

Repository Profiles for Atmospheric and Climate Sciences: Capabilities and Trends in Data Services

Chung-Yi Hou (hou@illinois.edu)¹, Cheryl A. Thompson¹, Carole L. Palmer²

1 - Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign; 2 – Information School, University of Washington



Introduction

As digital research data proliferate and expectations for open access escalate, the landscape of data repositories is becoming more complex. For example, DataBib currently identifies 980 data repositories across the disciplines, with 117 categorized under Geosciences. In atmospheric and climate sciences, there are great expectations for the integration and reuse of data for advancing science. To realize this potential, resources are needed that explicate the range of repository options available for locating and depositing open data, their conditions of access and use, and the services and tools they provide.

This study profiled 38 open digital repositories in the atmospheric and climate sciences, analyzing each on 55 criteria through content analysis of their websites. The results provide a systematic way to assess and compare capabilities, services, and institutional characteristics, and identify trends across repositories.

Research Objectives

- Profile capabilities and services of open digital repositories in atmospheric and climate sciences.
- Determine patterns and trends across repositories.
- Assess availability of information provided by repositories.

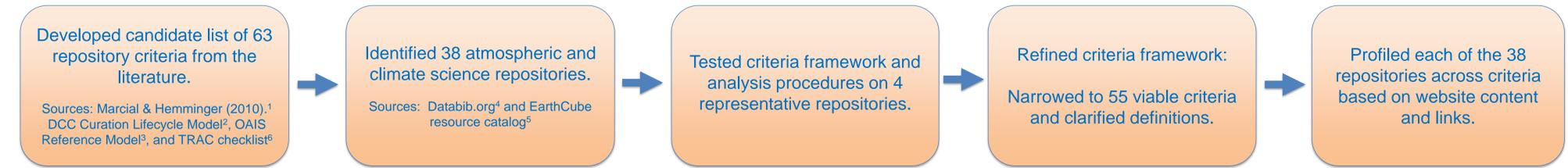
References

1. Marcial, L. H. & Hemminger, B. M. (2010). Scientific data repositories on the web: An initial survey. *Journal of the American Society for Information Science and Technology*, 61(10), 2029-2048.
2. Digital Curation Centre. (2014). *DCC Curation Lifecycle Model*. Retrieved from <http://www.dcc.ac.uk/resources/curation-lifecycle-model>
3. Consultative Committee for Space Data Systems (CCSDS). (2012, June). *Reference model for an open archival information system (OAIS)*. *Magenta Book, Issue 2*. Retrieved from <http://public.ccsds.org/publications/archive/650x0m2.pdf>
4. Databib. (n. d.). *Research Data Repositories*. Retrieved from http://databib.org/index_subjects.php# Geo
5. EarthCube Cross-Domain Interoperability Group. (2014). *EarthCube Resource Readiness Catalog*. Retrieved from <http://connections.earthcube.org/EcResources/Organizations>
6. Center for Research Libraries, and Online Computer Library Center (OCLC). (2007). *Trusted Repositories Audit & Certification: Criteria & Checklist Version 1.0*. Retrieved from: http://www.crl.edu/sites/default/files/attachments/pages/trac_0.pdf

Acknowledgments

This study was supported by the Data Curation Education in Research Centers (DCERC) project, funded by the Institute of Museum and Library Services (RE-02-10-0004-10).

