

### [What does it mean to Publish Data?](#) [1]

Submitted by NancyRitchey on Fri, 2014-10-10 15:16 Tuesday, January 6, 2015 - 15:30 to 17:00

**Event:** [Winter Meeting 2015](#) [2]

**Session Type:** [Breakout](#) [3]

**Expertise Level:** [Beginner](#) [4]

**Collaboration Area:** [Preservation and Stewardship](#) [5]

**Abstract/Agenda:** Recent national and international policies on open data and research results and the increase of similar policies from research funders and publishers have motivated the Earth Science community to enable data publishing through unique identifiers. In response, data repositories are minting unique identifiers such as a Digital Object Identifier (DOI) as part of their preservation services to enable data citations. However, publishers are keenly focused on uniquely identifying the data used to create the tables and figures within the publication, not necessarily the entire data record in its original form. This begs the question, "What does it mean to publish data?". This session will bring together speakers from Academia, Funders, Data Repositories and Publishing to understand their perspectives followed by a panel discussion addressing the ambiguous definition of "data publishing".

#### **Notes:**

Title: What does it mean to publish data?

Session Leads: Nancy Ritchey and Ruth Duerr

Speaker 1: Brooks Hanson ([bhanson@agu.org](mailto:bhanson@agu.org) [6]) of AGU and Kerstin Lehnert of ([lehnert@ldeo.columbia.edu](mailto:lehnert@ldeo.columbia.edu) [7]) EarthChem

Presentation Title: Coalition on Publishing Data in the Earth and Space Sciences, A Model to Advance Leading Data Practices in Scholarly Publishing

#### Notes:

- Substantial changes in the landscape of data publication 2014.
  - Ex: new data journals, such as AGU, data and data citation are increasingly being recognized as scholarly records, open access mandate is acknowledged, and cyberinfrastructure is being built.
- Publisher's perspective: several challenges, including lack of guidance, reinforcing compliance from the authors, concerns with repository funding and stability, and cost.
- Repository's perspective: data are valuable for discovery and integration; it is important to provide quality control and discovery; and currently there are poor connection to publications (often ad hoc), and therefore, there are needs for better integration with publication. There are cost concerns as well.
- Enabling Data Goals and Leverage: publication is a key value point for exposing data, and publishers/domain repositories would like to help enabling data and related best practices. However, in order to optimize achievable value for data, it is important to help authors early in the process.

## What does it mean to Publish Data?

Published on Commons (<https://commons.esipfed.org>)

---

- AGU hosted a meeting to bring together a coalition of publishers, funders, data facilities, and policy groups.
  - Result: A Statement of Commitment ([bit.ly/COPDESS-Statement](http://bit.ly/COPDESS-Statement))
  - Actions: Build an online directory of Earth and space science data repositories that can be used by journals and authors; develop work flows to support peer review, and promulgate metadata information.
- Additional trainings will become available in the near future.

### Comments/Questions from the Audience:

- Should the funding agencies be the ones that indicate how and where the data should be published?
  - It would be ideal from the publisher's point of view that the funders provide guidelines/requests regarding data publication. However, this might not ever become universal. As a result, it would also be helpful for others to help with the data publication effort. → The coalition indicated by the presentation could be a good way to influence all the factors/players in the data publication area.
- Currency, consistency and completeness are also a concern when building the directory of data repositories.

### Speaker 2: Robert Downs of CIESIN

#### Notes:

- Scientific Data Center Role in "Publishing Data": lifecycle approach, verify quality throughout the entire process, ensure developed data meet community needs, provide capabilities for diverse uses, and support ongoing data usage.
- How data are published at SEDAC:
  - Identifying, appraising, and selecting data
  - Developing data products and services
  - Disseminating data products and services

## What does it mean to Publish Data?

Published on Commons (<https://commons.esipfed.org>)

---

- Enabling data discovery, exploration, analysis
- Preserving the data for future use
- Challenges with “Publishing Data”:
  - Ensuring that competing demands don’t compromise data products and services offered (ex: deadlines between science team and data service team).
  - Providing capabilities to enable a variety of users to efficiently use data products and services.
  - Improving both internal processes and opportunities for users.
    - Feedback loops are provide for each major step.
- Minimum Requirements for Publishing Data
  - Plan (description of product or service and how it will be created)
  - Review (scientific and management)
  - Acquire (data rights)
  - Develop (creation of products/services and related information)
  - Quality Control (release in Alpha and Beta Configuration with Management Board reviews)
  - Archive (completed inventory, inspections, validation, packaging, and redundant storage)
  - Disseminate (discoverable and accessible online)
- Question from the Audience: Is DOI used?
  - Yes; also each data has a landing page with metadata and access method. However, the landing page is not machine readable in the sense that there is no content negotiation even though the page format is in HTML that is understandable to machines.

