TransparencyScience. Return on research investment, where do the funds go?

Lidia Contreras, Polytechnic University of Valencia
Cristina I. Font, Polytechnic University of Valencia
Paulina Morillo, Polytechnic University of Valencia
Diego Vallejo, Polytechnic University of Valencia

Abstract
The web application *www.transparencyscience.es* has been created in order to provide reliable information about public investment in science, in order to allow citizens to exercise their rights: to be informed in a transparent way, to control their government's actions and to bring their ideas to guide the country's policies on public investment in science. To achieve these goals, *www.transparencyscience.es* collects and process data from several open sources of the Spanish government. It uses different kinds of content and visualizations to facilitate the understanding of the Spanish public investment in science. It encourages citizen participation in three ways: a voting system; commenting system for collecting citizens' opinion in natural language; and finally, a crowdfunding system for proposed actions/petitions/etc.

The purposes of this paper are both to explain why we have designed and created the web application www.transparencyscience.es, and to describe how it works. It also reveals some added value in comparison with other projects in Spain.

Keywords: social network, vote system, crowdfunding, public investment, open data, open government, citizen participation, transparency, data visualization, data mining.
Citation: Contreras, L., Font, C.I., Morillo, P., Vallejo, D. (2015). TransparencyScience: Return on research investment, where do the funds go?. In *iConference 2015 Proceedings*.

Contact: liconoc@upv.es, crifonju@ei.upv.es, paumoal@inf.upv.es, dievalhu@inf.upv.es

1 Introduction

A large number of countries worldwide have joined the open government initiative. This initiative promotes public policies that seek transparency in their administrations, promotes citizen participation, and encourages government accountability (Calderón et al., 2010). Spain has also joined this initiative, and to this end open access tools and services are offered.

The Spanish Government's open access tools and services are still underdeveloped: huge amounts of raw data are scattered; format files are not reusable; it is not always possible to extract data simultaneously from different databases; the databases are not easily searchable; search engines do not retrieve them easily; websites usually lack usability; and data are not provided with clear explanations for citizens (Hernández-Pérez et al., 2013). These underdeveloped tools and services also affect information and data on the Spanish public expenditure on science and research.

The web application www.transparencyscience.es aims not only to alleviate this scattered data deluge in order to provide citizens, researchers and public policy makers with information about Spanish public expenditure on research and science, but also to let citizens give their opinion on public research funding priorities. Unlike other[1] web applications which provide, in an automatic way, information about where Spanish citizens' taxes are spent, *www.transparencyscience.es* provides information on research investment and funding à la carte, and gives citizens the chance to express their opinion on where tax collection should be spent with regard to scientific research. This latter feature of *www.transparencyscience.es* considers the general public an actor in decision making in research investment. Researchers and public policy makers should undoubtedly consider the general public's opinion on researcher funding and expenditure, as it seems that both scientists and research funding decisions are "divorced from social responsibilities" (Jack Stilgoe in "Against Excellence")[2]

[1] http://www.dondevanmisimpuestos.es/
[2] http://www.theguardian.com/science/political-science/2014/dec/19/against-excellence?CMP=share_btn_link

In order to achieve its two-way informative purposes, *www.transparencyscience.es* uses different tools and steps: data requests, data collection and data processing, visualization tools, and web 2.0 services for citizen participation.

2 Providing data to citizens

Data requests

Unlike other web sites, *www.transparencyscience.es* receives information requests à la carte from citizens. In order to request information, the general public can use either a blog or send an email message to the Webmaster. Later on, the requested information is provided on the website, and former and new visitors can both give their opinion, or request a more polished information request.

Other unique and important feature of *www.transparencyscience.es* is that it provides the information requested à la carte in a reusable format, e.g. .csv. Also, data provided are explained in a clear way, so that everyone can understand the data provided.

Data collection and processing

The web application *www.transparencyscience.es* gathers a myriad of data from different government Spanish institutions. Later the project managers screen the collected data and publish them in both the website, and social media sites.

Regarding data sources, data are collected from open data sources provided by the Spanish government. Some examples of information sources are Official State Gazette - BOE, National Institute of Statistics - INE, Spanish Foundation for Science and Technology – FECYT, Open Science Harvester - RECOLECTA, Ministry of Budget and Expenditure - SEPG, Ministry of Economy and Spain Open data Portal.

