23, no. 6.

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

Berends, J.; Carrara, W. and H. Vollers (2017). Barriers in working with Open Data, Analytical Report No. 5 - European Data Portal, European Commission, Brussels.

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

European Commission (2014): *Guidelines on recommended standard licenc-es, data sets and charging for re-use of documents*, Communication from the Commission 2014/C 240/01, Official Journal of the European Union.

European Commission (2015): *Creating Value through Open Data*, European Data Portal, European Commission, Brussels.

European Commission (2016): *Study on data sharing between companies in Europe, Final Report*, European Commission, Brussels.

European Commission (2017): *Open Data Goldbook for Data Managers and Data Holders-Practical guidebook for organisations wanting to publish Open Data*, European Data Portal, European Commission, Brussels.

European Commission (2017): *Recommendations for Open Data Portals: from setup to sustainability,* European Data Portal, European Commission, Brussels.

European Commission (2018): *Open Data Goldbook for Data Managers and Data Holders-Practical guidebook for organisations wanting to publish* Open Data, European Data Portal, European Commission, Brussels.

European Commission (2018): *Guidelines for the exchange of private sector data in the European data economy*, European Commission, Brussels.

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

Government of Spain, Red.es (22 October 2014): *Decálogo del Reutilizador de Datos Abiertos del Sector Público*, [news entry]. Retrieved from https://da-tos.gob.es/es/noticia/decalogo-del-reutilizador-de-datos-abiertos-del-sec-tor-publico.

Red.es (21 December 2017): *Guía para publicar datos abiertos de manera rápi-da y sencilla (con CKAN)* [news entry]. Retrieved from https://datos.gob.es/es/documentacion/guia-para-publicar-datos-abiertos-de-manera-rapida-y-sen-cilla-con-ckan.

European Open Data Portal (2018). *Open Data Goldbook for Data Managers and Data Holders Practical guidebook for organizations wanting to publish Open Data*. European Open Data Portal, January 2018. https://www.europeandatapor-tal.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.

Government of Spain; Red.es (2016): *Guía metodológica para planes open data sectoriales- August 2016*. Madrid: Ministry of Industry, Energy and Tourism, Ministry of Finance, and Public Administrations and Red.es. Retrieved from: https://datos.gob.es/sites/default/files/actualizacion_informe_iniciativa_sectorial_0.pdf

Hapuarhachchi, T.; Balley, B.; Dodds, L. and 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.

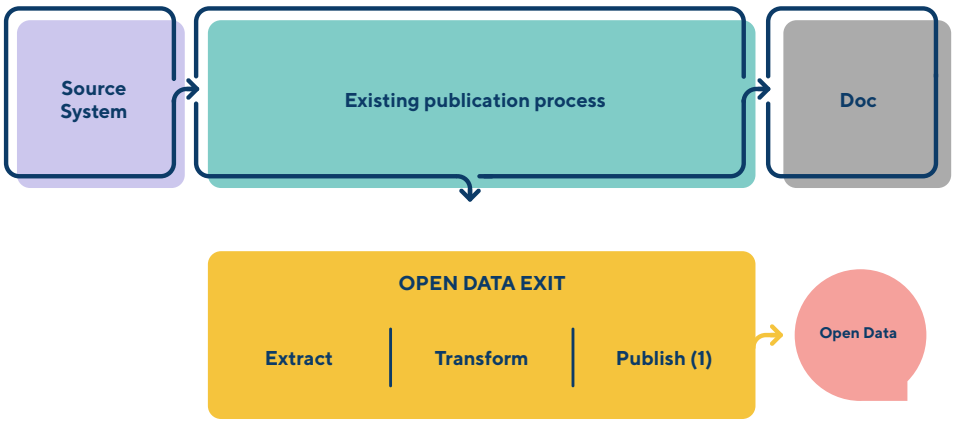
State Secretariat for Telecommunications and the Information Society (2009): *Guía Aporta sobre reutilización de la información del sector público-Proyecto Aporta.*

UK Data Service (2015): *Depositing shareable survey data*.

Williams, C. (5 April 2018): The future of Data Portals, [blog entry]. Retrieved from https://medium.com/vizzuality-blog/the-future-of-data-portals-1a9a20b20164.

ANNEX I: ETSD PROCESS SCENARIOS (EXTRACTION AND TRANSFORMATION)[57]

SCENARIO 1
ORGANISATION THAT PUBLISHES DATA IN PERIODIC REPORTS



ETP-Process with existing publication

An organisation (e.g. Gartner Reports) **publishes reports** (documents in Doc/Pdf/etc.) that contain data from different sources (internal or external), ready for publication. The data collected for the publication of reports can be transformed into an updated data stream shared by third parties and/or opened up:

- Existing data from previous publications: the existing publication process will be adapted to allow for the creation of an updated data stream to be shared/published at the stipulated frequencies.
- New reports: additional data entries will be added to the sharing/publishing process of the new data streams.
- Open Data: Data should be subjected to the open data publishing process (e.g. defining metadata, publishing the dataset or data stream preferably automatically, choosing the licence model, transforming the dataset using different formats, ensuring regular data updates, etc.).

SCENARIO 2
THE ORGANISATION PUBLISHES DATA IN DIFFERENT FORMATS

Description: The organisation already publishes datasets or data streams in different formats through its dissemination channels (e.g. the web). This data does not always meet the requirements of shared data or open data. However, with very little effort, data can be adapted for sharing or published as open data.

Extraction
Isolating and filtering the data in a uniform dataset. We can also offer filtered views of the dataset (less fields, anonymised, grouped) depending on whether it is an open data sharing or publication process.

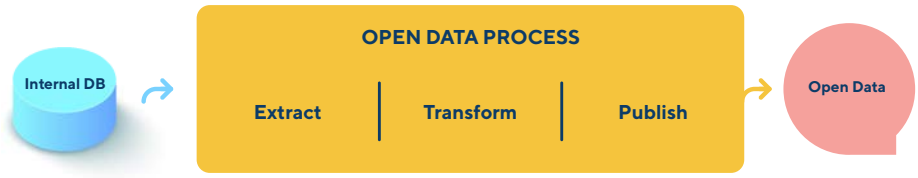
Shared/open data: Review existing processes to ensure that they meet all legal requirements for sharing with third parties or publishing them as open data. The latter are more restrictive than the former.

[57] The scenarios described in this annex are adaptations and, in some cases, they are the same cases appearing in Appendix 3 of the document https://www.europeandataportal.eu/sites/default/files/european_data_portal_-_open_data_goldbook.pdf

Transform
Apply the preparation process to transform published data into ready-to-open or linked data.

For the publication of open data, the data must be submitted to the organisation's publication process, as in scenario 1.

SCENARIO 3
THE DATA TO BE PUBLISHED/SHARED ARE IN A DATABASE



The ETP-Process starting from an existing database

Description: Extract data from a database and transform it into a data stream to share with third parties or publish in open. This scenario is different from the previous ones and assumes that the database is standard (e.g. Oracle, SQL server) and the organisation owns it or, in any case, can apply the extraction and transformation techniques to the tables managed by the database without any obstacles from third parties.

This is a typical case that requires applying extraction and transformation techniques.

Why is it necessary to extract data to a file? Because the data in the database (data in origin) is structured so that its access is optimal for the organisation's operational applications and it is not to be shared by third parties or published as open data. This structure is called OLTP (on-line transaction processing) and is relational in nature (the data is linked through relationships (keys) that allow the data to be linked between different tables.

If the data is in different databases, it is preferable to create a **logical data map** (destination dataset) that is abstracted from the existing physical relationships in the source databases and then apply the appropriate transformations (e.g. replace identifiers with values, make the references consistent - 0 = man; 1 = woman -, etc.) until the dataset corresponding to the logical data map is obtained. In any case, it is very rare to share or open the data as it appears in the starting database.

Additionally, if the data is in different source formats (e.g. databases, flat files, etc.), the scenario would be as follows (scenario 4).

Extract

Most database systems (e.g. ORACLE, SQL Server) have standard techniques for reading tables and writing to flat files. For example:

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

However, when baseline data changes frequently, a good alternative is to write programs that read the data directly from the Data Management System (DBMS) via ODBC – JDBC. ODBC (or direct SQL) allows to program more complex extractions.

Finally, it is also possible to use ETL tools.

From a practical point of view, if complex transformations are necessary, it is advisable to store the data in an intermediate database—temporary or permanent—, from which the final data stream will be obtained (to be shared or published in open). In this case, the two dimensions of the data stream must be taken into account:

- Exporting all the content of the database to be shared or published in open.
- Loading the modifications/delta with respect to the previous version and combine them with the data stream.

