



# RDA/WDS Publishing Data breakouts

## Research Data Alliance 2nd Plenary

**Tuesday, 17 September 2013**  
Salon B, Marriott Hotel, Washington DC

### **14:00-14:30 RDA/WDS Publishing Data Interest Group**

*Co-chairs: Michael Diepenbroek, Jonathan Tedds*

The Publishing [Data Interest Group](#) brings together all stakeholders involved in publishing research data including researchers, discipline specific and institutional data repositories, academic publishers, funders and service providers. Every effort will be made to get a good representation from related international programmes, their working groups and other private or institutional activities involved in this area. We will build on existing resources, reports and other shared experiences from the different stakeholders and will nurture more specific and targeted working groups addressing practical aspects in publishing research data.

As such, the Publishing Data Interest Group can be regarded as a broad and inclusive forum for interested individuals to contribute to and test, validate and promote the findings of the Working Groups. In particular we plan to address the implementation of workflows for publishing data and therefore help establish appropriate supporting infrastructure including bibliometrics, costings and other relevant services.

This session will update participants on the work to date through the joint RDA-WDS Working Groups proposed and invite further expressions of interest in these or other areas of common interest.

### **14:30-15:00 RDA/WDS Publishing Data Working Group Costs and Cost recovery**

*Co-Chairs: Ingrid Dillo*

This proposed RDA-WDS Joint Working Group has its focus on cost recovery models. How do you fund a long-term infrastructure when research funders are unwilling to take on recurring cost? This means we need cost recovery models for digital repositories: what has been done, and what has been shown to work? We also need to look at data publication and the relationship between publishers (traditional/non-traditional), data centres, libraries, and computer centres.

This approach of course requires a good understanding of the current work on cost models, but the focus of this group will be on understanding the various cost recovery models / business models that are available. We need to develop an understanding of the various business models that help provide this infrastructure. This means we need to answer questions like: What are the costs? What models exist for long-term sustainability? Under what conditions are different business models most appropriate/necessary?

The activities of the group will focus on the following components:

- Surveying current cost models (building on work of APARSEN and 4C);
- Filling in current knowledge about actual costs (taking a range of WDS and DSA members);
- Survey policies and implications for funding (what are the key policy developments and their implications);
- Researching and analysing various approaches to cost recovery/business models (survey data centres and data publication initiatives; how are costs recovered (phenomenon of DBs charging for access; and how are roles, responsibilities and costs divided within the data infrastructure, e.g. front office/back office model));
- Presenting conclusions about the potential appropriateness of different cost recovery models to different situations (what specific aspects of the data, the institution, or a combination of both work for which recovery models?) and about the potential of data publication initiatives fitting into a cost recovery strategy.

In this session the draft case statement of the group will be presented and the participants will be asked to give their comments and to provide clear examples of use cases and relevant cost recovery models.

### **15:00-15:30 RDA/WDS Publishing Data Working Group Bibliometrics**

*Co-Chairs: Kerstin Lehnert*

Bibliometrics are an essential ingredient for the broad adoption of data publication as part of scholarly communication. Bibliometrics provide incentives for investigators to publish their data, and will thus catalyze the necessary culture change toward comprehensive data sharing. The proposed RDA/WDS Publishing Data WG Bibliometrics aims to define the concept of data metrics as a way to quantitatively track the citation of published data in research and assess their use and impact. The WG will explore feasible options for indicators and their implementation across the broad range of stakeholders in digital data infrastructures, and develop a plan for adoption. Activities of the WG will be directed toward the following goals:

- Summarize current and emerging data citation practices of all stakeholders (journals, data centers, funders, societies), taking advantage of recent studies related to this topic;
- Understand and articulate necessary organizational and cultural changes in the scholarly publishing system needed to foster proper attribution of data sets;
- Evaluate and report on possible models how data citations can be successfully tracked and measured, based on existing and emerging approaches;
- Identify and report on possible barriers for the implementation and adoption of data citation and data metrics solutions.

In this session we will present the draft case statement of the group, gather feedback as well as information on existing practices, use cases, and ongoing or emerging studies, and solicit ideas for further engagement.

## **15:30-16:00 RDA/WDS Workflows for Publishing Data Working Group**

*Chair: Jonathan Tedds and Merce Crosas*

The objectives of this proposed RDA-WDS Joint Working Group are to investigate the range of possible workflows for publishing data, in particular reporting on:

- Current workflows for archiving and publishing data
- The role of Quality Assurance / Quality Control and peer-review in the publication process
- The role of science publishers and journals in the data publication process
- Barriers to implementation

In particular we will build on the work carried out to date in various international settings, including the UK Jisc funded PREPARDE project which provided reports on journal workflows from author submission of datasets and papers, through review to publication. This project also investigated data repository workflows from ingestion of data, through data centre technical review, to DOI assignment to dataset. While the focus was on mature examples in the Earth Sciences the reports also considered alternative paradigms which will be extended through the work of this group.

The group aims to identify workflow models for data publication across an appropriate range of use cases in different disciplines. We will provide the international research community with clear examples that may be adapted for use, in each case identifying the varying stakeholders and their different roles and responsibilities as well as the likely associated resource and cost implications, where possible.

This session will update participants on the work to date through the proposed joint RDA-WDS Workflows Working Group and invite further expressions of interest, e.g. in providing clear examples of use cases and relevant workflows.

## **16:00-16:15 Coffee break**

## **16:15-16:45 RDA/WDS Publishing Services Working Group**

*Chair: Hylke Koers*

The objective of this proposed RDA-WDS Joint Working Group is to address processes, workflows, and solutions that currently exist between individual parties within the data publication landscape, and investigate how these can be lifted to one-for-all services that increase interoperability, decrease systemic inefficiencies, and power new tools and functionalities to the benefit of researchers. Such processes may lie at different moments in the data publication workflow, including the submission, editorial, review, and publication process.

As a point of focus, the proposed working group takes particular interest in the problem of limited interoperability between data repositories, scholarly journal publication platforms, and tools for bibliometric analysis - in particular the absence of a common framework for cross-referencing data sets and published articles on different levels of granularity. This poses barriers and induces inefficiencies for the interlinking and contextualization of journal articles and data sets.

Key deliverables include:

- Inventory of interlinking, cross-referencing, and other tools and processes relevant to data publication currently in place. An analysis of pro's and con's, with an emphasis on scalability and do-ability.
- Gap analysis, including an analysis of needs & use cases for key stakeholders (data repositories, journal publishers, providers of bibliographic services, funding bodies, research institutions, the researcher).
- Recommendations for a one-to-all cross-resolving service that benefits the stakeholders in data publishing. These recommendations will include technical aspects, organizational, governance, and costs.
- An operational service (beta-release) for cross-referencing data sets and articles.

### **16:45-17:15 Reports from other groups**

*Chair: Simon Hodson*

### **17:15-18:15 General Discussion and Wrap up**

*Chair: Ingrid Dillo and Mustapha Mokrane*