We cannot, through specialization alone, master the pressing problems of modern mankind.
-- Ursula Hubenthal, International Interdisciplinary Scholar (Hubenthal, 1998)

1. ASSESSMENT OF NEED

Controlled vocabularies encode rich knowledge structures that aid cataloging and enhance access to digital resources by facilitating collocation, reducing the searching burden, and increasing the retrieval of relevant documents (Svenonius, 1986; Tudhope, et al., 2001; Drabenstott & Weller, 1996; Rowley, 1994; Greenberg, 2004). While controlled vocabularies have been proven to benefit the users within the library and related communities (Schatz, et al., 1996; Shiri & Revie, 2005), there remain a host of historical problems severely limiting their use in our growing and increasingly interdisciplinary world of digital libraries (e.g., Gnoli & Szostak, 2007). The majority of current problems can be defined by **cost**, **interoperability**, and **usability constraints.**

• **Cost**: Controlled vocabularies are expensive to create, maintain, and use within the current library infrastructure (Svenonius, 1986; Anderson, 1990; Will, 2004; Owens & Cochrane, 2004). Their creation requires domain experts and constant maintenance reflecting changes in knowledge (Greenberg, 2004; Nielsen, 2004). Unfortunately, libraries and digital initiatives generally only have human or financial resources supporting the use of one or perhaps two controlled vocabularies, despite the fact that multiple subject-specific controlled vocabularies could improve access, particularly for interdisciplinary collections. **Manual approaches to metadata generation** are costly (Greenberg, et al., 2001; Greenberg, et al., 2006). Cost constraints are also reflected in the **limited continuing education opportunities**. This limitation is a significant point in the recent "**On the Record"** report (Hill & Schottlaender, 2008), stemming from the **Future of Bibliographic Control Working Group**, and was emphasized in the LC Bibliographic Control Action plan produced in 2000 (Sandberg-Fox, A.M., 2001).

• **Interoperability**: Although many controlled vocabularies use the ANSI/NISO Z39.19-2005 standard, *Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies* (2005), they are generally not interoperable. Most controlled vocabularies are developed in silos—independent of other vocabularies. This results in inconsistent use of similar and identical semantics; different syntactic presentations of word forms; and the use of a wide variety of customized and in-house software applications for vocabulary representation that limits options for distributed searching across multiple vocabularies (Bergamaschi, et al., 1999; Nikolai, et al., 1998; Doerr, 2001).

• **Usability**: Vocabulary usability problems, stemming from interface design and functionality limitations, have been well documented as impeding cataloging and searching (Greenberg 2004; Johnson, 2004; Shiri & Revie, 2005; Macgregor & McCulloch, 2006). Vocabulary usability problems during cataloging extends beyond the traditional barrier of the physical library, as more digital archiving initiatives, such as the Open Archives Initiative, and institutional repositories require authors to assign subject metadata descriptors as part of the submission workflow. One of the reasons social tagging and folksonomy are flourishing in the digital world is because of the ease of access and use, which controlled vocabulary have not demonstrated.

These problems mirror **significant challenges** we are encountering in implementing a controlled vocabulary system for the Dryad repository[1] (Dube, et al., 2007; Carrier, et al., 2007). Dryad is a collaboration involving NESCent—the National Evolutionary Synthesis Center in Durham, North Carolina, and the Metadata Research Center at the School of Information and Library Science, University of North Carolina at Chapel Hill. Dryad is a digital repository that links data objects supporting published research in 10 leading evolutionary biology journals, including *The American Naturalist* and *Evolution*. Evolutionary biology is a highly interdisciplinary field drawing from a wide range of scientific disciplines, including ecology, developmental biology, genetics, molecular biology, paleontology, and systematics (Futuyma, 2005; Ridley, 2003).

As a first step in developing a controlled vocabulary system, we conducted a needs assessment by drawing close to 600 keywords from articles in Dryad partner journals. Keywords were categorized into facets (e.g., taxon,

[1] http://www.datadryad.org/ ; https://www.nescent.org/wg_digitaldata/Main_Page

geographic name, time period, topic), and mapped to terms in the following vocabulary sources: *Education Resources Information Center*, *Medical Subject Headings* (MeSH), the National Biological Information Infrastructure's *Biocomplexity Thesaurus* (NBII Thesaurus), *Library of Congress Subject Headings* (LCSH), the Getty's *Thesaurus of Geographic Names*, *WordNet*, *Alexandria Library Gazetteer*, *Gene Ontology*, *Integrated Taxonomic Information System* (ITIS), and *Universal Biological Indexer and Organizer* (UBIO). We found that 22% of the terms map directly to the NBII Thesaurus; 23% map to MeSH; and 33% map to LCSH, with roughly a 9% overlap among terms appearing in more than one vocabulary. These results indicate that an interdisciplinary subject, such as evolutionary biology, cannot be represented via a single vocabulary. However, multiple vocabularies combined, could provide greater standardized representation. Even with these findings, the cost of working with multiple vocabularies in a fairly manual environment is prohibitively expensive. Interdisciplinary repositories and digital libraries, such as Dryad, could greatly benefit from using automatic techniques to draw descriptors from multiple existing controlled vocabularies through a single system (Greenberg, et al., 2006).

