Colección Economía Digital

From infomediary sector to the data economy:

Characterization of the Infomediary sector. 2020 EDITION

EXECUTIVE SUMMARY

Iniciativa aporta

datos.gob.es
reutiliza la información pública

GOBIERNO DE ESPAÑA
VICEPRESIDENCIA TERCERA DEL GOBIERNO
MINISTERIO DE ASUNTOS ECONÓMICOS Y TRANSFORMACIÓN DIGITAL

ontsi observatorio nacional de las telecomunicaciones y de la SI

red.es

2020 EDITION

From infomediary sector to the data economy:

Characterization of the Infomediary sector

EXECUTIVE SUMMARY

The report: *From the infomediary sector to the data economy: characterization of the Infomediary Sector. Executive Summary*. *2020 Edition* has been prepared by the ONTSI team:

Ricardo Vázquez Martínez (Coordination)

María Martínez López

And the technical assistance of:

ICLAVES

Edit: Ministry of Economic Affairs and Digital Transformation

How to cite this report:

ONTSI (2020). *From the infomediary sector to the data economy: characterization of the Infomediary Sector. Executive Summary. Ministry of Economic Affairs and Digital Transformation*

Contents

1 EXECUTIVE SUMMARY

1.1 Infomediary sector characterisation

Infomediary sector delimitation

The infomediary sector is made up by very heterogeneous companies. Ranging from well-established business which use information to elaborate their products and services for many years, to start-ups specialised on new digital technologies such as Artificial Intelligence and Big Data.

The development of the data economy blurs the lines of the reusing of information. Thus, the infomediary sector delimitation is a growing challenge.

The first great difficulty characterising the sector is its own definition. After updating the census indirectly, as mentioned in the methodology, the infomediary sector has been set at **708 companies**.

Infomediary sector main characteristics

The infomediary sector is highly concentrated in two Spanish regions, Community of Madrid, and Catalonia. Half of companies are located in Madrid, while 18.4% are based in Catalonia.

Three quarters of the companies operate under the legal form of limited liability companies. Another 20.8% are public limited companies.

Regarding the age of the enterprises, **63.7% are active for more than 10 years**. More than a third are over 20 years old, showing the consolidation of the sector. However, it is not a stagnated sector. 36.2% of companies are less than 10 years old. The newest companies have emerged as response to the growing interest in the reuse of information as a business activity.

Most of infomediary companies (46.2%) are framed in two business codes (CNAE): **advertising and market studies (CNAE 73) and programming, consulting and other activities related to informatics (CNAE 62)**. The evolution of the CNAE according to the age of the companies shows how the younger ones tend to concentrate on more technological activities linked to informatics.

For the first time, the study has analysed the participation of foreign investment in the sector. The presence of foreign investment is still low, as only **8.4% of the companies declare that they have foreign shareholders**.

Activities developed by infomediary enterprises

Infomediary activity constitutes the **main line of business** for **57.5%** of the companies.

In order to analyse the positioning of the companies according to their infomediary activity, it has been split in four main business processes:

- Own data production
- Aggregation, processing and quality improvement of data sources
- Data analysis consultancy
- Data analysis and visualization

The business process in which the highest percentage of companies are focused is the **analysis and/or visualisation of data (65.5%)**, followed by the **aggregation, processing and quality improvement of data sources (52.9%)**. It is important to note that the business processes identified are not mutually exclusive. 73.5% of companies stated that their activity covered more than one of the defined business processes.

Regarding these four business processes, companies' activities materialise in specific value propositions, aimed at providing products and services to concrete sectors, or transversal solutions for any sector. **Research and consultancy** is the activity to which the highest percentage of companies declare to be dedicated (43.2%), followed by the provision of **corporate decision-making services** (32.1%) and the carrying out of **public opinion surveys and market research (23.5%).**

The analysis of the products and services offered by infomediary companies allows concluding that **business relations** are quickly **evolving** toward **services delivery models**. After a detailed revision of the solutions categories, the most sold ones are **processed data** (including aggregation and validation services), commercialised by **56.8%** of companies, **publications (55.6%) and analysis and strategic decision-making support services (50.6%)**.

Even though the way of commercialising solutions by infomediary companies is evolving towards service delivery, **revenue models** still depend on the individual sale of products. In fact, the most used revenue model is the **one-time payment for deliverable** (maps, databases, report, etc.). This model is used by **45.7%** of enterprises. **Subscription fees**, more linked to service delivery, are used by **22.2%** of infomediary companies.

