THE UNIVERSITY AS AN OPEN DATA ECOSYSTEM

JOSE JACOBO ZUBCOFF, LLORENÇ VAQUER, JOSE-NORBERTO MAZÓN, FRANCISCO MACIÁ,
IRENE GARRIGÓS, ANDRÉS FUSTER & JOSE VICENTE CARCEL
University of Alicante, Alicante, Spain.

ABSTRACT

Lately, the University of Alicante (Spain) started the OpenData4U project in order to create a methodology to easily develop open data portals for universities. Within this project, the open data portal of the University of Alicante was launched (http://datos.ua.es), which allowed us to develop a set of best practices in the implantation of an open data ecosystem for universities to foster reusing of Public Sector Information. The main aim of this paper is to describe our experience at the opening data process at the University of Alicante, as well as enumerating novel opportunities for social and economic impact in which a university could be a driving force.

Keywords: IT ecosystem, open data, open government.

1 INTRODUCTION

Public sector institutions (PSI) are moved to consider their own data as a valuable resource, not only for transparency or accountability but also for generating economical impact through data reusing. The University of Alicante, as a PSI, has been concerned during the last three years about all the implications of its data, and how opening it can be a driving force in generating social and economic impact (through an open government initiative named OpenData4U). In this paper, we describe the lessons learned from the University of Alicante together with the methodology for opening data that has been developed (that is applicable to other universities or even to other PSIs) under the umbrella of the OpenData4U project.

1.1 Government and open data

The open data movement becomes mainstream in 2009 after Obama's memorandum that highlights Open Government [1]. This memorandum arose the three building blocks of open government: (i) the government should be transparent, (ii) the government should be participatory and (iii) the government should be collaborative. Also, there exist several studies towards building an Open Government [2, 3] in which open data are considered as a cornerstone in reaching its goals. Moreover, during the last years, different definitions have been given about how data should be managed to be an open institution (i.e. an institution with Open Government) [4, 5]. To sum up, data must be open, which means that it must be freely accessible, reusable and redistributed. Thus, the open data initiatives must facilitate the reuse of open data to everyone without technical and legal restrictions [6]. Several works have been developed in this sense, e.g. "The Open Data Decalogue" [7] that proposes 10 steps plus one to open data. Also "The Open Data Handbook" [8] defined by the Open Knowledge Foundation (https://okfn.org/) discusses the legal, social and technical features of open data.

It is worth noting that open government and the open data movement are seamlessly supported by several laws, e.g. in the European context (Directive 2013/37 / EU of the European Parliament and

of the Council of 26 June 2013, by which modifies the Directive 2003/98 / EC relating to the reuse of public sector information [9]) and nationally context (Law 19/2013, of December 9, transparency, public access to information and good governance [10]).

1.2 Open university

We can define an Open University as a university that implements an open government initiative. To do so, a university must comply with the existing laws regarding transparency and data access, as well as promoting actions to generate social and economical impact from reusing open data (see Fig. 1). However, current laws and guides about how to open data in different public sectors, do not consider the complexity and idiosyncrasy of a university.These institutions have been established for many years and manage special kind of data (ranging from scientific data to teaching data, and administrative data). Actually, to the best of our knowledge, only one initial work has been developed for universities: "Towards and open university [11]" written by CRUE-TIC [12], focused on facilitating the open government in universities.

