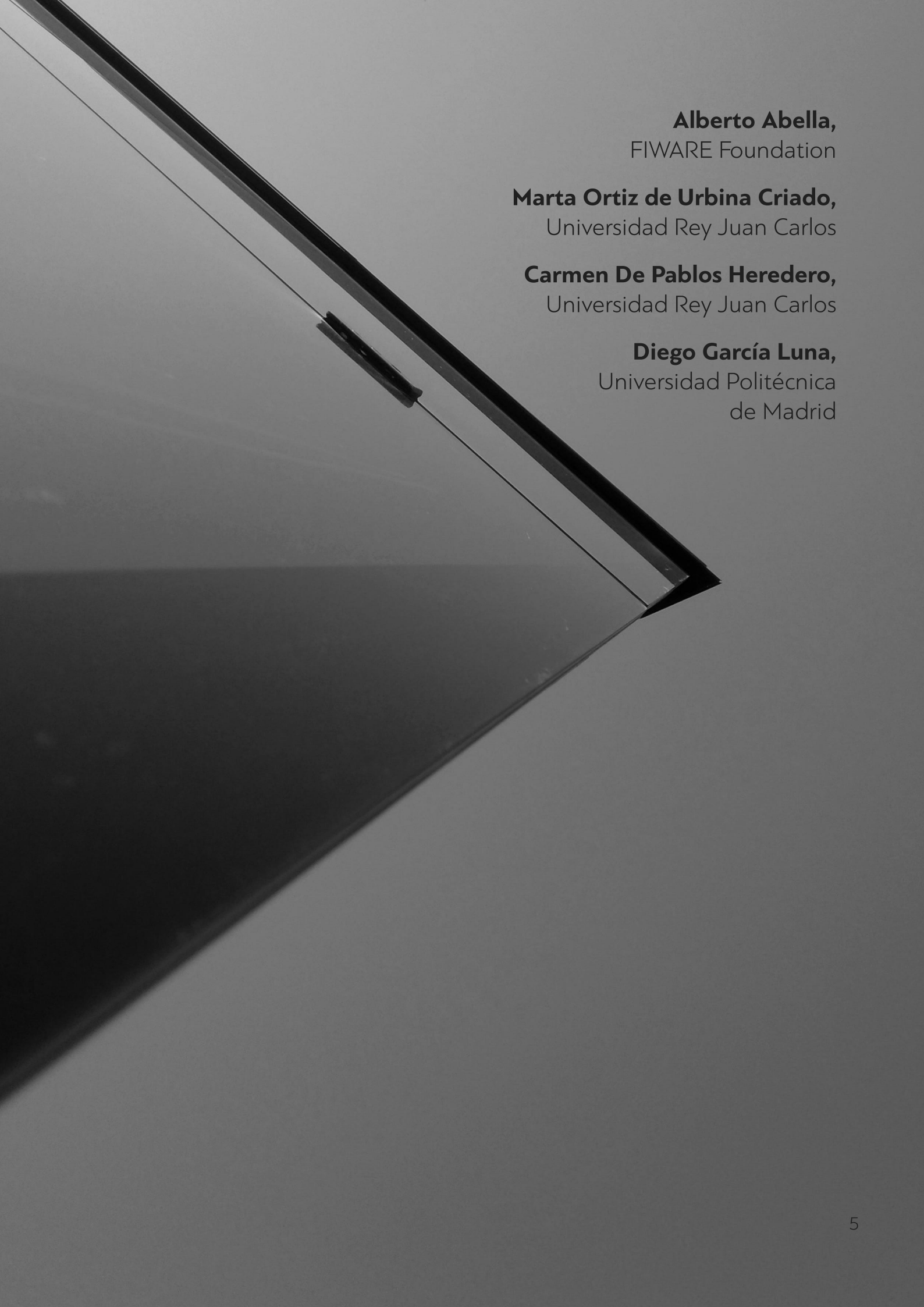
REUSING OPEN DATA IN SPAIN III

October 2021



Alberto Abella,
FIWARE Foundation

Marta Ortiz de Urbina Criado,
Universidad Rey Juan Carlos

Carmen De Pablos Heredero,
Universidad Rey Juan Carlos

Diego García Luna,
Universidad Politécnica
de Madrid

Reusing open data in Spain III. October 2021
Alberto Abella, Marta Ortiz de Urbina Criado,
Carmen De Pablos Heredero, Diego García Luna

Avda. de Valdenigrales, s/n,
28223 Pozuelo de Alarcón (Madrid)
Tel. 91 452 41 00
www.esic.edu/editorial

ISBN: 978-84-18944-60-4
Legal Deposit: M-7064-2022

Design and Layout: Gerardo Domínguez
Printed by: Gráficas Dehon

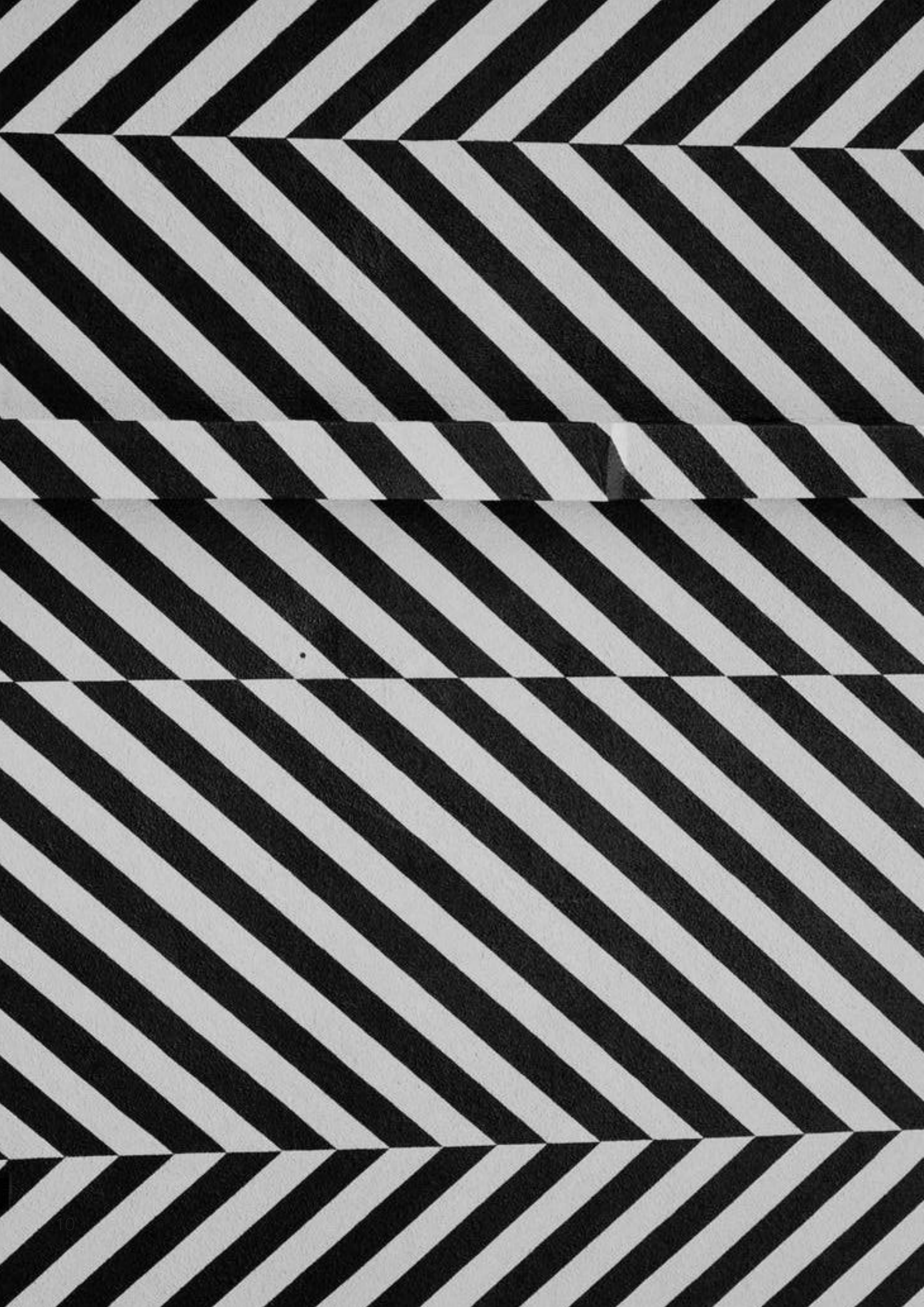
A book by
ESIC Editorial

Printed in Spain

This book has been printed with organic ink and sustainable paper.

Contents

Index of Figures



Index of tables

1

Executive Summary

Today's society is based on the mass processing of data and on the developing digitisation of cities (smart cities) and the objects around us (Internet of things). Public institutions and businesses are major providers of data, many of which can be reused to create value in the economic, social, and environmental spheres.

The reuse of data makes it possible to create value for society because of the range of new digital products and services and the stimulation of economic and business activity (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019). Among other things, the reuse of data has allowed the setting up of new businesses that, with few resources of their own, have implemented business models based on the creation of products and services enriched with value-added information (Marcos-Martín & Soriano-Maldonado, 2011; Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2014).

This report contains the third study of open data reuse with the aim of presenting the state of the question in 2021 and the progress made since the first report, drawn up in 2017. This will allow updated guidelines and recommendations to be made, to help generate business. To this end, the open data portals in Spain in 2021 were identified, and a sample of the datasets available and the services based on them was analysed. A questionnaire was also sent to those responsible for the portals to analyse some of the features of open data and their potential for reuse. Specifically, a diagnosis was conducted of their knowledge of the reuse of their data, of the type of innovation that can be developed with them, of activities to promote data use, of the services generated and of the value created in connection with the reuse of data. The latest version of the MELODA 5 metric was applied to assess the degree of reusability of the open data published on Spanish open data portals. All these diagnostics allowed an analysis of strengths, weaknesses, opportunities, and threats, on the basis of which some reflections have been included to help construct future public policy on data management.

The study carried out enables us to highlight the following facts and reflections about the data reuse ecosystem in Spain:

- STATISTICAL INPUT Statistical data sources have largely fed open dataset portals, being mainly responsible for the 45% increase in data generated since 2019, when the number of portals has only risen by 3.4%.

- AVERAGE DATA PUBLICATION FEATURES STEADY There has been slight percentage decrease in data publication on high-performance portals, from 74% in 2019 to 72% in 2021.

- INTERNAL CONSUMPTION At this moment in time, according to those responsible for the data portals, the biggest consumers of the data published are the public administrations themselves (64% often or always reuse data), as well as being the biggest generators of services (41%).

- LACK OF STANDARDISATION AND MODELS IN DATA PUBLISHED 80% of the data published do not include information on their structure or use standardised data models.

- UNSUITABLE TOOLS FOR PUBLISHING DATA The use of tools not intended specifically for managing an open data portal (63%) is limiting the development of services, for example because of the lack of automation of access to them.

- INFREQUENT UPDATING OF DATA 92% of the open data published are updated less that once a month, while the percentage of data published in real time is less than 0.3%.

- NON-GEOLOCATED DATA It is observed that 50% of the data published still do not contain any kind of geographical information.

- LACK OF MAINTENANCE OF OPEN DATA SERVICES 17% of the open data-based services listed in the portals are inactive or no longer exist.

- RANKING OF REPUTATION OPEN DATA PUBLISHERS The new version of MELODA (MELODA 5) requires assessment of the reputation of open data publishers. This ranking is included for the portals in which there was sampling of datasets, assessed by those responsible for the open data portals (see Annexe 12.3).

2

The 2021 report is available in two languages (Spanish and English), allowing wider circulation of the report, especially at European level.

Introduction to the report

2.1 Introduction to ESIC

ESIC aims to contribute to the transformation of students, so that they develop successfully in their professional life in a responsible way and anchored in ethical values by acting as agents of change in their organizations and in their social environment.

In this way, ESIC carries out activities in a responsible way to contribute to society, so through its Research Department and its publisher it is involved in the publication of this third edition of the report The reuse of data, that reflects how our society is being transformed with digitization and, above all, with the processing of data.

Our areas of training include official university degrees specialized in Business Administration and Management, Strategy and Organization, Marketing, Communication, Advertising and Public Relations and Digital Economy; the largest postgraduate training offering with master's degrees in Management, Marketing, Technology, International Trade and Business, Communication and Advertising, Human Resources, Sales and Commerce, Logistics and Finance; and, especially, the ESIC MBAs, which are among the best internationally and have AMBA accreditation, only within reach of the most prestigious business schools.

The Corporate Education unit develops training measures for companies, adapted to the needs of management and commercial teams, to facilitate the recycling of knowledge and learning of new areas of knowledge.

In line with the digitization of society, our online training highlights: masters, higher programs and specialized courses that can be taken from anywhere in the world with greater flexibility of time. In addition, at ESIC we teach the higher degree training cycles in the professional families of Commerce and Marketing, Administration and Management, Computing and Communications and Image and Sound.

In addition, beyond training, ESIC contributes to society by disseminating the knowledge of its community members. To this end, ESIC has its own publishing house, ESIC Editorial, which is the projection of ESIC in the world of publications, providing research and research on economic, business and marketing issues.

Its editorial fund, in constant renewal, has more than 800 living titles, whose authors are mainly university professors and professionals, experts with deep knowledge of our business environment. It has been editing professional

books on marketing, advertising, sales, human resources, managerial skills, finance, strategy, economics, mathematics and statistics, which a mix of rigor, experience, research and amenity. ESIC also has four scientific journals: ESIC Market, Economics and Business Journal; aDResearch-ESIC, International Journal of Communication Research; Spanish Journal of Marketing-ESIC, specialized in marketing; and EDEIJ, ESIC Digital Economy and Innovation Journal, in the field of the digital economy.

And finally, through the Professional Development Unit (UDP), ESIC opens the doors to the world to students, who can enjoy the job market for life, and the Entrepreneurship Department launches business ideas and provides advice. Definitively, ESIC and society together; committed to formation and ethics.

2.2 Introduction to COTEC

This report is the result of the extension and updating of one of the 24 projects selected —from a very wide set of proposals, totalling 1,839— in the second edition of the Open Innovation Programme launched by the Cotec foundation in 2017. Cotec is satisfied that the project has continued, with updates and improvements, on a biannual basis. The scale open data is achieving in today's world, in a context of digitisation of society, has been speeded up by the Covid-19 pandemic. It is therefore crucial to understand what its added value is and what policies are being followed by the authorities to drive it, as good management of these policies is critical in order to maximise their impact.

When we took on this project we did so because it fitted perfectly into Cotec's strategic direction, aiming to drive two transitions in Spain —the digital transition and the intangible transition— as well as the circular transition. Spain is part of the group of countries with the lowest investment in the assets characterising the knowledge economy, intangibles and, even though in the last two decades it has showed the best dynamic performance of all European economies, even narrowing the gap up to the 2008 crisis, it still has a long way to go to catch up. It is no coincidence that Spain, as an economy that lags in investment in intangibles, is also among the group of European countries with the most limited levels of productivity.

In this context, where the focus in increasingly on the value of intangible assets, data, as one of these, is of increasing value so that reusing it is more relevant and attractive than ever. Hence the great importance of projects like this, which not only identifies the limitations of the Spanish authorities' open data model as regards the reuse of data, showing up the fact that 80% of the data published follows no international standard (limiting the possibilities of putting them together with data from other sources), but also proposes

solutions to overcome these limitations (in the above case, taking agile standardisation initiatives to document data models).

2.3 Introduction to desideDatum

desideDatum Data Company, S.L.[1], generally referred to as "desideDatum", is a Spanish firm specialising in consulting and implementation of data-related projects, both for the authorities and for private organisations. Specifically, desideDatum offers services related to:

- Data governance.
- Openness and sharing of data.
- Transparency and accountability.
- Public innovation.

Regarding the area of making open data available it should be stressed that desideDatum is currently the company with the most references in Spain and the only one in the world that has collaboration agreements with the three main technology solutions for publishing open data: Socrata, OpenDataSoft and CKAN. Many public institutions have relied on desideDatum's services to make their data open (consulting and/or implementation), for example currently —October 2021— desideDatum has more the 20 projects under way to publish open data, highlights among which include:

- The regional government of Aragón
- The regional government of Catalonia
- The regional government of the Balearic Islands
- The regional government of Navarra
- The regional government of La Rioja
- The regional government of Valencia
- SABA Aparcamientos, S.A.

Moreover, desideDatum has always supported initiatives favouring the opening up and sharing of data, as in the case of this study, and also desideDatum employees are prominent activists in this area, often giving talks on the subject.

For desideDatum, opening up data means boosting the significant value of the data in question, an important way of empowering society and reactivating the economy thanks to data.

1 For further information you are recommended to consult https://www.desidedatum.com/


DIGITAL

3 Introduction

3.1 The Importance of Open Data

Open data are data that can be used, reused, and redistributed freely by anybody, and which are subject at most to the requirement that they be attributed and shared in the same form as they appear, as laid down by the Open Knowledge Foundation (2021). They are, therefore, strategic resources for governments, businesses, and citizens (Walter, Lovett, Maher, Williamson, Prehn, Bodkin - Andrews & Lee, 2021).

Having open data offers the chance to be able to reuse them to create more innovative, competitive products and services for end users. Public and private organisations, as well as citizens, must have the chance to access data so that they can be reused to create value. With reference to open government, Ferrer Sapena, Calabuig, Sánchez Pérez & Vidal Cabo (2020) point out that information must be well-structured "so that public efforts make possible, through appropriate reuse, the improvement of services to the public and at the same time act as a source of wealth creation and public sector modernisation" (p. 6).

Cadena (2019) points out that on an organisation's website there must be information that can be used by businesses and citizens. He admits that the lack of quality in the information published is a barrier to the process of mass reuse. The data created, stored, processed, exchanged, shared, aggregated and reused are of quality as long as they are useful, i.e., the expected value can be obtained from them. Furthermore, to guarantee the quality of the data in an organisation requires: 1) taking into account users' needs; 2) establishing a clear data generation process; 3) identifying the data life cycle; 4) appointing a person to manage the data and 5) assessing the quality of the data product.

Zuiderwijk, Pirannejad & Susha (2021) point out, referring to best open governance practices, that a critical factor characterising high levels of progress is providing facilities for citizens, businesses, and entrepreneurs to participate as key interested parties in open data projects.

The Covid-19 crisis has altered the public's routines. By imposing quarantine, avoiding face-to-face meetings, and closing the public offices responsible for receiving and processing documents, the pandemic revealed how problematic it is not to simplify and digitise certain interactions between citizens and governments (Cetina, 2021). For instance, the pandemic has shown the importance of having available data updated in real time to make decisions on public health. Cetina (2021) stresses that, to reduce the health impact of the crisis, data are essential in tracing and predicting the spread of Covid-19,

in guiding scientific research, in making diagnoses and in seeking treatments. He further stresses that to analyse the economic impact of the pandemic, data are crucial to direct economic assistance to those who need it most, to administer emergency public procurement and, gradually, to understand the relationship between public health, quarantine measures and economic activity.

