Publishing Data Workflows

Chairs:

Jonathan Tedds (University of Leicester)

Sünje Dallmeier-Tiessen (CERN) - presenter

RESEARCH DATA ALLIANCE

ICSU WORLD DATA SYSTEM

Agenda

- Our objectives
- Where are we now?
- Next steps!
- Role of publishers

RDA
RESEARCH DATA ALLIANCE

ICSU
WORLD DATA SYSTEM

Background and Motivation

- Policy pressure vs. lack of incentives for researchers
- Only a small fraction of research data is preserved and shared
  - often with a bare minimum of metadata
- We believe this is often due to the lack of "established" or "trusted" services and workflows

But there are established or emerging workflows!

- Usually in selected disciplines e.g. Earth Sciences
- Some provide credit via citation mechanisms

Relevance

- Policy pressure vs. lack of incentives for researchers
- Having information about workflows is crucial
  - for researchers
  - and the people/stakeholders supporting them

to understand the options available to practice open science

- That's why we propose to study and test workflows that allow
  - efficient and reliable reuse of research data
  - to enhance the possibilities for greater discoverability

across disciplines and stakeholder groups

The working group members (currently)

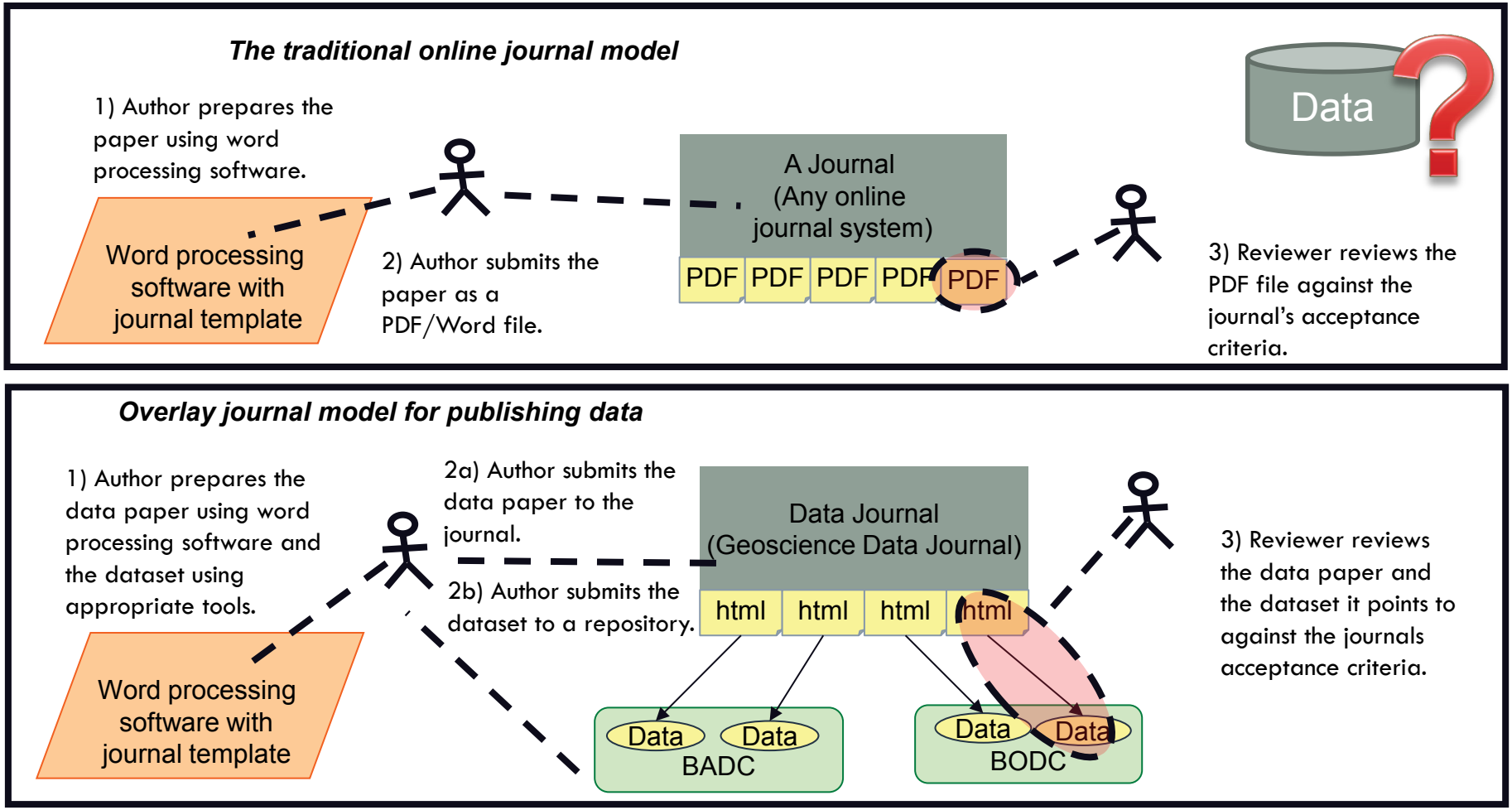
- Jonathan Tedds (UK, University of Leicester) [CO-CHAIR]
- Sünje Dallmeier-Tiessen (Switzerland, CERN) [CO-CHAIR]
- Merce Crosas (US, Harvard University)
- Michael Diepenbroek (PANGAEA)
- Kim Finney (Australia, AADC)
- John Helly (US, UCSD)
- Hylke Koers (The Netherlands, Elsevier)
- Rebecca Lawrence (UK, F1000 Research Ltd.)
- Fiona Murphy (UK, Wiley-Blackwell)
- Amy Nurnberger (Columbia University Libraries)
- Lisa Raymond (US, Library Woods Hole Oceanographic Institution)
- Johanna Schwarz (Germany, Springer)
- Mary Vardigan (US, ICPSR)
- Ruth Wilson (UK, Nature)
- Eva Zanzerkia (US, NSF)
- Angus Whyte (UK, DCC)
- Brian Hole (Ubiquity Press, UK)
- Varsha Khodiyar (UK, F1000 Research Ltd.)

