# Databib: Guidelines for Bibliographers

draft April 3, 2012

Wesbite:  http://databib.lib.purdue

Contact:  databib@gmail.com

The purpose of Databib is to create, share, and enhance bibliographic records that describe research data repositories.  Anyone is able to submit and enhance records, which will be reviewed and then committed by a volunteer editorial board.

What constitute "data" and "repositories" are open to interpretation and may vary from domain to domain.  This rubric gives some guidance to use to identify repositories to be included in Databib and some directions on creating consistent metadata.

We can begin to characterize research data as a primary inputs or outputs from a research project.  Examples of research datasets may include spreadsheets, transcripts, images, sensor and instrument output, gene sequences, survey responses, software source code, and video recordings.  In a general sense, anything could be considered or treated as data. For Databib, we are interested in data that support reproducibility or reuse in the context of research.

The term "repository" can also be slippery.  What is the difference between a data repository and an online database, a collection of data on the web, a data portal, or a set of analytical tools that have been packaged with data?

Our current, suggested practice is to look at the records that have already been created in Databib for comparison.  Some characteristics of a "data repository" may include:

- A system that provides online access to research data

- By an entity that is committed to providing access to the data

- That serves as a primary, authoritative source for data

- With a significant collection of like data or data within the scope of a management policy

This definition is also not fixed, and interpretations will vary from domain to domain.  An editorial board with representation from respective domains will help bibliographers interpret and apply this rubric.

As more repositories are identified and described in Databib, this rubric will be revisited and revised.

# Databib Metadata Record Schema

| Record fields | *denotes required field, leave unknowns blank. |
|---|---|
| Title of Repository* | Identify from title from homepage – follow title by acronyms of repository in parenthesis, authority acronyms in square brackets; Move "The" to end. |
| URL* | Of repository front page. |
| Authority* | Organization responsible for the repository. |
| Subjects (LCSH)* | Terminology within repository can be helpful for establishing subject headings, balance of broad coverage and specific topics if scope of repository is narrow. |
| Description | 2-3 _original_ concise sentences that summarize the contents of the repository and its designated community; include keywords, do not repeat terms from other metadata. The description should be interpretive and not be copied and pasted. Avoid adjectives and information about the organization (focus on what data the user will find there and who is served by the repository). |
| Access | Open/closed/licensed<br>• Open-Access by anyone anywhere at any time<br>• Closed-Access by registered users only<br>• Licensed-subscription database |
| Start date | Four-character year indicated earliest instance of repository. |
| Location | Country of responsible organization. Use of recognized abbreviation allowed, or fully name of country in capital letters. (Ex. USA or GERMANY) |
| Reuse | How users can reuse datasets (typically open or closed). Link to should be included here when available (e.g., Open-http://siox.sdsc.edu/usage.php). |
| Deposit | Who can deposit data. Appropriate terminology:<br>• Accepting data for deposit (Private submissions welcome)<br>• Credentialed individuals in…(*appropriate field*)<br>• Membership in (*specified*) organizations<br>• Closed (Responsible organization only, no outside submissions)<br>Include link to any pertinent information regarding data deposit |
| Repository type | • Institutional (hosted by a specific university or research program)<br>• Governmental (primarily supported by a federal agency)<br>• Disciplinary (data specific to an area of research or specialty)<br>• Commercial (private corporation)<br>Hybrids are acceptable (e.g., governmental/disciplinary) |