Managing & Sharing Your Research Data

Kate Anderson
MU Libraries
10/10/14

https://www.flickr.com/photos/intersectionconsulting/7537238368/in/set-72157614274686504/

Today's Session

- Why we're here…
- Data Lifecycle
- Data Management Plans
- MOspace & Data Repositories
- ORCID
- More resources

Caveats

- We’re here for 1 hour today
- Lots of different types of “data”
- Differences by discipline
- We’ll focus more on final data (rather than raw or intermediate data)
- It takes a village: DoIT; OSPA; Cyberinfrastructure Council; MU Libraries

Why we're here…

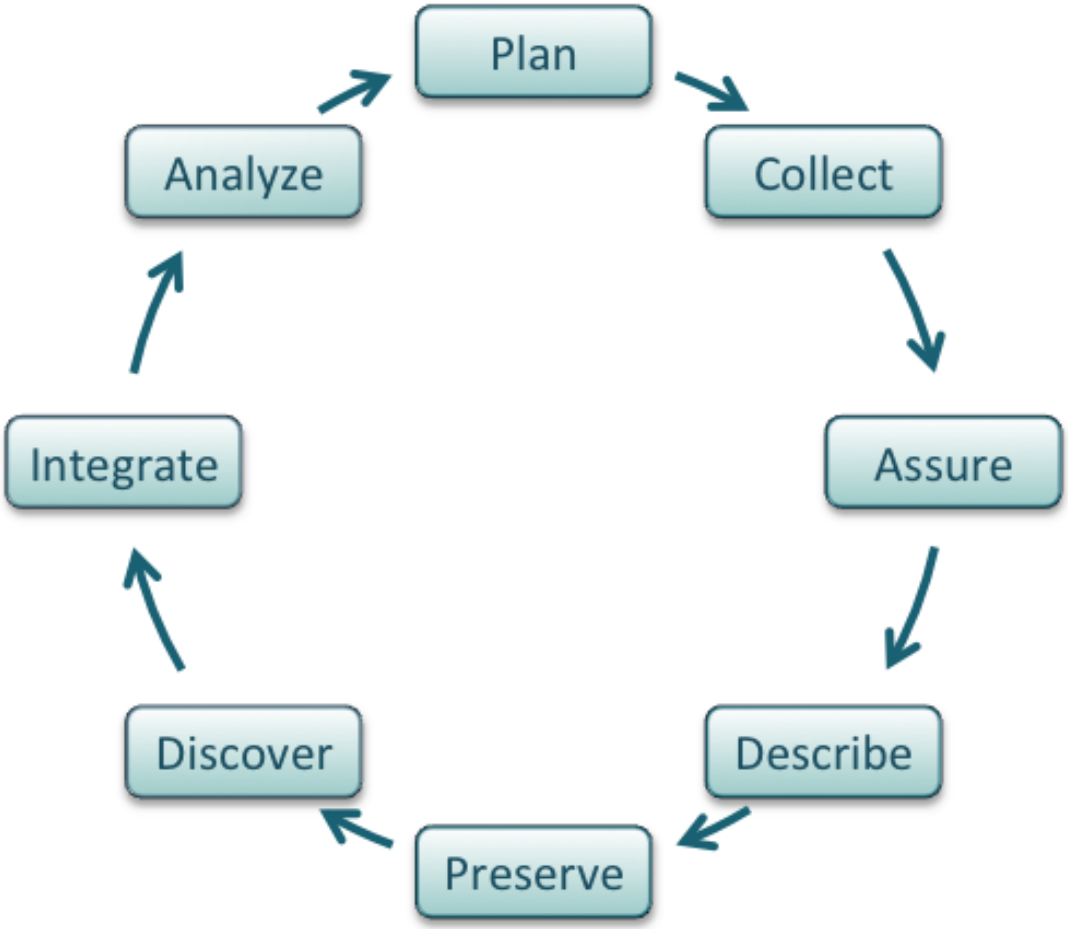
"The goal of data management is to produce self-describing data sets" (DataONE Primer)

- Data are important!
- Benefits you & your collaborators
- Benefits science & inquiry
- Funding Agencies and Journals Require Sharing of Data

Let's avoid this…

Data Sharing and Management SNAFU in 3 Short Acts
By Karen Hansen, Alisa Surkis & Karen Yacobucci
NYU Health Sciences Libraries
https://www.youtube.com/watch?v=66oNv_DJuPc

The Circle of Life…



(DataONE Primer)

The Circle of Life…

(DataONE Primer)

The White House

Office of the Press Secretary

E-Mail Tweet Share +

For Immediate Release May 09, 2013

Executive Order -- Making Open and Machine Readable the New Default for Government Information

EXECUTIVE ORDER

MAKING OPEN AND MACHINE READABLE THE NEW DEFAULT
FOR GOVERNMENT INFORMATION

By the authority vested in me as President by the Constitution and the laws of the United States of America, it is hereby ordered as follows:

Section 1. General Principles. Openness in government strengthens our democracy, promotes the delivery of efficient and effective services to the public, and contributes to economic growth. As one vital benefit of open government, making information resources easy to find, accessible, and usable can fuel entrepreneurship, innovation, and scientific discovery that improves Americans' lives and contributes significantly to job creation.

