

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

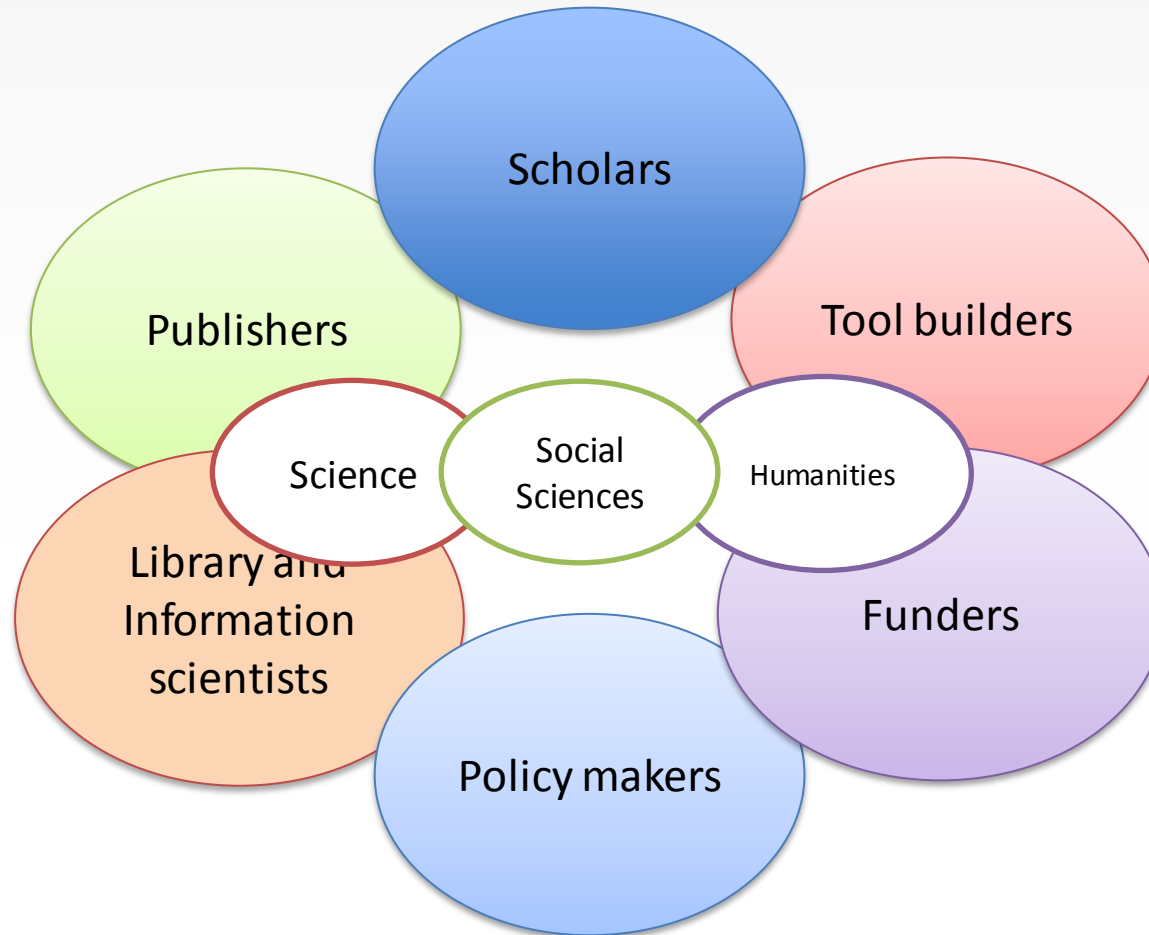
- Welcome and introduction to principles
 - Maryann Martone, Micah Altman
- Panel discussion on implementation issues
 - Christine Borgman, Joan Starr, Anita de Waard, Ruth Duerr, Joe Hourcle, Sarah Callaghan, Puneet Kishor
 - Break
- Community Discussion
 - Mark Parsons

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?

