datos.gob.es
reutiliza la información pública

Characterization of the Spanish Infomediary Sector

datos.gob.es
reutiliza la información pública

Edición 2016

Executive Summary
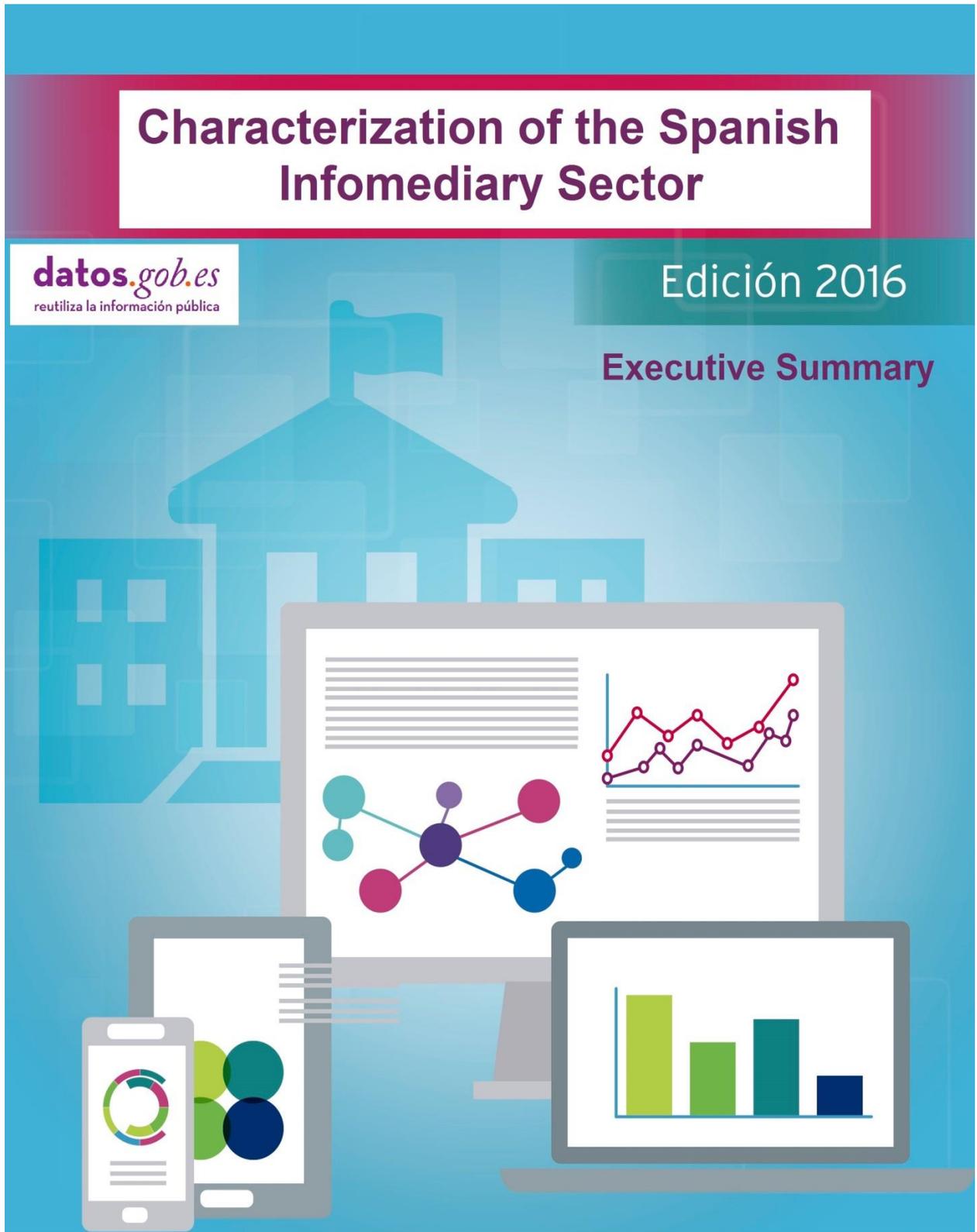
GOBIERNO DE ESPAÑA

MINISTERIO DE ENERGÍA, TURISMO Y AGENDA DIGITAL

ontsi

observatorio nacional de las telecomunicaciones y de la SI

STUDY OF THE INFOMEDIARY SECTOR 2016

EXECUTIVE SUMMARY

National Observatory of Telecommunications and the Information Society (ONTSI)

Public Corporate Entity Red.es

The State Secretary for Information Society and Digital Agenda.