Decades ago, the U.S. Government made both weather data and the Global Positioning System freely available. Since that time, American entrepreneurs and innovators have utilized these resources to create navigation systems, weather newscasts and warning systems, location-based applications, precision farming tools, and much more, improving Americans' lives in countless ways and leading to economic growth and job creation. In recent years, thousands of Government data resources across fields such as health and medicine, education, energy, public safety, global development, and finance have been posted in machine-readable form for free public use on Data.gov. Entrepreneurs and innovators have continued to develop a vast range of useful new products and businesses using these public information resources, creating good jobs in the process.

National Institutes of Health
Office of Extramural Research

http://grants.nih.gov/grants/policy/data_sharing/

Data sharing is essential for expedited translation of research results into knowledge, products and procedures to improve human health.

American Heart Association.
life is why™

The AHA requires grant applicants to include a data sharing plan as part of the application process. Any research data that is needed for independent verification of research results must be made freely and publically available in **an AHA approved repository** within 12 months of the end of the funding period (and any no-cost extension).

U.S. DEPARTMENT OF ENERGY | Office of Science

To integrate data management planning into the overall research plan, **all proposals submitted to the Office of Science for research funding are required to include a Data Management Plan (DMP)** of no more than two pages that describes how data generated through the course of the proposed research will be shared and preserved or explains why data sharing and/or preservation are not possible or scientifically appropriate. At a minimum, DMPs must describe how data sharing and preservation will enable validation of results, or how results could be validated if data are not shared or preserved.

National Science Foundation
WHERE DISCOVERIES BEGIN

- 2-page DMPs required
- (http://www.nsf.gov/bfa/dias/policy/dmp.jsp)
  - The **types of data**, samples, physical collections, software, curricular materials, and other materials to be produced in the course of the project;
  - The **standards** to be used for data and metadata format and content (where existing standards are absent or deemed inadequate, this should be documented along with any proposed solutions or remedies);
  - Policies for **access and sharing**, including provisions for appropriate protection of privacy, confidentiality, security, intellectual property, or other rights or requirements;
  - Policies and provisions for **re-use, re-distribution, and the production of derivatives**; and
  - Plans for **archiving data**, samples, and other research products, and for **preservation of access** to them.
- NEH Office of Digital Humanities aligns with NSF

NATIONAL ENDOWMENT FOR THE
Humanities

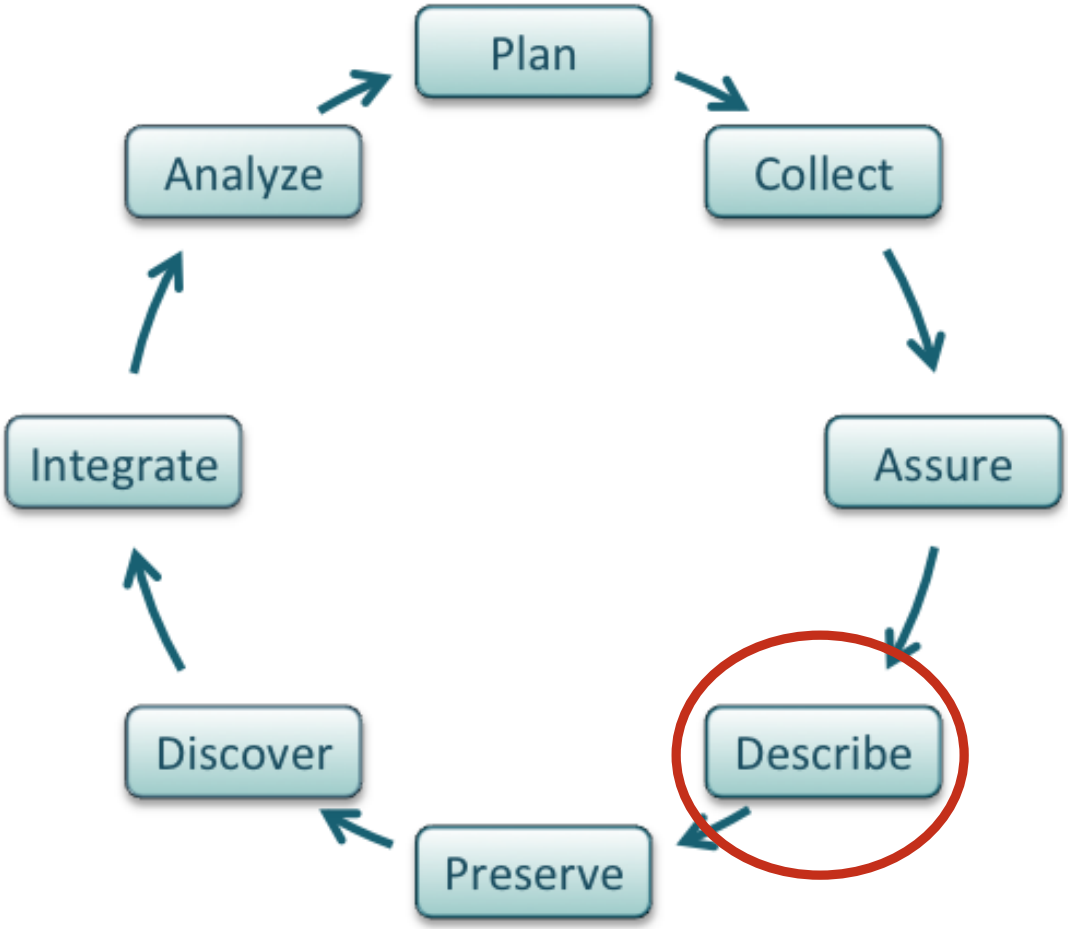
http://www.plosone.org/static/policies#sharing

