

# “Open Data Surveys: The issues, emerging survey findings, user interests and future directions”

## **Agenda and meeting summary - RDA 12th Plenary Meeting**

### **Hosted by the RDA IG Surveying Research Data Sharing Surveys**

*Tuesday Nov 6, 2019 11:30-13h, Room: Tsodilo B3*

**Background:** The Interest Group (IG) co-chairs believe that there are advantages to promoting a coordinated approach to the development and analysis of surveys examining open data. We seek to encourage comparative analysis of existing surveys in terms of their design, scope and findings. Through such analyses, we will promote a discussion on opportunities to develop modular and interoperable open survey(s) that interested stakeholders could utilize to track changes in practice overtime and promote policy learning.

At present, there is a great deal of experimentation: open data policy statements by government agencies, research funders, research performing organizations, professional societies and publishers are becoming common place. While they are united behind the goal of making research data open, incentives, institutional settings and infrastructure capabilities are some of the variables that influence the pace of change.

As open data policies are implemented, and data sharing practices evolve, comprehensive benchmarking and tracking of open data practices can serve to illuminate advances in data sharing (where and by whom) and help to understand the reasons for different data sharing practices.

We are experiencing the first wave of survey reports and are starting to see the gaps between policy and research practices. Some data sharing surveys provide insight into how policy and practice might be bridged but the granularity of surveys, their comparability and uneven geographical coverage needs addressing.

### **Meeting objectives**

This was the first RDA meeting following the establishment of the IG. As such, the meeting promoted an agenda-setting discussion to promote awareness of, and feedback on, the proposed goals and identify potential collaborators. As this meeting was the first RDA meeting in Africa, we promoted a discussion between African and international colleagues on how the scope of the IG might support their interests.

Case statement: <https://www.rd-alliance.org/group/open-questionnaire-research-data-sharing-survey-ig/case-statement/open-questionnaire-research>

## Meeting agenda

1. Outline, scope, rationale: Ingeborg Meijer, CWTS, the Netherlands
2. Learning from Surveys – 2 case examples
  - Research Data: The Researcher’s Perspective. Federica Rosetta (Elsevier)
  - Practical challenges for researchers in data sharing. Varsha Khodiyar (Springer Nature)
3. User cases
  - Japan, Kazuhiro Hayashi, NISTEP (Japan)
  - Austria, Paolo Budroni, University of Vienna (Austria)
  - University Infrastructure & Open Data, Nodumo Dhlamini, African Association of Universities (Ghana)
  - A Qualitative Findings from Southern Africa, Louise Bezuidenhout, University of Oxford (UK / S. Africa)
4. Future direction/discussion: David O’Brien, IDRC (Canada)

## Discussion Summary

*Note: the presenters agreed to make their presentations available on the Interest Group website. The following notes provide a summary of their remarks and discussion.*

1. **Opening remarks:** Ingeborg Meijer (University of Leiden) set the context with a presentation on the rationale for the RDA Interest Group and a framework for how researchers and practitioners could contribute to and benefit from its stated objectives.

The three goals of IG Charter are to: convene user communities; develop community-designed modular and interoperable open survey(s); and determine how such open survey(s) could be implemented and results analyzed globally.

### 2. Learning from Publisher Data Sharing Surveys:

#### i. Open Data: The Researcher’s Perspective. Federica Rosetta (Elsevier)

- Elsevier’s global survey focused on attitudes, ideas of ownership, research data management and visibility of research as primary topics.
- Attitudes of researchers: the majority of researchers believe that having access to research data is important but the proportion of authors who actually share their data is much lower than this number
- Major gap is lack of training for data sharing and lack of incentives

- Major benefits identified include increased potential for collaboration and higher visibility
- Main disincentive are time and financial costs, misuse of data and legal concerns
- Survey was important because it identified pain points and discussed how Elsevier can contribute to shaping the policy landscape to improve data sharing
- Complementary bibliometric analysis by CWTS revealed that in Africa, including datasets with papers increased the citation rate of African researchers exponentially
- Survey findings used to inform funders on how to change their data policies, which is important given the gaps that exists around the goals set in the European Union's Horizon 2020 program
- Surveys offer an opportunity to look at data sharing on a more granular level (ex: country/institution/discipline) and to do a deep dive on barriers to data sharing
- Comparing Africa to the global results showed that overall trends were similar. Some notable result include that Africans feel a strong sense of ownership about their data but still show a stronger desire to share their data compared to the global average.

## **ii. Practical challenges for researchers in data sharing. Varsha Khodiyar (Springer Nature)**

- Discoverability emerged as the most important driver for sharing data across disciplines
- Also found the majority of researchers surveyed already shared their data in supplementary material and/or repositories
- Found that barriers to sharing include: copyright, cost, size of data, unaware of repositories
- Also found differences across age: senior researchers don't have time whereas junior researchers lack the knowledge
- The survey informed the development of a research data support service at SpringerNature to address obstacles for data sharing
  - o Helping authors edit their data and instructing them on proper metadata to include
  - o Also building a training service that goes to institutions to train researchers
- Survey was limited in terms of geographic representation (majority of answers from EU, US)
- Re-ran surveys in Japan and China with a much higher response rate

## **3. Cases: National and Regional Surveys**

### **i. Japan Survey and Challenges of Interoperability. Kazuhiro Hayashi (NISTEP, Japan)**

- NISTEP's Open Science survey provided baseline understanding of practices and perceptions of researchers to research data management. Survey had a high response rate (~ 70%), which is much higher compared to the larger international surveys carried out by publishers (< 5%)
- The survey findings could not be directly compared to other survey results. Their analysis showed that very few questions across many published surveys could be compared.
- There is a need to work toward greater comparability.
- NISTEP sought to promote comparability in their second survey and collaborated with SpringerNature. Translation proved to be challenge between English and Japanese as concepts and terminology are not readily translatable.
- Additional biases may exists depending on the fields of research of survey users and these need to be addressed as best possible in the initial survey design

- Government authorities have recognized the survey and are working to address identified challenges.

