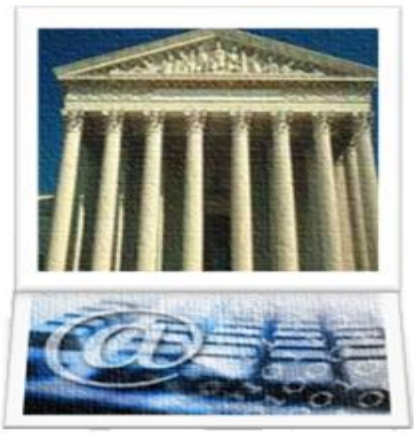
datos.gob.es
reutiliza la información pública

datos.gob.es
reutiliza la información pública



Estudio de caracterización del Sector Infomediario en España

Edición 2014

PARTE I. SECTOR INFOMEDIARIO PÚBLICO

Characterization Study of the Infomediary Sector 2014
Executive Summary

Characterization Study of the Infomediary Sector 2014

National Observatory of Telecommunications and the Information Society (ONTSI)

Public Corporate Entity Red.es

The State Secretary of Telecommunications and the Information Society The Ministry of Industry, Energy and Tourism

www.ontsi.red.es

Study carried out by the Ministries of Industry, Energy and Tourism and of Finance and Public Administration through the National Observatory of Telecommunications and the Information Society (ONTSI) of the Public Corporate Entity Red.es, with the assistance of ACAP (Asesores y Consultores de Administraciones Públicas) and Iclaves in research, field work and corporate analysis.

2014

Index

1 Executive summary

1.1 The infomediary sector

The aim of the 2014 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.

In this analysis, the term "infomediary" refers to the sector comprised by:

> **"A set of companies that create products or services in order to market them to third parties based on public sector information. 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** which constitute the key agent in the reutilization process.
- Delineating **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 services**.

The identity of the Spanish Infomediary Sector

Infomediary activity is not equivalent to any officially classified economic sector (within the CNAE). Businesses within many different economic sectors reuse information, obtained from both the public as well as the private sector. This makes it a complicated task to identify the relevant businesses given that the task requires knowing a priori the processes of production that characterize the sector. The complexity increases if one takes into account the fact that many of the businesses that carry out infomediary activity do not recognize themselves as businesses that reuse public or private information.

After the census update, **494 businesses were identified as carrying out infomediary activities on the basis of information from the private as well as the public sector**. In total, **413 businesses** were taken into account by the field work.

The general profile of infomediary businesses is that of a small business, 48% of them being microenterprises with less than ten employees and 25% of them with less than 25 employees.

Their most common legal status is that of a limited liability company. More than 50% of the businesses are from 5 to 20 years old.

32% of the businesses that responded to the survey indicated that they are part of a larger association.

General business activity

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

- **Geographic information –** which would include businesses dedicated to using geographic and land registry information (graphic as well as alphanumeric) including urban information and information on meteorological forecasts – represents 35% of the identified businesses.
- **Publishers –** which includes businesses dedicated to editing, drafting and printing books, commercializing databases, information services, publications, newspapers and magazines, distributing economic financial texts, etc. – are 32% of the businesses.
- **Market Studies –** which include businesses dedicated to activities connected with surveys of public opinion and research and market studies – make up 28%.
- **Directories** – which include businesses dedicated to creating directories and postal address guides for the purposes of locating and advertising to the businesses and people found in them and telephone subscriber lists – are 22% of the businesses.
- **Economy and Finance** – which includes businesses dedicated to commercial risk, credit and solvency information, credit bureaus, etc. – is the sector of another 22% of the businesses.

These sectors are not isolated from one another given that businesses frequently classify themselves as beloning to more than one of them at a time.