When submitting a manuscript online, authors must provide a *Data Availability Statement* describing compliance with PLOS's policy. If the article is accepted for publication, the data availability statement will be published as part of the final article. **Refusal to share data and related metadata and methods in accordance with this policy will be grounds for rejection**.

http://datadryad.org/

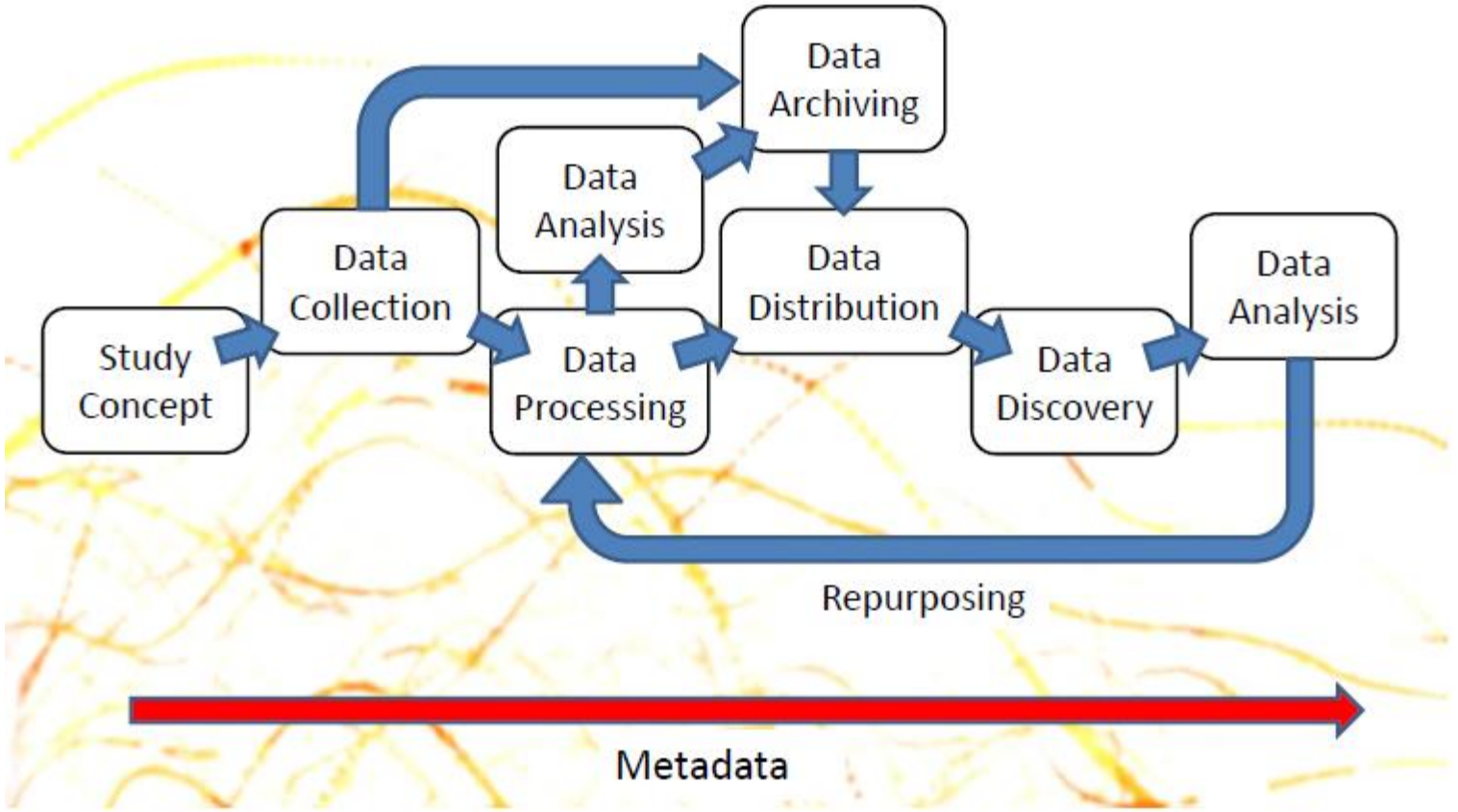
DataDryad.org is a curated general-purpose repository that makes the data underlying scientific publications discoverable, freely reusable, and citable. Dryad has integrated data submission for a growing list of journals; submission of data from other publications is also welcome.

