

RDA P11 - IG SHARC session 22 March 2018, Berlin.

SESSION SYNTHESIS AND ACTION POINTS

Chairs: Anne Cambon-Thomsen, Laurence Mabile, Mohamed Yahia, Romain David (remotely) and Martina Zilioli.

Approx. 50 people attended

Agenda

I. Brief introduction of Sharc group	1
II. Background paper's content presentation	1
III. Discussion of first draft of recommendations	2
IV. Community-specific recommendations: biodiversity case; geospatial data	9
V. Next steps from RDA P11	9

Session objectives:

- 1) meet and discuss with various RDA scientific communities and networks interested in the SHARC objectives of promoting sharing activities as recognized research outputs;
- 2) get feedback and input from other communities to the ongoing background paper;
- 3) network with relevant stakeholders involved in research assessment and publication.

Slides are available on the [RDA P11 meeting page](#).

I. Brief introduction of Sharc group

Problem: lack of recognition of the sharing activity itself that may be complex and not well identified is part of the barriers to sharing

Goal of the Sharc group: 1/ Explore mechanisms and instruments that could be put in place to encourage the sharing of data and material samples in different research domains by recognizing, crediting and rewarding the various steps of the work necessary to accomplish a reliable and useful sharing; 2/ provide suitable recommendations to various relevant stakeholders

II. Background paper's content presentation

- The background document will be the basis for concrete policy recommendations for rewarding the sharing activity, to national, supra-national and community bodies involved in the process of research output assessments and to suggest pilot projects to implement them.

- Specific objectives are 1) to describe the steps and actors necessary in the process of rewarding the sharing of data and resources; 2) to review existing reward mechanisms; 3) to underline the gaps; 4) to recommend new ways and tools that can be generalized, based on case studies.

See working document at :

https://docs.google.com/document/d/14_HxIrrkB0128EQpmTqrwXtuJTy3zQ_TLFODuoWaqe4

Everyone is invited to comment on the googledoc.

III. Discussion of first draft of recommendations: generic ones mostly / community-specific ones

Initial steps:

- Identifying blocking points

Recommendation 1. To institutions and researchers

Obstacles to data and/or samples sharing differ across scientific communities and data types. It is important to identify all the obstacles that lead to non-sharing or weak sharing and to qualify them. The willingness to share and the recognition of sharing are two concepts that can be strongly linked. If blocking points have not yet been fully identified consensually in a community, we recommend to investigate them further as a first step before designing any sharing policies.

Action: For any relevant scientific community where it does not already exist, a survey should be undertaken to explore the reasons for not sharing data. It could take the form of multiple-choice questions.

Comments:

- *Be aware that different responses may be obtained according to how the questions are addressed (e.g. more honest responses expected as part of a researcher's survey vs institution's survey)*
- *Responses may vary quickly over time; a baseline survey is needed but some monitoring of the situation should be done regularly (e.g. every 5 y), see DataOne's survey; should be added to Reco1*
- *Report of interest: from Springer, PRACTICAL CHALLENGES FOR RESEARCHERS IN DATA SHARING*
- *General comment: this recommendation addresses a very broad and difficult issue; surveys not good ways; the recommendation should focus on one area.*

- Training and education issues

Recommendation 2.

To researchers and institution and consortia administrators.

A general requirement for the reward of data sharing activities is the existence of an aware and expanding community of producers and users. Data are a crucial component of information and decision making, therefore data sharing needs to be incentivised. This also means that users should be educated to share data, use data-sharing infrastructures, as well as properly manage data and

infrastructures for the common good. Therefore, knowledge institutions and knowledge workers providing continuous training and education are crucial for the development of mutual understanding on definitions, types and sharing practices in the data community. Organisations such as RDA contribute to this mutual understanding and to the creation of such a community.

Actions:

- Research institutions and consortia should provide regular training sessions to scientists and PhD students on practices such as self-archiving, standards for the identification, formatting and curation of data and metadata to make data reusable, publishing venue and related licensing, metrics and acknowledging and crediting the reuse. In the era of big data, it is also essential to address their use comprehensively, including their societal, ethical, philosophical and regulatory aspects.

(in agreement with the 2017 Report of the EU Working Group on Skills for Open Science for categories of Open Science skills and expertise).

