

# Analyzing the employability and employment factors of graduate students in Spain: The OEEU Information System

Francisco Michavila  
Martín Martín-González  
Jorge M. Martínez  
UNESCO Chair in University Management and Policy  
Technical University of Madrid  
C/ Alenza N<sup>o</sup>4, 8 floor  
28003 Madrid. España  
+34 91 3366956

{francisco.michavila, martin.martin, jorge.martinez}@upm.es

Francisco J. García-Peñalvo  
Juan Cruz-Benito  
GRIAL Research Group  
Research Institute for Educational Sciences (IUCE)  
Department of Computers and Automatics  
University of Salamanca  
IUCE, Paseo de Canalejas 169  
37008 Salamanca, Spain  
+34 923 294500 ext. 3433  
{fgarcia, juanpcb}@usal.es

## ABSTRACT

This paper describes the technological approach for the development of the system that supports the Observatory for University Employability and Employment developed under the leadership of the UNESCO Chair in University Management and Policy. This observatory, nowadays, collects data from more than 50 Spanish universities and more than 134,000 graduate students in order to measure diverse factors linked to their professional career development and their employability. The paper explains many problems that this kind of project related to Academic Analytics and Institutional Intelligence needs to face and solve, as well as it explains some of the most important concerns and considerations that should be taken in the context of deploying a project like the described in a national scope. Also, the paper shows an introduction of how was planned the data strategy to gather the information, how was built the collector system, and how it will exploit the information from the perspective of Academic Analytics in order to provide great insights about the factors in graduate's employability in order to use it to make decisions and feedback the Institutional Intelligence processes.

## Categories and Subject Descriptors

D.3.3 [Information Systems]: Online Information Services - *Web-based systems*

K.4.3 [Computers and Society]: Organizational Impacts – *Employment*

J.1 [Computer Applications]: Administrative Data Processing - *Education*

## General Terms

Measurement, Documentation, Human Factors.

## Keywords

Employment, Employability, Graduate Students, Information Systems, Academic Analytics, OEEU, Spain.

## 1. INTRODUCTION

The Observatory for University Employability and Employment (*Observatorio de Empleabilidad y Empleo Universitarios, OEEU* in Spanish) <http://oeeu.org/>, is a technical unit that collects, produces, analyzes and disseminates information about employability and employment of university graduates in Spain, for which it has a network of researchers and technicians distributed throughout the country,

organized into ten teams working in coordination with a single methodology led by the UNESCO Chair in University Management and Policy (based at the Technical University of Madrid) and the ongoing advice of an Expert Council, composed of national and international academics and university experts. The project is developed also in collaboration with the "La Caixa" Foundation, the Conference of Rectors of Spanish Universities (CRUE) and the GRIAL Research Group of the University of Salamanca [8].

The vision of the Observatory is to become the information source of national and regional reference for understanding the behavior of variables related to employability and university employment, relying on information and indicators produced under international standards. Regarding that, the OEEU propose the following objectives:

1. To know the evolution of the employability and employment, and its characteristics related to university graduates.
2. To develop a system and a uniform methodology for measuring indicators of employability and employment of graduates.
3. To generate information on the employability and employment of university comparable between regions, branches of study (knowledge) and professional profiles.
4. To support the development of strategies and employment policies for universities.
5. To understand the mechanisms and actions of the Spanish Universities to promote employment and employability of their graduates.
6. To provide information to universities that favors the adjustment of academic supply and training demands of the labor market.

These goals of study and analyze the employability and employment in the Spanish Universities and Spanish graduates respond to the lack of public and transparent systems for tracking information about employment and employability in a wider scope than the current, the lack of culture about data-driven making decisions [17], the Institutional Intelligence and the Academic Analytics in the Spanish Universities. The study of employability and employment of Spanish graduate students and the possible results of the study like the identification of employability trends, the employment factors related to the education and competencies received from the University and the graduates' personnel context, can be used to develop new strategies [14] for enhancing and improving the employability and employment of graduate students as well as to take it into account in the Universities' curricula and education procedures, especially in this crisis context [1].

In order to achieve these goals and the meta-goal of becoming the referent information source in Spain for understanding the behavior of variables related to employability and university employment, the Observatory is developing, implementing and exploiting several parallel data-driven products [16] and strategies that will be publicly disseminated at the end of 2015.

