



Interoperable Thesauri: The Challenges and Experiences of the HIVE Project

CENDI/NKOS Workshop
National Agricultural Library
Beltsville, MD
October 22, 2009

JANE GREENBERG, PROFESSOR,
DIRECTOR OF THE METADATA RESEARCH CENTER
SCHOOL OF INFORMATION & LIBRARY SCIENCE
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL
janeg@email.unc.edu

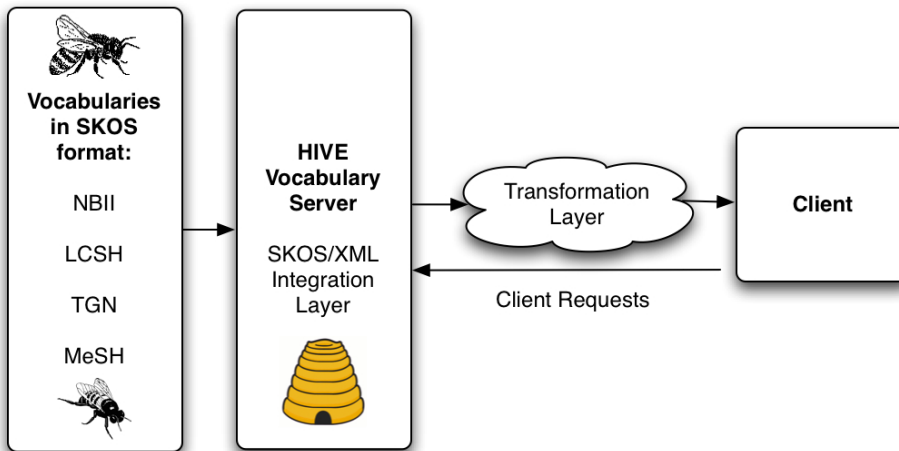


Overview

- HIVE—Helping Interdisciplinary Vocabulary Engineering
- Motivation—Dryad repository
- HIVE--Goals, status, and design
- Challenges
 - Technical and social
- Conclusion and questions



HIVE model

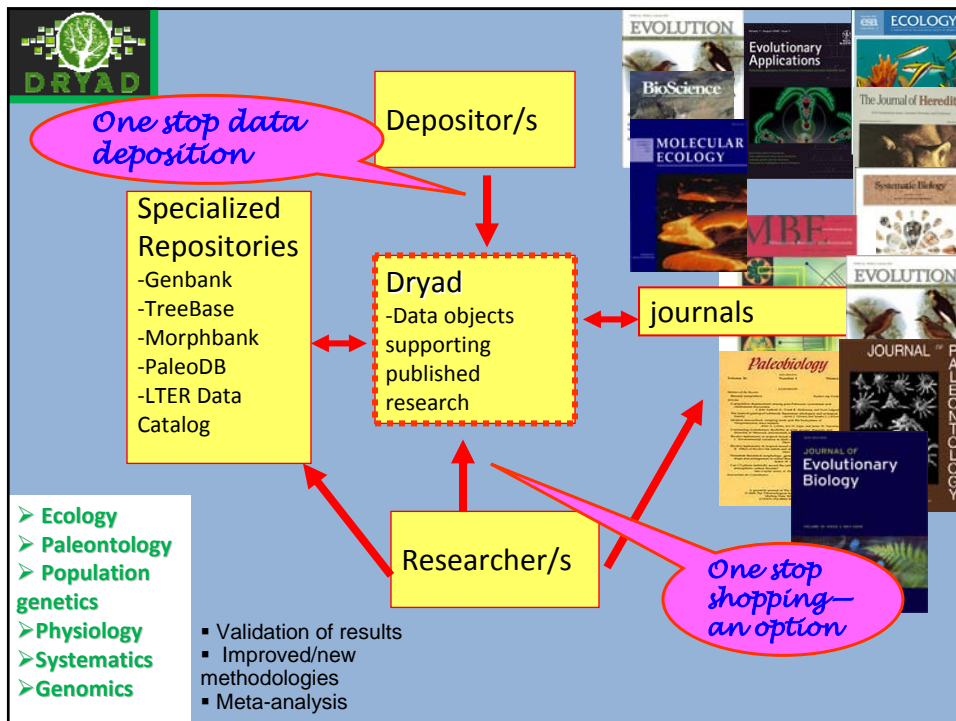
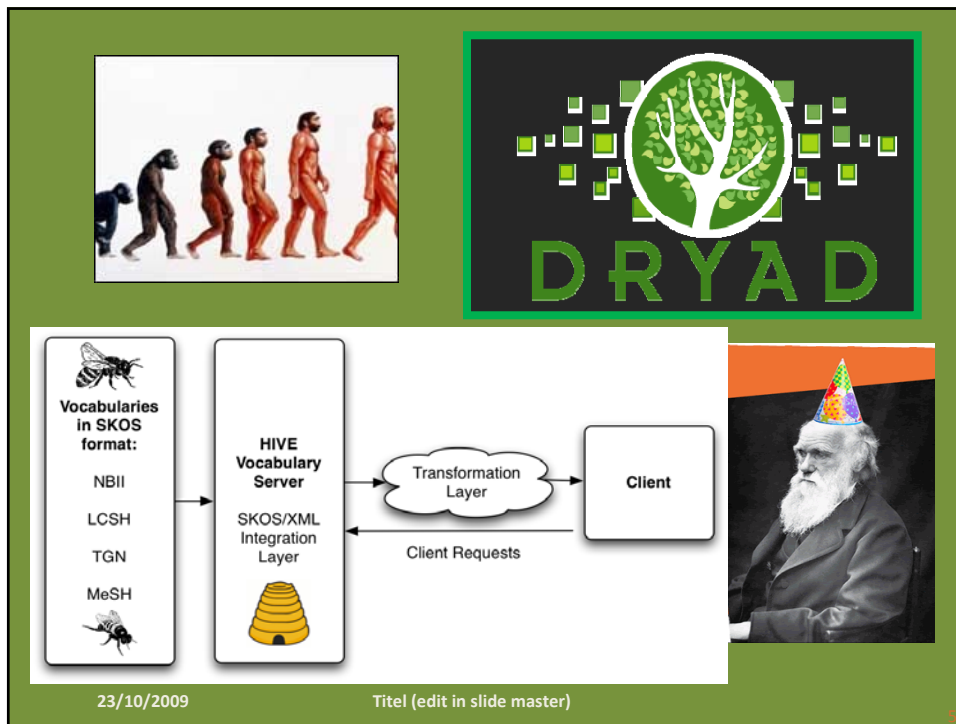


