Development of an Open Data Portal for a University
Experience from the University of Alicante

Jose Vicente Carcel, Andrés Fuster, Irene Garrigós, Francisco Maciá,
Jose-Norberto Mazón, Llorenç Vaquer, and Jose Jacobo Zubcoff
Universidad de Alicante, Alicante, Spain

Keywords: Open Data, University, Data Architecture, Data Management System, API.

Abstract: University of Alicante (UA), in Spain, is aligned with an Open Government strategy. Within this strategy, UA is carrying out the OpenData4U (Open Data for Universities) project which aims to provide mechanisms for opening data from universities, finding out how open data contributes to open government in universities. This project encourages reusing open data, not only for the sake of transparency, but also as a basis of novel data-intensive business models that universities can foster. This paper describes one of the outputs of the project: an approach for opening data from universities keeping in mind data quality criteria, and tailored to no specific technological scenario. This approach allowed UA to launch its open data portal http://datos.ua.es that is also reviewed in this paper. Also, some research challenges related to university open data are enumerated.

1 INTRODUCTION

Citizens demand an increasingly transparent behaviour of public institutions based on Open Government principles. Importantly, transparency implies that public data should be available with the aim of providing the greatest benefit to the wider society through an active participation of citizens (through watchdog organizations or consuming data directly or using some useful application fed with open data). Therefore, public data should be freely available to be easily used, reused and redistributed by anyone, i.e. data should be opened (OKFN, 2014).

Universities are complex institutions with a variety of heterogeneous data that must be collected, classified and opened, ranging from (anonymized) data about students and staff, to geographical campus data or financial data, etc. Importantly, universities are one of the main sources of knowledge for the society, such as open courses or open access research resources. This knowledge is mainly based on data produced in the own universities. Therefore, universities are an attracting and rich scenario in which open data initiatives must be promoted, not only for improving accountability, but also for better disseminating knowledge and for

contributing in the creation of novel business models around data.

Bearing in mind these considerations, University of Alicante (UA) has started a project called OpenData4U (Open Data for Universities) which aims to provide mechanisms for opening data from universities, finding out how open data contributes to open government in universities and encouraging reusing open data (since, reusing is at the core of the concept of data openness). One of the milestones of the project is the development of an open data portal. This software solution could be a challenge by a university, since adopting new software solutions often faces technological constraints for the sake of security, accessibility, and other non-functional requirements that must be accomplished. Therefore, an open data portal for a university must be based on an easily-to-deploy solution regardless the current adopted technology (i.e. suitable for any technological scenario), and effortlessly to maintain by the university in the long term.

Contributions of this paper are as follows: (i) describing the goals of the OpeData4U project related to transparency, collaboration and participation, (ii) sharing the experience in opening data and encouraging data reusing of a university through the OpenData4U project, (iii) defining the variety of university data and their nature, and (iv)

describing the technical solution for developing the UA open data (http://datos.ua.es).

2 AVAILABLE DATA MANAGEMENT SYSTEMS

Opening data of (public) institutions requires using a system that helps in providing mechanisms for exposing data in an easy manner and facilitating their reuse. These mechanisms include a friendly front-end with a data catalogue, application catalogue, APIs, and other resources for encouraging reusing data by citizens, developers and the own institution. Such systems are the so-called data management systems (DMS). Lately, several DMS have appeared based on different architectures and technologies that provide solutions for creating an open data portal. Table 1 shows the most important initiatives such as CKAN (http://ckan.org), DKAN (https://www.drupal.org/project/dkan) and Socrata (http://www.socrata.com) and summarizes their main technical features:

- Maturity: understood as stability of the system and frequency of released versions, size of its community of developers and support. CKAN and Socrata are ahead in this feature, while DKAN maturity is rather low.
- Interface: the most customizable interface and easy-to-develop system (since it has a lot of usable modules) is DKAN, since it uses the widely-known and well-documented Drupal system, while CKAN has customization limitations. Open version of Socrata does not provide an interface.
- Storage: storing catalogue metadata is an important point in a DMS. The most flexible solution is DKAN because various databases can be used according to specific project requirements.
- Language: DKAN uses PHP, while CKAN use Python. We have no data on which programming language is used by Socrata.
- License: CKAN and DKAN have open source licenses, and Socrata offers both open and proprietary.