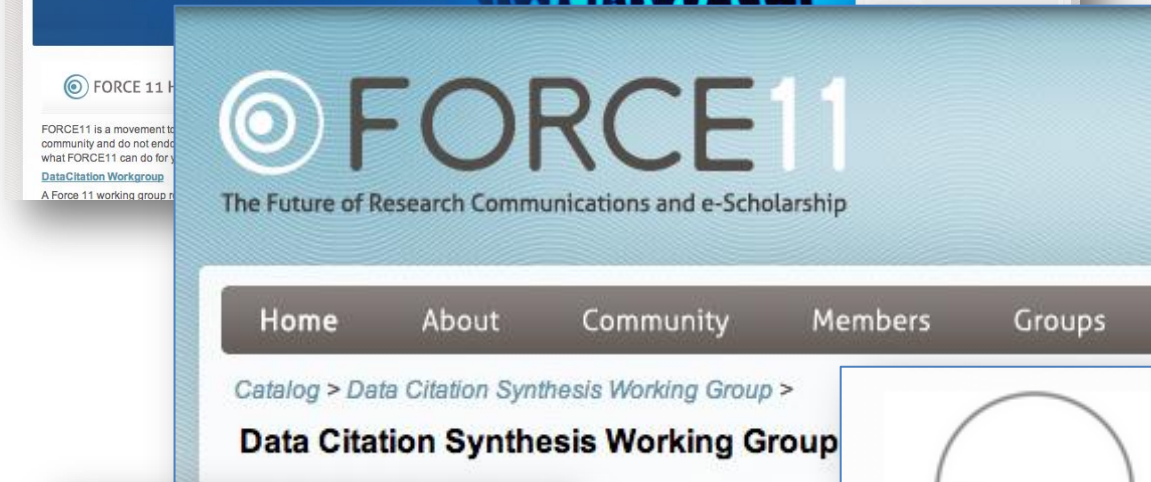


Anyone who has a stake in moving scholarly communication into the 21st century

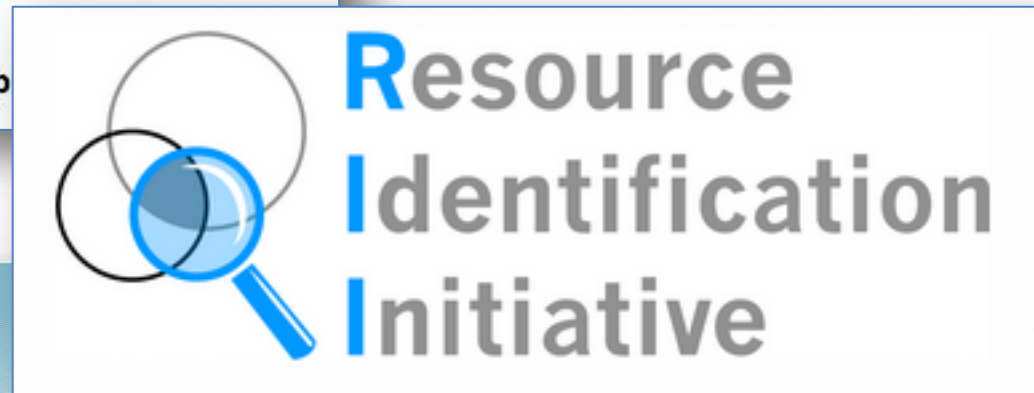
What does FORCE11 do?

>600 members from diverse stakeholder groups

- Community platform
 - Meetings
 - Discussions
 - Tools and resources
 - Blogs
 - Event calendar
 - Community projects
- Education
 - Scholarly communication 101