Other relevant information related to this activity includes:

- **71% of businesses indicate that infomediary activity is not carried out in isolation** but, rather, is integrated with other activities. In some cases, this activity is linked with software development.
- **Most infomediary businesses reuse public as well as private information (72%).** Only a minority of businesses who responded reuse **exclusively public (15%) or private information (13%)**.
- **The impact of infomediary activity.** The businesses pointed out that **when infomediary activity was combined with other activities it improved their results**, mainly thanks to an **increase of efficiency** in the production process.

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.** In this case, the sector provides for approximately **4200 to 4700 jobs**. (The personnel dedicated to infomediary activity

grew from 8-10% in the last year.) **The majority of these employees are highly qualified**, with 52% holding an advanced degree and 17% with a medium-level degree.

- It is worthy of note that **62% of businesses indicated having hired personnel in the last year**, which signals a recovery in the sector after the crisis.
- **The businesses indicated that it is not easy to find workers with adequate professional profiles in the job market.**

  The required profile is polyvalent: mathematician, with a degree in sociology, marketing or business management, with experience in data mining and analysis. Also included are developers who have experience with the Internet and geographic information systems, are oriented towards clients' needs and are able to provide end-to-end solutions.

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

  Only 16% said they payed for information, while only 7% of them quoted significant quantities. Some of these expenditures reached up to 300,000 euros.

- **Management information tools.** The most commenly used tools by infomediary businesses are **data analysis software**, such as **data and internet content search tools**.
  - Nearly **60% of businesses use cloud storage services**, which demonstrates needs linked to macro data usage.

Products and services

The businesses **diversified their production** to such an extent that the number of businesses that develop different types of products and servcices has grown. Similarly, the tendency seems to be to **develop products with greater added value**.

- **Products.** Concerning products, **processed data is on the rise**, above that of brute data. **71%** of businesses generate products through the processing of public data.
- **Services.** Concerning services, there is less preparation of personalized reports but **advisory services are on the rise**, which have a higher value given that they provide the client with specialized knowledge.

  There is also growth in services linked to **comparative analysis**, this last case essentially in the field of private data reuse.

- **Applications.** Concerning applications, 50% of businesses are providing this kind of solution, with growth in the sale of **mobile applications and alerting services**.
- **The sale of products and services in English and French has also risen**, intended for markets other than in Spain.
- **Client relations. Infomediary businesses** have an **active presence on the Internet**, based in blogs, social media, SEO/SEM positioning and other activities such as URL-forwarding.
  - 25% of businesses keep **private communication channels open with their clients over the Internet**.

Origin and type of reused primary information

98% of businesses use **national information** in opposition to **48%** that **also use international information**.

- **National information.** The General State Administration is the principal provider of national information.
    - **The INE is the main source of information**, followed by the BOE, the Land Registry Directorate, the National Geographic Institute, the Mercantile Registry, the Property Registry, etc.
- **The use of information obtained from international sources has risen**. In some cases, the use of international sources is due to the lack of availability of national data that, nonetheless, are available through international organizations. The European Commission is the principal international community source.
- The **type** of information that is **most reused** is **socio-demographic and socio-economic, as well as on transportaion and commerce**, which includes information on traffic. Next, in order of importance, there is information on urbanism and infrastructure, which includes land registry information as well as information related to the economy and revenue.
    - On average, businesses reuse from three to four kinds of public information to develop their products and services.
- **There is a tendency to reuse information from private sources but that have a public origin**, which supposes an increase of businesses that act as **data brokers**.
    - For businesses, this purchased information tends to be of a higher quality or with greater added value given that it supposes the processing of data obrained directly from the public sphere.

Access to data and primary information

- **The formats and channels of access to public information have been improved:**
    - **50% reuse information in an open format**.
    - Nonetheless, 65% still obtain information in a non-structured format, which requires subsequent transformation processes in order to be reused.
- Businesses understand that **purchased information is of a higher quality**, including private information based on that from the private sector. Some businesses are prepared to pay the marginal cost associated with generating higher quality information with the promise of a significant improvement in the quantity, access and form of presentation of public data.
- The way public data is accessed is more complex when it requires periodic updates or when the data is not available given that, in many cases, it requires a licensing agreement.
- Businesses indicate that **new models are needed for making information on demand available** when it is not accessible.