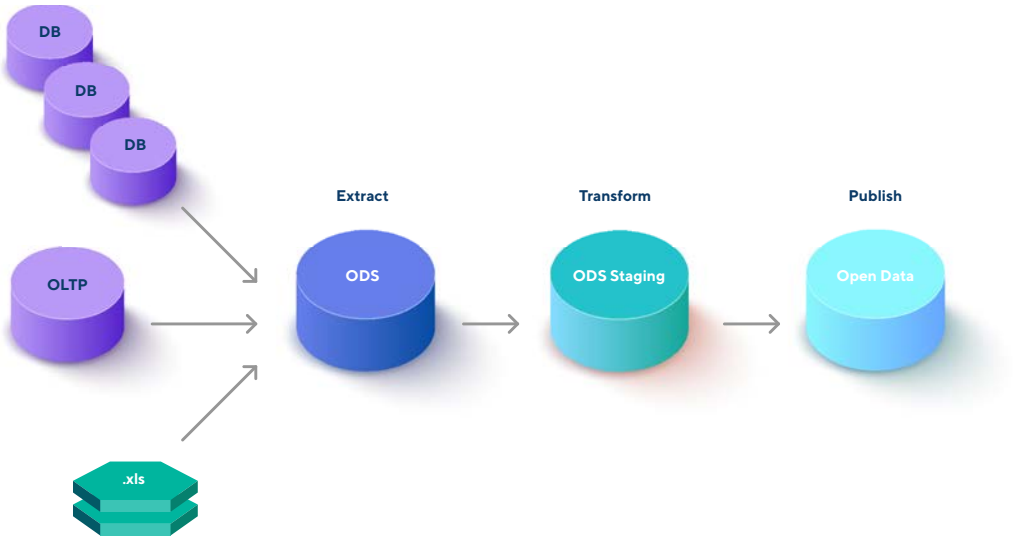
Transform

In this phase, as in the case of the data warehouse environment, the review of data quality is addressed. For example, using the same field names for the same data, transforming abbreviations into full names, saving data consistently, etc. Anonymisation processes, data grouping to achieve the same level of granularity, etc. are performed in this step.

Recommendation: If data changes very rapidly, it is preferable to make the data available to third parties as soon as possible rather than wasting time on transformations that are not strictly necessary for the purpose in question.

For the publication of open data, the data must be submitted to the organisation's publication process, as in scenario 1.

SCENARIO 4
ORGANISATIONS WITH CONSOLIDATED DATA FROM DIFFERENT SOURCES



Description: In this scenario, it is assumed that the organisation has the necessary experience and ETL tools. All the extraction and transformation techniques that are applied for the data warehouse or Business Intelligence are valid for the ETSD processes.

In an ETL process, the source data is extracted from different sources and stored in a consistent manner in an ODS (Operational Data Store). The data is also available at the lowest level of granularity. The ODS data is processed in different stages, which require the data to be stored in intermediate tables. The data is then aggregated and refined until it is consistent before being stored in the data warehouse. In the case of shared data or data to be published openly, the intermediate transformation steps and tables may be different.

In ETSD processes, it is not necessary to extract the data separately, so we can start directly from the ODS. In the transformation process, data streams can come from ETL processes. In some cases, it will be necessary to modify some phase of the ETL or replace one phase with another. It is recommended to re/use most of the ETL process. No additional tools are needed for these processes.

For the publication of open data, the data must be submitted to the organisation's publication process, as in scenario 1.

ANNEX II: PRINCIPLES THAT GUIDE RELATIONSHIP MODELS

B2B Business models

Transparency	Contractual agreements must transparently and clearly identify (i) the persons or entities that will have access to the data generated by the product or service, the type and detail of the data; and (ii) the purpose of the use of such data.
Protection of Commercial Interests	Contractual arrangements should protect both business interests and proprietary information and users of data.
Portability	Companies that offer a product or service that generates data as a by-product must allow the maximum possible portability of their data.
Shared Value	Contractual agreements should recognise that, when data is generated as a product derived from the use of a product or service, several parties have contributed to the creation of the data.
Fair Competition	Contractual arrangements should guarantee fair competition when exchanging sensitive commercial data.

Table 10 – B2B Business Model Principles

B2G Business models

Proportionality of Use	Requests for the provision of data from the private sector under preferential conditions for re-use must be justified by a clear and demonstrable public interest. The request for data from the private sector must be appropriate and relevant to the public interest purpose. The cost and effort required for the provision and re-use of private sector data should be reasonable in comparison to the expected public benefits.
Limited Purpose	The use of private sector data should be clearly limited to one or more purposes that are specified as clearly as possible in partnership contracts between companies and governments. The private sector company should receive specific assurances that the data obtained will not be used for unrelated administrative or judicial proceedings..
Protection of Trade Secrets	The organisation must guarantee that legitimate interests will be respected, specially the protection of trade secrets. The provision of business-to-government data should allow companies to continue to monetise the data in question.
Conditions for Re-use	Contractual arrangements should seek to be mutually beneficial while recognising the public interest objective. This can be accomplished by giving the public sector body preferential treatment over other clients. The agreed level of compensation could be linked to the expected public interest purpose. Data collaboration agreements between business and government which involve the same public bodies that perform the same functions should be treated in a **non-discriminatory manner.**
Reasonable Support	**Public bodies that benefit from private sector data provision** should offer reasonable and proportionate support to the companies that supply the data to, for example. help assess the quality of the data for the stated purposes, including the possibility of auditing or verifying the data. Businesses should not be required to invest in the improvement of data quality.
Transparency and Social Participation	Collaboration between companies and public bodies must be transparent. Best practices in corporate-government collaboration should be made available to the public as long as they do not compromise the confidentiality of data.

Table 11 – B2G Business Model Principles

ANNEX III: APPLICABLE LAW

GDPR (General Data Protection Regulation)

Regulation (EU) 2016/679 of the European Parliament and the Council[58], the new General Data Protection Regulation ("GDPR"), regulates the processing of personal data relating to individuals in the European Union (EU) by individuals, companies or organisations.

This regulation was subsequently updated in the corrected version of Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. It repealed Directive 95/46/EC (General Data Protection Regulation) (OJ L 119, 4.5.2016)[59].

From 25 May 2018, with the entry into force of the General Data Protection Regulation, there is a single set of data protection rules for all companies operating in the European Union (EU), regardless of where they have their headquarters.

The stricter rules on data protection mean that:

- People have more control over their personal data
- Businesses benefit from a level playing field

The following are some aspects that an organisation must take into account in order to comply with the regulations[60]:

IMPLEMENTATION OF THE REGULATION	
1.	Processing of personal data as part of the activities of one of its branches established in the European Union (EU), regardless of where the data is processed.
2.	The organisation is established outside the EU and offers products or services (paid or free) or observes people's behavior in the EU.
3.	It does not depend on the size of the company, but on the nature of its activities.
4.	The rules apply only to personal data, not to data on companies or any other legal person.

Table 12 – Implementation of the Regulation

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

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

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

PRINCIPLES OF THE REGULATION
1. The type and amount of personal data you may process depends on the reasons for the processing (legal reason used) and what you want to do with it: ▪ personal data must be processed lawfully and transparently ▪ specific purposes for the processing of the data must be stated and those purposes must be indicated to individuals when collecting their personal data ▪ only personal data needed to fulfill a specific purpose should be collected and processed ▪ it must be ensured that personal data is accurate and up to date ▪ it must be ensured that personal data is not kept longer than necessary for the purposes for which it was collected ▪ appropriate technical and organisational safeguards must be put in place to ensure the security of personal data
2. Principle of purpose limitation, the purpose for treating personal data must be known and the people whose data is processed must be informed.
3. The data may be used for purposes other than those initially defined,but only in cases where data has been collected on the basis of a legitimate interest, a contract or vital interests, and only after checking that the new purpose is compatible with the original purpose: ▪ the relationship between the initial purpose and the new or future purpose, ▪ the context in which the data was collected (what is the relationship between the company and the person?), ▪ the type and nature of the data (is it sensitive?), ▪ the possible consequences of further processing (how will it affect the person?), ▪ the existence of appropriate safeguards (such as encryption or pseudonymisation).
4. Personal data should only be processed when it is not reasonably possible to process it in any other way.
5. The data should be kept for as short a time as possible.
6. At the time of data collection, at least the following should be clearly stated to the individuals from whom data is being collected: ▪ identity (contact information and that of their data protection officer, if any), ▪ what use will be given to the personal data (purposes), ▪ the categories of personal data, ▪ the legal justification for the processing of the data, ▪ for how long the data will be kept, ▪ who else can receive the data, ▪ whether the personal data will be transferred to a recipient outside the EU, ▪ who are entitled to a copy of the data (right of access to personal data) and other basic rights in the field of data protection (see full list of rights[61]), ▪ the person's right to lodge a complaint with a data protection authority, ▪ the person's right to withdraw consent at any time, ▪ where appropriate, the existence of automated decisions and the logic applied should be stated, as well as the consequences they entail. See the full list of information to be provided[62].

Table 13 – Principles of the Regulation

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

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

LEGAL GROUNDS FOR DATA PROCESSING
1. You can only process data in one of the following circumstances: ▪ when you have the consent of the individuals concerned, ▪ when you have a contractual obligation (a contract between you and the customer), ▪ to comply with a legal obligation (as set out in the European Union or in the national legislation), ▪ where the processing is necessary for the performance of a task carried out for the public interest (as set out in the European Union or in the national legislation), ▪ to protect the vital interests of the data subject, ▪ to satisfy the legitimate interests of their organisation, but only after having verified that the fundamental rights and freedoms of the person whose data is processed are not significantly affected.
2. Where consent to the processing of personal data needs to be obtained, the following conditions must be met for such consent to be valid: ▪ to be given freely, ▪ to be informed, ▪ to be given for a specific purpose, ▪ having clearly established all the reasons for the treatment, ▪ being explicitly given and be given through a positive action (such as an electronic checkbox that the person concerned must explicitly check online or through the signature on a form), ▪ using clear and simple language and being easily visible, ▪ the organisation must have explained itself and given the individual the opportunity to withdraw consent (e.g. an unsubscribe link at the end of an email newsletter).
3. It must be as easy to give consent as it is to withdraw it.
4. Some flexibility is allowed regarding the level of specification and detail of consent in the context of scientific research.