The Circle of Life…



(DataONE Primer)

Metadata/annotation must be added throughout the lifecycle



University of Florida Data Lifecycle Management: http://ufdc.ufl.edu/IR00000801/00001

Metadata Basics

- Data about data
- Metadata lets others discover, understand, and use your data
- Basic Dublin Core Example: http://datadryad.org/resource/doi:10.5061/dryad.s2878/2?show=full

Metadata across the disciplines

Basic information to keep:

- Descriptive
  - What is it about?
  - Title, time, author, keywords
  - Relations to other data objects
- Administrative
  - Ownership and use permissions
- Provenance
  - Where does it come from?
  - History of changes to the data, versions

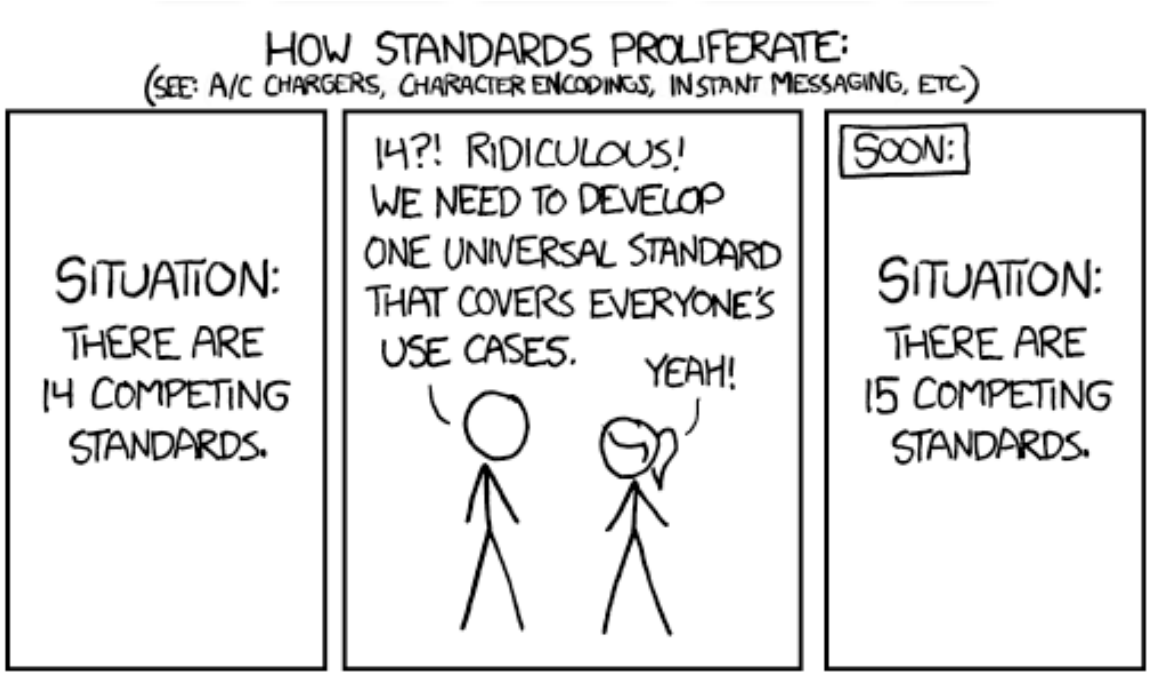
More specific information varies by discipline

University of Florida Data Lifecycle Management:
http://ufdc.ufl.edu/IR00000801/00001

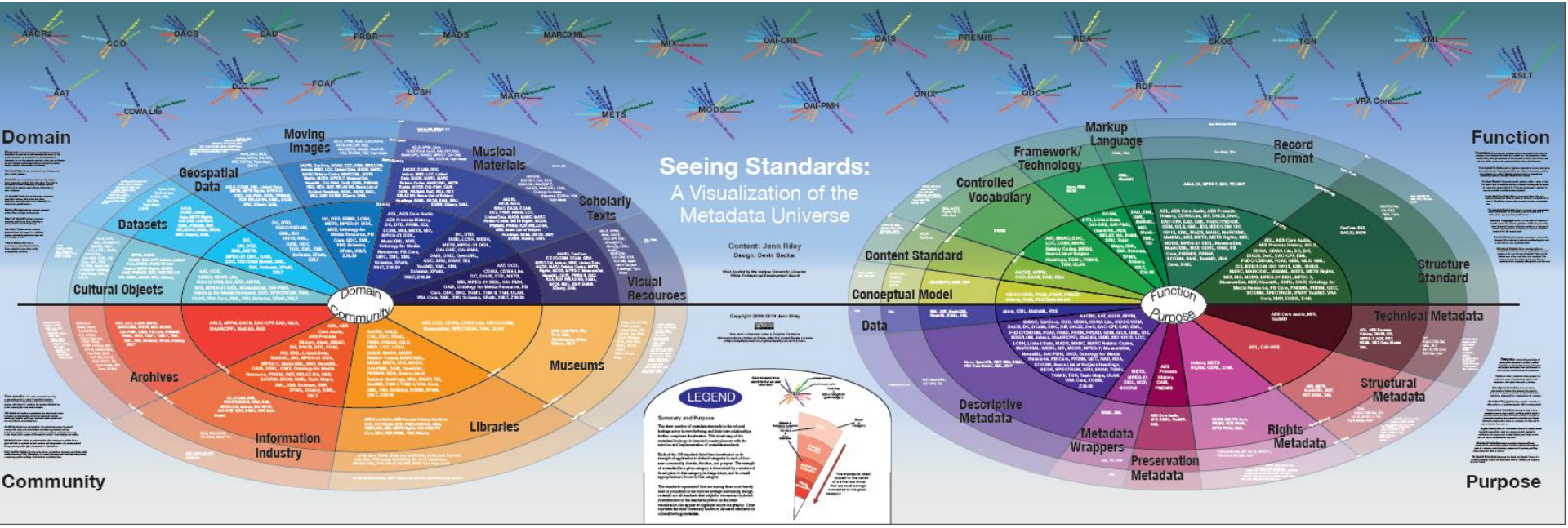
Metadata to Consider:
Who, What, Where, When, Why, How

