WDS/RDA Publishing Data Workflows: Working groups outputs
Who We Are

- Co-chairs: Sünje Dallmeier-Tiessen, Varsha Khodiyar, Fiona Murphy, Amy Nurnberger
- Members: 80 (!), representing diverse geographic, professional, and subject areas
Problem: Only a small fraction of research data is preserved and shared, often with a bare minimum of metadata – Often due to the lack of “established” or “trusted” services and workflows

Objectives:
- Provide an analysis of a representative range of existing and emerging workflows and standards for data publishing, including deposit and citation and providing reference models, a “classification”
- Test implementations of key components for application in new workflows
- Illustrate the benefits of the reference models for researchers and organisations
Stakeholders

▪ Researchers/Projects
▪ Data Publishers
  ▪ Institutional repositories
  ▪ Subject-specific repositories
  ▪ Journal publishers
▪ Research workflow development projects / tool developers

Because workflow information…
▪ Is crucial for researchers and other stakeholders to understand the options available for practicing open science, and to develop future options
▪ Illustrates possibilities for data sharing, leading to more efficient and reliable reuse of research data
▪ Builds trust in the data publication process, the data, and preservation
▪ Shows those involved in research data where they fit in the overall scheme of things
**Highlights of the Deliverable (1/2)**

### Recommendations

1. Start small, building modular, open source and shareable components
2. Implement core components of the reference model according to the needs of the stakeholder
3. Follow standards that facilitate interoperability and permit extensions
4. Facilitate data citation, e.g. through use of digital object PIDs, data/article linkages, researcher PIDs
5. Document roles, workflows and services

### Key components

<table>
<thead>
<tr>
<th>Required elements for published product</th>
<th>Additional elements for increased context, quality, visibility</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Primary data entry with PID</td>
<td>- Dataset Provenance/Provenance check</td>
</tr>
<tr>
<td>- Repository entry with persistent ID</td>
<td>- High documentation</td>
</tr>
<tr>
<td>- Metadata lifecycle review</td>
<td>- Link to results article</td>
</tr>
<tr>
<td>- Curation</td>
<td>- Link to data article</td>
</tr>
<tr>
<td>- Distribution/ discovery</td>
<td>- Link to software, code, simulations, etc.</td>
</tr>
<tr>
<td>- Editing process</td>
<td>- Peer review</td>
</tr>
<tr>
<td></td>
<td>- Submission support</td>
</tr>
<tr>
<td></td>
<td>- OAQC, Submission support</td>
</tr>
<tr>
<td></td>
<td>- Enable impact evaluation management platforms, metrics aggregations, author pages, 3rd parties</td>
</tr>
<tr>
<td></td>
<td>- Access support, value added services</td>
</tr>
</tbody>
</table>

**Definitions**


---


Highlights of the Deliverable (2/2)

Workflow model, for reference

Preliminary report:
- Connecting data publication to the research workflow: a preliminary analysis

http://bit.ly/1TvGe9v
http://dx.doi.org/10.5281/zenodo.34542
Impact of the Deliverable

This deliverable has the potential to:

▪ Impact **researchers** – who want to publish their data

▪ Impact **funders** – who want the data produced to be published

▪ Impact research **data repositories & publishers**

▪ Increase the practice of **open science/research**

▪ Increase efficient & reliable research data **reuse**

▪ Increase broad understanding of **roles & responsibilities**

▪ Increase **trust** in the process of long-term preservation
Giving Researchers Credit for their Data

- We’ve developed a helper app to support researchers in submitting data papers directly to journals, basic concept of which was grounded upon this WG’s survey and key findings on data publishing

Scientific Data

- Will hear from later

Academic Commons at Columbia University

- Columbia University's Academic Commons research repository is using the WDS/RDA Publishing Data Workflows Group reference model as a planning guide for technical, workflow, and policy development. The group's report is a valuable benchmarking tool in our effort to create a quality, open repository that can accept research data produced by Columbia researchers from many disciplines.

Elsevier Research Data Management Solutions

- Will hear from later
Research Space

The recommendations note that, "Many attempts to build infrastructure have failed because they are too difficult to adopt. The solution may be to enable infrastructure to develop around the way scientists and scholars actually work." RSpace adopts this recommendation by enabling convenient data capture and facilitating adding structure to the data without additional effort and at the natural time of data entry during the research process, and then supports deposit into a repository, retaining the structure, without requiring a duplicative effort by researchers to structure and prepare the data for deposit.