Table 14 – Legal Grounds for Data Processing

OBLIGATIONS
1. Roles and Responsibilities: ▪ The controller determines the purposes and means related to the processing of personal data—if you decide why and how the personal data should be processed, you are the controller. ▪ If you jointly determine why and how the personal data should be processed (together with one or more organisations), you are co-responsible for the processing. Those responsible for processing must enter into an agreement establishing their respective responsibilities regarding compliance with the rules of the General Data Protection Regulation. ▪ The data processor processes personal data only on behalf of the data controller. The data processor is usually a third party external to the company; however, in the case of groups of companies, one of them may act as data processor for another.
2. Another person (natural or legal person or any other body) may process personal data for you as long as there is a contract or other legal act.

3.	The General Data Protection Regulation (GDPR) is based on a risk-based approach—in other words, organisations processing personal data are encouraged to apply protection measures that are commensurate with the level of risk of their data processing activities.
4.	A data security breach occurs when the data for which you are responsible suffers a security incident that results in a breach of the confidentiality, availability or integrity of the data.
5.	A DPIA (Data Protection Impact Assessment) is required where the processing may involve a high risk to the rights and freedoms of individuals, at least in the following three cases: ▪ systematic and comprehensive assessment of an individual's personal aspects, including profiling, ▪ large-scale processing of sensitive data, ▪ large-scale systematic observation of a public area.
6.	A DPO (Data Protection Officer) should be appointed, whether they are a controller or a processor, whether their main activities relate to the processing of sensitive data on a large scale or whether their main activities relate to the routine and systematic observation of individuals on a large scale.
7.	The DPO may be a member of your organisation. You may also hire an external staff member under a service contract. DPO functions are as follows: ▪ to inform and advise the controller or processor, as well as their employees, of their obligations under the Data Protection Regulation, ▪ to monitor the organisation's compliance with all legislation relating to data protection, including audits, awareness-raising activities and training of personnel involved in processing operations, ▪ to provide advice when a data protection impact assessment has been carried out and to monitor compliance, ▪ to act as a contact point for requests from data subjects concerning the processing of their personal data and the exercise of their rights, ▪ to cooperate and act as a contact point with data protection authorities on data processing issues
8.	The protection offered by the General Data Protection Regulation (GDPR) moves along with the data, which means that the rules protecting personal data will continue to apply regardless of where the data ends up. This also applies when data is transferred to a non-EU State.
9.	The principle of proactive liability is a cornerstone of the General Data Protection Regulation (GDPR). According to the GDPR, every company or organisation is responsible for complying with all data protection principles, as well as demonstrating such compliance.

Table 15 – Obligations.

CITIZENS' REQUESTS	
1.	Individuals may contact the controller in order to exercise their rights under the General Data Protection Regulation (right of access, rectification, erasure, objection, portability and limitation to processing.
2.	**When someone requests access to their personal data, you should do the following:** ▪ confirm whether or not you are processing the data subject's personal data, ▪ provide a copy of the personal data you hold about that person, ▪ provide information on the processing (such as purposes, categories of personal data, recipients, etc.).

3.	**The General Data Protection Regulation gives individuals the right to request that their data be erased; organisations have an obligation to do so, except in the following cases:** ▪ the personal data you possess is necessary to exercise the right to freedom of expression, ▪ when a legal obligation requires the data to be retained, ▪ for reasons of public interest (such as public health, scientific or historical research purposes).
4.	Individuals have the right to object to the processing of personal data if they state the specific reasons affecting them.
5.	Individuals have the right to receive from their organisation the personal data provided in a structured, machine-readable format and to have it transmitted to another company or organisation (right to data portability).

Table 16 – Citizen Requests.

COMPLIANCE AND SANCTIONS	
1.	One of the functions of the Data Protection Authority (DPA) is to publish expert advice on data protection issues.
2.	The EDPB (European Data Protection Board)[63] is a European Union (EU) body responsible for the application of the General Data Protection Regulation (GDPR) from 25 May 2018. It is composed of the Director of each Data Protection Authority (DPA) and the European Data Protection Supervisor or their representatives. The European Commission participates in the meetings of the EDPB without voting rights. The EDPB secretariat shall be managed by the European Data Protection Supervisor.
3.	If data processing is carried out in different countries, the competent DPA (which will be the main authority in its relations with other DPAs involved in the EU) will be the DPA of the EU country in which it has its main establishment. Its central administration in the EU shall be the place where decisions on the purposes and means of the processing of personal data are made, and where the power to implement these decisions is located. If it does not have a central administration in the EU, the main establishment will be the place where the actual and effective exercise of the management activities determining the main decisions, such as the purposes and means of processing, takes place.
4.	The General Data Protection Regulation (GDPR) provides different tools for data protection authorities (DPAs) in case of non-compliance with data protection rules: ▪ possible infringement: e.g. a warning can be issued, ▪ infringement: the possibilities include a warning, a temporary or definitive ban on processing and a fine of up to EUR 20 million or 4% of the total annual worldwide turnover.
5.	Individuals can claim compensation if an organisation has not respected the Data Protection Regulation (GDPR) and if he or she has suffered material damage (such as economic loss) or immaterial damage (such as anxiety or loss of reputation).

Table 17 – Compliance and Sanctions

e-privacy

The proposal[64] for a Directive on privacy and electronic communications ensures the protection of fundamental rights and freedoms—particularly, the respect for privacy, the confidentiality of communications and the protection of personal data in the electronic communications sector. It also ensures the free movement of data, equipment and electronic communications services within the European Union. It incorporates the fundamental right to respect privacy in the communications sector into the EU's secondary legislation, as laid down in Article 7 of the Charter of Fundamental Rights of the European Union[65].

Privacy legislation must be adapted in order to align with the new European Data Protection Regulation (GDPR).

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

[64] At the time of publication of this guide, this is a proposal for a Directive that has not yet been adopted.

[65] https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=CELEX%3A52017PC0010

The proposal for the new e-privacy regulation aligned with the European Commission's GDPR includes[66] aspects such as:

- **New players:** In the future, privacy rules will also apply to new players who provide electronic communications services, such as WhatsApp, Facebook Messenger and Skype. This will ensure that these popular services guarantee the same level of confidentiality of communications as traditional telecommunications operators.

- **Stricter standards**: All individuals and businesses in the EU will enjoy of the same level of protection for your electronic communications through of this regulation. Companies will also benefit from a single set of standars across the EU.

- **Stricter rules:** All individuals and businesses in the EU will enjoy the same level of protection for their electronic communications through this regulation. Businesses will also benefit from a single set of rules across the EU.

- **Communications content and metadata:** The privacy of metadata and the content of communications, e.g. time of a call and location, is guaranteed. Metadata has a high-privacy component and should be anonymised or deleted if users have not given their consent, unless the data is necessary for billing

- **New business opportunities**: Once consent is given for the communications data (content and/or metadata) to be processed, traditional telecommunications operators will have more opportunities to provide additional services and develop their businesses

- **Simpler rules for cookies**: The new rule will be easier to use, as the browser settings will provide an easy way to accept or decline the tracking of cookies and other identifiers. The proposal also states that no consent is needed for non-private intrusive cookies to enhance the Internet experience (e.g. to remember shopping cart history) or cookies used by a website to count the number of visitors.

- **Protection against spam**: This proposal bans unsolicited electronic communications by e-mail, SMS and automated calling machines. Under national law, individuals will be protected by default from receiving marketing calls. People calling for marketing purposes must display their telephone number or use a special prefix indicating that it is a marketing call.

- **More effective enforcement**: Enforcement of confidentiality rules will be the responsibility of the data protection authorities, which are already in charge of the rules under the GDPR.

Database Directive

The Database Directive[67] stipulates that the creators of databases have the right to control access to and re-use of their content, but only in those cases in which it represents a large economic investment. In May 2017, the European Commission launched a public consultation[68] to assess the impact of the Directive on users and identify the adjustments needed to improve its applicability.

The preliminary results of this public consultation were released in October 2017[69]. The consultation received a total of 113 responses, 18 of which came from companies. On the basis of the responses received, the majority of respondents considered that the original objectives of the Directive are still valid today, although there is no unified opinion on the extent to which the Directive has achieved them.

The majority of respondents were of the opinion that the *sui generis* right es-

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

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

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

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

tablished by the Directive sufficiently protects the investments made in the creation and maintenance of databases, but opinions are divided as to whether the current scope of this right is sufficiently broad today. Views also diverge when it comes to the ability of the Directive to strike a balance between the rights and interests of the owners and users of the Database, and there is no consensus as to which approach should prevail.

PSI Directive and Open Data

There is a specific European Directive for the re-use of data in the public sector which, although it is not initially the focus of this guide, should be taken into account for those cases in which the sharing of data is carried out with public bodies, especially since its last version of 20 June 2019, where it becomes known as the "Open Data and PSI Directive"[70].

The Directive on the re-use of public sector information provides a common legal framework for a European market for government data (public sector information). It is built around two fundamental bases: transparency and fair competition.