In general to adapt any of the above DMS for a specific project, the most flexible solution seems to be DKAN, since development can be done by using PHP within Drupal and it can be connected to any database. Also, it has an open source license. However, this DMS has a rather low maturity level

which makes it hard to use for a public institution project.

Table 1: Summary of DMS technical features.

	CKAN	DKAN	SOCRATA
Maturity	high	low	high
Interface	Pylons	Drupal	-
Storage	postgreSQL	any	postgreSQL
Programming language	Python	PHP	-
License	open	open	open or proprietary

All public institutions, including universities, have a strategic plan that should be supported by using certain technology. Therefore, IT department must decide which is the best technology to use in each project to fulfill the goals and requirements stated on the strategic plan. Technology adoption must be driven not only by functional requirements, but also by non-functional requirements such as the IT department experience, security issues, IT budget, etc. Therefore, opening data must be aligned with the strategy of the university and selected technology must ensure the continuity of the project through time. The key issue is thus considering the technological scenario of the institution in order to select the right DMS. Actually, the technological scenario at the UA has been developed by following a long-term strategic plan and it is responsibility of the IT department:

- There is an in-house developed CMS called Vualá (https://vuala.ua.es) at the UA where all institutional websites and web applications are hosting.
- IT department at UA has expertise on Oracle DBMS.

This specific technological scenario at UA makes difficult to adopt one of the aforementioned DMS for its open data portal. Therefore, the open data portal of the UA requires a flexible DMS solution for leveraging the university infrastructures, resources and technologies adopted to achieve its strategic plan. Taking into account the collected facts about existing DMS summarized in Table 1, the following decisions were made:

- Discarding using CKAN because its Pylons front-end framework is not fully customizable to be integrated with Vualá. Also, CKAN database storage is only postgreSQL.
- Although DKAN allows us to implement a database layer with any database, its maturity level hinders its use in the UA.

Integrating Drupal CMS with Vualà may also present some difficulties.

- Socrata has a high maturity level but its database storage is postgreSQL and it has a commercial license.

Therefore, UA decides to create its own DMS for developing its open data portal aligned with its strategic plan in order to leverage its infrastructures and technology know-how.

3 OPEN DATA AT UNIVERSITY OF ALICANTE: DATOS.UA.ES

Open Government initiative of University of Alicante (UA) is a crucial aspect of IT issues within its strategic plan 2014-2019 (PEUA40, http://web.ua.es/es/peua/plan-estrategico-ua-40.html). One of the priority projects is called OpenData4U (Open Data for Universities) and it aims to encourage reusing data from UA. One of the first milestones of this project is the development of the UA's open data portal: http://datos.ua.es.

3.1 Openness before Open Data Portal

From some time ago, UA has been launching several initiatives related to open government. These initiatives are instilling an "openness" spirit in the university community that is getting aware of the convenience of having a transparent, participative and collaborative university. These initiatives are supported by a strategic IT project portfolio. This project allows UA board to support an overall decision making process on the IT projects to develop. Also, university community is informed about criteria, decisions and current status of the projects.

There are several IT projects related to datos.ua.es and they strategically complement each other. For example, "Universidad en cifras" (http://utc.ua.es/es/datos/la-ua-en-cifras.html) where several quality indicators are shown and analyzed for the sake of improving university; Geographical Information System of University of Alicante (SIGUA, http://sigua.ua.es/), Future Education Portal (UAedf, http://uaedf.ua.es/) where all teaching knowledge that belong to UA are brought together and opened independently of their content, nature or platform (OCW, http://ocw.ua.es/ or MOOCs), or the UA Repository (RUA, http://rua.ua.es/) that offers open access to documents (teaching or research) generated by UA

staff and students doing degree, master or PhD. thesis.

3.2 OpenData4U Project Overview

The OpenData4U project is being carrying out by the vice-chancellor office for information technology and the WaKe research group from UA and it is supported by several UA actors: IT department, vice-chancellor office for infrastructures and sustainability, vice-chancellor office for quality, university library, GIS department, etc.