The aim of this paper is to present this OEEU's approach to collect, analyze and exploit the data and information about employment and employability factors of the spanish university graduates. For that reason, the paper is structured as follows: the second section (Main considerations of the Observatory) explains the most relevant theoretical, technical and management issues of the Observatory. The third section (Information sources and data acquisition strategies) explains how is gathered the data used by the Observatory. The fourth section (The OEEU's information system) explains how the information system that supports the operations of the Observatory is built and organized. The fifth section (First results) explains the initial results of the information system. To end the paper, there is another sixth section with the conclusions of the work.

## **2. MAIN CONSIDERATIONS OF THE OBSERVATORY**

This section presents the main considerations, issues and challenges related to the Observatory; they could be assumed as the objective framework in several ways: theoretical, technical and management procedures.

From the theoretical point of view, the Observatory project is a data-driven project that aims to serve as a knowledge management system [6] about the current state of many key aspects of the Spanish University and the students' employability. The focus of the data-driven approach is to gain knowledge and wisdom from the data retrieved from universities and students [1; 18; 19]. Regarding the possible use of the knowledge and wisdom retrieved, the Observatory aims to help the Spanish University System to plan and improve its organizational structures, metrics and results regarding to the students preparation [10], employability and satisfaction with their higher education. So, it is possible to affirm that the project is in the scope of emergent areas like the Academic Analytics [2; 4; 5; 9] or Institutional Intelligence.[7; 15]

The data that the OEEU tries to collect and exploit is very sensitive, they contain information of students like gender, studies, employment, wage, the state where they live, etc. For this reason, the Observatory must to anonymize and aggregate the information as much as possible, so no one can distinguish the individuals behind the provided data [3]. Not only there are sensitive data about students, but also about universities: knowing the studies performed by students, their wage, employment rate, employability factors, etc. The targets that will know the conclusions of the OEEU's studies could infer the best universities, degrees and conditions for later get employment. This data, that should be public, is nowadays not public in Spain, and can affect the Universities' enrollment ratios and not all universities accept this kind of transparency (at least at this moment of crisis). These two examples are not the unique sensible information that introduces real risks for the Observatory, at least in the management.

Among the risks related to the sensitive information, there are not only aspects related to the bad (or non-desired) use of data themselves, there are also many other implications:

- The Universities, that want to protect their most sensitive data, trust the Observatory for keeping them safe and anonymized. If the Observatory does not assure these conditions, the Universities will not trust on the project and they will not provide their information and data to the Observatory. Also, there are universities that are not involved in the first stage of the project, but if the

Observatory accomplishes these tasks and considerations, and knowing the real value of the information, metrics and analysis of the Observatory, they could be involved in the future.

- The data should be gathered and saved complying the Spanish Personal Data Protection Act (usually named in Spanish as LOPD [11]). This requires that the information provided of the Observatory to users, general audience, etc. couldn't allow identifying the individual person behind the data without the strict permission of the data owner. For that reason, if the data allow identifying a person, the Observatory and its Information System is responsible for anonymizing and aggregating the information in order to avoid these legal issues. Also, the Observatory is avoiding the most sensitive data like ID Numbers, names, etc.
- Considering that the Observatory's data-driven product is a singular initiative in the national scope (there are no others like this running in Spain at this moment, only regional and local initiatives), the data gathered by the OEEU is very valuable, so the security, confidentiality and access control of the information is mandatory [3].

These and other considerations (like proving the real value of the information, the inclusion of the metrics and results of the Observatory analysis in the public Higher Education workflow, etc.) are the initial postulations of the project and the Observatory, and are taking into account very carefully by all the stakeholders involved.

### 3. INFORMATION SOURCES AND DATA ACQUISITION STRATEGIES

In order to retrieve the information required for the Observatory purposes, in the design of the OEEU actions were defined the data acquisition strategies and the ways to collect the data. These strategies define the timeframes, conditions and information sources that will feed the Observatory's Information System and will be the *fuel* of the studies and analysis that it will perform.

