

DataDryad.org and the interoperability continuum.

Repositories and Interoperability

2nd National Data Service Consortium Workshop (NDS2)
October 24, 2014



Jane Greenberg
Professor, College of Computing & Informatics
Director, Metadata Research Center

**Dryad...a curated
general-purpose
repository...makes
data discoverable,
freely reusable,
and citable.**

"...enables scientists to
validate published findings,
explore new analysis
methodologies, repurpose
data for research questions
unanticipated by the
original authors, and
perform synthetic studies."
(<http://datadryad.org/>)



This

**Not
this →**



Author



prepare manuscript and related data files



DRYAD

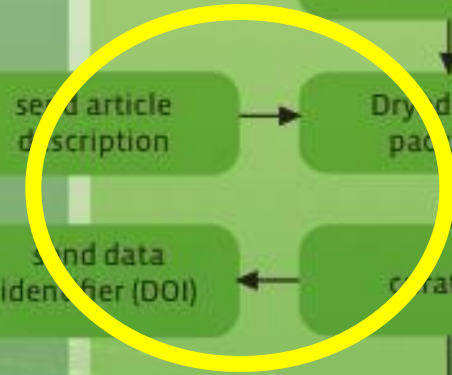
upload data

Dryad data package

creation

data curator

published data (with article citation)



JDAP

The American Naturalist

EVOLUTION

ECOLOG

BioRisk

JOURNAL OF Evolutionary Biology

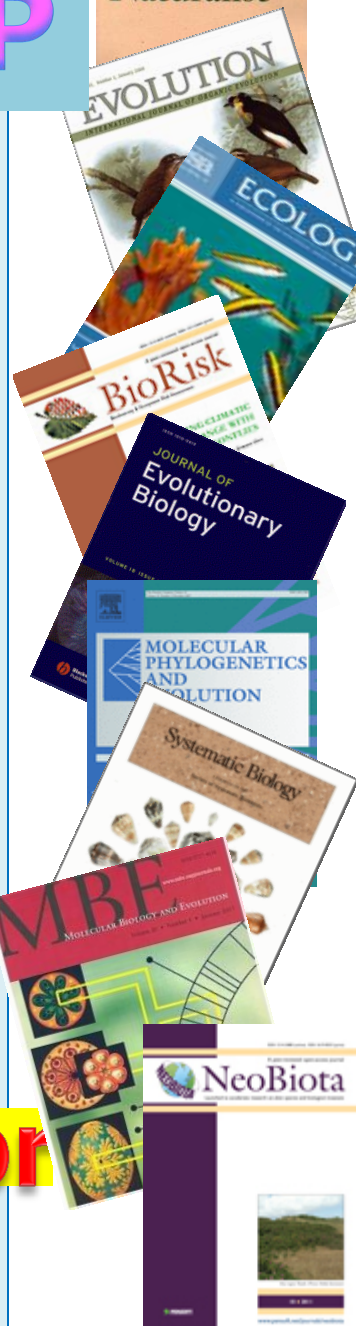
MOLECULAR PHYLOGENETICS AND EVOLUTION

Systematic Biology

MBE Molecular Biology and Evolution

NeoBiota

Curator



workflows

Journal	Re. Wrkfl	Blackout
AmNtrl	N	N
MBE	N	N
BioRisk	Y	N
BMJ Open	Y	N
....		Y

- Journals (77...PLOS): <http://datadryad.org/pages/integratedJournals>
- **X > 10GB** = \$15, \$10+
- Google Dryad FAQ

statistics

Type	Total	30 days
Data packages	6564	209
Data files	19985	773
Journals	359	80
Authors	23251	2905
Downloads	599566	15555



Interoperability

Technology

DSpace

DOIs via CDL/DataCite

CC0 (<m> + data)

Integration with specialized repositories and databases

- Federated searching with TreeBASE and KNB LTER
- TreeBASE submission (OAI-PMH)
- GenBank (currently in development)

Governance

"non-profit status, 12 member Board of Directors"

- Sets policy, goals
 - science, journals, societies, OCLC, MS
- **2006** Dryad development – NESCent + <MRC>
 - Stakeholders: journals, publishers and scientific societies, and researchers.
- **2009-2012**: Interim Board

\$ PAYMENT-Sept. 1,2014

...about interoperability



DREXEL UNIVERSITY
**Metadata
Research Center**
College of Computing & Informatics