Elsevier Research Data Management Solutions

Will hear from later
ISPS Data Archive

The Yale Institution for Social and Policy Studies (ISPS) Data Archive adopts the recommendations of the RDA/WDS Publishing Data Workflows group. ISPS is currently developing software which structures data curation and publication workflows and will look to the model for further improvements to the software. In particular, the data publishing reference model and recommendations inform and confirm this work by presenting a coherent framework for data publishing which incorporates data quality assurance, a bedrock of the Archive. The software will be made available later in 2016.

Digital Curation Centre

The DCC (Digital Curation Centre) is drawing on the reference model for a new title in its series of How-to guides, which offer practical guidance to research organisations on delivering support for Research Data Management.
Edinburgh University Data Library: Edinburgh DataShare

Edinburgh University Data Library welcomes the RDA Publishing Data Workflow reference model and adopts its recommendations for Edinburgh DataShare by continuing to adopt relevant standards (e.g., DataCite DOIs, ORCIDs), remaining in good standing with Data Seal of Approval, maintaining interoperability with internal and external systems, supporting and improving our metrics transparency in landing page viewing statistics and other information about re-use and citations, and improving our documentation for depositors to encourage them to consider how to make their data as re-usable as possible ("Checklist for deposit"), acquaint them with our workflows in advance of deposit, and raise awareness around quality and data longevity (Data Seal of Approval).

GigaScience
Adopter: Scientific Data
Varsha Khodiyar, Data Curation Editor

www.nature.com/scientificdata/
Adopter: Scientific Data
Implementation of the required and additional elements

<table>
<thead>
<tr>
<th>Required elements for published product</th>
<th>Additional elements for increased context, quality, visibility</th>
</tr>
</thead>
</table>
| Primary data entry with PID            | Context
| Repository entry with persistent ID    | Parallel descriptions / dependent objects                    |
| Metadata generation/review             | Rich documentation                                          |
| Curation                               | Link to results article                                    |
| Distribution/discovery                  | Link to data article                                        |
|                                        | Links to software, code, simulations, etc.                  |

| Quality QA/QC, Submission support      | Editing process                                            |
|                                        | Enhanced curation by domain experts                        |
|                                        | Peer review                                               |
|                                        | Submission support                                        |

| Visibility/Accessibility                | Enable indexing/machine readability                        |
|                                        | Access support, value added services                      |
|                                        | Push to impact/evaluation management platforms, metrics aggregators, author pages, 3rd parties |

Implemented by Scientific Data
Under wider consideration by Springer Nature
ISA-explorer: A demo tool for discovering and exploring Scientific Data’s ISA-tab metadata

December 17, 2015 | 2:42 pm | Posted by Varsha Khodiyar | Category: Featured, Guest Posts, Metadata

Guest post by Dr Alejandra Gonzalez-Beltran, Research Lecturer, Oxford e-Research Centre, University of Oxford

blogs.nature.com/scientificdata/2015/12/17/isa-explorer

Search by:
• Data Repositories
• Experiment design
• Measurements made
• Technologies used
• Organism
• Environment types
• Geographic locations
Elsevier’s RDM Solutions is an early adopter of the WDS/RDA Publishing Data Workflows reference model

Elements already adopted by Elsevier RDM Solutions via Mendeley Data, Research Elements (data/software journals, like for example SoftwareX) and Data Linking Program

Elements currently under development or investigation for adoption
Elsevier’s RDM Solutions: an example of an updated publication workflow (since its last year evaluation)
### Who can use this deliverable?

<table>
<thead>
<tr>
<th>Do you:</th>
<th>Can you use this?</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Generate publishable research data?</td>
<td>YES</td>
</tr>
<tr>
<td>• Publish research data?</td>
<td>YES</td>
</tr>
<tr>
<td>• Fund publishable research data?</td>
<td>YES</td>
</tr>
<tr>
<td>• Have interest in any of the above?</td>
<td>YES</td>
</tr>
</tbody>
</table>

### How?

- Sign on as an adopter

http://dx.doi.org/10.5281/zenodo.34542
Next Steps and Contact Information

Next Steps:
- Distribute outputs broadly:
  http://dx.doi.org/10.5281/zenodo.34542
- Further develop report on connecting data publication to the research workflow

Further interest:
- WDS/RDA Publishing data interest group (3Mar, 0900, conference room 2)

Contact:
- data-pub-workflows@icsu-wds.org to reach any of the co-chairs or other contributors