Moreover, as Marc Garriga Portola of desideDatum mentions, many citizens and even journalists have discovered open data thanks to Covid. The use of open data—in the most mature portals—has significantly increased thanks to Covid though, after the initial spurt, considerable use of open data continues without all these being related to the pandemic. Examples are the portals of the regional governments of Catalonia and Castilla y León, which are two of the leaders in opening up data about Covid.

3.2 Open Data in Europe

The Digital Economy and Society Index (DESI) is considered an indicator of the digital transformation in Europe and reflects its progress in European countries in terms of productivity. This index is made up of the following dimensions: connectivity, human capital, Internet use, integration of digital technology and digital public services (European Commission, 2020a). The evolution of these dimensions depends to a great extent on the quality of the open data available to be reused (Hrustek, Furjan & Pihir, 2021). It is no coincidence that the countries that score highest for the maturity of open data are also the best positioned in this index, as in the case of Ireland, Netherlands Spain and Malta (European Commission, 2020b).

The European data portal (https://data.europa.eu) seeks to improve access to open data from any European country and to foster the practice of publishing data at regional, national, and international level. Every year since 2015, comparative data have been presented on the evolution and use of open data in European countries. The results of these comparisons are shown in the open data dashboard, which represents a highly practical tool to compare the level of maturity of open data in the different European Union countries. Cecconi & Radu (2018) compare best practices in Europe in open data management and highlight the cases of Ireland, Spain, and France. The European Commission (European Commission 2020b) updated this report on Open Data Maturity, observing that Ireland, Spain, and France maintain their dominant position in open data management, while the group is joined by Denmark, Estonia, Poland and Austria. The methodology used by Cecconi & Radu (2018) has been improved over time to assess the level of maturity of

open data in different countries, taking into consideration four dimensions, as explained below:

1. Open Data Policy:
This dimension centres on the existence of specific policies and strategies to foster open data at national level. The dimension assesses the governance structures that allow participation by public and private stakeholders, as well as the implementation of measures to allow open data initiatives at national, regional and local level.

2. The Impact of Open Data:
The second dimension assesses the level of understanding of open data among public sector bodies, the activities that go on to supervise reuse and their impact. This centres on five areas:

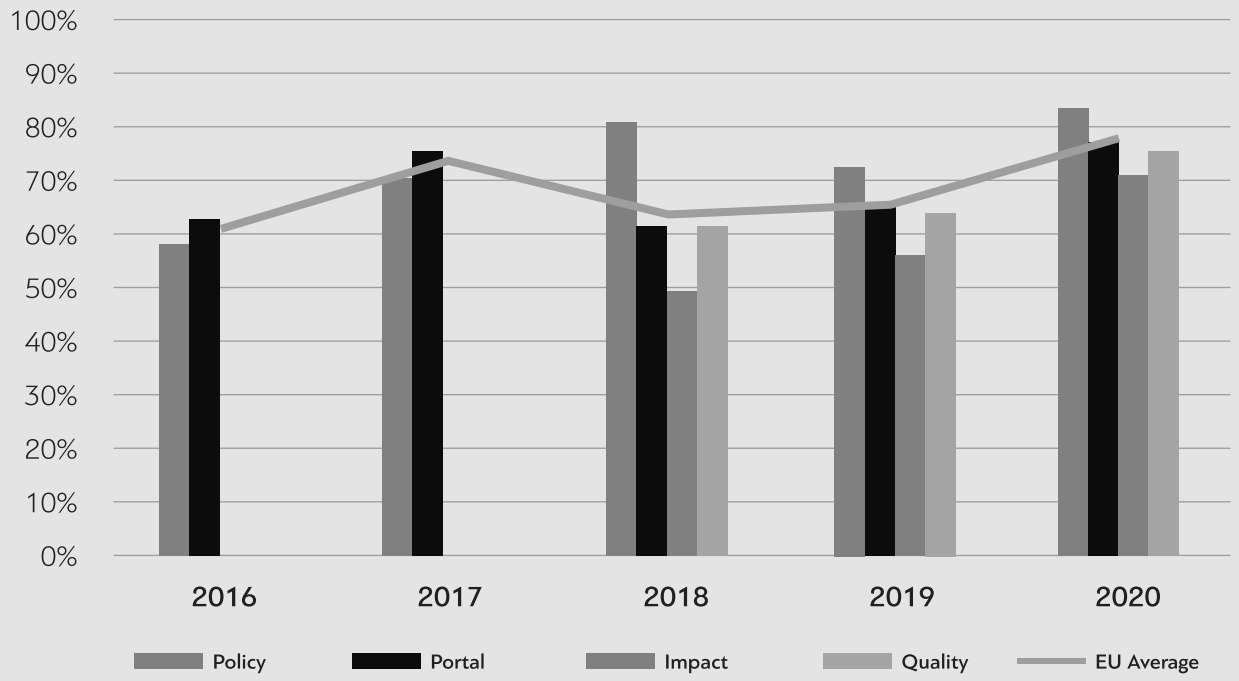
- Strategy for measuring and monitoring impact
- Political impact
- Social impact
- Environmental impact
- Economic impact

3. Open Data Portal:
The third dimension centres on advanced features of the portal to allow users to access the open data and foster interaction between publishers and reusers, for example through discussion groups. This dimension also assesses the use portal managers make of analytical tools to better understand users' needs and behaviour and to update the online portals in the light of these analyses. It also assesses the coverage of open data in different domains, as well as on-site rating to assure the sustainability of the open data portal.

4. Quality of open data
The fourth dimension centres on the measures taken by the portal managers to assure a systematic feed by the data sources in the country. It monitors the level of compliance with the DCAT-AP metadata standard to determine how far the data and metadata are updated and complete. This dimension offers incentives to data managers and decision-makers to adopt best practices, including linked data elements such as uniform resource identifiers (URIs).

According to the latest open data maturity report (European Commission, 2020b), it can be stated that (Figure 1):

- European countries are increasingly mature in all these dimensions. Countries' maturity ratings are concentrated more on the upper end of the spectrum of results.
- The average open data maturity rating for the 27 EU countries is 78%, an increase of 12 percentage points in comparison with 2019.
- Denmark comes top in the assessment for the first time this year.



Source: European Commission (2020b, p. 6)

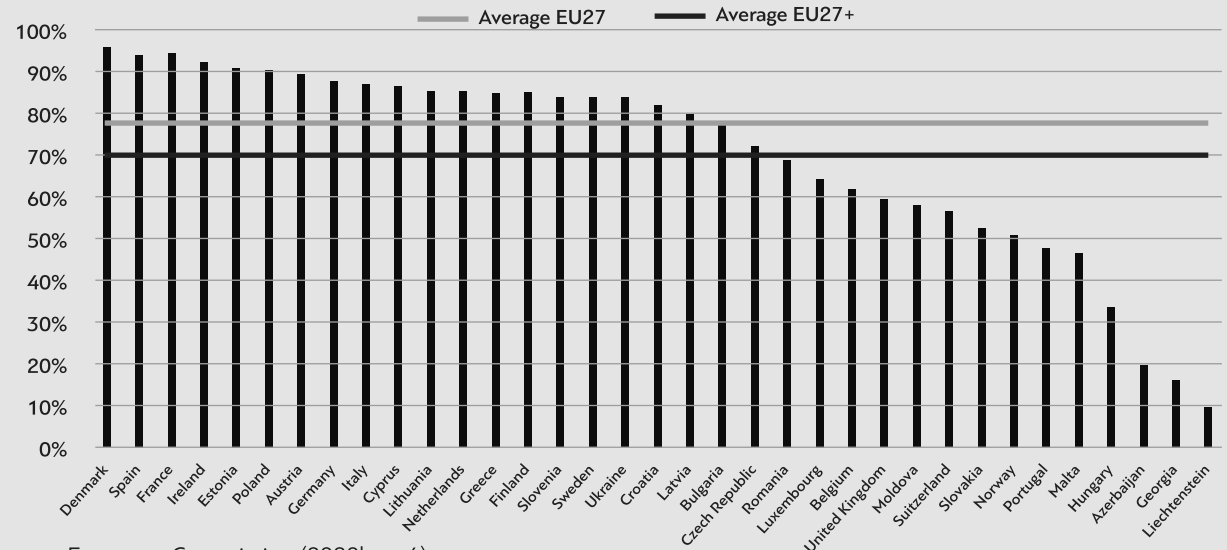
Figure 1: The evolution of the dimensions of open data maturity in Europe.

Meanwhile, Ireland, Spain and France continue their trend from last year. The group is joined by Estonia, Poland and Austria.

- Many of the member states (18) of the EU 27 are above the average for the EU 27.

Figure 2 below shows countries' overall open data maturity scores according to the 2020 Open Data Maturity Report (European Commission, 2020b).

Figure 2: General ratings for open data maturity in the 2020 assessment.



Source: European Commission (2020b, p. 6)

Figure 3 shows the clusters resulting from this maturity index. The countries are shown from lower to higher in the index, in four categories: established, fast, followers and beginners. This results in the following conclusions:

- European countries show a considerable rise in their maturity levels, resulting in a higher concentration of countries at the top end of the spectrum of results.
- The cluster that establishes trends consists of seven countries. Ireland, Spain and France keep their leading positions and the group is expanded with Denmark, Estonia, Poland and Austria.
- The fast group consists of 13 countries and is the largest of the four groups.

Figure 3: Groups of countries in the open data maturity index (2020).

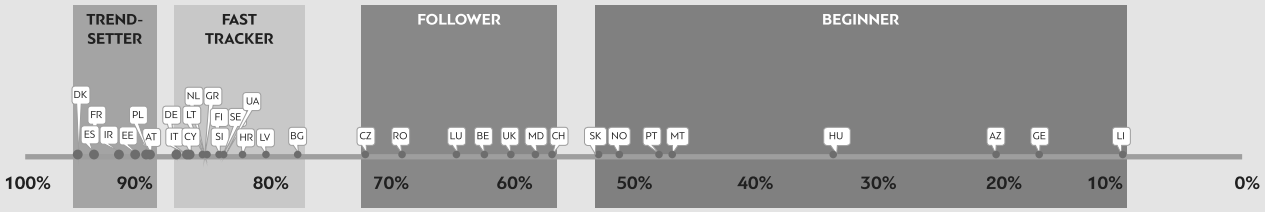
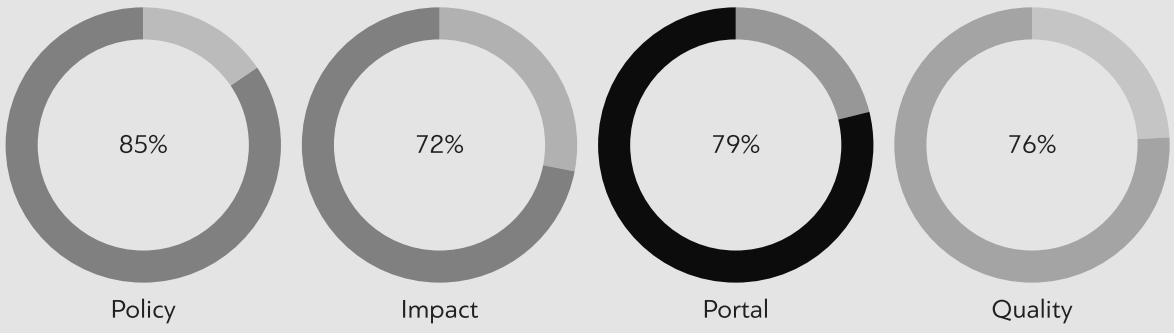


Figure 4 shows the average maturity level of the EU 27 in each of the four dimensions, giving the following conclusions:

- Policy is the most mature open data dimension, with an average score of 85%.
- National portals are increasingly advanced, with an average maturity rating of 79%.
- The average score of 76% for quality shows that there have been big improvements in terms of guaranteeing the quality of data and metadata.
- Impact is the least mature open data dimension, with an average score of 72%. This highlights the need for a strategic focus to supervise and measure open data reuse and the impact this has.

Figure 4: Average maturity level for the EU 27 in each of the four dimensions.
Source: European Commission (2020b, p. 7)



In addition, the European report (European Commission, 2020b) highlights other trends that represent good opportunities to develop and improve open data management in the European context:

The Covid-19 pandemic has shown up the real need to have data available. This year European countries show a considerable rise in their maturity levels, indicating that Europe is heading in the right direction as regards the objectives set for open data and making them available for citizens to reuse. The need to cope with the emergency led many countries to start publishing data and to pursue initiatives to make data easier to access and understand (European Commission, 2020b).

As European countries' open data policies mature, their attention has shifted from the amount of data available to also guaranteeing their quality. Moreover, quality is not considered in isolation, but as an interoperability factor in terms of capacity to collaborate within countries and across borders by facilitating the pooling of data between computer systems (European Commission, 2020b).

According to the Open Data Barometer (Web Foundation, 2018), data interoperability between governments and countries would allow a more coordinated response and more accurate decision-making globally, nationally, and locally. For instance, Alamo Reina, Mammarella & Abella (2020) point out that access to and use of open data in the struggle against Covid-19 was hindered by having different data formats, changing, non-uniform measurement criteria and constant changes in the structure and hosting of databases. Alamo et al. (2020) also stress that an area where this heterogeneity is obvious is in calculating the mortality rate from the virus, which is subject to a multiplicity of criteria for being recorded and made available to the public.

As Worthy (2015) states, reforms do not just depend on authorities' gross data. It is hoped that a mixture of private and voluntary data intermediaries will innovate to create easy-to-use portals and applications offering interoperability to allow users to combine different data and tools in new ways.

Another relevant aspect is measuring impact. Generating a positive impact on society and the economy through the publication of open data has always been the ultimate goal of the broad multinational efforts throughout Europe. Measuring impact is a complex task and there is still no shared understanding of how to do it better. Many European countries are successfully engaging in activities to understand and capture how far open data are reused and how value is created, through interaction with user communities. The European Commission plans to take this as a basis for developing a shared impact framework in the coming years.

The 16th Analytical Report (European Commission, 2020c) analyses some relevant practices in leading countries in open data management, specifically

the cases of Cyprus, France and Ireland. One of the common themes highlighted in this report concerns open data policy in each of these three countries. They have specific legislation on open data and a clear strategy. To make strategy more tangible and viable, an action plan helps these countries to closely monitor the progress of the different measures to be taken and to make the parties responsible accountable for them.

There is also a trend for national portals to become a one-stop shop for open data. This includes not just public sector data, but also data published by the private sector of value to the public. Eventually, all necessary, useful information regarding open data, their publication, reuse and so on should be available from this common platform.

The features of national open data portals go beyond simply allowing users to find available datasets. There is a common focus on the interaction between data publishers and reusers through discussion forums, specific data feedback systems and rating systems. Through examples, portals offer valuable cases of reuse of open data.

The goal of national open data teams is to increase impact by not only fostering but also supervising and analysing the reuse of open data. They examine the topics currently appearing in the news and their aim is to ensure that the pertinent datasets are published and given special attention in the portal. End users can also contact the portal privately to request datasets.

The best practices in each of the open data maturity dimensions discussed in this report may be highly beneficial for all the countries in Europe to seek inspiration, learn and apply in order to improve their own practices.

3.3 Open Data in Spain

The latest edition of the Infomediary Sector Report by ASEDIE includes the results of an analysis of 529 businesses that have data-based business models (ASEDIE, 2021). The Spanish National Telecommunications and Information Society Observatory (ONTSI) has published reports describing the infomediary sector from 2011 to the present (Red.es 2021). The Aporta initiative (https://datos.gob.es/es/blog-tags/iniciativa-aporta) also supplies information about the current state of the infomediary sector and the products and services on offer, which are essential factors in reuse, as well as analysis of the features of the range of information on offer, from considering the data as raw materials to turning them into useful products and services (Aporta, 2021).

ASEDIE (2021) makes clear that digitisation, progress in artificial intelligence and the Internet of things are realities that are developing ever faster and causing a transformation in the economic system. The data, their management and analysis, have become a necessary part of business progress, making the infomediary sector one of the most influential factors in our economy.

This same report highlights how the year 2020 saw progress and new challenges in digitisation and artificial intelligence, as a result of the crisis caused by Covid-19, making clear the importance of progress in these areas to growth in the economy and society (ASEDIE, 2021).

It also recognises that there is a lack of consolidated "culture of data", not only in society but at the different levels of government. The lack of digitisation and poor quality of data are some of the barriers to access and use of information (ASEDIE, 2021).

Infomediary companies are more active in some zones than others in Spain. The Community of Madrid in particular stands out, with 38% of the infomediary businesses based in this autonomous community. A second group of three autonomous communities (Catalonia, Andalusia and Valencia), make up between 9 and 13% each, accounting for 32% of infomediary companies. The other 12 autonomous communities and the Autonomous City of Melilla account for 30% of infomediary businesses (ASEDIE, 2021).

Out of all the infomediary businesses, 40% (281) were set up in the last 10 years. The sub-sector in which most businesses have been set up over this period was the economic and financial one, with 58 firms. In proportional terms, the sub-sector in which the most businesses have been set up is meteorology, at 60%. The trend in the infomediary sector has been positive, with growth 6.4% above that in gross domestic product, which was 3.4% (ASEDIE, 2021).

96% of respondents state that they reuse data to create their value-added products or services. 70% of respondents state that they also reuse data for internal purposes, while 30% say their use is purely external. 61% of the businesses interviewed use artificial intelligence to create value-added products or artificial intelligence-based products. 100% of the businesses interviewed state that opening data in a structured way boosts the value of the database itself (ASEDIE, 2021).

Intermediary businesses agree that the standardisation of information is important and necessary at all levels (central government, the 17 autonomous communities, 52 provinces, local government, etc.). Homogenisation and quality have become a standard. Any initiative to allow access to regulated data that is as standardised as possible is rated positively. Unlike in previous years, the businesses interviewed state that an increase in open databases and access to these conditions positive development in the sector, and over 60% see it as a priority (ASEDIE, 2021).

The two main barriers encountered by businesses to reusing information are inadequate updating of the databases available and the fact that the information is only available in some autonomous regions or local authorities (ASEDIE, 2021).

In this respect, with a view to ameliorating these two issues, we have detected a significant increase in federated data in comparison to our previous report (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019). The federation of data allows a set of open data to be redistributed from its original domain either entirely or in part. This is a way of including external data sources in domains that are regularly visited in a more active way.

Regulations are also considered a significant obstacle to the reuse of data. Not really knowing how to prioritise between co-existing regulations, as well as confusion between regulations when accessing the information sometimes leads to restrictions on access or even literal bans on its reuse. An example of this confusion is that caused by data protection regulations, which are being interpreted wrongly, applying them not only to natural persons but also to legal entities, so blocking access to this information and prejudicing business transparency (ASEDIE, 2021).