This project aims to develop a successful information reusing model at UA. This model must favor the growth of a market of entrepreneurs (last year students and former students) with the goal of taking advantage of open data to build a successful business plan. In this way, OpenData4U tries to support students to find a quality job and to develop an increasingly productive and sustainable IT sector by means of supporting the creation and commercialization of added-value services. Specifically, OpenData4U project raises a set of actions, namely:

- Alignment with national and international initiatives as the CRUE-TIC (http://www.crue.org/TIC), OKFN (https://okfn.org) group on open universities or the Spanish national open data portal (http://datos.gob.es).
- Development of a data map, including a process to identify and classify those data prone to open.
- Starting datos.ua.es: the open data portal which collect and expose all the data coming from UA and related applications.
- Boosting an innovative market through a set of actions oriented to instill open data philosophy in potential reusing actor. These actions include ideas contests, hackathons, unconferences, and also meetings and courses about open data and open government in a university.

At the core of OpenData4U project is the open data portal of the University of Alicante which will be next described in detail.

3.3 Motivation and Goals of datos.ua.es

By publicly exposing their data through datos.ua.es, UA makes a commitment to provide transparency and trust and to help in generating value to the society. These data can be consumed by several

actors: university community, former students, citizenship, companies (or entrepreneurs) that reuse data to add value, and the own university. Encouraging data reusing is a crucial mission for datos.ua.es, having a twofold goal:

- Generating a social value to citizenship, by fostering transparency and accountability through easy access to data generated in UA.
- Fostering a positive economic impact in the UA environment through novel innovative business models based on intensive reusing of data. These business models should aim at adding value to open data in order to get useful services for society. Importantly, data openness is a good mean for getting opportunities for entrepreneurship, not only among the university community, but also among citizenship (including entrepreneurs and SMEs).

Furthermore, as described in Fig. 1, active listening is done in order to know the most demanding data for openness. Also, the UA internally reuses their own open data for improving process, services and support decision making.
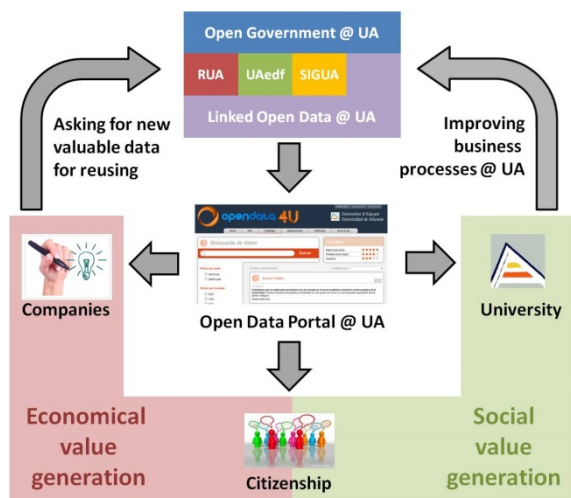

Figure 1: Open data portal within UA.

3.4 Data Catalogues: Classification, Quality and Formats

Open data portal of UA exposes its own data, making it easy to reuse. The typology of these data is wide, so the first step is developing a classification schema for data from universities.

Currently, we found interesting the following classification (CRUE-TIC, 2014):

- Organizational information:
  - Organization chart.
  - Personnel data (teachers, researchers, administrative assistant, etc.).
  - Job offers from university.
- Economic information:
  - Account status.
  - Budget.
  - Grants.
  - Tenders.
- Communication information:
  - Events organized by university.
  - News related to university.
- Infrastructure and services:
  - University buildings and other university geospatial entities (rooms, departments, etc.)
  - Rooms' equipment (capacity, accessibility information, etc.).
  - Cafeteria menus.
  - IT services
  - Library services
  - Services to students (accommodation, cafeterias, vending machines, sports facilities, cultural events, etc.
- Education:
  - Degree catalogue.
  - Calendar and timetable.
  - Mobility programs.
  - Educational contents.
- Student:
  - Profile of registered students.
  - Grants.
  - Academic performance (qualifications, etc.)
  - Student satisfaction tests.
- Research:
  - Journal and conference articles
  - Experiments data
  - Thesis
  - Projects

One of goal of the open data portal is to maximize reusing data. To do so, the so-called 5 star schema is considered in order to determine the quality of data regarding their potential for reusing. There are several data quality criteria that have been studied (Batini et al 2009) and they can be applied to open data as well, however, other data quality criteria

must be considered with special emphasis on reutilization of data, i.e., allow that data is available for someone and that someone can use and reuse them. In this sense, Tim Berners-Lee proposes the five star model, which is related to get Linked Open Data (LOD) (Bizer et al 2009). This model establishes five levels labeled with stars according to their fitness for reutilization:

1. Data are available on the Web, independently of the format (e.g., pdf file).
2. Data are published on the Web in a machine-readable structured format (e.g. using Excel format instead of a scanned image of a table).
3. Data are published in a non-proprietary format (e.g. CSV instead of Excel).
4. Data are identified by URIs (e.g. by using W3C standards such as RDF) in such a way that they are easily and persistently accessible.
5. Data are linked with other data in such a way that they are contextualized.