Our point of view is considered a university as an open data ecosystem. This concept by M. Heimstädt et. al in "Conceptualizing Open Data Ecosystems" [13]. This helps us to see the university as an ecosystem, made up of many actors and small organizational structures that should recognize data like the raw material that is in a cycle and is capable to feed the ecosystem providing benefits to all parties. Figure 1 depicts the initial considerations before developing a process for opening data at the University of Alicante within an open government approach. First of all, the government of the university must be committed to ensure that open data initiative is well-supported and required coordination for analysis current data and opening it is easily achieved. Also, there are other actors involved, such as (i) different units that manage the data of the institutions as well as (ii) IT staff
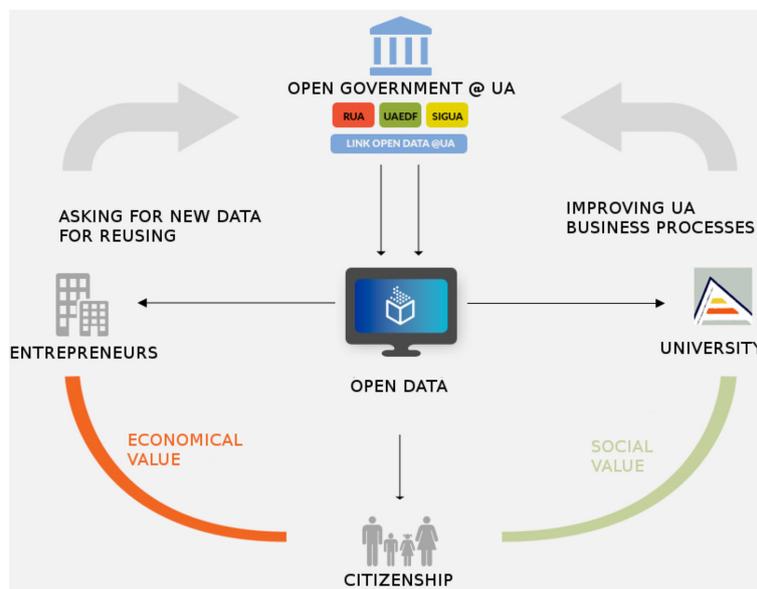

Figure 1: Overview of open data ecosystem at University of Alicante.

related to data, and also (iii) former projects related to "openness", such as institutional repositories of open access, massive online open courses platforms, or geographical information systems. Open data aims to generate social and economic value to both entrepreneurs and citizenship, which implies that feedback is obtained about which data should be open. Also, the university itself that may improve its own business process by considering insights providing by opening the data.

1.3 Contributions

When open government principles are applied in the university context, it is required to define an open data ecosystem for a university. In this way, the success of open data (and open government) projects are maximized. Therefore, the main aim of this paper is to describe the experience of opening data at the University of Alicante through the OpenData4U project, which defines an initial ecosystem, as well as enumerating novel opportunities for social and economic impact in which a university could be a driving force.

2 REQUIRED ACTIONS FOR OPENING DATA AT UNIVERSITIES

A university is a complex institution that concerns many processes ranging from teaching (including enrollment of students, assessment, etc.) to research and also economical and organizational issues. Therefore, there are different departments with different tasks from several domains and different kinds of daily activities that are responsible for a variety of data.

Therefore, one of the key issues is that the government of the university must be committed in order to ensure that open data initiative is well-supported. Furthermore, two initial actions should be started:

1. Definition of a clear project with well-defined objectives and actions, as well as a clear roadmap. The University of Alicante carried out such a project, whose main actions developed to achieve the objectives are enumerated in Fig. 2 and described next.
2. Creation of a working group with experience in open data issues. This group would be in charge of the daily activities required for accomplishing a successful open data initiative. These activities are twofold: (i) development and deployment of a technological ecosystem for publishing open data from a university and (ii) dissemination activities to promote open data philosophy and entrepreneurship around data, thus obtaining feedback from the university community. Also, the university context allows the project to count on a working group with a research profile to take advantage starting novel research lines on open data and big data related to the project.
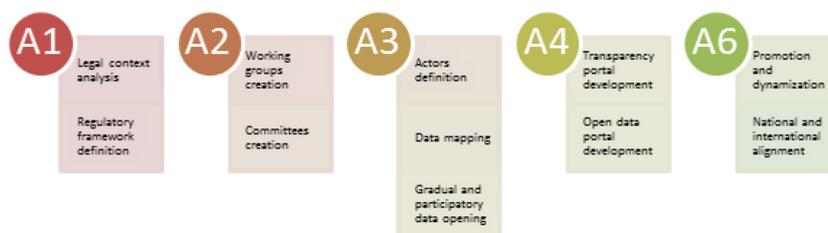
Figure 2: Actions included in the OpenData4U Project.

First of all, the project must study the existing laws and regulations regarding transparency and open data at different levels (regional, national and European) to start action A1 (see Fig. 2) and define the own regulatory framework for opening data at the university. To do so, some committees (with different levels of responsibility) are suggested to be created as follows (see action A2 in Fig. 2):

1. Transparency Committee (TC), it is in charge of ensuring the compliance with the transparency standards and regulations, as well as making decisions. It is formed by the Vice Chancellors of different areas of knowledge, thus being at the executive level.
2. Technical Committee on Transparency (TCT), it is an advisory board in charge of support TC to establish criteria for publication of open data. It is formed by heads of different units and services of the university. The TC will also elaborate guides to provide autonomy to the units, thus accelerating the open data process.

