Data Librarianship and Small (& Large) Collections Support

SPNHC 2016

Brian Westra

Science Data Services Librarian

University of Oregon

UNIVERSITY OF OREGON Libraries

Survey question

Go to **www.menti.com** and use the code **56 57 16**

https://www.menti.com/77e300

Results

Why libraries are involved in Research Data Services

Expertise and philosophy similar to other services for scholarly communications/research products:

- Organize
- Describe
- Preserve
- Provide access

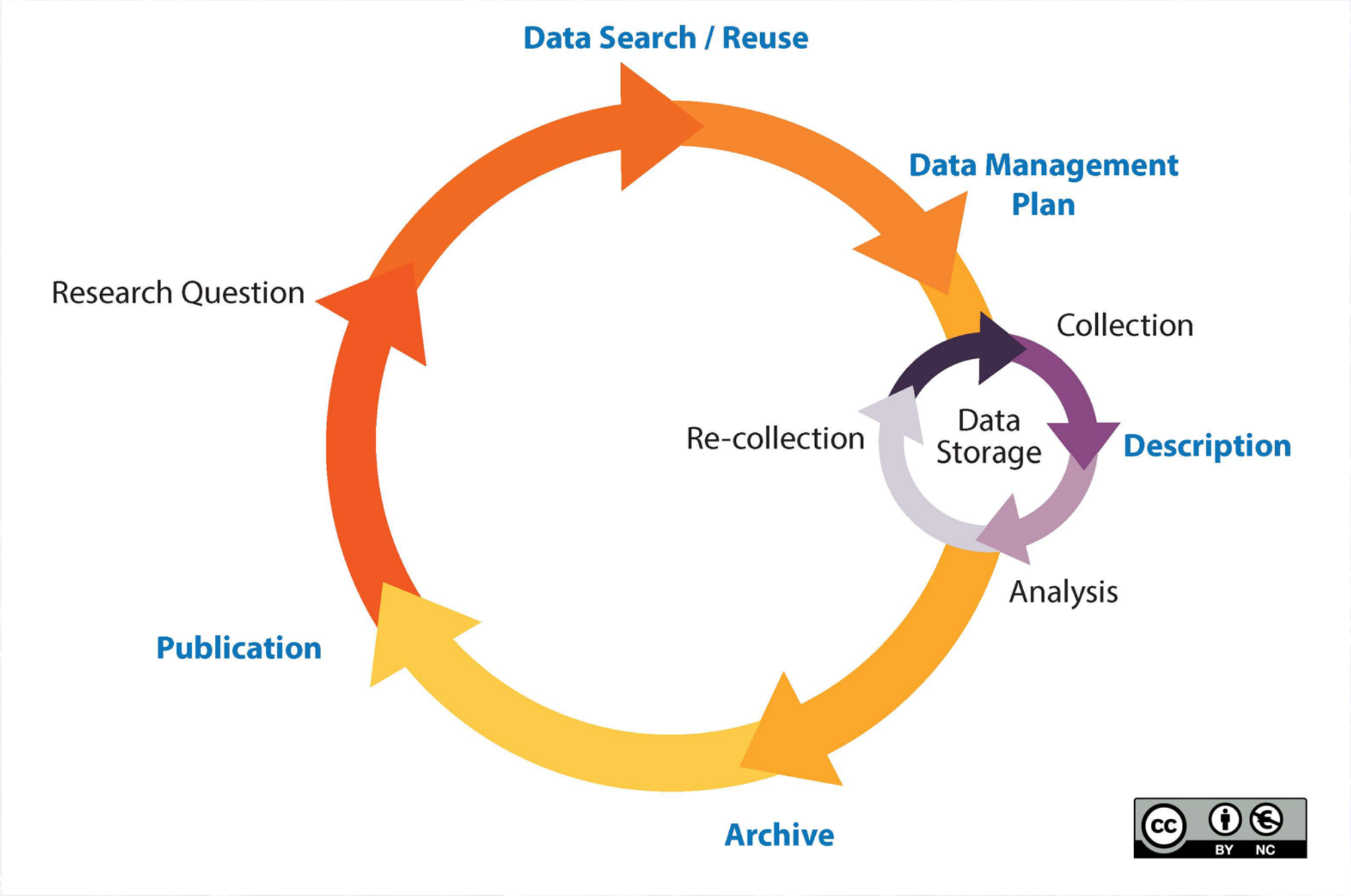


Image from UCSC Data Management guide, at http://guides.library.ucsc.edu/datamanagement/ accessed June 21, 2016

Data services and libraries

- Data Management Plan (DMP) consultations
- Data information literacy
- Infrastructure for organizing, preserving and sharing data
- Metadata support
- ‘Curation-friendly’ tools and practices that facilitate data sharing and preservation