- Researchers should ensure the training of members of their teams on those issues.
- Workshops should be specifically undertaken to precisely define such educational content and tools. Ways to release them to the relevant communities should be addressed (e.g. E-learning). *Role for RDA?*

Comments:

Anne-Sophie Archambeau proposes a use-case from GBIF where such training sessions are provided.

- **Policy / Legal aspects:**

Recommendation 3.

To various stakeholders.

The multiplicity of heterogeneous rules regarding data sharing hampers their implementation. We recommend to harmonise policies in use for all relevant stakeholders (researchers, funders, research institution administrators, publishers, governmental policy makers) in order to standardize practices and send a clear message to the various communities as regards the importance of data sharing.

➤ **Action recommended to the RDA governance:**

Organise consensus-building workshops with representatives of the various groups of stakeholders to identify points of convergence and divergence and strive to build consensus to design common rules (Delphi methodology). The results of such workshops should be communicated to international bodies that shape international regulation such as INGSAs...

➤ **Actions recommended to funders:**

Harmonise incentives provided by all funders and ensure feasibility, in particular:

- Provide specific funding to enable not only data reuse but also to finance costs related to implementing FAIR data
- Propose targeted grant awards towards enhancement of collected data ex: Funding of replicating studies
- Enforce data sharing as one mandatory condition for obtaining funds
- Require from researchers a data management plan specifying data sharing arrangements following end of project

- Monitor systematically research projects to ensure that DMPs have been implemented. Otherwise, restricting new financing to researchers who have not done it.
 - **Actions recommended to publishers-editors:**
- Require that data and Data Management Plans (DMPs) are open access to allow replication and checks
- Reject articles which datasets are not archived openly, unless a proper justification is provided by authors.
 - **Actions recommended to research Institutions:**
- Take data sharing activities into account the in the researcher's career assessment and evaluation.
- Require researchers design and implement a DMP specifying data sharing method after research project completion.
- Systematically monitor research projects to ensure that DMP have been implemented; Otherwise set up a sanctioning process.
- Require acknowledgement or citation of reuse in DTA/MTA or in some policies accompanying data usage.
 - **Formal acknowledgement** of the data providers and/or funding agencies in all disseminated work making use of the data
 - **Formal citation** of the data providers and/or funding agencies in all disseminated work making use of the data. Ex. CoBRA
 - **Co-authorship** on publications resulting from use of the data
 - The **opportunity to collaborate on the project** (including, for example, consultation on analytic methods, interpretation of results, dissemination of research results, etc.
 - Results based (at least in part) on the data could not be disseminated in any format without the **data provider's approval**.
 - The data provider is given a **complete list of all products that make use of the data**, including articles, presentations, educational materials, etc (in MTA/DTA)
- **Actions recommended to researchers:**
- Comply to the sharing policies
- Whenever possible, associate free open licences to datasets (such as Creative Commons) submitted to an appropriate repository.

Comments:

- *One important stakeholder is missing: data repositories (and materials infrastructures such as biobanks).*

- *Open access DMPs exist a little in US as it is part of the regulation of the government funding but permission is required from the proposer. See the DMP tool: 'it helps researchers answer various data management questions relating to their project, such as how data will be documented and made available for public or secondary uses, how data quality will be assured, backup procedures, and preservation plans. It also aids institutions in identifying costs associated with data management, and helps with forward resource planning.'*

<https://www.dataone.org/software-tools/dmp-tool>

- *In France DMPs are only recommended so far by the national funding agency (ANR) but not mandatory (MY).*

Preliminary steps required for rewarding

Recommendation 4.

To researchers.

For an effective inclusion of sharing activities in rewarding processes, the data to be shared must be visible on the scholarly digital information system. This implies that they must abide by the FAIR principles: specifically, they must be present on the web, traceable, reusable and the assessment of their use should be possible.