The Ministry of Energy, Tourism and the Digital Agenda.

www.ontsi.red.es

www.red.es

2016

Executive Summary

The infomediary sector

The aim of the 2016 study of the Spanish infomediary sector is to further our knowledge of it by setting out in greater detail the results of the ongoing analysis being carried out since 2011.

> ### INFOMEDIARY is:
>
> "A set of companies that become intermediaries of general information in such a way that they analyze and treat public and/or private sector information in order to create value-added products for third party businesses or the public that serve efficient decision-making processes.
>
> This definition includes companies created for this purpose as well as those that were not created solely for this purpose but have an area and/or specific department devoted to the creation and marketing of new products and/or services based on public sector information."

The current study aims to show the current state of the infomediary sector that reuses public sector information in Spain and to study its evolution since 2012.

In particular, the study focuses on:

- Discovering **the main features of infomediary companies and their offered products and services**.

- Identifying **the main characteristics of the primary information supply**, which constitute the materia prima from which the products and services are generated.

- Learning about **the market and the demand for applications, products and service**

Identification of the Spanish Infomediary Sector

Infomediary activity is not equivalent to any officially classified economic sector, considered from the point of view of the National Classification of Economic Activity (CNAE). On the contrary, there are many business activities that reuse information.

Therefore, infomediary businesses are not identified on the basis of a classification of said activities but rather their processes of production and the value-added products and services that they generate.

Many infomediary activities have been carried out traditionally by businesses, whether or not they have ever been identified as such, either

because the value of data and information has been increasingly recognized in the information society or the use of data and information has modified their processes of production.

To these traditional activities have been added others that previously did not exist, innovative products and services that allow new players to enter the value-added market based in data and information.

All of this supposes that the attempt to identify and characterize infomediary businesses is an approximation.

This study was based in a **survey of 652 businesses in 2016**.

Finally, **have participated 535 business** in these investigation.

51% of infomediary businesses are located in the **Community of Madrid**, 19% in Catalonia and 6% in the Basque Country.

The typical profile of infomediary businesses in as a **small business**, 56% of them micro-enterprises with less than 10 employees and 22% of them with less than 25.

The most common legal framework is that of a **limited liability company** (75%).

82% of the businesses in the sector have been in operation for **more than 5 years**. Likewise, between 62% and 65% of them have been carrying out a **data re-using activity** for more than five years.

In any case, **18% of these businesses have been in operation for less than five years**, which demonstrates the sector's capacity for recovery and the existence of new initiatives associated with data reuse. It is no coincidence that **37% of the surveyed infomediary businesses started after the first European Directive and Spanish Law 37/2007 on reutilization of public sector information were passed**.

31 % of respondents indicated that they are part of an **association**.

652

Number of businesses identified as **potential data re-users**

535

final participants in the survey

General business activity of infomediary businesses

The responses to the survey indicate that the most predominant sectors of activity are the following, in order of importance:

- **Market Studies and publicity, 30 %**: includes carrying out public opinion surveys, research in and analysis of markets, consumers and lifestyles, media and communication, as well as other publicity and marketing services.

- **Research and consultancy in the public and/or private sector, 34 %.**

- **Publishers, 20 %**: includes editing, drafting and printing books, commercializing legal and/or contract databases, general information services, publications, newspapers and magazines.

- **Geographic information, 15 %**: includes generating geographic and land registry information, graphic as well as alphanumeric, including urban and real estate information.

- **Economy and Finance, 13 %**: includes generating information on commercial risk and marketing, credit and solvency information, credit bureaus, information on financial markets, investments, insurance companies, etc.

