Data Citation:
framing the discussion and global context

Dr Simon Hodson

Executive Director, CODATA

www.codata.org
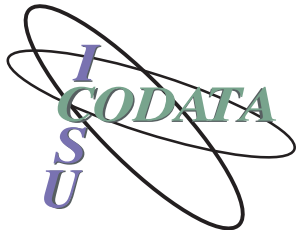
What do we mean by data citation / referencing

1. **Citation as reference: a citation is the means by which we indicate in scholarly work that a given assertion relies upon a given source. It points the reader to the evidence.**

   a) Pointing to the date, where appropriate accessible via a trusted digital repository, is good research practice.

   b) Assists verification, reproducibility, reuse.

   ➢ Desideratum: incorporate data more effectively into the process of scholarly communications.

2. **Citations as a means of crediting research contributions.**

   a) Citations as credit exert considerable influence over researcher motivations both for recognition and reward (career advancement).

   b) Data citations are not currently used in metrics around research contributions.

   ➢ Desideratum: encourage data citations to credit research contributions.
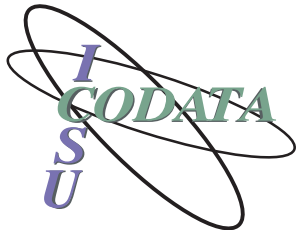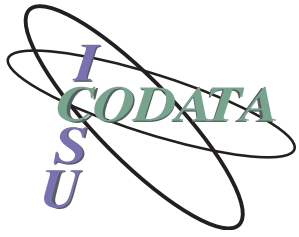
Data and Scholarly Communications

Royal Society Report: Science as an Open Enterprise

CODATA President Geoffrey Boulton, FRS

- Outlines the response required to epoch-making changes in technology that are fundamentally affecting the conduct of research.

- Responding to these changes requires more than just Open data, it requires 'Intelligent Openness': data should be accessible, assessable, intelligible, usable.

- 'As a first step towards this intelligent openness, **data that underpin a journal article should be made concurrently available in an accessible database.** We are now on the brink of an achievable aim: for all science literature to be online, for all of the data to be online and for the two to be interoperable.'
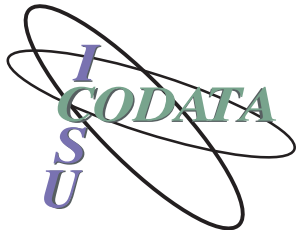
First Statement on Open Research Data

- *Nullius in Verba*
- (Take nobody's word for it)
- Science is about only believing things for which there is evidence!
- Verify all statements by an appeal to facts determined by experiment.
- Motto of the Royal Society, adopted 1663

PHILOSOPHICAL TRANSACTIONS:

GIVING SOME

ACCOMPT

OF THE PRESENT

Undertakings, Studies, and Labours

OF THE

INGENIOUS

IN MANY

CONSIDERABLE PARTS

OF THE

WORLD.

Vol I.

For *Anno* 1665, and 1666.

In the SAVOY,

Printed by *T. N.* for *John Martyn* at the Bell, a little without *Temple-Bar*, and *James Allestry* in *Duck-Lane*, Printers to the *Royal Society*.

Scientific Malpractice

Failing to make data available and failing to cite data is scientific malpractice.

'publishing an article without at the same time making the data or evidence available is scientific malpractice'
[Geoffrey Boulton at KE Workshop, 'Making Data Count'
http://www.knowledge-exchange.info/ Default.aspx?ID=576]

'Publishing research without data is advertising, not science'
[@McDawg, Twitter, #solo12data]

Sarah Callaghan
@sorcha_ni

Following

Audience comment: Publishing research without data is advertising, not science #solo12data

Reply   Retweet   Favorite   More

9 RETWEETS

4:06 PM - 12 Nov 12

Reply to @sorcha_ni

Graham Steel @McDawg   12 Nov
That was moi !! MT @sorcha_ni Audience comment: Publishing research without data is simply advertising, not science #solo12data
Details

Credit for Data: from a **vicious** to a virtuous circle?