Pre-required actions:

- Datasets must be archived in open digital repositories that warrant long term preservation. Recommended data repositories are listed by FAIRsharing <fairsharing.org>
- Physical resources used in research (such as physical samples) must be embedded in an organised storage and identification infrastructure (such as a biobank for human samples)
- Datasets and resources should be archived under open licences (such as *Creative Commons* <www.creativecommons.org>)[1]
- Datasets -or material resource description- must be identified uniquely, persistently and connected to the scholarly digital sphere: attribution of a permanent identifier (PID) such as Digital Object Identifier, DOI from Datacite <www.datacite.org>
- Re-use of data and physical resources should be done following given instructions if they exist, or by using standards (such as the CoBRA guideline for citing bioresources in journal articles <www.equator-network.org/reporting-guidelines/cobra>)
- We recommend that researchers get ORCID so that DOIs of cited resources are attached to their ORCID profile.
- If available, contributions such as data/samples sharing should be specified in publications through the CRediT-CASRAI initiative < <http://docs.casrai.org/CRediT>>

Comments

- Regarding repositories, from Shelley Stall, AGU, sstall@agu.org:

Add re3data.org as an option as well as [FAIRsharing.org](http://fairsharing.org). [Re3data](http://re3data.org) is mentioned by EC however it is better to recommend TDR (Trusted Data Repositories) because [re3data](http://re3data.org) and [fairsharing](http://fairsharing.org) have limitations.

AGU's draft Statement of Commitment (open for review through end of March) - Researcher tenants:

Individual Researchers will:

- Enable their research outputs to be FAIR compliant and cite data sources as references in publications and proposals.
- Place their research outputs (data, software, sample information) in domain repositories that are FAIR compliant whenever possible.
- Follow leading practices for including provenance information on research output.
- Help educate their colleagues in practices that enable FAIR research outputs.
- Support development of FAIR standards and practices in their institutions and organizations, and in scholarly publishing as authors, reviewers, and editors.

To give feedback on the project Commitment Statement:

- [The link to the “Commitment to Enabling FAIR Data in the Earth and Space Sciences” cover sheet and document.](#) (view only permission).
- [The link to the comment and feedback form.](#) In this form, please add your name, email, document reference and comment/feedback.
 - The recommendation could be more general in some parts and doing so could cover Humanities.
 - There is some discussion about whether the recommendations should be addressed to researchers too as it could have negative effects. To be more pro-active: motivation instead of recommendation? Or address the recommendation to researchers as evaluators or reviewers; should be oriented on how to treat the colleagues...
- This recommendation needs to be implemented at the institutional level at the same time.

Recommendation 5.

To publishers/editors.

Supporting the pre-required actions (reported above):

By the mean of editorial policies addressed to authors and reviewers, editors can efficiently encourage or mandate the pre-required actions needed for rewarding processes.

Actions:

- Include in instructions to authors information that will help to choose a proper repository for archiving data, a unique and persistent identifier for datasets and data citation standards.
Follow the RDA Data Policy standardisation and implementation IG.

Recommendation 6.

To publishers/editors.

Better visibility of data. Although data journals have appeared in the last decade, their number is still limited. This is even more true for journals publishing descriptions of physical resources such as collections of samples.

Action: Create additional meta-journals to publish detailed descriptions of every kind of datasets and resources that will constitute a research data record.

Comments:

- Publishers ask more and more for data papers at the same time as research papers. It is becoming the current practice.
- Instead of creating more data journals, we should support the creation of more repositories.
- Yes but provided that data are described properly and homogenously.
- Data with paper is a much valuable process. It is up to journals and communities to establish consensus about archiving data in repositories and linking to the research paper.
- Still, papers do lack for materials description.

- In some countries such as China, they do not share much but data papers do stimulate the process. It depends on the culture of the country. The discussion is very European-centric, the problem is much larger than that.

Recommendation 7.

To funders and research institutions administrators

Fair measuring of impact.

Article-level metrics are not adapted to measuring the re-use of data and physical resources. Data-level metrics reflecting real usage are being developed and should be used preferentially.

Actions:

- Citation based metrics can be used for data published in data journals and alternative usage based-metrics (such as Altmetrics) can be used to measure attention and uptake of a dataset, while waiting for output from RDA the Fairmetrics and Make Data Count initiatives -*RDA Data Usage Metrics WG*.

Rewarding mechanisms for sharing activities

Evaluation related-criteria

Recommendation 8.

To funders and institution administrators.

There is an urgent need to acknowledge and reward the FAIR management and sharing of data and materials as a **first-class research output**. One powerful driver is to take sharing activities into account in the **evaluation** scheme.

