

Mercè Crosas

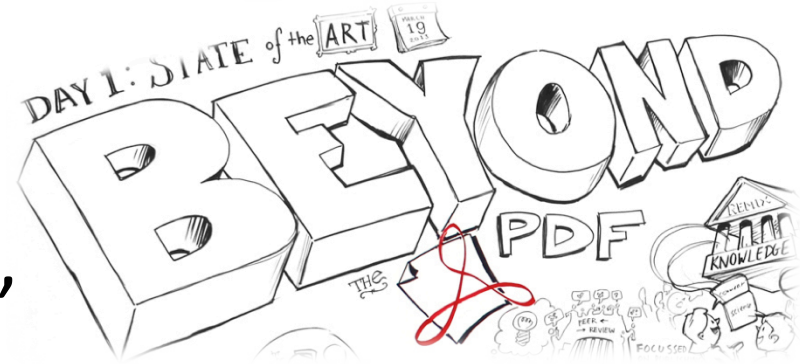
Director of Data Science, IQSS, Harvard University

**DECLARATION OF DATA CITATION  
PRINCIPLES**

~~AMSTERDAM MANIFESTO~~

# A little history

- Beyond the PDF in Amsterdam, March 2013, organized by **Force11**



- 1K Challenge:

*What would you do to change scholarly communication?*

# Amsterdam Manifesto:

## A 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:...”

those who  
and potentially

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 other Principles ...

## 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.

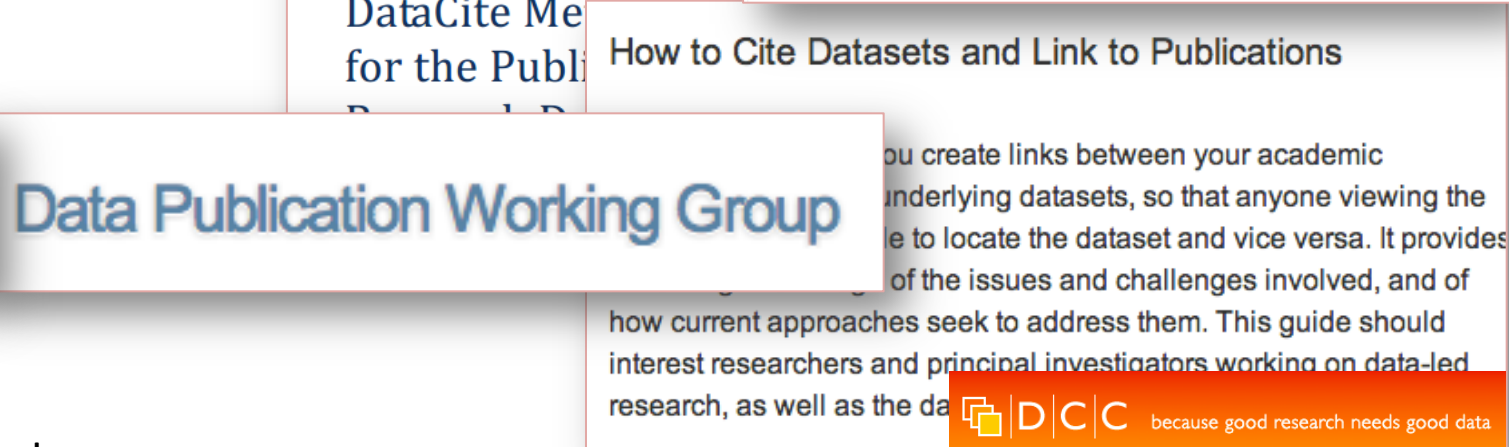
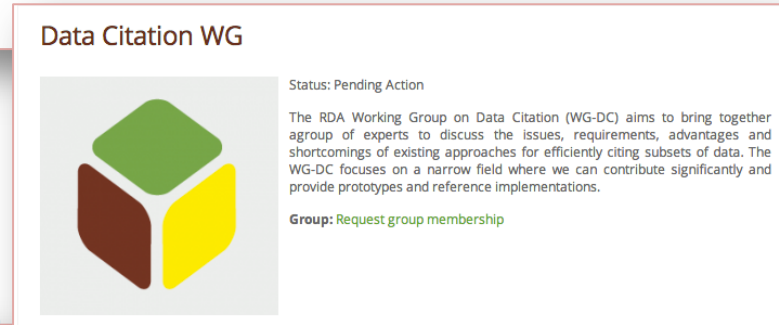
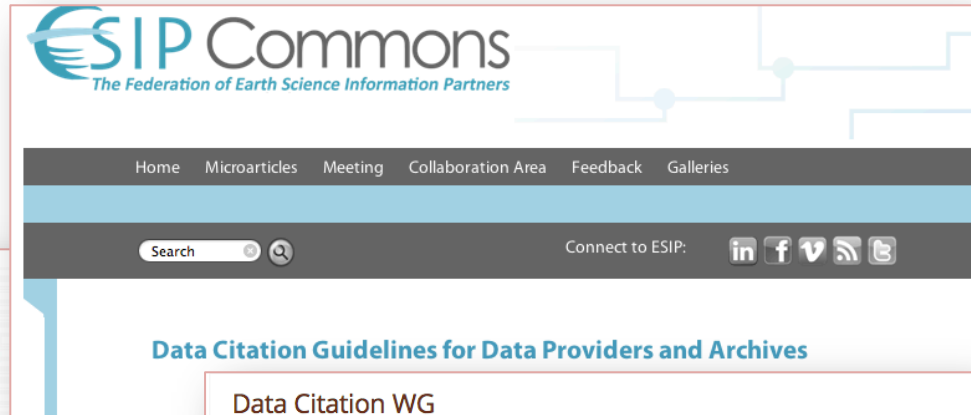
These principles are offered as guides to implementers, not as an endorsement of any particular implementation.