- **Directories, 13 %**: includes the creation of directories and postal address guides for locating and advertising to business and people.

- **Cultural businesses, 13 %**: includes the circulation and/or commercialization of goods, services and cultural activities that facilitate the public's access to them.

> **These sectors cannot be considered as hermetically sealed compartments. This is due to the fact that these businesses frequently classify themselves as pertaining to more than one sector.[1]**

For **73%**, infomediary activity is one business line amongst others.

74%

Reuse public and private information.

Other interesting data relevant to the sector include:

- 73% of businesses indicate that **their infomediary activity is not carried out in isolation** but, rather, is integrated with other activities. In some cases, this activity is linked to software development.

- **Most infomediary companies reuse public as well as private information** (74 %). Only a minority of businesses indicated that they reuse only public (8%) or only private (17%) information. **More than 80% companies are re-use public data companies.**

- The impact of infomediary activity. Businesses indicate that infomediary activity, when combined with other activities, **has allowed them to increase their profits, largely due to the diversification of new products and services and the discovery of new business opportunities**.

---

[1] It is possible that the market studies sector is broader than that represented in the study carried out here given that the majority of surveyed businesses that do not recognize themselves as infomediary are associated with it.

Business resources

Analysis was carried out on Human Resources, the expenditure on public sector information and the tools used by businesses.

HR assigned to infomediary activity.

EMPLOYMENT

5000-5200
employees working in the sector

53%

Have **hired personnel in the last year**

In this case, **the sector provides for approximately 5000 to 5200 jobs**. (The personnel dedicated to infomediary activity **grew from 10-16%** in the last year.) The majority of these employees are highly qualified, with 61% holding an advanced degree and 17% with a medium-level degree

o It is worthy of note that **53% of businesses indicated having hired personnel in the last year**, which indicates the capacity of the sector to stimulate job growth.

o **Businesses indicated that their activities were based on knowledge, thereby grounding their existence in academic capital and the experience of their employees**.

The profiles required by the sector include **doctors and/or graduates with degrees in sociology, computer programming, engineering mathematics, physics, economics, journalism and architecture**.

They also frequently stress the need for **technological profiles focused on product development and** the demand for **marketing experts**.

Employees need to combine a mastery of techniques and technological tools that permit successful data mining and a familiarity with the kind of products and services produced in the sector. For this reason, many businesses employ multidisciplinary teams (techniques, tools, sector).

The costs of reusing information

- Most businesses that responded to the survey indicated that the public information they reused was **free**.

  It is worth noting that the number of businesses that make a significant investment in public and/or private information has grown. This shows the value that data and information are acquiring in the market, as well as the growth of the data management market itself.

Tools for managing information

- The tools most used by infomediary companies are **software and data analysis processing tools**, as well as visualization and report presentation software and cloud data storage services.

  The growing importance of the use of cloud storage services indicates that the sector needs a large storage capacity due to the use of macro data. Infomediary companies are consumers of software and very specialized ICT services which, in some cases, they have developed on

their own but, in others, they have purchased or leased from third parties.

> **The need for specialized software and ICT services is seen as a barrier to market access and business growth for the first time in this study.**

The majority of businesses combine three types of tools in the development of their products and services.

Products and services

Businesses have diversified their production as they have grown and thereby have developed different kinds of products, services and applications. Similarly, the **tendency seems to be to develop products with greater added value**.

The current trend is to develop value-added products and services

Portfolio

- **Products.** In the case of products, **the amount of treated and raw data generated has decreased and publications have increased**. This reduction is connected to the tendency to develop value-added services that substitute more traditional products.

- **Services.** Concerning services, businesses offer support in order to **facilitate their clients' decision-making processes**, thereby offering specialized consulting services of various types. In the same way, **predictive analyses** have also risen in significance.

  The number of businesses dedicated to product and services comparisons has diminished proportionately.

- **Applications.** The top applications are programs used for **data visualization**, mentioned by 41% of the businesses surveyed, which reinforces the growing tendency to use value-added services dedicated to decision-making processes.