The metadata hook

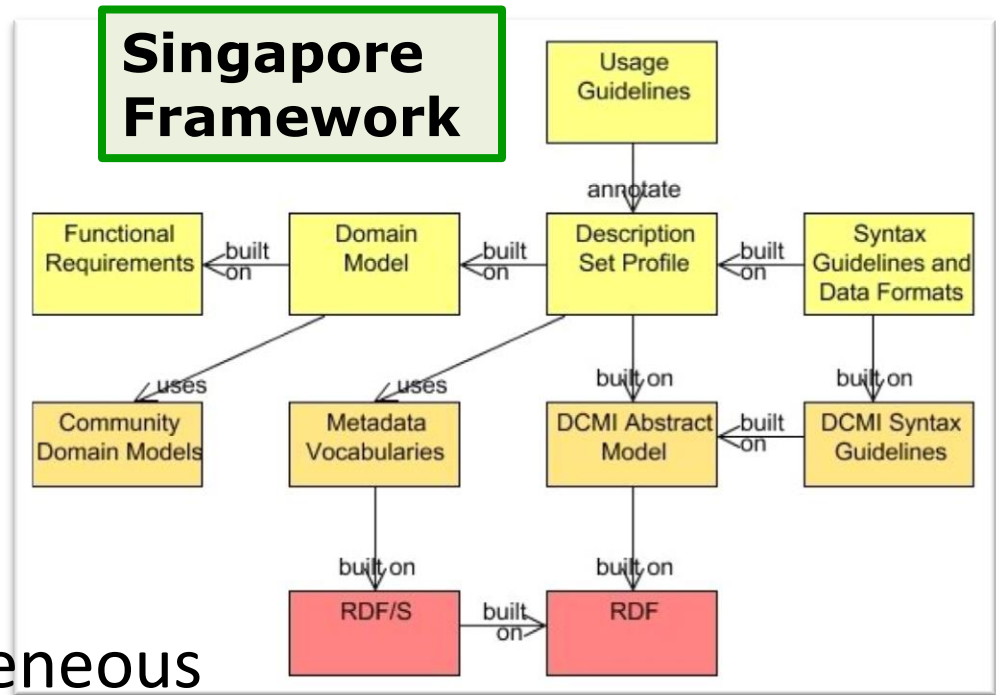


Dryad DCAP, ver. 3.0

- [bibo](#) (The Bibliographic Ontology)
- [dcterms](#) (Dublin Core terms)
- [dryad](#) (Dryad)
- [DwC](#) (Darwin Core)

Vision

1. **Simple:** automatic metadata gen; heterogeneous datasets ***Data-package centric**
2. **Interoperable:** harvesting, cross-system searching
3. **Semantic Web compatible:** sustainable; supporting machine processing



Greenberg, et al, 2009, Metadata Best Practice for a Scientific Data Repository, JLM, **DOI:10.1080/19386380903405090.**