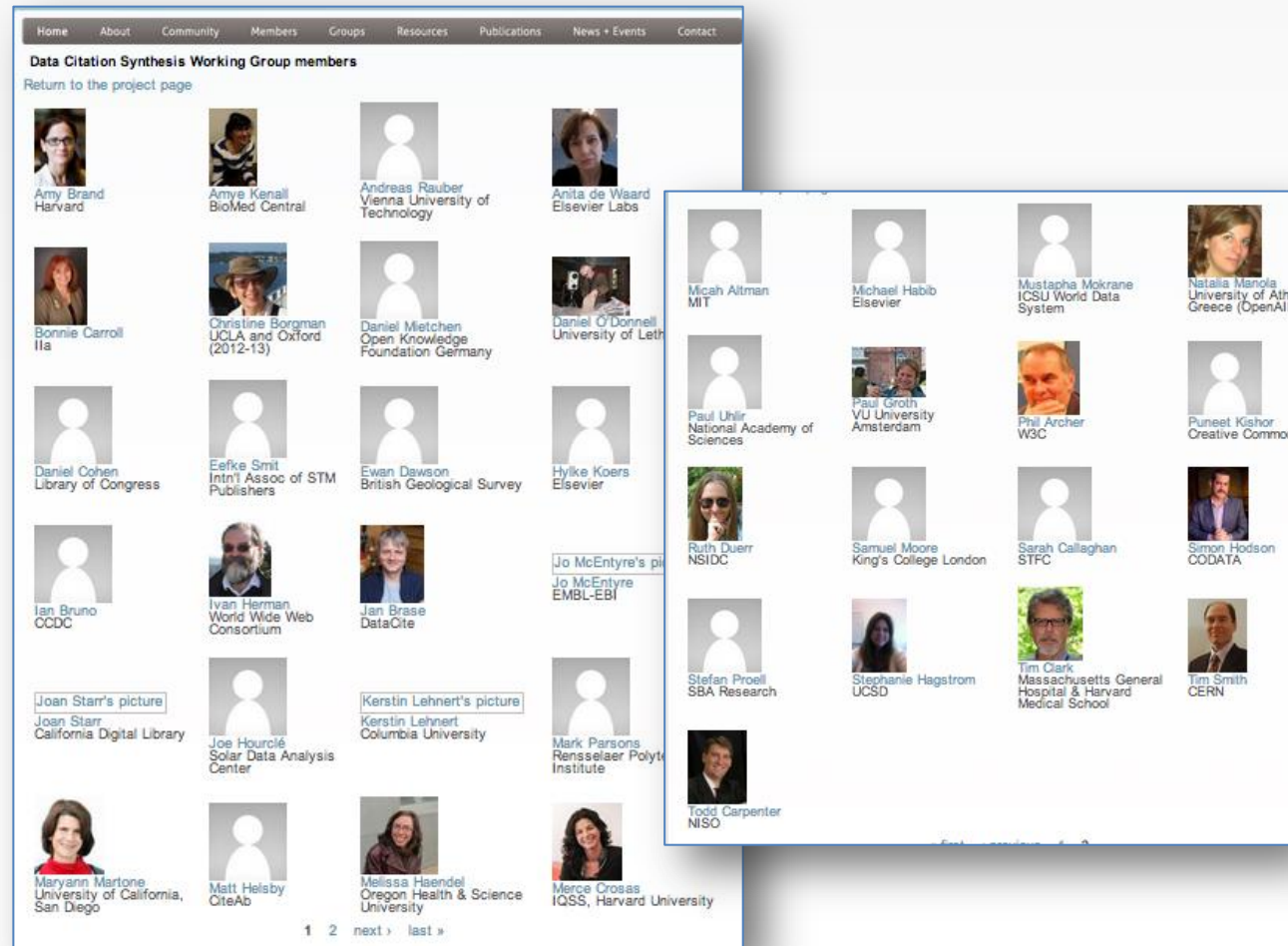


The Future of Research Communications and e-Scholarship



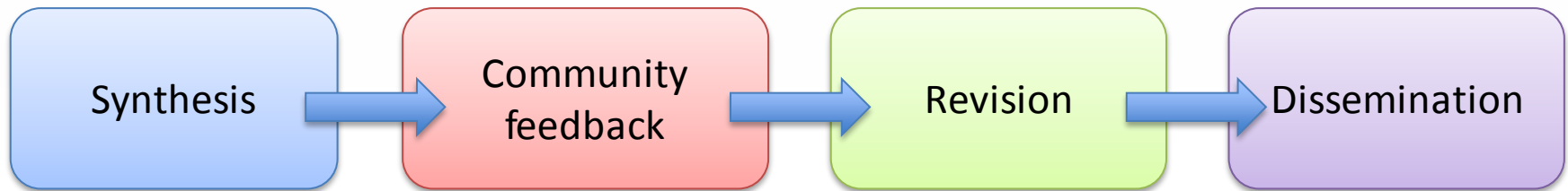
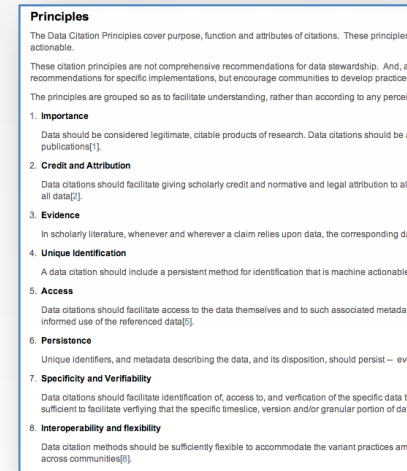
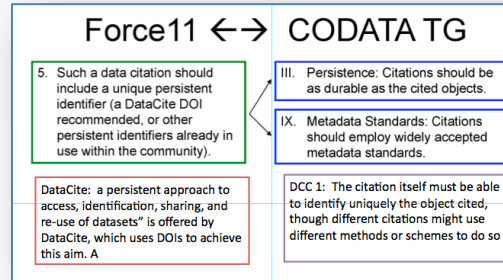
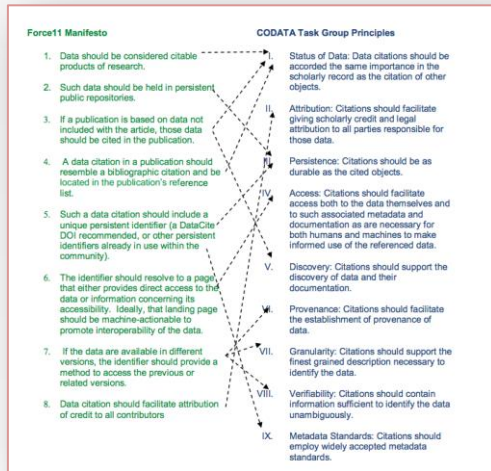
Data Citation Synthesis Group

- Leaders:
 - Paul Uhler, NAS
 - Merce Crosas, Harvard
- 41 members
- ~20 organizations



<http://www.force11.org/datacitationsynthesisgroup>

