Engaging Scientists in Metadata Ownership: Framing the Questions

International Symposium on Knowledge Communities 2012
December 13 – 15, 2012
Research Center for Knowledge Communities
University of Tsukuba

Jane Greenberg,
Metadata Research Center <MRC>
SILS/UNC-CH

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE
Metadata Research Center <MRC>

Outline

1. Assumptions
2. Motivation
3. Overriding goals and objectives
   - Dryad
   - HIVE
   - DataONE—PAMWG work
4. Conclusions and framing questions
5. Q&A

Assumptions

Prevailing metadata generation methods result in advantages and limitations

	+	–
Automatic	Efficient, consistent	Disambiguation challenges
****Manual (info. professional)**	Able to disambiguate	Costly, inefficient, not as intimate w/subject
Manual (author)	Intimate, …	Quality limitations
Collaborative/ combinatory	Best of all worlds	Challenges magnified, costly to find right combination
Social/annotation	Cheap, additional views	Inconsistent

*Formalized, standard schemes, vetted on some level.

Assumptions

1. Prevailing metadata generation methods result in advantages and limitations
2. More than one way to skin a cat
   - Complementary, alternative approaches
   - Social technology
3. Ownership appeal
   - Empowerment and sustainability

Motivation

1. Metadata Generation Research; AMeGA 2001-2005, Metadata bottleneck

   27 scientists; don't touch my metadata

2. COPD (Chronic Obstruction Pulmonary Disease) ontology ~ NIH

   Need + attract domain experts; sustainability

3. Dublin Core / proliferation of metadata schemes (Riley, 2009-2010; Willis, et al, 2012)

Long tail →

4. Interoperability and data reuse

   Dryad, DataONE, …

Outline

1. Assumptions
2. Motivation
3. Overriding goals and objectives
   - Dryad
   - HIVE
   - DataONE—PAMWG work
4. Conclusions and framing questions
5. Q&A

Dryad

DRYAD

Submit Data Now!

See how to submit

Account

Login or Register

owse

Authors

Journal Title

formation

Depositing Data

Using Data

Dryad Members

Journal Archiving Policy

About Dryad

Dryad Blog

Dryad Documentation

Dryad is a nonprofit organization and an international repository of data underlying scientific and medical publications.

The scientific, educational, and charitable mission of Dryad is to promote the availability of data underlying findings in the scientific literature for research and educational reuse.

The vision of Dryad is a scholarly communication system in which learned societies, publishers, institutions of research and education, funding bodies and other stakeholders collaboratively sustain and promote the preservation and reuse of data underlying the scholarly literature.

As of Dec 11, 2012, Dryad contains **2396 data packages** and **6482 data files**, associated with articles in **175 journals**.

Recently Published Data

Brassey CA, Margetts L, Kitchener AC, Withers PJ, Manning PL, Sellers WI (2012) Data from: Finite element modelling vs. classic beam theory comparing methods for stress estimation in a morphologically diverse sample of vertebrate long bones. *Journal of the Royal Society Interface* doi:10.5061/dryad.9ct2f

Hernandez RR, Mayernik MS, Murphy-Mariscal ML, Allen MF (2012) Data from: Advanced technologies and data management practices in environmental science: lessons from academia. *BioScience* doi:10.5061/dryad.cv86385c

Delcourt M, Blows MW, Aguirre JD, Rundle HD (2012) Data from: Evolutionary optimum for male sexual traits characterized using the multivariat Robertson–Price Identity. *Proceedings of the National Academy of Sciences of the United States of America* doi:10.5061/dryad.d7g00

Behie SW, Bidochka MJ, Zelisko PM (2012) Data from: Endophytic-insect parasitic fungi translocate nitrogen directly from insects to plants. *Sc* doi:10.5061/dryad.6pv0v

Popat R, Crusz SA, Messina M, Williams P, West SA, Diggle SP (2012) Data from: Quorum sensing and cheating in bacterial biofilms. *Proceed of the Royal Society B* doi:10.5061/dryad.vg0b5

Caravas J, Friedrich M (2012) Data from: Shaking the Diptera tree of life: performance analysis of nuclear and mitochondrial sequence data partit *Systematic Entomology* doi:10.5061/dryad.f7m

Hefley T, Hygnstrom S, Gilsdorf J, Clements G use on white-tailed deer. *Journal of Fish and W*

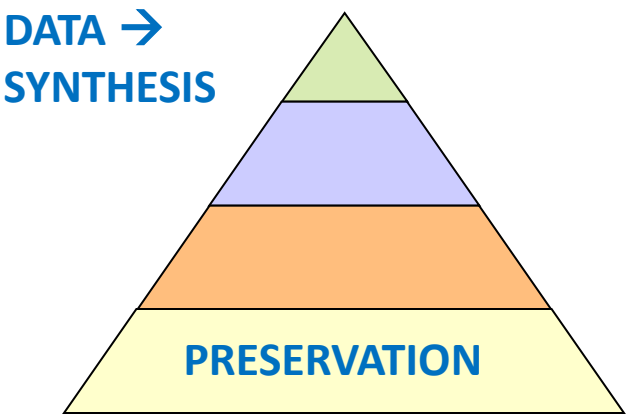
Newcomer TA, Kaushal SS, Mayer PM, Shield

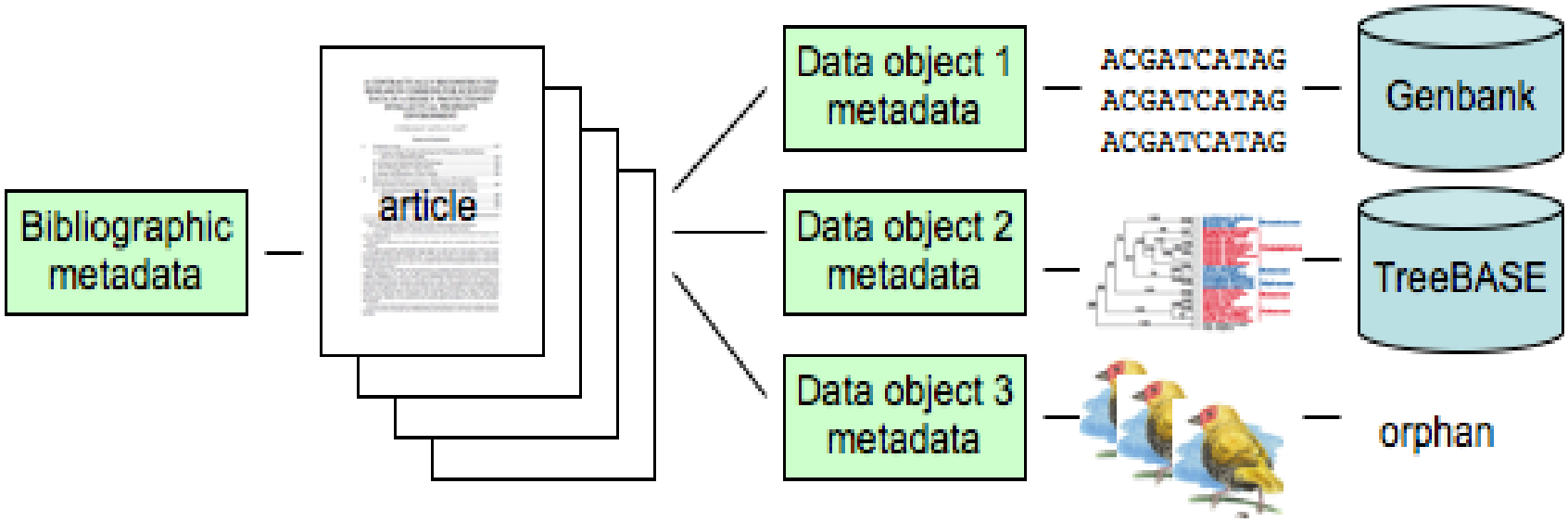
- DSpace repository software (open source)
- DOIs via California Digital Library/DataCite
- CCZero (CC0)
- Dryad DCAP (Dublin Core Application Profile), ver. 3.0

Dryad’s goals

Dryad's Goals

- One-stop deposition/access for data objects supporting published research…
- Acquisition, preservation, discovery, and reuse of heterogeneous digital datasets
- Allow journals and societies to pool their resources





Dryad development and governance

- Dryad development - a joint project of NESCent, the UNC Metadata Research Center, and a growing number of partner organizations.
  - Stakeholders: journals, publishers and scientific societies, and researchers
- Governance
  - Dryad is a nonprofit organization
  - Governed by member organizations, including journals, publishers, scientific societies, funding agencies, and other stakeholders.
  - Board: Sets policy and long-term strategic goals
    - Reps from science, journals, societies, OCLC, MS, etc.

Charter Dryad members

- The American Naturalist (Am. Soc. of Naturalists)
- BMJ Open (British Medical Association)
- The Biological Journal of the Linnean Society (Linnean Society of London)
- BioMed Central
- Ecology Letters (Recherche Scientifique)
- Evolution (Society for the Study of Evolution)
- Evolutionary Applications
- Heredity (The Genetics Society)
- British Ecological Society
- Journal of Evolutionary Biology (European Society for Evolutionary Biology)
- Journal of Fish and Wildlife Management
- Journal of Heredity (The American Genetic Association)
- Journal of Paleontology and Paleobiology (Paleontological Society)
- Molecular Biology and Evolution (Society for Molecular Biology and Evolution)
- Molecular Ecology and Molecular Ecology Resources
- Molecular Phylogenetics and Evolution
- Oikos (Nordic Society Oikos)
- Oxford University Press
- Pensoft Publishers
- Public Library of Science
- Science (American Association for the Advancement of Science)
- Systematic Biology (Society for Systematic Biology)
- Wiley-Blackwell