## **ii. Austrian Survey and Use. Paulo Padroni (University of Vienna)**

- E-Infrastructure Austria undertook a national survey of researchers at public universities
- The process also involved workshop and national meetings to discuss the findings
- The survey identified significant training needs to support the research community, including training on data management and assistance interpreting policy/legal documents
- The survey had an impact on universities and the government agencies. It led to new policies and support programs within universities and changes in funding guidelines.
- At the national level, the survey focused attention on the need for an integrated e-infrastructure.

## **iii) Data infrastructure in Africa. Nodumo Dhlamini (Association of African Universities, Ghana)**

- Nodumo Dhlamini was unable to attend but her slides were presented.
- The AAU sponsored a survey of university infrastructure to support open research data.
- Survey focused on researchers' awareness of data repositories and whether their universities provided or were planning to provide support for sharing data
- The survey generated considerable interest from universities who are keen to understand how to adjust their infrastructure to support open data
- Their survey was directed toward supporting academic administrators build the requisite support and infrastructure.

## **iv) Qualitative findings from interviews with Southern African researchers - Louise Bezuidenhout**

- Presentation based on qualitative research findings of a study examining barriers to data sharing specific to the African continent
- Some of the major cultural and physical barriers identified included: high teaching loads and researchers having to pay to acquire and share their data
- The study identified different levels of awareness and understanding of 'open data', a challenge that is more readily addressed in qualitative research. Those planning survey research will likely encounter misunderstanding of key concepts if efforts are not made to clarify concepts.
- Use of findings: Further work in this area needs to address the questions how the survey data will be used? Is the purpose to characterize the state of open data or change the context. There is an ethical issue surrounding survey use: can designers and sponsors of surveys provide support to communities after challenges have been identified?

## **4. Discussion & next steps**

To start the discussion, the chair asked the following questions of participants:

- What is the state of knowledge on research data sharing among the constituencies you represent?
- Are you attending the meeting because you see an opportunity / interest to utilize surveys to identify perceptions, awareness, organizational capabilities, etc?
- What suggestions or directions would you like to see the IG pursue? What might you / your organizations contribute to the IG to realize those interests?

## Comments for clarification

- Some participants asked for clarity on terminology. For example, what did we mean by modularity? The organizers clarified that the term is used to describe a block of questions with a thematic focus.
- What does the term 'sharing' imply? Is it data open to all in a repository or by request? The organizers agreed that greater specificity is needed. Some surveys provide numerous data sharing categories with different degrees of access whereas in others, sharing could include very different kinds of access: e.g., willingness to send a dataset on request and data archived in an open access repository.
- Presenters used the term granularity to suggest a difference between global and national surveys. Would this be the only distinction? If disciplinary research cultures were influential in shaping open data trends, surveys should provide enough granular detail to tease out these differences. The organizers agreed and noted that the sponsors of open data / open science surveys have different objectives and they will likely ask questions that help them address their questions.

## Comments on future directions

### Compile resources and facilitate comparison

- There was encouragement for the IG to compile a list of all surveys, questionnaires and datasets currently available. While many are available online, having them all in one place would be a useful resource.
- A RDA member from the San Diego Super Computing Centre said his organization built SUAVE, which could help visually compare survey data. The SUAVE program currently houses the Belmont Forum open data survey. ([http://suave-dev.sdsc.edu/main/file=zaslavsk\\_Belmont\\_Forum\\_Open\\_Data.csv&views=111010&view=bucket](http://suave-dev.sdsc.edu/main/file=zaslavsk_Belmont_Forum_Open_Data.csv&views=111010&view=bucket))

### Promote standards

- Participants encouraged the IG to promote the interoperability of survey instruments through common questions and consistent language. Was this focus also going to be extended to include the data and supplementary material (e.g. codebooks to help understand the data and survey)?
- On a similar note, it was also suggested that the IG might propose key performance indicators that survey designers may adopt.

### Support other RDA groups

- A RDA co-chair from another IG mentioned that RDA had just approved their IG on research data architecture. They would be keen to receive feedback on their questionnaire and provide access to their findings.
- There was a similar request from another IG member who indicated they had a pilot survey they would appreciate feedback on. Following this, a representative for RDA noted the opportunity to collaborate between RDA groups that rely on the use of survey instruments.

### Provide guidance

- A number of the presentations identified how their survey results were disseminated and what impact they have had to date. A participant asked whether the IG intended to track the policy or programmatic impacts surveys have had as this information could be useful for agencies who sponsor open data surveys as well as those who design them. For example, what have universities, government agencies, academic societies, etc. done to reduce gaps between policy intent and actual practices? Tracking policy options / solutions might be an interesting contribution.