Languages and client relations

- **Languages.** There has been a notable increase of the presence of products in different languages such as English (in 69% of businesses) that offer infomediary products and services.

  German and Italian have also been mentioned and the use of French is on the rise. All of this could indicate a **growing international presence of Spanish infomediary products and services** since 2014.

  o It is worth mentioning that 88% of the businesses surveyed with 10 to 50 employees offer their products in English.

59%

Develop products /services in English

- **Client relations**. Infomediary companies have an active presence on the internet, through web pages, social networks, SEO/SEM positioning and other means, such as the re-directing of links.

  o References to private channels for client relations, such as an extranet, dropped 20%, but there are now references to new data access and cloud content services.

The position of the businesses in the value chain

71 % of businesses in the sector occupy are located in the final link of the value chain, which is to create value-added products and services, such as reports for decision-making, programs, apps, etc.

- **45%** of businesses provide services related to the **analysis of data and the posterior visualization** of the results.

- **49%** are dedicated to the **elaboration of data**, focusing on diverse tasks such as aggregation, anonymization, filtering, integration or data enrichment.

- Another important group, about **40%**, indicates that they **commercialize treated data**.

And so, in general:

- **Businesses that only reuse public data mainly commercialize treated data.**

- Businesses that **only reuse private data mainly produce data-on-demand**.

- Businesses that **reuse public as well as private data** are mainly dedicated to providing **value-added services**.

Origin and type of reused primary information

95%
of businesses reuse **public national information**

40%
of businesses reuse **information from all three levels** of administration (national, regional and local).

Both 95% of the businesses that reuse public information as well as 89% that reuse private information also use national information, compared to 48% and 50%, respectively, that use international information.

- **National information. The National State Administration is the main provider of national information.**

  o 40% of businesses reuse information from all three levels of Spanish administration: The National State Administration, the Autonomous Communities and the local administrations.

    The INE is the main source of information, followed by the Mercantile Registry, Eurostat, the Center for Sociological Investigation (CIS), the Ministry of Energy, Tourism and the Digital Agenda, the BOE, etc.

- The **private information** that is most reused has its origin **in private sources accessible on the internet**, as well as in the use of **one's own client's data**, but also those from other **infomediary data providers (data brokers)**.

    The organization most mentioned is the **Association for Communication Media Research** (AIMC) which runs the General Media Survey (EGM). In second place are several different private sources accessible to the public on the

internet (e.g. social networks) and in third place are businesses that provide treated data.

- 71% of businesses indicate that they use information from different countries belonging to the European Union. It is significant that 59% of businesses use information from other countries. The **European Commission** is the main source from within the Union.

- The most reused types of information are **socio-demographic and socio-economic information, as well as information involving education, science and technology**. After them, in order of importance, are geographic information and information involving businesses and economic sectors.

- **Businesses indicate that their products and services are the result of combining data and information from different sources and of different types**:

  o Concerning public information, most businesses reuse more than 4 types in the development of their products and services; while businesses that reuse private information mention using one or two types.

  o 24% of businesses use more than 20 different public sources, which shows the **effort required to treat this information**. In contrast, 32% reuse from one to five public sources, which shows, in addition, how **important it is that data be put at the disposal of the public in an effective manner** – as well as the dangers posed by the possible discontinuity of said information.

Access to data and primary information

67%

access data in **open formats**

Vs

70% access data that is **unstructured**

- Concerning the format in which public information is accessed, **the number of businesses that reuse data in open formats has risen** compared to the previous survey, reaching 67%.

  o In contrary fashion, 70% continue to access unstructured data and only 24% access URLs that allow for large scale data transfer.

- The sector needs more available data and information related with economic subsectors, foreign trade activities, as well as the availability of data on existing certifications. Fiscal information on legal entities is also required.

24%
access URLs that allow for large scale data transfer

40%
use "crawler" tools

73%
The web site of the data provider is the main communication cannel

⚠️

68%
Tiene que pedir datos no disponibles

o Businesses need desegregated information on the characteristics of the neighborhoods and postal districts of the resident population, more accessibility to the data on transactions carried out with notaries, more information on transport and its routes and timetables, as well as a more updated survey of existing equipment. The need for structured data on financial support recipients, by help, theme or regional programs was also reiterated.