Meanwhile, the working group would start to study existing data that the university manages to further define which data from the university are likely to be opened (action A3 in Fig. 2). To do so, three tasks should be performed:

1. Identify actors that are involved in managing existing data to be opened. There are two kinds of actors: those experts in a domain but without IT skills (data domain experts with a non-technical profile) and those with IT skills that are in charge of domain data (data responsible with an IT profile). Both kinds of actors must understand the benefits of open data. Also, communication between the working group and the actors should be fluid. Actually, these relationships between actors and the working group are highly important due to the fact that data responsible should grant access to data in a secure manner, and data domain experts should define the metadata from data to provide the best information to the users. Therefore, these communications will determine the quality of open data to a large extent.
2. Create a data mapping. For that, a list with most important topics of university (organizational information, teaching, students, infrastructure and services, the economy and communication) must be set and related to their data responsible and data domain experts, with the aim of contacting them and studying the feasibility of opening data, technology requirements and the complexity of datasets.
3. Define a process to open data in a gradual and participatory manner. It is hard to open all available at once, therefore, at the University of Alicante decided to start providing a reduced data map and proposing some dissemination activities and initiatives to encourage the university
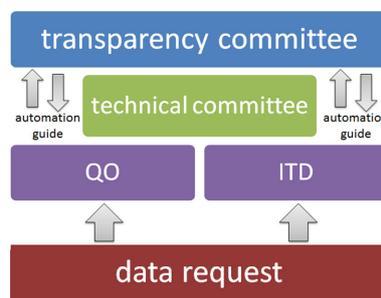

Figure 3: Process for opening data at University of Alicante.

community to reuse open data to get feedback and prioritize which data are more demanding. This process is described in Fig. 3. Both the request of data and data that is proposed in data map to be opened must be processed by experts and ultimately responsible for opening data. The Quality Office (QO) and the IT Department (ITD) of the university have a role to allow the opening of data, always with the consent of the commissions (CT and TCT). In this figure, data request also refers to the different data that are included to be opened in the different laws as an active advertising policy.

To provide open data, it is necessary several technological tools that allow everyone to access, reuse and apply for data. These tools should be aware and oriented to the users, this means that the provided data should be easily understandable and reusable. The University of Alicante has developed several tools to access data (action A4): the Open Data Portal (ODP) (http://datos.ua.es/) launched in April 2014, that offers data on open and reusable formats processable automatically by machines, perfect for developers. These portals also include an API (https://dev.datos.ua.es/) to access data using web services. But we will explain these features in the next section. The Transparency Portal (TP) (http://transparencia.ua.es/es/), launched in October 2015 and concerned to provide data of the university regardless of format. This portal offers a channel so that users can request data from university.

Finally, an important action within the process of opening data is the dissemination actions, that will give us greater visibility and pick up a very useful feedback to improve services and the opening of the priority data. Some of these actions that have been implemented at the University of Alicante are (i) a competition in which participants proposed ideas for reusing university data, this contest was a success with sixty proposals, this allowed to have knowledge about which data is the most requested by users; and (ii) a competition of application, in which participants create applications that consume data from university provided in the open data portal. Last action proposed in 2015 was an open data meeting (http://datos.ua.es/es/encuentro-de-datos-abiertos.html), where the University of Alicante invited experts in open data for discussing the current status and future trends of both (i) publishing open data and (ii) reusing open data for generating an economical and social impact. Figure 4 shows some of the existing applications that reuse open data of the University of Alicante. They can be downloaded and used. First of all, GradUAte is an application that allows potential students to select the most convenient study to enroll to (https://play.google.com/store/
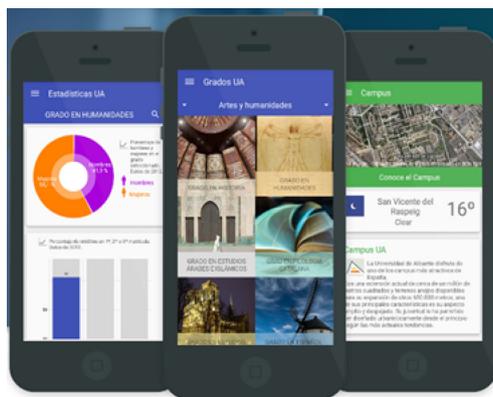
Figure 4: GradUAte mobile app that reuses open data from the University of Alicante.