Speaker 3: Ruth Duerr of NSIDC

Presentation Title: Data Publication at NSIDC

Notes:

- Definition of “publishing data”: process whereby a data repository ingests data and make it available to a designated community(ies).
  - This can be a multi-step process that involves many parties (including assessment of data and the user community, update/creation of metadata and documentation, implementation of access methods, generate collateral, ingest data into archive, potentially creation of alternate method, and “switch publication on”).
- How rigorously the data publication is done depends on the funding agencies.
- Challenges: local and traditional knowledge and social conditions as well as funding requirements (sometimes lack thereof).
- What is the data?: Information collection could be dynamic or static, but what information should be included and considered as “data”?
- How are we handling it?: NSIDC has developed a process that helps them handle the different types of data formats/types.
- Minimum requirements for your organization?: NSIDC is currently working through this question at this time.
- A question from the audience: How much interactions do we have between the publishers and the data centers?
  - There are not a lot of interactions in the case of NSIDC; however, data and publication is not a one to one relationship, so there are a variety of issues that need to be worked through.
- A question from the audience: Do journal publishers consider data publication as “publication”?
  - There is still a distinction between peer-reviewed, reputable publications and those that do not meet the traditional criteria. The main concerns are the changes and the authenticity relating to “data publication”.
  - Also, the data publication in the journal setting is more focused on the datasets that are related specifically to the published journals.

## What does it mean to Publish Data?

Published on Commons (<https://commons.esipfed.org>)

---

- A question from the audience: Enabling data mining through data publication?
  - Defining the scope of the service would help defining what types of applications or uses that could be done with the published data.
- A comment from the audience: Citation and publication might not serve the same purpose.

Speaker 4: Peter Fox of RPI

Presentation Title: Is Data Publication the Right Metaphor?

Notes:

- Is “P/publication” with a capital P or small case p?
- Attributes of current data management paradigms are presented (from Peter and Mark’s paper): <http://dx.doi.org/10.2481/dsj.WDS-042> [8]
- Pros and cons of the different paradigms are also presented including data publication, Big Iron, and linked data.
- “Why data citation currently misses the point” poster by Peter and Mark: <http://www.slideshare.net/MarkParsons/parsons-citation-agu2014> [9]
- A question from the audience: Definitions will drift over time. How relevant are definition drifts?
  - It is part of the conversation regarding what “term” (or metaphor) we need to use in order to garner the most traction to encourage the discussion to define and implement data publication.
- A comment from the audience: “Publication” has different meanings/rules with different organizations. This is another reason why it is important for the data publication discussion to continue.

### Session Leads:

**Name:** [Nancy Ritchey](#) [10]  
**Organization(s):** [NOAA/NCDC](#) [11]  
**Email:** [nancy.ritchey@noaa.gov](mailto:nancy.ritchey@noaa.gov) [12]

**Name:** [Ruth Duerr](#) [13]  
**Organization(s):** [The Ronin Institute](#) [14]

### Notes takers:

**Name:** [Sophie Hou](#) [15]  
**Organization(s):** [UCAR/NCAR](#) [16]

## What does it mean to Publish Data?

Published on Commons (<https://commons.esipfed.org>)

---

### Participants:

In person:

- Viv Hutchison,
- Shannon Rauch,
- Audrey Mickle,
- Ken Casey,
- John Kusterer,
- Rama Ramapriyan,
- Melissa Zweng,
- Greg Janee,
- Sara Graves,
- Brooks Hansen,
- Joe Hourcle,
- Robert Downs,
- Laura McNulty,
- Don Collins,
- James Frew,
- Sarah O'Connor,
- Rahul Ramachandran,
- Nancy Ritchey,
- Ruth Duerr,
- Sophie Hou.

On webex:

- Aaron Sweeney,
- Ann Wilson,
- Bruce V,
- Heather Brown,
- Kathy Martinolich,
- Krisa Arzayus,
- Mark Parsons,
- Robert Summers,
- Steve Aulenbach,
- Chung-Lin Shie.

**Creative Common License:** Creative Commons Attribution 3.0 License

**Teaser:** Bring together data repositories and publishers to understand their perspectives and to begin the discussion needed to answer this question

**Accepted:**

**Keywords:** [publishing data](#) [17]

**Source URL:** <https://commons.esipfed.org/node/2719>

### Links

[1] <https://commons.esipfed.org/node/2719>

[2] <https://commons.esipfed.org/2015WinterMeeting>

[3] <https://commons.esipfed.org/session-type/breakout>

[4] <https://commons.esipfed.org/taxonomy/term/260>

[5] <https://commons.esipfed.org/collaboration-area/preservation-and-stewardship>

[6] <mailto:bhanson@agu.org>

[7] <mailto:lehnert@ldeo.columbia.edu>

## What does it mean to Publish Data?

Published on Commons (<https://commons.esipfed.org>)

---

- [8] <http://dx.doi.org/10.2481/dsj.WDS-042>
- [9] <http://www.slideshare.net/MarkParsons/parsons-citation-agu2014>
- [10] <https://commons.esipfed.org/node/2696>
- [11] <https://commons.esipfed.org/taxonomy/term/1492>
- [12] <mailto:nancy.ritchey@noaa.gov>
- [13] <https://commons.esipfed.org/node/295>
- [14] <https://commons.esipfed.org/taxonomy/term/2011>
- [15] <https://commons.esipfed.org/node/7872>
- [16] <https://commons.esipfed.org/taxonomy/term/2486>
- [17] <https://commons.esipfed.org/taxonomy/term/1511>