About the data that will be gathered, it was necessary to define the proper information characteristics and conditions. Among these characteristics and conditions, it is possible to highlight the following:

- The Observatory will collect the data in the first stage from two main information sources: the universities and the students. The universities will provide administrative data about the graduated students by uploading a file with the information to the Observatory's Information System, and the students will complete this information providing their *personal* data, regarding to their employment, knowledge gained during the degree, acquired competences, etc. by filling a questionnaire. This questionnaire could vary depending their university, employment data, etc. and try to gather all the possible information about them.
- The information collected by the Observatory is organized in more than 500 variables; of these variables the Observatory previously defined more than 100 as bank of knowledge and another more than 400 variables could be retrieved from the universities and students. The bank of knowledge gathered by the Observatory includes information about Spanish universities, degrees, competencies, employment characteristics, wage rates, employment business fields, etc. The variables retrieved from the universities and students include data about studies (duration, average score, etc.), professional competencies acquired, professional experience, work experience gained during these years, etc. Table 1 shows the different main areas of data and information that the Observatory covers in this moment (left row), and the different information sources that contribute to each of them.
- The information will be segregated and collected regarding the academic year when students end their degrees. This differentiation in years will help the Observatory to measure and track the evolution of information over the years allowing also to incorporate to the studies external variables like changes in education plans, demographics, etc. The collection process of this information will follow up also this segregation; each new round of data collection will cover certain academic courses.
- It was also defined the collection conditions (formats, structures, relationships, etc.) in order to assure the quality of the data and the proper conditions to manage and analyze it.
- Regarding the timeframe, the Observatory delimits that the data retrieved from the universities and students will be segregated by the academic course when the students finished their degrees.
- The collection period is not closed to specific dates. It is available for all universities that share their data with the Observatory for a flexible period, depending on each case and the easiness of contacting with the graduate students, their availability for responding the questionnaires, etc.

**Table 1. Main areas of data variables in the Information System and information sources that contribute to each of them**

	Data variables provided by the universities	Data variables provided by students	Data variables and classifications present in the OEEU's bank of knowledge	Data calculated from the other information sources	Total
Personal context and social-demographic data	5	3	5	1	14
Data related to parents	6	0	8	0	14
Residence	2	2	5	1	10
International experience after studies	0	4	0	4	8
Data about studies and university	3	0	13	0	16
Performance in studies	5	0	1	1	7
Practices	7	0	0	6	13
Participation in national/international exchange programs	7	0	5	1	13
International experience during the studies	0	1	0	5	6
Employment during the studies	0	3	0	3	6
Other studies	0	12	1	2	15
Data about opinions and assessment of learning and methods of the University	0	10	0	0	10
Satisfaction with the studies	0	7	1	0	8
Generic competencies	0	105	0	9	114
Competencies related to finding a job	0	16	0	0	16
Data about foreign languages	0	19	0	0	19
Employment path and current status of employment	0	34	6	12	52
Adequacy of studies and job	0	6	0	0	6
Satisfaction with the employment	0	12	0	0	12
Data about finding a job and getting a contract	0	100	0	0	100
Company information	0	5	6	4	15
Other variables contained in the bank of knowledge	0	0	60	0	60
Total	35	339	111	49	534

#### 4. THE OEEU'S INFORMATION SYSTEM

The Observatory's information system is the technological product behind the acquisition, management and analysis of the data. This information system ensures the veracity and quality of data, as well as the proper distribution of information and mainstreaming of the OEEU's analysis results. Several tools that aid the Observatory to achieve its informational objectives compose it. Among these tools and systems could be highlighted the following:

- **Administrative Data Processing Unit:** This subsystem allow the universities to upload their data in CSV format, validating it and incorporating it to the Observatory's database for the further integration with other data sources and the analyze it.
- **Questionnaires Unit:** This subsystem generates personalized questionnaires for each student based on their university, administrative data, and other issues. In order to preserve the students' privacy, the information system generates the questionnaire with a unique URL for each student, and their university takes over the task of sending each student its personalized questionnaire (at least the URL). These questionnaires can present between 30 and 70 questions (that could include other subquestions and suboptions) to the students depending on their responses, university, current employment status, etc.
- **Data Processing Unit:** This subsystem ensures the proper administration of data; it is responsible for example of the data cleaning, aggregation, anonymization, indexing, classification, etc.