- Name of the data set and data files
- Date of creation and last modification
- Software used to create file (including version)
- Data processing performed
- Who collected the data
- Contact information of responsible party
- Sponsor or funding agencies
- Why the data were collected (abstract; keywords; controlled vocabulary); when and where
- Instrumentation; experimental conditions; calibrations
- Units of measure
- Taxonomic details
- Known problems that limit data use
- **How to cite the data set**

TIP: Use non-proprietary formats (e.g., .txt not .docx)

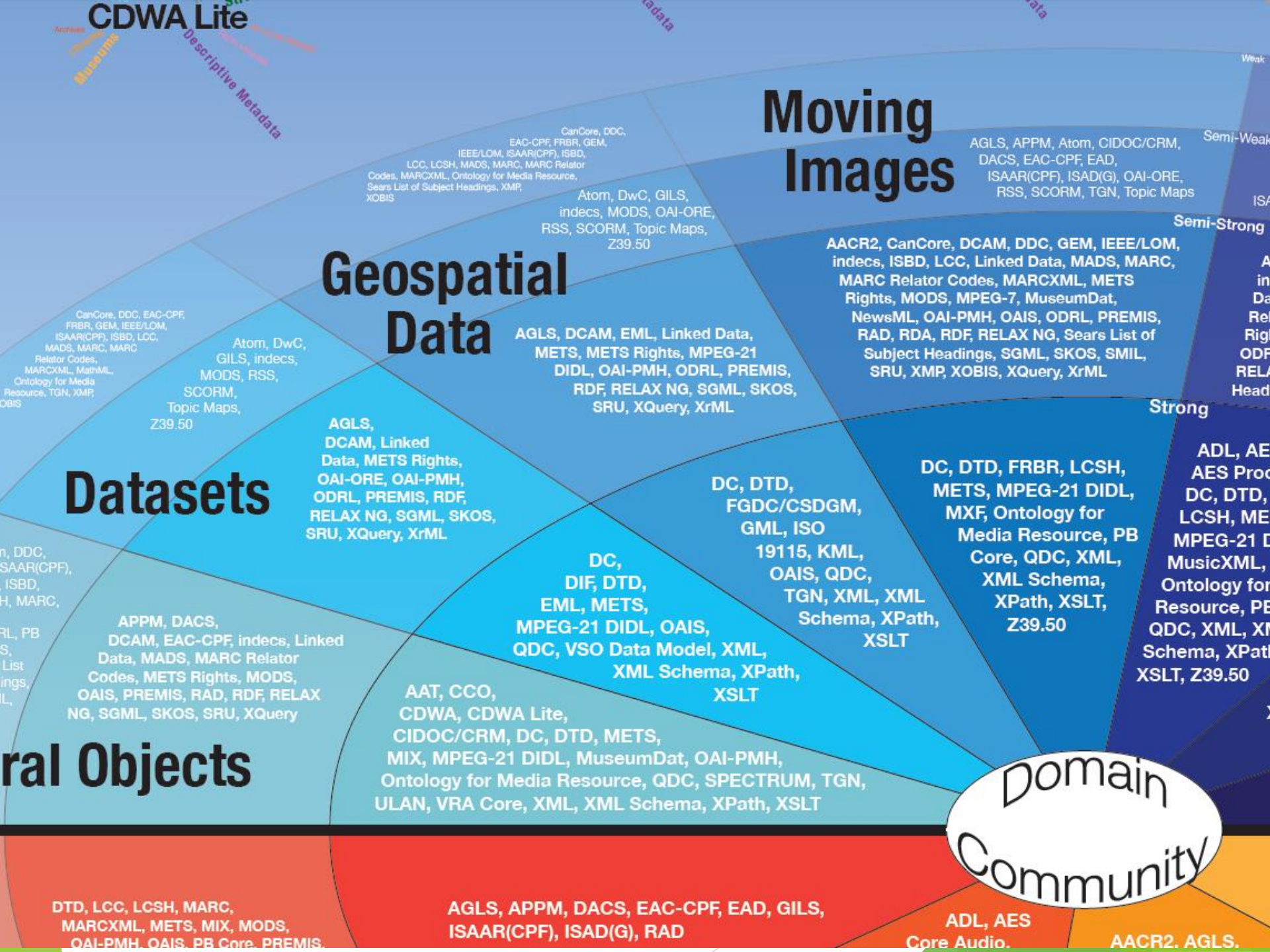


http://imgs.xkcd.com/comics/standards.png

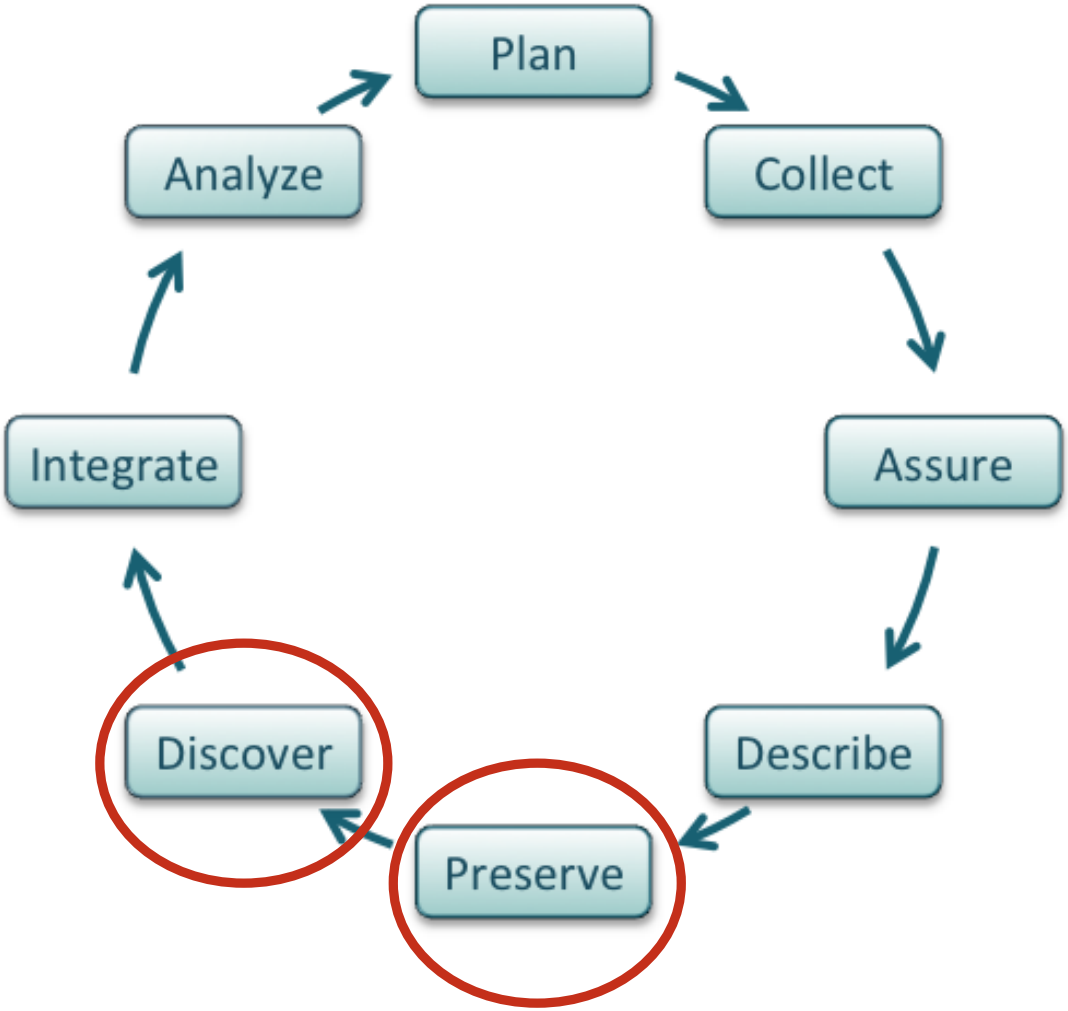


Seeing Standards: A Visualization of the Metadata University

http://www.dlib.indiana.edu/~jenlrile/metadatamap/


CDWA Lite
Descriptive Metadata
Moving Images
AGLS, APPM, Atom, CIDOC/CRM, DACS, EAC-CPF, EAD, ISAAR(CPF), ISAD(G), OAI-ORE, RSS, SCORM, TGN, Topic Maps
CanCore, DDC, EAC-CPF, FRBR, GEM, IEEE/LOM, ISAAR(CPF), ISBD, LCC, LCSH, MADS, MARC, MARC Relator Codes, MARCXML, Ontology for Media Resource, Sears List of Subject Headings, XMP, XOBIS
Atom, DwC, GILS, indecs, MODS, OAI-ORE, RSS, SCORM, Topic Maps, Z39.50
Geospatial Data
AACR2, CanCore, DCAM, DDC, GEM, IEEE/LOM, indecs, ISBD, LCC, Linked Data, MADS, MARC, MARC Relator Codes, MARCXML, METS Rights, MODS, MPEG-7, MuseumDat, NewsML, OAI-PMH, OAIS, ODRL, PREMIS, RAD, RDA, RDF, RELAX NG, Sears List of Subject Headings, SGML, SKOS, SMIL, SRU, XMP, XOBIS, XQuery, XrML