In 2021 over 60% of respondents consider that, after three years in force, data protection regulations have a negative impact on their business, an opinion that is more widely held than in previous years (57% in 2020 and 53% in 2019) (ASEDIE, 2021).

88.24% of respondents felt the effects of Covid-19 over the year 2020, resulting in a boost to digitisation processes, increased use of open data portals and increased demand and opening up of new datasets, many of them related to Covid-19. It is important to point out that, in prioritising datasets, the public interest at any given time is of considerable importance. Thus, today there is a lot of interest in databases related to the Covid-19 pandemic (healthcare, economics, employment, social assistance, vaccination, etc.) (ASEDIE, 2021).

The main barrier to opening up databases in 12 of the 17 autonomous communities is the lack of training of public employees in the benefits of open data: “A culture of open data needs to be encouraged.” The lack of a culture of open data means there is sometimes “fear of sharing and opening up data; departments see data as their property and are unwilling to share them”. The human resources situation in government is seen as a barrier as resources cannot be permanently assigned to coordinating the opening up of data and as there is a lack of staff with the training necessary to do this (ASEDIE, 2021).

Lack of digitisation is the next barrier mentioned for 10 of the 17 autonomous communities, where it was made clear that sometimes information cannot be extracted from databases because it is not digitised. For 9 of the 17 autonomous communities the next barrier faced is the poor quality of data, followed by technical barriers, mentioned for 7 out of 17. It is also stressed that the complexity of databases is often the excuse for not opening them up (ASEDIE, 2021).

The passing in 2020 of the 4th Open Government Plan (Ministry of Regional Policy and Public Service, 2020) was a chance for this ASEDIE report (2021) to ask those responsible for open data in the 17 autonomous regions about the main aspects of the commitments taken on by their regions under this plan. These commitments are to be established and implemented in the years up to 2024, which is the horizon of the 4th plan.

Three key aspects emerge from the responses: transparency, open data and public participation. Out of the 17 autonomous communities, 8 specified that one of the main aspects of the commitment they have taken on concerns increased transparency through different mechanisms such as drafting new laws or strategic plans, facilitating access to transparency portals, implementing systems for integrity and accountability, etc. (ASEDIE, 2021).

With regards to open data, 7 out of 17 specified driving the opening up of data and their reuse as a commitment they have taken on. Drawing up open data strategies aimed at prioritising data and boosting interoperability, developing digital administration, fostering reuse and interaction with reusers are the main aspects mentioned. Of the 17 autonomous communities, four mentioned developing public participation through programmes of attention to the public, involving society in the constant improvement of public services, etc. (ASEDIE, 2021).

Cadena (2019) points out that data quality must be a multidimensional process on a web portal, starting from the moment when the information is digitised, and considers the requirements for their reuse in terms of the following dimensions (Cadena, 2019):

1. Accuracy: i.e., data that correctly represent the entity or event in the real world.
2. Timeliness: data that represent the real situation and agility in publishing them.
3. Consistency: they do not contain contradictions.
4. Exhaustiveness: they include all the data elements that represent the entity or event.
5. Availability: they are accessible now and in time.
6. Conformity: they adhere to accepted standards.
7. Credibility: they are reliable sources.
8. Accessibility: they are machine-readable.
9. Relevance: the right amount of data.

According to Cadena (2019), the first four quality dimensions depend on the source information systems, and dimensions 5 to 9 depend on the way the data portals are managed.

Ferrer Sapena, Calabuig, Sánchez Pérez & Vidal Cabo (2020) highlight how a certain way of processing data published in raw form, which they admit

few people use, can be important as a symbol of local development, good public services and social improvement. They argue that, in isolated cases, access to official data does not involve their real exploitation, but rather a demonstration that making their use possible represents social progress.

With reference to the development of governmental portals, Gil-García, Gasco-Hernández & Pardo (2020) point out that while there are successful cases, the speed at which the different projects have been carried out and the "need" to implement open government projects due to national or international pressures has led to a certain confusion and ambiguity. Specifically in the case of Spain, we have identified a large number of what we call pretender portals, as they are open data portals containing data that are not suitable for reuse.

Zuiderwijk, Pirannejad & Susha (2021) explain that when the actual implementation and impact of open data are examined taking a critical approach, it can be observed that most of the points of reference assessed to not measure the actual use of open data. Further work is therefore necessary on what Cetina (2021) calls "purposeful data", to identify in advance the most urgent challenges open data can help to meet.

To do this, it is very important to have reliable metrics that allow the quality of the data hosted on open data portals to be assessed with a view to reuse. This report assesses the open data portals in Spain by applying the latest version of the MELODA metric (Abella, Ortiz-De-Urbina-Criado & De-Pablos-Heredero, 2019).

3.4 The Data Reuse Model

MELODA is a metric for assessing the quality of open data, rating the information, and assessing its reusability (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2014). MELODA 4 is the version used in the reports of 2017 (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2017) and 2019 (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019). In its current version, MELODA 5 assesses two more dimensions and alters the levels and calculations (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2019), as shown in Table 1.

In version 5 of MELODA some of the ideas put forward by experts were examined and the weighting of each level reviewed. In this version each level is assigned the accumulated value it has (1, 3, 6, 10 and 15). To assess the degree of reuse of each dataset, two measurements are used: 1) the sum of the scores for each dimension, and 2) for each dimension there is a descriptive analysis of the frequency of each level. The first measurement gives a ranking of datasets according to their level of reuse; while the second gives a more detailed picture for each dataset and identifies which dimensions need improvement (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2019: 6).

Table 1: Dimensions and levels in MELODA 5

DDimensions (maximum 61 points)	Levels
Legal license (max. 6 points)	1. private use 2. non-commercial reuse 3. commercial or unrestricted reuse
Access to the information (max. 6 points)	1. access to the dataset through website or unique URL parameters 2. unique access to the website with parameters involving individual data 3. query language or API
Technical standard (max. 6 points)	1. reusable closed or non-reusable open standard 2. reusable open standard 3. open standard, with individual metadata
Level of standardisation (max. 10 points)	1. own standardisation model 2. own or ad hoc published standardisation model (harmonisation) 3. local standardisation 4. global standardisation
Geolocated content (max. 6 points)	1. no geographical information 2. simple or complex text fields 3. with full geographical information or coordinates
Data update frequency (max. 15 points)	1. over a month between updates 2. monthly. With update periods between 1 month and 1 day 3. daily. With update periods between 1 day and 1 hour 4. every hour. With update periods from 1 hour to 1 minute 5. in seconds Update period less than 1 minute
Dissemination (max. 6 points)	1. non-systematic communication / dissemination 2. resources available on updates (i.e., feed on social networks) 3. proactive dissemination / push dissemination (information automated in a certain time)
Reputation (max. 6 points)	1. there is no information on the reputation of the data source 2. statistics or reports are published according to users' opinions 3. rankings or indicators based on the reputation of the data source

Source: Abella, Ortiz de Urbina Criado & De Pablos Heredero (2019: 6)

Furthermore, to classify the datasets based on MELODA 5, three categories for the degree of reuse were created (Table 2): from 8 to 23 points (the bottom end is the sum of category 1 of the 8 dimensions) is inadequate; from 24 to 47 points (the bottom end is the sum of category 2 of the 8 dimensions) is basic; and from 48 to 61 points (the bottom end is the sum of category 3 of the 8 dimensions) is advanced.

Table 2: Ease of reuse rating ranges in MELODA 5

MELODA 5 ranges	8-23	24-47	48-61
MELODA 5 category	Inadequate	Basic	Advanced

Source: The authors

3.5 Aim of the Report

This report conducts a study of the reuse of data in Spain with the aim of presenting the state of the question and identifying guidelines and recommendations to help foster the use of data and generate business. It sets out from the two previous reports drawn up in 2017 (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2017) and 2019 (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019) to analyse the changes that have come about in the last two years. To this end, the open data portals in Spain were identified to analyse a sample of the datasets they publish and some of the services generated. A questionnaire was also conducted with those responsible for the portals to analyse some of the features and activities in relation to their open data. Specifically, a diagnosis was conducted of the portals' knowledge of the reuse of their data, of the type of innovation that can be developed with them, of activities to promote data use, of the services generated, of the value created by reusing the data and of the portals' reputation. The latest version of the MELODA 5 metric was applied to assess the degree of reuse of the open data published on Spanish open data portals. All these diagnostics allowed, through a SWOT analysis, a diagnosis of strengths, weaknesses, opportunities and threats, on the basis of which some reflections have been included to help construct future data management policy to foster the creation of business.

4 Methodology

4.1 Methodology for study of data publishing portals

330 data portals have been identified, from the following sources:

- Previous 2019 report on the state of open data in Spain (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019).
- Data from datos.gob through its list of initiatives[1].
- Open Administration Consortium of Catalonia.
- Complementary research by the team carrying out the report.

Through the consolidation of these sources, the availability of each of them was validated one by one, detecting in 41 cases that either the portal was not available (i.e. 404 error) or the portal did not publish data[2] .

In accordance with Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), the following values were identified for each portal:

- Availability of publication mechanisms for data updates.
- Availability of a resource catalogue, number of datasets available, and whether the catalogue is downloadable.
- Existence of direct connection mechanisms with the data (API) or query language (i.e. SPARQL).
- Availability of a portal where services and/or applications can be identified based on the data of the portals and the number of identified services.
- Use of a specific data publication and reuse tool: Data Management System (DMS).
- Number of published datasets.
- Which of the published datasets are unique to the portal and which are federated from other portals.
- Identification of the autonomous community of the entity that publishes the portal, or if it belongs to a national entity.
- The mechanisms for getting in contact with those responsible for the portal (e.g. an email address or web form).

1 https://datos.gob.es/es/iniciativas

2 The portals the Consorci d'Administració Oberta de Catalunya creates automatically for all public entities in Catalonia (1841) have been identified. Those where the public entity had assigned their own manager for the portal were included (17). We would like to thank them for their excellent collaboration in preparing this report.

4.1.1 Simplified maturity model of data publishing portals

Following the same methodology as Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), a maturity model has been defined, simplified based on the model defined in the pan-European data portal initiative by Carrara, Nieuwenhuis & Vollers (2016), introducing the following elements for consideration:

- The population of datasets that exceeds 30 items.
- The availability of a source (RSS channel or equivalent) with data updates.
- The availability of an application programming interface (API) that allows automated access to data by external users.
- The use of a data management system (DMS). For the purpose of this research, the following tools have been considered: CKAN, Socrata, DKAN, OpenDataSoft, ESRI Open Data and AOC.
- The availability of an application portal with services developed based on published data.

An analysis of the maturity of the data portals has been carried out with the information gathered. For this, the metric developed by Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017) has been used. The aspects analysed and their importance are presented in Table 3.

Table 3: Metrics to analyse the degree of maturity of data portals

Concept	Conceptin European report	Weight
Having more than 30 datasets	*Spread of data*	20%
Having a source with catalogue updates	*Usability of the portal*	10%
Use a *data management system* (DMS)	*Usability of the portal*	15%
Availability of an application programming interface (API) for automated interaction with *datasets*	*Reusability of data*	25%
Applications/services portal based on open data	*Reusability of data*	30%

Source: Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017) based on Carrara, Nieuwenhuis & Vollers (2016)

Portals with a maturity degree below 25 are considered inadequate, 25 to 50 are basic, 50 to 75 are adequate, and above 75 are optimal.

4.2 Methodology for study of published datasets

Of the 289 valid portals, a statistically representative sample of the population of datasets was selected with an interval of 10 points with 95% confidence using the survey software tool (http://www.surveysoftware.net/sscalce.htm), which resulted in a sample number of 300.

In the previous section, all the datasets that were specific to each portal were identified, which made it possible to make a list of the 58,318 datasets collected, in this way establishing the numeration intervals that corresponded to each portal. After this step, by means of a random generator, 300 different numbers were randomly identified, so that each number could be assigned to the portal according to the previous ranges.

A total of 58,318 datasets were identified in the 289 selected portals. Taking into account the range of the population under study, a sample of 300 datasets was carried out. Following Abella Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), for each dataset, each of the data reuse dimensions described in the MELODA metric (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2014) and updated to the latest version of MELODA 5 (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2019) were into consideration:

- Legal reuse license.
- Technical standard in which the information is presented.
- Information access mechanisms.
- Data model used.
- Geographic content of the information.
- Update frequency.
- Dissemination.
- Reputation.

The MELODA 5 metric reflects some of the recommendations given by the OECD in its latest document (OECD, 2021):

- From point 1 of the OECD (2021), "Empower and pro-actively engage all relevant stakeholders alongside broader efforts to increase the trustworthiness of the data ecosystem," it picks up the need for trust in the agents of the data ecosystem and is evaluated in the reputation dimension of open data publishers.

- From point 5 of the OECD (2021), "Further improve conditions for cross-border data access and sharing with trust," it picks up a fair number of elements corresponding to point 2 of the recommendation, such as having an open license, effective access mechanisms, shared data models, etc.

- It is especially aligned with point 6 of the OECD (2021), "Foster where appropriate the findability, accessibility, interoperability and reusability of data across organisations, including within and across the public and private sector," seeing as the ease of access to data, interoperability and data reusability are central elements of the metric and also affect each of its dimensions.

4.3.Methodology for study of the reuse of published data

A survey was carried out on open data portals during the months of June and July 2021. The survey was sent to 269 portals whose managers had been identified, a contact form had been requested or some form contact mechanism was found.

35 complete responses to the questionnaire were obtained, which represents a response rate of 13%. Each respondent had to answer a 10-question questionnaire about the reuse of data on their data portal, their knowledge of reusers and their use promotion policies (Annexe 10.1). A manual analysis of 330 open data portals was also carried out to identify their characteristics and their degree of maturity. Observations concluded that there are some that are no longer operational, specifically, 41 of them (12.42%).

To analyse the degree of reuse of open data, the datasets published on open data portals were reviewed. Specifically, 300 datasets were sampled, of which 280 (93.3%) resulted valid and the rest were observed to have no content. The MELODA 5 metric was applied to analyse the degree of reuse of open data (datasets) published in the sampled portals. In the new version of the MELODA metric, the reputation of these portal's data is evaluated, but since there are no rankings on this aspect, in order to solidly implement this aspect, a second survey was carried out among the 269 managers of the data portals on the reputation of published datasets (Annexe 10.2), which was launched on August 28 and ran until 30 September 2021. 35 responses were obtained (each with 10 qualified portals), therefore a rate of 13.0%[3].

On this occasion, only the portals that had been sampled in the sampling of datasets were enquired about (58 of the 289)[4]. The groups of portals to be qualified were divided into groups of 10 and each of the groups was assigned about 48 portals, making certain that no portal could vote for itself. In this way, six surveys were launched for the 58 portals, with different portals to vote on, yet with the same content.

This survey enquired about their knowledge of other open data portals and the prestige they presented according to their opinion. They were all classified into three categories. Their final reputation was obtained by means of the most frequent value voters gave on the reputation of other open data portals, analysed based on the knowledge the respondents claimed to have about the portal being voted on.

4.3.1 Data federation

Regarding the selection of datasets, the study observed that the federation of datasets is a phenomenon that has grown recently. Data federation is the

3 It is a coincidence that the response rate of the survey on data reuse coincides with that of the reputation ranking of open portals. The surveys were launched at different times and the coincidence of respondents cannot be confirmed seeing as both surveys are anonymous.

4 This small percentage of portal sampling occurs due to the disparity in the amount of data hosted in each one of them, while there are some portals with more than 3000 datasets, others barely exceed 30. The latter therefore have a probability of being sampled much lower, proportional to the number of published datasets.

republishing, on any given portal's catalogue, of datasets belonging to other portals. This publishing can be either complete, with the duplicated data, or through linking to the original portal.

This phenomenon, which in previous editions was only significant for the national Spanish portal datos.gob.es and for open data Euskadi, now occurs in 45 of the 289 portals. In addition, the identification of this federation is not standardised, therefore requiring manual investigation to determine different ones.

4.3.2 Consolidation of the portals of the General State Administration

In previous reports (2017 and 2019) it was standard for each public entity of the General State Administration to have its own open data portal. Yet in 2021, the majority had this function delegated to the data portal datos.gob.es. For this reason, the portals were examined to identify the datasets that are federated, and to establish criteria for their sampling. Specifically, only those that are federated in datos.gob.es have been taken as a quantity, seeing as the classification of these portals is not the same. For example, this criterion has been applied for the National Centre of Geographic Information (CNIG), and based on our analysis, they have 80 datasets, while the National Library, following the same criteria, has 285. For the INE's datasets, said institution defines statistical operations (217 short-term operations, 444 structural and 64 of another type), each of which can generate a varied number of datasets, but with the means at hand it is not possible to know how many these are. For the cadastre, we observed that it has many sources that are not open data (they require identification to access), but those that are, are structured by years and provinces in many cases. In this case, it was counted as only one, independent of years or provinces, and we noted that it is not federated with datos.gob.es. Therefore, the 8,507 datasets federated in datos.gob.es and generated by INE were analysed. Note that there may be many more, but we have no reasonable way of estimating them.

4.4 Methodology for study of the reuse of published data

To complete the information, a direct analysis of the applications and services that the data portals themselves provide as accredited data reusers was also carried out, in this way identifying the developers and extracting data from each of their corporate portals. 62 applications and services were sampled from randomly chosen portals that had inventoried services based on open data. This implies a statistically significant sample for an interval of 12 points with a confidence level of 95%, following the same approach and tool as in the previous point. As for the services developed, it was detected in the sample of 62 applications/services listed in the application sections of the portals that 17 of them were no longer available, representing a percentage of 27%.

Based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), the data extracted for each of these services was:

- The themes of the service according to NTI-RISP classification, from which its equivalent in DCAT-AP (European Union, 2017) can be extracted.
- Geolocation features.
- The real-time features of the service.
- The type of application developer.
- A sustainability mechanism, depending on if it has one or not. Sustainability describes the economic viability of the service in the medium term, either because there is an entity that supports the costs or because the service has its own mechanisms for generating income. For example, in the case of a corporate service, the entity that publishes it assumes its costs despite not generating directly related income. If it has a business model that includes a source of income other than that of the entity that publishes the data and, if so, what type. The categories used are:

    - Ads (ads support the cost of the application/service).
    - Supporting entity (i.e. a public entity financing it).
    - Freemium (one part of the service is free while the other is paid).
    - Marketing entity (SME). Used to advertise the developer.
    - Pay per use. The application must be paid for.
    - No business model.

5

Diagnosis

5.1 Diagnosis of portals that publish data

330 portals were sampled, of which 41 are no longer active. Therefore, an analysis of 289 portals was carried out. If we analyse which Autonomous Communities these portals are associated with (Table 4), we observe that 39.8% are in Extremadura, 17.0% in Catalonia[5], 5.5% in Andalusia and the Canary Islands, and 4.5% in Madrid. Of all of them, 9.3% are considered national.