In connection with the channels businesses use to access information, **73% indicate that they get their data from the information provided on web pages**.

- **68% need to request data that is not readily available**. 35% have data sharing agreements and 40% use web crawlers.

- **59 % of businesses make reference to obtaining data via API or a web service (large scale transfer)**, which requires automated file reading and download processes. This indicates the providers have a high level of advanced service. **The fact that, amongst businesses that only reuse public information, this kind of service has dropped to 14% could indicate that public administrations are offering this kind of service with less frequency**, a warning signal that should be confirmed in subsequent studies.

- **Businesses attest to an improvement in gratuities.** Public data that were once for purchase have now been made available for free.

- **Businesses would be prepared to assume increased prices brought on by marginal costs in the processing of public data in order to get quality service or assure the data's availability.**

Evaluation of primary information

3,1/5

Evaluation of how information is made available by Public Administration

Making information available

Public administration bodies received a **high approval rating (3,1/5)** concerning the information they make available, compared to previous editions of the study.

- The most highly valued characteristics of data are being **free of charge and available in structured and/or open formats.**

- The evaluation is not so positive in the case of **update frequency**, **the level of disaggregation of the data or its accessibility**. There is a large room for improvement in all of these areas.

  o The most highly evaluated information included urban planning and infrastructure, land registries, the environment and rural areas, which includes geographic and meteorological information.

  o The worst evaluation was given to information concerning the economy, internal revenue and employment. This was followed by information on legislation and the courts and the very public sector itself, which expressly includes contract information.

DATA ACCESS SERVICES

Positive:
- Data always available and downloadable

Improvable:
- Channels to request new/more information
- Large scale download services

OBSTACLES

- **Lack of homogeneity**
- **Lack of Admin's response to the needs of the sector**
- **Needs for technological investment**
- **Little development of large scale downloads**

Services for making information available

In connection with **services linked to making information available**:

> - Businesses give their **approval** of **making data always available and downloadable in file format** – the aspect which they gave the **highest rating**.

Despite the fact that existing services are normally available, there are not as many as there should be. For this reason, the following negative conclusion can be emphasized:

> - There is a **lack of specific channels** to to request the publication of new data. **Data, in general, are not readily available**. 45% request **more large scale download services**.

Obstacles to the devlopment of new services

- **The main obstacle to the development of new services is the lack of homogeneity amongst the information of the different autonomous regions,** something which could also be said of the information provided at a local level. The lack of homogeneity limits the development of global products.

- Another obstacle was found in the lack of response by and interest of administrative bodies in becoming familiar with and understanding the needs of infomediary businesses and **the administration's lack of response to requests for currently unavailable information**.

  o The need for **greater investment in technology** is seen as a barrier in 56% of the cases, which might have to do with the effort required to treat public data such that they become useful.

  o 36% of businesses identify **the lack of large scale download services** as a barrier, linked to the need for technological investment. The lack of such services creates the need to treat the data, instead of the administrations offering more adequate service.

The infomediary product market and its revenue model

FINANCIAL TURNOVER OF THE INFOMEDIARY SECTOR

Reutilization
600-750 M€

Global Market sector
1.550-1.750M€

Financial turnover

The overall financial turnover of the sector in 2015 was from 1.550 M€ to 1.750 M€, a number much higher than others offered by alternative studies.

The turnover derived directly from infomediary activity would be somewhere between 600 and 750 million Euros, higher than 2014, when it oscillated between 550 and 700 million Euros.

datos.*gob*.es
reutiliza la información pública

As noted before, more than 80% of the companies are re-use companies of public information and the offer of products and services and therefore their business, depends on, such reuse.[2]

- The crisis provoked a price adjustment in the sector, driven by a need to stay in the market. This allows one to suppose that the positive results of these last two years is related more to an increase of demand than an increase in prices.

- **In general, businesses have seen improvement, which supposes a sound recovery, as is shown by increased employment and demand.**

- **81% of infomediary revenue derives from the direct sale of products**, while more than 19% derives from other sources of revenue such as advertising, the brokerage of product sales and third-party services, etc.