Method



Results (n = 38 repositories) - 31 of 55 analysis criteria shown, including key elements in the DCC Curation lifecycle model

| Percent (n) | Metadata & Description |
|-------------|--|
| 94.7% (36) | Expect data providers to contribute descriptive metadata at the time of submission. |
| 42.1% (16) | Use own institutional standard. |
| 31.6% (12) | Provide metadata services or best practices/guidelines. |
| 21.1% (8) | Identify and follow specific metadata standards (e.g., ISO 19115, ISO 19139, FGDC, and Dublin Core). |
| 5.3% (2) | Invite community members to provide and improve metadata. |

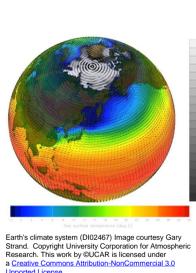
Consistent metadata improves users' ability to identify, evaluate, and reuse data of interest.

| Preservation Procedures and Policies | |
|--------------------------------------|--|
| • 15.8% (6) | Data appraisal and selection procedures. |
| • 15.8% (6) | Plans for damage or loss from environmental factors and disasters. |
| • 13.2% (5) | Explicit preservation policy. |
| • 2.6% (1) | Long-term sustainability plan. |
| Services | |
| • 97.4% (37) | Provide information on services and support functions. |
| • 68.4% (26) | Active in outreach and education, presentations, seminars, and customer service. |
| • 47.4% (18) | Offer some level of data processing. |
| • 44.7% (17) | Provide software and tools. |

Availability of preservation information enhances organizational transparency and builds user confidence.

| Percent (n) | Data Citation & Identifiers |
|-------------|---|
| 78.9% (30) | Request users cite and/or acknowledge datasets and the data center. |
| 52.6% (20) | Use a data identification system. |
| 18.4% (7) | Apply data identifiers consistently. |
| Percent (n) | Standards & Best Practices |
| 10.5% (4) | Evidence of compliance with certification or accreditation standard (e.g., Information Fair Trade Scheme, ISO 9001, and ISO 14001). |

Data citation promotes data sharing. Repositories can encourage citation best practices, in step with ongoing work on next generation data citation methods.



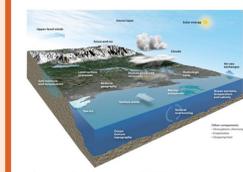
Earth's climate system (D02467). Image courtesy Gary Strand. Copyright University Corporation for Atmospheric Research. This work by IOLCAR is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

| Data submission | |
|-----------------|---|
| • 71.1% (27) | Indicate data types accepted (e.g., instruments, research topics, or science discipline). |
| • 60.5% (23) | Acquire data through deposit by community members. |
| • 31.6% (12) | Require data providers to contact repository for data submissions and address transfer of rights. |
| • 21.1% (8) | Receive data directly from instruments. |
| • 18.4% (7) | Accept physical forms (e.g., microfilms, CDs, and hard drives). |

Many repositories provide detailed information on data types accepted and submission processes.

| Access & Use | |
|--------------|---|
| • 65.8% (25) | Utilize distributed access model (vs. centralized access). |
| • 42.1% (16) | Allow access without registration or authorization. |
| • 23.7% (9) | May require fees under certain access and/or use conditions. |
| • 10.5% (4) | Allow embargo periods and address processes for data with confidentiality concerns. |
| • 7.9% (3) | Indicate when data are free from copyright. |
| • 2.6% (1) | Require authentication in person. |

While repositories continue to work towards open data, user registration and embargo periods increase user assurance and control of data they submit.



Community Earth System Model (CESM). Copyright University Corporation for Atmospheric Research. This work by IOLCAR is licensed under a Creative Commons Attribution-NonCommercial 3.0 Unported License.

GRADUATE SCHOOL OF LIBRARY AND INFORMATION SCIENCE
The iSchool at Illinois

Information School
UNIVERSITY of WASHINGTON

Limitations

The profiles were based exclusively on information made accessible through repository websites. More comprehensive coverage and validation could be achieved by engaging repository personnel in review and elaboration of the profiles.

Conclusion

- Most repositories in the atmospheric and climate sciences provide a significant amount of information to facilitate data sharing and access.
- Most repositories provide clear information on the scope of data accepted and citation expectations; other services and support functions are less widely developed.
- Aspects that are particularly important for building trust, such as certification and confidentiality assurance, are limited across the sample.
- Guidelines and policies on metadata and transfer of rights are uneven, and metadata services appear underdeveloped.