AGLS, DCAM, EML, Linked Data, METS, METS Rights, MPEG-21 DIDL, OAI-PMH, ODRL, PREMIS, RDF, RELAX NG, SGML, SKOS, SRU, XQuery, XrML
CanCore, DDC, EAC-CPF, FRBR, GEM, IEEE/LOM, ISAAR(CPF), ISBD, LCC, MADS, MARC, MARC Relator Codes, MARCXML, MathML, Ontology for Media Resource, TGN, XMP
Atom, DwC, GILS, indecs, MODS, RSS, SCORM, Topic Maps, Z39.50
Datasets
AGLS, DCAM, Linked Data, METS Rights, OAI-ORE, OAI-PMH, ODRL, PREMIS, RDF, RELAX NG, SGML, SKOS, SRU, XQuery, XrML
DC, DTD, FGDC/CSDGM, GML, ISO 19115, KML, OAIS, QDC, TGN, XML, XML Schema, XPath, XSLT
DC, DTD, FRBR, LCSH, METS, MPEG-21 DIDL, MXF, Ontology for Media Resource, PB Core, QDC, XML, XML Schema, XPath, XSLT, Z39.50
APPM, DACS, DCAM, EAC-CPF, indecs, Linked Data, MADS, MARC Relator Codes, METS Rights, MODS, OAIS, PREMIS, RAD, RDF, RELAX NG, SGML, SKOS, SRU, XQuery
DC, DIF, DTD, EML, METS, MPEG-21 DIDL, OAIS, QDC, VSO Data Model, XML, XML Schema, XPath, XSLT
AAT, CCO, CDWA, CDWA Lite, CIDOC/CRM, DC, DTD, METS, MIX, MPEG-21 DIDL, MuseumDat, OAI-PMH, Ontology for Media Resource, QDC, SPECTRUM, TGN, ULAN, VRA Core, XML, XML Schema, XPath, XSLT
Domain Community
DTD, LCC, LCSH, MARC, MARCXML, METS, MIX, MODS, OAI-PMH, OAIS, PB Core, PREMIS,
AGLS, APPM, DACS, EAC-CPF, EAD, GILS, ISAAR(CPF), ISAD(G), RAD
ADL, AES
Core Audio,
AACR2, AGLS,
Semi-Weak
Semi-Strong
Strong

