Publishers’ Policies for Data Citation: Do they ease data discovery and use?

Christine Malinowski  
*Massachusetts Institute of Technology*

Chris Sherratt  
*Massachusetts Institute of Technology*

Follow this and additional works at: [https://escholarship.umassmed.edu/escience_symposium](https://escholarship.umassmed.edu/escience_symposium)

Part of the Scholarly Communication Commons

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).

**Repository Citation**

Malinowski, C., & Sherratt, C. (2016). Publishers’ Policies for Data Citation: Do they ease data discovery and use?. *University of Massachusetts and New England Area Librarian e-Science Symposium*.  

Creative Commons License

This work is licensed under a [Creative Commons Attribution-Noncommercial-Share Alike 4.0 License](https://creativecommons.org/licenses/by-nc-sa/4.0/).  
This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in University of Massachusetts and New England Area Librarian e-Science Symposium by an authorized administrator of eScholarship@UMMS. For more information, please contact Lisa.Palmer@umassmed.edu.
Publishers’ Policies for Data Citation: Do they ease data discovery and use?

Objective

Publisher policies have long guided researchers on how to cite publications, but now many publishers have adopted policies for data citation and sharing. This project examines the data sharing and citation practices of MIT authors in atmospheric sciences, a field that has seen a recent rise of publisher data policies. Through a multi-pronged approach, we sought to understand how research groups in this discipline find and cite data used in their research, share their own produced research data, and what variables (specifically funder and publisher mandates) may alter or influence this behavior.

Methods

We identified 7 atmospheric sciences research groups to include in this study. For each of these groups, we approached our research question via the following methods:

- **Bibliometric Analysis.** Research articles published by each of the research groups between January 2011 and November 2015 were included in the study, totaling 231 publications in our dataset. This time period was selected to encapsulate pre- and post-implementation of major publisher data policies.

- **Group Interviews.** We aimed to systematically review each article, utilizing a Google Form, identifying source and produced data types, their acknowledgement/citation within the paper and shared data availability.

- We expanded this interview format to include members of the research groups beyond the faculty PI, acknowledging that the groundwork for data management and sharing may be carried out by additional members of the research group.

- We are still in the process of completing both of these activities. This poster presents our work completed thus far.

Current Results

36.6% (84) of the articles have been reviewed

0 of these articles provide information on how to find their ‘underlying’ data

Interview Takeaways

- Personal contact (asking individuals directly) is the major mechanism by which this community shares data.
- There is a desire to communicate directly with those who want to use the data to help them interpret it correctly.
- Is it the underlying data that is useful or is it the models that produce this data?

Current Results

<table>
<thead>
<tr>
<th>Publisher (top 5 in study set)</th>
<th>Publications in Study Set</th>
</tr>
</thead>
<tbody>
<tr>
<td>American Geophysical Union (AGU/Wiley)</td>
<td>66 (28.6%)</td>
</tr>
<tr>
<td>European Geosciences Union (EGU)</td>
<td>53 (22.9%)</td>
</tr>
<tr>
<td>American Meteorological Society (AMS)</td>
<td>31 (13.4%)</td>
</tr>
<tr>
<td>American Chemical Society (ACS)</td>
<td>17 (7.4%)</td>
</tr>
<tr>
<td>Nature</td>
<td>16 (6.9%)</td>
</tr>
</tbody>
</table>

These top 5 publishers account for ~79% of the total (231) publications in the study set.

Current Results

0 of these articles provide information on how to find their ‘underlying’ data

- There’s a reason why data is worked up and written into a paper. If we just shared the data, why would I write it up and put it into a paper?
- “Maybe 1 out of 100 papers I read [has a citation to a data set].”
- “I do think it’s good that they [publishers] are pushing in that direction [requiring sharing data]. You can misinterpret your own data too...extra eyes can always help with that.”

Current Results

0 of these articles provide information on how to find their ‘underlying’ data

- “There’s a reason why data is worked up and written into a paper. If we just shared the data, why would I write it up and put it into a paper?”
- “Maybe 1 out of 100 papers I read [has a citation to a data set].”
- “I do think it’s good that they [publishers] are pushing in that direction [requiring sharing data]. You can misinterpret your own data too...extra eyes can always help with that.”

Challenges

- Defining our source & produced data types to provide consistent coding for more in-depth analysis
- Securing time with research groups to conduct interviews

Preliminary Conclusions

Preliminary results show that while researchers are not yet consistently providing persistent identifiers to datasets or making full data sets publicly available, they perceive their data sharing efforts to be in line with the expectations and needs of their community. Such inconsistencies in publisher language and researcher behavior highlight the need to further explore definitions of data sharing, underlying data, etc., across stakeholders.

It is very possible that as the publisher data policies mature, this research community will begin to adjust the mechanisms and documentation surrounding their obtained and derived data, but it may be too early yet to observe such impacts of these policies.

Next Steps

- Finish our coding of the remaining research articles
- Finish the interviews of the remaining research groups
- Examine articles’ approaches to data sharing more directly with their publishers’ policies, data types used, etc.

Christine Malinowski, Research Data Librarian, MIT Libraries
Chris Sherratt, Atmospheric and Oceanic Sciences Librarian, MIT Libraries