- <AMG> approach for integrating discipline CVs
- Model addressing C V cost, interoperability, and usability constraints (interdisciplinary environment)

3

Motivation





No *single* Evolutionary Biology Controlled Vocabulary



■ Vocabulary analysis

NBII Thesaurus, LCSH, the Getty's TGN, Gene Ontology

- 600 keywords, Dryad partner journals
Facets: taxon, geographic name, time period, topic
- 431 topical terms, exact matches
NBII Thesaurus, 25%; MeSH, 18%
- 531 terms
LCSH, 22% found exact matches, 25% partial
Need multiple vocabularies + which vocabularies



Dryad's Goals

1. One-stop deposition and shopping for data objects supporting published research...
2. Support the acquisition, preservation, resource discovery, and reuse of heterogeneous digital datasets
3. Balance a need for low barriers, with higher-level ... data synthesis

Dryad Team NESCent



- Hilmar Lapp
Assistant Director for Informatics
- Ryan Scherle
Data Repository Architect
- Todd Vision, Associate Director of Informatics

UNC/SILS/MRC

- Jane Greenberg, Professor
- Lina Huang, MSIS Student, Research Assistant
- Robert M. Losee, Professor
- Jose R. Pérez-Agüera, Clinical Assistant Professor
- Hollie White, Metadata Research Center Doctoral Fellow



North Carolina State University, University of New Mexico/LTER, Yale University, + partner journals and societies

HIVE Goals, design, and status



HIVE—Helping Interdisciplinary Vocabulary Engineering

- Address CV (controlled vocabulary) cost, interoperability, and usability constraints
 - Controlled vocabularies are expensive to create, maintain, and use
 - Controlled vocabularies very often developed in silos
 - Vocabulary usability problems, stemming from interface design and functionality limitations, have been well documented

> Mirror significant challenges faced in implementing CV system in Dryad / bioportal, OBO foundry, phenoscape

HIVE...as a solution



- Automatic metadata generation approach that dynamically integrates discipline-specific controlled vocabularies encoded with the [Simple Knowledge Organisation System \(SKOS\)](#)
 - *provide efficient, affordable, interoperable, and user friendly access to multiple vocabularies during metadata creation activities*
1. *Building HIVE*
 - *Vocabulary Development*
 - *Server preparation*
 - Primate Life Histories Working Group
 - Wood Anatomy and Wood Density Working Group
 2. *Sharing HIVE* continuing education
 3. *Evaluating HIVE* examining HIVE in Dryad