Data sharing and data reuse cannot be measured making it difficult to credit and reward

Scholars don't share because it is not credited and not worth the expenditure of time and effort

Data sharing and reuse transactions are low, making credit and metrics relating to data difficult or impossible

See The Value of Research Data: Metrics for datasets from a cultural and technical point of view
http://www.knowledge-exchange.info/datametrics
and Knowledge Exchange Workshop: 'Making Data Count':
http://www.knowledge-exchange.info/Default.aspx?ID=576

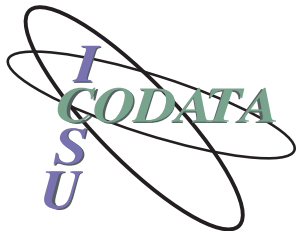# Credit for Data: from a vicious to a **virtuous** circle?

Scholars share data as a matter of course because it is accepted practice and it is credited

Data sharing and reuse transactions are high, making credit, metrics and other assessment around data effective

Data sharing and data reuse can be measured making it easy to credit and reward

See The Value of Research Data: Metrics for datasets from a cultural and technical point of view
http://www.knowledge-exchange.info/datametrics
and Knowledge Exchange Workshop: 'Making Data Count':
http://www.knowledge-exchange.info/Default.aspx?ID=576

Building a Culture of Data Citation

CREATE

USE

3

4

doi

doi is used in data citation

Australian researcher creates a research dataset and a publication related to the dataset

Dataset is stored in a publicly accessible repository

Research community use the **doi** to access the dataset and carry out related research

doi

Research community generate new publications using the **doi** to reference the dataset

1

Researcher uses ANDS services to mint a Digital Object Identifier (**doi**) for the dataset

doi

ands
AUSTRALIAN NATIONAL DATA SERVICE

2

Citation metrics services (eg Scopus, Web of Knowledge) accumulate citation references to the dataset and publication

doi

Researcher future funding and promotion influenced by dataset citation metrics

5

Funding and research groups review publication and dataset citation metrics

7

REWARD

6

MEASURE

ands
AUSTRALIAN NATIONAL DATA SERVICE

ands.org.au

http://www.ands.org.au/guides/data_citation_poster.pdf

CODATA Working Group and DataCite

- Presentation and discussions of ideas at CODATA Conference 2000, Michael Lautenschlager and Joachim Waechter.

- Followed by a concept paper and agreement from DFG to fund CODATA Task Group on 'Possibility of Citing Scientific Primary Data', early 2001.

- Aims 'to discuss between librarians and scientists the concept of scientific data publication and come up with an implementation plan'.

- **Final Report: 'Conception of citing scientific primary data' delivered 29 May 2002.**

- From 2003: Pilot Implementation; from 2006 developed as a service, assigning DOIs for data; 2009 foundation of DataCite.

- *Thanks to Michael Lautenschlager*

- http://datacite.inist.fr/IMG/pptx/datacite-keynote-annconf-nancy-25-260814-final-rev.pptx

17th International CODATA Conference (2000)
Baveno, Lago Maggiore

Tischvorlage CODATA Landesausschußsitzung
29.11.00 in Bonn:

Publikation und Zertifizierung von
wissenschaftlichen Daten

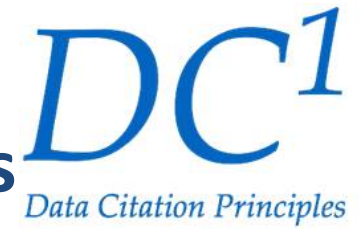Michael Lautenschlager (MPIM/M&D)
Joachim Wächter (GFZ)

Der Fortschritt der modernen wissenschaftlichen Forschung geht mit einem enormen Gewinn an Daten einher. Die einzelnen Disziplinen, z.B. Biologie oder die Geowissenschaften sind dabei dieses Problem zu betrachten. Von zentraler Bedeutung sind wissenschaftliche Netzwerke, mit denen die Ergebung, das Verfügbarmachen, sowie die Integration und Nutzung der großen, heterogenen, interdisziplinären Datenmengen unterstützt wird.