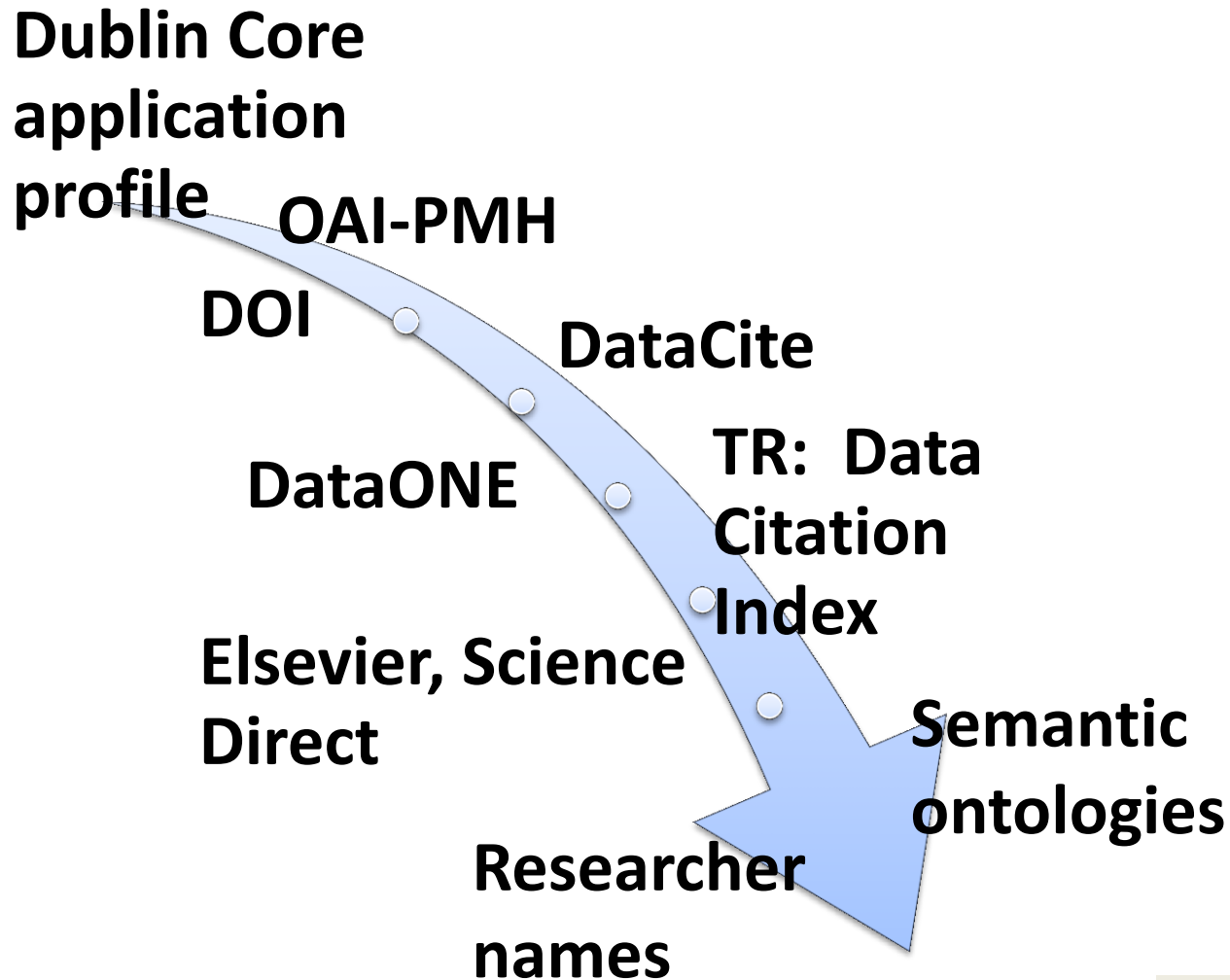
Metadata research & development

1. Curation workflow - [cognitive walkthroughs](#)
2. Dryad metadata scheme development - [crosswalk analyses](#) (Dube, et al, 2007; Carrier, et al, 2007; White et al., 2008, Greenberg, et al, 2010; Greenberg 2009; 2010)
3. Metadata reuse - [content analysis](#) (Greenberg, IDCC Research Summit, 2010)
4. Instantiation - multi-method study ([comprehensions assessment](#)) (Greenberg, RDAP, 2010, UNAM 2012)
5. Name-authority control - [exploratory study](#) (Haven, 2009, INLS 720)
6. KO/metadata community practices - Concurrent triangulation mixed methods ([survey + simulation experiment](#)) (White, 2010, ASIST, 2010 *JLM*)
7. Metadata functions - [quantitative categorical analysis](#) (Willis, Greenberg, and White, 2010, CODATA, 2012, JASIST)
8. Vocabulary needs (HIVE) - [mapping study](#) (Greenberg, 2009, CCQ; Scherle, 2010, Code4Lib)
9. Metadata theory - [deductive analysis](#) (Greenberg, 2009)





Interoperability continuum



Agency/
institution

Refine by: Author

[<< Previous Page](#)

Now showing items 31-40

[Next Page >>](#)

Tango, Jazzlyn M. (1)

Taniguchi, Fumiya (2)

Taper, Mark, L. (1)

Tarailo, David (1)

Taraschewski, Horst (1)

Tarone, Aaron (1)

Tartarini, Stefano (1)

Tarver, James E. (1)

Tarver, James (1)

Tattersall, Glenn J. (1)