Figure 5: BrujUA mobile app that reuses open data from the University of Alicante.

apps/details?id=com.pgonzalezcarrizo.comienzaua). Secondly, BrujUA is an application that helps users to move within the campus (https://play.google.com/store/apps/details?id=labrujua.labrujua).

After describing the experience and the process the University of Alicante follows for opening data, we will focus on describing the required IT ecosystem in the next section.

3 IT ECOSYSTEM FOR OPEN DATA AT UNIVERSITY OF ALICANTE

As mentioned above, an IT-based solution is at the heart of the novel concept of "Open University". Specifically, opening data requires to define a dedicated IT ecosystem to collect, store and publish data and metadata, as well as giving facilities to reuse data. Furthermore, the design of this IT ecosystem should be aware about data releases along the time, so ecosystem should contain a layer in which data could be easily updated. Besides, it should be able to consume data from a heterogeneous set of data sources. Taking into account these features and considering other specific details of the University of Alicante, we have developed a platform for opening data of universities.

After studying the technological and data scenarios of the University of Alicante (UA) in detail, we decided to develop our own data management system (DMS) instead of deploying an open-source DMS like CKAN (http://ckan.org/), DKAN (https://www.drupal.org/project/dkan) or Socrata (https://www.socrata.com/). These DMSs provide solutions for creating open data portals and they were analyzed, but they do not meet our non-functional requirements as we explained in [14]: to sum up, an open data portal at UA must use a specific CMS used at UA for the sake of homogeneity and easiness of use for everyone (called Vualà) and also it must use an Oracle DBMS, since IT staff from UA are highly qualified for managing it.

There is a centralized database in UA (i.e. a data warehouse solution) containing consolidated data that is very valuable for opening and increasing the social impact (through transparency and accountability), for example financial data or several PKIs regarding teaching. However, from our own experience, we strongly believe that there are other sources containing data that are more useful for reusers to generate economic impact through innovative business models around novel data-intensive services or products, e.g. data coming from teaching schedule, data from services in the university (price of a cafeteria or cultural events and activities), etc. These data are distributed through data silos across different departments and units of the university.
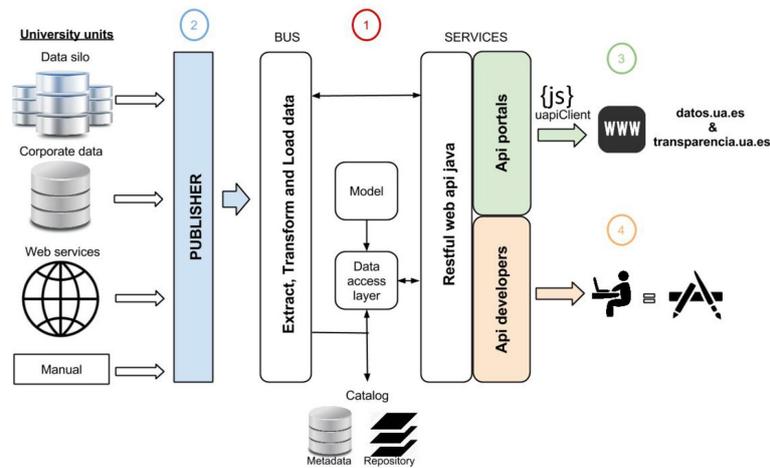
Figure 6: Architecture for opening data at University of Alicante.

In order to take into account every requirement, we have developed an IT-solution based on four parts as shown in Fig. 6.