DFG Deutsche
Forschungsgemeinschaft
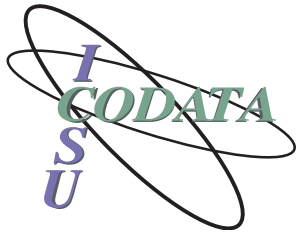
CODATA-ICSTI Task Group
Data Citation, Standards and Practices

DC¹
Data Citation Principles

Out of Cite, Out of Mind

http://bit.ly/out_of_cite
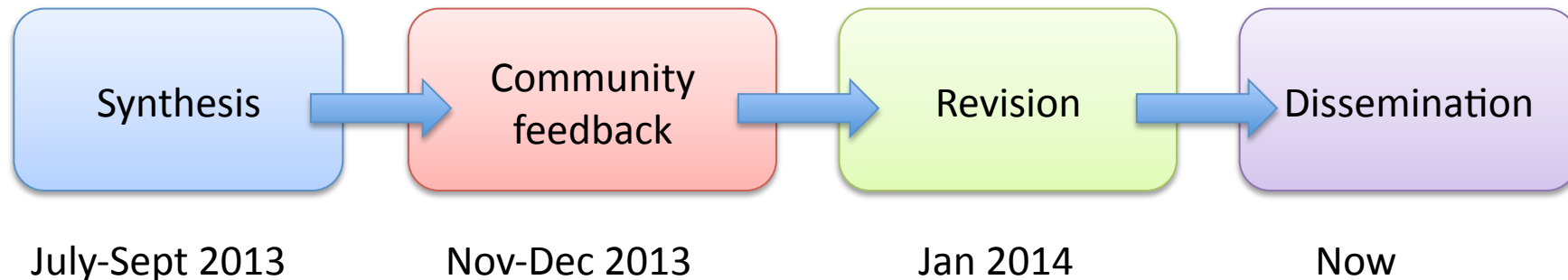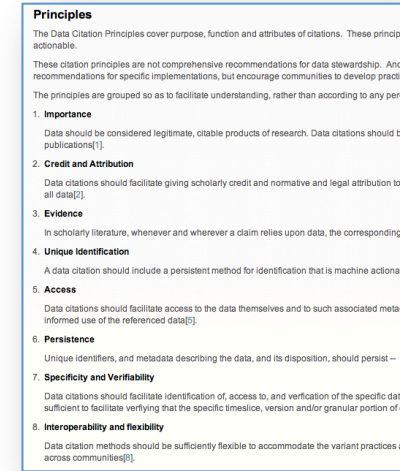
Joint Declaration of Data Citation
Principles:
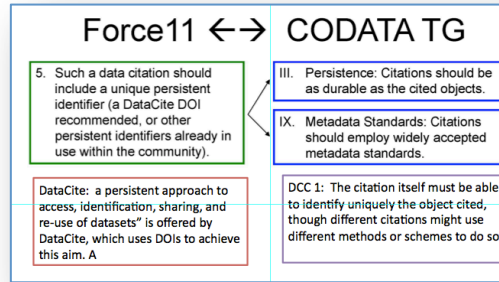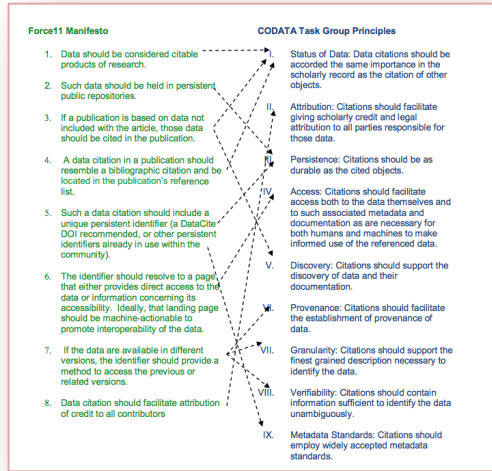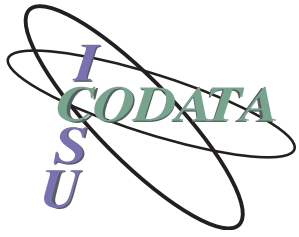https://www.force11.org/datacitation