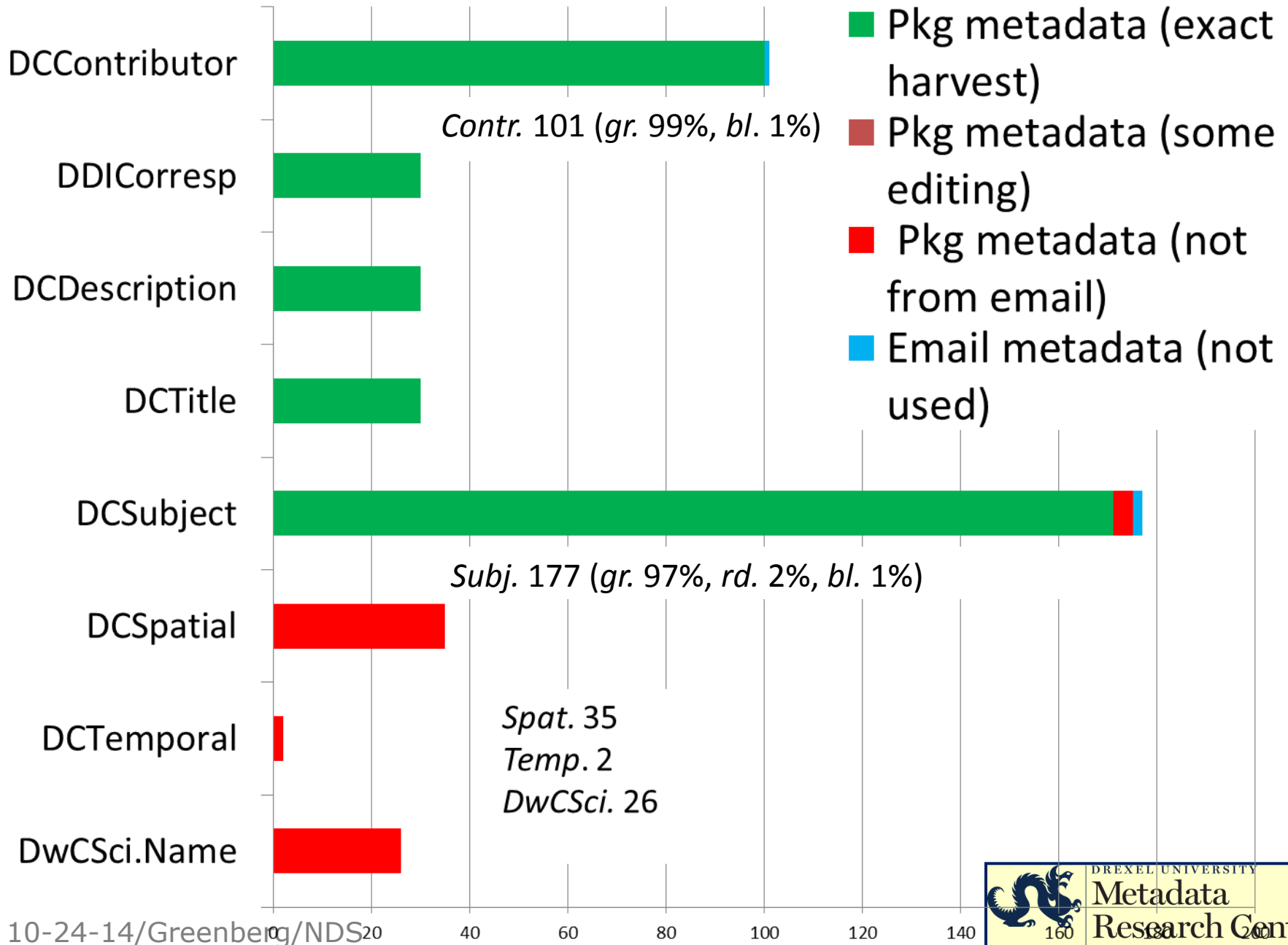
[<< Previous Page](#)

