

anonymized for research purposes [3]? And what about the re-purposing of data for ends very far away from the original research agenda – for military or even criminal purposes? There are no easy answers to these questions, and the culture of ethics surrounding good research is making some communities tread warily.

Our survey highlights a lack of knowledge about the legal aspects of data sharing and data reuse, in particular around intellectual property rights, copyright and licensing, which can act as a barrier not only for opening data but also for re-using someone else's data. Choosing the right licence, for instance, can be a daunting task for some researchers who don't necessarily understand the implications of their actions.

While researchers are naturally keen to see their research published as widely as possible, in an interesting contrast to the open access scholarly paper movement, open data is viewed differently. Often research groups invest significant time and effort in collecting "hard to get data" which can then be used to build careers, offering what can only be termed a competitive advantage over those who do not have access to the same data. This

issue of credit and consequent career progression is a real concern in many communities.

The way forward

While aware of, and supportive of, the open access data agenda, many research communities are looking for guidance about the practicalities of doing it; training on managing the legal issues, for instance. They also feel that these issues should be addressed at cross-disciplinary level, perhaps rendering the tasks even more challenging. And while much of the open access focus is on coordination efforts, training needs and policies, researchers also stress the importance of developing the right tools and services to enable these policies and, ultimately, the sharing and reuse of data; this is seen as particularly crucial for handling sensitive data.

Some final words

Compared to scholarly publications, open access to research data is both less developed and more difficult to implement. Although open access to research data has only just begun, the broad spectra of expectations on EUDAT and other initiatives show that research communities have the notion that open access to research data cannot be solved through isolated activities or actions;

instead it needs to underpin the whole system, reaching from strategic planning and overall policies to the mindset and everyday practice of the individual researcher.

Link:

EUDAT – European Data project:
<http://www.eudat.eu/>

References:

- [1] European Commission: "Guidelines on Open Access to Scientific Publications and Research Data in Horizon 2020".
- [2] G8 Science Ministers Statement, 13 June 2013, available at <https://www.gov.uk/government/news/g8-science-ministers-statement>
- [3] Article 29 Data Protection Working Party: "Opinion 05/2014 on Anonymisation Techniques", adopted on 10 April 2014.

Please contact:

Marie Sandberg
CSC, Finland
Tel: +358 9 457 2001
E-mail: marie.sandberg@csc.fi

Rob Baxter
University of Edinburgh, UK
Tel: +44 131 651 3579
E-mail: r.baxter@epcc.ed.ac.uk

Providing Research Infrastructures with Data Publishing

by Massimiliano Assante, Leonardo Candela, Paolo Manghi, Pasquale Pagano, and Donatella Castelli

The purpose of data publishing is to release research data for others to use. However, its implementation remains an open issue. 'Science 2.0 Repositories' (SciRepos) address the publishing requirements arising in Science 2.0 by blurring the distinction between research life-cycle and research publishing. SciRepos interface with the ICT services of research infrastructures to intercept and publish research products while providing researchers with social networking tools for discovery, notification, sharing, discussion, and assessment of research products.

Data publishing approaches, namely the "data as a paper" ones [1], are mainly inspired by scientific literature communication workflows, which separate the place where research is conducted, i.e., Research Infrastructures (RIs), from the place where research is published and shared. In particular, research products are published "elsewhere" and "on date", i.e. when the scientists feel the products obtained so far are sufficiently mature. In our opinion, this model does

not fit well with other kinds of research products, for which effective interpretation, evaluation, and reuse can be ensured only if publishing includes the properties of "within" the RIs and "during" the research activity.

To enable effective scientific communication workflows, research product creation and publishing should both occur "within" the RI (as opposed to "elsewhere") and "during" the research

activities (as opposed to "on date"). To facilitate this, research infrastructure ICT services should not only be devised to provide scientists with facilities for carrying out their research activities, but also to support marketplace like facilities, enabling RI scientists to publish products created by research activities and other scientists to discover and reuse them. In other words, RIs should not rely on third-party marketplace sources to publish

their products, but rather should integrate them into the RI.

Unfortunately, current repository platforms are not suitable to implement this vision, as they are designed not to integrate with existing RI ICT services but instead to support today's notion of the "elsewhere" and "on date" research marketplace. We propose an innovative class of repositories: Science 2.0 Repositories (SciRepos).