Infomediary products and services are mainly provide in Spanish (96.3% of companies). Nevertheless, more than half of the enterprises (54.3%) also offer them in English. Among co-official languages in Spain Catalan stands out, with 17.3% of companies providing services and products in this language.

The infomediary activity analysis ends with the characterisation of companies according to the time they have dedicated to it. **Almost half of them (48.2%) have been dedicated for more than 10 years to infomediary activities**. However, 35.5% have started infomediary activities in the last 5 years, proof of the important dynamism of the sector and the outstanding development of infomediary activity, which coincides in time with the political boost to the data economy at European and national level.

Employment in the infomediary sector

The new approach used in current report edition had led to reviewing the companies' census, which in turn impedes the analysis of the employment evolution regarding previous versions of the report. Despite this, the traditional methodology to estimate the number of employees has been followed. There are between **15,000 and 16,000** employees in infomediary sector. **The number of employees involved in reusing information processes ranges between 7,700 and 8,400.**

Regarding the size of the companies, **94% are considered small and medium-sized enterprises (SME)**. Less than half (47.6%) are considered micro-enterprises (0-9 employees).

When considering the distribution of employees in infomediary activities, 63.9% of companies have less than 10 employees involved in those activities.

For the first time, this study analysis sector employment from a gender perspective. The good news is that employee's distribution by gender is equal. **49.9% of employees are women**. If only infomediary activities are considered the presence of women is even greater (51.9%). The bad news is that women's participation decreases dramatically in younger companies where the technological component is much more accentuated, in line with what happens in the whole digital ecosystem, where female employment is in a clear minority.

Last two years, the evolution of recruitment in the infomediary sector has been positive. In 2017, the percentage of business recruiting new staff was 60.2%. In 2018, this percentage grew till 66.3%. In 2020, a decrease in recruitment is expected, as only 38.6% of enterprises intend hiring new staff.

1.2 Primary information characterisation

Information sources

Information provenance can be analysed from two perspectives: entities providing data and information and their geographic origin. Regarding providers, **69.9% of companies use data from both the public sector and private agents.**

Regarding geographic origin, **95.2%** of companies use information from **national sources**, while **57.8%** use **international sources**.

74.7% of companies using national data sources, obtain it from public entities, while 79.7% reuse information from private sources. Among companies using international data sources, percentages are very similar (75% from public sources and 79.2% from private agents).

Considering information from public sources, at national level most of the companies use information from the central state administration (86.4%). At international level, the European Union is the information source for 86.1% of companies.

Types of information used

When surveying companies about types of information used, the following two main criteria have been considered:

- Free or paid information
- Standardisation or personalisation level of the information

According to these criteria, there are six possible types of information. The free sources of standardised information are the most used (61.4% of companies), followed by free sources of raw information (59%). **The fact that the information is free is a very relevant factor when choosing one information or another.**

Regarding the subject matter of the information and its source (public or private), the main categories reused from public sector are commerce (47.1% of companies), environment (42.6%), economy (42.6%) and demography (42.6%). Considering private information, the main categories used are commerce (47.9%), economy (42.5%), tourism (39.7%), science and technology (39.7%).

Information access modalities

The **access and download of information without need of previous request to provider** is the most common modality of access to information (**66.3%** of companies). 54.2% of companies choose to ask providers for data when they need.

Regarding technical mechanisms of information access, 74.7% of business obtain data through web consulting services or APIs.

Formats used to provide information

Structured data in no-proprietary formats is the most common way to access information (csv, xml, json, etc.). 65% of companies declare that they almost always or often access data in these formats. Less frequent is the format of data linked to its context. Only 27.7% of companies declare to access them almost always or often.

Unstructured data (pdf, jpg, etc.) are still very common. Its presence constitutes one of main complaints from companies to data providers due to its difficulty of processing.

1.3 Characterisation of infomediary products and services market

Costumers origin

97.5% of infomediary companies have national customers. 28.4% sell products and services to clients within the European Union, and 23.5% commercialise their products out of the EU. **The internationalisation of the Spanish sector is, therefore, in an incipient state and it should be promoted by the sector to stimulate its economic growth**.

Within the EU, the main client countries for Spanish companies are Portugal, Germany, France and the United Kingdom. Out of the EU, Latin America is the region where most Spanish companies sell.

For the first time, companies have been asked about the mechanisms used to develop their international activity. **72.4% develop their international activity from Spain**. 24.1% have local partners in foreign markets. And 20.7% have created new enterprises abroad.