As the information requests are very specific, automatic data collection is not possible. Therefore manual data collection is required. The data search can follow different steps and sometimes need different strategies.

With regard to data processing, most data related to public investment expenditure are in PDF format, so data scraping is necessary for extraction and exploitation of the data. Once data are extracted, they are provided in manageable formats such as .xlsx or .csv.

A paramount activity in this step is improving data legibility. We erase superfluous data details, so that the information provided is clear and lacks of redundant and unnecessary details.

Data visualization

The web application uses different tools in order to display collected, filtered and processed data: a blog, RSS, and a Twitter service.

BLOG. Posts on data visualizations. We use a Google Charts plugin for Wordpress in order to generate the interactive graphics, which show the data requested by a web user. Below there is an example of what is provided in the blog.

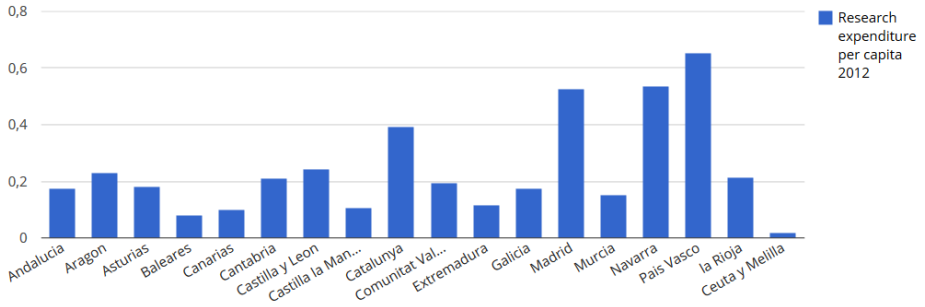

Figure 1: Expenditure in thousands of euros. 2012 by community / population.

All content published in the blog are automatically broadcasted on Twitter[3] and Facebook[4]. Citizens can stay informed about changes and proposals generated by the application through several applications (website and social media). So, citizens are free to choose the tool to visualize the requested information.

RSS. Last Scientific Works Visualization System. - On the website an RSS system shows users the 10 latest published scientific papers in different public repositories. This RSS system is fed from the website Recolecta (Fecyt), which is connected to about 80 repositories of scientific publications. This service is totally automatized and does not require any manual work by the webmasters.

#research Tweets Visualization System.- It also displays tweets that have the hashtag "research". So, Twitter news is available to both Twitter users and non-users.

3 Citizen participation

The application promotes citizen participation through two ways: a voting system and a comment system.

Vote System. - The voting system is similar to the system likes of Facebook. Initial research topics are proposed in Transparency Science. These topics are hierarchically organized according to the UNESCO science classification. This way we have research fields, disciplines and subdisciplines grouped into bubbles. The more votes that a field or discipline has, the bigger the bubble gets.