- The sector **needs more available data and information** related with the following areas: public contracts, the consumption of public services, land and equipment ordinance, demographic data on social behavior, administrative authorization processes and subsidies. They also request more information on **the local sphere**.

Evaluation of primary information

Public administration bodies were given a **high approval rating** concerning the information they make available. In addition, businesses indicated that **the available information has substantially improved in its quantity as well as its quality**.

- The **most highly valued** characteristics were that the information is **free-of-charge** as well as the structured and/or open **formats** in which the data and information can be obtained.
- The evaluation was **not as positive** concerning **the level of disaggregation of data**, its **complexity**, the frequency of its **updates** or additional **information** on the **context** accompanying it. There is ample room for improvement in all of these areas.
    - The most highly evaluated information included that of urbanism and infrastructure associated with land registries, the environment and rural areas – which includes geographic and meteorological information – and information on legislation associated with different official bulletins.
    - The worst evaluation was given to information concerning the economy and finance, demographic and socio-economic information and information referring to the public sector itself.
- Concerning **services** linked to making information available, businesses also gave a positive approval rating to **the most highly evaluated charactersitic: always being available and functioning correctly**.

  On the other hand, the aspect that received the lowest approval rating was that administrative bodies do not always provide the necesary data, which prevents them from developping their products or forces them to access the data through other – possibly non-public – sources.
- The **main obstacle** to the development of new services is **the lack of homogeneity amongst the information of the different autonomos 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.
- Another obstacle was identified in **non-structured formats** in so far as the processing of this data **requires the development of tools whose cost would imply an investment**.
    - This investment acts as an **entry barrier** to new businesses that want to enter into the market.

The infomediary product market and its revenue model

The **financial turnover** of infomediary businesses in 2014 was from **450 to 500 million euros**.

The results of the 2014 survey have allowed us to make our estimate more precise and to provide a narrower billing interval, establishing a minimum that is significantly higher than that given in previous reports. (In 2012, the quantity was situated between 330-550 M€).This turnover would mean a growth around 5-6%.

- The sector was not unaffected by the crisis and the reduction of internal demand. This had a positive effect on businesses' **offer diversification** and provoked **greater competition and a decrease in the unit prices** of products and services in the hopes of not losing clients and staying in the market.
- **82% of infomediary revenue derives from the direct sale of products**, while more than 17% 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, which implies that the sector's global invoicing volume is somewhere between 1000 and 1200 M€, a number that is very similar to those provided by other reports on the sector, and could mean a growth equally around 5-6%.
- **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 62%** of businesses.
    - This is the model that is most common in smaller businesses. Those that are larger or with a greater number of clients are developing other models, such as ones based in subscriptions, freemium, or free-of-charge with advertising.

Demand for infomediary products and services

95% of businesses indicate that they sell their products and services in Spain. Even so, 41% sell across Europe and 37% in the rest of the world, which shows that **there has been an international expansion since the survey carried out in 2012**. Part of this is due to the need to enter new markets given the decrease of internal demand.

- **France is the main consumer** of Spanish infomediary products, followed by Germany, the United Kingdom and Portugal. The United States and Latin America are the most important non-European markets for Spanish products.
- This internationalization is not only the case for larger businesses, but **small businesses are also entering international markets**.
- **Businesses are the principal consumers of infomediary products and services** in so far as they need them in support of their decision-making processes about business development or investments in new economic activities. **Public administrations are the second largest group of consumers** of infomediary products and services. (**68%** have public administrations as their own clients.)

The evolution of demand and opportunities

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

- **39% 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.
- In the middle term, greater **opportunity is expected with respect to open data through the development of Smart Cities** and, because of this, the possibility to develop **products intended for the general public based in real-time information**, something that – without a doubt – would require improving the current ways in which public sector information is made available.