Table 4: Open data portals by Autonomous Community

Autonomous Community	Frequency	Percentage
Andalusia	16	5,5
Aragon	5	1,7
Balearic Islands	3	1
Canary Islands	16	5,5
Cantabria	3	1
Castilla and León	4	1,4
Castilla-La Mancha	5	1,7
Catalonia	49	17.0
Community of Madrid	13	4,5
Navarra	2	0,7
Community of Valencia	6	2,1
Extremadura	115	39,8
Galicia	3	1
La Rioja	1	0,3
Melilla	1	0,3
Basque Country	10	3,5
Principality of Asturias	3	1
Region of Murcia	4	1,4
Valencia	3	1
National	27	9,3
Total	**289**	**100**

Source: The authors

5 It should be noted that Catalonia has automated the creation of an open data portal for all public entities in the region, which not only includes municipalities but also consortiums, associations, and other public law entities, more than 1,800. These have not been included in this version of the report.

The presence of Extremadura as the leading community is a result of the more than 100 portals generated automatically for as many towns in the province of Cáceres. It should be noted that these portals have around thirty datasets, do not use a DMS, do not have APIs and do not present significant changes in their content since the previous report.

5.1.1 Updating of data and API availability

In 2021, 17.3% of portals had a channel that allowed knowing when a dataset had been updated; slightly higher than the percentage in 2019, when it was 17.2%, but lower than in 2017 when the percentage was 28%. In 2021, 37.7% had an automated data access mechanism (API or semantic interrogation point); while in 2019 it was 33.3%, and in 2017, 46%.

One out of every four portals publishing open data have an alert mechanism for its data actualization, and slightly more than a third have an automated access (API).

In 2021, as occurred in 2019 and 2017, approximately 63% of portals did not have an automated access mechanism (API), which poses a similar problem, seeing as it forces reusers to download all the data from a dataset, even though they only need part of it, and to process it in order to extract the data they really need.

5.1.2 Data management system

Regarding the system used, in 2021, 29.07% correctly used a data management system (CKAN, Socrata, DKAN, ESRI open data and AOC); while, in 2019 it was 25.2%. On the other hand, in 2021, 55.71% used a non-specific solution such as certain content managers including Joomla, Wordpress and Liferay, a figure that was lower in 2019 (49.8%). On the other hand, in 2021, in 13.49% of the cases it was not possible to identify the platform manager, seeing as they did not belong to the previous groups. The figure was similar in 2019 (12.9%).

Only 29% of data portals use a specific tool for data publishing (DMS).

5.1.3 Development services portal

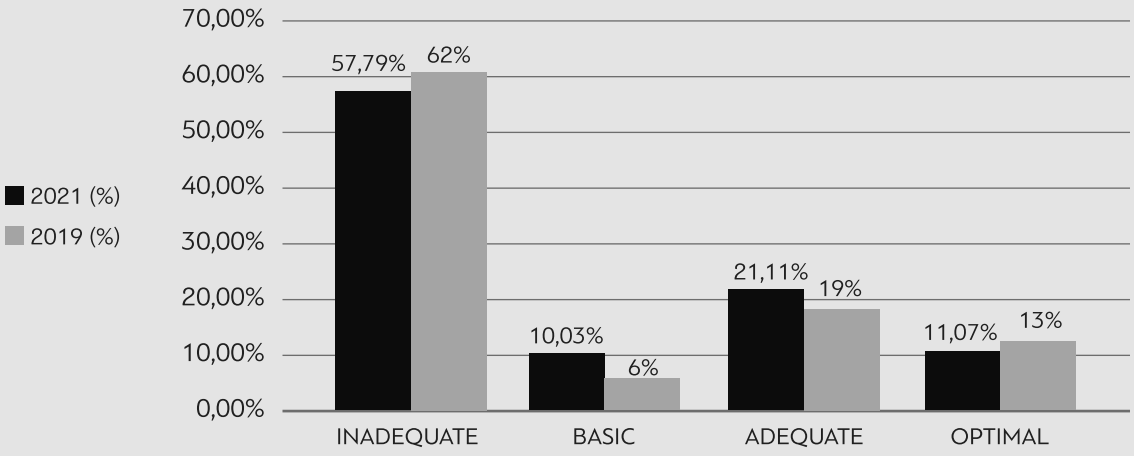
Regarding the development of services, in 2021, 15.6% of portals had information on the services developed based on their data; while in 2019, the figure was somewhat higher (19.4%) and in 2017 it was 40% (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019).

5.1.4 Portal maturity according to methodology

Figure 5 shows the percentage of portals based on their maturity for data publication according to the methodology described in point 3.1. In 2021, the maturity level improved a little, although most are still inadequate (57.79% of

Figure 5: Distribution of data portal maturity in Spain in 2019 and 2021.

cases); in 2019, that figure was 62%. In 2019, only 13% had an optimal level, a figure that is similar for 2021 (11.07%). In addition, in 2021, 21.11% of portals had an adequate maturity level, similar to 2019 (19%)



Source: The authors

5.2 Diagnosis of published datasets

300 datasets have been analysed, of which 6.7% cannot be considered as such either because they are not in recoverable data format or were not available (404 error). For this reason, the analyses carried out in this section refer to the 280 that can be considered valid.

5.2.1 Distribution by degree of maturity of portals that publish datasets

Figure 6 reflects the sum of datasets found in the portals according to their degree of maturity in 2021. If we add the datasets located in the portals that are in each maturity interval, we can observe in Figure 6 that in 2021, 88% of the datasets were located in portals with degrees of optimal or adequate maturity, while in 2017 it was 91%, and in 2019, 72.75%.

5.2.2 Distribution according to portal maturity

Figure 6 shows that 39% of portals are in the optimal category. On the other hand, 49% of datasets are in portals with an adequate level of maturity.

Almost three out of every four published datasets are located in the portals with a higher maturity in data publishing.

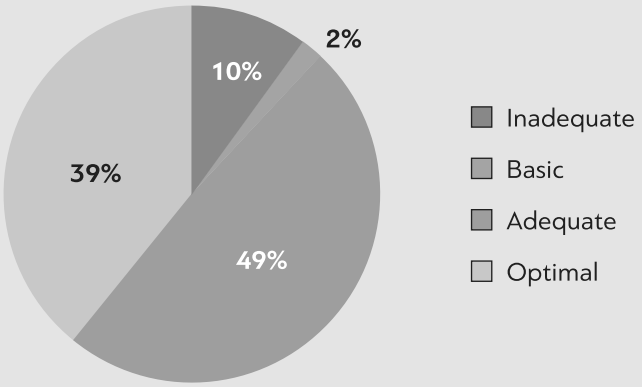
5.2.3 Categorisation by reuse license

Figure 7 shows the distribution of datasets according to the user license under which the data is published. In 2021, 95% published data under licenses that allow commercial reuse or without restrictions, and only 5% maintained their data under licenses for private use.

DATASETS BY DEGREE OF THE MATURITY OF THEIR PORTALS 2021 (%).

Figure 6: Distribution of datasets according to the maturity of the portals that publish them (2021)

Source: The authors

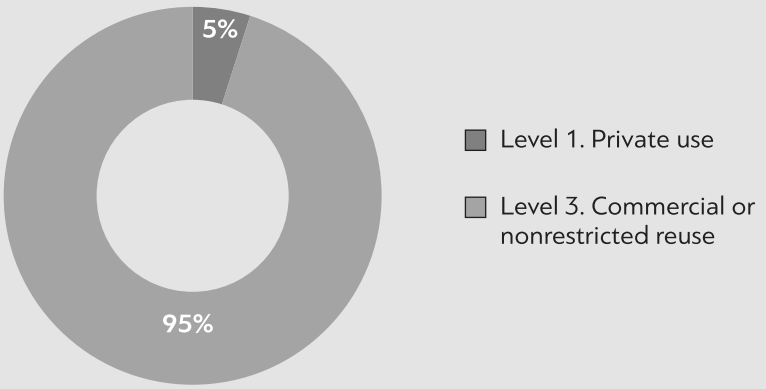


5% of published dataset have a license that restricts the commercial use of data.

LICENCE 2021

Figure 7: Distribution of datasets by type of license 2021 (MELODA 5).

Source: The authors



5.2.4 Categorisation by data model

Figure 8 shows the distribution of datasets according to data model. In 2021, only 2% had a global data model, and the majority, 78%, had their own data model but without documentation. And 15% had a local but documented data model. The European Union has various projects[6] that seek the adoption of recognised and open standards. This signifies a clear barrier for the reuse of data.

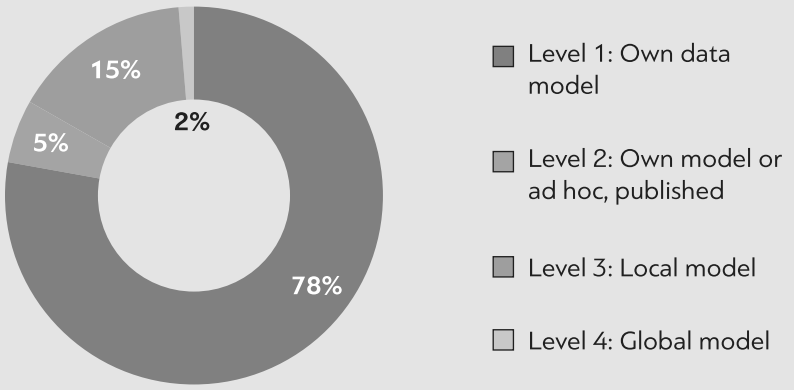
Only 2% of published datasets have a global data model related to an international standard.

6 https://ec.europa.eu/cefdigital/wiki/display/CEFDIGITAL/Data+Models

DATA MODEL 2021

Figure 8: Distribution of datasets by data model 2021 (MELODA 5).

Source: The authors



5.2.5 Categorisation by technical standard utilised

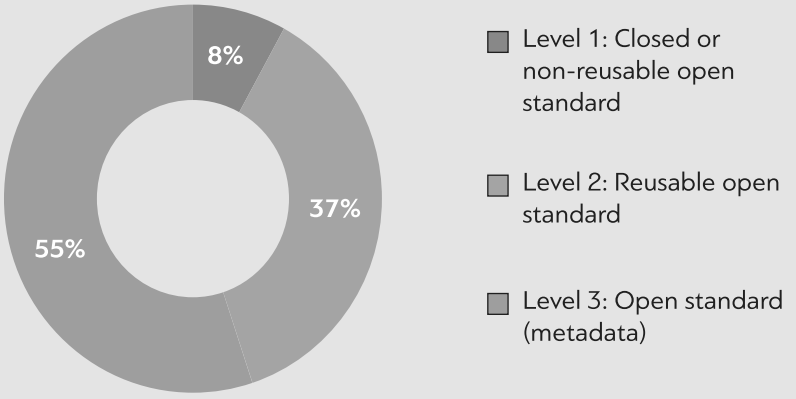
Figure 9 shows the distribution of datasets according to the technical standards with which the data was published. In 2021, 55% were at the highest level of technical standard. It is also worth noting that only 7.9% had a closed or non-reusable open standard.

Approximately 8% of the datasets are not reusable due to the technical standard utilised.

TECHNICAL STANDARD 2021

Figure 9: Distribution of datasets according to storage standard 2021 (MELODA 5).

Source: The authors



5.2.6 Categorisation by access mechanisms necessary to access data

In 2021, it is worth noting that 70% of datasets were at level 3 (API or query language) and 15% at level 2. Another 15% provided access via website or unique parameters (Figure 10).

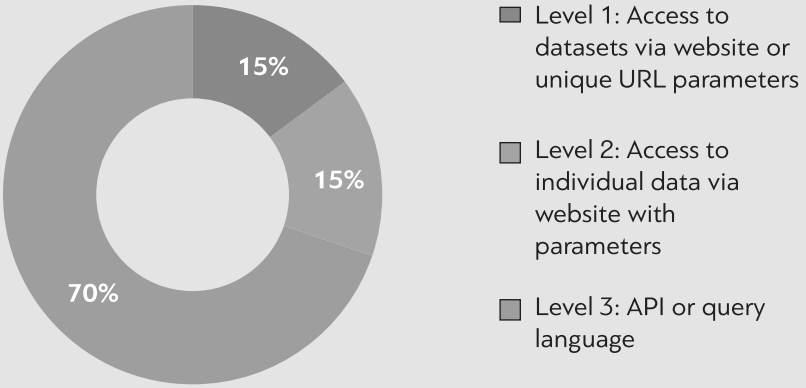
In 2021, 49% of datasets did not have any type of geographic information. On the other hand, 17% had complete coordinates or information associated with published content (Figure 11).

5.2.7 Categorisation by geographic content

ACCESS MECHANISM 2021

Figure 10: Distribution of datasets by access mechanism used 2021 (MELODA 5)

Source: The authors



GEOLOCATION 2021

Figure 11: Distribution of datasets by geographic content of information 2021 (MELODA 5).

Source: The authors

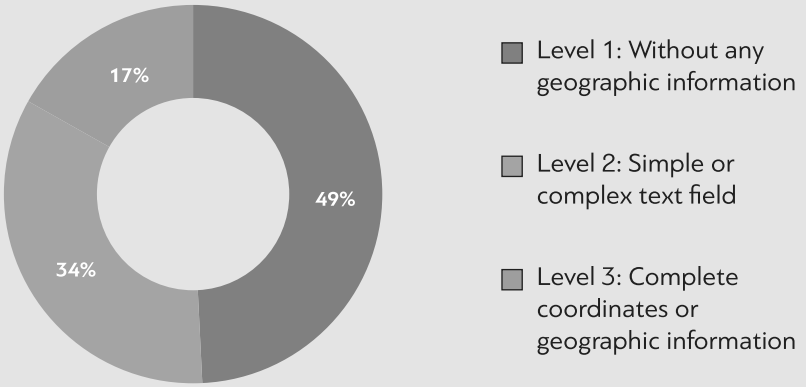


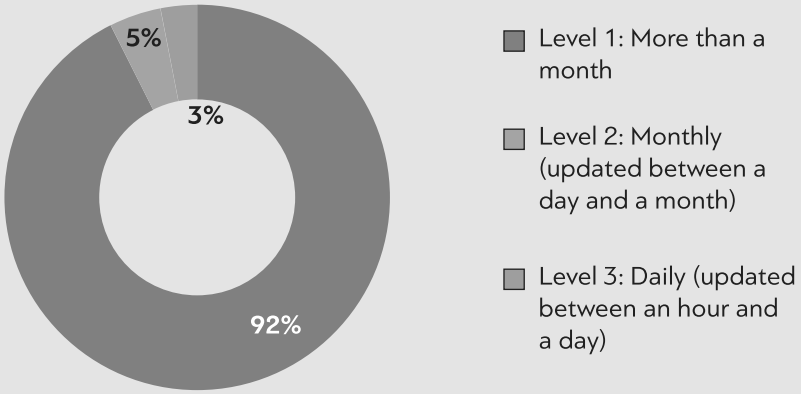
Figure 12 shows the update frequency of the sampled datasets. Especially noteworthy is the fact that 92% are in the category of carrying out updates more than one month apart, and that only 3% carry out a daily update.

5.2.8 Categorisation by update frequency

UPDATE FREQUENCY 2021

Figure 12: Distribution of datasets according to their update frequency 2021 (MELODA 5).

Source: The authors



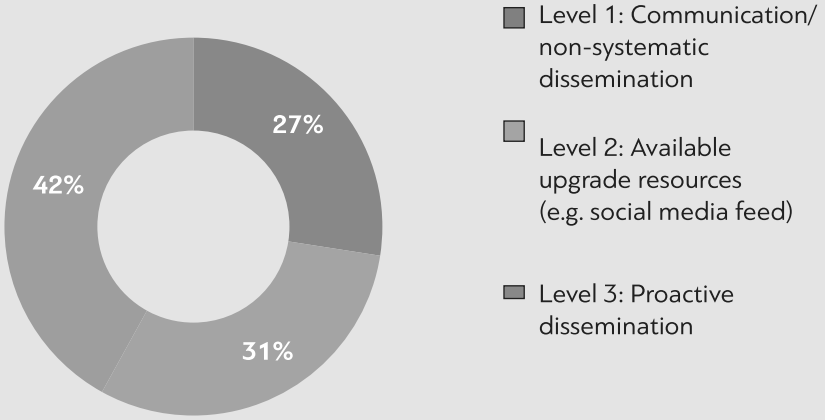
5.2.9 Categorisation by dissemination

Figure 13 shows the spread of the sampled datasets. It stands out that 42% were in the highest category with proactive dissemination, while 27% did not carry out systematic dissemination.

DISSEMINATION 2021

Figure 13: Distribution of datasets according to their dissemination 2021 (MELODA 5).

Fuente: Elaboración propia



5.2.10 Categorisation by reputation

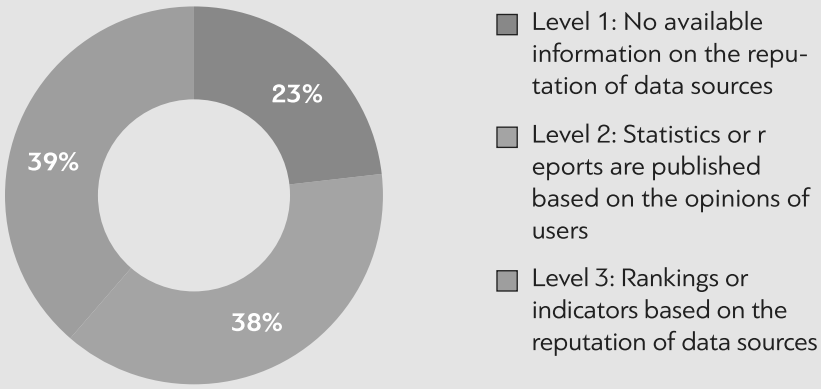
Figure 14 shows the reputation of portals on which the sampled datasets were published. Especially noteworthy is the fact that 39% of the datasets were on portals with a high reputation level, while 23% have a low level.

In addition to analysing the MELODA 5 reputation aspect, the responses to the survey sent to those responsible for the portals were used to rank them, considering the two variables analysed: knowledge of the portal (1: low level of knowledge; 2: medium level of knowledge; 3: high level of knowledge) and reputation (1: emerging portal; 2: mature portal; 3: model portal). These rankings can be found in Annexe 12.3.

REPUTATION 2021

Figure 14: Distribution of datasets based on reputation 2021 (MELODA 5).

Source: The authors



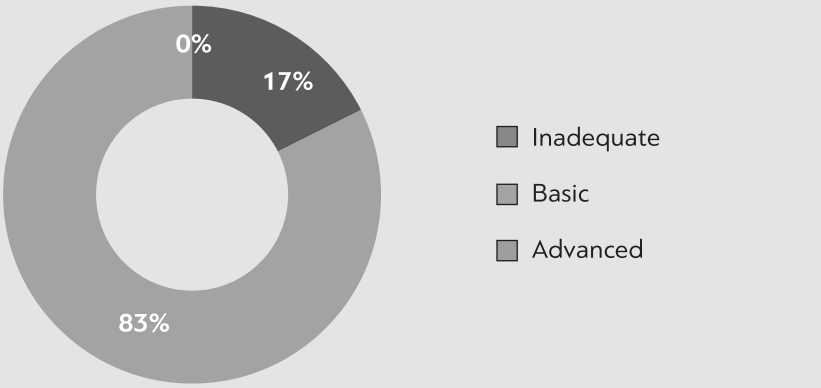
5.2.11 Categorisation by global reuse

In accordance with the MELODA 5 metric, the sampled data has been qualified according to its eight categories (reuse license, technical standard, access mechanism, data model, data geolocation, update frequency, dissemination, and reputation). Figure 15 shows the obtained results.