- **Businesses also indicated that 43% of their activity originates from reused public information.**

Revenue model

Businesses indicated that they employ very different kinds of revenue models. The main one is the **pay-per-use, per-access or per-task** model which is employed by 69% of businesses.

○ The revenue model is not necessarily dependent on the size of the business. All the revenue models are more or less present amongst all the businesses independently of the number employees they have. This means that **the revenue model is related to sales strategy or the type of products sold**.

○ Some infomediary activities are integrated into very complex revenue models such that **the service is linked to other commercial activities**, serving as support in the search for new clients or as a value-added element that accompanies the service.

The main clients are other businesses but there is still an important presence of *Public Administrations* and *more citizens are end customers* of the sector

Type and number of clients and their main requests

The **heterogeneity** of the revenue model of the infomediary sector also becomes evident when we look into the number and type of clients:

- **64% of businesses have their activity concentrated in less than 50 clients, 28% of them with 10 clients or less. On the other extreme, 21% of businesses have more than 500 clients.**

---

[2] IMPORTANT METHODOLOGICAL NOTE. The estimation of business directly related to the information public, and only dependent on it, it becomes a calculation complex because 74% of companies, products and services derived from reuse involves integration of public and private source data, without be able to calculate what would have more value in each case, and in total 82 percent of them would be reusing information public in some way or another. We understand that products and services can be generated by the fact that both sources of information are used together and if it were not so, such products could generate no of course, marketing.

- **The majority (90%) of businesses in the sector have other businesses as clients.**

- Public Administration is in second place at 58%, having fallen in importance compared to the previous study. And in third place are citizens at 49%, which is more significant than in the previous edition.

- Concerning the kind of information requested, described according to the type of client: businesses that direct their products at other businesses consume proportionately more information related to legislation, economic sectors, economy and finance, or the public sector itself. Businesses that direct their products at Public Administrations consume information related to legislation, the environment, geographic information, tourism and culture. Those who have citizens as clients use information on legislation, urban development, wellness and health.

- Products. The products most sought-after by **businesses are raw data**, while **administrations would like more maps and publications and citizens want maps above all other products**.

- About services, there is a **high demand for comparative analyses for business and administrative decision-making**, although, for these last-mentioned, **training services** are also significant.

- Finally, concerning applications, **there is a greater demand by businesses for alerts and systems of geographical information by public administrations**.

Demand for infomediary products and services

97% of businesses indicate that they sell their products and services in Spain. Even so, **34% sell across Europe** and 34% in the rest of the world, which shows that international activity may have slowed since 2014, although it is still above the figures for 2012.

The main area of business is the national market.

34% of businesses operate internationally.

- **54% of businesses indicate that they only sell their products in Spain versus 19% that operate in the three areas (Spain, the EU and the rest of the world.)**

- France is the main consumer of Spanish infomediary products, followed by Germany, Portugal and the United Kingdom. The United States and Latin America are the most important non-European markets for Spanish products.

- Internationalization is not only the case for larger businesses, but small businesses are also entering international markets.

- For some smaller businesses, participation in European projects provides an opportunity to further their internationalization.

For some businesses, their international activity derives from partnerships with clients or national or international investors.

The growth of demand and opportunities

OPTIMISM FOR THE SECTOR'S FUTURE

46%
Have increased their client base

New elements:

- Services for **real-time decision-making**

- Specific fields: **SOCIAL MEDIA / SMART CITIES**

- Technology: **tools that manage data automatically and the growth of BIG DATA.**

Businesses are optimistic concerning the evolution of the market in the short and long term.

- **46% of businesses indicate that their number of clients has grown this last year, versus 15% that say it has decreased.** This positive trend is supported by the fact, already mentioned, that there has been an increase in the hiring of personnel in the last year.

- **The impression is worse amongst smaller businesses that were probably the most affected by the crisis.**

- The outlook on the future and growth of the sector remains positive, with companies identifying opportunities to develop their business based in open data.