Background and Developments:
http://bit.ly/data_citation_principles

If publications are the stars and planets of the scientific universe, data are the 'dark matter' – influential but largely unobserved in our mapping process

Process

Synthesis → Community feedback → Revision → Dissemination

July-Sept 2013   Nov-Dec 2013   Jan 2014   Now

Data Citation Principles: Open for Endorsement

Joint Declaration of Data Citation Principles (Overview); Slide: Sarah Callaghan
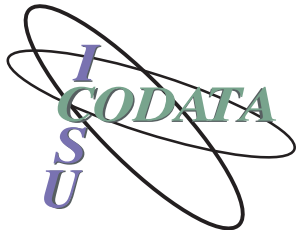
Joint Data Citation Principles
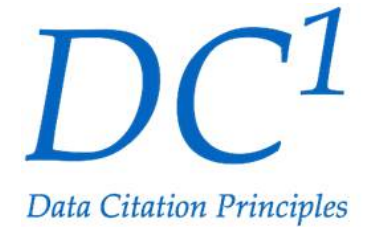
Preamble

Sound, reproducible scholarship rests upon a foundation of robust, accessible data. For this to be so in practice as well as theory, data must be accorded due importance in the practice of scholarship and in the enduring scholarly record. In other words, data should be considered legitimate, citable products of research. Data citation, like the citation of other evidence and sources, is good research practice and is part of the scholarly ecosystem supporting data reuse.

In support of this assertion, and to encourage good practice, we offer a set of guiding principles for data within scholarly literature, another dataset, or any other research object.

The Data Citation Principles cover purpose, function and attributes of citations…
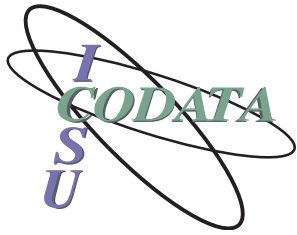
Joint Data Citation Principles
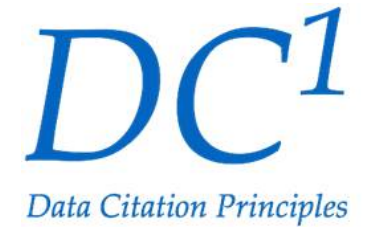
Purpose

1. Importance. Data should be considered legitimate, citable products of research. **Data citations should be accorded <u>the same importance in the scholarly record</u> as citations of other research objects, such as publications.**

2. Credit and attribution. 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.

3. Evidence. In scholarly literature, whenever and wherever a claim relies upon data, the corresponding data should be cited.

Function

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

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

Attributes

6. Persistence. Unique identifiers, and metadata describing the data and its disposition, should persist -- even beyond the lifespan of the data they describe.

7. Specificity and verifiability. Data citations should facilitate identification of, access to, and verification of the specific data that support a claim. **Citations or citation metadata should include information about provenance and fixity sufficient to facilitate verifying that the specific timeslice, version and/or granular portion of data retrieved subsequently is the same as was originally cited.**

8. Interoperability and flexibility. 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].

Endorse the Data Citation Principles

https://www.force11.org/datacitation/endorsements

Add Endorsement

Joint Declaration of Data Citation Principles

Post to Twitter

Individual Endorsements

120 Endorsements

First Name	Last Name	Affiliation	Endorsement Date
Alberto	Accomazz	NASA Astrophysics Data System	2014-02-27 16:21
Donat	Agosti	Plazi	2014-03-04 13:25
Micah	Altman	MIT	2014-02-27 09:56
Martin	Alvarez Espinar		2014-02-27 03:52
Eva	Amsen	F1000Research	2014-03-10 11:04
Roger	Barry	NSIDC/CIRES, Univ. of Colorado	2014-02-28 08:07
Rob	Baxter	EPCC, University of Edinburgh	2014-02-28