- Data Analytics Unit: This subsystem leads the process of the data analysis. This processing unit performs the mathematical, statistical and data processing operations on the retrieved information by the collector units (Administrative Data Processing Unit and Questionnaires Unit). This unit responds to the OEEU's needs and goals on data analysis.
- Data Presentation Unit: The data presentation unit is the responsible of presenting the data to users and general audience of the Observatory. This unit manages the data presentation slants depending on the user profile and access rights to the OEEU's knowledge. Regarding users and their access rights, there are three main user roles:
  - Administrators: They have access to make queries on OEEU database, the defined data experiments and the results of the performed data analysis.
  - Universities: They have the same access rights of the Administrators, but only in the scope of their own data (administrative data of the university and data gathered from its students).
  - Main public: Non-registered users have access rights over the most relevant public information and aggregated results of the data analysis on the Observatory's less sensitive data.

Also, apart of these main user roles, other ones could exist that mix the permissions of those stakeholders above defined.

At this moment, this unit only provides organized information and feedback to the users in a low-level way (data tables, main metrics and some graphics), but in the next months this presentation unit will allow the users to explore the data and analysis results using the Visual Analytics principles and techniques [12; 13].

- Data Interoperability Unit: This logical unit is responsible of the proper distribution and communication of the data between the subsystems included in the OEEU's information system. At this moment, this unit does not allow the use of its APIs and other resources by third applications and systems, but in the future could be the manager of the public information endpoints and data APIs.

## 5. FIRST RESULTS

Despite the Observatory is running from 2013, its information system started to be built in 2014 and started to gather data at the end of 2014 (collecting administrative data of the universities) and beginning of 2015 (collecting data about students through the questionnaires).

In these months of data gathering, the OEEU's information system has reached some promising initial results:

- The Observatory has involved 49 Spanish Universities (public and private) in the initial stage of the project and first data analysis steps. Many others were interested in participate, but finally did not do.



Indica qué importancia tenían en tu universidad los siguientes métodos de enseñanza y aprendizaje:

1 Poca importancia, 4 Mucha importancia

	1	2	3	4
Asistencia a clase	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trabajos en grupo	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Participación en proyectos de investigación	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Prácticas en empresas o instituciones	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Conocimientos prácticos y metodológicos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Teorías, conceptos y paradigmas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El profesor era la principal fuente de información	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
El aprendizaje se basa en proyectos o problemas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Trabajos escritos	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Exposiciones orales	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Si pudieras volver en el tiempo, con la experiencia y conocimiento que has acumulado, ¿volverías a cursar los mismos estudios?

\_\_\_\_\_

Indica el nivel en el que te encuentras satisfecho/a con la formación recibida

1 Nada satisfecho, 4 Muy satisfecho

	1	2	3	4
El profesorado	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los contenidos de tu titulación	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Las metodologías educativas	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Los servicios de apoyo al estudiante (orientación, búsqueda de empleo, etc)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
La formación recibida en general	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Continuar

**Figure 1. Example of the Observatory's questionnaires for students**

- These 49 universities provided administrative data to the information system of more than 134,000 graduate students (graduated in the 2009-2010 season). These administrative data include around 40 data variables that include gender, age, nationality, parents' studies, employment conditions of parents, place where students live, average score of the students in their studies, data about students' mobility, internships and placements, etc.
- Of the 134,000 students, around 12,000 started to respond the questionnaire, and of them more than 9,000 end it (the finalization rate varies between 73-74%). These students provide the most valuable information to the information system, they provide about 400 data variables related to their employment (type, relation with their studies, wage, methods used to get the employment, data about their employer, etc.), competencies (acquired in their studies, acquired in other contexts, required by their employment, etc.), data about other studies they made, opinions (about their university, their studies, etc.), ratings about their studies, mobility over the world during and after their studies, languages spoken, etc.
- The Observatory has built a knowledge bank that contains information about Spanish degrees, universities, fields of employment, relevant information of companies as employers, information about countries over the world (used for students' nationality, or mobility around the world related to studies or work, etc.), etc. Some of the information prepared and organized by the Observatory was public previously, but other was not organized never before for this exploitation or use in an information system like this. Some of the information prepared and organized by the Observatory was public and properly set previously by organizations like the

Spanish Government or the ISO's rules, but other information was not prepared never before for its exploitation or use online in an information system like this.