This Directive on the re-use of public sector information, also known until June of this year as the "PSI Directive" (Directive 2003/98/EC[71]) entered into force on 31 December 2003. It was revised by Directive 2013/37/EU[72], which entered into force on 17 July 2013. Its latest version is called: *Directive (EU) 2019/1024 of the Parliament and of the Council of 20 June 2019 on open data and the re-use of public sector information*. This comprehensive review of the 2013 Directive also takes into account the results of the public consultation carried out by the European Commission at the end of 2017 and entered into force on 26 June 2019. The PSI Directive focuses on the economic aspects of the re-use of information rather than on citizens' access to information. It encourages Member States to make as much information available for re-use as possible. The latest revision of the Directive also encourages Member States to go beyond the minimum requirements set out in the Directive by applying the requirements of the Directive *to documents held by public undertakings relating to activities which, by virtue of Article 34 of Directive 2014/25/EU of the European Parliament and of the Council, are directly exposed to competition*. Member States may also decide to apply the requirements of the Open Data Directive to private companies, specifically to those providing services of general interest.

In July 2014, the Commission issued guidelines to help Member States transpose the revised rules and indicate best practices for the re-use of public sector information in each country. In Spain, Law 37/2007 on the re-use of public sector information of 16/11/2007, BOE no. 276 of 17/11/2007, no. 19814, p. 47160 – 47165, as subsequently amended by Law 18/2015 of 9 July, BOE no. 164/2015 of 10 July 2015, p. 57436[73], is created, and we hope that it will soon be modified in accordance with the new Directive 1024/2019.

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

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

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

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

ANNEX IV: DESCRIPTION OF SOME KEY TECHNICAL CONCEPTS IN THE FIELD OF OPEN DATA

This section describes some relevant concepts related to open/shared data, as well as open data techniques necessary for the opening and sharing of data following existing good practices in the public domain.

Linked data

Linked data is a technology that involves a set of design principles for sharing and linking data on the Internet or the Data Web, and involves an evolution of the idea of the Semantic Web linked to data and metadata. Linked data involves not only the standard publication of datasets, but their interconnection or linking through graphs. The data is tagged with uniform descriptors, which convert the data into dynamic and interpretable information, regardless of the language it is in or the type of file that contains it. These technologies and the associated standards (RDF, JSON-LD, etc.) improve interoperability and make it easier for third parties to find the data or dataset (discoverability).

Linked data uses various techniques, including RDF, vocabularies and URIs. Many catalogues of open data portals that publish linked open data use a predefined vocabulary.

In the field of linked data, it is essential to model the data to reflect the nature and value of the data and to establish semantically relevant relationships with other datasets. Modelling must be able to represent the defined thematic or organisational area or domain. For this purpose, a specific modelling language, a standardised vocabulary, or even the local design of a vocabulary or ontology of its own is necessary. Since semantic relationships can be established between datasets, the modelling of descriptive metadata of datasets acquires special relevance and is very important for establishing global interoperability. The process of assigning licences is another very relevant aspect and depends on the conditions of publication of the data that the organisation has. This may require from hardly restrictive licences or completely open ones (Linked Open Data, LOD) to the publication of unopened data (Linked Closed Data).

As defined by Tim Berners-Lee in his five-star incremental scheme (which we saw in section 4.4.2), if we really want to achieve efficient re-use and exploitation of business data, the objective should be to comply with the principles of linked data or LOD beyond the mere publication of data sets in proprietary formats not prepared or intended for re-use.

More information about linked data: https://www.w3.org/standards/semanticweb/data. Also recommended is the e-book arising from the EUCLID project, Using linked data effectively.

Metadata, metadata standards and publication of metadata catalogues

Metadata is data, data about data, structured information that describes, explains, locates or, in any case, facilitates the retrieval, use or management of information.

Metadata must be understandable to both people and machines (machine-readable). Metadata is also essential for the contextualisation of data, as well as for linking/mapping users and software with data available on the Internet.

Metadata helps, among other things, to:

- Discover (find) the dataset
- Understand the nature and structure of dataset
- Know the origin of the dataset and under what terms it can be used.

And above all, metadata helps to make the data useful and of quality. The better the metadata (accurate, adequate, complete, etc.), the better the data and its chances of recovery and re-use.

Metadata Standards

Standards are essential for the contextualisation of metadata. They provide a common structure for metadata and facilitate interoperability between systems, as well as accessibility and discoverability of datasets.

We can classify metadata, vocabularies or metadata vocabularies, mixing different levels and criteria (systems, practices, models, schemas, elements, uses and trends). From the multiple criteria available when categorising metadata, some of them are applicable at the level of metadata registration (record) and others at the level of schema. Let's analyse now only those that are applicable at schema or format level.

- Depending on the purpose or field of application of metadata, we can distinguish between general purpose metadata (such as CD) and metadata for specific purposes (such as domain-oriented metadata, e.g. ISO 19115/FGDC to describe geospatial information).

- Depending on the use of metadata. This is perhaps the vaguest and yet the most important criterion for classifying metadata types, since the main purpose of metadata, as we have already said, is to make the data useful. Thus, any type of metadata application and any utility attributed to it can be included in this category. Depending on the application or functionality of the metadata, we can indicate two fundamental categories: metadata for the retrieval of information and metadata for the use of the resource.

- Taking into account the **level of standardisation**. Although all metadata initiatives and all vocabularies are born to be standard, we can classify the metadata models, schemas, etc., also according to the level of standardisation they reach. In a generic way, we can distinguish:

    - *De jure* metadata standards, usually recognised nationally or internationally, and are metadata models or infrastructures for the creation of vocabularies that have acquired formal approval thanks to the standardisation process carried out by a standardising institution. For example, the case of Dublin Core, which is an ISO 15836-2008 standard; or ISO 19115-2003 for the description of geospatial information, etc.

    - *De facto* metadata standards, which are usually Public Access Specifications or PAS, which are those that deal with one or several aspects of a metadata policy (elements, values, good practices, etc.) that have achieved some recognition or dominant position and whose consultation is open to everyone, because they are publicly disseminated.

Generally, before a standard becomes a de facto or de jure standard, we like to say (Méndez & Van Holland, 2014) that metadata initiatives are standards "by se-

duction", which implies that a community, a project or a scientific domain adopts a model or a standard because it is useful, simple, basic, extensible, etc. Establishing levels of standardisation for our metadata and vocabularies is also crucial in any project to **create and maintain** a data sharing or open service/system.

Some of the most used standards for data opening in the public sector are:

- **DCMIDC** – *Dublin Core/ Dublin Core Metadata Initiative*[74]: General purpose metadata standard developed since 1995. It was initially used to describe the contents of web pages, and later, it was used in the field of digital libraries and Open Access archives, among other multiple domains, to promote interoperability. The simple Dublin Core format is an ISO 15836 standard that collects and explains the semantics and scope of 15 elements of this metadata model, without going into coding or syntax details.

- **DCAT** - *Data CATegory Vocabulary*[75]: DCAT is an RDF vocabulary designed to facilitate interoperability between data catalogues published on the Web, including public sector information (PSI) data portals. It is a metadata standard developed in the context of the Web Consortium and used for data portal datasets published on the Web. It is based on the Dublin Core philosophy, but it has the specific purpose of describing public sector datasets and information. DCAT is currently being extended to other types of data on the web through different metadata application profiles or other vocabularies. If DCAT is used to describe datasets from different data catalogues, those responsible for publishing them (e.g. private sector companies) increase the possibilities of access and interoperability between multiple metadata catalogues.

- **CERIF** – *Common European Research Information Format*[76] is a metadata standard for sharing data on research in Europe, linked to the CRIS (Current Research Information Systems) of universities and other research organisations.

- **Inspire**: is a European standard used for the description of geospatial information and data.

It is recommended that metadata be published in machine-readable formats and that standard terms be used to define the metadata. In addition, the general characteristics of the dataset should be described by using information about local parameters, licences, origin and quality of the dataset. Selecting standard metadata formats guarantees its interoperability and re-use. The standard or schema of linked metadata. The DCAT-AP standard is recommended.

Publication of metadata

Although metadata is linked to the datasets it describes, it may be useful to have the metadata associated with a dataset in multiple sites. For example, the open data portals of the European Union share the metadata of the open data portals of the Member countries. In some cases, only a subset of the metadata associated with the datasets is shared (those corresponding to the subset of the dataset to be shared). In others, in addition to all or part of the metadata, the dataset itself is shared. This practice of open data with public data could be extrapolated to private or business data. Sharing metadata across multiple sites substantially improves the availability of datasets. This sharing can be done manually or automatically through open data portals by extraction or metadata harvesting.

See https://www.europeandataportal.eu/en/providing-data/how-to-be-harvested-by-us to find out how to federate datasets in the European open data portal.

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

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

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

It is recommended to publish metadata as linked data whenever possible[77]**.** It improves ease of data access and interoperability. Many portals incorporate a software solution that allows to easily create the metadata associated with datasets.

Other Metadata Vocabularies

Metadata generally describes different attributes or properties of data and other digital objects, giving them meaning, context, and organisation in a standardised way. The first categorisation of metadata appeared 20 years ago, and it studies metadata by analysing the different functionalities supported by existing metadata elements. However, the identification of the many uses and dimensions of metadata leads to building a very broad typology. Most traditional types refer to the functional use and objectives of metadata rather than the inherent qualities of metadata **schemas**, metadata elements or any value of such elements (**schemes**). Additional information and training on metadata can be found in:

- 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

Compiling: best metadata practices for the description of datasets:

- Publishing the metadata next to the data.
- Using machine-readable formats for metadata.
- Using standard terms to define metadata.
- Describing the general characteristics of the dataset.
- Providing information on local parameters (date, duration, value, types, language).
- Providing information on the licence associated with the dataset.
- Providing information on the origin of the dataset.
- Providing information on the quality of the dataset

Ontologies
RDF (Resource Description Framework)

RDF (Resource Description Framework) is a standard data coding language that allows to describe and exchange resources on the web through metadata—not only documents, but all types of entities and digital objects that have a web presence, including people, organisations, products, services, and so on. RDF is an essential tool for making the Web a global infrastructure for sharing and re-using data and documents between different types of users and systems. RDF is a standard for standardised coding of data and metadata on the Web. It is the key language for semantic web and linked open data.

