Data Citation in the Wild

Valerie Enriquez[1,2], Sarah Walker Judson[1,3], Nicholas M. Weber[1,4]

and mentors Suzie Allard[1,5], Robert B. Cook[1,6], Heather A. Piwowar[1,7], Robert J. Sandusky[1,8], Todd J. Vision[1,7,9], Bruce Wilson[1,5,6]

Data attribution is important

- Rewards those who contribute
- Allows measurement of reuse to demonstrate value and evaluate policy

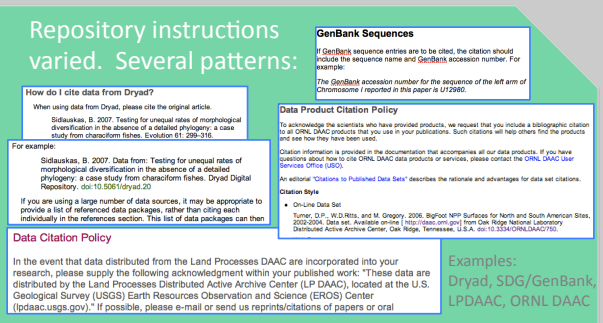
How is data reuse currently attributed?

We investigated the **policies, practices,** and **implications** of data attribution practices in the **environmental sciences.**

Policies

We reviewed policies of various stakeholders within the environmental sciences, looking for ***evidence of data citation policies*** and ***attribution suggestions***.

Repository instructions varied. Several patterns:



Examples: Dryad, SDG/GenBank, LPDAAC, ORNL DAAC

Policies in our sample that mention data citation

Repositories	8 of 26,	**31%**
Journals	16 of 307,	**6%**
Funders	1 of 52,	**2%**

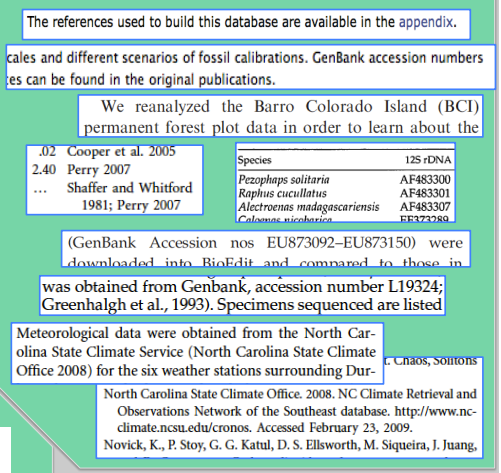
Journal policy content also varied:

- how to hyperlink accession numbers,
- which DBs are allowed in reference lists
- "If you use data, [..] acknowledge fully"

Practices

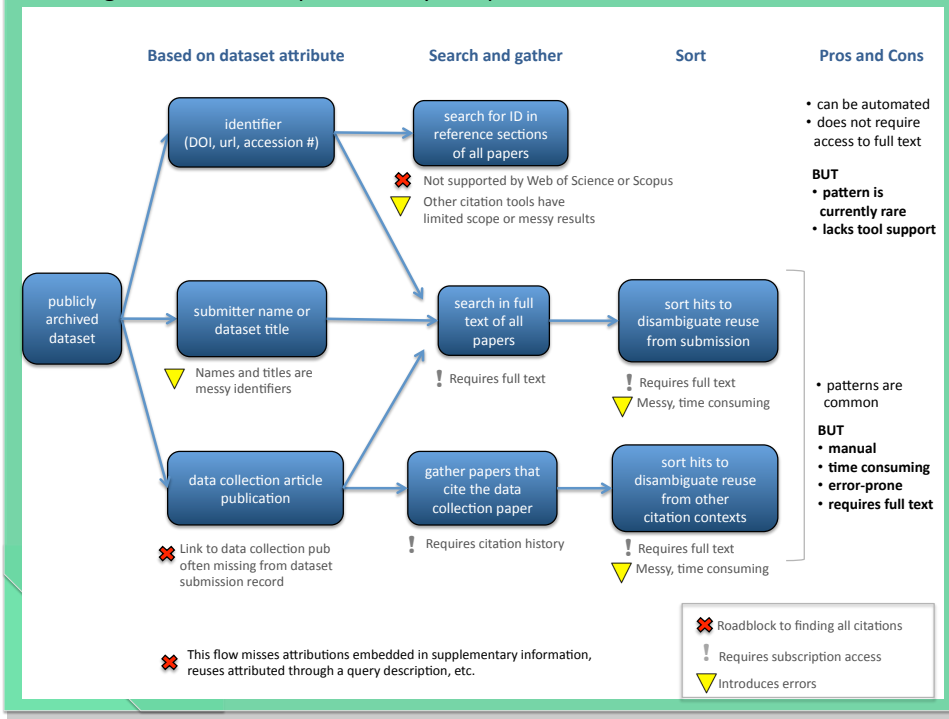
We reviewed 500 articles in six leading environmental science, ecology and evolutionary biology journals, 2000-2010. **Data attribution patterns** varied widely in the main body of articles (n=221):

Mentioned author ref	Mentioned repository	Mentioned dataset ID	Pattern frequency
			17%
	✓		**22%**
✓			**36%**
✓	✓		**12%**
	✓	✓	**7%**
✓	✓	✓	**5%**
53%	**47%**	**13%**	



Implications

Because of **tool limitations** and these **diverse data attribution practices**, tracking dataset reuse is prohibitively complicated:



We found that data citation best practices are rarely mentioned, and attribution practice is very diverse.

As a result of diverse practices and tool limitations, data citations are currently very difficult to track.

Call to action:

1. **Issue unique dataset identifiers**
2. **Adopt data citation policies**
3. **Enforce data citation policies**
4. **Add support for data citation tracking to existing tools**
5. **Demonstrate the win**

Acknowledgements

This work was conducted as part of **the DataONE Summer 2010 internship program.**
Funded by INTEROP: Creation of an International Virtual Data Center for the Biodiversity, Ecological and Environmental Sciences, NSF grant #0753138,
and the Data Observation Network for Earth (DataONE) NSF cooperative agreement #0830944.

Affiliations

1: DataONE; 2: Simmons College; 3: Brigham Young University;
4: University of Illinois Champaign-Urbana; 5: University of Tennessee, Knoxville;
6: Oak Ridge National Laboratory; 7: National Evolutionary Synthesis Center;
8: University of Illinois at Chicago; 9: University of North Carolina at Chapel Hill

This research was done as open science: ask us about it!
http://openwetware.org/wiki/DataONE:Notebook/Summer_2010

For more information: hpiwowar@nescent.org

CC BY