Others are very welcome ☺

Objectives

- To provide an analysis of a representative range of existing and emerging workflows and standards for data publishing
  - including deposit and citation
  - provide reference models, a “classification”
- To test implementations of key components for application in new workflows.
- To illustrate the benefits of the reference model to researchers and organisations

Example: How to publish data in GDJ

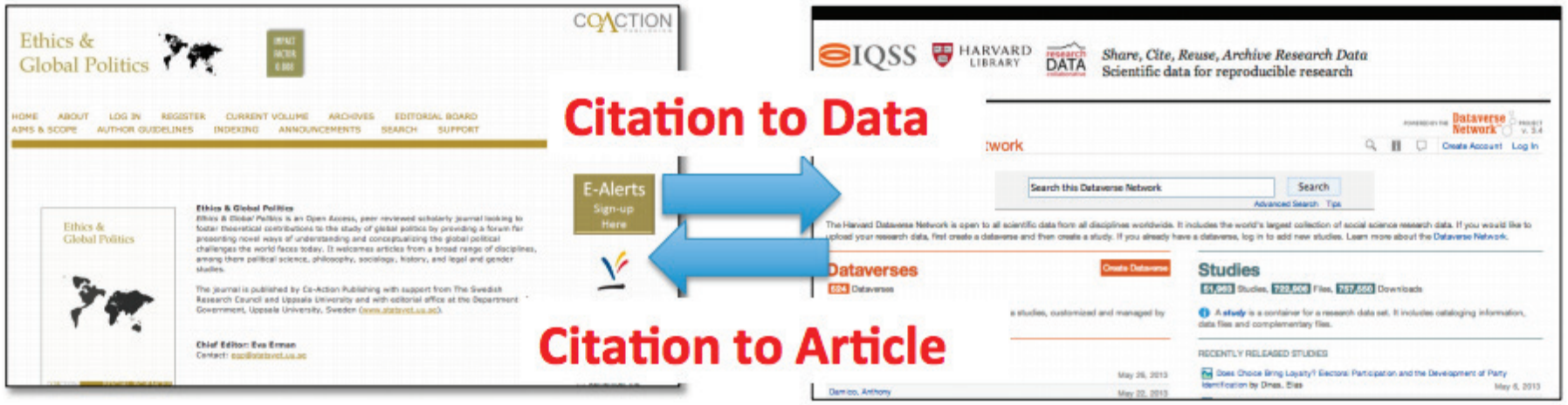


RESEARCH DATA ALLIANCE

WORLD DATA SYSTEM

Example:

- Integrate PKP's Open Journal System with Dataverse
- OJS plugin for: Data + metadata + supporting files, sent via SWORD API to the Dataverse
- http://projects.iq.harvard.edu/ojs-dvn



OJS Journal

Harvard Dataverse Network

More detailed work programme

- Identification of a smaller set of reference models covering a range of such workflows to include:
  - For example **when and where** QA/QC and data peer-review fit into the publishing process
  - **Who** does what and when…
  - Automated vs. “manual” processes
- Selection of key use cases and organizations in which components of a reference model can be **implemented** and tested for suitability
  - For example: dedicated data peer review
  - For example: metadata checks

Where are we now?

- Ongoing work right now:
  - Survey planning
  - Collecting and categorizing workflow examples

➔ **Share your own data publishing workflows**

- How do you choose, collaborate with or link to data repositories?
- Do you have data peer review?
- Do you have recommendations for data citation?

Current categories

1. Based on standards such as OAIS
2. Refers to a data life cycle
3. Includes PID assignment to data set
4a. Peer review of data
4b. Peer review of metadata
4c. Includes technical checks, e.g. for integrity
5a. Number of formats covered
5b. Formats covered
5c. Loss of data through normalisation
6. Number of people involved
7a. Links to grants
7b. Links to author PIDs
8a. Link to paper
8b. Standalone data
8c. Applicable for both cases
9. Final product

What are we missing?

Next steps: from theory to practice

- Are you interested in exploring the “data publication workflows” more?
  - Let us know!
- We would like to facilitate the “reuse” and transfer of trusted components
  - We are looking for test environments to implement
  - Let us know if you consider options, e.g. data peer review or guidelines for referencing data.

Role of publisher

- Becomes crucial in the second part of the initiative
- We will build reusable components that can be applied, in a journal for example
- Your editors and authors can benefit from “tested” workflows

→ Please let us know – if you have an interest in contributing and testing!