Data citation in CSIRO
Building a culture of data citation

Anne Stevenson | Research Data Services Support
Presented by Cynthia Love, 29th January 2014

Image: http://xkcd.com/285/
The big picture

• Data citation is part of the bigger picture:
  • Data management
  • Citing sources
• Appeal to researchers to describe and manage their data, for
  • Accountability
  • Reproducibility
  • Reuse
  • Reputation
Carrots

• Recognition for managing data: get your data cited:

- Carrot: linking to data from publications can raise citation rates
  http://dx.doi.org/10.7717/peerj.175

- Example of cited data:
  [Image of data citation example]
In place

An automated process to:

• Deliver a DOI for a publically available dataset
  – a PID (a handle) for a “non-public” dataset

• Present an attribution statement within the collection:
  – Visible during the creation of the record in the Data Access Portal (DAP)
  – Any changes made to the record during the deposit process are reflected in the attribution
  – Can be copied for use in publications

• Bidirectional linking between data and publications
Attribution statement in the Data Access Portal

Data Citation

These fields will be the components of the Attribution Statement for this Data Collection in the following format:

<Lead Researcher>, <Contributors> (Published Year): <Collection Title as in Description tab>, <version>, <Publisher> <resource type>, <PID/DOI>

The Publisher will be CSIRO.
If there are more than 5 contributors "et al" will be shown in the Attribution Statement.

The DOI (Digital Object Identifier) will be generated when the data collection has been published if the Public can view metadata and data. A PID (Persistent Identifier) will be created for all data collections.

Citation Order

*Lead Researcher: 1 Stevenson, Anne (CSIRO IM&T, Newcastle)

Do you want CSIRO listed as Lead Researcher?

Contributors: 2 Jim Hindley Energy Technology

Publication Year: 2012

Bidirectional linking

Time-series for plankton and physio-chemistry of Logan’s Dam

Field of Research: Freshwater Ecology, Surface Water Hydrology, Water Resources Engineering

DOI: 10.4225/06/50F12E10535

Data Start Date: 21 Jul 2009

Data Collectors: McLean, David;Burn, Stewart

Related Materials: Urban Water Security Research Alliance

Water quality of Logan’s Dam as a function of ecosystem variables, climatic fluctuations and application of a monolayer

Date of Publication: 2012-08-19

Publication Type: Conference Material

Conference Name: UWSRA Science Forum, Brisbane, 19-20 June 2012

Papers: 59-63

Editor: Bogthe, D.L. and Wakim, S

Abstract: Time series for ecosystem variables and water quality of Logan’s Dam, water storage were obtained for 2009-2011. Sampling was designed to include one year prior to application of an evaporation-reducing monolayer and one year after. Continuous application of monolayer did not eventuate after the first year; however, short term trials were undertaken more
In place

• Training sessions for depositing and managing data sets in the Data Access Portal (DAP)
  – We can almost hear the attention level increase when we mention DOI minting
• Adding Google Analytics reports to wiki (shows most viewed, most downloaded)
• Providing altmetric data for inclusion in annual reports
• Distributing the ANDS Data Citation leaflet into CSIRO’s libraries, information centres and canteens
• Exploring tracking and metrics options such as ImpactStory and Thomson Reuters Data Citation Index
In place

• Contacting authors who have submitted journal articles:

Hello Yucang,

We have noticed that you have a manuscript recently approved in ePublish:
EP120237: Revisiting rolling and sliding in discrete element models

Are there datasets associated with this publication that you may need to:
• Archive for future discovery?
• Share internally?
• Deposit data to meet journal or funder requirements?
• Make publicly available, generating a DOI?

Please feel free to contact me or email the Research Data Service.

Anne Stevenson
Research Data Service | CSIRO Library Services
Information Management & Technology
CSIRO
E: anne.stevenson@csiro.au T: +61 2 49660687
PO Box 880, Newcastle, NSW 2300
www.rds.csiro.au

Publish your data
Tabulated data, images, video, audio — you name it. Archive your work, share it with specific colleagues, or open it to the wider research community.

Cite your collection
Create a Digital Object Identifier, making your data easier to cite and for others to find. Have your data ready and cited in your paper for people to download when your paper is published. Linking your data to your paper can improve your paper’s citation rate.

Easy to use

Easy to find
In place

• Tapping into the internal professional writing and proposal preparation sessions to encourage referral of data sources
• A dedicated intranet section (this is currently brief and under review)
• Using internal media channels such as Yammer and newsletters opportunistically
• Participating in ANDS’ Project Columbus
Project Columbus

The steps:
• Weekly report on new collections from the DAP.
• New collections with DOIs identified.
• We search our publications system for the data collection authors, looking for papers that aren’t published yet.
• We then contact the authors that match to talk about data citation generally, and Project Columbus specifically.

Results so far:
• Very early days. 4 candidates identified.
• Candidates are extremely keen, however.
• Additional datasets have been deposited by each candidate
But … resistance and reservations

“Do I publish my data then cite it in my journal article, or do I publish my journal article and then publish my data?”

“I wouldn’t release the data before I publish the related article, would I?”

Challenges

DOI management

• Researchers “get” DOIs; they understand their relevance & importance for publications – Natasha Simons
  • Object level, collection level, changes to the data, additions to the collection (eg time series)
  • Collaborative data & CSIRO data stored outside CSIRO

Cultural change

• Tracking citations is a great objective, however it’s important that researchers are citing the data they use, and they aren’t necessarily aware that they should – Steve McEachern
  • Pose the question in discussions with researchers: what happens if you can’t find your data? – Steve McEachern
Plans

Feed the beast!

• Include data collections/citations in internal assessments (Science reviews)
  – Build cultural change in research projects and themes (programs)
  – Leverage off funders’ requirements
• Maintain resources on impact and citation metrics
  – Tap into our Communications group
• Check the parent Library Services intranet: any content on citing publications can be expanded to include data
  – Tools to manage your data citations
• We will attempt to recruit “champions” as we are aware that researchers are more likely to hear this message from other researchers than from support staff
 Plans

• Communications
  – Link to publications AND data in media releases, news items
  – Use the Google Analytics reports to notify top-viewed, top-downloaded depositors
  – News items on data reuse

• Reviewing our DOI resolution to link to CSIRO data regardless of it’s location to ensure that ALL CSIRO data has the capability to be cited easily.
Further plans

It’s all about context

• Provenance and its relationship to citation
  – Work on integrating provenance into research and data workflow
• Software
  – It’s relationship to the data. How it too can improve data use and thus citation rate
  – Software citation in relation to data citation
• Linking up our systems
  – ePublish, DAP, Staff profiles, Research profiles
  – In context exposure – increase likelihood of reuse => citation
• Automating feeds to other portals
  – In context exposure – increase likelihood of reuse => citation
CSIRO Research Data Support Service

Anne Stevenson
Cynthia Love
David Benn
Dominic Hogan
Sue Cook

Plus roving data advocate: John Morrissey
Thank you

CSIRO IM&T
Anne Stevenson
Information Specialist

t +61 2 4960 6087
e anne.stevenson@csiro.au
w www.csiro.au