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*



Advisory Board



- Jim Balhoff, NESCent
- Libby Dechman, LCSH
- Mike Frame, USGS
- Alistair Miles, CCLRC Rutherford Appleton Laboratory
- 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, UNC Chapel Hill
- Lisa Zolly, USGS

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



HIVE Construction



- HIVE's technological infrastructure *stores millions of concepts from different vocabularies* and is preparing to make them available on the Web by a simple HTTP
- Vocabularies are imported in HIVE using SKOS/RDF format
- HIVE is divided in two different modules:
 - HIVE Core**
 - SKOS/RDF storage and management (SESAME/Elmo)
 - Automatic Metadata Extraction and Topic Detection (KEA++ and MAUI)
 - Concept Retrieval (Lucene and MG4J)
 - HIVE Web**
 - Web user Interface (GWT—Google Web Toolkit)
 - Machine oriented interface (SOAP and REST)

FROM NBII



```
<rdf:Description rdf:about="http://thesaurus.nbii.gov/Mud">  
<rdf:type  
rdf:resource="http://www.w3.org/2004/02/skos/core#Concept"/>  
<skos:broader rdf:resource="http://thesaurus.nbii.gov/Sediments"/>  
<skos:prefLabel>Mud</skos:prefLabel>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Clays"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Mud-flats"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Oozes"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Silt"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Slimes"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Sludges"/>  
<skos:related rdf:resource="http://thesaurus.nbii.gov/Soils"/>  
<skos:scopeNote>ASF Aquatic Sciences and Fisheries LSC Life  
Sciences</skos:scopeNote>  
</rdf:Description>
```



Check all headings that apply to this publication.
To see broader/narrower terms, click the link for the respective vocabulary.

Bechner JC, Nguyen N, Alberts SC, Altmann J. 2006. The endocrinology of pregnancy and fetal loss in wild baboons. *Hormones and Behavior* 49:688-699.

Abstract: An impressive body of research has focused on the mechanisms by which the steroid estrogens (E), progestins (P), and glucocorticoids (GC) ensure successful pregnancy. With the advance of non-invasive techniques to measure steroids in urine and feces, steroid hormones are routinely monitored to detect pregnancy in wild mammalian species, but hormone data on fetal loss have been sparse. Here, we examine fecal steroid hormones from five groups of wild yellow baboons (*Papio cynocephalus*) in the Amboseli basin of Kenya to compare the hormones of successful pregnancies to those ending in fetal loss or stillbirth. Using a combination of longitudinal and cross-sectional data, we analyzed three steroid hormones (E, P, GC) and related metabolites from 5 years of fecal samples across 188 pregnancies. Our results document the course of steroid hormone concentrations across successful baboon pregnancy in the wild and demonstrate that fecal estrogens predicted impending fetal loss starting 2 months before the externally observed loss. By also considering an additional 450 pregnancies for which we did not have hormonal data, we determined that the probability for fetal loss for Amboseli baboons was 13.9%, and that fetal mortality occurred throughout gestation (91 losses occurred in 656 pregnancies; rates were the same for pregnancies with and without hormonal data). These results demonstrate that our longstanding method for early detection of pregnancies based on observation of external indicators closely matches hormonal identification of pregnancy in wild baboons.

Keywords: Fetal loss; Miscarriage; Fecal steroids; Estrogens; Progesterins; Glucocorticoids; Baboon; Papio; Pregnancy

- Abortion, Spontaneous [USE FOR Miscarriage] ([MESH](#))
- Amboseli National Park ([TGN](#))
- Baboon (Musical group) ([LCSH](#))
- Baboon Creek ([TGN](#))
- Baboons ([LCSH](#))
- Estrogens ([NBII](#), [MESH](#))
 - Broader: Sex hormones
 - Narrower: Phytoestrogens
 - Related: Estrus
- Estrogens, Catechol ([LCSH](#))
- Glucocorticoids ([MESH](#), [LCSH](#))
- Kenya ([TGN](#))



HIVE vocabulary server
Sign in | Register | About HIVE | Help | Contact Us | Feedback

Help with Interdisciplinary Vocabulary Engineering

Home
Concept Browser
Indexing

Welcome to HIVE Vocabulary Server!

