

## **The emergence of scanR**

interview to Emmanuel Weisenburger (Ministère de l'Enseignement supérieur, de la Recherche et de l'Innovation) et Loïc Petit (Data Publica) conducted during the kick off meeting of the Re-search Alps project ([researchalps.eu](http://researchalps.eu)).

6<sup>th</sup> June 2017

### **What? When? Who?**

scanR (<https://scanr.enseignementsup-recherche.gouv.fr/>) is a kind of disruptive concept. The goal of the product is to characterize research and innovation activities of French research labs and French based private companies through their patenting activity, their scientific publications, their participations in various research consortium financed by public authorities, etc. The main idea behind this project is to combine structured data and web-crawling in a large-scale industrial fashion. A lot of technology is needed to integrate such a diversity of information. Some of the skills required for scanR are mastered internally in the Ministry, others are not. It was necessary to associate with a partner with complementary skills.

The idea of the project emerged in 2014 or 2015 in a specific team of the ministry of Higher education, research and innovation (<http://www.enseignementsup-recherche.gouv.fr/>) which is highly innovative. The team inside the Ministry is a team of inspired people, not really a research team: some of them are statisticians, some of them have competences in informatics, others are pure IT developers or economist of research and innovation. The team was originally made by three people (ten years ago), now they are ten plus two. Collectively, they decided to make things different from the past.

They team inside the Ministry the permission to launch this project as an experimental process with no perspectives, no certainty about the future existence of the products. However, they had a small amount of money (less than €100.000, for two years) and the chance to work.

In addition to this project, the team had many other innovative projects in open data and data visualization.

### **Implementation of scanR and characteristics of the team**

A public tender occurred: there were different proposals. Eventually, they considered that C-Radar (Data Publica) was the best choice (<https://www.c-radar.com/>). But, notice that a crucial component of the project was that also the Ministry staff worked a lot for the development of the project, in close connection with Data Publica.

With regard to the competences involved in the project, first of all, an integration between the Ministry's team and Data Publica occurred. Four people of the Ministry worked close to Data Publica, including one who spent 100% of his time on the project. Other people of the Ministry worked on data management and data sets (patents, H2020 participations, ...) From Data Publica, one person was in charge for the project management, while other two persons were main developers. But, multiple other contributions were made by other people from Data Publica.

### **The core of the activity**

Core of the action has happened just in four months. Basically, the product started in Jan 2016, and it was developed (and finished) at the end of Apr 2016. Then, there was some minor adjustments, refinements and the preparation of datasets. And administrative stuff, too. On July 6, 2016, there was a press statement on the product.

Ministry's team has supported the overall projects in terms of data quality. They were managing data cleaning. Data Publica took care of the technical engine of the process.

They currently have funds for a two-year maintenance of the platform. After that, for all the operations made about it (to avoid any possible shut down), more money will be needed.

### **scanR: impact**

The project has delivered several major goals.

Which were the Ministry's expected outcome at the beginning? There was a large need for a change, but to which extent, the collaboration with Data Publica has produced a change?

Nothing more but nothing less than written in the Tender call.

Data Publica, answering the Tender, was aligned with Emmanuel's vision.

However, the output of the activity has changed expectations further. Now, the Ministry is able to answer to questions such as who is working on what with who? This project has produced changes in other parts of the Ministry, in terms of better awareness of data at the higher level of the hierarchy. The demonstration of the high-performance scanR led the Ministry to generate a sustainable budget for the next four years. The fact remains that scanR has fewer resources than most IT projects of the Ministry of comparable ambition.

Future changes to be implemented will not directly affect the Ministry, but - with the implementation of scanR - additional research bodies (CNRS, universities...) have now major acknowledge about the existence of data and this provides new improvements/implementations, such as, a lot of contacts with these networks.



This is a great side effect. Other institutions may be interested in giving new information to improve their SEO (Search Engine Optimization) and use it.

With regard to the involvement of the team behind scanR in the Re-Search Alps project (<http://researchalps.eu/>), it decided to enter the partnership in order to share its experience in an international framework, to participate in the emergence of disruptive tools at European level and to benefit from all the advances (data, developments, methods) of the project for its own one.

*by Margherita Russo and Francesco Pagliacci (Unimore, IT), revised by Emmanuel Weisenburger*