It is noteworthy that according to the MELODA 5 categories, no dataset has an advanced level of reuse, yet the majority, 83%, have a basic degree.

Figure 15: Distribution of sampled datasets by MELODA 5 categories 2021.

Source: The authors



5.3 Diagnosis of data reuse

5.3.1 Analysis of knowledge regarding entities that reuse published data

The results obtained in the survey carried out in 2021 are presented below and have been compared with those obtained in the 2017 (Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero, 2017) and 2019 (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019) reports.

Respondents were asked if they had any knowledge of the public or private entities (companies, NGOs, academics, students, individual citizens) that reuse the data they publish on their open data portal.

In 2021, almost every portal responsible knows the reusers of the data they publish.

As can be seen in Figure 16, in 2017, more than 77% of those surveyed stated that they had knowledge of at least one of the reusers, while this figure was 75.6% in 2019 and 97.2% in 2021. These findings show that awareness of data reuse has improved substantially.

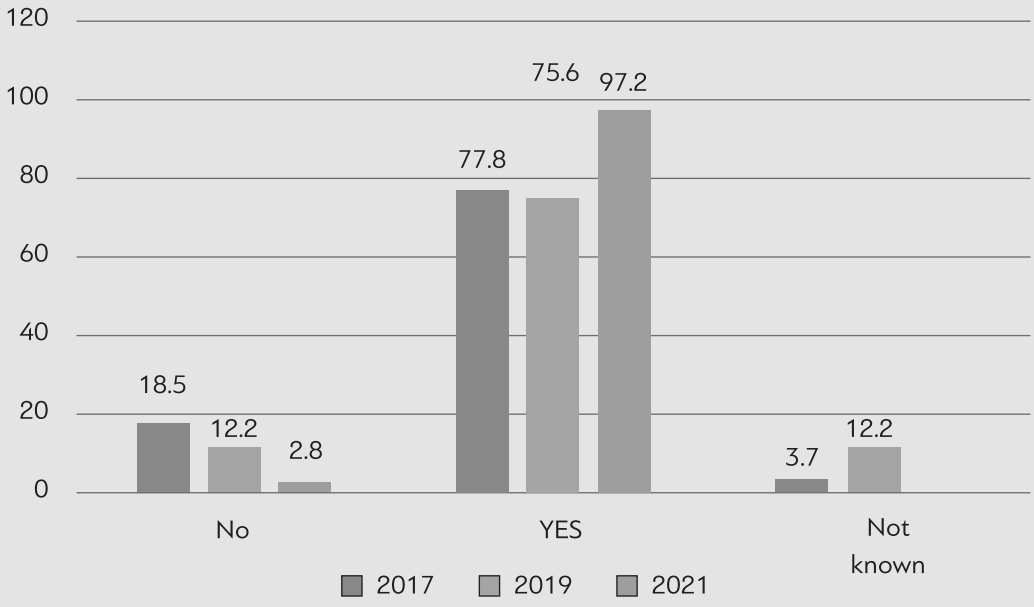


Figura 16. Conocimiento de los reutilizadores de datos (%).

Source: The authors based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

To find out the respondents' opinions on the degree of reuse of open data for each type of reuser, they were asked to what extent each type of entity reused the open data on their portal. Table 5 shows the obtained results.

Based on the opinion of those responsible for the portals, it was obtained that citizens increasingly reuse open data. In 2017, only 25.9% used it frequently or always. In 2019, the figure was even lower, 19.6%, but in 2021, the figure increased to 38.80%.

In the case of for-profit professional reusers (infomediaries and individual for-profit developers), we observed that, in 2019, there was an increase in the frequency of use (36.6% use it frequently or always) compared to the 25.9% that was obtained in 2017, and that figure remains similar in 2021 (33.30%). Regarding the use of open data by professional non-profit reusers (NGOs, foundations, individual non-profit developers and other social initiatives), the figures dropped in 2021, with 13.90% using it frequently or always, while this number stood at 31.7% in 2019 and 22.2% in 2017.

Table 5: Knowledge of the type of data reusers

Scale		Never	Hardly ever	Sometimes	Frequently	Always	Not known/ No response	Total	N.º
Citizens	2017	3,70%	22,20%	22,20%	11,10%	14,80%	25,90%	100%	27
	2019	9,80%	12,20%	29,30%	9,80%	9,80%	29,30%	100%	41
	2021	5,60%	22,20%	30,60%	19,40%	19,40%	2,80%	100%	36
For-profit professionals	2017	0%	18,50%	18,50%	18,50%	7,40%	37%	100%	27
	2019	0%	12,20%	17,10%	24,40%	12,20%	34,10%	100%	411
	2021	2,80%	11,10%	41,70%	19,40%	13,90%	11,10%	100%	36
Non-profit professionals	2017	11,10%	7,40%	11,10%	11,10%	11,10%	48,10%	100%	27
	2019	0%	12,20%	12,20%	19,50%	12,20%	43,90%	100%	41
	2021	6%	13,90%	25,00%	11,10%	2,80%	41,70%	100%	36
Academics	2017	0%	14,80%	25,90%	25,90%	11,10%	22,20%	100%	27
	2019	0%	12,20%	26,80%	17,10%	17,10%	26,80%	100%	41
	2021	3%	16,70%	44,40%	13,90%	11,10%	11,10%	100%	36
Own organisation	2017	7,40%	3,70%	22,20%	22,20%	22,20%	22,20%	100%	27
	2019	0%	2,40%	24,40%	26,80%	29,30%	17,10%	100%	41
	2021	6%	8,30%	16,70%	22,20%	41,70%	5,60%	100%	36
Other public entities	2017	7,40%	14,80%	22,20%	7,40%	14,80%	33,30%	100%	27
	2019	0%	9,80%	22%	14,60%	12,20%	41,50%	100%	41
	2021	2,80%	13,90%	44,40%	5,60%	8,30%	25,00%	100%	36

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

Researchers and academic personnel (including students) maintained similar figures in 2017 (37% frequently or always) and 2019 (34.2%) but dropped to 25% in 2021. Regarding the organisations themselves, they are the agents that most frequently use open data. They were frequent or habitual users in 2017 (44.4%); a figure that increased in 2019 (56.1%) and in 2021 (63.90%). In contrast, for other public entities, the frequency of use is lower: 22.2% used them frequently or always in 2017, 26.8% in 2019, and in 2021 it dropped to 13.90%.

Also noteworthy is the percentage of cases that do not know if the data published on their portal is being reused. In 2017, the highest percentage was

for non-profit professional reusers (48.1%), as well as those who are for-profit (37%). In 2019, the highest values were for non-profit professionals (43.9%) and other public entities (41.5%). And in 2021, for non-profit professionals (41.7%) and other public entities (25%). In addition, in 2019, only 17.1% of organisations did not know if their own data was being used; a figure that was greatly reduced in 2021 (5.6%).

5.3.2 Analysis of data reuse by activity sectors

Regarding the reports by Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017) and Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), we can observe that in 2017, the transport and storage sector and the public administration and defence sector were the sectors with the highest frequency of daily use, although in absolute terms it was still low (11.1%). In 2019, there was an increase in the frequency of daily use in these two sectors: 14.6% in storage and transportation and 17.1% in public administration and defence. In addition to these sectors we find information and communications with 19.5%, while in 2017 its figures were very low (3.7%). However, in 2021, these figures fell in all sectors, the most frequent users being information and communications (8.3%) and other services (8.3%) (Tables 6.1, 6.2 and 6.3).

Table 6.1: Data reusers by activity sector (I)

Reusers by sector	Sector /year	Agriculture, forestry and fishing	Manufacturing industry	Extractive industries	Electricity, gas, steam and air conditioning suppliers	Water supply, sanitation activities, waste management and decontamination	Construction	Information and communications
Daily	2017	3,70%	0%	NA	NA	NA	NA	3,70%
	2019	0%	0%	2,40%	0%	7,30%	0%	19,50%
	2021	0%	0%	0,00%	3%	0,00%	0%	8,30%
Frequently	2017	0%	0%	NA	NA	NA	NA	11,10%
	2019	9,80%	0%	0%	4,90%	4,90%	0%	9,80%
	2021	2,80%	3%	0%	0,00%	2,80%	6%	2,80%
Sometimes	2017	3,70%	0%	NA	NA	NA	NA	0%
	2019	9,80%	9,80%	0%	2,40%	7,30%	4,90%	9,80%
	2021	8,30%	0,00%	0%	0,00%	5,60%	2,80%	8,30%
Exceptionally	2017	3,70%	0%	NA	NA	NA	NA	7,40%
	2019	9,80%	7,30%	12,20%	9,80%	9,80%	12,20%	7,30%
	2021	2,80%	2,80%	2,80%	0,00%	2,80%	0,00%	8,30%
Never	2017	7,40%	3,70%	NA	NA	NA	NA	0%
	2019	2,40%	2,40%	4,90%	4,90%	4,90%	9,80%	2,40%
	2021	5,60%	0,00%	0,00%	2,80%	2,80%	5,60%	16,70%

Note: NA: not available.

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

Table 6.2: Reusers by activity sector (II)

Reusers by sector	Sector /year	Whole-sale and retail	Transport and storage	Hospita-lity and tourism	Finan-cial and insurance activities	Real estate	Professio-nal, scientific and tech-nical activities	Adminis-trative activities and auxiliary services	Public adminis-tration and defence
Daily	2017	0%	11,10%	3,70%	0%	3,70%	0%	0%	11,10%
	2019	2,40%	14,60%	9,80%	2,40%	2,40%	4,90%	7,30%	17,10%
	2021	2,80%	5,60%	11,10%	0,00%	2,80%	2,80%	2,80%	0,00%
Fre-quently	2017	3,70%	0%	7,40%	0%	0%	7,40%	3,70%	11,10%
	2019	4,90%	4,90%	14,60%	2,40%	4,90%	17,10%	7,30%	14,60%
	2021	5,60%	5,60%	2,80%	5,60%	2,80%	13,90%	0,00%	8,30%
Some-times	2017	0%	0%	11,10%	3,70%	7,40%	7,40%	0%	11,10%
	2019	7,30%	4,90%	12,20%	4,90%	14,60%	17,10%	7,30%	17,10%
	2021	2,80%	8,30%	8,30%	0,00%	0,00%	5,60%	5,60%	19,40%
Excep-tionally	2017	0%	0%	3,70%	0%	0%	7,40%	0%	7,40%
	2019	9,80%	9,80%	4,90%	7,30%	0%	4,90%	4,90%	2,40%
	2021	0,00%	2,80%	5,60%	0,00%	6%	11,10%	5,60%	5,60%
Never	2017	0%	0%	3,70%	0%	0%	0%	0%	0%
	2019	2,40%	2,40%	2,40%	2,40%	2,40%	4,90%	4,90%	2,40%
	2021	2,80%	11,10%	8,30%	2,80%	5,60%	11,10%	5,60%	19,40%

Note: NA: not available.

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

Table 6.3: Reusers by activity sector (III)

Reusers by sector	Sector/ year	Educa-tion	Health and social service activities	Artistic, recreational and enter-tainment activities	Activities of extraterrito-rial organi-sations and bodies	Activities of households as employers of domestic staff	Other services
Daily	2017	0%	0%	0%	0%	ND	3,70%
	2019	2,40%	0%	4,90%	0%	0%	ND
	2021	2,80%	0%	2,80%	6%	0%	8,3
Frequently	2017	3,70%	0%	0%	0%	ND	0%
	2019	14,60%	2,40%	4,90%	2,40%	2,40%	NA
	2021	2,80%	0,00%	2,80%	0,00%	0,00%	5,6
Sometimes	2017	7,40%	0%	3,70%	3,70%	NA	3,70%
	2019	7,30%	9,80%	7,30%	12,20%	0%	NA
	2021	13,90%	5,60%	2,80%	2,80%	0%	2,8
Exception-ally	2017	7,40%	3,70%	0%	3,70%	NA	0%
	2019	7,30%	7,30%	7,30%	2,40%	7,30%	NA
	2021	8,30%	0,00%	0,00%	0,00%	0,00%	2,8
Never	2017	0%	0%	0%	0%	NA	0%
	2019	2,40%	2,40%	4,90%	4,90%	4,90%	NA
	2021	8,30%	2,80%	0,00%	0,00%	0,00%	0

ote: NA: not available.

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

In 2017, the information and communications sector and the public administration and defence sector were the ones that stand out in the frequent use of data compared to other sectors, but their frequency was also low (11.1%, in each case). However, in 2019 there was an increase in the figures in the "frequent use" category in several sectors: hospitality and tourism (14.6%), professional, scientific and technical activities (17.1%), education, and public administration and defence (14.6%). And in 2021, the sectors with the highest frequency were professional, scientific and technical activities (13.9%), and public administration and defence (8.3%).

5.3.3 Analysis of data reuse by territorial scope

To find out the scope of action, those responsible for the portals were asked about the territorial spheres in which the reusers of their data worked (Table 7).

Table 7: Spheres of action of open data reusers

Scale	Year	Never	Hardly ever	Sometimes	Frequently	Always	Not known /No response	Total
Local	2017	3,70%	11,10%	18,50%	11,10%	14,80%	40,70%	100%
	2019	2,40%	7,30%	14,60%	9,80%	29,30%	36,60%	100%
	2021	5,60%	8,30%	27,80%	25,00%	22,20%	11,10%	100%
Regional	2017	0%	11,10%	22,20%	7,40%	11,10%	48,10%	100%
	2019	2,40%	4,90%	14,60%	29,30%	2,40%	46,30%	100%
	2021	2,80%	8,30%	36,10%	22,20%	11,10%	19,40%	100%
National	2017	7,40%	4%	14,80%	14,80%	0%	55,50%	100%
	2019	2,40%	2,40%	19,50%	22%	7,30%	46,30%	100%
	2021	2,80%	25,00%	33,30%	14%	2,80%	22,20%	100%
European	2017	11,10%	11,10%	7,40%	3,70%	3,70%	62,90%	100%
	2019	4,90%	4,90%	17,10%	7,30%	4,90%	61%	100%
	2021	19,40%	27,80%	2,80%	2,80%	2,80%	44%	100%
Others	2017	7,40%	7,40%	0%	3,70%	7,40%	74,10%	100%
	2019	2,40%	4,90%	4,90%	2,40%	4,90%	80,50%	100%
		5,56%	8,33%	5,56%	0,00%	2,78%	77,78%	100%

Own elaboration based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

Note: Due to rounding of the figures, the sum of the figures in the left-hand columns may differ the total column.

In 2017, it was observed in all areas of action that in almost half of the cases no information was available, the local sphere being the one with the lowest figure, 40.7%. However, this trend began to improve in 2019, lowering the numbers a bit in almost all spheres. And then in 2021, it dropped even further in all cases.

If we analyse, for each sphere, the cases in which use is frequent or always, we can observe that the local sphere had a percentage of 25.9% in 2017, 39.1% in 2019 and 47.2% in 2021, while in the regional (18.5% in 2017, 31.7% in 2019 and 33.3% in 2021), national (14.8% in 2017, 29.3% in 2019 and 16.8% in 2021) and European (7.4% in 2017, 12.2% in 2019 and 5.6% in 2021) the figures were lower, decreasing in 2021.

Also noteworthy are the results obtained for the frequency "never". In this case, it is the European sphere that has the highest figures, 11.1% in 2017, 4.9% in 2019 and 19.4% in 2021, with the national standing at 7.4% in 2017, 2.4% in 2019 and 2.8% in 2021, and the local at 3.7% in 2017, 2.4% in 2019 and 5.6% in 2021.

5.3.4 Analysis of the types of innovation for the reuse of open data

To find out the types of innovation that can be achieved with the use of open data, the portals were asked to what extent the types of innovation that appear in Table 8 were used.

Table 8: Types of innovation for the reuse of open data

Escala	Product innovation			Process innovation		
	2017	2019	2021	2017	2019	2021
Never	7,40%	9,80%	8,30%	7,40%	4,90%	8,30%
Hardly ever	25,90%	9,80%	27,80%	22,20%	14,60%	30,60%
Sometimes	7,40%	12,20%	25,00%	11,10%	17,10%	27,80%
Frequently	14,80%	14,60%	8,30%	11,10%	9,80%	11,10%
Always	0%	2,40%	2,80%	3,70%	4,90%	0,00%
Not known/No response	44,40%	51,20%	27,80%	44,40%	48,80%	22,20%
Total	100%	100%	100%	100%	100%	100%

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

By types of innovation, the results obtained show that in 2017, in 44.4% of cases, the respondents did not know if the open data on their portal was being used to carry out product or process innovation. In 2019, the figures were even higher: 51.2% for product innovation and 48.8% for process innovation. However, in 2021 the figures dropped substantially, standing at 27.8% and 22.2%, respectively. In addition, we can observe that for the categories "frequently" and "always", the percentages were low for all years: 14.80% in 2017, 17% in 2019 and 11.10% in 2021 for product innovation, were similar for process innovation: 14.80% in 2017, 14.70% in 2019 and 11.10% in 2021.

The responsible people for the data portals have little knowledge of the product or process innovations generated with their data.

5.3.5 Analysis of availability of data access records

Table 9 shows the information available according to those responsible for the portals based on access login records.

We can observe that in 2017, 59.3% of portals had knowledge of the records of access to their data. In 2019 and 2021, this figure was approximately 63%. This analysis is interesting seeing as data access records are an essential element for managing and promoting demand.

Table 9: Availability of data access records

Frequency/year	2017	2019	2021
No	25,90%	14,60%	36,10%
Yes	59,30%	63,40%	63,90%
Not known/No response	14,8%	22%	0%
Total	100%	100%	100%

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

More than 60% of the portals confirms that they have data access records. In 2021, every portal knows if there is an access to their data.

5.3.6 Analysis of activities promoting the use of open data

To gather insight on the activities that promote the use of open data, portals were asked to what extent they used certain types of promotional activities. Table 10 shows the obtained results.

Regarding activities promoting the use of open data, in 2017 and 2019 there were no cases in which an activity was used always, but in 2021, that figure rose to 3%. However, there were more activities that were used frequently.

Meetings with reusers were the most common in 2017 and 2019 (25.9% and 19.5%, respectively), but were less common in 2021 (5.6%). Frequent use of internal activities stood at 18.5% in 2017, 19.5% in 2019 and decreased to 16.7% in 2021. However, the decrease in 2021 of external events organised by other entities to promote available data was even greater: from 22.2% in 2017 and 17.1% in 2019 to only 11.1% in 2021.