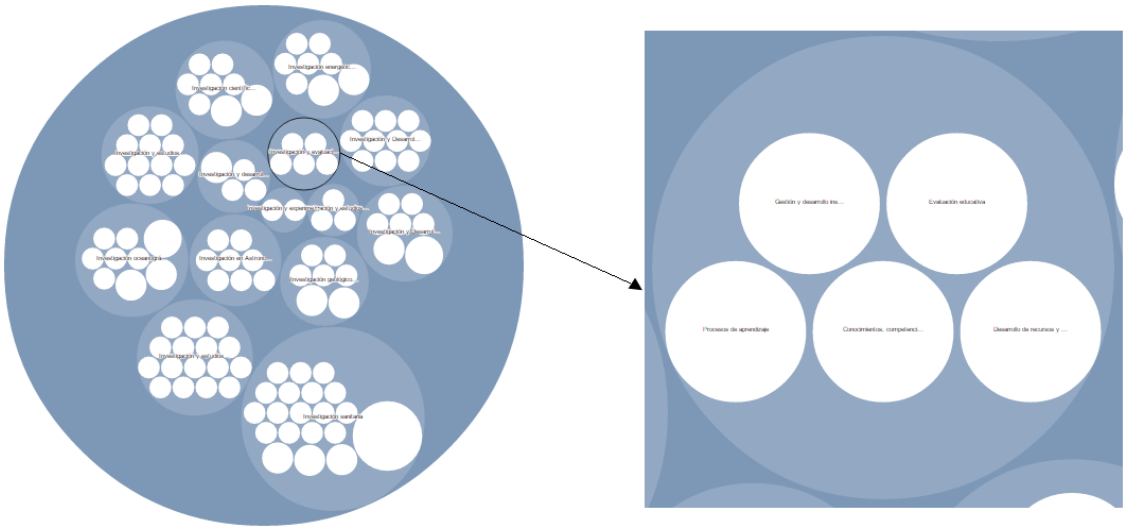

Figure 2: Topics of interest and Participation tool.

Comments System.- The application has also a blog where citizens can participate actively by leaving their comments, which can be read and commented by others.

4 Future considerations and improvements

We will include both a form to send suggestions for improvements regarding the web application, and to send data requests on public investment in research.

[3] https://twitter.com/trans_science
[4] https://www.facebook.com/transparencyscience

Votes on fields and disciplines will be collected in a database, which will export the preferences for research funding in a .csv file. This way, the vote system can be reused for any other purposes by public policy makers, researchers or other interested person in knowing the general public's willingness on research.

We will automatize the most common and simple data requests, as this is done manually right now.

We will promote the site in different ways (informative sessions, online marketing campaigns, etc.) so that policy makers, research centers, universities, government staff know the citizens' opinion on where the public funds should go. We will also use social media such as LinkedIn to promote the relationship between research and real social needs.

5 Conclusions

The *www.transparencyscience.es* web application will empower the general public to learn about and participate in discussions about Spanish public science expenditures. It will serve as the first source of information for citizen feedback and participation in their governments' investment in science, creating an appropriate link between the scientific sector and the public.

As the project continues to develop it should help build relationships between society and the different entities engaged in research and open science. Traditional agents in science management, such as universities, are complemented by citizen initiatives like the Open Knowledge Foundation and the OpenData platforms. This project brings voice to Spanish citizens, and the validation of this methodology will allow other countries to implement it.

References

Bird, Steven; Klein, Ewan; Loper, Edward; "Natural Language Processing with Python", 2009.

Calderón, César; Sebastián, Lorenzo; "Open Government: Gobierno abierto", 2010. Jaén, Spain. Pages: 11-19. http://libros.metabiblioteca.org/handle/001/163

Chandrajit Bajaj; "Data visualization techniques", 1999.

Ferrer-Sapena, Antonia; María Dácil Marín Garcia; Fernanda Peset Macebo; Francisco Manuel Solís Cabrera; Enrique Teruel Doñate y Diego Álvarez Sánchez (2014) . ¿Y las Universidades qué? Universidad Abierta: algo más que Gobierno Abierto y Open Data. Actualidad administrativa, 7-8, pp. 841-849

Ferrer-Sapena, Antonia; Peset, Fernanda; Aleixandre-Benavent, Rafael. Acceso a los datos públicos y su reutilización: open data y open government. El profesional de la información. May-June 2011, vol. 20, nº 3, pp. 260-269. http://www.elprofesionaldelainformacion.com/contenidos/2011/mayo/03.html

Hernández-Pérez, Tony; García-Moreno, María-Antonia; "Datos abiertos y repositorios de datos: Nuevo reto para los bibliotecarios", 2013. Madrid, Spain. Pages: 259-263.

López-Borrull, Alexandre and Canals, Agustí La colaboración científica en el marco de nuevas propuestas científicas: Open Science, e-Science y Big Data., 2013. In La colaboración científica: una aproximación multidisciplinar, Valencia (Spain), 21-23 November 2013.

Peset, Fernanda; Fernández-López, Antonio-Lázaro (2013) Carencias informativas de los datos abiertos en España. Anuario ThinkEPI, 2014, v. 8, pp. 318-321 http://www.thinkepi.net/carencias-informativas-de-los-datos-abiertos-en-espana

Table of Figures

Figure 1: Expenditure in thousands of euros. 2012 by community / population…… 2
Figure 2: Topics of interest and Participation tool…….. 3