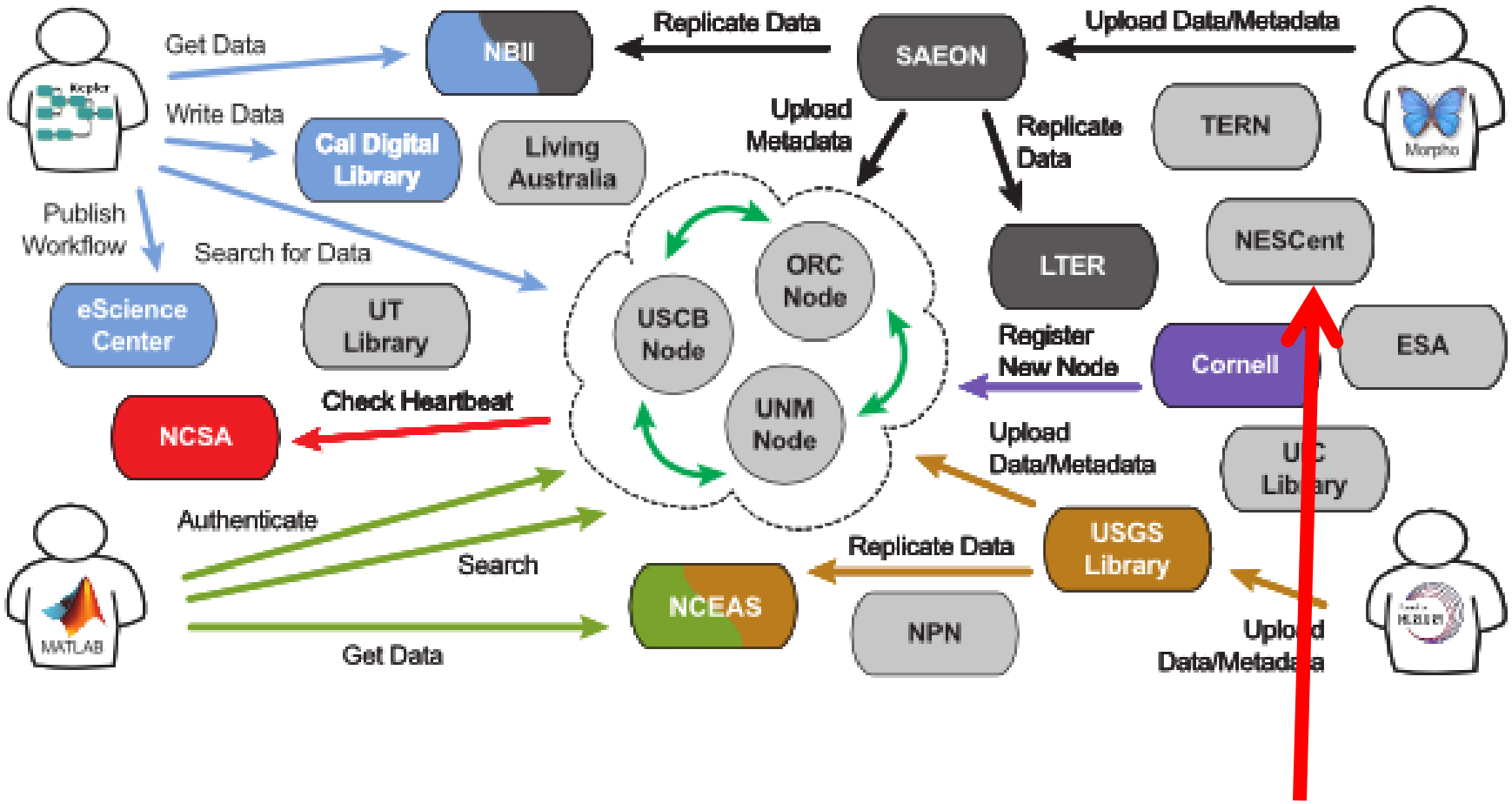
Partner repositories: Knowledge Network for Biocomplexity, NCBI GenBank, TreeBASE, DataONE

Joint Data Archiving Policy

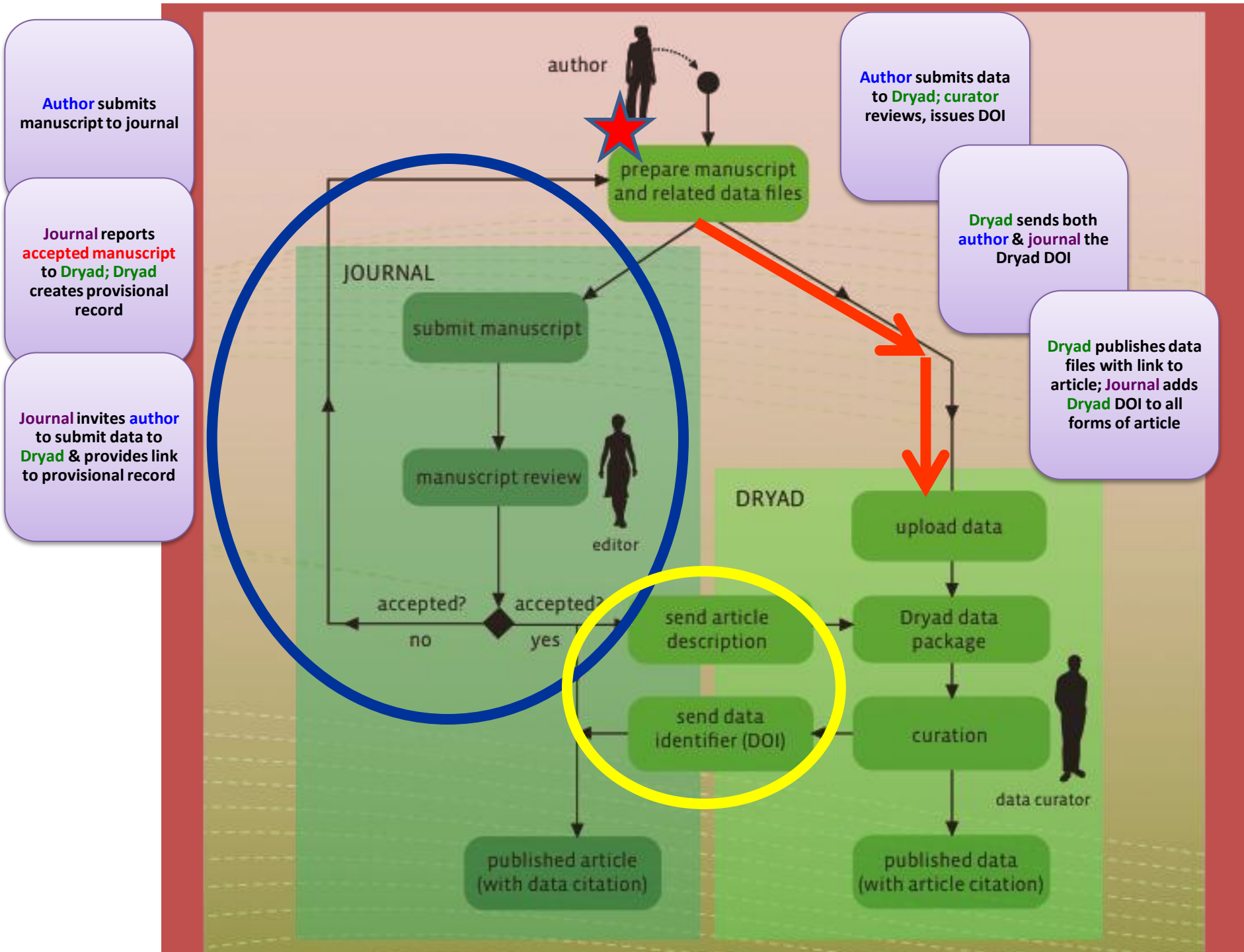
(http://datadryad.org/jdap)

<< Journal >> requires, as a condition for publication, that data supporting the results in the paper should be archived in an appropriate public archive, such as **<< list of approved archives here >>**. Data are important products of the scientific enterprise, and they should be preserved and usable for decades in the future. Authors may elect to have the data publicly available at time of publication, or, if the technology of the archive allows, may opt to embargo access to the data for a period up to a year after publication. Exceptions may be granted at the discretion of the editor, especially for sensitive information such as human subject data or the location of endangered species.

- Whitlock, M. C., M. A. McPeek, M. D. Rausher, L. Rieseberg, and A. J. Moore. 2010. Data Archiving. American Naturalist. 175(2):145-146. DOI:10.1086/650340


Get Data
Write Data
Publish Workflow
Search for Data
NBII
Cal Digital Library
Living Australia
eScience Center
UT Library
Replicate Data
SAEON
Upload Data/Metadata
Upload Metadata
Replicate Data
TERN
Morpho
LTER
NESCent
ORC Node
USCB Node
UNM Node
Register New Node
Cornell
ESA
Check Heartbeat
NCSA
Upload Data/Metadata
UC Library
Authenticate
Search
Get Data
MATLAB
NCEAS
Replicate Data
USGS Library
NPN
Upload Data/Metadata

Dryad's Workflow


Author submits manuscript to journal
Journal reports accepted manuscript to Dryad; Dryad creates provisional record
Journal invites author to submit data to Dryad & provides link to provisional record
Author submits data to Dryad; curator reviews, issues DOI
Dryad sends both author & journal the Dryad DOI
Dryad publishes data files with link to article; Journal adds Dryad DOI to all forms of article
author
prepare manuscript and related data files
JOURNAL
submit manuscript
manuscript review
editor
accepted?
no
accepted?
yes
send article description
send data identifier (DOI)
published article (with data citation)
DRYAD
upload data
Dryad data package
curation
data curator
published data (with article citation)

From: managing.editor@molecol.com
Date: April 19, 2011 3:09:22 PM EDT
To: Author
Cc: journal-submit@datadryad.org
Subject: Dryad entry for MEC-11-0140.R1

Dear Author

Many thanks for agreeing to participate in the Dryad project. To upload your data, please click the link below- it will take you directly to your entry in the Dryad database.

http://datadryad.org/submit?journalID=MolEcol&manu=223330

<deleted text>

Once you have uploaded your data please include the Dryad identifier in your manuscript. Please let me know if you have any questions about this process.

All the best,
Tim Vines,
Managing Editor, Molecular Ecology

Describe publication

Submitting data to Dryad consists of three simple steps:

1. **Describe your publication**
2. Upload and describe your data files
3. Approve data for publication

Please describe your publication in as much detail as possible. Providing a detailed description will make it easier for othe
data in Dryad. Please describe the **publication only**. Do not enter information specific to your data files on this page.

Fields marked with an asterisk (*) are required. For more information on expected contents for a field, hold your mouse o
question.

Publication metadata

Title*: Adaptive responses and disruptive effects: how major wildfire

Authors*:

Last name, *e.g. Smith* | First name + initial, *e.g. Donald F.* | Add

- Banks, Sam
- Blyton, Michaela
- Blair, David
- McBurney, Lachlan
- Lindenmayer, David

Remove selected

Journal name*: Molecular Ecology

Abstract:

Environmental disturbance is predicted to play a key role in the evolution of animal social behaviour. This is because disturbance affects key factors underlying

Pre-populated metadata field

Data file *

Please upload your data file or provide the identifier of a file located in another re

Browse...

External file identifier

(please select a repository)

(please select a repository)
TreeBASE
GenBank
KNB

Data file description

Title*:

Description:

VOL. 177, NO. 4 THE AMERICAN NATURALIST APRIL 2011

Multiple Benefits Drive Helping Behavior
Breeding Bird: An Integrated A

Sjouke A. Kingma,[1,*] Michelle L. Hall,[1,2,3] and Anne Peters[1,4]

1. Max Planck Institute for Ornithology, Vogelwarte Radolfzell, Schlossallee 2, 78315 Radolfze
Sanctuary, Australian Wildlife Conservancy, PMB 925, Derby, Western Australia 6728, Austral
Australian National University, Canberra, Australian Capital Territory 0200, Australia; 4. Sch
University, Clayton, Victoria 3800, Australia

Submitted July 23, 2010; Accepted January 3, 2011; Electronically published March 10, 2011

Dryad data: http://dx.doi.org/10.5061/dryad.8210.

DRYAD

Search Data

Submit Data Now!