NSF Data Management Plan requirement - 2011

Some of the challenges for proposal writers:

- Data and metadata formats and standards
- Where and how to share data
- Backup and storage ≠ Preservation

How authors propose to share their data

	All	BIO	CISE	ENG	GEO	MPS	SBE	Scale
Journal / supplement	36	27	23	45	35	54	18	80
Data center or repository	34	75	14	8	66	25	42	70
On request	30	23	30	38	24	34	26	60
Personal website	25	13	44	31	25	20	12	50
Other method	22	27	30	15	23	18	22	40
Institutional repository	17	6	12	20	6	28	20	35
Conference / proceedings	13	8	11	23	8	13	8	30
Did not specify	8	0	18	9	2	8	4	25
Thesis / Dissertation	3	0	0	5	2	6	2	20
Not planning to share	3	0	5	3	0	1	10	10
Book	2	2	2	3	2	2	2	0

DMPs as a Research Tool (DART): https://osf.io/kh2y6/files/

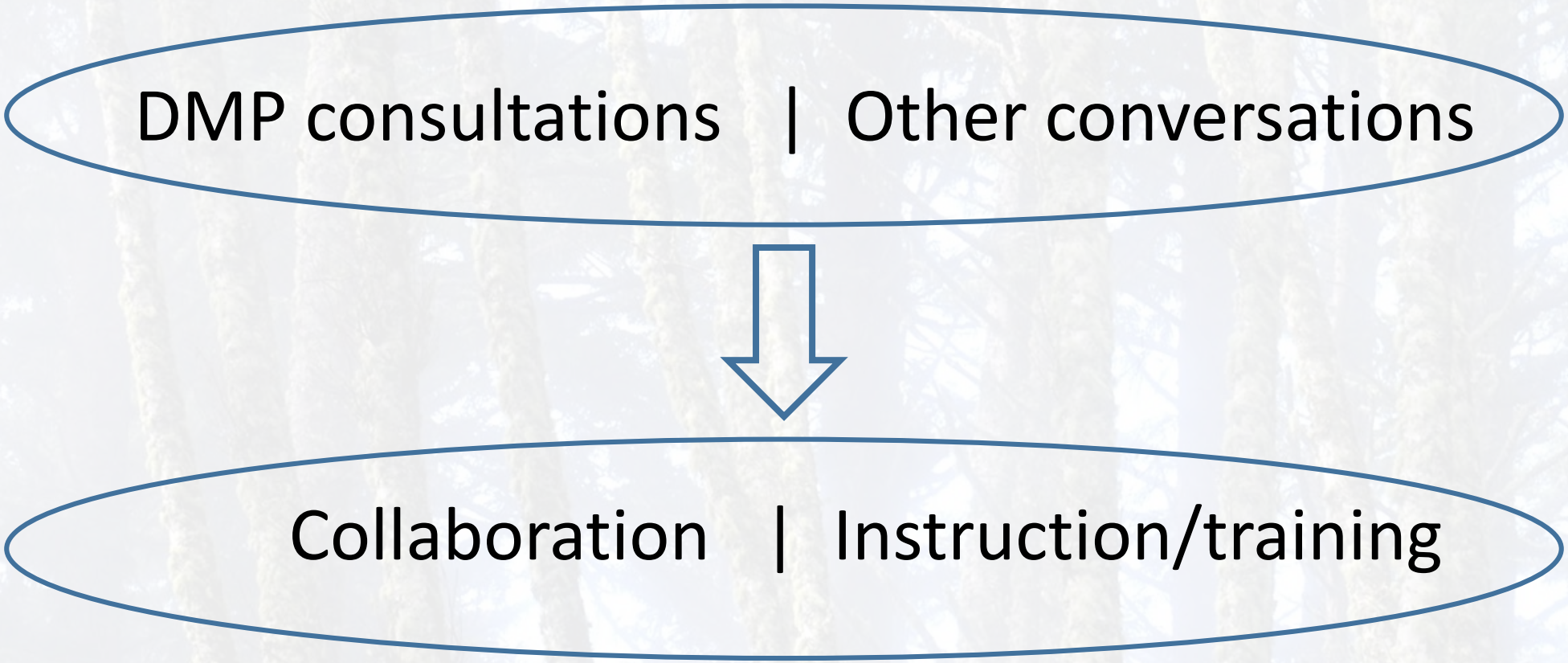
What kinds of metadata do the plans propose?

