

WDS-RDA Publishing Data Interest Group

To be submitted for RDA Council review as a joint RDA and WDS Interest Group

The **Publishing Data Interest Group** is focused on all issues related to the data publication concept. It brings together all stakeholders involved in data publication activities such as data repositories, science publishers, and service providers. Every effort will be made to get a good representation from the major international programmes, their working groups and other private or institutional activities involved in this area. The **Publishing Data Interest Group** 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 of the data publication concept. The **Publishing Data Interest Group** can be regarded as a broad and inclusive forum for interested individuals engaged in testing, validating and promoting the findings of the Working Groups.

The main objectives of the **Publishing Data Interest Group** are to promote and establish the data publication concept among scientists, data repositories, science publishers, and bibliometric service providers by addressing the issues of workflows for publishing data and establishing corresponding services as part of scholarly publishing.

Publishing Data

In the empirical sciences, data has traditionally been an integral part of scholarly publishing. In the last decades, rapid technical developments, such as digital data and high-throughput techniques, dramatically changed the scholarly publishing paradigm. This requires new approaches in order to ensure availability and usability of science data. However, existing approaches to address this issue are mostly technically dominated and lack success because they do not supply the necessary benefit for data producers. Instead, the concept of Data Publication is undergoing a renaissance as part of scholarly communication and on the base of new and proven technologies. Publishing data is a new and strong incentive for scientist to share their data and has positive effects on the data quality. The impact on citation rates could be shown in recent bibliometric studies on science articles having supplementary data.

Publishing data can be similar to the conventional publication of articles in journals that includes online submission, quality checks, peer-review, editorial decisions, and an equivalent of 'page proofs'. In fact, storage of data in public repositories and the ability to reference the datasets is getting increasingly important. It is already mandatory for the acceptance of peer-reviewed publications in specific fields of research such as molecular sciences or ecology. However, Data Publication as a generally accepted new publication type self-standing or supplementary to literature is not without controversy.

For data centres, science publishers, and service providers it is a challenge in terms of organization, technical developments, and funding. Compared to science articles the economic value of data is generally higher but they also need more resources for production, processing, long-term archiving and publication. If published data are to be usable as reliable as peer-reviewed science articles, they should not only meet

scientific requirements. Archiving and publishing procedures have to be transparent and accepted as part of the science culture. Moreover, published data should conform to current content and interoperability standards thus allowing for efficient usage and integration of data from various sources.

Initial Working Groups

The following 4 initial Working Groups are currently developing Case Statements under the umbrella of the **Publishing Data Interest Group**. New Working Groups can also be formed or join the Interest group as it develops.

1. Workflows for archiving and publishing data

- Investigate current workflows for archiving and publishing data
- The role of QA/QC and peer-review in the publication process
- The role of science publishers and journals in the data publication process

Deliverable: Provide generic workflow models for data publication

2. Bibliometrics including published data

- General requirements for citability of scientific data (granularity, citation information and persistent identification)
- Current citation practice in data centres and literature

Deliverable: Recommendations for data publishers & science publishers

3. Data publication services

- Existing service components to be used as building blocks
- Relevant content and interoperability standards
- Interoperability requirements for data centres (registration, metadata & data services)

Deliverable: Infrastructure and organization for a one-for-all cross-referencing service for science publishers and providers of bibliometric services

4. The costs for publishing data

- Investigate current cost structures for archiving and publishing data.
- Elaborate a business model based on open access which compensates for the additional costs due to data publication

Deliverable: Recommendations for funding organizations

Membership and meetings

The Co-Chairs will be responsible for the coordination of the IG work and updating the RDA online platform. They will report to the RDA Council as appropriate. Physical meetings will be organized during regular RDA meetings. The Interest Group will use the RDA structures for online collaboration and consultation between these meetings.

Co-Chairs:

- **Jonathan Tedds** (UK, University of Leicester)
- **Michael Diepenbroek** (Germany, PANGAEA)
- **Eefke Smit** (The Netherlands, STM)

Initial Membership:

- Euan Adie (UK, Altmetric.com)
- Jan Brase (Germany, DataCite)
- Ross Cameron (Scopus)
- David Carlson (UK, ESSD)
- Cyndy Chandler (US, Woods Hole Oceanographic Institution)
- Michael Diepenbroek (Germany, PANGAEA)
- Ingrid Dillo (The Netherlands, DANS)
- Kim Finney (Australia, AADC)
- Bettina Görner (Germany, Springer)
- John Helly (US, UCSD)
- Francisco Hernandez (Belgium, Flanders Marine Data Centre)
- Simon Hodson (UK, JISC)
- Yannis Ioannidis (Greece, University of Athens)
- Hylke Koers (The Netherlands, Elsevier)
- Kerstin Lehnert (US, IEDA)
- Caroline Martin (Eaux & Territoires Journal)
- Mustapha Mokrane (Algeria/France, ICSU World Data System)
- Fiona Murphy (UK, Wiley-Blackwell)
- Fiona Nielsen (UK, DNAdigest.org)
- Amy Nurnberger (Columbia University Libraries)
- Lyubomir Penev (Bulgaria, Pensoft Publishers)
- Lisa Raymond (US, Library Woods Hole Oceanographic Institution)
- Nigel Robinson (UK, Thomson Reuters)
- Sergio Ruiz (Germany, DataCite)
- Eefke Smit (The Netherlands, STM)
- Jonathan Tedds (UK, University of Leicester)
- Mark Thorley (UK, NERC)
- Frank Toussaint (Germany, DKRZ-WDC Climate)
- Mary Vardigan (USA, ICPSR)
- Anita de Waard (The Netherlands, Elsevier)
- Juanle Wang (China, WDC for Renewable Resources and Environment)
- Eva Zanzerkia (US, NSF)

Initial Working Groups membership

- **Workflows** : Jonathan Tedds, Kim Finney, John Helly, Hylke Koers, Fiona Murphy, Amy Nurnberger, Lisa Raymond, Mary Vardigan, Juanle Wang, Eva Zanzerkia
- **Bibliometrics** : Kerstin Lehnert, Euan Adie, Jan Brase, Ross Cameron, Cyndy Chandler, Fiona Murphy, Lyubomir Penev, Fiona Nielsen, Nigel Robinson, Mary Vardigan
- **Publishing Services** : Hylke Koers, David Carlson, John Helly, Francisco Hernandez, Caroline Martin, Lyubomir Penev, Nigel Robinson, Johanna Schwarz, Eva Zanzerkia
- **Cost Models and Policies** : Ingrid Dillo, Simon Hodson, Barbara Sierman, Frank Toussaint, Mark Thorley (observer), Kim Finney, Anita de Waard, Eva Zanzerkia