See how to submit

My Account

Login or Register

Browse

Authors

Journal Title

Information

Depositing Data

Using Data

Dryad Partners

Archiving Policy

About Dryad

Dryad Blog

Data from: Patterns of morphological and plastid DNA variation in the Corallorhiz species complex (Orchidaceae)

When using this data, please cite the original article:

Barrett CF, Freudenstein JV (2009) Patterns of morphological and plastid DNA variation in the Corallorhiza striata species complex (Orchidaceae). Systematic Botany 34(3): 496-504. doi:10.1600/036364409789271245

Additionally, please cite the Dryad data package:

Barrett CF, Freudenstein JV (2009) Data from: Patterns of morphological and plastid DNA variation in the Corallorhiza stria species complex (Orchidaceae). Dryad Digital Repository. doi:10.5061/dryad.1013

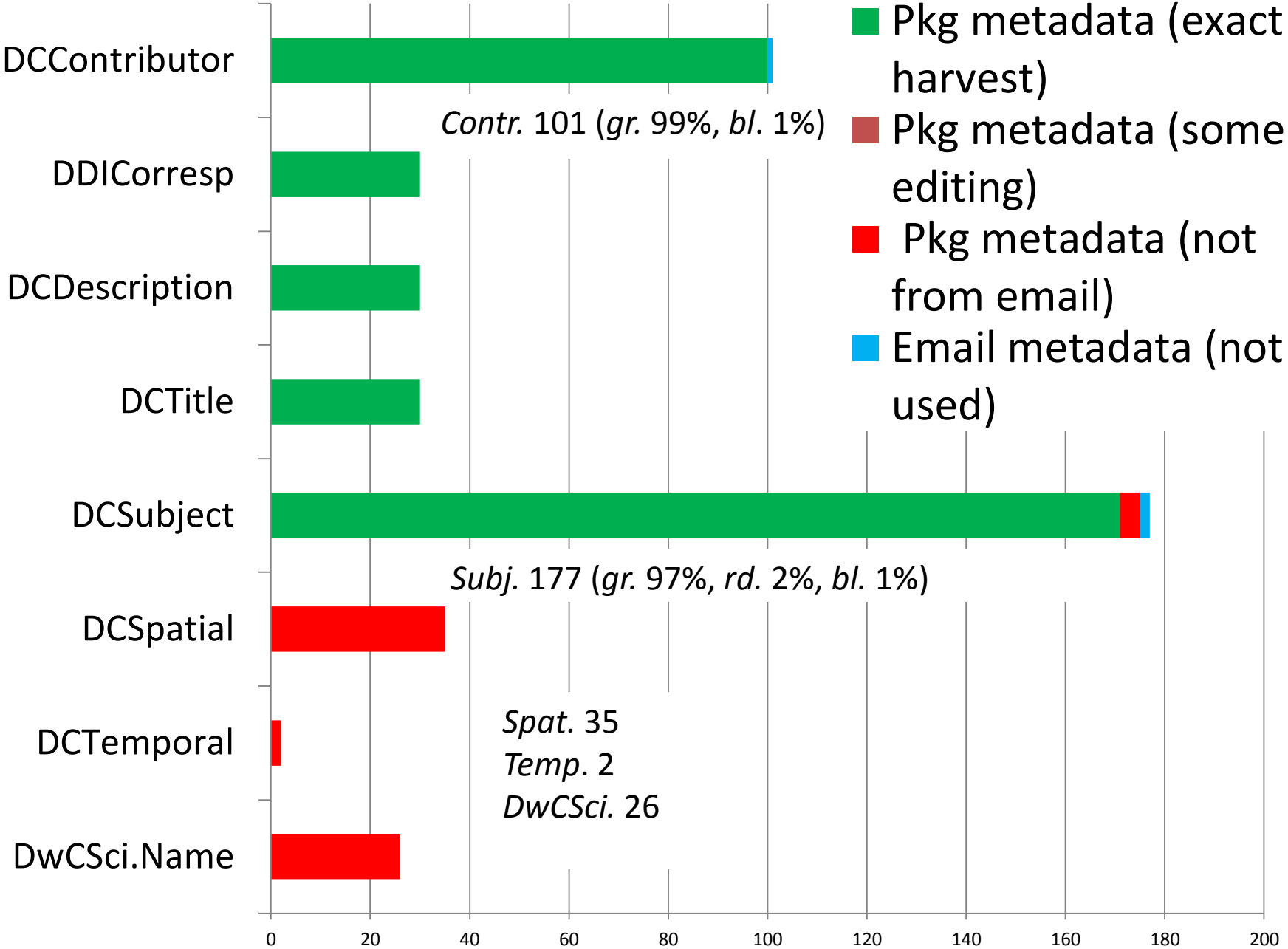
Cite | Sh

Dryad Package Identifier	doi:10.5061/dryad.1013	44 views
Individual Data Files	Supplementary Figure 1	37 views 9 downloads
	Supplementary Figure 2	34 views 11 downloads

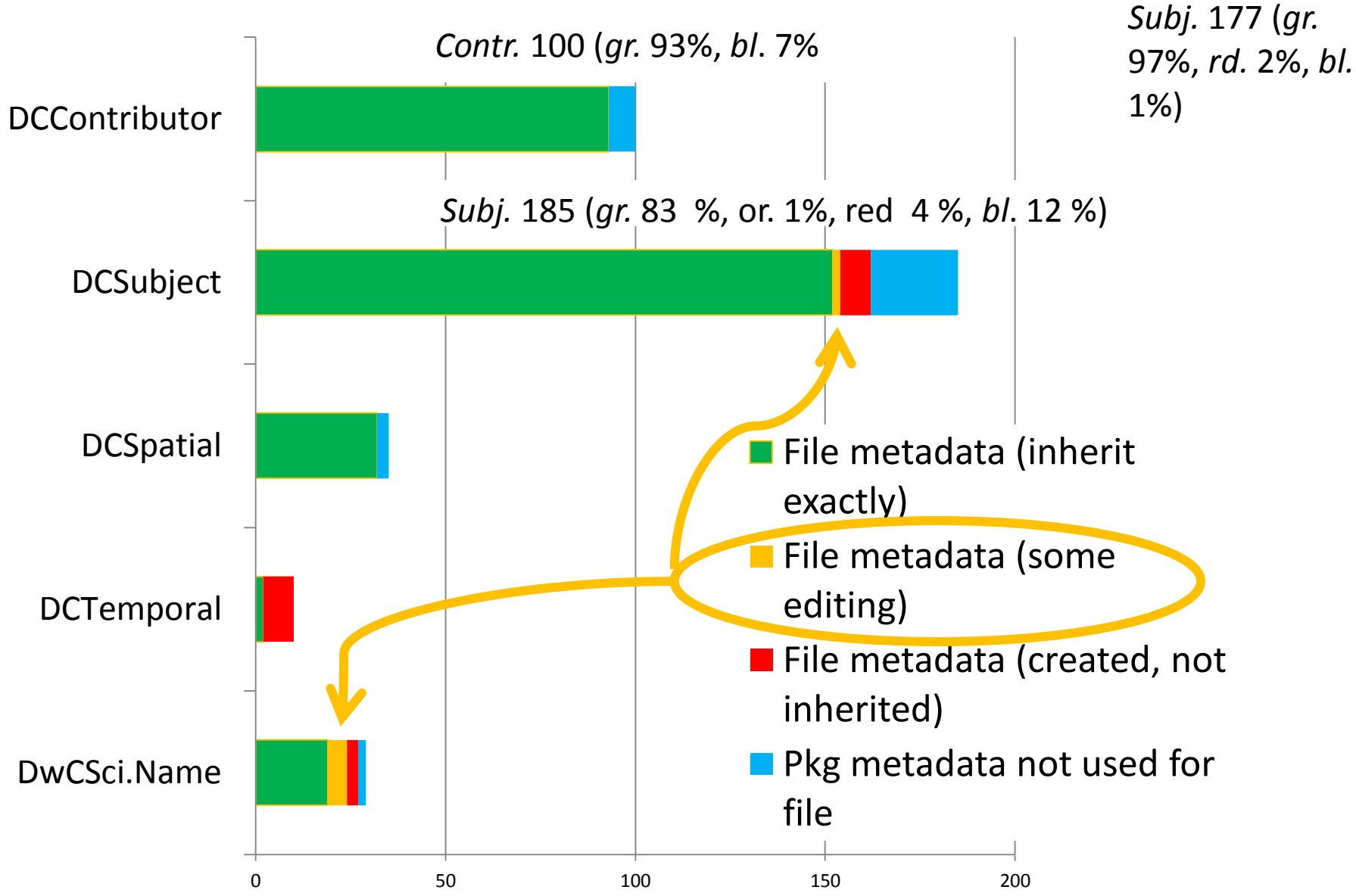
Abstract

Corallorhiza striata is a wide-ranging, morphologically variable, mycoheterotrophic species complex distributed acro North America. Objectives of this study were to assess relationships and test validity of previously delimited varieties striata, including the recently described C. bentleyi. Two plastid DNA regions were sequenced for individuals from s populations across North America, identifying four major clades. The large-flowered C. striata var. striata (northern U.S.A., southern Canada) was sister to the smaller-flowered var. vreelandii (southwestern U.S.A., Mexico), and thes sister to a Californian clade with relatively intermediate-sized flowers. C. striata var. involuta (Mexico) and the endan C. bentleyi (eastern U.S.A.) shared a close relationship, sister to the remaining C. striata. Principal Components An and Nonparametric Multivariate Analysis of Variance on nine qu astid DN

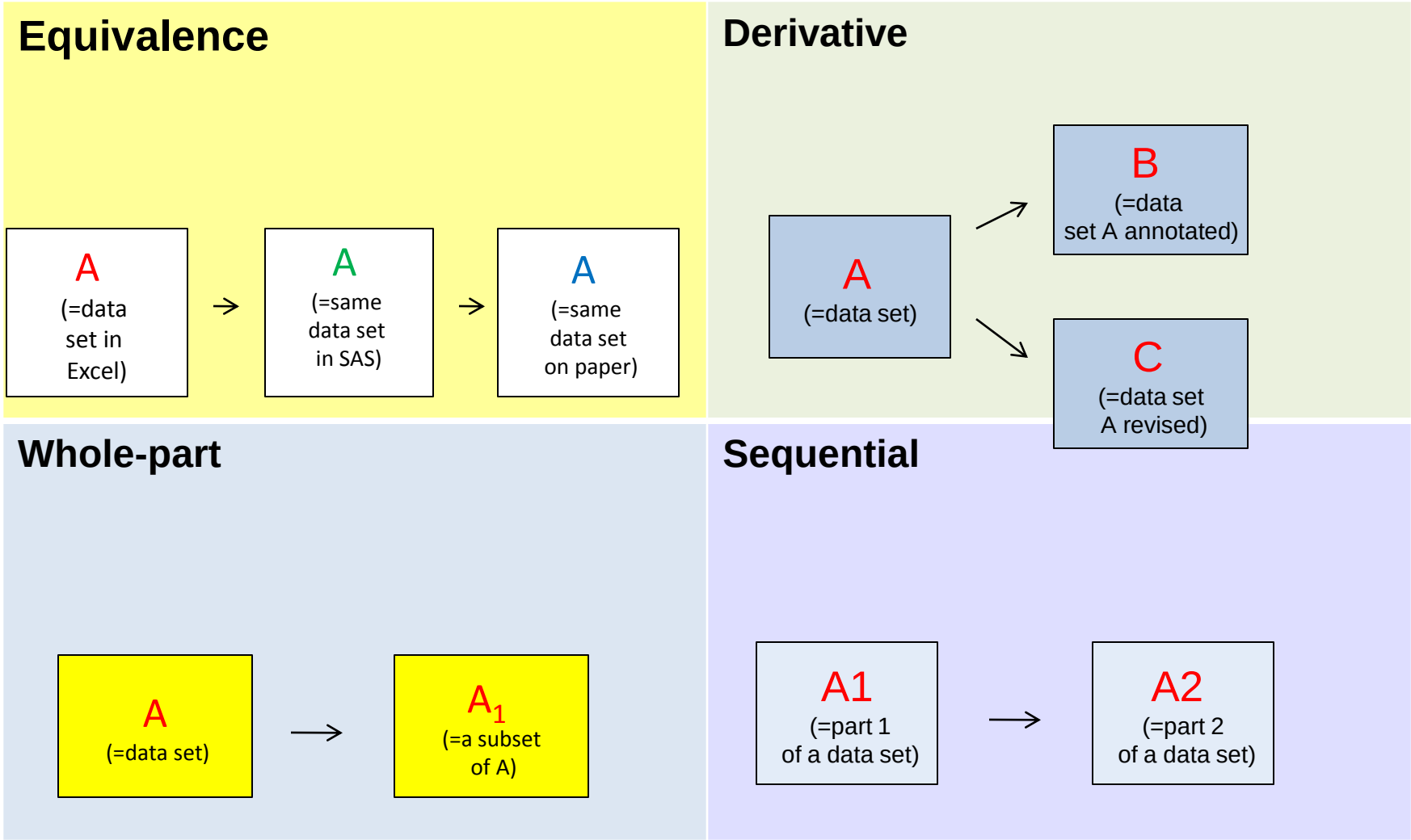
Package metadata harvested from email



File metadata harvested from package metadata



Data object relationships



Instantiation; notion of "a work"