	Identifies metadata standards/formats					
	Complete / detailed	Addr., but incomplete	Did not address	Most common metadata standards (N)	Yes (%)	Scale
All	19.4	23.7	57.0	Dublin Core (23), CSDGM*/ISO 19115 (12), EML^ (11)	14.9	60
BIO	38.5	21.2	40.4	EML (8), Darwin Core (5), CSDGM/ ISO 19115 (4), Dublin Core (3)	38.5	50
CISE	9.1	28.8	62.1	DIF# (2), Dublin Core (1), EML (1)	6.1	40
ENG	15.1	30.2	54.7	Dublin Core (7)	9.4	30
GEO	18.1	18.1	63.9	CSDGM/ISO 19115 (3), EML (2), ACADIS~ (2)	12.2	20
MPS	18.8	22.4	58.8	Dublin Core (9), FITS% (3)	15.3	10
SBE	26.0	16.0	58.0	CSDGM (5), ODM& (2) Dublin Core (2)	16.3	0

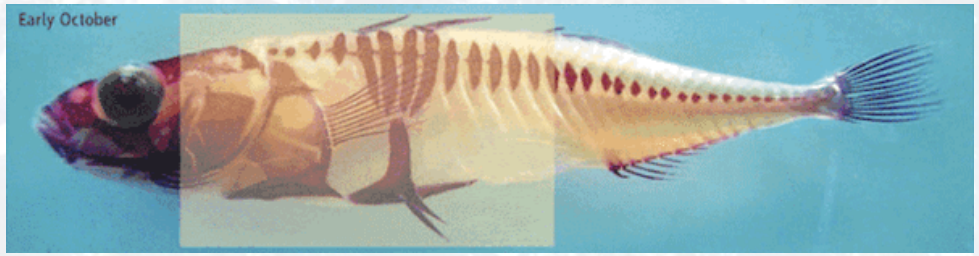
*Content Standard for Digital Geospatial Metadata; ^Ecological Metadata Language; #Directory Interchange Format; ~Advanced Cooperative Arctic Data and Information Service; %File Information Tool Set; &Observations Data Model