1. The "Open Data Engine" is the core of the DMS. It has been developed in Java and served by Tomcat. The parts of this core are: (i) the catalog, formed by the repository where data files are stored and the database where are stored the metadata of the open data resources including the link to repository; (ii) the data model; (iii) the data access layer to communicate the (iv) REST-FUL Web API with database using the model; (v) two APIs (one of them to provide data to open data portal and transparency portal and the other to provide data to developers); and our (vi) logical ETL-based bus to extract, load and transform data that stores small processes to consume and transform data from different data sources as required. It is worth noting that our approach has a unique data catalog that feeds both open data and transparency portal in order to encourage that data that must be available in the transparency portal (due to the existing laws and the regulations established in the UA) could be effectively provided in reusable and structured formats. Thus, all data from UA can be easily reused and it will increase not only the accountability but also the business opportunities.
2. The publisher takes care to submit the new data in our system, and these data can be accessed from both portals and from the developer's API if it is structured data. This publisher is designed to allow the people with permits to submit data to open data portals, establishing a quality control by different experts (mentioned in section 2) before the data are published.
3. The ODP and TP, that feed their information using an API client, developed with Javascript, and the web services of Open Data Engine explained in the first point.
4. The API for developers, explained in https://dev.datos.ua.es/ is designed to allow developers to consume data and metadata providing the results with JSON to facilitate the reusability.

The result is a DMS that fits perfectly in a university scenario due to the definition of decoupled layers. DMS provides a set of facilities to deploy an open data core engine in the context of a university by considering in any technological scenario and taking into account the classification of data at the universities.

4 CONCLUSIONS AND FUTURE WORK

After more than 3 years of work, the University of Alicante is at a turning point. Our tools are defined and are proven to be useful in our OpenData4U project. Now, our plan is to share our tools to the community for the sake of opening data of universities. The more universities open their data, the more interesting services and products will be developed based on data, so data will be the fuel that will move the open-data ecosystem cycle of universities.

The establishment of this open data ecosystem supports the open data government movement, offering the data accessibility from universities to the society, thus achieving social and economic impact.

Next steps go beyond the great initial effort, and it is required in future initiatives to empower the project and to make it more visible and useful. Within UA interesting initiatives have been already launched and collaborations with entrepreneurship movements has started (since open data projects can bring benefits to students as future entrepreneurs).

Finally, future work to carry out is the harmonization of catalogs between universities, since most data generated by universities have similar nature, but formats and structures may differ, so an effort must be done in order to publish these data.

REFERENCES

[1] https://www.whitehouse.gov/the_press_office/TransparencyandOpenGovernment
[2] McDermott, P., Building open government. *Government Information Quarterly*, **27**(4), pp. 401–413, 2010.
    http://dx.doi.org/10.1016/j.giq.2010.07.002
[3] Lathrop, D. & Ruma, L., Open government: collaboration, transparency, and participation in practice. " O'Reilly Media, Inc.", 2010.
[4] Ubaldi, B., Open Government Data, 2013.
[5] Shadbolt, N., O'Hara, K., Berners-Lee, T., Gibbins, N., Glaser, H. & Hall, W., Linked open government data: Lessons from data. gov. uk. *IEEE Intelligent Systems*, **27**(3), pp. 16–24, 2012.
    http://dx.doi.org/10.1109/MIS.2012.23
[6] Auer, S., Bizer, C., Kobilarov, G., Lehmann, J., Cyganiak, R. & Ives, Z., *Dbpedia: A Nucleus for a Web of Open Data*. Springer Berlin Heidelberg, pp. 722–735, 2007.
[7] http://red.gnoss.com/comunidad/OpenData/recurso/Open--Data-Decalogue/4e4352a8-e034-4750-a0b7-da1abf5d6968
[8] http://opendatahandbook.org/
[9] https://www.boe.es/diario_boe/txt.php?id=DOUE-L-2013-81251&lang=en
[10] https://www.boe.es/buscar/act.php?id=BOE-A-2013-12887
[11] http://www.crue.org/TIC/Documents/Haciaunauniversidadabierta.pdf
[12] http://www.crue.org/TIC/Paginas/default.aspx
[13] Heimstädt, M., Saunderson, F. & Heath, T., Conceptualizing Open Data Ecosystems, 2014.
[14] Carcel J.V, Fuster A., Garrigós I., Maciá F., Mazón J-N, Vaquer L. & Zubcoff J.J., Development of an open data portal for a university - experience from the university of Alicante. *DATA,* pp. 297–304, 2014.
    http://dx.doi.org/10.5220/0005146602970304