- ***Bibliographic relationship (*Tillett, 1992, 1992; Smiraglia, 1999, 2000+.; Coleman, 2002)**

Challenges

motivating challenges…

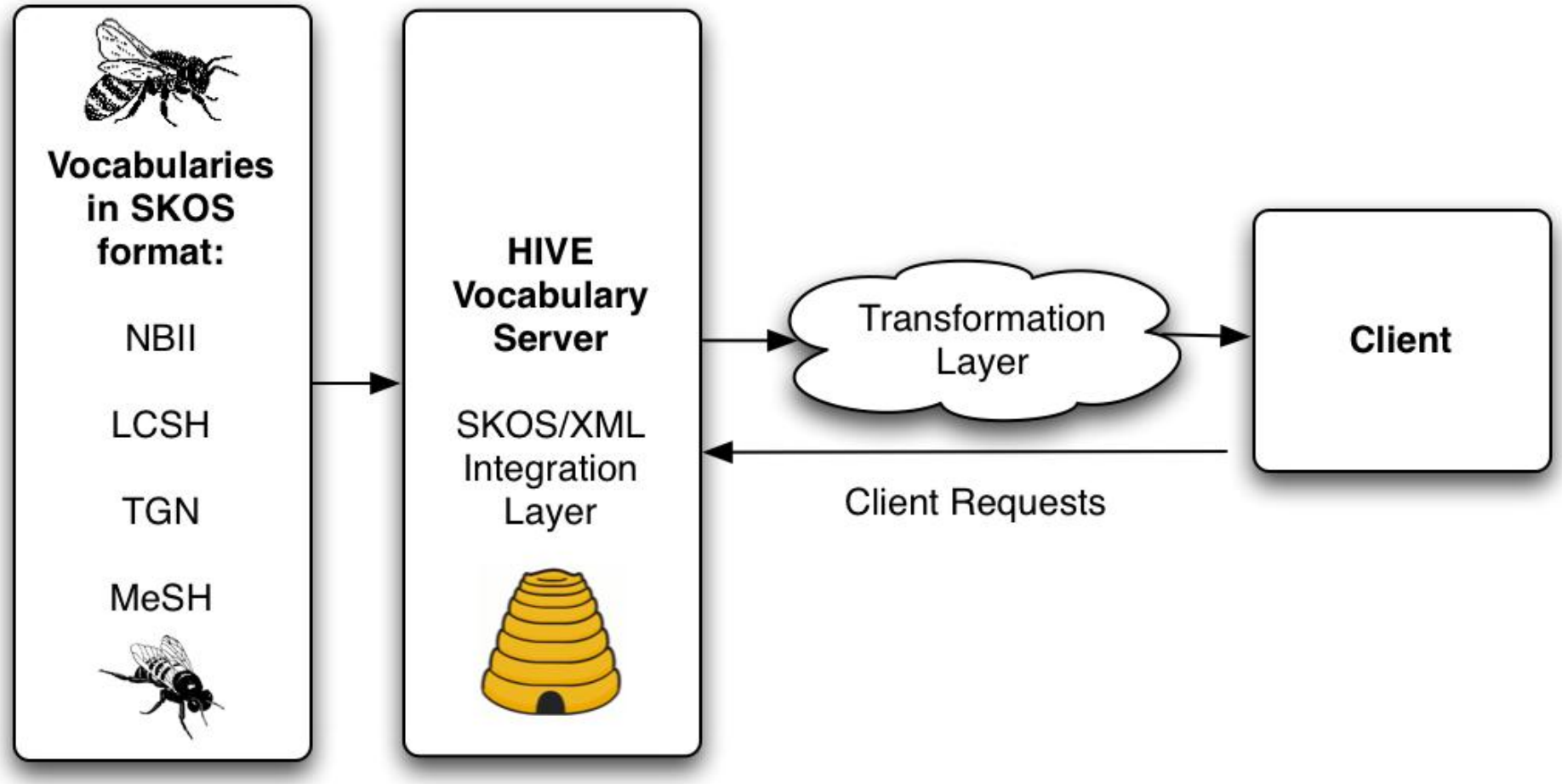
- Operational with ongoing development

  R&D, metadata, and team logistics

- Different workflows
- Growth and sustainability

HIVE

Helping Interdisciplinary Vocabulary Engineering (HIVE)



- <AMG> approach for integrating discipline CVs
- Model addressing **C V cost, interoperability, and usability constraints** (interdisciplinary environment)

Building, Sharing, Evaluation the HIVE….

HIVE Team


José R. P. Agüera
Lina Huang
Lee Richardson
Bob Losee
Madhura Marathe
Hollie White
Jane Greenberg
Craig Willis
Ryan Scherle
INSTITUTE of Museum and Library SERVICES
National Evolutionary Synthesis Center
NESCent
UNC
SCHOOL OF INFORMATION AND LIBRARY SCIENCE
Metadata Research Center <MRC>

Vocabulary needs for Dryad

- Vocabulary analysis
  - 600 keywords, Dryad partner journals
    - Vocabularies: NBII Thesaurus, LCSH, the Getty's TGN, ERIC Thesaurus, Gene Ontology, IT IS (10 vocabularies)
    - Facets: taxon, geographic name, time period, topic, research method, genotype, phenotype…
- Results

  431 **topical** terms, exact matches
  - NBII Thesaurus, 25%; MeSH, 18%

  531 terms (**topical terms, research method and taxon**)
  - LCSH, 22% found exact matches, 25% partial
- Conclusion: Need multiple vocabularies

HIVE Partners

Vocabulary Partners

- Library of Congress: LCSH
- the Getty Research Institute (GRI): TGN (Thesaurus of Geographic Names)
- United States Geological Survey (USGS): NBII Thesaurus, Integrated Taxonomic Information System (ITIS)
- National Library of Medicine and the National Agricultural Library

Advisory Board

- Jim Balhoff, NESCent
- Libby Dechman, LCSH
- Mike Frame, USGS
- Alistair Miles, Oxford, UK
- William Moen, University of North Texas
- Eva Méndez Rodríguez, University Carlos III of Madrid
- Joseph Shubitowski, Getty Research Institute
- Ed Summers, LCSH
- Barbara Tillett, Library of Congress
- Kathy Wisser, Simmons
- Lisa Zolly, USGS

WORKSHOPS HOSTS: Columbia Univ.; Univ. of California, San Diego; George Washington University; Univ. of North Texas; Universidad Carlos III de Madrid, Madrid, Spain

HIVE
Vocabulary Server

Helping with Interdisciplinary Vocabulary Engineering

Home | Concept Browser | Indexing

pened vocabularies: ✕AGROVOC ✕LCSH ✕MESH ✕NBII +Add

animals Search

AGROVOC | LCSH | MESH | NBII

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 [0-9]

- Additives
- Administration
- Africa
- Agents
- Aggregate data
- Agricultural structure
- Agroindustrial sector
- Alcohols
- Aldehydes
- Alkaloids
- Americas
- Amides
- Amino acids
- Amino compounds

Your search for animals returns following concepts:

AGROVOC Aquatic animals
LCSH Pottery animals
LCSH Laboratory animals
LCSH Animals
AGROVOC Noxious animals
LCSH Animals--Wintering
LCSH Food animals
LCSH Cannibalism in animals
AGROVOC Draught animals
AGROVOC Performing animals
AGROVOC Wild animals
AGROVOC Meat animals
AGROVOC Laboratory animals
AGROVOC Newborn animals
LCSH Working animals
LCSH Feral animals
LCSH Nocturnal animals

Filter the result
- ☑ AGROVOC
- ☑ LCSH
- ☑ NBII
- ☑ MeSH

AGROVOC->Aquatic animals

View in SKOS

Preferred Label	Aquatic animals
URI	http://www.fao.org/aos/agrovoc# c_552

HIVE
Vocabulary Server

Helping with Interdisciplinary Vocabulary Engineering

Home | Concept Browser | Indexing

HIVE vocabulary server provides functionality to identify concepts from given document or text. You need only two easy steps to get the concepts that are relevant to document:

- Step 1:Select the vocabulary source
- Step 2:Upload your document OR Enter the URL of your document
- Step 3:Click on Start Processing