SciRepos are characterized by the following features:

- Integration with RI ICT services in order to intercept the generation of products within research activities and to publish such products, i.e. making them discoverable and accessible to other researchers;
- Provision of repository-like tools so that scientists can access and share research products generated during their research activities;
- Dependence on social networking practices in order to modernize (scientific) communication both intra-RI and inter-RI, e.g., posting rather than deposition, "like" and "open discussions" for quality assessment, sharing rather than dissemination.

The SciRepo supports scientists with two kinds of end-user functionalities:

- Repository-oriented facilities: offering typical repository facilities on the information graph such as search and browse allowing search by product typology, but also permitting navigation from research activities to products and related products. Ingestion facilities are provided, allowing scientists to manually or semi-automatically upload "external" products into the repository and associate them with a research activity, thus including them in the information graph. Examples are publications, but also alternative scientific products, such as web sites, blogs, slides, documentation, manuals, etc. Ingestion allows scientists to complete the action of publishing a research activity with all products that are connected to it but generated out of the boundaries of the RI. The way scientists or groups of scientists can interact with products (access and reuse them) is ruled by clear rights management functionalities. Rights are typically assigned when products are generated in the RI or

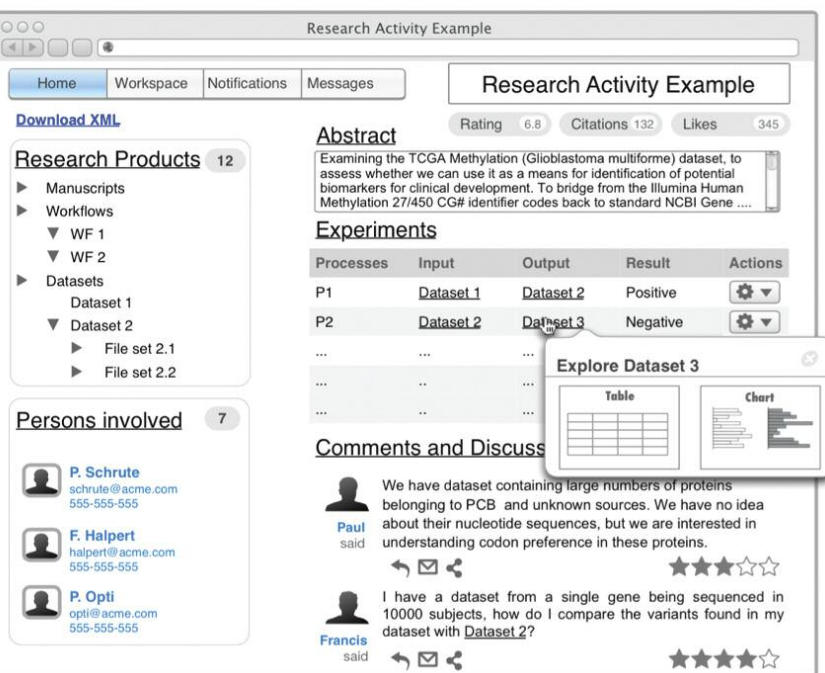


Figure 1: Repo: An example of a SciRepo Activity Web Page.

ingested by scientists, but can vary overtime.

- Collaboration-oriented facilities: offering typical social networking facilities such as the option to subscribe to events that are relevant to research activities and products, and be promptly notified, e.g., the completion of a workflow execution, the generation of datasets obeying to some criteria. Users can reply to posts and, most importantly, can express opinions on the quality of products, e.g., "like" actions or similar. This goes in the direction of truly "open" peer-review. More sophisticated assessment/peer-review functionalities (single/double blind) can be supported, in order to provide more traditional notions of quality. Interestingly, the posts themselves represent a special typology of products of the research activity and are searchable and browsable in the information graph.

In order to implement a SciRepo, RIs should develop their own software, thereby investing in a direction that requires different kinds of skills and dedicated funds. To facilitate this process we are designing and developing a SciRepo platform, devised to support the implementation of SciRepos at minimum development cost for the RIs. This platform builds upon previous works and experience [2][3].

References:

- [1] L. Candela et al. "Data Journals: A Survey", Journal of the Association for Information Science and Technology, 2014.
- [2] M. Assante et al.: "A Social Networking Research Environment for Scientific Data Sharing: The D4Science Offering", The Grey Journal, Vol. 10, 2014.
- [3] A. Bardi and P. Manghi: "A rationale for enhanced publications", LIBER Quarterly, 2014.

Please contact:

Massimiliano Assante
ISTI-CNR, Italy
E-mail:
massimiliano.assante@isti.cnr.it