The Circle of Life…



(DataONE Primer)

Data Repositories

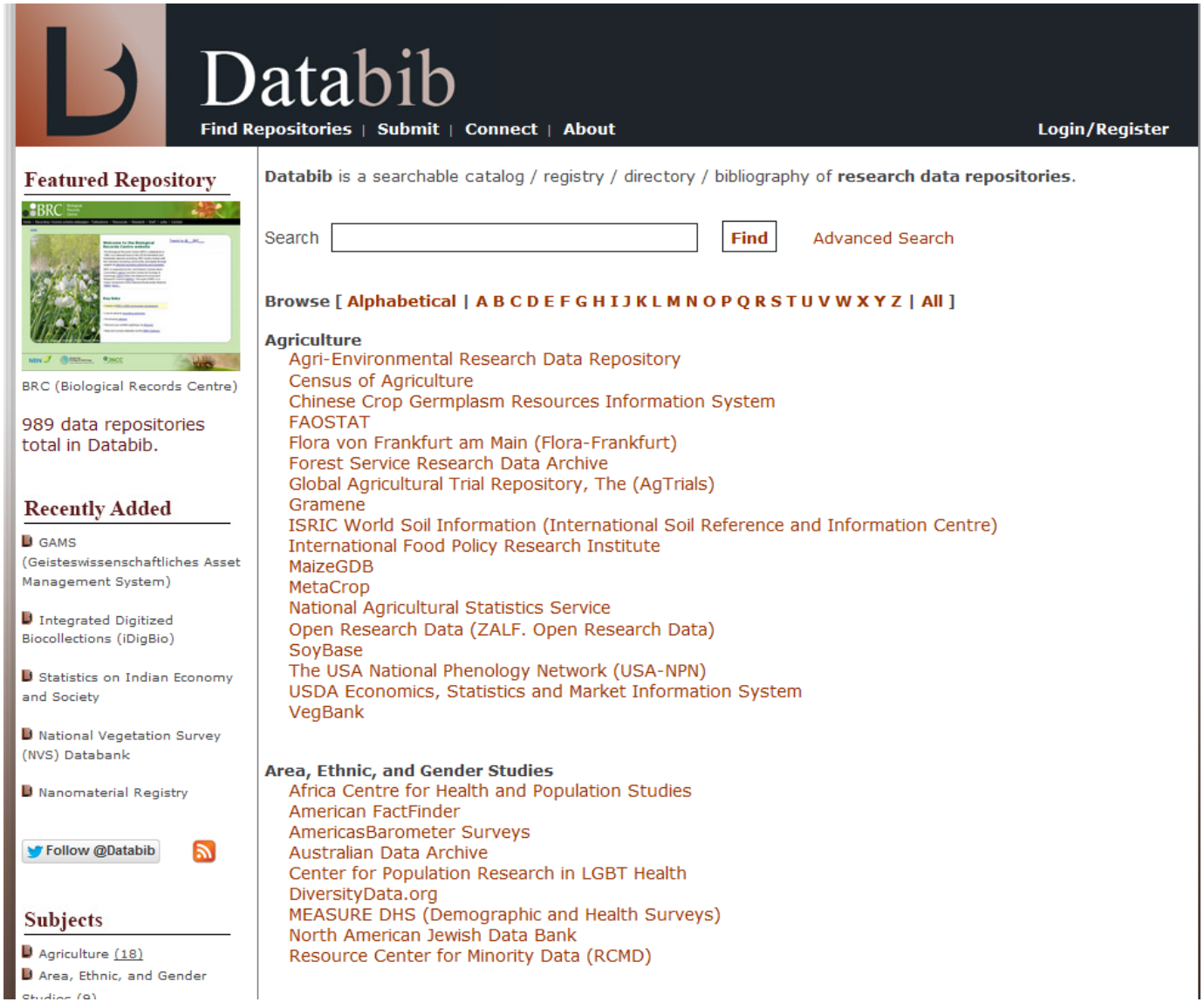
Domain Repositories

- Data stored with similar items
- Researchers in your area are familiar with the repository
- Subject-specific / data-type specific needs addressed
- More computational tools available

MOspace

- Subject repository may not exist
- Preserves link to institution with guarantee of support from the university
- Domain repositories can shut down once the grant ends

Domain Repositories: So Many Choices!