HIVE Automatic Concepts Extractor

1 Select vocabulary source — Select

2 Upload a document — Choose File no file selected — Upload

OR Enter the URL

Hide advanced settings

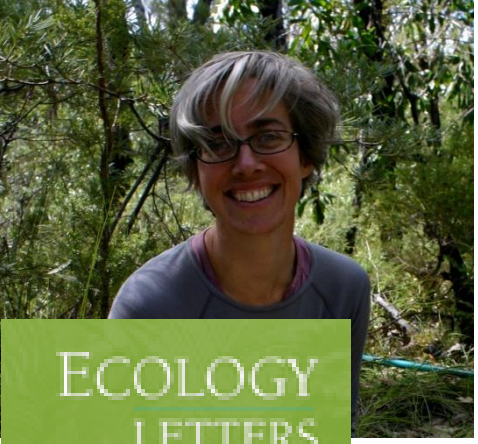
0 Number of hops

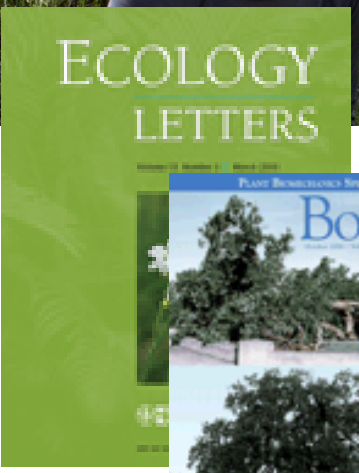
10 Maximum number of terms

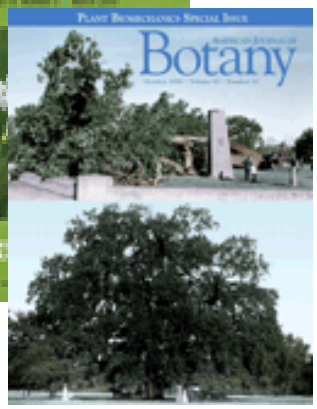
3

Start Processing

Powered by KEA







~~~~Amy

- Meet Amy Zanne. She is a botanist.
- Like every good scientist, she publishes, and she deposits data in Dryad.

Family	Binomial	A (mm^2)	F (mm^2/mm^2)	N (mm^-2)	S (mm^4)
Caprifoliaceae	Abelia biflora	0.002375829	0.924197654	389.0	6.10753E-06
Caprifoliaceae	Abelia dielsii	0.00115375	0.357418211	331.0	3.48565E-06
Caprifoliaceae	Abelia integrifolia	0.001134115	0.240432369	212.0	5.3496E-06
Caprifoliaceae	Abelia mosanensis	0.000855299	0.632065665	739.0	1.15737E-06
Caprifoliaceae	Abelia serrata	0.000706858	0.206402637	292.0	2.42075E-06
Caprifoliaceae	Abelia spathulata	0.000804248	0.230819095	287.0	2.80226E-06
Malvaceae	Abutilon fruticosum	0.001452201	0.137959114	95.0	1.52863E-05
Malvaceae	Abutilon pannosum	0.003117245	0.124689812	40.0	7.79311E-05
Fabaceae	Acacia albida	0.012271846	0.049087385	4.0	0.003067962
Fabaceae	Acacia ataxacantha	0.013069811	0.169907541	13.0	0.00100537
Fabaceae	Acacia borleae	0.004071504	0.061072561	15.0	0.000271434
Fabaceae	Acacia burkei	0.008992024	0.053952141	6.0	0.001498671
Fabaceae	Acacia caffra	0.010207035	0.214347725	21.0	0.000486049
Fabaceae	Acacia cyanophylla	0.009160884	0.201539452	22.0	0.000416404
Fabaceae	Acacia davyi	0.008332289	0.099987469	12.0	0.000694357
Fabaceae	Acacia erioloba	0.015174678	0.091048067	6.0	0.002529113
Fabaceae	Acacia erubescens	0.008824734	0.07059787	8.0	0.001103092
Fabaceae	Acacia exu	[illegible]	[illegible]5839	16.0	7.08822E-05
Fabaceae	Acacia galp	[illegible]	[illegible]0257	8.0	0.001509535
Fabaceae	Acacia gerr	[illegible]	[illegible]3581	7.5	0.001543255
Fabaceae	Acacia grar	[illegible]	[illegible]7175	7.0	0.000929126
Fabaceae	Acacia hae	[illegible]	[illegible]4417	19.0	0.000264555
Fabaceae	Acacia hebeclada	0.008659015	0.043295074	5.0	0.001731803
Fabaceae	Acacia hereroensis	0.003959192	0.047510306	12.0	0.000329933
Fabaceae	Acacia karroo	0.020867244	0.16693795	8.0	0.002608405
Fabaceae	Acacia luederitzii	0.007542964	0.105601495	14.0	0.000538783
Fabaceae	Acacia mangium	0.016933724	0.130928066	7.7	0.002208747
Fabaceae	Acacia melanoxylon	0.011976733	0.072419798	6.0	0.001996122
Fabaceae	Acacia mellifera	0.007697687	0.107767624	14.0	0.000549835
Fabaceae	Acacia montis-usti	0.005410608	0.043284864	8.0	0.000676326

Amy's data

Ecology Letters, (2009) **12**: 351–366 doi: 10.1111/j.1461-0248.2009.01285.x

REVIEW AND SYNTHESIS

Towards a worldwide wood economics spectrum

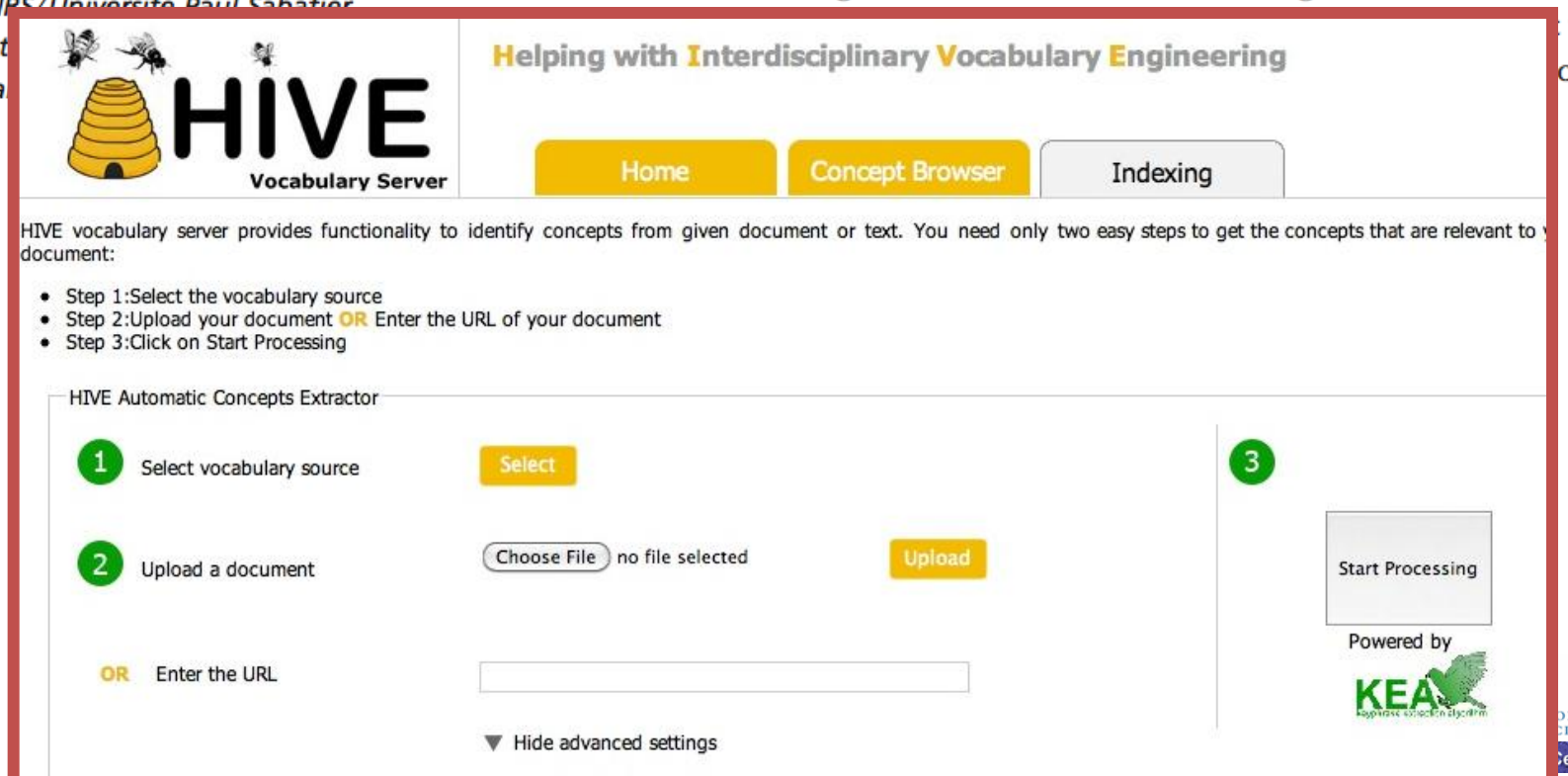
Jerome Chave,[1]* David Coomes,[2] Steven Jansen,[3] Simon L. Lewis,[4] Nathan G. Swenson[5] and Amy E. Zanne[6,7]