practices, we have

arly record as the

# And others ...

Many groups working on similar documents, each from a slightly different perspective





Paul Uhlir “...a plea to come together”

Photo: Flickr

# Data Citation Synthesis Group

- 36 members, ~ 20 organizations
- Weekly meetings for 3 months, led by Force 11
- Comparison of 4 sets of principles:
  - Amsterdam Manifesto
  - CoData Data Citation Principles
  - DataCite Data Citation Principles
  - Digital Data Center Data Citation Principles
- Merge them to create a synthesis set of principles - high level and simple

# **THE CONSENSUS: 8 PRINCIPLES**



Data should be considered legitimate, citable products of research. Data citations should be accorded the same importance in the scholarly record as citations of other research objects, such as publications.

## **1. IMPORTANCE**

Data citations should facilitate giving scholarly credit and normative and legal attribution to all contributors to the data, recognizing that a single style or mechanism of attribution may not be applicable to all data.

## **2. CREDIT AND ATTRIBUTION**

Where a specific claim rests upon data, the corresponding data citation should be provided.

### **3. EVIDENCE**

A data citation should include a persistent method for identification that is machine actionable, globally unique, and widely used by a community.

## 4. UNIQUE IDENTIFIERS

Data citations should facilitate access to the data themselves and to such associated metadata, documentation, and other materials, as are necessary for both humans and machines to make informed use of the referenced data.

## 5. ACCESS

Metadata describing the data, and unique identifiers should persist, even beyond the lifespan of the data they describe.

## **6. PERSISTENCE**

Data citations should facilitate identification and access to different versions and/or subsets of data.

Citations should include sufficient detail to verifiably link the citing work to the portion and version of data cited.

## **7. VERSIONING AND GRANULARITY**

Data citation methods should be sufficiently flexible to accommodate the variant practices among communities but should not differ so much that they compromise interoperability of data citation practices across communities.

## **8. INTEROPERABILITY AND FLEXIBILITY**



**WHAT NEXT**

# Review Process

- Now to End of October:
  - Draft document by synthesis group:
    - Preamble
    - Principles (“as is”)
    - Supplementary materials (example + glossary + references)
- November – January:
  - Open document to wider review (publishers, data repositories, libraries, funders, gov agencies, ...)

# Dissemination Plan

- Beginning January:
  - Integrate comments from stakeholders/community
- Mid January:
  - Publish final Principles in web site
  - Initiate endorsement requests
  - Press releases, dissemination to scholars within institutions, social media, media kit
  - Transition to other groups next steps after Declaration is endorsed widely



Amy Brand  
Harvard



Amye Kenall  
BioMed Central



Andreas Rauber  
Vienna University of  
Technology



Anita de Waard  
Elsevier Labs



Paul Groth  
VU University  
Amsterdam



Phil Archer  
W3C



Puneet Kishor  
Creative Commons



Ruth Duerr  
NSIDC



Bonnie Carroll  
IIA



Christine Borgman  
UCLA and Oxford  
(2012-13)



Daniel O'Donnell  
University of Lethbridge



Daniel Cohen  
Library of Congress and  
Board on Research  
Data and Information  
(Nat'l Academy of  
Sciences))



Sarah Callaghan  
STFC



Simon Hodson  
CODATA



Stefan Proell  
SBA Research



Stephanie Hagstrom  
UCSD



David Shotton  
University of Oxford



Eefke Smit  
Intn'l Assoc of STM  
Publishers



Hylke Koers  
Elsevier



Ivan Herman  
World Wide Web  
Consortium



Tim Clark  
Massachusetts General  
Hospital & Harvard  
Medical School



Tim Smith  
CERN



Todd Carpenter  
NISO



Willy S



Jan Brase  
DataCite



Jo McEntyre  
EMBL-EBI



Joan Starr  
California Digital Library



Kerstin Lehnert  
Columbia University



Mark Parsons  
Rensselaer Polytechnic  
Institute



Maryann Martone  
University of California,  
San Diego



Melissa Haendel  
Oregon Health &  
Science University



Merce Crosas  
IQSS, Harvard  
University



Micah Altman  
MIT



Mustapha Mokrane  
ICSU World Data  
System



Natalia Manola  
University of Athens,  
Greece (OpenAIRE)



Paul Uhler  
National Academy of  
Sciences

**THANK YOU!**  
The Synthesis Group

<http://www.force11.org/node/4381>