DMPs as a Research Tool (DART): https://osf.io/kh2y6/files/

Bridging communities



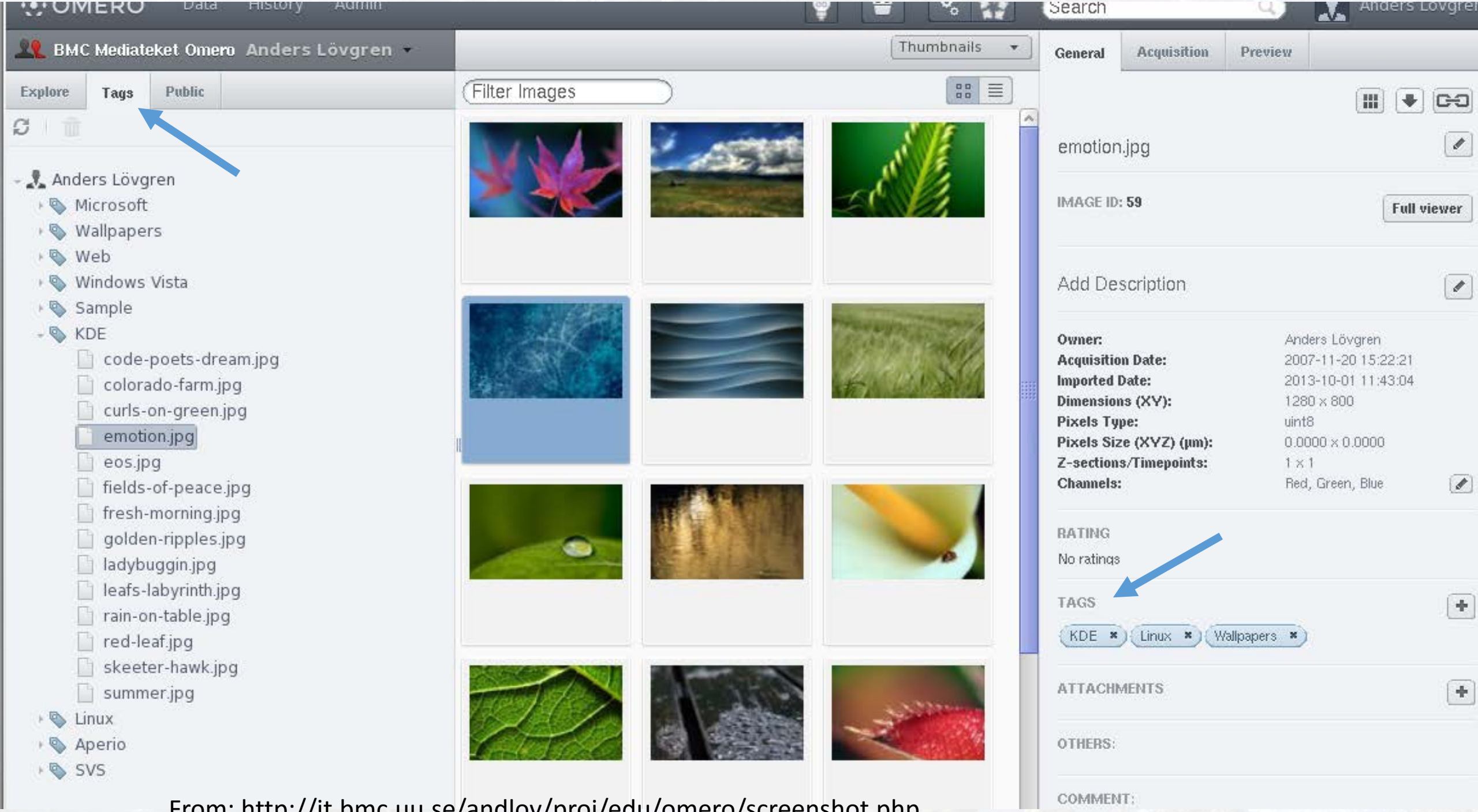
Managing Images in an Evo Bio team



- Stickleback fish
- Wanted workflow-friendly tool to improve:
  - Organizing, describing, sharing, and analyzing images
- UO + partners at UAlaska Anchorage

OMERO – Open Microscopy Environment

- Server + Desktop Client
- Virtual folder and file structure to organize images
- Tags and other descriptive information
- ImageJ integration
- BioFormats integration



From: http://it.bmc.uu.se/andlov/proj/edu/omero/screenshot.php

OMERO – Outcomes

- ~1000 images in the system
- Less time spent looking for an image
- Share images and views of images across multi-institutional team
- Add tags and other descriptive information
- BioFormats integration – handles proprietary formats

OMERO – Lessons learned

- Decentralization
  - Independence comes with a cost
- Champions act within trusted relationships
- Unstructured metadata – ease of use but costs later on

Managing collections data

Biodiversity research collection

- Insect - fungi - substrate interactions
- Needs:
  - Data sharing
  - Image preservation

Managing collections data

Using  to:

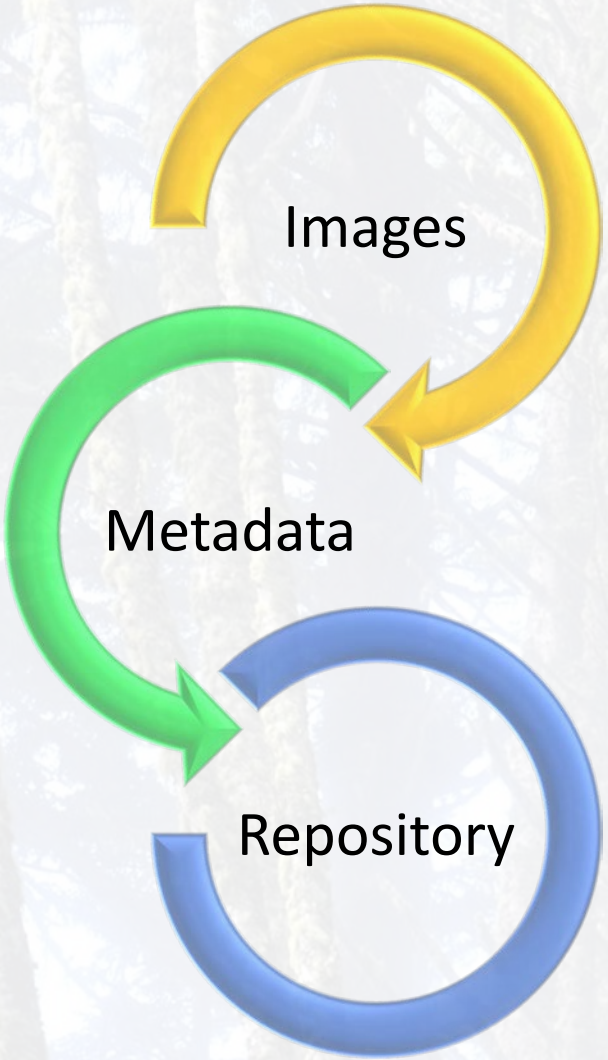
Organize collection objects

Manage structured metadata

To publish data to iDigBio

Next steps?

Archive images in repository



Managing collections data – Lessons

Transition to a new system takes time

Leverage external support

- iDigBio sponsored Specify training
- Specify team for schema/mapping, Specify Cloud (hosting)

Many other opportunities

- Curriculum-integrated data information literacy
- Follow-up on DMPs: intention > practice
- Build/share infrastructure/suite of tools and services:
  - Active data processes;
  - Storage and backup;
  - Post-project data preservation and sharing

Thank you!

Brian Westra | bwestra@uoregon.edu

http://library.uoregon.edu/datamanagement

This presentation at SPNHC thanks to support from a University of Oregon Foundation/Solari Incentive award and iDigBio

DART project funded by Institute of Museum & Library Services grant number LG-07-13-0328