Table 10: Activities promoting the use of open data

Scale		Never	Hardly ever	Sometimes	Frequently	Always	Not known/No response	Total
Application contests	2017	40,70%	14,80%	11,10%	11,10%	0%	22,20%	100%
	2019	29,30%	22%	12,20%	4,90%	0%	31,70%	100%
	2021	44,40%	11%	22,20%	2,80%	0%	19,40%	100%
Internal promotion activities	2017	7,40%	29,60%	29,60%	18,50%	0%	14,80%	100%
	2019	7,30%	26,80%	17,10%	19,50%	0%	29,30%	100%
	2021	25,00%	22,20%	27,80%	16,70%	3%	5,60%	100%
External promotion activities	2017	7,40%	29,60%	22,20%	22,20%	0%	18,50%	100%
	2019	14,60%	24,40%	22%	17,10%	0%	22%	100%
	2021	16,70%	30,60%	31%	11,10%	0%	11,10%	100%
Meetings with reusers	2017	25,90%	25,90%	3,70%	25,90%	0%	18,50%	100%
	2019	29,30%	22%	9,80%	19,50%	0%	19,50%	100%
	2021	25,00%	33%	22,20%	5,60%	0%	13,90%	100%
Others	2017	18,50%	0%	3,70%	7,40%	0%	70,40%	100%
	2019	24,40%	0%	0%	14,60%	0%	61%	100%
	2021	19,40%	2,80%	2,80%	5,60%	0,00%	69,40%	100%

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

We also discovered cases in which respondents considered that certain types of activities were never used. The figures obtained for application contests (40.7% in 2017, 29.3% in 2019 and 44.40% in 2021) and for meetings with reusers (25.9% in 2017, 29.3% in 2019 and 25% in 2021) stand out.

The most frequent activities promoting the use of open data were internal events.

5.4 Diagnosis of generated services

To identify the generated services, on the one hand, those responsible for the portals were asked about the most common uses of their portal's data (Annexe 10.4) and, on the other, the most used datasets (Annexe 10.5).

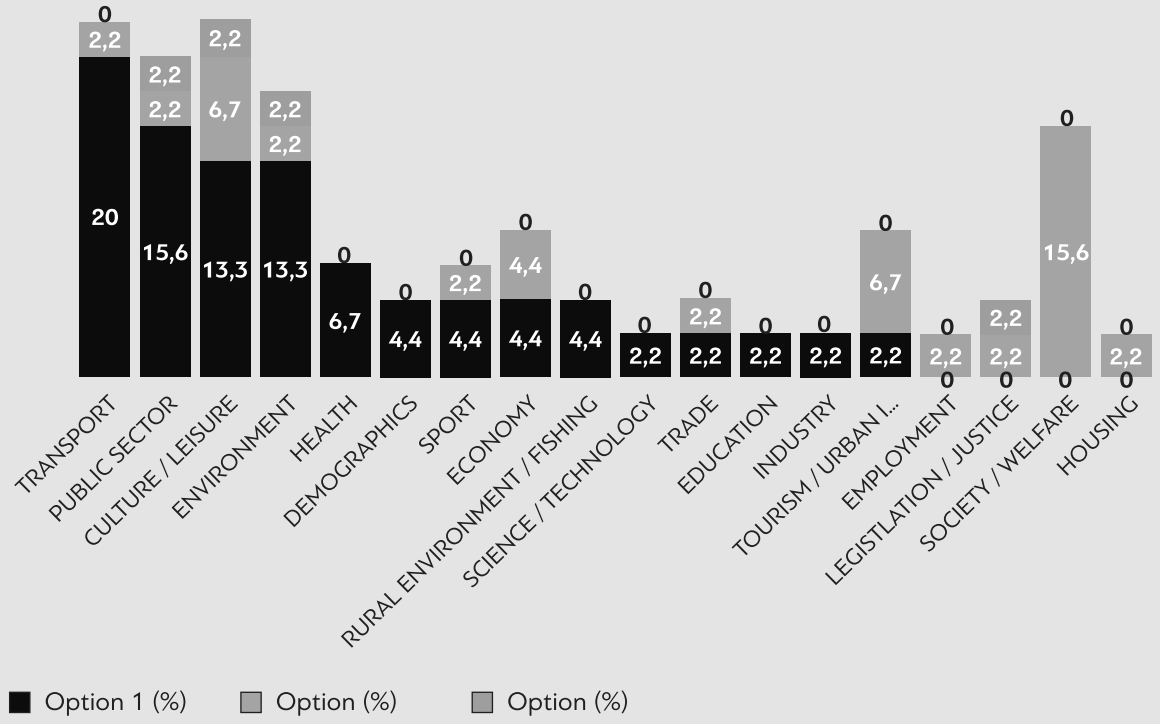
5.4.1 Analysis of service themes

To obtain more detailed information on the generated services, an analysis of them was carried out based on the information obtained in the portals' datasets. 62 applications were sampled, and we observed that 17 of them were not valid (27.42%), that is, they had ceased to be operational. Therefore, the analysis was carried out on the 45 that were still active. For 2021, Table 11 and Figure 17 display the themes following the NTI-RISP classification of the generated applications. In some cases, there were several themes that have been collected as options 2 and 3. We can observe that the most frequent was transport (20%), followed by the public sector (15.6%), culture/leisure (13, 3%) and the environment (13.3%). In contrast, in 2019, the most frequent categories were culture and leisure (14.5%), followed by the environment (12.9%), transport (12.9%) and urban planning and infrastructure (9.7%).

Table 11: NTI-RISP application themes in 2021 (Source: The authors)

NTI-RISP	Option 1	Option 1 (%)	Option 2	Option 2 (%)	Option 3	Option 3 (%)
Transport	9	20,0	1	2,2	0	0,0
Public sector	7	15,6	1	2,2	1	2,2
Culture/leisure	6	13,3	3	6,7	1	2,2
Environment	6	13,3	1	2,2	1	2,2
Health	3	6,7	0	0,0	0	0,0
Demographics	2	4,4	0	0,0	0	0,0
Sport	2	4,4	1	2,2	0	0,0
Economy	2	4,4	2	4,4	0	0
Rural environment/fishing	2	4,4	0	0,0	0	0,0
Science/technology	1	2,2	0	0,0	0	0,0
Trade	1	2,2	1	2,2	0	0,0
Education	1	2,2	0	0,0	0	0,0
Industry	1	2,2	0	0,0	0	0,0
Tourism/urban planning/ infrastructures	1	2,2	3	6,7	0	0,0
Employment	0	0,0	1	2,2	0	0,0
Legislation/justice	0	0,0	1	2,2	1	2,2
Society/welfare	0	0,0	7	15,6	0	0,0
Housing	0	0,0	1	2,2	0	0,0
No response	1	2,2	21	46,7	41	91,1
Total	45	100	45	100	45	100

Figure 17: NTI-RISP application themes in 2021.



Source: The authors

5.4.2 Analysis of service sustainability and business model

We also analysed whether the developed services had any sustainability mechanism in place. In 2021, 31.1% of applications did not have any sustainability mechanism, institutional sponsorship or income generation mechanisms, be that through direct, indirect or freemium payments or advertising. This figure was 30.6% in 2019 and 52% in 2017. On the other hand, we also observed that in 2021, 68.9% of services had an identified business model, while this only occurred in 19.4% of cases in 2019, and in approximately 25% of cases in 2017.

5.4.3 Analysis of service generators

A notable aspect is that in most cases the generators of services are public organisations, both those that publish data and those that are related: 40% in 2021, 61.2% in 2019 and 43% in 2017. These are followed by professional reusers (companies): 35.6% in 2021, 24.2% and 30% in 2017. Citizens also have a growing presence with 20% in 2021, compared to 9.7% in 2019 (Figure 18).

Public administrations are the greatest generators of open-data-based services (41%).

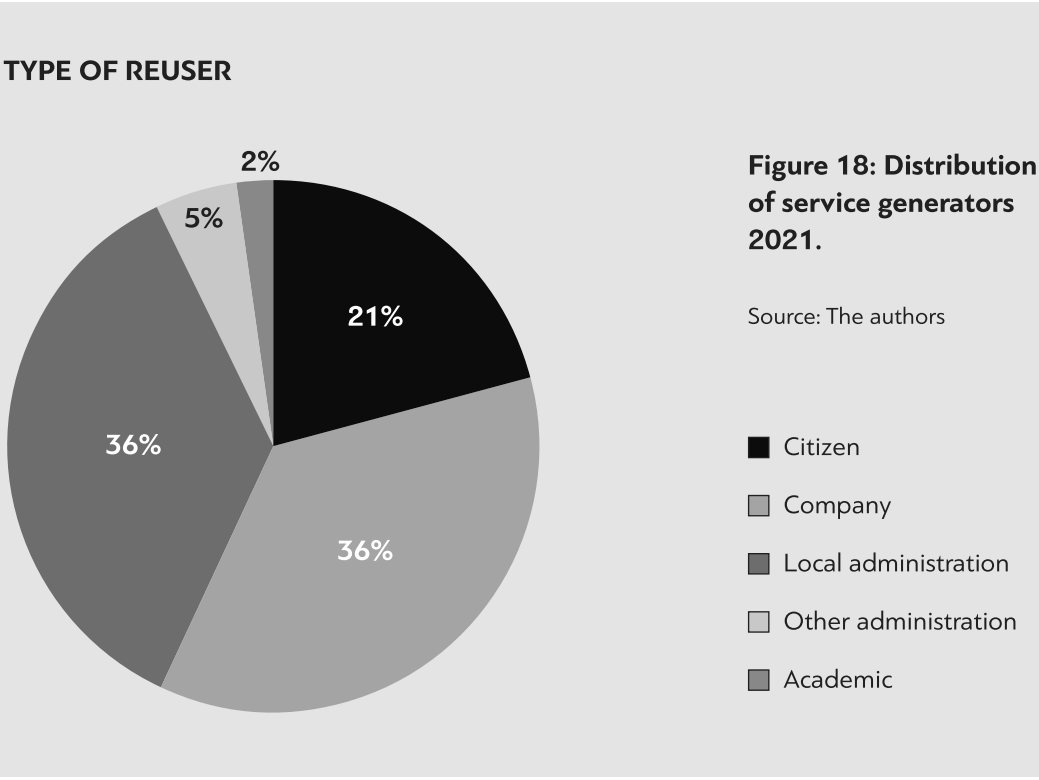


Figure 18: Distribution of service generators 2021.

Source: The authors

5.4.4 Analysis of other service characteristics

In 2021, 73.3% of generated applications were geolocated, while 42.2% provided information in real time. In general, there was a high tendency for services to be geolocated (74.2% in 2019 and 72% in 2017). However, only 25.8% (2019) and 35% (2017) were services with real-time information.

5.4.5 Analysis of the value creation of services

To analyse value creation, portals were asked to what extent the reuse of data has created value in the following areas, both for the entities that use the data and for their clients/users (Table 12).

One of the most relevant aspects when analysing the reuse of open data is the value creation that can be achieved. However, carrying out an analysis of value creation is not an easy process, and in many cases is a pending task.

In the survey carried out, the percentages obtained for the "not known" option or "no response" stand out: 51.8% in 2017 and 56.1% in 2019 for client satisfaction, although in 2021 there was a drop to 30.60%. For citizen satisfaction: 40.7% in 2017, 31.7% in 2019 and 13.90% in 2021. In the case of environmental improvement: 55.5% in 2017, 58.5% in 2019 and 27.80% in 2021. For the improvement of infrastructures: 55.5% in 2017, 46.3% in 2019 and 27.80% in 2021. The highest figures were for the creation of companies: 66.6% in 2017, 56.1% in 2019 and 41.70% in 2021; and for others: 85.2% in 2017, 92.7% in 2019 and 83.40% in 2021.

In 2017, 37% of those responsible for portals considered that citizens frequently or always felt satisfied with the publication of data, while in 2019 that

Table 12: Creation of value through the reuse of data

Scale		Never	Hardly ever	Sometimes	Frequently	Always	Not known/ No response	Total
Satisfied clients	2017	11,10%	3,70%	3,70%	18,50%	11,10%	51,80%	100%
	2019	2,40%	4,90%	12,20%	19,50%	4,90%	56,10%	100%
	2021	8,30%	13,90%	19,40%	25,00%	2,80%	30,60%	100%
Satisfied citizens	2017	7,40%	7,40%	7,40%	25,90%	11,10%	40,70%	100%
	2019	2,40%	7,30%	19,50%	29,30%	9,80%	31,70%	100%
	2021	5,60%	13,90%	41,70%	25,00%	0,00%	13,90%	100%
Environmental improvement	2017	14,80%	0%	11,10%	18,50%	0%	55,50%	100%
	2019	4,90%	7,30%	17,10%	9,80%	2,40%	58,50%	100%
	2021	11,10%	13,90%	36,10%	11,10%	0,00%	27,80%	100%
Improvement of infrastructures	2017	7,40%	0%	7,40%	14,80%	14,80%	55,50%	100%
	2019	7,30%	9,80%	7,30%	22%	7,30%	46,30%	100%
	2021	16,70%	13,90%	27,80%	14%	0,00%	27,80%	100%
Creation of companies	2017	14,80%	0%	18,50%	0%	0%	66,60%	100%
	2019	7,30%	7,30%	19,50%	9,80%	0%	56,10%	100%
	2021	16,70%	16,70%	13,90%	11,10%	0%	41,70%	100%
Improvement of public administration services	2017	7,40%	3,70%	33,30%	11,10%	22,20%	22,20%	100%
	2019	2,40%	4,90%	19,50%	24,40%	22%	26,80%	100%
	2021	8,30%	11,10%	38,90%	22,20%	8%	11,10%	100%
Others	2017	11,10%	0%	3,70%	0%	0%	85,20%	100%
	2019	2,40%	0%	2,40%	2,40%	0%	92,70%	100%
	2021	8,30%	0,00%	8,30%	0,00%	0,00%	83,40%	100%

figure rose to 39.1% and then dropped to 27.8% in 2021. The improvement of public administration (AAPP) services is the criterion for which most information is available. We found that in 2017, in 33.3% of cases, improvements were frequently or always achieved. That figure stood at 46.4% in 2019, and at 30.5% in 2021. Nevertheless, the reuse of open data sometimes (18.5% in 2017, 19.5% in 2019 and 13.9% in 2021) lead to the creation of companies, becoming frequent in 11.1% of cases in 2021, according to the data.

Own production based on Abella, Ortiz-de-Urbina-Criado & De-Pablos-Heredero (2017), Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 survey.

About 30% of the portals think that reusing data always or frequently leads to a public administration improvement.

6

Qualitative estimation of innovative services

6.1 Types of business models identified

Five business models have been identified in the sampled services:

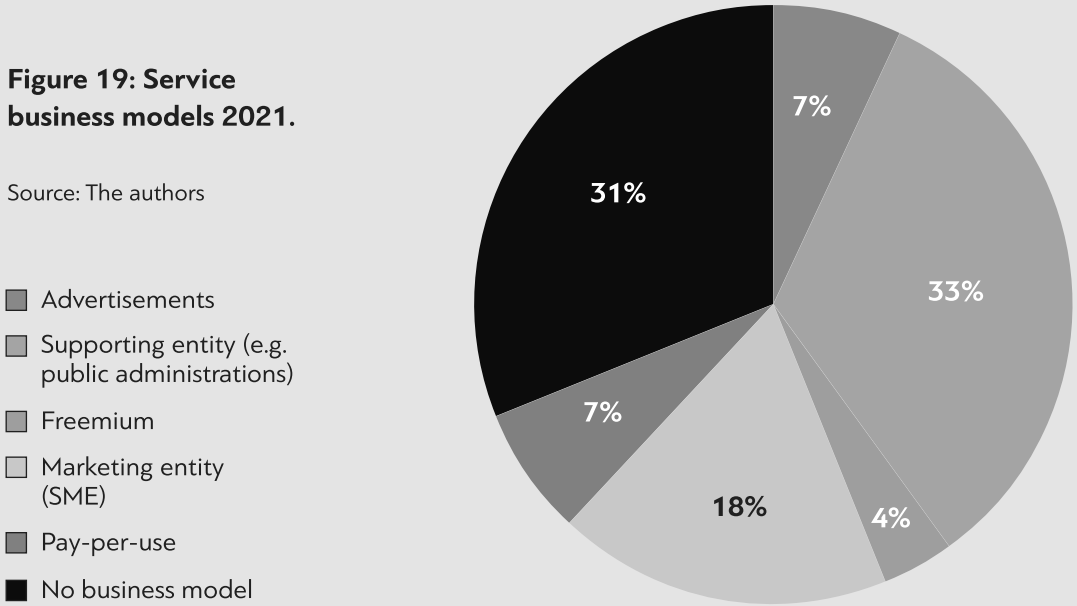
- Promotional (marketing entity)
- Supporting entity
- Freemium
- Pay-per-use
- Advertisements and additionally
- Non-profit

Figure 19 shows the results obtained in 2021. Regarding the 2019 report (Figure 20), there is an increase in the number of cases with a business model (31, 68.9%) and 33.3% with a promotional/marketing entity model. In 2019 we found 12 cases, of which 66.7% were promotional models and for the rest of the models, a percentage of 8.3% was obtained in all cases.

BUSINESS MODELS

Figure 19: Service business models 2021.

Source: The authors



Source: The authors

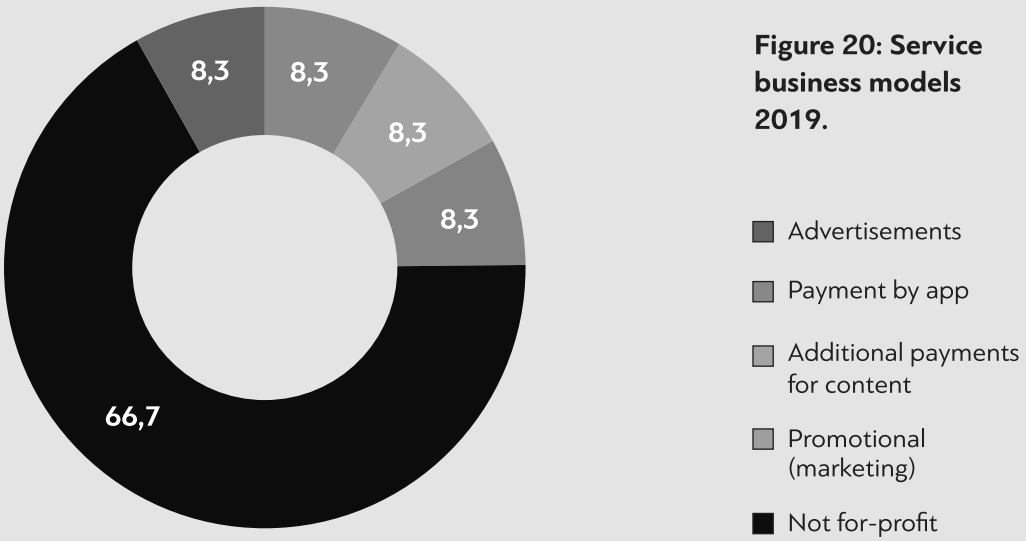


Figure 20: Service business models 2019.