The highest open data quality according to this 5-star schema (datasets scoring 5 stars) corresponds to LOD and it should be the goal of an open data initiative.

Although, at this moment, most of the open data formats that you can find in datos.ua.es are: CSV, JSON, XML and PDF (scoring from 1 to 3 stars), the goal of the open data portal of the UA is reaching the highest level of quality, having a semantic infrastructure which allows the Linked Open Data management in the different university scopes. The best final scenario would be an open data portal that provides the open data to be downloaded and accessed via Web services and also having a semantic entry point to get all UA data as LOD. However, we must not forget that most available data that can be open do not have the highest quality criteria and scoring 5 stars becomes a stepwise process. UA is committed to open all available data, while working in technologies that allow evolving the quality of the current data and the data generated daily and getting UA data in some semantic formats such as RDF/XML, OWL/XML, Turtle, JSON-LD among others.

Besides the important 5-stars schema to facilitate data reuse, a quality model for open data portals, developed within the University of Alicante (Oviedo et al 2013), has been applied. This model specifies a set of measures to consider in an open data initiative beyond the 5-star schema.

3.5 Application Catalogue

Open data portal at UA has also an application catalogue which aims to help in the search of applications that have been developed by reusing open data from the university. A good example of these kinds of applications is UAlojamientos (http://datos.ua.es/es/ficha-apps.html?idApp=1) a smartphone application developed by the IT department of the UA that reuses data from the accommodation and bureau of the UA. As a matter of fact, it is worth mentioning that open data from a university can be reused for a plethora of potential applications: degree recommender according to potential student profile, menu planner on cafeterias in the campus, etc.

3.6 Architecture of datos.ua.es

Once UA decided to develop their own DMS (recall motivation in section 3), the first step to do is to study the technological and data map scenarios of the UA in detail. In this way, every aspect is considered to design an adequate architecture that supports the goals of the project. To sum up, two conclusions are obtained:

- *Technological scenario*: the CMS used in the UA is called Vualá and it uses an Oracle DBMS. Staff at IT department are expert in this technology, so the envisioned DMS must handle this scenario: (i) front-end of the open data portal must be integrated with Vualá, and (ii) the data catalogue must be deployed in Oracle DBMS.
- *Data map scenario:* different kinds of data are hosted in isolated databases implemented in different DBMS (e.g. SQLServer). In fact, there is a centralized database containing most data to be open but certain types of data are distributed through data silos across different departments and units of the university. Therefore, the envisioned DMS must also support the access to different kind of databases.

Therefore, data belonging to UA are scattered, heterogeneous and come from different sources. In order to overcome these drawbacks, UA has developed a DMS architecture having a high level of decoupling based on Web APIs. Our DMS provides a set of facilities to develop an open data portal in the context of a university by considering in any

technological scenario and taking into account the classification of data at the universities. Figure 2 shows the designed DMS architecture.
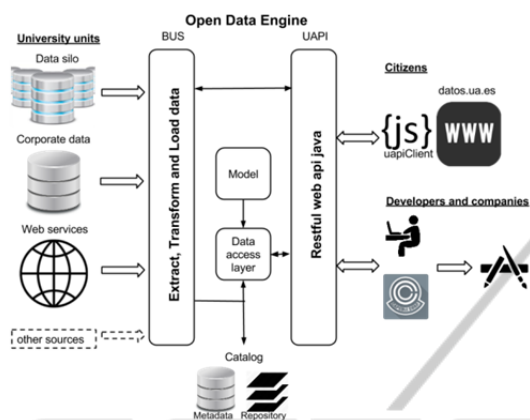

Figure 2: DMS architecture for datos.ua.es.