[1]*Laboratoire Evolution et Diversité Biologique, UMR 5174, CNRS/Université Paul Sabatier*
Bât
Fra

Abstract

Wood performs several essential functions in plants, including mechanically supporting aboveground tissue, storing water and other resources, and transporting sap. Woody tissues are likely to face physiological, structural and defensive trade-offs. How a plant optimizes among these competing functions can have major ecological implications, which have been under-appreciated by ecologists compared to the focus they have given to leaf function. To draw together our current understanding of wood function, we

wood
omical



Ecology Letters, (2009) **12**: 351–366 doi: 10.1111/j.1461-0248.2009.01285.x

REVIEW AND SYNTHESIS

Towards a worldwide wood economics spectrum

Jerome Chave,[1]* David Coomes,[2] Steven Jansen,[3] Simon L. Lewis,[4] Nathan G. Swenson[5] and Amy E. Zanne[6,7]

[1]*Laboratoire Evolution et Diversité Biologique, UMR 5174, CNRS/Université Paul Sabatier Bâtiment 4R3 F-31062 Toulouse, France*

Abstract

Wood performs several essential functions in plants, including mechanically supporting aboveground tissue, storing water and other resources, and transporting sap. Woody tissues are likely to face physiological, structural and defensive trade-offs. How a plant optimizes among these competing functions can have major ecological implications, which have been under-appreciated by ecologists compared to the focus they have given to leaf function. To draw together our current understanding of wood function, we identify and collate data on the major wood functional traits, including the largest wood density database to date (8412 taxa), mechanical strength measures and anatomical

Usability

Formal usability study 4 biologist, 5 information professionals

~ Tasks, usability ratings, satisfaction ranking

- Average time to search a concept:

  Librarians: 6.53 minutes

  Scientists: 3.82 minutes

  ~ consistent w/research at NIEHS, 2 times as long

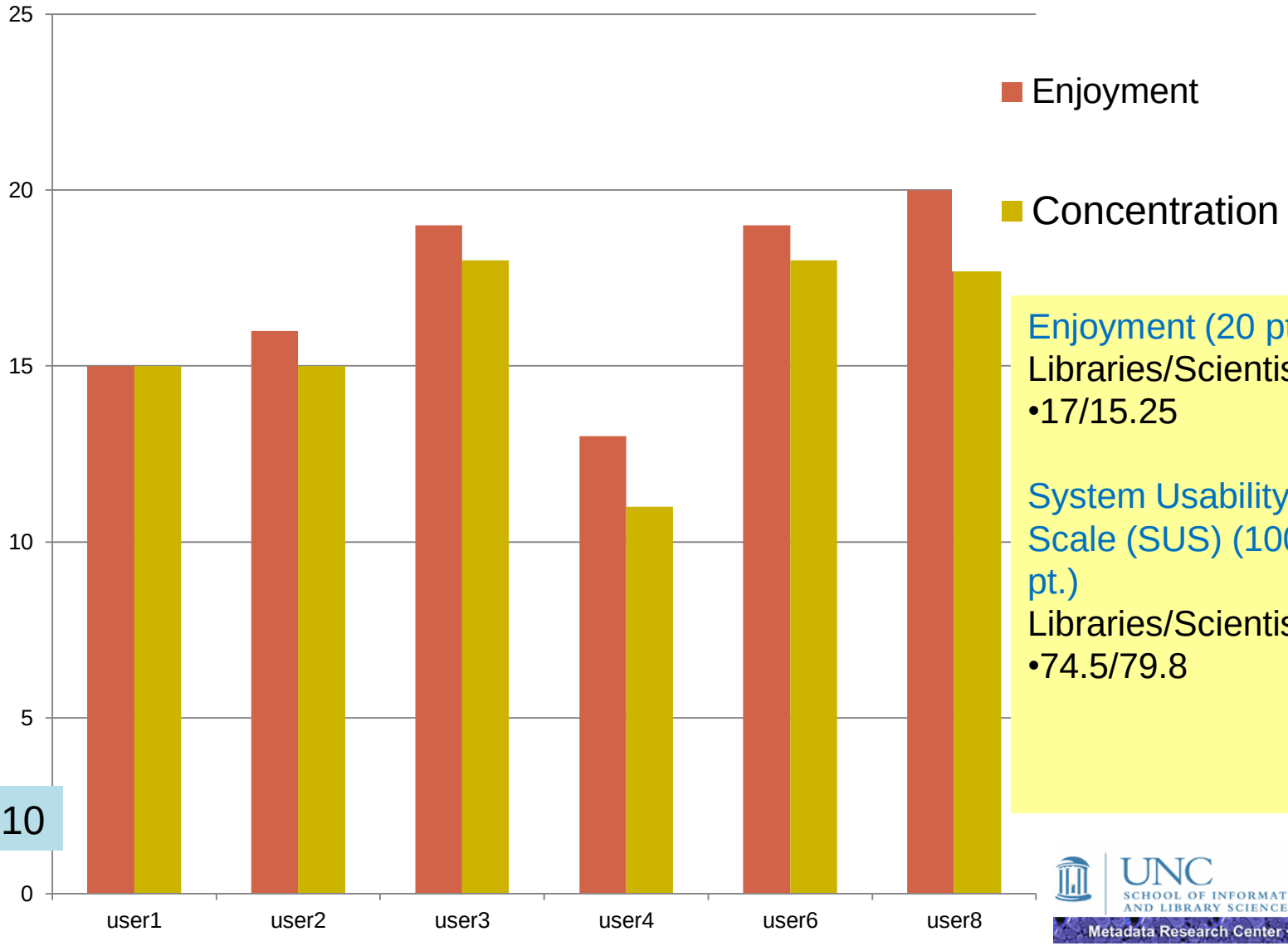
- Average time for automatic indexing sequence

  Librarians: 1.91 minutes

  Scientists: 2.1 minutes

Huang, 2010

System usability and flow metrics



Enjoyment (20 pt.)
Libraries/Scientists
•17/15.25

System Usability Scale (SUS) (100 pt.)
Libraries/Scientists
•74.5/79.8

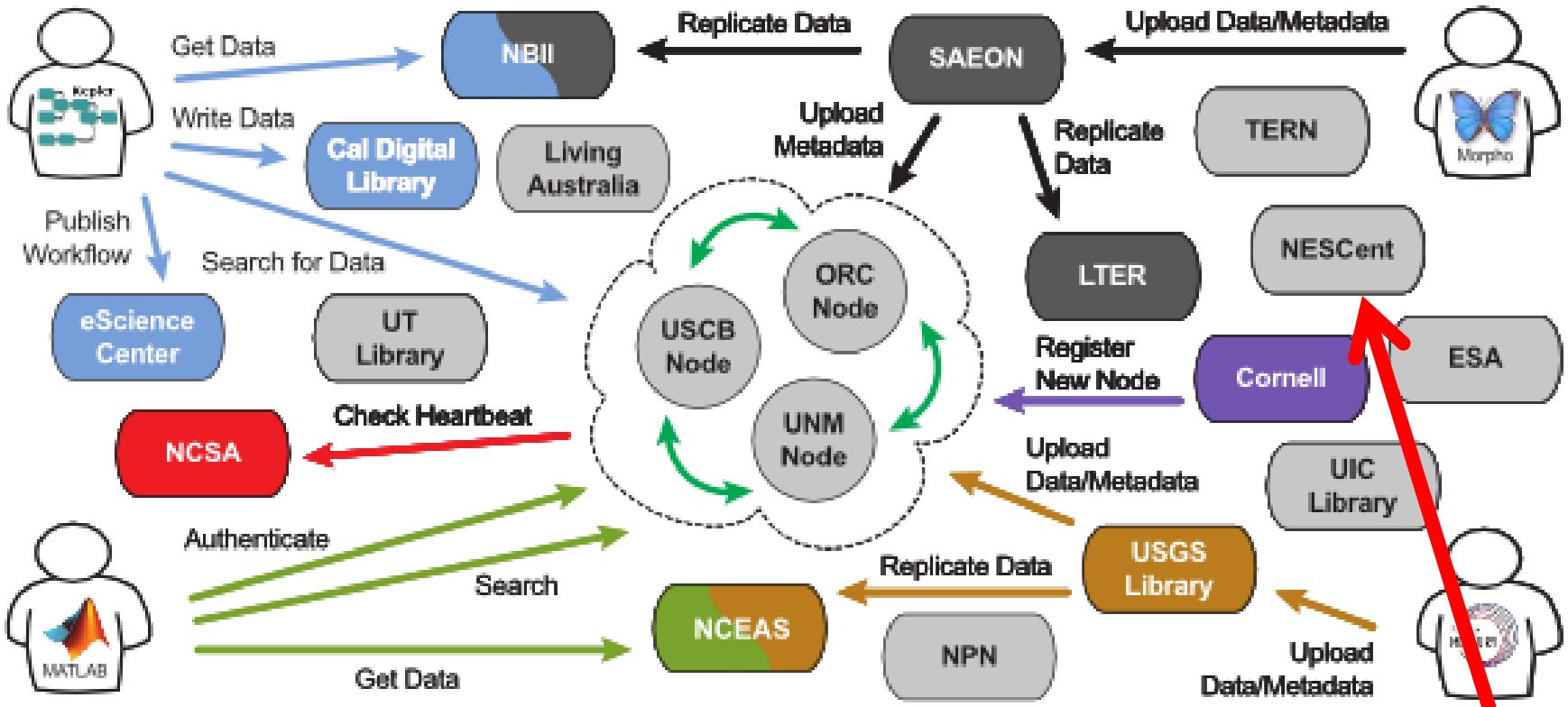
Huang, 2010

Challenges

- Building vs. doing/analysis
- Interoperability = dumbing down ontologies
- Proof-of-concept/ illustrate the differences between HIVE and other vocabulary registries (NCBO and OBO Foundry)
- People wanting a service
- General large team logistics, and having people from multiple disciplines (also the ++)

DataONE

Data Observation Network for Earth (DataONE)DataONE



- Distributed framework for sustainable cyberinfrastructure