Source: Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019)

6.2 Analysis of services by business model

In 2021, 74.2% compared to 83.3% in 2019, and 87%, in 2017, of the services with a business model are geolocated and only 41.94% of them are in real time (in 2019 it was 58.3% and in 2017 it was 66%).

Of the applications with a business model, in 2021, transport (22.58%), environment (16.13%) and public sector (12.9%) stand out. While in 2019, the most frequent are health (25%), economy, environment, and transport (16.7%) and in 2017, transport (47%) and meteorology (27%).

7 SWOT

Based on the analyses carried out, threats and opportunities, as well as strengths and weaknesses have been identified. In the SWOT presented in Table 13, the internal environment has been considered to include the managers of the portals and the reusers of all types of data, while the external environment is made up of society as the recipient of the services and the public managers who allocate the resources to keep the data portals up and running. Table 13 compares to the SWOT of the 2019 report (Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena, 2019).

Table 13: SWOT Analysis

Weaknesses 2021	• Lack of use of data portal management tools (DMS) in less mature portals. • The lack of digitisation of data and processes. • The low capacity for data reuse. • Poor updating of data. • Difficulty in identifying truly federated data in different domains. • There is still a	percentage of published datasets that are not reusable due to the technical format of publication. • Almost half of the datasets do not have any kind of geographic information. • More than a quarter of the data portals have a low level of reputation. • No dataset has an	advanced degree of reuse. • There is a percentage of open data managers who still do not know whether or not they are reusing data from their portals. • Those responsible for open data portals are not aware of the product and process innovations that are	generated with their data. • Difficult contact with those responsible for the portals. • 36% of the portals do not know if their data is accessed or not. • High closing rate of open portals. • High drop-out rate of open data-based services.
Weaknesses 2019	• Lack of use of data portal management tools (DMS) in less mature portals. • Weakly geolocated information and with low update frequency.	• Limited knowledge of reusers • Registration of access and knowledge of the use of published data. • Lack of systematics in promoting the use of	published data. • Percentage of services with sustainable business models. • Lack of awareness by portal managers of value creation	Misalignment between the characteristics and themes of the data published and the services generated.
Threats 2021	• Failure to maintain clear open data policy specific legislation. • Current data protection regulations.	• Pretender portals: Portals with political momentum but no strategy or prior planning.	• The lack of regulations for the use of shared data models for publication.	• Lack of business models for services based on open data.
Threats 2019	• Lack of demonstration of impact and its effect on financing.	• Consideration of the amount of data, but not its use for resource allocation.	• Misaligned objectives between the organisation and the data portal managers.	• Difficulty demonstrating innovation generated by data reuse.

Strengths 2021	• The knowledge of data reusers by data portal managers has improved substantially. • At local and regional levels, the reuse of data has increased. • Adequate metrics are in place to measure the reuse capacity of open data in portals (MELODA 5). • Increasing the maturity level of open data	portals. • Almost all portals publish data under licenses that allow commercial or unrestricted reuse. • About 40% of the datasets are on portals with a high reputation. • Citizens and the publisher itself have increased the number of services developed with open data.	• The number of services that have an identified business model has increased. • An interesting percentage of portals think that reusing data leads to frequent or permanent improvements in public administrations. • A high percentage of the datasets are in the highest category with proactive dissemina-	tion. • Existence of a reputation ranking of open data portals.
Strengths 2019	• Maturity of the open data sector in Spain.	• Most datasets are found on portals with a higher degree of maturity.	• Dissemination of data by academic reusers.	
Opportunities 2021	• The Covid-19 pandemic has emphasised the real need for data. • Greater dissemination of data thanks to the trend towards federated open data.	• Agile standardisation enables standards to be adopted quickly. • Specific European funding for open data.	• Dumping of statistical data on open data portals • Increased demand for data by society (digital literacy).	
Opportunities 2019	• Greater communication with reusers. • Coordination of publication activities and promotion of the use of all portals at	national level. • Development of data-driven services for business start-ups. • Development of more data-driven services	to meet citizens' expectations.	

Source: The authors based on Abella, Ortiz-de-Urbina-Criado, De-Pablos-Heredero, Vidal-Cabo & Ferrer-Sapena (2019), along with the results of the 2021 report

8 Conclusions

8.1 Conclusions about portals

There have been 41 portal cancellations, yet there are 10 new portals, which means the birth of 51 portals in the two-year period. The General State Administration has consolidated many portals on datos.gob.es and is responsible for a large part of these cancellations.

The number of portals that do not use a specific tool for this purpose (DMS) is more than 60%, which limits the possibilities for users to reuse their data.

In addition, the percentage of portals that are assessed as having inadequate performance for reuse still exceeds 50% according to the study methodology, although this figure is down from 62% in 2019.

45 portals (15.6%) have been identified that federate data among themselves (they replicate links to data), which initially facilitates their dissemination.

Assessing the reputation of open data publishers was one of the recommendations made by an international group of experts for MELODA version 5. As rankings on this aspect were not available in this study, a survey to portal managers, about the reputation of their portals, was carried out. Their results show an even distribution between benchmark portals (38%), mature portals (39%) and emerging portals (23%).

8.2 Conclusions on data

Data federation has evolved from being a one-off phenomenon to a regular occurrence. This means that the same dataset can be accessed from different points, making it difficult to identify the replica.

There is a massive publication of data from statistical sources on open data portals, 27.5% of the total.

Despite the maturity of the sector, there is still a high percentage of data in formats not suitable for reuse (6.7%) and data with licenses that do not allow commercial reuse (5%).

Almost 80% of the published data does not contain the description of the data model used (it does not follow a public standard), which makes it difficult to consolidate common data from different sources.

Integration with spatial data infrastructures (SDI) is increasingly common.

8.3 Conclusions on portal managers

It is worth highlighting that data managers have significantly improved their knowledge of reusers (97%) compared to 75% in 2019. However, the number of managers who do not have access to their data reuse statistics remains constant (36%), a percentage similar to previous editions.

Both own events (44%) and external events (42%) are the most popular mechanisms disseminating the data on the open data portals.

8.4 Conclusions on services generated

The most frequent topics of the services generated are culture and leisure (22.2%), transport (22.2%), public sector (20%) and environment (17.7%).

The services generated are published in real time in 42% of cases, a clear improvement from 26% in 2019, and 73% are geolocated, a similar percentage to 2019 (74%).

Applications' federation, similarly, but on a smaller scale than in datasets, also occurs.

The number of services without a business model reaches 31%. In terms of authors, there is a tie between the companies responsible for 36% of the services and the administrations themselves, another 36%. It is worth noting the increase of citizens, who reach 20% of the services compared to 9.7% in 2019.

9 Recommendations

- **Unique datasets identifier.** Data federation allows the republication of datasets on different portals. The lack of a unique identifier and explicit mechanisms to identify federated datasets makes it impossible to obtain a true picture of the data that a portal publishes, and globally of the actual number that is published.

- **Pretender portals.** Still 50% of open data portals have minimal reuse features, a very low number of datasetss and extreme difficulties in interacting with portal managers.

- **Data standardisation.** 80% of the published data do not follow an international benchmark and do not publish their own data model, which limits their possibilities of reuse, and only 2% follow an international standard. Agile standardisation initiatives[7] allow data models to be documented as a preliminary step to standardisation by classical standardisation bodies.

- **Rising of *data spaces*.** Data spaces are data exchange mechanisms and places where data is exchanged both openly and for a fee. Data spaces are widely promoted in the European Union[8] and open data portals could become data providers for these spaces.

- **Dataset consolidation.** In some portals, out-of-logic fragmentations in datasets have been detected. This means that the same dataset is divided, for example, by years and by categories of the same item, which artificially multiplies the number.

- **Approach of clear strategies in Spain for the development of the sector.** The need, following the example of other EU countries, to establish a clear strategy for portals, taking into account their expected use. It is recommended to work with open innovation tools to better understand the needs of reusers and to better match data to these needs.

7 For example, the Smart Data Models Program (https://smartdatamodels.org).

8 For example, the GAIA-X initiative. https://www.data-infrastructure.eu is specifying, among other elements, an open architecture of data spaces.

10

Future lines of work

- **Continuous collection of open data in Spain.** The work of the report provides a more up-to-date and detailed list of open data portals, their contents, and the potential for reuse of the open data published on them. The reports that have been done so far are biannual (2017, 2019, 2021), but it is within reach to develop a way to automate it and offer it continuously or at least more frequently. Some developments[9] allow this work to be carried out with limited resources.

- **Applications and services that use open data.** Their impact analysis and the innovation they generate is an unexplored area. In this regard, Jaime Obregon has carried out a first analysis of apps in the public sector,[10] in which he highlights the lack of exploration that exists.

- **Covid data.** The information available can be used to analyse how open data on Covid is doing and how it has helped or will help decision-making. This territory has hardly been explored although there are already some studies on the subject, such as https://www.datadista.com/coronavirus/.

- **Data for sustainability.** The circular economy and climate change are indisputable trends in terms of their necessity. However, analysing its effects, especially, in relation to the Sustainable Development Goals (SDGs), is an issue to be developed as there is still no consensus on how to assess it and what data to use. Therefore, at the national level it is interesting to know what is being published as open data. Moreover, in the private sphere from a corporate social responsibility point of view, there is also no consensus on how to translate this into common metrics to determine the progress of private entities.

9 IDRA https://github.com/FIWARE-GEs/idra allows the aggregation of different open data sources.

10 https://www.businessinsider.es/hacker-charlataneria-tecnologica-politicos-886683

11

Bibliography

Abella, A.; Ortiz de Urbina Criado, M.; De Pablos Heredero, C. (2014). "MELODA, a metric to assess open data reuse". *El profesional de la información*, 23(6), 582-588. https://doi.org/10.3145/epi.2014.nov.04

Abella, A.; Ortiz de Urbina Criado, M.; De Pablos Heredero, C. (2017). *La reutilización de datos abiertos: Una oportunidad para España*. COTEC. Report prepared as a result of the project selected within the 2016 call of the Open Innovation Programme (PIA) of the Cotec Foundation for Innovation. ISBN: 978-84-929-3338-9. Available at http://informecotec.es/media/INFORME_REUTILIZACION-DE-DATOS.pdf.

Abella, A.; Ortiz de Urbina Criado, M.; De Pablos Heredero, C. (2019). "MELODA 5: A metric to assess open data reuse". *El profesional de la información*, 28 (6), e280620. e-ISSN: 1699-2407. https://doi.org/10.3145/epi.2019.nov.20.

Abella, A.; Ortiz de Urbina Criado, M.; De-Pablos Heredero, C.; Vidal Cabo, C.; Ferrer Sapena, A. (2019). "La reutilización de datos abiertos en España II". DesideDatum. Available at: https://www.desidedatum.com/wp-content/uploads/2019/11/La_reutilizacio%CC%81n_datos_abiertos_espan%CC%83a_2019.pdf.

Alamo, T.; Reina, D.G.; Mammarella, M.; Abella, A. (2020). "Covid-19: Open-Data Resources for Monitoring, Modeling, and Forecasting the Epidemic". Electronics, 9 (5), 827. https://doi.org/10.3390/electronics9050827.

ASEDIE (2021). *2021 Sector infomediario*. Report. Available at: https://www.asedie.es/es/informes-anuales.

Aporta (2021). Reports. Available at: https://datos.gob.es/es/blog-tags/iniciativa-aporta

Cadena, S. (2019). "Marco de referencia para la publicación de datos abiertos comprensibles basados en estándares de calidad". Available at: http://rua.ua.es/dspace/handle/10045/102507.

Carrara, W.; Nieuwenhuis, M.; Vollers, H. (2016). *Open data in Europe 2016. Insights into the European State of Play.* European Commission. Directorate General for Communications Networks, Content and Technology. European Data Portal. https://www.europeandataportal.eu/sites/

default/files/edp_landscaping_insight_report_n2_2016.pdf.default/files/edp_landscaping_insight_report_n2_2016.pdf

Cecconi, G.; Radu, C. (2018). *Open data maturity in Europe.* European Commission report. Available at: https://www.europeandataportal.eu/sites/default/files/edp_landscaping_insight_report_n4_2018.pdf

Cetina, C. (2021). "Gobernanza de datos y capacidades estatales para la pos-pandemia". Policy Brief #28, The Andean Development Corporation (CAF). Available at: https://scioteca.caf.com/handle/123456789/1765

European Commission. (2020a). *Digital Economy and Society Index Report 2020*. Available at: https://ec.europa.eu/newsroom/dae/document.cfm?-doc_id=67079

European Commission. (2020b). *Open Data Maturity Report 2020*. Available at: https://data.europa.eu/sites/default/files/edp_landscaping_insight_report_n6_2020.pdf

European Commission. (2020c). *Analytical Report 16. Open Data Best Practices in Europe: Learning from Cyprus, France, and Ireland.* Available at: https://europeandataportal.eu/

European Union (2017). "Open Data Maturity in Europe 2017". Available at: https://data.europa.eu/en/highlights/open-data-maturity-europe-2017

Ferrer Sapena, A.; Calabuig, J.M.; Sánchez Pérez, E.A.; Vidal Cabo, C. (2020): "Gobierno abierto y acceso a la información: un estudio de caso sobre el impacto en la economía local". BiD: textos universitaris de biblioteconomia i documentació, 45- https://doi.org/10.1344/BiD2020.45.17.

Gil-García, J. R.; Gasco Hernández, M.; Pardo, T. A. (2020). Beyond transparency, participation, and collaboration? A reflection on the dimensions of open government. *Public Performance & Management Review*, 43(3), 483-502. https://doi.org/10.1080/15309576.2020.1734726

Hrustek, L.; Furjan, M. T.; Pihir, I. (2021). "Political Participation in the Information Society and Impact of Open Data on It". In 2021 IEEE Technology & Engineering Management Conference-Europe (TEMSCON-EUR) (p. 1–6). IEEE. https://ieeexplore.ieee.org/document/9488626

Marcos Martín, C.; Soriano Maldonado, S. L. (2011). "Reutilización de la información del sector público y open data en el contexto español y europeo. Proyecto Aporta". *Profesional de la información*, 20 (3), 291–297. https://doi.org/10.3145/epi.2011.may.07

Ministry of Territorial Policy and Public Function (2020). IV Open Government Plan. Available at: https://transparencia.gob.es/transparencia/dam/jcr:d306cd62-cc0f-40a1-9be8-fe24eeeee10d/IVPlanGobiernoAbierto_ES_vdef.pdf.

International Observatory of Telecommunications and the Information Society (ONTSI) (2021). Informes. Reports. Available at: https://www.ontsi.red.es/es/estudios-e-informes/Informes%20anuales

OECD (2021). OECD Recommendation on Enhancing Access to and Sharing of Data (EASD). Available at: //legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0463

Open Knowledge Foundation (2021). "What is open data?". Available at: https://opendatahandbook.org/guide/es/what-is-open-data/

Red.es (2021). Datos abiertos. Available at: https://sede.red.gob.es/datos-abiertos

Walter, M.; Lovett, R.; Maher, B.; Williamson, B.; Prehn, J.; Bodkin-Andrews, G.; Lee, V. (2021). "Indigenous Data Sovereignty in the Era of Big Data and Open Data. *Australian Journal of Social Issues*, 56 (2), 143–156. https://doi.org/10.1002/ajs4.141.

Web Foundation (2018). Open Data Barometer – Leaders Edition. Washington. DC: World Wide Web Foundation. Available at https://opendatabarometer.org/doc/leadersEdition/ODB-leadersEdition-Report.pdf

Worthy, B. (2015). "The Impact of Open Data in the UK: Complex, Unpredictable, and Political". *Public Administration*, 93 (3), 788–805. https://doi.org/10.1111/padm.12166.

Zuiderwijk, A.; Pirannejad, A.; Susha, I. (2021). "Comparing open data benchmarks: Which metrics and methodologies determine countries' positions in the ranking lists?". *Telematics and Informatics*, 62, 101634. https://doi.org/10.1016/j.tele.2021.101634.

12 Annexes

12.1 Questionnaire to reusers

Scale 1-5, where 1: not at all / never and 5: all / always	1	2	3	4	5
1. Do you know of any organisations that reuse the data they publish on their open data portals?					
Companies					
NGOs					
Academics					
Students					
Individual citizens					
2. Indicate the extent to which each type of these entities reuse your data	1	2	3	4	5
Individual citizens					
Professional for-profit reusers (infomediaries and individual for-profit developers					
Professional non-profit reusers (NGOs, foundations, individual non-profit developers and other social initiatives					
Researchers and academic staff (including students)					
The organisation itself that publishes the data (other areas or departments)					
Other public entities than the one publishing the data					
3. To which sector do these entities belong? Choose the five that reuse the most	1	2	3	4	5
Agriculture, Livestock, Forestry and Fishing					
Extractive industries					
Manufacturing industry					
Electricity, gas, steam and air conditioning supply					

Scale 1-5, where 1: not at all / never and 5: all / always	1	2	3	4	5
Water supply, sanitation, waste management and decontamination activities					
Construction					
Wholesale and retail					
Transport and storage					
Hospitality and tourism					
Information and communications					
Financial and insurance activities					
Real estate activities					
Professional, scientific and technical activities					
Administrative activities and ancillary services					
Public Administration and Defense					
Education					
Health and social work activities					
Artistic, recreational and entertainment activities					
Activities of households as employers of domestic workers					
Activities of extraterritorial organisations and bodies					
4. Indicate the extent to which reusers operate in the following areas	1	2	3	4	5
Local (city)					
Autonomous region					
National					
European					
Others					

Scale 1-5, where 1: not at all / never and 5: all / always	1	2	3	4	5
5. Indicate the extent to which your organisation has carried out any kind of activity to promote the use of your data	1	2	3	4	5
Application contests					
Own events (organised by your organisation) to present the available data					
External events (organised by other organisations) to present the available data					
Meetings with reusers (of any kind)					
Others (indicate in the notes at the end of the survey)					
6. Indicate the extent to which these types of innovations are produced by the reuse of data	1	2	3	4	5
Product innovation					
Process innovation					
7. Indicate to what extent the reuse of data has created value in the following areas for the organisations using the data or for their customers / users	1	2	3	4	5
More satisfied customers					
More satisfied citizens					
Environmental improvement					
Improvement of infrastructures: safety, health, transport, etc.					
Business creation					
Improvement in the services offered by the public administration					
Others					

8. Could you indicate the three most common uses of your portal data by reusers?
1
2
3
9. Could you indicate which three datasets (in order of most accessed to least) are the most used from your portal?
1
2
3
10. Data is available from the reusers access point
NOTES (any other comments you wish to make)
The questions below this line are to increase the scope of study.
An association of open data repositories would be useful.