Databib

Find Repositories | Submit | Connect | About Login/Register

Featured Repository

BRC (Biological Records Centre)

989 data repositories total in Databib.

Recently Added

- GAMS (Geisteswissenschaftliches Asset Management System)
- Integrated Digitized Biocollections (iDigBio)
- Statistics on Indian Economy and Society
- National Vegetation Survey (NVS) Databank
- Nanomaterial Registry

Follow @Databib

Subjects

- Agriculture (18)
- Area, Ethnic, and Gender Studies (9)

Databib is a searchable catalog / registry / directory / bibliography of **research data repositories**.

Search Find Advanced Search

Browse [Alphabetical | A B C D E F G H I J K L M N O P Q R S T U V W X Y Z | All]

Agriculture

- Agri-Environmental Research Data Repository
- Census of Agriculture
- Chinese Crop Germplasm Resources Information System
- FAOSTAT
- Flora von Frankfurt am Main (Flora-Frankfurt)
- Forest Service Research Data Archive
- Global Agricultural Trial Repository, The (AgTrials)
- Gramene
- ISRIC World Soil Information (International Soil Reference and Information Centre)
- International Food Policy Research Institute
- MaizeGDB
- MetaCrop
- National Agricultural Statistics Service
- Open Research Data (ZALF. Open Research Data)
- SoyBase
- The USA National Phenology Network (USA-NPN)
- USDA Economics, Statistics and Market Information System
- VegBank

Area, Ethnic, and Gender Studies

- Africa Centre for Health and Population Studies
- American FactFinder
- AmericasBarometer Surveys
- Australian Data Archive
- Center for Population Research in LGBT Health
- DiversityData.org
- MEASURE DHS (Demographic and Health Surveys)
- North American Jewish Data Bank
- Resource Center for Minority Data (RCMD)

Find Data Repositories: http://libraryguides.missouri.edu/c.php?g=28117&p=173330

https://mospace.umsystem.edu/xmlui/

Zip file example:
https://mospace.umsystem.edu/xmlui/handle/10355/40433

Automated analysis of temperture dataloggers to determine hydroperiods of vernal wetlands

Excel file example:
https://mospace.umsystem.edu/xmlui/handle/10355/40553

Pond-Breeding Amphibian Community Composition in Missouri

So, what do I say in my DMP?

Remember that NSF DMPs are subject to peer review, so the nature of the plan will be specific to your project.

"[X type of data] will be deposited in MOspace, the University of Missouri's digital institutional repository. MOspace is based on MIT's DSpace technology and is a joint venture of the University of Missouri's Division of Information Technology and the University Libraries. MOspace items will include appropriate metadata and a permanent URL. Items will be freely available via the MOspace web site at https://mospace.umsystem.edu and will be searchable via Google and other search engines."

DMP Examples: http://data.library.arizona.edu/data-management-plans/data-management-plan-examples

DMP Tool: https://dmp.cdlib.org/

Depositing Data to MOspace

- mospace@missouri.edu
- Include ReadMe text file with information on:
  - Author Name(s)
  - Project Title
  - Software needed to read file
  - 2-5 keywords
  - other repositories that house the data

https://twitter.com/LegoAcademics/status/505361805309640705
Accessed 3 Sept 2014

ORCID

Connecting Research and Researchers

FOR RESEARCHERS | FOR ORGANIZATIONS | ABOUT | HELP | SIGN IN

http://orcid.org

DISTINGUISH YOURSELF IN THREE EASY STEPS

ORCID provides a persistent digital identifier that distinguishes you from every other researcher and, through integration in key research workflows such as manuscript and grant submission, supports automated linkages between you and your professional activities ensuring that your work is recognized. Find out more.

1 REGISTER — Get your unique ORCID identifier Register now! Registration takes 30 seconds.

2 ADD YOUR INFO — Enhance your ORCID record with your professional information and link to your other identifiers (such as Scopus or ResearcherID or LinkedIn).

3 USE YOUR ORCID ID — Include your ORCID identifier on your Webpage, when you submit publications, apply for grants, and in any research workflow to ensure you get credit for your work.

LATEST NEWS

Mon 10/06/2014
ORCID Competition for Best Tweet and Best Quote

Wed 10/01/2014
ORCID in Andalucia! CBUA joins ORCID as consortium member

Fri 09/19/2014
Last chance to comment on Peer Review citation data

More Resources

- MU Libraries Guide on NSF Data Management Plans: http://libraryguides.missouri.edu/datamanagement
- DataONE: Primer on Data Management: What you always wanted to know* (*but were afraid to ask): https://www.dataone.org/best-practices
- MIT Libraries. Data Management and Publishing: http://libraries.mit.edu/guides/subjects/datamanagement/index.html
- UW-Madison Research Data Services: http://researchdata.wisc.edu/
- University of Arizona Libraries Data Management Resources: http://data.library.arizona.edu/

Questions?

Kate Anderson | AndersonKat@missouri.edu | @ZalkLibrary