RDF provides a general and flexible method to decompose any dataset into small pieces called triplets, with some semantic rules to link those pieces. Each of these triplets are sentences or statements about the dataset or any other entity that you want to describe in a normalised way. RDF is based on the idea that the things described have properties that, in turn, have values, and that the resources can be described by statements such as the one in the example, which specify what those properties and values are. The terminology used to name those parts of the statement is: subject, predicate and object.

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

- **Subject**: identifies the resource or dataset on which the statement is about
- **Predicate**: identifies the property or characteristic of the subject that is expressed by this statement (e.g. language of the data, date of creation, etc.).
- **Object**: identifies the value of the property to which the predicate refers.

More information about RDF: *http://www.linkeddatatools.com/introducing-rdf*

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

Uniform Resource Identifiers (URIs) are a type of persistent identification of resources or datasets. A URI refers either to a text, e.g. URN (Uniform Resource Name), or to a location, URL (Uniform Resource Locator). Through triplets and graphs, URIs define each of the elements of a declaration or triplet in the field of linked data and its coding languages (RDF, JSON-LD, etc.). The URI is used to identify the subject of the declaration (the dataset), which allows to locate it and access it on the network. However, many of the elements that might be part of an RDF declaration do not necessarily have to have a location on the network, so RDF proposes the use of IRI (International Resource Identifier) (a name that replaces that of URI (Uniform Resource Identifier) in the new version of RDF 1.1). IRIs are a special type of URL that allow to refer to anything that needs to be mentioned in an RDF expression, including datasets, especially if these are only shared, but not expressly open. RDF uses IRIs as the basis for its identification mechanism for subjects, predicates and objects.

More information about URIs in the field of dataset re-use
https://joinup.ec.europa.eu/sites/default/files/c0/7d/10/D7.1.3%20-%20Study%20on%20persistent%20URIs.pdf

ANNEX V: NON-EXHAUSTIVE SELECTION OF FREE SOFTWARE TOOLS AND APPLICATIONS FOR THE IMPLEMENTATION OF OPEN DATA

Atlas (Apache): https://atlas.apache.org
ATLAS is a free APACHE software tool that facilitates the integration of an organisation's data ecosystem by using Hadoop-compatible metadata.

APACHE ATLAS provides open metadata management and allows organisations to build their dataset catalogue, classify datasets and govern datasets. It also includes functionalities that facilitate collaboration with data researchers (e.g. big-data experts) and data managers in the organisation.

CKAN : http://ckan.org
It is a tool developed by Open Knowledge and it has become a standard for open data portals. It is a freely available free software application supported by a continuously improving community. The CKAN catalogue system includes functionalities such as harvesting, publishing and auditing of data, as well as having its own integrated data storage system. CKAN needs a CMS to work. It is recommended to use it with Drupal, Wordpress, and Django.

Datatank: http://thedatatank.com
It is a free software tool, such as CKAN, Drupal or Elastic Search, which transforms datasets into http API format.

DKAN: http://demo.getdkan.com
It is also a free software catalogue system developed by Drupal. It is recommended by open data portals in the USA. It works well with CKAN and has an integrated CMS based on Drupal, although it can be easily integrated with other CMS.

ODI Certificate Tool: https://certificates.theodi.org
ODI (Open Data Institute) has developed a self-assessment tool that allows to identify the maturity of the datasets with respect to their technical and legal openness. It also evaluates the steps that must be taken so that datasets can be found more easily by third parties. Self-evaluation is a source of relevant information to identify improvement actions in the data sharing/opening initiative in an organisation. That application also helps the institution that uses it for certification with the open data maturity model.

OPENREFINE : http://openrefine.org
It is a tool developed by Google that allows to clean and analyse datasets.

OPEN DATA SOFT: https://www.opendatasoft.com
Data sharing platform that teams use to access, re-use and share data that makes the business grow.

SNORQL: http://snorql.nextprot.org
This tool is useful to learn how to make queries in SPARQL.

SOCRATA: https://socrata.com
It is a secure API platform that allows government organisations to modernise their organisations' data infrastructures. It drives a repeatable process by which staff can take data from multiple source systems, automate its flow to the cloud, organise it uniformly, and then turn it into an ever-active service for downstream data reuse.

SpagoBI. Dashboards I (Fi-ware): http://www.stratebi.com/spagobi
It is an Open Source Business Intelligence platform that covers and satisfies all BI requirements, both in terms of analysis and data management, handling and security.

For the analytical part, SpagoBI offers solutions for reporting, multidimensional analysis (OLAP), data mining, dashboards and ad-hoc queries. It also has its own modules for the management of collaboration processes and geo-reference analysis.

SpagoBI is an integration platform (and not a product platform), since it is not built and it revolves around a predefined set of tools. It has a modular structure in which all modules are related to the core system, which ensures the harmony of the platform along with its evolutionary capacity.

VIRTUOSO: https://dbpedia.org/sparql
It is one of the best known tools to execute queries in SPARQL.

ANNEX VI: QUESTIONNAIRE ON OPEN DATA INITIATIVES IN THE PRIVATE SECTOR

Justification and objectives

COTEC's Open Data Group, coordinated by SPRI, has the fundamental goal of contributing to the promotion of the opening up of data as a valuable resource in private companies. In order to encourage these data opening processes, and as a starting point, we decided to draw up a *Guide to Opening and Sharing Data in the Business Environment*. In a good number of cases, organisations do not undertake these processes because there is no clear awareness of the benefits they may entail and/or because the protocols and intermediate steps to be taken are unknown.

In this sense, companies that have already embarked on this type of initiative have previous experience that may encourage other companies to follow this path. The examples of these other companies (included in the *Guide* as case studies), as well as the knowledge of the real reasons or obstacles that prevent companies from getting involved in this type of initiative, the factors that would help to change this approach and the company's plans with respect to the open data in the near future are essential questions that promote these opening processes.

Thus, in order to get to know all this valuable information first hand and integrate it into the *Guide to Opening and Sharing Data in the Business Environment*, we ask for your collaboration in answering this questionnaire[78], which is valid whether or not you have carried out open data projects. The questionnaire clearly specifies which questions have to be answered in each case, although it has been structured as a logical unit. The full reading of the questionnaire aims to serve as a learning experience for all types of companies. Public Administrations or associations can also contribute by sharing their experiences or the experiences from other companies in their environment.

The *Guide to Opening and Sharing Data in the Business Environment* will allow to share the terminology and processes that lead to a successful opening of data, as well as to support the definition of an open data strategy. Flexibility is its main feature, since it will address any sector and any size of company. *The Guide* describes the main patterns of data sharing/opening, as well as the main reasons that lead companies to share/open their data. It also addresses the methodological framework of data opening/sharing, focusing on the description of the different models of data opening, the necessary consideration of the legal framework when dealing with any initiative of this type, and the detailed consideration of the steps that need to be taken to address opening projects (strategy). Finally, it will include the main recommendations given by different companies that have carried out data opening processes. The advice has been collected through this questionnaire and interviews.

Thank you very much for your collaboration!

[78] Complied by the author - Based on:
https://publications.europa.eu/en/publication-detail/-/publication/004fcf02-49c7-11e8-be1d-01aa75ed71a1/language-en/format-PDF/source-69800191

1. PRIOR INFORMATION

Name of the company/organisation

_____

Name and surname(s) of the person answering this questionnaire

_____

Position in the company

_____

INFORMATION ON THE DATA OPENING INITIATIVE

2. Title and brief description of the open data initiative *(objectives, justification of opportunity). If there is information available online on that initiative, please also include the address of the website(s)*:

2.1. When was the initiative launched?

_____

2.2. How long did it last?

_____

2.3. Is it still active? ☐YES ☐NO

2.4. Has this initiative led to further data opening initiatives ☐YES ☐NO

3.a What type of data have you opened?

☐ Location/positioning data.

☐ Transaction data (customers, suppliers).

☐ Internal operational data (quality, production, ...).

☐ Data in real time (or almost real time).

☐ I don't know.

☐ Other (specify):

_____

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

3.b. ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. Should you become involved in a data opening initiative in the future. what type of data do you think your company is most likely to open? *(You can check several options)*

☐ Location/positioning data.

☐ Transaction data (customers, suppliers).

☐ Internal operational data (quality, production, ...).

☐ Data in real time (or almost real time).

☐ I don't know.

☐ Other (specify):

_____

4. What have been the main users of your organisation's open data? *(multiple choice)*

☐ Manufacturing.

☐ Distribution/Logistics.

☐ ICT services, including app/software developers.

☐ Healthcare.

☐ Automotive and transport.

☐ Financial services and insurance.

☐ Tourism and hospitality.

☐ Public Administrations.

☐ Education.

☐ Construction.

☐ Research.

☐ Media and advertising.

☐ Construction.

☐ Utilities.

☐ Primary sector.

☐ Other:

_____

5. Data users' relationship with your company

☐ Customers.

