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/)
### Journals

- **Journals (77...PLOS):**

- **X >10GB = $15,$10+**

- Google Dryad FAQ

### Statistics

<table>
<thead>
<tr>
<th>Type</th>
<th>Total</th>
<th>30 days</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data packages</td>
<td>6564</td>
<td>209</td>
</tr>
<tr>
<td>Data files</td>
<td>19985</td>
<td>773</td>
</tr>
<tr>
<td>Journals</td>
<td>359</td>
<td>80</td>
</tr>
<tr>
<td>Authors</td>
<td>23251</td>
<td>2905</td>
</tr>
<tr>
<td>Downloads</td>
<td>599566</td>
<td>15555</td>
</tr>
</tbody>
</table>
Interoperability

Technology
DSpace
DOIs via CDL/DataCite
CC0 (<m> + data)
Integration with specialized repositories and databases
  - Federated searching with TreeBASE and KNBLTER
  - 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
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
3. Metadata reuse - content analysis (Greenberg, IDCC Research Summit, 2010)
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)
8. Vocabulary needs (HIVE) - mapping study (Greenberg, 2009, CCQ; Scherle, 2010, Code4Lib)
9. Metadata theory - deductive analysis (Greenberg, 2009)
Interoperability continuum

- Dublin Core application profile
- OAI-PMH
- DOI
- DataCite
- DataONE
- TR: Data Citation Index
- Elsevier, Science Direct
- Researcher names
- Semantic ontologies
- Agency/institution
Package metadata harvested from email

**DCContributor**
- Contr. 101 (gr. 99%, bl. 1%)

**DDICorresp**
- Pkg metadata (exact harvest)

**DCDescription**
- Pkg metadata (some editing)

**DCTitle**
- Pkg metadata (not from email)

**DCSubject**
- Email metadata (not used)

**DCSpatial**
- Subj. 177 (gr. 97%, rd. 2%, bl. 1%)

**DCTemporal**
- Spat. 35
  - Temp. 2
  - DwCSci. 26

**DwCSci.Name**
Helping Interdisciplinary Vocabulary Engineering (HIVE)

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

**Building, Sharing, Evaluation** the HIVE....
Towards a worldwide wood economics spectrum

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

Reaction wood
Wood--Figure
Wood--Discoloration
Calaviccii, Al (Fictitious character)
Lat,
al- (Arabian deity)
Murphy, Al (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
Infiltration water
Carbon
taxes
Growth
Fetus--Growth
Color
Drinking water

LCSH

NBII
Interoperability continuum

- Dublin Core
- application profile
- OAI-PMH
- DOI
- DataCite
- DataONE
- Elsevier, Science Direct
- TR: Data Citation Index
- Semantic ontologies
- Researcher names
- 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