RDA EUROPE AND INDUSTRY

SUMMARY REPORT

The overall majority of RDA members (both individual and organizational) come from the academic field and public research institutions. Representation of the private sector, though slowly growing, is still quite low.

Industry engagement is an important priority: sharing data between researchers and innovators with the aim to tackle the big challenges of society is in the core of RDA's vision. RDA, as a community-based initiative, aims to become a common ground in which academia and industry can collaborate.

THE INDUSTRY ADVISORY BOARD

RDA-EU3 is a Coordination and Support Action (CSA) to promote the RDA mission in Europe and support a stronger participation of Europeans in RDA. For the long-term sustainability of RDA in Europe and worldwide, industrial uptake of RDA output is essential.

In this context, the Industry Advisory Board is an entity of RDA-EU3, which has collected a number of advisors from the industry sector to get their input to RDA through the RDA-EU support action. Fabrizio Gagliardi (BSC) chaired the Board. The members have been chosen to represent a mix of industrial IT solutions providers, major data consumers and SMEs, which derive their business models from Big Data. The members contributed by critically reviewing the outputs that RDA has produced, by performing a gap analysis of the services and the activities that RDA is carrying out, which should be relevant for industry and made suggestions for the long term sustainability of RDA.

THE INTERVIEWS

Thanks to the contacts established with several RDA Working Groups and Interest Groups (Use Cases Group, BioSharing Registry WG, Scholix WG, Weather, Climate and Air Quality IG, BioDiversity Data Integration IG, Publishing

Data Workflows WG, Data Citation WG, Data Fabric IG, Big Data IG), we were able to contact a number of Industry representatives. Some of them became members of the RDA Europe Industry Advisory Board.

We have conducted seven interviews to private sector representatives, who offered their valuable insights regarding the data landscape in Europe. We interviewed representatives of IBM, Predictia, Terradue, EDP Renováveis, Engineering SPA, Rasdaman and Springer Nature in this context.

Moreover, considering that RDA Europe will continue to organize activities addressed to the private sector in the region with the future RDA-EU4 CSA, our aim is to pave the road to offer interesting and useful activities for data-intensive companies. For that reason, with RDA Europe's future activities in mind, we tried to get an overview from the private sector's perspective, in order to understand the needs and expectations of the Industry and the way in which RDA could address their demands.

The interviews focused in four areas: the role of data in the companies' activity, the recurring problems that the companies encounter in this scenario, the initiatives that aim to solve these recurring problems, and the opportunities for academia and industry to collaborate in their mutual benefit.

DATA AND BUSINESS MODELS

With the interviews we conducted, we tried to cover three different types of companies, regarding their relationship with data:

- Companies that own the data they produce themselves;
- Companies that offer technologies and services to work on external data;
- Companies that offer consultancy services and expertise.

Most of the companies we interviewed are working with external data sources, both from their own clients' repositories and / or from public repositories. For some of the interviewees, open data (especially scientific data) is a very important resource: on the one hand, it can be used to do experiments and tests and, on the other, it can even be considered as a "business driver". Some companies derive their own business models from open data and, while doing so, foster collaboration between different stakeholders, beyond their specific business.

> THERE ARE HIGH EXPECTATIONS RELATED TO THE DEVELOPMENT OF BUSINESS MODELS THAT COULD BUILD UPON OPEN RESEARCH DATA. KEEPING DATA CLOSED IS A DRAWBACK.
>
> JOAQUÍN BEDIA, PREDICTIA

Our interviewees stressed the fact that "it is of utmost importance to continuously be aware of (and extend) the state of the art". Therefore, participation in collaborative initiatives with academia is key.

> WE BELIEVE THAT THE PRODUCTS FROM RESEARCH DATA SHOULD BE SYSTEMATICALLY INCLUDED AS PART OF A 'VALUE-CREATION WHEEL' THAT THE DIGITAL ECONOMY NOW ALLOWS FOR.
>
> HERVÈ CAUMONT, TERRADUE

Even though most of the interviewees make use of public data sources under open data licenses, one of them pointed to the fact that usage restrictions are usually applied to this kind of data in some cases (open access for non-commercial use only, for instance). He added, "These are interesting cases because it means very often that the data provider is also 'open' to discuss terms and conditions for a commercial use".

RECURRING PROBLEMS & BOTTLENECKS

A recurring bottleneck is the skills gap, and finding suitable skilled staff tends to be a problem for the companies that participated in this survey. Moreover, it was stressed that leadership teams and other workers (such as editors in the Publishing industry) should be aware of data-related terminology, methods and tools.

> I GUESS IT IS A CULTURAL CHANGE TO CONSIDER DATA AS BEING INTEGRAL TO WHAT WE ARE DOING IN SCIENCE PUBLICATIONS.
>
> VARSHA KHODIYAR, SPRINGER NATURE

Data traceability is a very relevant requirement for the data-services community, especially for those who work with several data sources, both public and private. Traceability is linked to data authenticity and security, which are identified as main concerns for some companies' customers. Quality

assurance and quality control should be easier to use and less expensive, according to some interviewees and "efforts are still needed on the technology providers side to help in this goal".

Compliance with privacy rules (GDPR was specifically mentioned by some of our interviewees) is considered as a potential problem.

The usage of different types of data sources is usually linked to an important issue: standards. The adjustment to specific standards is sometimes identified as a customers' demand and, therefore, being aware of the market's demand for specific standards is of utmost importance for the private sector.

Whereas the interviewees recognize that, there is usually sufficient documentation and metadata available to enable meaningful sharing and reuse "a plethora of local conventions impede exploitation". Some interviewees point to the fact that some regions tend to offer more open datasets than others do and, therefore, the standards implemented differ. In addition, since standards are continually evolving, it is necessary to consider that "technology and the software implementations sometimes follow these evolutions with disparate timeframes". For these reasons, the interviewees claim that standards should be ideally global.

SOLVING PROBLEMS THROUGH COLLABORATION - WHAT COULD RDA DO?

Active participation in relevant standardization bodies is an approach followed by some companies in order to foster the application of standards and tackle this problem. Besides implementing standards, some interviewees have collaborated as editors of such open standards.

> WE BELIEVE THAT IN THE NEAR FUTURE STARTUPS WILL BLOOM THAT CONCENTRATE ON THEIR CORE BUSINESS, BET ON OPEN STANDARDS, AND DO NOT TRY TO REINVENT SILO SOLUTIONS.
>
> PETER BAUMAN, RASDAMAN

Moreover, some of them have taken active roles in community-based organizations such as RDA, in order to promote data sharing.

Regarding RDA output, some of the companies that participated in this survey have implemented RDA recommendations and output. It was pointed out that general principles, guidelines and proofs of concept have been much easier

to implement than technical outputs. In this line, it was stressed that, for outputs to be easily implemented, they must be maintained for the longer term, and funded accordingly.

Another positive outcome of collaborating with others is the fact that, "people working together better understand limits and constraints when reusing data". These experiences can, therefore, encourage standardization efforts and avoid "reinventing the wheel" through in-house developments.

The interviewees point to RDA as a forum in which computer science and data experts can work together with the public and private sectors and foster collaboration. For that reason, they encourage RDA to advertise its output and organize activities in which academics and Industry representatives can get together and find opportunities to collaborate in their mutual benefit.

They also suggested that some pilot test cases on the use in their companies of the RDA outputs could have been very useful if RDA-EU had identified the necessary resources to support this exercise. Maybe to be considered in the next RDA-EU4 CSA.