Types of customers

The commercial activity of infomediary companies has a marked Business-to-Business (B2B) character, as 92.5% of them declare that they sell to other enterprises. 47.5% sell product and services to public administrations and 32.5% to end users.

In this edition of the report, special attention to the **public administrations as customers** have been paid. The companies were asked about the typology of clients from the public sector. 81.6% of companies sell products to regional entities, 78.9% to local entities and 65.8% to central state entities. The products and services most required by public administration are specific publications (51.4%), own data processing (45.9%), and data visualisation tools (40.5%).

Considering public administrations as customers, they show very concrete needs**. This fact pushes infomediary companies to adapt their products or even create ad hoc products and services**. 37.8% of the companies surveyed that sell to public authorities claim to have developed a specific solution for the public sector

The study analyses the weight of the turnover from the sale of products and services to the public sector in relation to the overall turnover. For one-third of the companies, the revenues from the public sector represent more than 50% of their total incomes.

Market volume

The market volume of the infomediary sector grew 14.7% in 2018, reaching **1.987 million euros**. However, not all revenue comes from infomediary activities. **Turnover from information and data re-**

use activities reached 718 million euros in 2018, 11.1% more than in 2017.

Another market indicator that allows to have an idea of its growth is the **evolution of the number of clients**. This evolution has been positive for 60.3% of the companies, which states that the number of clients has increased.

1.4 The sector opinion

Data characteristics that enable business creation

Updating and completeness are the main data characteristics with most impact when it comes to generate business from them. Three out of four companies consider that both characteristics have great influence to consider data as reusable.

Other important characteristics are **detail and disaggregation level** and **free of charge**.

The companies interviewed considered of paramount importance that data and information are opened according to **international recognised standards**.

Regarding formats, companies' representatives were unanimous: data must be shared in **reusable, legible and suitable formats to be processed by machines**, that requiring as little work as possible on cleaning and treatment. In this line, companies also consider essential that data owners assess **the impact that any modification** (change of format, sharing channel, etc.) **may have on users before undertaking it**.

Accessibility, understood as easy identification of data repositories and the existence of aggregation platforms, is another strategic characteristic highlighted by the companies.

According to the companies, **greater availability and access to data and information has a direct impact on business generation**. This would enable the development of new products and services.

Characteristics of public services characteristics to open data

The characteristics of public services aimed at opening data play a key role to facilitate business generation. The characteristics most valued by companies are the **full availability and the proper functioning of services**. Another relevant characteristic is the **availability of web services that allow massive download of data**. The possibility of **data downloads automatisation** is a key element to develop the industry.

Companies **miss wider awareness among political leaders at all levels to boost data opening**. Companies noticed a very big boost

5 or 6 years ago, coinciding with the entry into force of Law 19/2013, of Transparency, Information Access and Good Governance. But now they feel that the interest in open data is slowing down at political level. Despite this, infomediary enterprises value very positively the work done by the Secretary of State for Digital Advancement and Red.es to promote the culture of reuse within public administrations, mainly through the Aporta project.

Another remarkable point for companies is the **change of mentality** in public administrations. Today, there is a **more cooperative attitude**, as administrations have understood that infomediary companies' role is to add value to public information which, in turn, benefits the whole economy and society.

Barriers to sector growth

More than 75% of companies surveyed point as the main barrier the **lack of information that should be availability**.

71.4% of enterprises consider as an important obstacle the **lack of homogeneity in the information provided by different administrations.**

61.9% of companies also consider the **lack of public information updating** as an obstacle to its use.

The **difficulty to find concrete data sets and the lack of accessibility, interoperability or data standardization** are relevant barriers for 55.6% of companies.

The use of non-reusable formats difficult the automation of information acquisition process.

The companies consider that some **regulations**, allegedly aimed to guarantee the proper use of data and information, **are being used as an excuse to deny the access to information**. This is the case, for instance, of the General Data Protection Regulation (GDPR).

The companies consider that some **regulatory uncertainty** regarding the use of information still exist. In the opinion of the companies, the current regulatory framework is very strict, but it can also be interpreted, which undoubtedly reduces its effectiveness and leads the sector to a certain degree of legal uncertainty when it comes to reusing information.

Considering public administrations as customers of information and data-based products and services, the companies believe that **public procurement models are not adapted** to this type of solutions.

The last great barrier highlighted by infomediary companies is the **difficulty to find professionals with appropriate skills and knowledge**.