☐ Suppliers.

☐ Other companies within a business group.

☐ Industry association.

☐ Other (specify):

_____

6. Are you aware of the primary use these users have made of your data?

☐ YES

Can you specify which?

_____

☐ NO

7.a Your organisation's open data initiative can be catalogued as *(check several if they have been done)*:

☐ Monetisation of Data: Obtaining income through the data generated by the organisation.

☐ Data Marketplace: Trusted intermediaries that bring together data providers and data users, providing a secure exchange platform.

☐ Industrial Data Platform: Collaborative and strategic approach to exchanging data between a restricted group of companies.

☐ Open Data: Free data opening to encourage the development of new products and/or services.

☐ B2G (business to government) models: data donation, awards, etc.

☐ Technical Facilitators: They allow the exchange of data through a technical solution. Revenues are obtained from setting up, using and/or maintaining the solution.

7.b. ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. Should you become involved in a data opening initiative. En caso de **involucrarse en una iniciativa de apertura de datos**, what type of initiative do you think is most likely to be the focus of your efforts? (*Check up to two*)

☐ Monetisation of Data: Obtaining income through the data generated by the organisation.

☐ Data Marketplace: Trusted intermediaries that bring together data providers and data users, providing a secure exchange platform.

☐ Industrial Data Platform: Collaborative and strategic approach to exchanging data between a restricted group of companies.

☐ Open Data: Free data opening to encourage the development of new products and/or services.

☐ B2G (business to government) models: data donation, awards, etc.

☐ Technical Facilitators: They allow the exchange of data through a technical solution. Revenues are obtained from setting up, using and/or maintaining the solution.

REASONS FOR OPENING DATA

8. For what main reason did the company decide to carry out the data opening initiative? (*check the three most important ones)*

☐ Improving internal process efficiency.

☐ Diversifying the company's offer of products and/or services through better knowledge of the market and the client.

☐ Solving a specific internal management problem.

☐ Supporting open and collaborative innovation.

☐ Creating new business models.

☐ Pioneering these processes (strategic positioning).

☐ Improving company transparency/accountability.

☐ Social purpose, public common good.

☐ Contributing to the resolution of a local or global challenge.

☐ Other (*specify*):

_____

9. Once the initiative has been implemented, has the result obtained **corresponded to the expectations** raised?

☐ Has it exceeded expectations.

☐ Has it been in line with expectations.

☐ Has it been below expectations.

10. If it has been above or below expectations, indicate the main reason that justifies your response *(open question)*

OBSTACLES TO THE OPENING OF DATA

11.a What were the **main obstacles** you faced? (*check the three most important ones).*

☐ Lack of knowledge regarding the benefits associated with the initiative and/or the value of the data.

☐ Regulatory/legislative risks (e.g. conflict with GDPR or trade secret).

☐ Ignorance of intermediate steps/protocols to be taken into account.

☐ Cybersecurity risks.

☐ Lack of adequate skills to manage initiatives of this type in the company.

☐ Leak of sensitive information to competing companies.

☐ High cost of initiative (in data preparation, data extraction and processing, etc.).

☐ Technical problems/lack of interoperability or standardisation of data.

☐ Lack of appropriate contractual arrangements to ensure proper use of data, ownership, etc.

☐ Lack of interest in the organisation's data by third parties.

☐ Other (*specify*):

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

11.b ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. Why hasn't the company participated in data opening initiatives to date? (*check up to four*).

☐ Lack of knowledge regarding the benefits associated with the initiative and/or the value of the data.

☐ Privacy issues.

☐ Regulatory/legislative risks (e.g. conflict with GDPR or trade secret).

☐ Ignorance of intermediate steps/protocols to be taken into account.

☐ Cybersecurity risks.

☐ Lack of adequate skills to manage initiatives of this type in the company.

☐ Leak of sensitive information to competing companies.

☐ Technical problems/lack of interoperability or standardisation of data.

☐ Lack of appropriate contractual arrangements to ensure proper use of data, ownership, etc.

☐ Lack of interest in the organisation's data by third parties.

☐ Economic costs of the opening of data (e.g. costs of making data available in the desired format, infrastructure costs related to data collection, data curing costs, etc.).

☐ Reputational costs in case of misuse of open data.

☐ Other (specify):

SUCCESS FACTORS FOR PROPER DATA OPENING

12.a. What are the success **factors** in data opening initiatives? (*check the two most important ones).*

☐ Having a clear legal framework regarding the protection of personal data and intellectual property.

☐ Making a good initial investment or having external support funding.

☐ Simplicity and ease of use of the data (access, interoperability, etc.).

☐ Understanding the demand for data (needs, use cases, users, etc.).

☐ Having staff with the necessary qualifications and competences.

☐ Creating collaboration agreements with companies with complementary knowledge (win-win).

☐ Being a pioneer (by type of company, sector, type of project, etc.).

☐ Creating a framework of trust, guaranteeing data security and control.

☐ Exchanging experiences and knowledge in data sharing.

☐ Other (specify):

12.b. ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. What would increase the likelihood of your company undertaking data opening initiatives? (*check up to three*).

☐ Greater certainty of how to open up the data from a contractual point of view.

☐ Greater legal clarity regarding what can be done and what cannot be done with open data.

☐ Greater control over the use of data once it has been opened.

☐ Greater availability of technical qualifications to exploit the value of the data.

☐ Increased levels of interoperability/infrastructure to facilitate the opening of data.

☐ Increased financial compensation for the opening of the data.

☐ Other (specify):

IMPACT OF DATA OPENING AND OUTLOOK

13. What is the degree of quantitative significance of data opening activities in the company's activity? (percentage of data opened monthly on all data generated by the enterprise) *(check only one):*

☐ Less than 1%

☐ Up to 5%

☐ Up to 10%

☐ Up to 25%

☐ Up to 50%

☐ If more than 50%, indicate the percentage

---

14 What is the degree of qualitative importance of data opening activities in the company's activity? *(check only one):*

☐ The opening of data is one of the main activities of my company (example: the company sells sensors for temperature control, and obtains most of the turnover from the valuation of data obtained by these sensors, not from the sale of sensors).

☐ The opening of data is a secondary activity of my company (example: the company produces vacuum cleaners that collect information from surfaces in order to optimise cleaning. The company obtains its highest income from the sale of robots, but also, in a secondary way, from the sale of the collected data to other interested companies, thus obtaining an additional income).

☐ The opening of data is currently a secondary activity, but it is likely to be a major activity in the next five years.

☐ The opening of data is limited to the initiative described here and there is no

intention for it to increase its status in the company's activity in the near future.

☐ The opening of data is limited to the initiative described here, but other initiatives are planned in the near future.

15. ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. To what extent do you consider it likely that your company will start the opening of data in the next five years? (*check only one).*

☐ Highly unlikely.

☐ Unlikely.

☐ Likely.

☐ Highly likely.

☐ I don't know.

RECOMMENDATIONS FOR THE OPENING OF DATA

16. From experience in the world of open data, **where should a company that wants to undertake such a project start?** Any other final **recommendation**?

17. ONLY for COMPANIES THAT HAVE NOT undertaken data opening initiatives. What kind of guidance would you need to be able to start a data opening initiative? *(open question).*

ANNEX VII: DESCRIPTION OF CASES OF OPENING AND SHARING OF DATA

PRIMAFRIO

In the company, the opening of data is a strategic element that still has a relatively secondary and exploratory character, but it is expected to take on a main role in the next five years.

For this reason, the Primafrio Group, a company with more than fifty years of experience specialised in road transport and refrigerated transport, has introduced open data as an element of value in its operations, specifically in the areas of location and positioning (real-time data of parameters that the client may need, such as the registration number of the truck, position, driver, etc.) and related to transactions they carry out (invoicing or accounting on their own account), which are managed through platforms that have been created internally ad hoc.

There is clear evidence of how the customers have used this data, and this is related to the optimisation of order tracking and also the ability to advance billing.

Objective:
This initiative has fundamentally sought to improve the efficiency of the company's internal processes, while at the same time solving a management problem (order tracking) by strategically positioning the organisation. However, the initiative has transcended these limits and become a tool that also seeks to be incorporated into open innovation processes, insofar as they are aimed at the customer and there is a clear willingness, through active listening to their needs, to be able to undertake co-innovation dynamics. Although no formal evaluation has been carried out, those in charge state that it has been in line with the expectations set.
In the future, the company is already set to undertake new developments and technology platforms that will benefit the value chain of the transport sector.

Risks:
In any case, the initiative has not been without its associated difficulties and risks. Among them, the most relevant (which can be extrapolated to a good part of data opening projects) are regulatory/legislative risks (for example, conflict with GDPR or trade secrets), cybersecurity risks and the relatively high cost of the initiative. Related to this, the most critical success factors have to do with the provision of a clear, transparent legal framework to move away from these regulatory/legislative and trustworthy risks to ensure data security and control. Having the right personnel and using user-friendly frameworks are also key.

Recommendations:
The main recommendation for companies that have decided to undertake open data initiatives is to have a well-defined and contractual relationship with the end user to avoid any type of incident, given the importance of the data to be shared. It is also important to have a good structure and support that allows the data of interest to be transferred to the final customer.