Do you belong to an organisation similar to an association of open data repositories/publishers? Indicate which.
Given the list of respondents. Are you missing any other open data portal? Include a reference to them (URL or name of the website or organisation or contact email).

12.2 Questionnaire to managers of open data portals on the reputation of published data

Reputación de portales de datos abiertos

Esta encuesta pregunta por la repurtación de los datos de portales de datos abiertos de España para el III informe sobre el estado del open data en España.

Sólo se pregunta por un grupo de portales (10).

El conocimiento de la entidad se refiere a su familiaridad con el portal sobre el que se le pregunta

- ***Alto conocimiento del portal***: Vd conoce bien el portal de datos y sus prestaciones. P.e. lo usa con frecuencia y ha descargado sus datos y los ha utilizado.
- ***Medio conocimiento del portal:*** Vd conoce el portal de datos y sus prestaciones. Lo ha usado puntualmente y/o alguna vez ha utilizado sus datos.
- ***Bajo conocimiento del portal***: No conocía su existencia o alguna vez lo ha visitado sin utilizar sus datos.

Se debe clasificar en tres niveles la reputación

- ***Portal referente del sector***.- Los datos del portal son una referencia en el sector y/o aparecen como referencia reconocida en rankings del sector o en informes de reputación
- ***Portal maduro***.- Los datos del portal son conocidos y su calidad es valorada y/o se conocen las opiniones de sus usuarios
- ***Portal incipiente***.- No se tiene información sobre la reputación de los datos del portal

	Conocimiento del portal				Nivel de reputación				
	Alto conocimiento del portal	Medio conocimiento del portal	Bajo conocimiento del portal		Portal referente del sector	Portal maduro	Portal incipiente		Sin respuesta
http://www.torrecilladelosangeles.es/open-data	○	○	○		○	○	○		●
http://www.getxo.eus/es/gobierno-abierto/opndata	○	○	○		○	○	○		●
https://www.seu-e.cat/es/web/santcugatdelvalles/dades-obertes	○	○	○		○	○	○		●
https://seu-e.cat/es/web/santjustdesvern/dades-obertes	○	○	○		○	○	○		●
http://atlastenerife.es/portalwebide	○	○	○		○	○	○		●
https://datos.parcan.es/dataset?q=	○	○	○		○	○	○		●
https://idem.madrid.org/catalogocartografia/srv/spa/catalog.search#/home	○	○	○		○	○	○		●
https://info.igme.es/cartografiadigital/portada/default.aspx?mensaje=true	○	○	○		○	○	○		●

12.3 Ranking of portals according to reputation

Below is the list of portals ordered from low to high knowledge (survey results in Annexe 12.2). An average of the scores obtained for each portal has been calculated, taking into account that value 1 is low knowledge of the portal, 2 is medium and 3 is high knowledge of the portal.

Table 14. Ranking of portals by level of knowledge

Portal	Ranking by level of knowledge
[http://datos.gob.es]	3,00
[http://www.bcn.cat/opendata/]	2,57
[http://opendata.aragon.es]	2,50
[http://opendata.euskadi.EUS]	2,38
[http://www.opendatalapalma.es/]	2,00
[http://www.amb.cat/es/web/area-metropolitana/dades-obertes]	2,00
[http://centrodedescargas.cnig.es/CentroDescargas/]	2,00
[www.ine.es/datosabiertos]	2,00
[http://www.juntadeandalucia.es/datosabiertos/portal.html]	2,00
[http://dadesobertes.gencat.cat]	2,00
[http://datos.madrid.es]	1,89
[http://datosabiertos.malaga.eu/]	1,86
[http://www.gipuzkoairekia.eus]	1,78
[https://info.igme.es/cartografiadigital/portada/default.aspx?mensaje=true]	1,67
[http://www.santacruzdetenerife.es/opendata/]	1,67
[https://www.opendatabizkaia.eus]	1,67
[datos.gijon.es]	1,63
[https://datos.alcobendas.org/#]	1,56
[http://www.gobiernodecanarias.org/istac/datos-abiertos/]	1,50
[http://datosabiertos.sevilla.org/]	1,50

Continued on next page ›

Table 14. Ranking of portals by level of knowledge (continuation)

Portal	Ranking por nivel de conocimiento
[datosabiertos.torrent.es/]	1,50
[http://datosabiertos.laspalmasgc.es/]	1,50
[https://www.seu-e.cat/es/web/santcugatdelvalles/dades-obertes]	1,50
[https://idem.madrid.org/catalogocartografia/srv/spa/catalog.search#/home]	1,50
[http://www.bilbao.net/opendata/]	1,44
[http://www.gobiernoabierto.navarra.es/es/open-data]	1,43
[datosabiertos.malaga.es]	1,43
[http://abertos.xunta.es/portada]	1,43
[https://seu-e.cat/es/web/santjustdesvern/dades-obertes]	1,33
[http://www.puertodesantacruz.es/open-data]	1,33
[https://data-crtm.opendata.arcgis.com/]	1,33
[http://datos.ua.es/]	1,29
[http://zaguan.unizar.es/collection/OpenData?ln=es&as=1]	1,29
[https://www.donostia.eus/datosabiertos/]	1,29
[https://opendata.pamplona.es/]	1,29
[http://datosabiertos.castillalamancha.es/]	1,29
[https://www.tarragona.cat/dadesobertes]	1,29
[http://www.getxo.eus/es/gobierno-abierto/opndata]	1,25
[http://opendata.manresa.cat/]	1,17
[http://datos.icane.es/]	1,14
[https://datosabiertos.ayto-arganda.es/#]	1,14
[http://opendata.terrassa.cat/]	1,14
[https://web.larioja.org/dato-abierto]	1,00

Continued on next page ›

Table 14. Ranking of portals by level of knowledge (continuation)

Portal	Ranking por nivel de conocimiento
[https://seu-e.cat/es/web/castelldefels/dades-obertes]	1,00
[http://datosabiertos.regiondemurcia.es/]	1,00
[http://www.fega.es/es/PwfGcp/es/accesos_directos/datos_abiertos/index.jsp]	1,00
[http://www.torrecilladelosangeles.es/open-data]	1,00
[http://atlastenerife.es/portalwebide]	1,00
[https://datos.parcan.es/dataset?q=]	1,00
[http://datos.lorca.es/]	1,00
[opendata.dadesobertesmanlleu.cat]	1,00
[http://www.arroyomolinos.es/open-data]	1,00
[https://idechg.chguadalquivir.es/]	1,00
[http://www.botija.es/open-data]	1,00
[http://www.torreorgaz.es/open-data]	1,00
[http://www.jaraicejo.es/open-data]	1,00
[http://www.navasdelmadrono.es/open-data]	1,00

Source: The authors

Below is the list of portals ordered according to the variable measuring reputation (results of the survey in Annexe 12.2). An average of the scores obtained for each portal has been calculated, taking into account that value 1 is an emerging portal, value 2 is a mature portal and value 3 is a benchmark portal in the sector.

Table 15: Ranking of portals by level of reputation

Portal	Ranking by reputation
[http://datos.gob.es]	3,00
[http://dadesobertes.gencat.cat]	2,71
[http://opendata.euskadi.EUS]	2,63
[http://centrodedescargas.cnig.es/CentroDescargas/]	2,50
[http://opendata.aragon.es]	2,50
[http://www.bcn.cat/opendata/]	2,43
[datos.gijon.es]	2,40
[www.ine.es/datosabiertos]	2,29
[http://www.juntadeandalucia.es/datosabiertos/portal.html]	2,20
[https://datos.alcobendas.org/#]	2,17
[http://datos.madrid.es]	2,13
[http://www.amb.cat/es/web/area-metropolitana/dades-obertes]	2,00
[http://datosabiertos.malaga.eu/]	2,00
[https://data-crtm.opendata.arcgis.com/]	2,00
[http://www.gipuzkoairekia.eus]	1,80
[http://www.bilbao.net/opendata/]	1,80
[http://www.gobiernoabierto.navarra.es/es/open-data]	1,71
[https://idem.madrid.org/catalogocartografia/srv/spa/catalog.search#/home]	1,67
[https://www.opendatabizkaia.eus]	1,67
[http://datosabiertos.castillalamancha.es/]	1,67
[https://www.tarragona.cat/dadesobertes]	1,67
[http://www.opendatalapalma.es/]	1,50
[https://web.larioja.org/dato-abierto]	1,50
[http://www.gobiernodecanarias.org/istac/datos-abiertos/]	1,50
[http://datosabiertos.sevilla.org/]	1,50
[datosabiertos.torrent.es/]	1,50
[http://datosabiertos.laspalmasgc.es/]	1,50

Continued on next page ›

Tabla 15. Ranking of portals by level of reputation (continuation)

Portal	Ranking por grado de reputación
[https://seu-e.cat/es/web/castelldefels/dades-obertes]	1,50
[datosabiertos.malaga.es]	1,50
[https://www.donostia.eus/datosabiertos/]	1,50
[http://www.getxo.eus/es/gobierno-abierto/opndata]	1,50
[https://www.seu-e.cat/es/web/santcugatdelvalles/dades-obertes]	1,50
[http://abertos.xunta.es/portada]	1,50
[http://opendata.manresa.cat/]	1,50
[http://datos.icane.es/]	1,43
[https://datosabiertos.ayto-arganda.es/#]	1,43
[http://datos.ua.es/]	1,33
[https://opendata.pamplona.es/]	1,33
[http://www.santacruzdetenerife.es/opendata/]	1,33
[http://zaguan.unizar.es/collection/OpenData?ln=es&as=1]	1,29
[http://opendata.terrassa.cat/]	1,20
[http://datosabiertos.regiondemurcia.es/]	1,17
[http://www.fega.es/es/PwfGcp/es/accesos_directos/datos_abiertos/index.jsp]	1,17
[http://www.torrecilladelosangeles.es/open-data]	1,00
[https://seu-e.cat/es/web/santjustdesvern/dades-obertes]	1,00
[http://atlastenerife.es/portalwebide]	1,00
[https://datos.parcan.es/dataset?q=]	1,00
[https://info.igme.es/cartografiadigital/portada/default.aspx?mensaje=true]	1,00
[http://datos.lorca.es/]	1,00
[opendata.dadesobertesmanlleu.cat]	1,00
[http://www.arroyomolinos.es/open-data]	1,00
[https://idechg.chguadalquivir.es/]	1,00
[http://www.puertodesantacruz.es/open-data]	1,00
[http://www.botija.es/open-data]	1,00
[http://www.torreorgaz.es/open-data]	1,00
[http://www.jaraicejo.es/open-data]	1,00
[http://www.navasdelmadrono.es/open-data]	1,00

Source: The authors

12.4 Response on the three most common uses that reusers make of their portal data

Table 16: Responses from the three most common uses of portal data

Most common uses made by reusers 2017	• Cultural activities • Some newspaper to publish expenditure information • Parking • Search for information • Mapping • Mapping, urban planning, geo-positioning • Complete own databases • Citizen communication • Know the agenda of cultural and sports activities in the city • Consult information • Tourist data enquiries • Public budget consultations	• Administrative structure consultations • Contracting technical services • Creating maps • Comply with the Transparency Law • City planning data • Transport planning data • Mobile application development • Data statistics • State of services: beaches, parks, public transport • Market research • Management of points of interest • Bus schedules	• Academic information • Real-time information • Real-time traffic information: navigators, routes • Real-time information: public transport, management of public services • Budget information • Ranking information • Tourist information • Inform the public • Data journalism work • Tenders • Environmental • Mobility • Job offers • Offer transportation information	• Spatial planning • Route planner • Prices of fruit and vegetable products • Budgets • Implementation of applications • Reuse of information by the organisation itself • Transform your own organisation • Use of mapping for a variety of uses • Added value for existing applications or services
Most common uses made by reusers 2019	• Culture • News published in the media by journalists • Buses • Search for information • Mapping • Access to public information • Platform on Administrative Procedures • Cross data with information from other administrations • Leisure and free time, orientation for outdoor activities • Information enquiry • For information related to tourism	• Internal consumption in corporate applications and websites • Information on local government • Studies and reports • Tourism: Festivals, fairs • Creation/updating of directories of tourism companies in the region • Population statistics • Real-time urban transport data for apps and applications • Improving Apps • Population statistics • For studies with population data	• Better knowledge of the market in which companies participate • Citizen mailbox activity on social networks • GIS analysis and generation of derived products • Studies and research • Analysis of the information accessed via APIs • Mobile application consultation • Transport • Generation of studies and dashboards • Internal performance statistics • Data analysis • Water	• Data journalism • Integration into reports and publications • Environmental applications • Mobility • Use by micro-enterprises and SMEs • Planning transport routes around the city based on traffic information • Synchronisation of data related to municipal spatial planning • Organisation and planning of annual calendars related to fairs and festivals in the region

Continued on next page ›

	• Exploitation • Data visualisation • Creation of applications • Internal management within the institution itself • The organisation itself: municipal website	• Applications • New products, services and apps • Personal projects of citizens • Publication of third party data • Reuse of static datasets	• Master's Thesis for university students • Research and teaching Education (University, Secondary Education, Training courses) Scientific • API information consumption	• Improving the product they have by incorporating open data information • Private developments
Most common uses made by reusers 2021	• Academics, scientific or informative • Database update • Database feeding • Data analysis • Analysis and development of own content within our organisation • Analysis and proposal for the creation of new products/services • Geographic intelligence application for the city • Weather application • Trail applications • Computer applications • Health related applications (Air quality) • Tourist applications • Agri-food information apps • Press articles • Audit/control of municipal management • Air quality • Marketing campaigns • Inter-municipal comparisons • Add-on for solution development • Verification of the entity's solvency • Competitions, challenges	• Knowing the state of drought • Knowing where there is feasibility for water catchments • Getting to know the administration better: transparency • Knowledge of environmental regulations • Data query • Creation of apps • Website creation • Real-time public transport service data • University access and enrollment data for citizens in the decision-making process for career choice after the University Entrance Exam (EBAU) • Statistical data • Population and street statistics • Geographical data on accommodation during the course and national/international student mobility by student accommodation management companies • Teaching • Development of dashboards for the internal management of the city council • Mapping • Scales • Research studies • Statistical studies • Studies, analysis and geography • Evaluation and study • Underpinning policy action • Geolocation of noise • Management of services • Treasury • General information to the public through available visualisations • Graphic information on municipal budgets • Information on public services • Information on municipal procedures • Integrations of APIs in apps • Integration of visualisations in geographic portals • Location of services • Maps • Road improvement • Improved decision-making • Goods	• Organisation of production • Student internships • End of degree projects • Publication of news in the media • Accountability Municipal services-Transparency • Reuse for applications • Reuse for web viewing • Decision-making • Transport • Tourism • Use of traffic data for operators such as TomTom, Moovit, Here or Google Traffic • Using transport data to make a virtual assistant with Alexa • Visualisations

Source: The authors

12.5 Responses to the three most frequently used datasets on your portal

Tabla 17. Table 17: Comparison of the three most used datasets 2017, 2019, 2021

Datasets 2017	• Cultural and leisure activities • Agenda • Cultural agenda • Daily agenda • Tourist agenda • Cultural agenda information • Mayors of the municipalities of the province of Alicante Legislature 2015 • Buses • GTFS of bus schedules • La Palma Island Council bus timetable • List of lines • Public transport routers (bus and taxi) • Arrival time of buses • TITSA -- Transportation system information • Bike • Working calendar • Air quality • Mapping • Mapping information • Check timetables and frequency • Contracting technical services • Contracts awarded • Staff calls • Job offers • List of jobs	• Demographic • Minor local entities and districts of the Province of Alicante • State of the beaches • Beaches in the province of Cádiz • Default status • Public budget information • Budget • Administrative infrastructures (educational centres, agricultural offices, registry offices, etc.) • List of associations • Maps • Number plates	• Digital elevation model • Mobility • Selective collection points • Ranking • Educational resources • Results of union elections in the administration itself • Trails in the province of Cádiz • Municipal services and facilities • Traffic
Datasets 2019	• Agenda of activities (city, events, etc.) • General agenda of the Provincial Council • Directories • BOPB (Official Gazette of the Province of Barcelona) • Companies in the city • Minor local authorities • General database of local entities • Buses • City data, parking, cycling, etc. • Open data catalogue • Street map • Stations • Real-time information on TUVISA buses • Bicing stations in Barcelona city (mechanical and electric) • Bicycles • Number of incidents on public roads on the Green Line • Air quality datasets • Mapping • 'INAGA' GIS Mapping of Hunting Grounds of Aragon (in 2018) • Timetables • Recruitment • Contracts under tender • CIDO – Civil Service Examinations • Job offers • List of jobs • Local residents • Government of Aragon Contracts (in 2018) • Leisure • GTFS • Economic • Economic-financial ratios • Budget • General Administration Directory • Associations • National Topographic Map (scale 1:25,000) • Parking spaces • MDT5 • Graphic information bike lane • Citizen petitions • Electoral data of municipalities • Monuments and museums • Municipalities and counties • Lidar (altimetric information) • Incidents on public roads • Traffic datasets	• Resolutions of complaints and claims • Mobile Numbering Consultation Service (CNMC) • LPISQA • Information systems • Subsidies • Bus and API waiting time • Registered unemployment by municipalities • Weather • Most current orthoimage from the PNOA (National Aerial Orthophotography Plan) • Register of inhabitants per neighbourhood • Bus stops • Municipality population • Payments

Continued on next page ›

Datasets 2021				
• Economic activities • Actions of the municipal works brigade • Cultural agenda • City agenda • Bicing • Real Estate • Official Gazette of the Mercantile Registry (BORME) • Air quality • Street map • Mapping • Record of discharges • CIDO (Search Engine For Information and Official Journals) • Solvency check • Solid waste containers • Small contracts • Real-time data of the urban bus network • Covid data • Demographic data	• Statistical data • Geographical data • Real-time weather data • General Administration Directory • Administrative division of municipal districts • Municipal companies and businesses • Tourist tickets • Drought status • Evolution of coronavirus disease (Covid-19) • Official street names guide • Inhabitants per portal • Incidents on public roads • Real bike information • Real-time bus information • Complaints and suggestions report	• Inventory of CARM Senior Officials • Territorial limits • List of educational centres • Groundwater bodies • Number plates • Picnic areas • Monuments • Mobility • Municipalities in the province of Barcelona • Museums • Public offer of employment • Registration assistance offices • Payment to suppliers • Personnel-PAS • Parking spaces • Budgets • Points with water abstraction authorisation • Trail network	• Register of associations • Register of cooperatives • List of jobs • Repairs on public roads • Payments • Health • Headquarters of educational centres in the region of Murcia • Mobile Numbering Consultation Service (CNMC) • Traffic • Transit • Agricultural demand units • Protected areas	

Source: The authors


desideDatum
01100100 01100001 01110100 01100001
COTEC
Universidad
Rey Juan Carlos
esic
9 788418 944604