- Science and society support ~ open, persistent, robust, and secure access to well-described, easily discovered Earth observational data.

Overriding goals and objectives

Develop an approach supporting metadata ownership; community driven

(fairly applied)

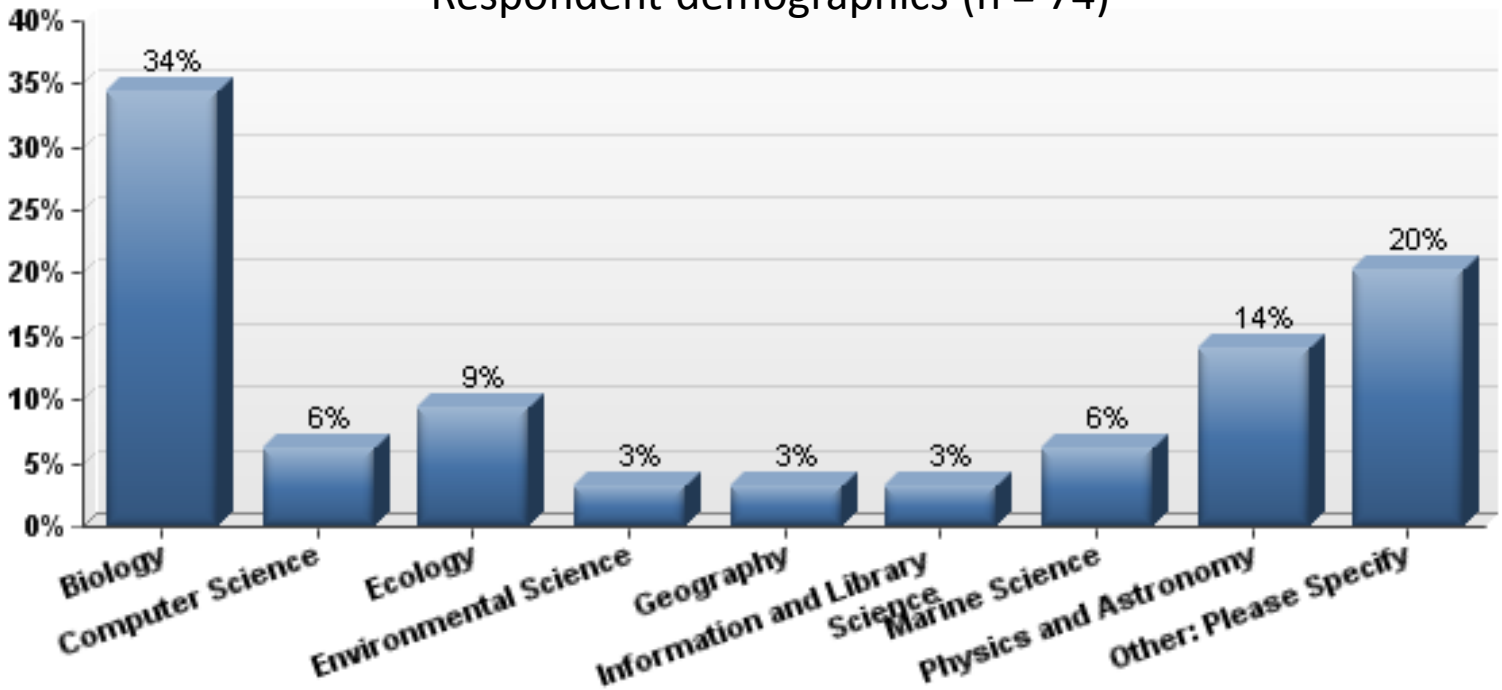
Evaluate ownership impact:

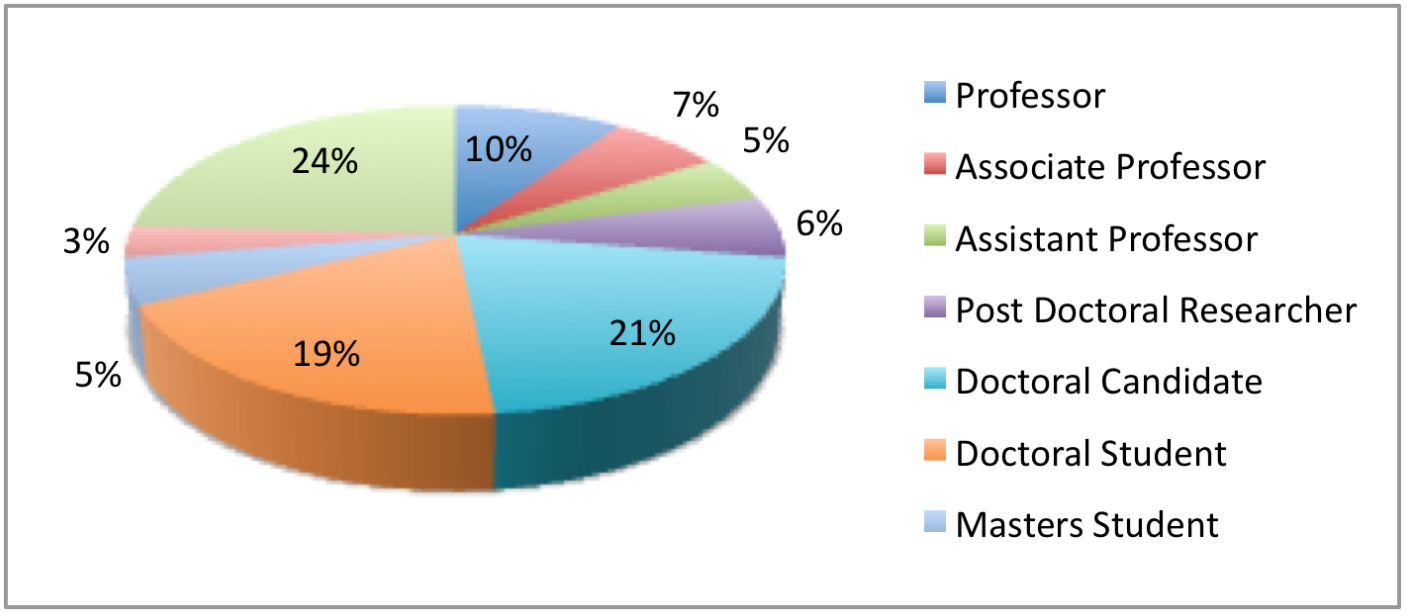
1. on empowerment and sustainability
2. as a complement to predominant metadata approaches
3. for DataONE interoperability and data reuse

DataONE summer intern program
DataONE Preservation and Metadata WG (PAMWG)

- Components of Successful Metadata Registry Frameworks (**A. Murillo**, J. Greenberg, & J. Boone, MRC/SISL/UNC-CH, and J. Kunze, CDL)
- Multi-method approach
  1. Literature review
     - confirm value; necessary for long tail science
  2. Registry evaluation
     - Access and services most important
     - Challenge in access → limited use; workflow
  3. Survey

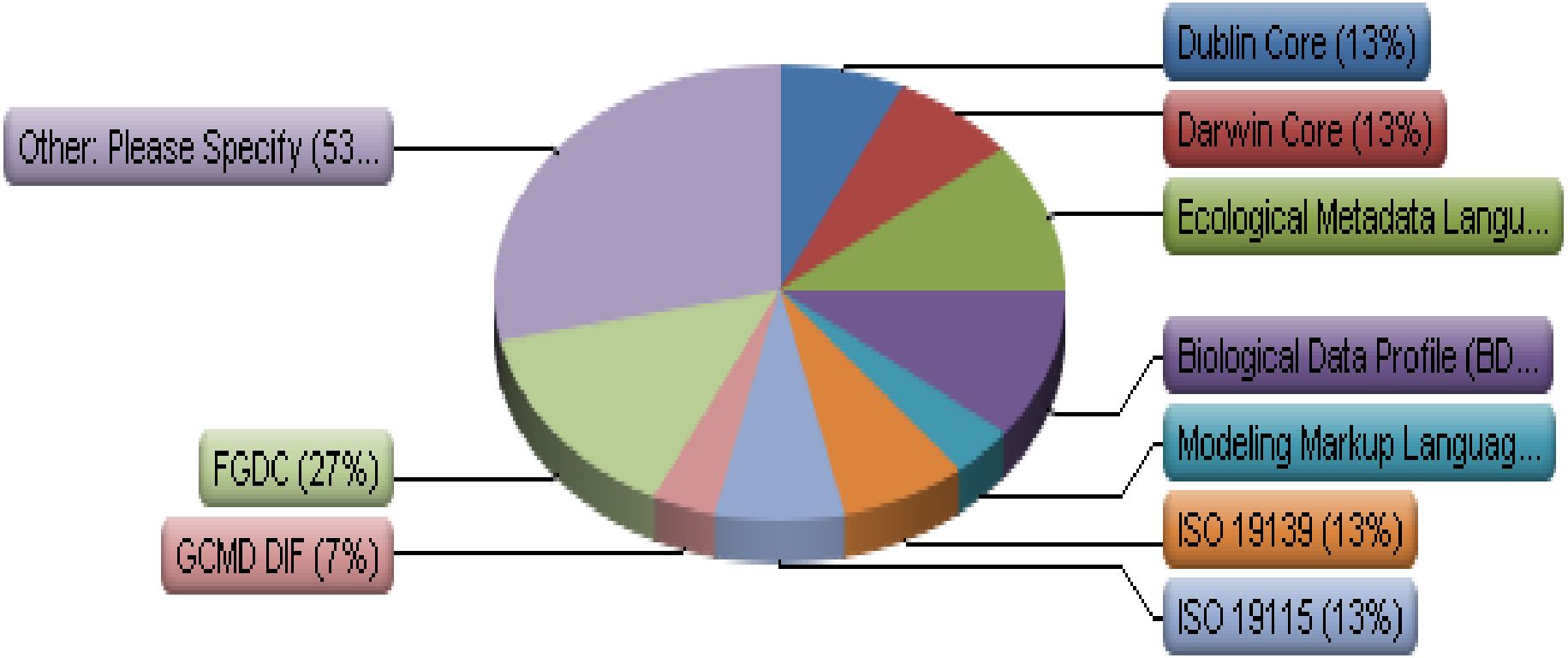
Respondent demographics (n = 74)





Types of metadata created: descriptive, administrative, etc.

14 standardized schemes used, lots of in-house



Metadata Vision

- One dictionary
- Crowd sourced plus lightly supervised canon
- Anyone can look up terms
- Any part of “metadata speech”
- Anyone can propose and refine their terms
- Strong terms rise, weak terms decline

DataONE all hands Sept. 17-20, Albuquerque New Mexico

- Met
- Laughed, Talked, Cried, Hugged
- Conquered

Translating a vision to principles

Low barrier for contributions.
Transparency in the review process.
Collective review, with rotating responsibilities among community members (scientists, developers, organizations, curators, etc.)
Consideration of elders (experts) to guide the review process and maintain thoughtful, balanced discussion.
Voting capacity of all users on the candidacy of terms submitted and their use.
Collective ownership of any user or organization.
Stakeholder engagement in the design and review process.

Prototyping

- Collective ownership
- Voting
- Good rises to the top
- Tracks history

- How to populate?
- How to ensure + sustain ownership?
- How to measure?

Support and contradictions

- Support
  - Data on the Long-tail
  - National and international data sharing policies
- Contradictions
  - NASA scientists
  - Global data meeting (US/EU)
    - Simple set → need detailed metadata

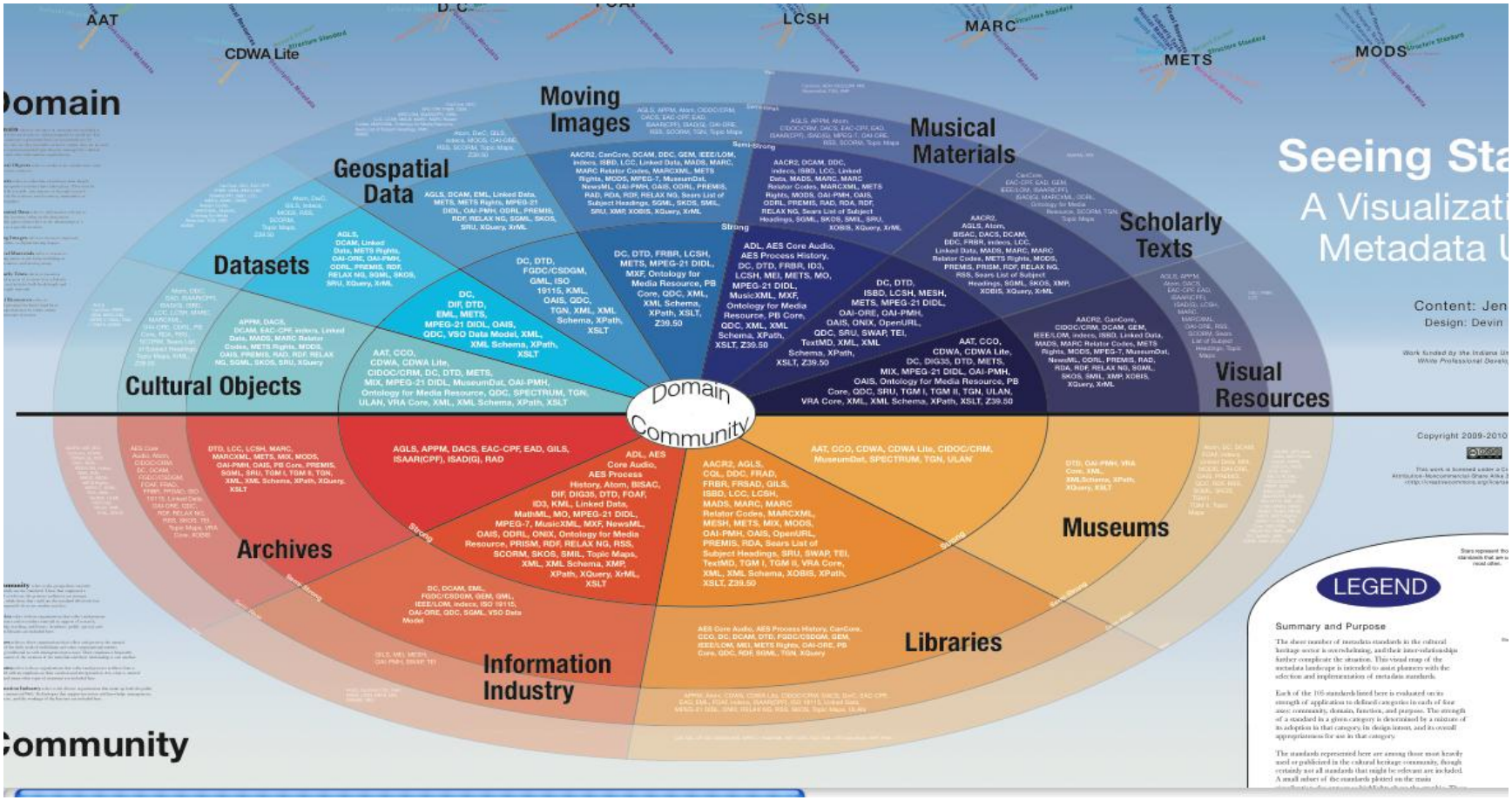
Outline

1. Assumptions
2. Motivation
3. Overriding goals and objectives
   - Dryad
   - HIVE
   - DataONE—PAMWG work
4. **Conclusions and framing questions**
5. Q&A

Assumptions

1. Prevailing metadata generation methods result in advantages and limitations
2. More than one way to skin a cat
   - Complementary, alternative approaches
   - Social technology
3. Ownership appeal
   - Empowerment and sustainability