Organization Endorsements

40 Endorsements

Organization	Endorsement Date
BioMed Central	2014-03-21 01:30
CODATA	2014-02-27 09:59
CODATA-ICSTI Task Group on Data Citation Standards and Practices	2014-02-27 07:59
DANS	2014-03-07 05:07
DataCite	2014-03-26 04:50
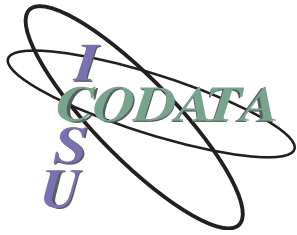
What does a data citation look like?

What is best practice in data citation?

Citing data in academic publications isn't new; the UK Data Archive and other data repositories around the world have been requiring it as part of the standard user agreement for many years.

But citing data using persistent identifiers (such as DOIs) is an area that is rapidly growing in importance especially given the increased need to track citations as a way to understand impact. In 2012, DOIs became part of a new international standard, ISO 26324.

Many journals and style manuals already use DOIs for formally citing publications. The next step is to include standard electronic citation of research data.

The ESRC recommends using a citation format for data that is based on DataCite metadata properties.

Essential properties
- Identifier
- Creator
- Title
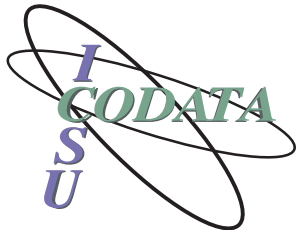- Publisher
- Publication year

Optional properties
- Resource type
- Version

As with publications, these citations are often automatically generated. Data users need only copy and paste.

Example citation from the ESRC's Economic and Social Data Service (ESDS)

University of Essex. Institute for Social and Economic Research and National Centre for Social Research, *Understanding Society: Wave 1, 2009-2010 and Wave 2, Year 1 (Interim Release)*, 2010 [computer file]. *3rd Edition*. Colchester, Essex: UK Data Archive [distributor], February 2012. SN: 6614, http://dx.doi.org/10.5255/UKDA-SN-6614-3

ESRC-UKDA-Data Cite Guide 'Data Citation: what you need to know' https://www.ukdataservice.ac.uk/media/104397/data_citation_online.pdf

What does a data citation look like?

Most Downloaded Items

Forstmann BU, Keuken MC, Schafer AS, Bazin P, Alkemade A, Turner R (2014) Data from: Multi-modal ultra-high resolution structural 7-Tesla MRI data repository. *Scientific Data* http://dx.doi.org/10.5061/dryad.fb41s

Number of Downloads

1170029

Data Citations

Abstract · **Background & Summary** · **Methods** · Data Records · **Usage Notes** · **Additional information** · **References** · Data Citations · **Acknowledgements** · **Author information**

1. Forstmann, B. U. *NITRC* http://www.nitrc.org/projects/atag_mri_scans/ (2014).
2. Forstmann, B. U. *Dryad* http://doi.org/10.5061/dryad.fb41s (2014).

Data Records

Abstract · **Background & Summary** · **Methods** · Data Records · **Usage Notes** · **Additional information** · **References** · Data Citations · **Acknowledgements** · **Author information**

All data records listed in this section are available from NITRIC (Data Citation 1: NITRC http://www.nitrc.org/projects/atag_mri_scans/) or Dryad (Data Citation 2: Dryad http://doi.org/10.5061/dryad.fb41s). A README file with a detailed description of the content of all downloads is available in Dryad. Additional material and information are also provided in Data Citation 1: NITRC http://www.nitrc.org/projects/atag_mri_scans/ and Data Citation 2: Dryad http://doi.org/10.5061/dryad.fb41s.
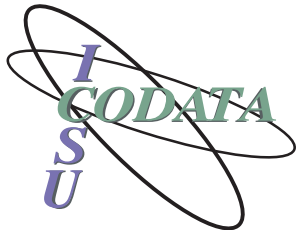
What does a data citation look like?