The "Open Data Engine" is at the core of the DMS. It has been developed in Java and served by Tomcat. In terms of functionality, it has two main parts:

1. UAPI (exposing data): our engine provides a REST API called UAPI for querying the metadata (JSON format) and accessing to open data files. UAPI communicates with catalogue through a data access layer. In addition, a Javascript client API called uapiClient has been developed to easily implement an interface that offers a high degree of decoupling for communicating with UAPI. Data are exposed to two main actors:
   a. *Citizens*: the website of the open data portal (http://datos.ua.es) has been developed within "Vualá" by using the Javascript uapiClient.
   b. *Developers (including companies)*: our REST API has been developed to allow developers to consume data and develop applications.
2. BUS (collecting data): our engine has a logical bus to extract, transform and load (ETL) data from a variety of databases and web services from UA in a transparent way. Our bus uses data integration technologies (ETL processes) for addressing these heterogeneous sources. Also, through the UAPI the bus can be launched to collect data. This feature enables the development of a data federator.

In summary, this architecture meets the decoupling requirement (by means of the open data engine and its APIs). It also solves the scattered and heterogeneous data sources issue by using the bus. Also, it is worth mentioning that this solution would be exportable to other universities because their problems are pretty similar.

4 SOME DETECTED RESEARCH CHALLENGES

When carrying out the development of the open data portal, we found some research challenges that is worth mentioning as an avenue for future work. They are briefly described in this section.

Model-Driven Development For Automatic Code Generation: metadata model should be generic for every software that handles open data, because they have many features in common. E.g. all open data portals need a database which allows store metadata into a filesystem to store different kinds of files and this information must be accessible by any client (browser, application or service). All these common features could be analyzed and modeled in order to automatically generate the required code within a data architecture. Also, this model-driven development approach could generate the code required to publish open data (i.e., a complete CRUD for open data).

Semantic Lifting To Get 5 Stars In University Data: impact that Semantic Web technologies have in the universities must be analyzed. Semantic web can aid to improve education and research because linking different data sources and content (internal or external) may enrich the ways in which educational and research resources are used. There are some interesting projects in this sense, e.g., ResearchObjects (http://www.researchobjects.org) that aims to package research results and workflows in a semantically-rich format (a research object) for the sake of sharing and reusing research. However, semantic annotation of university data is not a trivial task, since interoperability schemas need to be defined for any of the university domain (education, research, etc.).

Data Search Engines: as happened with the search of documents at the beginning of the WWW, there are no effective tools for searching open data. There are only portal-based open data tools such as

http://publicdata.eu. These are containers of structured and categorized links to facilitate access, thus providing a simple filter for areas or categories. The same scenario already happened with Web portals, like Yahoo!, in the 90's. This procedure is not very scalable and effective and may return data sets of low quality (inaccurate, incomplete, etc.). Therefore, search mechanisms for accessing and reusing useful datasets are needed. These tools should be independent of the existing open data portals and should be guided by predefined quality criteria (Oviedo et al, 2013).

5 RELATED OPEN DATA INITIATIVES FROM UNIVERSITIES

Many national governments are opening their data not only for the sake of transparency, but also, to encourage co-creation of valuable services among entrepreneurs. Data is thus an asset for participation and collaboration in order to increase economy. However, main actors of "open data" are not only national governments and other public institutions have an important role to play, since they have some interesting data to reuse. e.g., http://datos.gob.es/, http://opendata.aragon.es/, http://www.opendatalatinoamerica.org, http://opendataforafrica.org/, https://www.data.gov/, https://open-data.europa.eu.

One of the institutions that better fits with "open data" philosophy is a university. The university community produces a lot of data that can be reused at different levels (ranging from scientific to education) to improve the region and country where settled. Specifically, OpenData4U project aims to improve services in the University of Alicante and provide the university community with all generated data and mechanisms to reuse it.

Interestingly, the great part of the initiatives of open data portals in universities is concentrated in Europe. The country having more initiatives is UK, with sites like http://data.southampton.ac.uk (University of Southampton), https://data.ox.ac.uk/ (University of Oxford) or http://data.bris.ac.uk/ (University of Bristol). Spain has also universities that bet on open data portals: http://www.upo.es/datos-abiertos/ (Universidad Pablo de Olavide), http://datosabiertos.unileon.es (Universidad de León), http://dkh.deusto.es/ (Universidad de Deusto), http://data.upf.edu/ (Universitat Pompeu

Fabra). Finally, our initiative from Universidad de Alicante: http://datos.ua.es/ . Outside Europe we can find few open data portals from universities. Actually, one of the few examples is the University of Waterloo in Ontario (Canada), https://api.uwaterloo.ca/ that provides an API to access some data via web services: https://api.uwaterloo.ca/

Next, main open data portal for universities are briefly described in order to check differences between them and highlight their most important features.