  In the middle term, greater opportunity is expected with respect to the development of products or **services intended for decision-making in real time**, based in the **analysis of FIELDS such as SOCIAL NETWORKS or the data of SMART CITIES and new tools** that allow for the automatic management and interpretation of data.

  In this sense, the advances of BIG DATA and the tools for its utilization and support will be crucial for the future development of the sector.

- There has been a change in the trend of infomediary businesses. They no longer need to make so much pedagogical effort when selling their product given that, thanks to the crisis, **their clients have become aware of the value of information when taking decisions**.

1.1   Proposals and sector requirements

The sector needs:

> ### Greater homogenization of Autonomous Community information, local institutions and all Administration in general.

> The barrier to accessing homogenous information in the Autonomous Communities and local institutions impedes the creation of economies of scale *restricts the development of products and services and has a negative impact on economic growth,* which is the objective of Spain's information reutilization policy.
>
> Further, this situation *could have a negative impact of Spanish businesses' ability to compete,* given that businesses from other countries could hold an advantage over Spanish ones when developing products and services that require an intensive use of data from different levels of the Administration.
>
> The lack of homogeneity in the standards applied to data offered by different administrations, and the lack of large scale download services, *oblige businesses to develop processes or tools for each of the administrations from which they acquire their data,* leading businesses to claim that "*in many cases, this makes the development of certain services impracticable, especially at the local level, given the multitude of municipalities in Spain.*"

> ### Proactive efforts in Administrations' information management

> *For businesses, making an active request for data wouldn't be the exception rather than the rule if the data were made readily available and more essential to their needs, which would oblige the public administration to look into businesses' needs and determine which data, if put made available, could have a positive impact on a region's economic growth.*

The development of ICT tools to treat public data

Specific programs might be needed to stimulate the development of ICT tools that help automate the access, cleansing, treatment, enrichment and analysis of public data, with the goal of lowering the barrier of investing in the technology needed to take on new data reutilization initiatives.

Improved availability and quality of public data

This would require overcoming the following **CHALLENGES:**

* *Empower the right to access public data*. According to regulations, this is an obligation.

* *Ensure interoperability* and the standards of public data access services. Guarantee semantic interoperability connected with data interpretation.

* Ensure *data homogeneity* on different administration levels*, unify sources.*

* Drive the development of *open reusable formats* and large scale download services.

* Establish known and predictable *Service-Level Agreements* (*SLA*). Open data available for all under any circumstances.

* *Revise legislation on statistical privacy* to liberate data protected by this restriction.

* Make sure that *privacy restrictions* linked to natural persons *are not applied to legal entities.*

Areas of the sector that need improvement in data accessibility

* **From business information:**

  o Independent businessmen
  o Dis-aggregated data of economic subsectors
  o Fiscal data on legal entities
  o Data on foreign trade
  o Business certifications

* **From the sociodemographic point of view:**

  o Resident population characteristics of neighborhoods and districts.
  o Data on the behavior of different immigrant groups.

- **Area of urban planning and real estate information:**
  - Data on transactions carried out with notaries.
- **Public service area:**
  - Information on transport, routes and timetables of transport subject to public concession.
  - In the urban environment, a more up-to-date census is needed on equipment, including the infrastructure belonging to the Spanish Port System. Aggregated data on concessions through help, theme and regional programs.
  - Improve the availability of AEMET meteorological information.
  - There is a clear demand for the historical data included in the Agroclimatic Information System for Irrigation (SiAR) of the Ministry of Agriculture and the Environment.

Additional conclusions:

The Administration could improve by reusing its own data

Businesses also refer to the utility that could be provided were the Administration to reuse public information to improve its own processes and services. They *call for Public Administration to implement data reutilization processes to improve their own internal operations and the services they provide to the citizenry, ensuring, in addition, greater effectiveness and efficiency.*

Public-private collaboration should be encouraged so that infomediary services and their market can be developed.

Businesses are ready to contribute to the development of new services through public-private collaboration. All of this while keeping in mind the need to ensure free and fair competition such that business in the sector can develop their services under equal conditions.