Actions:

- Include FAIR Sharing activity as a set of evaluation criteria, making it a strong requirement in all proposals.
- To assess researchers' proposals or career, use a set of criteria drawn from the Open Science Career Evaluation Matrix (OS-CAM) designed by the EU WG on OS rewards: more criteria for sharing samples? Which ones?
- Avoid the exclusive use of quantitative assessment mechanisms such as impact factors in assessment processes.
- For quantitative assessment, use data-level metrics, and FAIR metrics whenever they are available.
- Use ORCID profiles in evaluation (they encompass many research activities, including samples/data sharing)
- Use Credit-CASRAI badges in the researcher's activity assessment
- add sharing criteria to key performance indicators in institutional evaluations.
- Dedicate significant and coordinated effort to the training of evaluators, selection committees and researchers themselves, so that they would take sharing practices and FAIR principles implementation into account in evaluations.

Comments:

- Opportunities will be missed if those issues are thought out of the RDA Sensitive Data WG; need to collaborate with them.
- The various levels of researchers (junior/ senior...) should be taken into account as well as the specificity of some evaluation criteria related to the culture of the discipline.

Allocating dedicated funding

Recommendation 9.

To funders and institution administrators.

As the implementation of the FAIR principles may induce extra-costs for the sharer, a way to encourage it concretely is to allocate specific funding to sharing initiatives.

Actions:

- Dedicate specific additional funding to the FAIR sharing activities that cannot be diverted towards any other expenses.
- Promulgate consequences for not complying such as being temporarily ineligible for further funding.

Comments:

- This sounds good but unrealistic; it won't work in practice.
- ACT: The whole picture must be seen.

Allocating support to facilitate implementation issues

Recommendation 10.

To institution administrators.

As the implementation of FAIR sharing may be in some communities very time-consuming and may require skills that researchers do not have, backing this activity by providing human and structural support is essential.

Actions:

- Provide sustainable human, financial and infrastructural support to be made available within research institutions.
- Undertake more large-scale collaborative efforts to make storage infrastructures and dedicated trained personnel available to all researchers regardless of their affiliation
- Organise help from trained personnel to design and implement data/sample sharing in their research endeavors.

General discussion:

- As for projects, the recommendations should be SMART (specific, measurable, attainable, relevant, time-bound). How much are they impacting? The most important actions that will have a measurable impact should be short listed. That will help prioritising.
- Promoting to cite data in a proper way should appear more clearly from the recommendation.
- There should be more criteria of choice for repositories, see Datacite/Clarivate instructions to choose repositories.

- Indexes of interest are emerging from publishers, ex: Data Citation Index.
- All elements are in the recommendations but should be sorted differently.
- A lot is said on what to do but not enough on how to do it.
- See GITHUB to track everyone's contribution, exercising control with reuses tracked, comparing open data principles to social conventions of open source, very practical.
- ACT: we need to be realistic with already existing solutions but at the same time we need to keep some space for imaginative new ways, to allow evolution and not be locked in the current system...
- + post-meeting comments:
 - It would be helpful to have on one scientific workflow scheme every actor involved and what they can change to improve the crediting processes.
 - What about service providers (ex: software designers), they are also willing to get some credit for their contribution in the sharing chain.
 - Fairmetrics will develop a framework for others to develop metrics; they will not design new metrics.

IV. Community-specific recommendations:

- **The biodiversity case:** See Romain David's presentation [herein](#).

Comments:

Florencia Grattarola is very interested by the recommendations towards researchers and by joining the group for the biodiversity specificities. She could provide a case-study for Uruguayan researchers that point to the difficulties they encounter to share data (no DMP, no tools, no relevant skills...) and is keen to contribute to elaborate recommendations she wishes to provide there.

- **Geospatial data case** from Martina Zilioli, not presented during the session but see slides [herein](#).

Comments are very welcome.

V. Next steps from RDA P11

Input and feedback can be done in many ways:

- as comments on the google doc or by mail to laurence.mabile@univ-tlse3.fr
https://docs.google.com/document/d/14_HxIrrkB0128EQpmTgrwXtuJTj3zQ_TLFODuoWage4
- by joining the group if you wish to follow up.
- by attending ESOF 'Sharing science towards new horizons', Toulouse FR, July 2018
<https://www.esof.eu/en/>

If enough people from the Sharc IG attend, a small meeting will be organised there.

Action points:

- Integrate as much as possible suggestions in the recommendations (LM)
- Prioritise actions and sort recommendations accordingly and per stakeholder (LM, ACT)
- Constitute a 'task force' that will elaborate a set of specific criteria for the sharing activity assessment
- Join Usage metrics WG and Sensitive data WG (LM)

submitted by L. Mabile,
10 Apr 2018