Helping Interdisciplinary Vocabulary Engineering (HIVE) is an IMLS funded project involving the Metadata Research Center (MRC) at the School of Information and Library Science, University of North Carolina at Chapel Hill, and the National Evolutionary Synthesis Center (NESCent) in Durham, North Carolina. HIVE is an automatic metadata generation approach that dynamically integrates discipline-specific controlled vocabularies encoded with the Simple Knowledge Organisation System (SKOS), a World Wide Web Consortium (W3C) standard. HIVE Vocabulary Server is a web based system for searching and browsing concepts in interdisciplinary vocabularies, and providing cataloging aids by automatically extracting concepts for a given document.

Search Concept

Search

[Go to Concept Browser](#)

Annotate Document

Upload

[Go to Annotation](#)

Formal Queries

SPARQL
 SERQL

Execute

Vocabulary Statistics

Vocabulary	Concepts	Relationships	Date Added
LCSH	10,236	5,635	08-13-2009
NBII	21,129	5,635	08-13-2009
GRI	21,129	5,635	08-13-2009
USGS	21,129	5,635	08-13-2009
MeSH	21,129	5,635	08-13-2009
ITIS	21,129	5,635	08-13-2009
WordNet	21,129	5,635	08-13-2009
TGN	21,129	5,635	08-13-2009
UBIO	21,129	5,635	08-13-2009
Gene Ontology	21,129	5,635	08-13-2009

Last updated: 08-21-2009

HIVE All Rights Reserved.

HIVE web interface HIVE web service Sign in | Register About HIVE Help Contact Us Feedback

HIVE Vocabulary Server

Help with Interdisciplinary Vocabulary Engineering

Home Concept Browser Indexing

Current vocabularies sources: X LCSH X NBII X MeSH + Add

Searching for Concept

ABCDEFGHIJKLMNOPQRSTUVWXYZ
abcdefghijklmnopqrstuvwxyz

LCSH NBII MeSH

- aallonpiluus_kooste
- aaqa_mires
- aandark
- abandoned_houses
- abandonment
- abbreviations
- abdomen
- abductions
- ability
- ability_to_write
- abnormality
- abortion
- abrasion_resistance
- abrasion_tests
- absences
- absent-mindedness
- absorption
- absorption
- absorptive_disorders
- abstract_art
- abstraction_of_water
- abstraction_of_water
- abstract_publications

Retrieved Concepts:

- NBII - Body
- LCSH - Body & Society
- NBII - Body fluids
- NBII - Body conditions
- LCSH - Body & Mind
- NBII - Body cavity
- LCSH - Body and soul in literature
- NBII - Beer body
- NBII - Body
- LCSH - Body & Society
- NBII - Body fluids
- NBII - Body conditions
- LCSH - Body, Agnes, 1866-1952
- LCSH - Body & Mind

[View more...](#)

Current Concept: NBII->body

Preferred Label	Body
Broader Term	No broader term
Narrower Term	Abdomen, Appendages, Attachment organs, Bladders, Body cavity
Related Term	Anatomy, Animal morphology, Biomechanics
Scope Notes	LSC Life Sciences
Alternative Label	Animal body Regions, Body regions
URI	http://www.nbii.com/body

Context viewer

HIVE web interface HIVE web service Sign in | Register About HIVE Help Contact Us Feedback

HIVE Vocabulary Server

Help with Interdisciplinary Vocabulary Engineering

Home Concept Browser Indexing

Current vocabularies sources: X LCSH X NBII + Add

HIVE vocabulary server provides functionalities to identify concepts from a given document and automatically assign subject metadata to the document. Working with HIVE to assign metadata from authoritative vocabularies is easy and simple. You just need to upload the document and select the vocabularies you are interested in.

HIVE Automatic Metadata Extractor

[Help](#)

Current vocabularies: X LCSH X NBII

Prioritize the vocabularies: 1. LCSH 2. NBII
Drag the name of the vocabulary to change the order.

Option 1:
 Upload the document:

Option 2:
 Enter text here:

HIVE All Rights Reserved.

Challenges

- Combining many vocabularies during the indexing/term matching phase is difficult, time consuming, inefficient.
 - NLP and machine learning offer promise
- Interoperability = dumbing down
 - ontologies
- Proof-of-concept/ illustrate the differences between HIVE and other vocabulary registries
- General large team logistics, and having people from multiple disciplines (also the ++)



Conclusion

- Linking data
- Dynamic vocabulary integration, a solution...
- Dryad and HIVE are real-world applications using Semantic Web technology

Links

- HIVE
 - <http://ils.unc.edu/mrc/hive/>
- Metadata Research Center <MRC>
 - <http://www.ils.unc.edu/mrc/>
- Dryad
 - <http://datadryad.org/>
- National Evolutionary Synthesis Center (NESCent)
 - <http://www.nescent.org/index.php>