Table 2: Main features of related open data initiatives from universities.

	LOD	SPARQL endpoint	Apps	Filesystem	Community
University of Münster	YES	YES	NO	NO	NO
U. of Southampton	YES	YES	YES	YES	NO
U. Pompeu Fabra	SOME	YES	NO	YES	YES
U. Pablo Olavide	SOME	YES	NO	YES	YES
Linked Open Aalto Data Service	NO	YES	NO	NO	NO
U. of Mexico	NO	NO	NO	YES	NO

Open Data Portal at the University of Münster, German (http://lodum.de/). One of the most interesting issues of this portal is its SPARQL Endpoint. The SPARQL language provides an efficient way to access Linked Open Data sources. However, this portal misses an application catalogue.

Open Data Portal at the University of Southampton, UK (http://data.southampton.ac.uk**).** It has a complete website that includes all functionality like SPARQL Endpoint, catalogue of applications, data catalogue. However, a smooth communication with users is missing.

Open Data Portal at the University Pompeu Fabra, Barcelona, Spain (http://data.upf.edu/). This university has used CKAN to develop the open data portal, this is a easy and quick solution to open the data because this DMS provides almost every requirement needed to open the data. But not all of public institutions can use CKAN. Also, CKAN miss a section with applications that reuse data. Also it has a participate section for the users.

Open Data Portal at the University Pablo Olavide, Sevilla, Spain (http://www.upo.es/datos-abiertos/**).** It is one of the pioneer universities to

launch an open data portal in Spain, and it is very complete since it fulfill basic requirements and more that should have any open data portal like SPARQL Endpoint, a channel for users to participate, a display panel data. Still they don't have an apps section, but sure that they will have it soon.

Linked open Aalto data service portal, Helsinki, Finland (http://data.aalto.fi/). This portal aims to open their data for reuse and transparency, fulfills with standards of w3c semantic web to provide rich and interlinked data and allows consume web services and get datasets to make easy applications development. But they don't have a way to easy participate, the people will be able to reuse the data, must be expert in the domain. They have all linked data and sparql endpoint to consume this data.

Open Data Portal at the University of Mexico, Alburquerque, Mexico (http://opendata.unm. edu/). They have few open data, but all of them are available in a structured format and open. The data have an enclosure that explain the metadata with datasets. They have a singularity and is that you can submitted your own datasets. But the fault is that people can only participate through a feedback email and neither have a section with data reuse experiences like apps.

6 CONCLUSIONS

The University of Alicante (UA), in Spain, is aligned with an Open Government strategy. Within this strategy, UA is carrying out the OpenData4U (Open Data for Universities) project which aims to provide mechanisms for opening data from universities, finding out how open data contributes to open government in universities and to encourage reusing open data. The goal the project is twofold; (i) improving transparency and accountability, and (ii) promoting novel data-intensive business models among entrepreneurs. This paper describes one of the outputs of the project: an approach for opening data from universities. Universities that open data often deal with the following problems: (i) data should be opened fast at the same time that high data quality is demanded; and (ii) complex environments, such as universities, have technological constraints for the sake of security, accessibility, and other non-functional requirements. Our approach overcomes these drawbacks by developing a data architecture that is tailored to no specific technological scenario. This approach allowed UA to launch its open data

portal http://datos.ua.es that is also reviewed in this paper. Also, some research challenges related to university open data are enumerated.

REFERENCES

Batini, C., Cappiello, C., Francalanci, C., Maurino, A. (2009). Methodologies for data quality assessment and improvement. ACM Comput. Surv. 41(3).

Bizer, C., Heath, T., Berners-Lee, T. (2009): Linked Data - The Story So Far. Int. J. Semantic Web Inf. Syst. 5(3): 1-22.

CRUE-TIC (Sectorial de Tecnologías de la Información de la Conferencia de rectores de las Universidades Españolas). Hacia una universidad abierta. Recomendaciones desde la sectorial TIC de la CRUE. 2014.
http://www.crue.org/TIC/Documents/Haciaunauniversidadabierta.pdf

OKFN - Open Knowledge Foundation (2014). Open data definition https://okfn.org/opendata/

Oviedo, E.; Mazón, J-N.; Zubcoff, J.J. (2013): Towards a data quality model for open data portals. CLEI 2013: 1-8.