Data Citation Principles: A Synthesis

The Data Citation Synthesis Group

Maryann Martone
President, FORCE11
Professor of Neuroscience, University of California, San Diego

Supported by a grant from the Alfred P. Sloan Foundation
Agenda

• Welcome and introduction to principles
  – Maryann Martone, Micah Altman

• Panel discussion on implementation issues
  – Christine Borgman, Joan Starr, Anita de Waard, Ruth Duerr, Joe Hourcle, Sarah Callaghan, Puneet Kishor
  – Break

• Community Discussion
  – Mark Parsons
What is FORCE11?

- **Future of Research Communications and E-Scholarship**
  - A grass roots effort to accelerate the pace and nature of scholarly communications and e-scholarship through technology, education and community
- **Why 11?** We were born in 2011 in Dagstuhl, Germany
- **Principles laid out in the FORCE11 Manifesto**
- **FORCE11 launched in July 2012**
Who is FORCE11?

Anyone who has a stake in moving scholarly communication into the 21st century

- Scholars
- Tool builders
- Funders
- Policy makers
- Social Sciences
- Humanities
- Science
- Library and Information scientists
- Publishers
What does FORCE11 do?

>600 members from diverse stakeholder groups

- Community platform
  - Meetings
  - Discussions
  - Tools and resources
  - Blogs
  - Event calendar
  - Community projects

- Education
  - Scholarly communication 101
Data Citation Synthesis Group

- Leaders:
  - Paul Uhlir, NAS
  - Merce Crosas, Harvard

- 41 members
- ~20 organizations

http://www.force11.org/datacitationsynthesisgroup
Process

Synthesis

Community feedback

Revision

Dissemination

July-Sept 2013

Nov-Dec 2013

Jan 2014

Now

Data Citation Principles: Open for Endorsement

http://www.force11.org/datacitation/endorsements
Many thanks to our sponsors for this workshop
“We wish to promote best practices in data citation to facilitate access to data sets and to enable attribution and reward for those who publish data. Through formal data citation, the contributions to science by those that share their data will be recognized and potentially rewarded. To that end, we propose that:...”
But there were others...

Out of Cite, Out of Mind:  
The Current State of Practice, Policy, and Technology for the Citation of Data

CODATA-ICSTI Task Group on Data Citation Standards and Practices

For the purpose of clarity and simplicity, we have prioritized creating a list of principles, which, for many, may appear to be a laundry list, but we believe it is important to do so, as we attempt to evolve from practices, we have

1. Traceability: Citations should allow for the traceability of the citation to the actual objects of concern.

2. Attribution: Citations should facilitate giving scholarly credit and legal attribution to all parties responsible for the data.

3. Persistence: Citations should be as durable as the cited objects.

4. Access: Citations should facilitate access to data by humans and by machines.

5. Discovery: Citations should support the discovery of data and their documentation.

6. Provenance: Citations should facilitate the establishment of provenance of data.

7. Granularity: Citations should support the finest-grained description necessary to identify the data.

8. Verifiability: Citations should contain information sufficient to identify the data unambiguously.

9. Metadata Standards: Citations should employ widely accepted metadata standards.

10. Flexibility: Citation methods should be sufficiently flexible to accommodate the variant practices among communities.

These principles are offered as guides to implementers, not as an endorsement of any particular implementation.
And others...

- Many groups working on similar documents, each from a slightly different perspective.
“...a plea to come together”

• Groups pledged to work together through FORCE11
• Review existing guidelines
• Come up with a set of consensus principles
  – “Data Citation Declaration”
  – Widely disseminate
Mission

• The data citation synthesis group is a cross-team committee leveraging the perspectives from the various existing initiatives working on data citation to produce a consolidated set of data citation principles (based on the Amsterdam Manifesto, the CODATA and other sets of principles provided by others) in order to encourage broad adoption of a consistent policy for data citation across disciplines and venues. The synthesis group will review existing efforts and make a set of recommendations that will be put up for endorsement by the organizations represented by this synthesis group.
Scope

• The synthesis group will produce a set of principles, illustrated with working examples, and a plan for dissemination and distribution. This group will *not* be producing detailed specifications for implementation, nor focus on technologies or tools.