1.2 Proposals and sector requirements

FOR THE SECTOR

MAKE A GREATER EFFORT TO EDUCATE CONCERNING THE BENEFITS OF REUSING OPEN DATA

An effort to educate seems to be necessary. Associations that represent businesses making significant use of public data and information to develop their products and services need to become sensitive to these businesses' needs.

This effort also needs to be carried out with the different public administrations with the aim of improving the processes they use to make information available. It should also be carried out with the general public with the aim of presenting and promoting infomediary products and services as well as their utility in taking informed decisions.

CONTINUE IN THE EFFORT TO IDENTIFY INFOMEDIARY BUSINESSES AND ACTIVITIES

It seems that an effort is also needed to further the process of identifying businesses and activities that reuse public and private information, as well as related sectors (like software development used in infomediary processes).

This process should be comprehensive and transparent. The effort should include the collaboration of all public administrations, associations and economic agents with an interest in or related to the sector.

HUMAN RESOURCES ADAPTED TO BUSINESS NEEDS

The course contents of degree studies such as mathematics, sociology, marketing, business, computer science and other higher education fields related to the profiles needed in the infomediary sector should be adapted with the aim of promoting the development of data analysts and developpers with the requisite professional skills.

FOR PUBLIC ADMINISTRATION

IMPROVE THE WAY INFORMATION IS MADE AVAILABLE FOR BUSINESSES

Businesses indicate that public administration bodies are not clear about the model they need to implement concerning data and information that is not available.

They do not know if the information can be made available, if they should charge for it, if it should be subject to licensing agreements, etc.

Businesses indicate that they would understand if payment models for obtaining information need to be introduced to cover the marginal costs associated with making it available, as long as the data is of high quality and it responds to business needs.

This would also imply better familiarity with business needs and, for this purpose, the establishing of defined and concrete communication channels between businesses and public administration bodies in order to identify the relevant businesses and establish a succesful means to achieve their goals.

IMPROVE ADMINISTRATION'S PREDELICTION TO GIVE ACCESS TO PUBLIC SECTOR INFORMATION

For businesses in the sector, access to public sector information for the purposes of its reuse is not perceived as a right which can be fully exercised in Spain.

In this case, as in the previous point, more attention needs to be paid to business needs and greater empathy needs to be developed.

The regulation is adequate and complete. The problem is the lack of predeliction on the part of public administrative bodies to provide the requested data or the unjustified reference to intellectual property rights and/or personal data protection laws.

It is understood that public administration needs to work on its organization, resources and management to find the optimum way to provide open data that can be used by infomediary businesses.

BETTER COORDINATION PROVIDING FREE DATA

There is a lack of coordination between different public administrative bodies.

It is proposed that a statutory body be created to coordinate the provision of free data by the different public administative organizations and bodies. This could be supported by a strategic plan accompanied by sufficient resources that would allow for an expansion of the process.

The plan should allow for the normalization of:

- **Formats to be used when publishing the same type of data.**
- **The kind of data whose periodical publication by different administrative bodies is mandatory.**
- **The update period for data according to type.**

THE CHALLENGE IS TO GENERATE QUALITY DATA

Quality should be given priority over quantity. The current standard of data disaggregation is not adequate. Further, on many occasions, the data are not complete, are not up-to-date, or are lacking context information as to how they were generated or simple things like when they were last updated or the frequency of the updates.

Therefore, it is necessary to:

- **Develop data configuration standards and avoid market entry barriers.**
- **Normalize data and information semantics across the same subject matters generated by the different administrative bodies in order to produce global products.**
- **Correctly update in a periodic and formalized manner.**
- **Encourage data completeness.**
- **Preserve context information on the data: their origin, field, scope, characteristics, etc., to allow for a better understanding and interpretation by its users.**
- **Improve disaggregation. For example: on a municipal level, census categories, by district or postal codes.**
- **Provide more information on public services and the use or consumption of them that indicates trends and social behaviors.**
- **Provide real-time information to create opportunities for new value-added products and services.**