Process



July-Sept 2013

Nov-Dec 2013

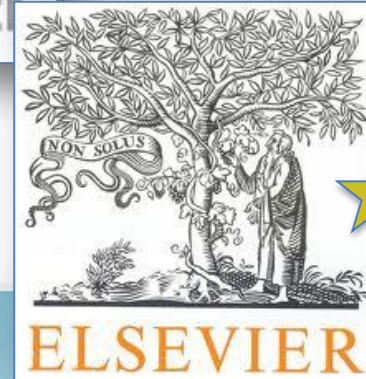
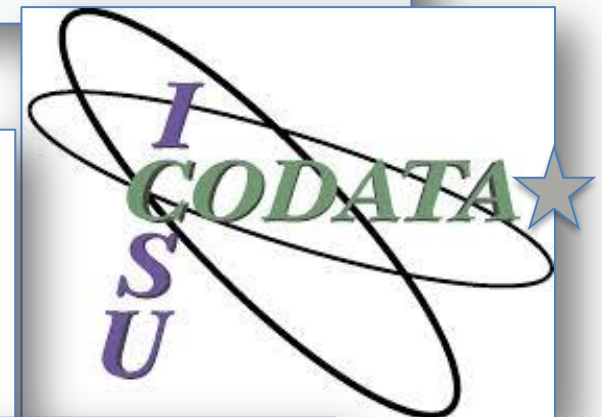
Jan 2014

Now

Data Citation Principles: Open for Endorsement

<http://www.force11.org/datacitation/endorsements>

Many thanks to our sponsors for this workshop



Amsterdam Manifesto: Simple set of data citation principles

Amsterdam Manifesto

“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:...”