Now showing items 31-40

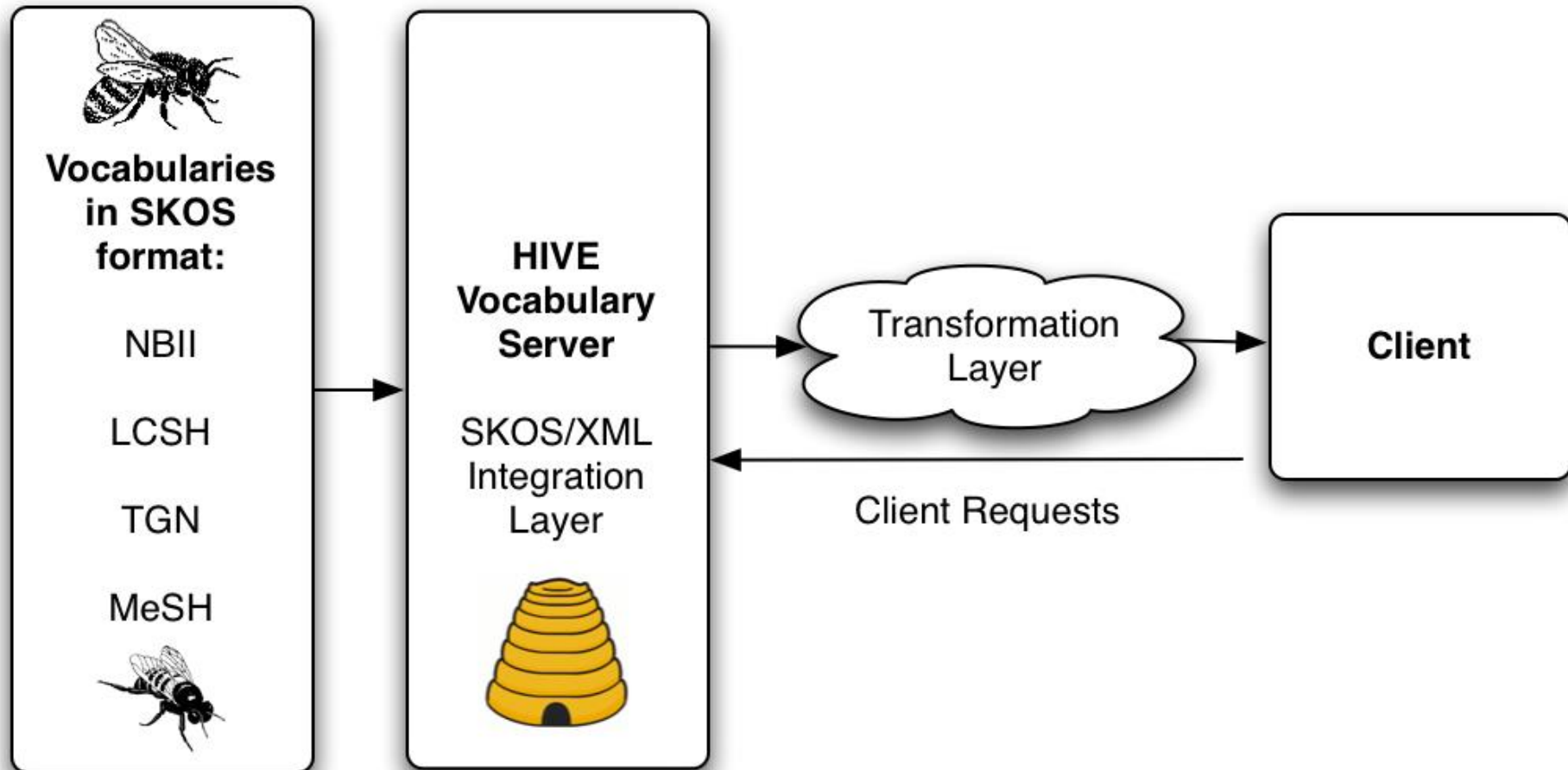
[Next Page >>](#)



Package metadata harvested from email



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....

REVIEW AND SYNTHESIS

Towards a worldwide wood economics spectrum

Jerome Chave,^{1*} David Coomes,²
Steven Jansen,³ Simon L. Lewis,⁴
Nathan G. Swenson⁵ and Amy E.
Zanne^{6,7}