Type of collaboration and open data type	• Open/shared data. • Data: location/positioning transactions (customers, suppliers) real-time
Sectors of Primafrio's main clients	• Distribution/Logistics. • ICT services, including app/software developers. • Automotive and transport
Monetisation of the business:	There is no direct monetisation: this is data offered as a value-added service to the customer, as offering better customer service impacts the optimisation of customer loyalty in the medium term and, therefore, ultimately increases the monetisation capacity. • Improving internal process efficiency. • Solving a specific internal management problem. • Pioneering these processes (strategic positioning). • Supporting open and collaborative innovation.
Reason(s) for opening:	• Having a clear legal framework regarding the protection of personal data and intellectual property. • Simplicity and ease of use of the data (access, interoperability, etc.).
Success factors:	• Having staff with the necessary qualifications and competences. • Being a pioneer (by type of company, sector, type of project, etc.). • Creating a framework of trust, guaranteeing data security and control.
Obstacles:	• Regulatory/legislative risks (e.g. conflict with GDPR or trade secret). • Cybersecurity risks. • High cost of initiative (in data preparation, data extraction and processing, etc.).

EUSKALTEL GROUP

The Euskaltel Group's initiative to tackle the field of data sharing arose in 2017, when it began to study the possibilities that could arise from the analysis of the information contained in its CDR (Call Detail Records), which collect information on the connections of mobile terminals to antennas.

The study of the potential of this information concluded that very interesting indicators on the behaviour of the group's customers can be generated for different businesses, and even for the public sector, and they can produce valuable indicators for improving their performance.

This led to the use of a Big Data solution developed within the group: Mileva. Through its components, Mileva provides solutions for Big Data data storage, visualisation and advanced data analysis. The Mileva tool also has the ability to bring together different data sources (both public and private) that allow to enrich this information, adapting it to each of the individual interests that may arise from different businesses or governments.

The information generated from the source data is always properly anonymised and aggregated, and so, complying with this data protection regulation prevents any cases of re-identification of users. This, together with the fact that the Big Data solution has been created within the Big Data team of the group itself, makes it so that the data does not leave the company's systems.

Depending on the case, for the algorithms, the organisation will seek the help of technology partners and external suppliers specialised in each of the areas which will generate a proposal of information that adds value to the business. In this case, those in charge of carrying out the relevant analyses will be technological partners of the group. All the external elements that come into contact with the information will follow the GDPR requirements of compliance in order to guarantee privacy, confidentiality and the correct use of the data of the Euskaltel Group. The agreements will be specific to each project.

The final client will only be able to access final reports under the conditions they have decided on (temporal range, geographical area), but at no time will they have access to the raw data used for the generation of the report they have contracted, so that the raw information of our clients is not exposed at any time.

Type of collaboration:	• B2B • B2G (the government's public administrations)
Sectors of the main clients:	• Transport • Public sector • Retail/commerce
Monetisation of business:	The price of the service depends on the value given to the client. Criteria to determine the value of the data: • Level of data enrichment • Geographical area • Time period of the study
Main obstacles:	• Adapting the organisation's own infrastructure to the particular case. • Convince potential customers of the value that the data to be shared will have for their business.

TELEFÓNICA

TELEFONICA is a company with extensive experience in the field of data management. As an example, and as proof of this, it was recognised in 2018 by the FORRESTER Report (Wave Leaders 2018. Specialized Insights. Service Providers) as a leading provider of insights (information extracted from large amounts of anonymised and aggregated data). The bastion of Telefonica's benchmark position is LUCA, the business unit specialising in Big Data, which offers services that combine business insights, tools, infrastructure and consultancy and analytics adapted to its customers. LUCA offers solutions and services to support companies in their journey towards digital transformation and from its initial stage.

As a data provider, LUCA mainly shares insights generated from data collected from the mobile network, such as crowd movements, location data, etc., guaranteeing customer privacy at all times and strictly complying with existing data protection legislation. In less than three years, LUCA has implemented several hundred projects (around 30% in the Public Sector), addressing very diverse realities and sectors, such as tourism, transport, commerce, financial services, advertising and media and also cross-sectors (such as energy saving or fleet management).

As such, TELEFONICA does not have open datasets, although it is willing to open up data on demand when there is a social purpose for its use. From time to time, it shares data for hackathons.

Of all the initiatives carried out with data, the leading role of "**Big Data for Social Good**" stands out for its social commitment and its open nature. This initiative makes use of TELEFONICA's internal data—together with other external data—to return the value of the data to society, and thus contributing to the Sustainable Development Objectives set by the UN for 2030. Within this initiative, we can highlight projects such as the study of mobility patterns in communities vulnerable to extreme weather episodes; the OPAL (Open Algorithms) project to support the National Planning Department in Colombia in the exploitation of private sector data in a way that has a scalable and economically sustainable social impact, prioritising data security; or the prediction of the spread of diseases such as Dengue and Zyka in Colombia and Brazil through UN Global Pulse. The content of these projects is analysed in greater detail at https://luca-d3.com/es/big-data-social-good-es/index.html.

Type of collaboration:	• B2B (mainly TELEFONICA customers) • B2G (government's public administrations and data donation in case of Big Data for Social Good)
Sectors of LUCA's main clients:	• Advertising and media • Financial services • Retail trade • Tourism • Transport • Government
Cross-sector services	• Optimisation of communications • Energy saving • Fleet management • Mobile identity • Marketing and advertising
Monetisation of business:	Charges for insight according to the value it brings to the client. Offers technological solutions and consulting services for the establishment of a data-driven strategy in the client companies.
Main obstacle:	Make the market understand the true potential of data.
Success factors:	The close relationship with its clients allows the organisation to be aware of their needs and offer them the data sharing solutions that best fit them (modularity). Mobile network data is a differential asset that other companies do not have.

SUEZ

Suez, a multinational company that operates in the water distribution and treatment industry, as well as the waste treatment industry, developed in 2017 an interesting initiative for the opening of data which originated from the hosting of a hackathon, called Hackath20n. More information is available at http://www.hackath2on.es/.

It was an event in which, over a weekend, 20 teams met to tackle, from a collaborative perspective, three water-related challenges, specifically: 1) excellence in customer experience; 2) promotion of the social value of water; and 3) new services for the administration.

The initiative has been useful to the company, since carrying it out helped the company to design new products and services, and it encouraged the launch of other open data initiatives in the immediate future. It was therefore in line with expectations.

The main reasons for hosting the hackathon were to seek the diversification of the company's offer of products and/or services through a better knowledge of the market and of the client, and the creation of new business models, as well as a social purpose, a common public good.

According to the company, the main obstacles in open data initiatives are related to regulatory/legislative risks, as well as with the lack of appropriate contractual agreements that guarantee the adequate use or ownership of the data. In line with the latter, one of the main success factors relates to the provision of an appropriate legal framework for the protection of personal data and intellectual property, as well as issues of access and interoperability (simplicity and ease of use of data) and the understanding of data demand (needs, use cases, users, etc.).

With regard to recommendations, emphasis is placed on the need to set targets when a data opening initiative is carried out in order to ensure its success. It is also very important to provide the necessary resources in the field of security and protection of personal and sensitive data from the initial stage, as not taking this element into account may mean having to abort the initiative without being able to complete it.

Type of collaboration and open data type	• Open Data: Free data opening to encourage the development of new products and/or services. • Data: - Transaction data (customers, suppliers) - Internal operational data (quality, production, ...). • Diversifying the company's offer of products and/or services through better knowledge of the market and the client. • Creating new business models.
Reason for the opening of data	• Social purpose, public common good.
Success factors	• Having a clear legal framework regarding the protection of personal data and intellectual property. • Simplicity and ease of use of the data (access, interoperability, etc.). • Understanding the demand for data (needs, use cases, users, etc.).
Obstacles	• Regulatory/legislative risks (e.g. conflict with GDPR or trade secret). • Lack of appropriate contractual arrangements to ensure proper use of data, ownership, etc.

ITI

The Instituto Tecnológico de Informática (ITI), is a Technological Centre for Research, Development and Innovation in ICT located in the Valencian Community and active since 1994. Although ITI is not an institution focused on data generation, its participation in an H2020 project on open data gives rise to a series of lessons and recommendations that should be considered.

This is the Transforming Transport project (https://data.transformingtransport.eu/), still in force, in which ITI has participated for the creation of a pilot initiative in collaboration with several national and international companies. The project as a whole has involved several data opening pilots in the transport and logistics sectors. From this perspective, the different pilot projects created can be ascribed within the category of industrial data platform, i.e., a collaborative and strategic approach to exchanging data between a restricted group of companies.

The two fundamental reasons that supported the implementation of the project were the contribution to the resolution of a challenge (local or global), which resulted in the creation of the pilot initiatives, and the support of open innovation. As for the main obstacles, they cite regulatory/legislative risks (especially those related to trade secrets), the leak of sensitive information to competing companies, and economic (high cost of the initiative) and technical (lack of interoperability or standardisation of data) issues. In fact, one of the reasons why the results of the project have not met the set expectations has to do with the fact that most of the participating companies have not really opened data, as the data was already public before the implementation of the project. Only in very few cases, and by means of very concrete agreements and the previous signing of confidentiality clauses, was there a process of data opening.

Consistent with the economic and technical obstacles, success factors cited are the right initial investment (or having external support funding) and the simplicity and ease of use of the data, in terms, for example, of access and interoperability. In addition, there is the creation of a framework of trust to ensure data security (which again highlights the relevance of regulatory risks), as well as having personnel with the necessary skills and competences.