Regarding the technological product (the information system) that supports the tasks and goals of the Observatory, the firsts results are:

- The Observatory has tools for collecting raw administrative information from the universities and its students.
- The Observatory has a system that generates personalized questionnaires for each student registered by the universities (figure 1). This personalization is based on the questionnaires responses, university, employment information of the student, etc. The Observatory has a system for data presentation and data analysis that deals with the information provided by the different stakeholders (universities, students, Observatory) and offers to the different user roles dashboards (figure 2) and websites that show the knowledge discovered by the system.

## 6. CONCLUSIONS

The Observatory for University Employability and Employment is developing methods and strategies for analyzing the employability of Spanish graduate students and become the reference source of information and knowledge about it in Spain.

To achieve that, the Observatory is building an information system that collects, treat and analyze administrative data from the universities and information gathered directly from the students through questionnaires. This information system deals with more than 500 data variables that cover many information areas like students' competencies, their employment, their preparation for professional career, etc. During the first months of technical development, the Observatory has built several tools and subsystems that enable collecting and processing data, and initial ways to discover knowledge from this data. Using the collected information and the results of the data analysis, the Observatory is able to discover the lacks of the Spanish Universities and the Spanish Educational System regarding employability, and propose future ways to improve them from the point of view of the Management and Policy of the universities also giving to the society new tools and ways of gain knowledge and improve the transparency of this institutions.

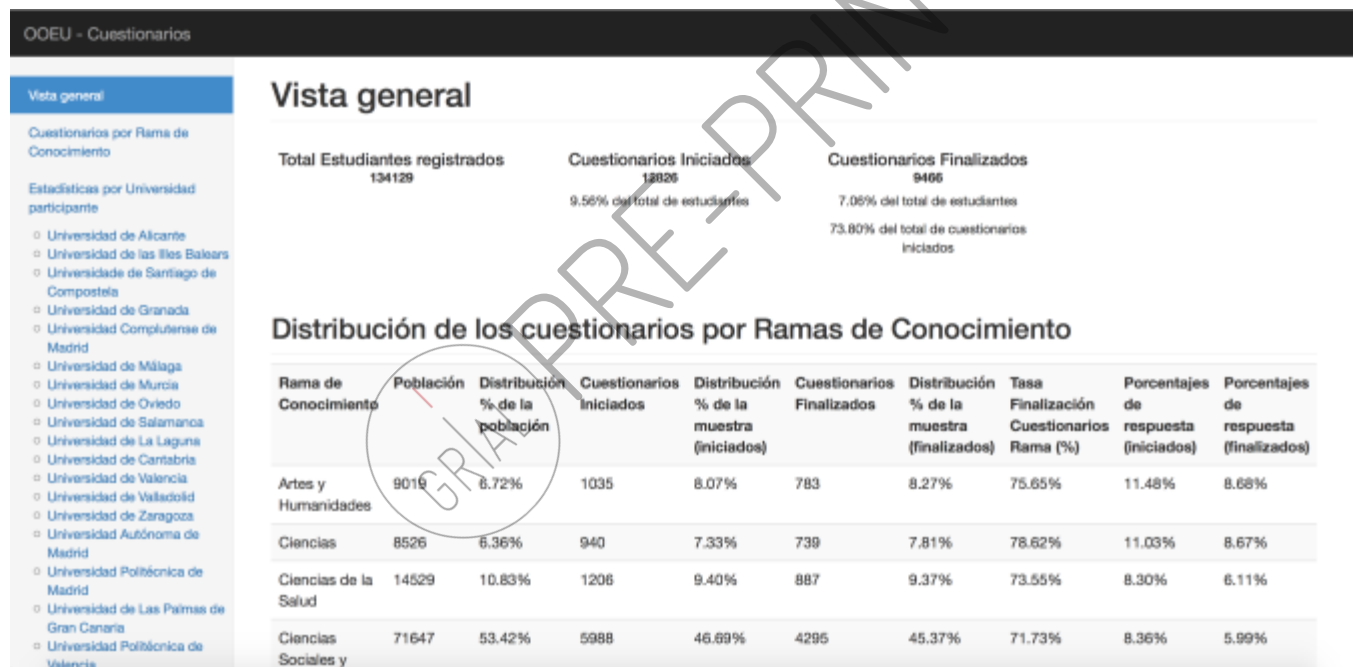


Figure 2. An example of the main dashboard for administrators and managers of the Observatory

## 7. ACKNOWLEDGMENTS

Article written under the scope of the "Observatory for University Employability and Employment" project, developed by the UNESCO Chair in University Management and Policy, the GRIAL Research Group of the University of Salamanca, and the Conference of Rectors of Spanish Universities (CRUE), and supported by "La Caixa" Foundation. This publication only reflects the views of the authors, and the participant organizations cannot be held responsible for any use that may be made of the information contained therein.