¹Laboratoire 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

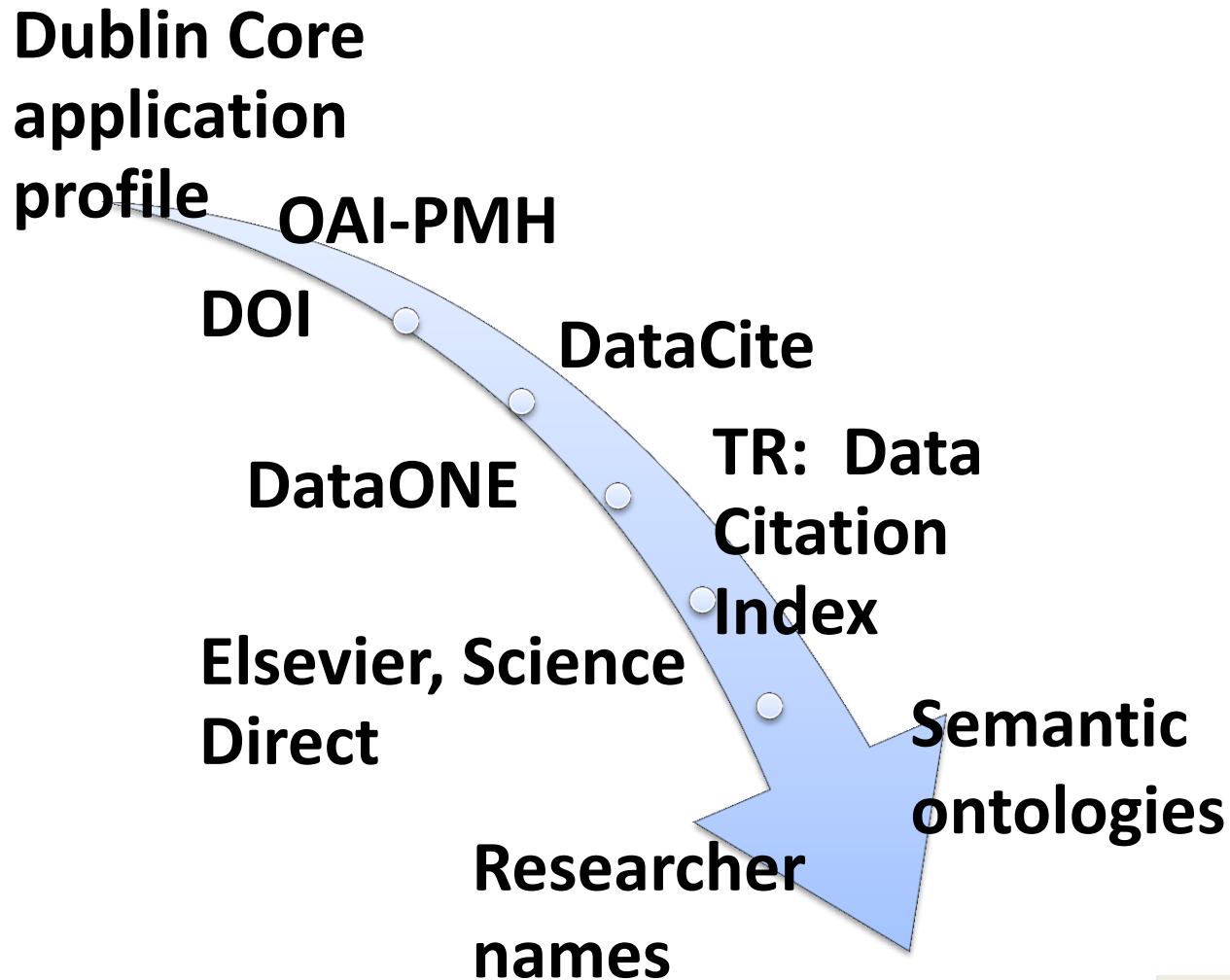
Extracted Concepts Cloud

AGROVOC
LCSH
NBII

Reaction wood Wood--Figure Wood--Discoloration Calavicci, AI (Fictitious character) Lāt,
al- (Arabian deity) Murphy, AI (Fictitious character) Density Soils--Density Population
density Recessive traits Traits (genetics) Dominant traits Associated species Species
diversity Numbers of species Plant anatomy Plant litter Plant condition Leaf
spots Leaf prints Leaf blowers Brushes, Carbon Electrodes, Carbon Carbon
taxes Growth Fetus--Growth Growth (Plants) Infiltration water Water--
Color Drinking water



Interoperability continuum



Agency/
institution


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



Concluding comments

- A contribution, have to start somewhere...
 - Good timing, the right discipline
- Confirmed use
- Machine capabilities
- An educative commons, intellectually engaging

An aerial photograph of a city, likely in the Middle East, featuring a large, modern white stadium with a curved roof. The stadium is situated near a road and a body of water. The background shows dense urban development and a clear blue sky.

<http://datadryad.org>
<http://blog.datadryad.org>
<http://datadryad.org/wiki>
<http://code.google.com/p/dryad>
dryad-users@nescent.org
Facebook: Dryad
Twitter: @datadryad
<http://ils.unc.edu/mrc/hive/>
<http://code.google.com/p/hive-mrc/>