PANGAEA

Willmes, S et al. (2009): Onset dates of annual snowmelt on Antarctic sea ice in 2007/2008. doi:10.1594/PANGAEA.701380
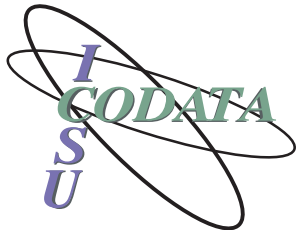
Dryad

Kingsolver JG, Hoekstra HE, Hoekstra JM, Berrigan D, Vignieri SN, Hill CE, Hoang A, Gibert P, Beerli P (2001) Data from: The strength of phenotypic selection in natural populations. Dryad Digital Repository. doi:10.5061/dryad.166

Dataverse

Frederico Girosi; Gary King, 2006, 'Cause of Death Data', http://hdl.handle.net/1902.1/UOVMCPSWOL UNF:3:9JU+SmVyHgwRhAKclQ85Cg== IQSS Dataverse Network [Distributor] V3 [Version].

Figure 2: Data citation formats suggested by repositories

DCC Guide 'How to Cite Datasets and Link to Publications': http://www.dcc.ac.uk/resources/how-guides/cite-datasets

Data Citation: From Principles to Practice

- CODATA Task Group on Data Citation 'Data Citation: From Principles to Practice, A Focus on the Research Policy and Funding Community':
http://www.codata.org/task-groups/data-citation-standards-and-practices

- Organising an international series of implementation and adoption workshops.

- **Promote the implementation of data citation principles in the research policy and funding communities throughout the world.**

- **Stakeholders** include: government, funders, research performing institutions, research administrators, research librarians, researchers, learned societies, publishers, data archives, journal editors …

    - What is the policy environment for data citation?

    - What are current attitudes to data citation?

    - What infrastructure currently exists to support data citation?

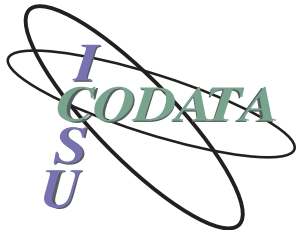   - What specific plans for implementation were identified?