In a forward-looking way, ITI proposes to carry out other data opening initiatives in the future, and in terms of the main recommendations, it stresses the need to assess the cost of maintaining an open data infrastructure—which is costly, especially if the data is consumed constantly and by multiple customers. Another interesting challenge is the necessary anonymisation of sensitive business data in a way that is compatible with preserving the value of the data.

Type of collaboration and open data type	• Industrial data platform • Data: - Varies depending on the pilot project (location, transaction data...), but fundamentally varies depending on the description of the data sources used and how to access them. There is private data of the categories mentioned above. Most of the "open" data were already public before.
Sectors of application	• Distribution/Logistics • Transport

Reason(s) for opening	• Contributing to the resolution of a local or global challenge . • Supporting open and collaborative innovation.
Success factors	• Making a good initial investment or having external support funding. • Simplicity and ease of use of the data (access, interoperability, etc.). • Having staff with the necessary qualifications and competences. • Creating a framework of trust, guaranteeing data security and control.
Obstacles	• Regulatory/legislative risks (e.g. conflict with GDPR or trade secret). • High cost of initiative (in data preparation, data extraction and processing, etc.). • Leak of sensitive information to competing companies. • Technical problems/lack of interoperability or standardisation of data.

EDP

EDP's experience in opening data stems from its deep understanding of the role they play as a tool of value and promotion of innovation for the company. Its experience in open data is fundamentally included in the proposal to launch challenges, in force since 2018. Their approach relies on the opening of datasets—which can be accessed with just a click (through the link: opendata.edp.com)—and on the innovative management of these datasets by people and teams through online "hackathons". These are, therefore, open innovation initiatives that use datasets as raw material to address challenges, which are awarded with rewards (which can be monetary or testimonial). By way of example, two of these challenges are the following:

- Detection of failures in wind turbines. It is a challenge with a reward of $5,000. It involved determining, on the basis of Big Data models, when such equipment is most likely to fail and, consequently, allows turbine maintenance tasks to be optimised.
- Prediction of solar energy. In this case, the reward was testimonial (the recognition of working to make a more sustainable world). Based on EDP's SunLab project (a laboratory that aims to measure the performance of different photovoltaic technologies installed in different configurations and environments), its aim was to construct an algorithm capable of predicting solar energy production for the first seven days of 2018 from a dataset of meteorological variables.

Also following this background of open innovation, there is a second modality in which there is no challenge to be tackled as such, but which consists of the promotion of all those works (theses, publications, creation of new algorithms, etc.) that have been produced thanks to the opening of the datasets, thus promoting collaboration with academic institutions, mainly Universities. An example is the Doctoral Thesis on angular distribution of solar radiation (University of Sciences of Lisbon), which uses datasets from the SunLab project at its headquarters in Faro (Portugal).

As a summary, EDP's main reasons for opening data are to support open innovation, to contribute to the resolution of a local or global challenge and to improve the efficiency of internal processes.

For its part, one of the main obstacles faced by the organisation has to do with the lack of knowledge regarding the true potential (both internal and external) of the data and the benefits that these initiatives can generate. In fact, one of the main success factors is involving agents of all kinds in the use and creation of value related to the data, thus generating very dynamic and fruitful co-creation mechanisms. Other relevant obstacles are regulatory risks and the lack of adequate skills to manage these types of initiatives in the company.

Type of collaboration and open data type:	• Open data: Free data opening to encourage the development of new products and/or services. • Internal operational data (quality, production...). In short, it is also intended to create real-time data derived from some of its assets.

Main users:	• The data has been used for R&D purposes. Therefore, it has been used from a triple perspective:
	  • ICT services, including app/software developers
	  • Education
	  • Research

Reason(s) for opening:	• Improving internal process efficiency. • Supporting open and collaborative innovation. • Contributing to the resolution of a global or local challenge.
Main obstacle(s):	• Lack of knowledge regarding the benefits associated with the initiative and/or the value of the data. • Regulatory/legislative risks (e.g. conflict with GDPR or trade secret). • Lack of adequate skills to manage initiatives of this type in the company.
Success factors:	• Understanding the demand for data (needs, use cases, users, etc.). • Exchanging experiences and knowledge in data sharing.

ARCELOR MITTAL

Arcelor Mittal, together with TECNALIA, organised on 24 and 25 September 2018 a datathon aimed at finding "useful and creative" solutions based on data analysis to solve a real problem with their steelworks. More specifically, the datathon had two objectives:

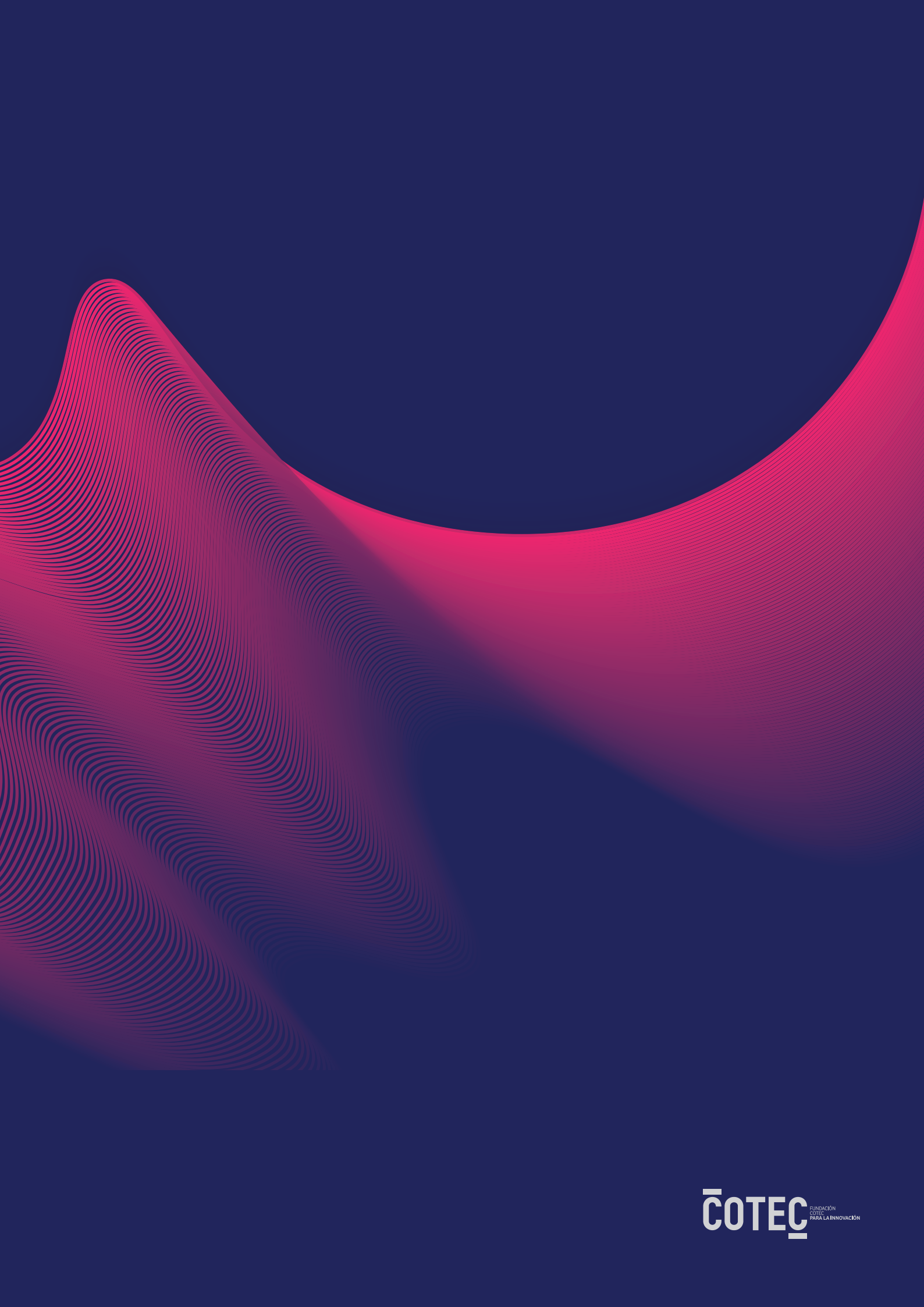
- Optimise the ferrous performance of each type of scrap by referencing real data from the melting process of the steel mill and taking into account a set of factors that affect performance.

- Predict the final copper content in liquid steel at a point in time for each type of scrap mixed.

The initiative had more than 20 participants, including researchers from Tecnalia, the University of the Basque Country and BCAM (Basque Centre for Applied Mathematics). The proposals were analysed by a jury made up of researchers from Arcelor Mittal and TECNALIA, who assessed aspects such as technical quality, novelty and contribution to the "state of the art", industrial applicability, creativity and innovation. The best proposal was rewarded with the development of a proof of concept for a possible subsequent implementation.

ArcelorMittal deemed this open innovation initiative as very positive, since it allowed the organisation to tackle a wide range of innovative solutions that would have an impact on the improvement of its processes.

Type of collaboration and open data type	• Data from the use of the scrap blocks generated in situ (up to 16), from the melting process of the steelworks and from the results: from 07/2017 to 04/2018, 4,000 consecutive samples.
Monetisation of business	• Distribution/Logistics • ICT services, including app/software developers • Automotive and transport
Monetisation of business **Reason(s) for opening**	• No direct monetisation. • Exploring new analytical solutions for the interrelationship of the production process. • Supporting open and collaborative innovation.


COTEC
FUNDACIÓN
COTEC
PARA LA INNOVACIÓN