Infomediary companies' contribution to society

The impact of reusing goes beyond the merely economical, having a remarkable social impact.

First social impact highlighted by companies is the **increase of transparency**. Closely linked to transparency, companies also consider that their work contribute **to democratising public information access**, which was formerly very difficult for citizens.

From a social point of view, the greatest positive impact may be the contribution of infomediary companies **to improve the decision-making process** in emergency services and healthcare.

The surveyed companies highlight three main economic benefits with direct impact on society:

- To give **legal security to commercial transactions** favouring economic growth.
- **To improve competitive intelligence process** optimising technology watch.
- **To facilitate access to added value information** for customers.

Future trends

One of most remarked trends for companies is the **evolution** that is taking place from the mere sale of products (data, databases, or information generated from them) to the **provision of services.** According to enterprises, the four most relevant business related to data and information for the next years are:

- Consulting
- Data analysis
- Infrastructure and technology for data opening
- Security

In the field of data infrastructures, there is an unstoppable trend to **providing services from the cloud**.

The huge amount of existing information, which make it difficult for data companies to identify and extract value from it, is leading these companies to generate so-called "data lakes", with raw information for customers to experiment with and detect possible uses for their business.

Current emerging markets of data products and services are characterised by fragmentation. The sector will probably face soon a **process of concentration**. This process will allow companies, by means of acquisitions or strategic alliances, offering integrated services related with data, from initial consulting to the development of AI algorithms to predict clients' behaviour.

All the companies interviewed agree that the **integration of data analysis as a key corporate process** will be a fact, just as it happened in the past with processes such as marketing or logistics, which started with a secondary role and are now an essential part of any company's corporate strategy.

1.5 Recommendations for infomediary sector development

Main challenges

The main challenge, which in a way encompasses all the others, is to **consolidate the growth path** that has been started. However, the **current configuration of the market**, which is **highly atomised** and has many small and medium-sized specialised companies **may become a barrier to growth**.

The search for growth undoubtedly involves the **integration of the most advanced digital data processing technologies** (big data solutions and artificial intelligence), since they are the ones that provide the most value.

Another great challenge is the **lack of staff with appropriate technological knowledge**.

The **existence of suitable communications and infrastructure** is a key factor to achieve the integration of data analysis technologies.

The development of new products faces two main challenges. The first one is the **difficulty to standardise products** for massive sales. The second one is the **market immaturity** due to the lack of knowledge of data value by most of the managers of the client companies or administrations.

Another fundamental challenge to create new markets is the **internationalisation of companies**. It is a challenge shared with the whole Spanish economy.

Other important challenge is to achieve a **balance between respecting citizens' privacy and the use of data to generate business**.

Finally, the growth of the sector depends on supporting a **cultural change in companies and administrations**, so that management and administration are no longer based exclusively on the experience or intuition of management professionals, but on the use of information and data to support their decisions.

How can collaborate the Public Administration to boost the sector?

Companies request **to deepen into data opening,** providing more data sets.

Regarding data quality, companies demand that the information provided by the Administration be **structured and homogeneous**. In this sense, companies consider it very necessary that the data have a **national coverage.**

Homogenisation is not only demanded for data or information opened up by administrations, but also for the services that access this information.

Companies continuously demand for data updating. In the current edition of the report companies go further requiring **access services to information in real time**.

The companies demand that the information is provided according to **recognised international standards**.

The last great requirement related to data is that it should be free. This characteristic allows companies to improve their economic results which, in turn, benefits society in form of taxes.

Regarding coordination, infomediary companies suggest the creation of the **National Centre for Data Management and Diffusion**.

The enterprises also suggest **coordinating data opening strategies** in the Administration, arguing that isolated initiatives have very little impact. In addition, companies require a better **communication of public initiatives regarding data opening**. There are hardly any channels to keep informed about new released data sets.

The companies require more **collaboration** between the Administration and themselves to define priorities in the opening of **data**.

Proposals for public-private collaboration:

- Defining **economic and financial support policies** for the sector.
- **Developing competitions** (similar to the Aporta challenge) challenging start-ups and other entities to provide concrete solutions to real problems.
- **Adapting public procurement models** to facilitate the acquisition of data-based products and services.
- **Updating information systems** in public administrations.
- Defining a framework to achieve a **balance between regulatory restrictions** (for instance GDPR) **and innovation** of products and services.
- **Improving connectivity in rural areas** so that they can also benefit from data analysis systems applied to the rural world.