DC¹
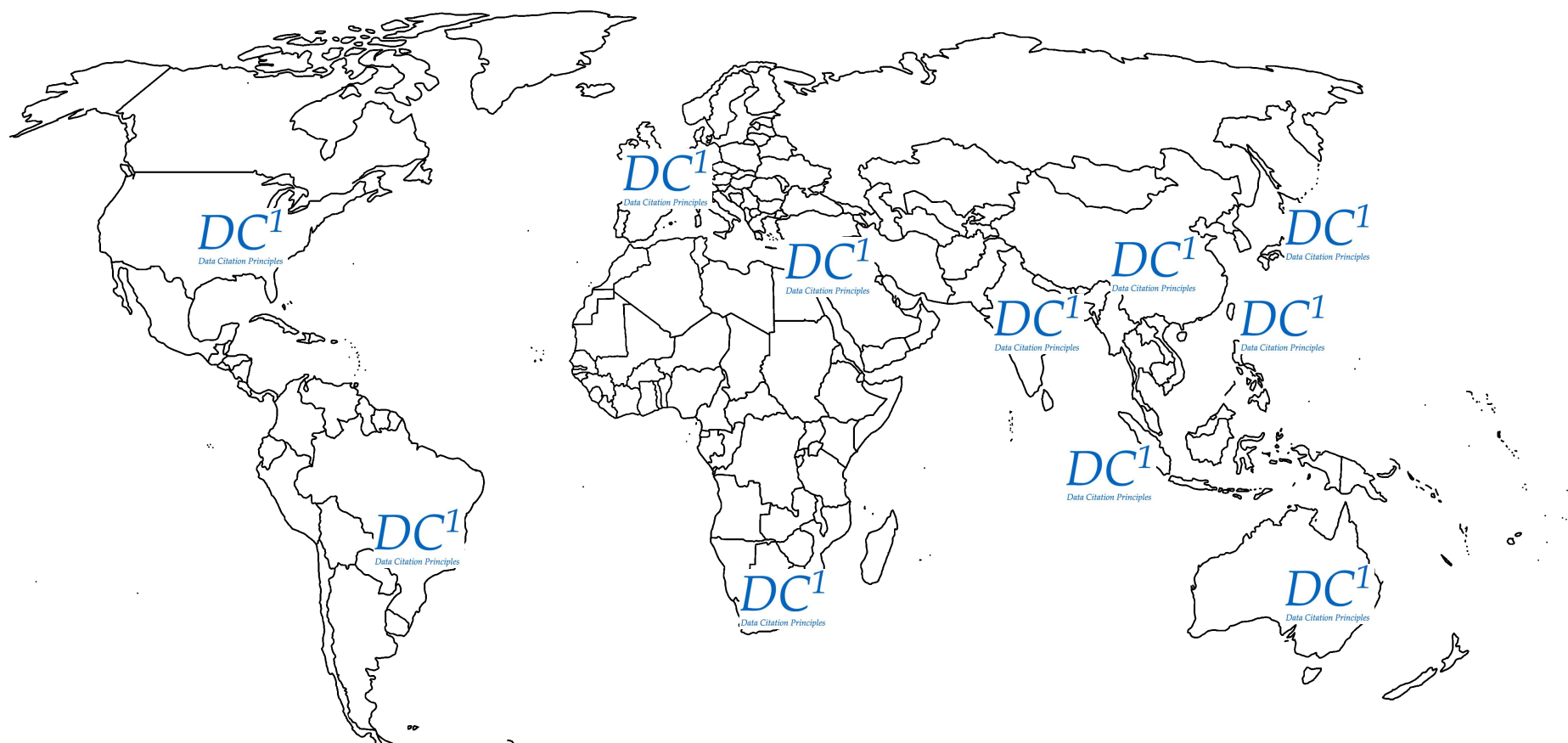
Data Citation Principles

We are taking Data Citation workshops on a world tour!

China… then Australia, Japan, India and South Africa. Plus: USA, Taipei, Korea, Indonesia, Brazil, EC, France, Israel…

Guides to Data Citation

- DCC Guide 'How to Cite Datasets and Link to Publications':
  http://www.dcc.ac.uk/resources/how-guides/cite-datasets

- ESRC-UKDA-Data Cite Guide 'Data Citation: what you need to know'
  https://www.ukdataservice.ac.uk/media/104397/data_citation_online.pdf

- BL Workshop Series on Data Citation:
  http://www.bl.uk/aboutus/stratpolprog/digi/datasets/workshoparchive/archive.html

- ANDS Series on Data Citation
  http://www.ands.org.au/training/data-citation.html

- RDA Data Citation of Evolving Data:
  https://rd-alliance.org/system/files/documents/RDA-DC-Recommendations_150924.pdf
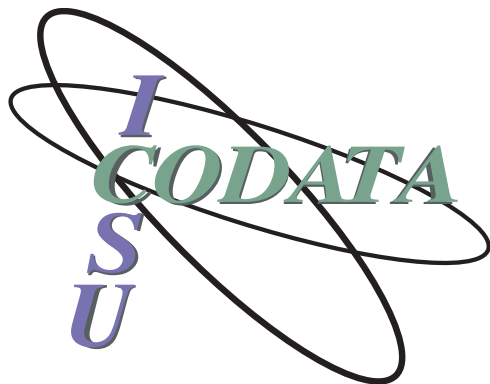
Thank you for your attention!

Slide credits: Louise Corti

Simon Hodson
Executive Director CODATA

www.codata.org
http://lists.codata.org/mailman/listinfo/codata-international_lists.codata.org
Email: simon@codata.org
Twitter: @simonhodson99
Tel (Office): +33 1 45 25 04 96 | Tel (Cell): +33 6 86 30 42 59

CODATA (ICSU Committee on Data for Science and Technology), 5 rue Auguste Vacquerie, 75016 Paris, FRANCE