## 8. REFERENCES

- [1] Alcolea Picazo, J. and Pavón De Paula, S., 2013. Los datos como recurso estratégico Libro Blanco Inteligencia Institucional en Universidades (pp. 17-43). Madrid, Spain: OCU (Oficina de Cooperación Universitaria).

- [2] Baepler, P. and Murdoch, C.J., 2010. Academic analytics and data mining in higher education. *International Journal for the Scholarship of Teaching and Learning* 4, 2, 17.
- [3] Barlow, M., 2014. *Innovation, Security, and Compliance in a World of Big Data*. O'Reilly.
- [4] Bichsel, J., 2012. *Analytics in higher education: Benefits, barriers, progress, and recommendations*. EDUCAUSE Center for Applied Research.
- [5] Campbell, J.P., DeBlois, P.B., and Oblinger, D.G., 2007. Academic analytics: A new tool for a new era. *Educause Review* 42, 4, 40.
- [6] García-Holgado, A., Cruz-Benito, J., and García-Peñalvo, F.J., 2015. Comparative analysis of the Knowledge Management in Spanish Public Administration. In *Proceedings of the III Congreso Internacional sobre Aprendizaje, Innovación y Competitividad - CINAIC 2015* (Madrid, 2015).
- [7] García-Peñalvo, F.J. Inteligencia Institucional para la Mejora de los Procesos de Enseñanza-Aprendizaje. *GRIAL Research Group*, from <http://repositorio.grial.eu/handle/grial/406>.
- [8] García-Peñalvo, F.J., Rodríguez-Conde, M.J., Seoane-Pardo, A.M., Conde-González, M.A., Zangrando, V., and García-Holgado, A., 2012. GRIAL (GRupo de investigación en InterAcción y eLearning), USAL. *IE Comunicaciones: Revista Iberoamericana de Informática Educativa*, 15, 85-94.
- [9] Gómez Aguilar, D.A., García-Peñalvo, F.J., and Therón, R., 2014. Analítica Visual en eLearning. *El Profesional de la Información* 23, 3, 233-242.
- [10] Greller, W. and Drachsler, H., 2012. Translating learning into numbers: A generic framework for learning analytics. *Journal of Educational Technology & Society* 15, 3, 42-57.
- [11] Jefatura Del Estado. España, 1999. Ley Orgánica 15/1999, de 13 de diciembre, de Protección de Datos de Carácter Personal (LOPD).
- [12] Keim, D., Andrienko, G., Fekete, J.-D., Görg, C., Kohlhammer, J., and Melançon, G., 2008. *Visual analytics: Definition, process, and challenges*. Springer.
- [13] Keim, D.A., Kohlhammer, J., Ellis, G., and Mansmann, F., 2010. *Mastering The Information Age-Solving Problems with Visual Analytics*. Florian Mansmann.
- [14] McAfee, A., Brynjolfsson, E., Davenport, T.H., Patil, D.J., and Barton, D., 2012. Big data. The management revolution. *Harvard business review* 90, 10, 61-67.
- [15] Oficina De Cooperación Universitaria. Ocu, 2013. Libro Blanco Inteligencia Institucional en Universidades Madrid: OCU.
- [16] Patil, D.J., 2014. *Data Jujitsu: The Art of Turning Data into Product*. O'Reilly.
- [17] Patil, D.J. and Mason, H., 2014. *Data Driven. Creating a Data Culture*. O'Reilly.
- [18] Rowley, J., 2007. The wisdom hierarchy: representations of the DIKW hierarchy. *Journal of information science* 33, 2 (April 1, 2007), 163-180. DOI= <http://dx.doi.org/10.1177/0165551506070706>
- [19] Zeleny, M., 1987. Management support systems: Towards integrated knowledge. *Human systems management* 7, 59-70.