2. Such data should be held in persistent public repositories.
3. If a publication is based on data not included with the article, those data should be cited in the publication.
4. A data citation in a publication should resemble a bibliographic citation and be located in the publication's reference list.
5. Such a data citation should include a unique persistent identifier (a DataCite DOI recommended, or other persistent identifiers already in use within the community).
6. The identifier should resolve to a page that either provides direct access to the data or information concerning its accessibility. Ideally, that landing page should be machine-actionable to promote interoperability of the data.
7. If the data are available in different versions, the identifier should provide a method to access the previous or related versions.
8. Data citation should facilitate attribution of credit to all contributors

About

This Manifesto was created during the [Beyond the PDF 2 Conference](#) in Amsterdam, 20 March 2013.

Original authors are Mercè Crosas, Todd Carpenter, David Shotton and Christine Borgman.

Original document with comments from BTPDF2

Mercè Crosas, Todd Carpenter, David Shotton and Christine Borgman

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

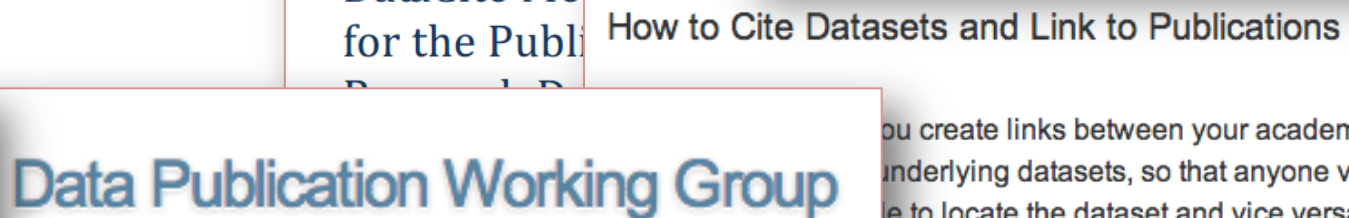
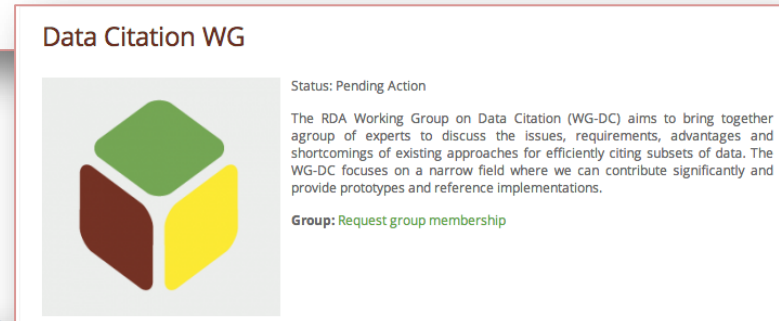
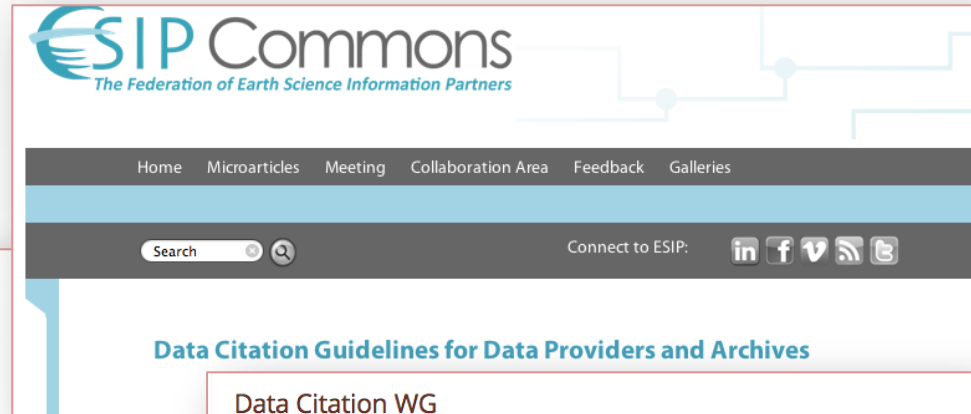
citation of other objects.

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