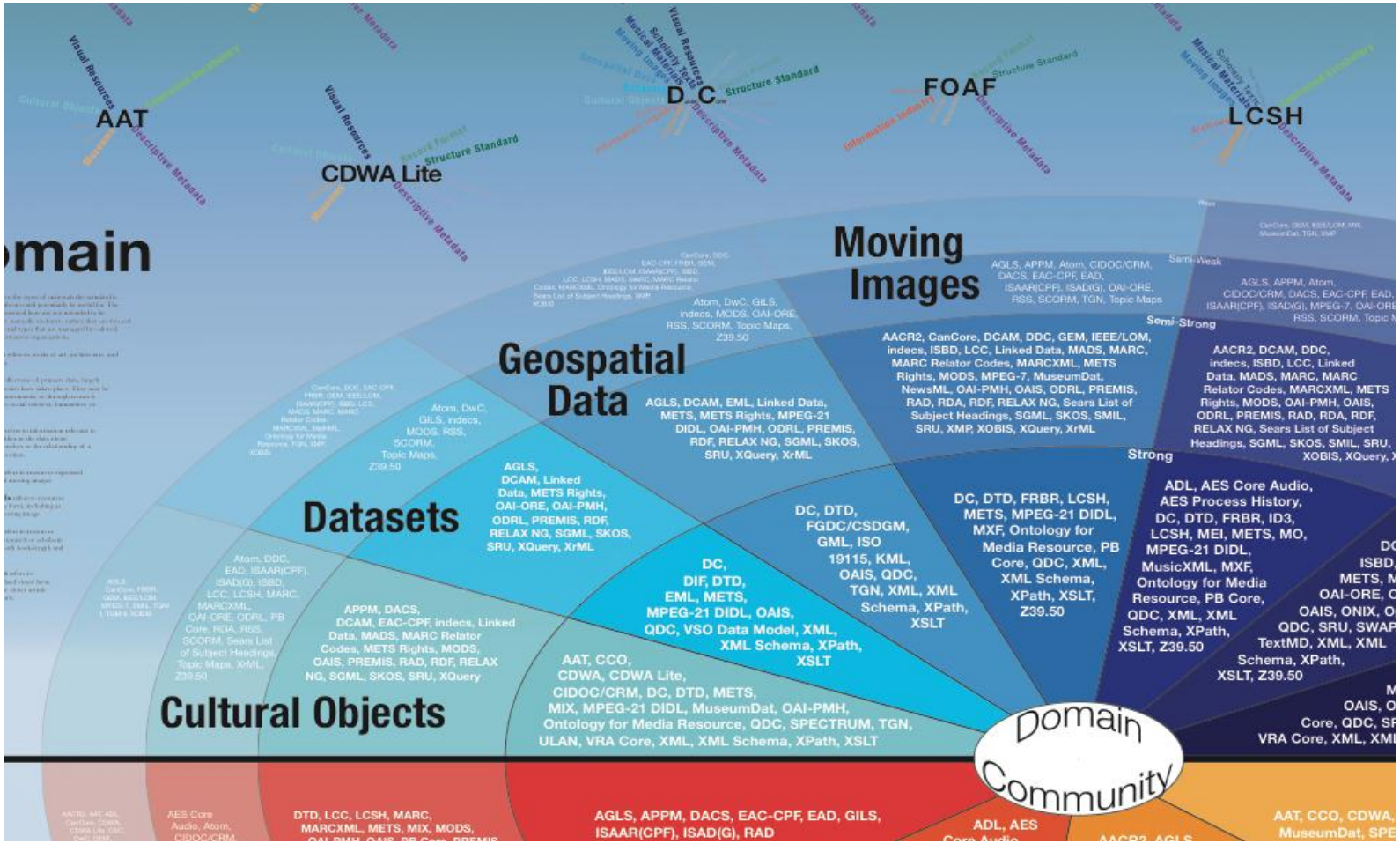
The Metadata Universe



Jenn Riley, Metadata Universe

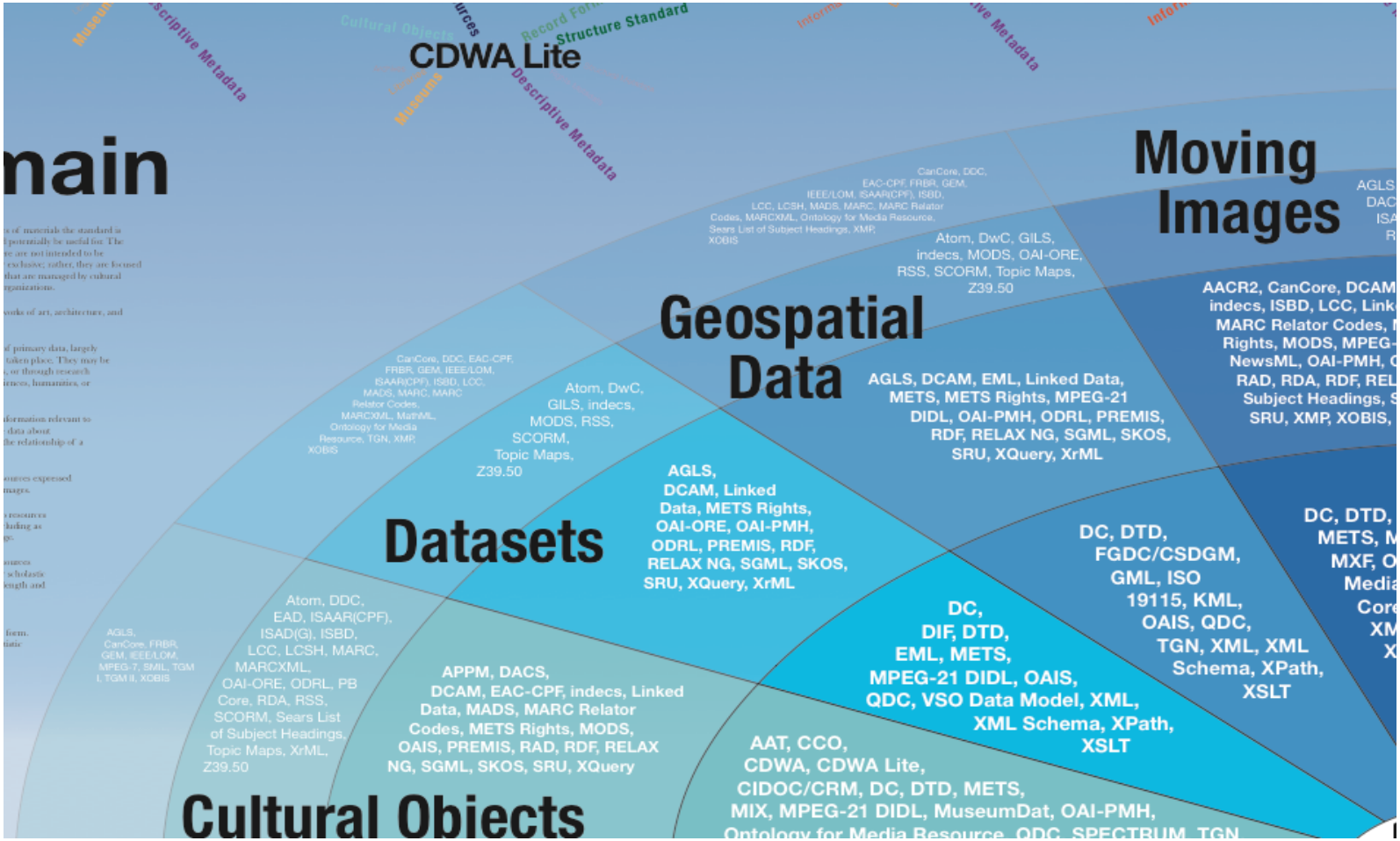
credit to John Kunze, CDL for this slide, and next 3

The Metadata Universe



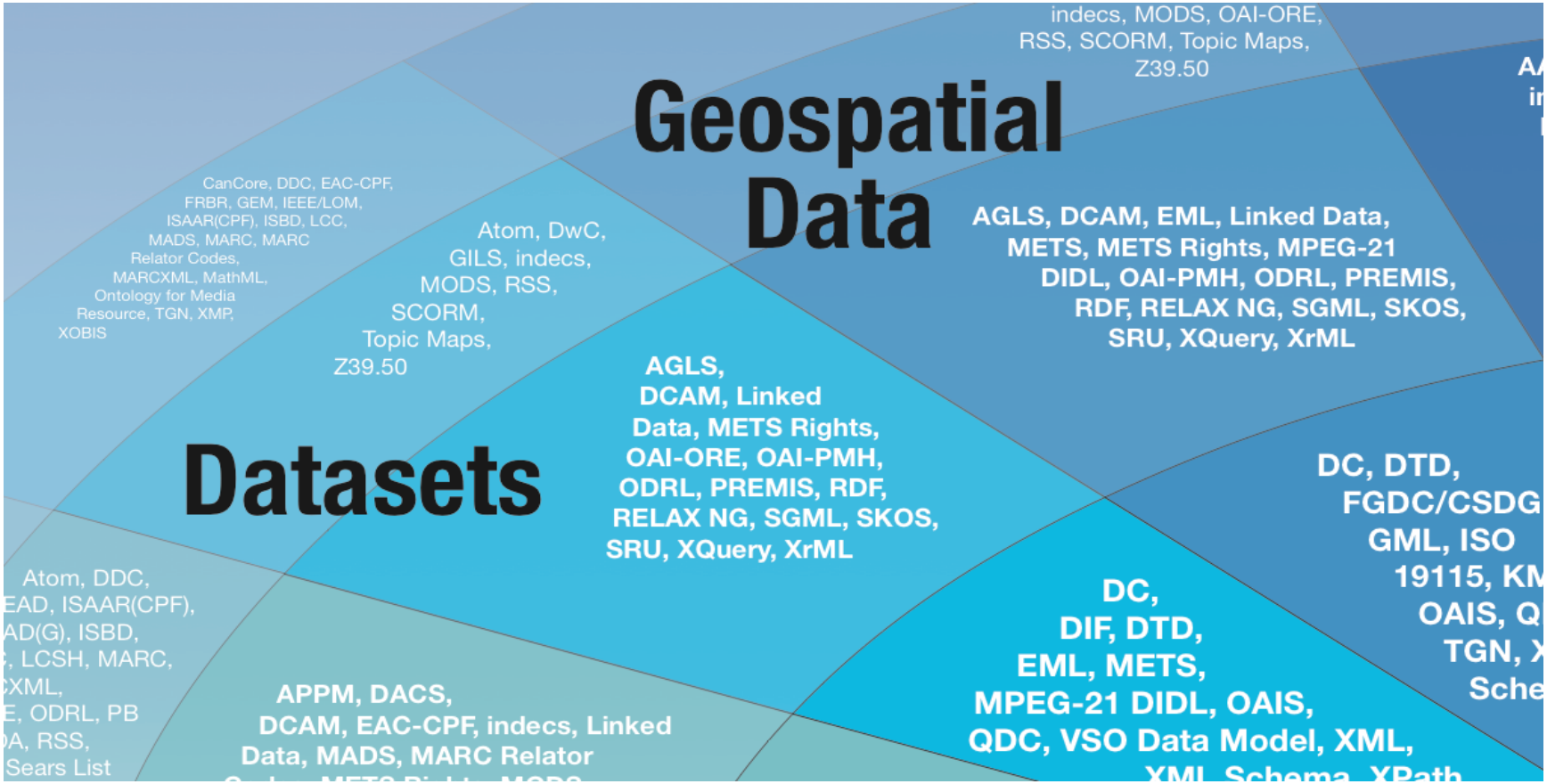
 Jenn Riley

The Metadata Universe



Jenn Riley

The Metadata Universe



 Jenn Riley

Framing questions

- What next…?
- How can we encourage scientists to:
  - Generate metadata?
  - Care about metadata quality?
  - Follow standards?
- Is there a threshold of expectation?
- Where do automatic applications best fit?
- How can we study this topic more?
  - Need to balance applied work and basic research

Concluding comments

- A contribution, have to start somewhere…
  - Good timing, the right discipline
- Confirmed success on some level
- Machine capabilities, eScience/data synthesis
- An educative commons, intellectually engaging

Acknowledgements *all around*

Dryad Acknowledgments

- Dryad Consortium Board, journal partners, and data authors
- NESCent: Kevin Clarke, Hilmar Lapp, Heather Piwowar, Peggy Schaeffer, Ryan Scherle, Todd Vision (PI)
- UNC-CH <Metadata Research Center>: Jose R. Pérez-Agüera, Sarah Carrier, Elena Feinstein, Lina Huang, Robert Losee, Hollie White, Craig Willis
- U British Columbia: Michael Whitlock
- NCSU Digital Libraries: Kristin Antelman
- HIVE: Library of Congress, USGS, and The Getty Research Institute; and workshop hosts
- Yale/TreeBASE: Youjun Guo, Bill Piel
- DataONE: Rebecca Koskela, Bill Michener, Dave Veiglais, and many others
- British Library: Lee-Ann Coleman, Adam Farquhar, Brian Hole
- Oxford University: David Shotton

HIVE Acknowledgments

- HIVE Development Team
- Dryad Repository Team
- Former SILS Masters students: Lina Huang and Jacquelynn Sherman

National Science Foundation
WHERE DISCOVERIES BEGIN

INSTITUTE of Museum and Library SERVICES

DRYAD

UNC SCHOOL OF INFORMATION AND LIBRARY SCIENCE
Metadata Research Center <MRC>

PAMWG Acknowledgments

U.S. National Science Foundation (Grant #OCI-0830944).

DataONE

The Subpopulations and Intermediate Outcomes in COPD Study (SPIROMICS) is funded by contract from the National Heart, Lung, and Blood Institute (NHLBI) to the University of North Carolina (HHSN268200900020C).

Metadata Subgroup Members

- Co Chairs: J.Greenberg + J.Kunze
- Sarah Callaghan, British Atmospheric Data Centre
- Greg Janee, Digital library research specialist, Earth Research Institute, University of California Santa Barbara
- Nassib Nassar, Senior Research Scientist, Senior Research Scientist RENCI, University of North Carolina Chapel Hill
- Angela Murillo, UNC-CH
- Karthik Ram, Environmental Science, Policy & Management, University of California Berkeley
- Jim Regetz, National Center for Ecological Analysis and Synthesis (NCEAS)
- Tim Robertson, Global Biodiversity Information Facility (GBIF)