To address these needs, and noted controlled vocabularies limitations, we are proposing HIVE (Helping Interdisciplinary Vocabulary Engineering), an automatic metadata generation approach that dynamically integrates discipline specific controlled vocabularies and assists content creators (or professionals) with subject cataloging.

HIVE will address the need for:

1. **Affordable automatic approaches for generating subject metadata.**
2. **Interoperable vocabulary using the Simple Knowedge Organisation Systems[2] (Miles & Pérez-Agüera, 2007).**
3. **Usable and functional systems that will aid resource catalogers and authors creating metadata.**
4. **Empowered professionals with knowledge of new standards.**

Our overriding goal is to demonstrate HIVE, and make our process, procedures, and SKOS server technology accessible to any library or digital initiative that may benefit from access to multiple controlled vocabularies for indexing and accessing collection holdings.

2. NATIONAL IMPACT

HIVE will have a national impact in three key areas: the **utility of controlled vocabularies in information environments**; **facilitating access to interdisciplinary collections;** and **empowering information professionals.**

• **Utility of controlled vocabularies in information environments.** The information profession has invested, and continues to invest, substantial human and financial resources into controlled vocabulary development, maintenance, and use. HIVE will allow controlled vocabularies to be used in efficient, flexible, and less costly ways, making them much more accessible to the information community. The standardization of controlled vocabularies in SKOS will also enable greater automatic metadata generation and resource sharing opportunities.

• **Facilitates access to interdisciplinary collection.** The HIVE model is meant to be used in ways that exploit the true richness of collections extending beyond traditional disciplinary boundaries. The growth in interdisciplinary studies over the last several decades, together with new digital technologies, have led to an increase in digital collections, archives, and museum projects covering interdisciplinary topics. A recent informal assessment of model Institute of Museum and Library Services (IMLS) projects represented in the IMLS Digital Collections and Content (DCC) repository[3] found that nearly all of the projects in this repository could benefit from an interdisciplinary vocabulary system. Two specific examples are Discover Babylon[4], which uses gaming to promote studying Mesopotamia's contribution in writing, math, literature and law, and promotes the use cultural artifacts to "encourage interdisciplinary collaboration and Code City[5], a website that

[2] http://www.w3.org/2004/02/skos/
[3] http://imlsdcc.grainger.uiuc.edu/
[4] http://www.discoverbabylon.org/
[5] http://www.tenement.org/codecity/

looks at different legal policies from a social perspective while integrating the use of text, photographs, and maps.

• **Empowering information professionals.** Continuing education is essential for all members of the information profession. Through innovative workshop sessions, HIVE will empower librarians, archivists, and museum professionals with knowledge and skills about new approaches for tackling basic information challenges. In daily resource description activities, information professionals are forced to assign subject headings that accurately represent resource content, because they do not have access to the appropriate vocabularies. Through dynamic vocabulary integration techniques, HIVE will alleviate the stress of trying to find a perfect vocabulary match within one confined vocabulary by providing access to multiple vocabularies.

3. DESIGN

The **HIVE project design** includes three parts: **I. Building HIVE**, **II. Sharing HIVE**, and **III. Evaluating HIVE**. These components have been developed to address the needs articulated above in "Section 1, Assessment of Need," and to support an evaluation of HIVE. Part I., Building HIVE, addresses the need to provide efficient, affordable, interoperable, and user friendly access to multiple controlled vocabularies during metadata creation activities. Part II., Sharing HIVE, addresses the need to educate library, museum, and archival professionals about new enabling technologies that can assist them with developing and using controlled vocabularies. Part III., Evaluating HIVE, supports the evaluation of HIVE using automatic information retrieval approaches, studying resource authors and scientists contributing to the Dryad repository, and by surveying members of the LIS professional community about the usefulness of HIVE for their daily work.

Our project design will result in a procedural model, a technical architecture, and open source applications—all of which will be evaluated. Project outcomes, including our evaluation results, will be made available to any digital initiative wanting to employ automatic metadata generation techniques to draw descriptors from multiple controlled vocabularies in order to enhance access to collection holdings. In the following section, we state our project's goals; describe our intended interdisciplinary sample; outline the objectives and tasks for each part of the project; and present our evaluation plans.

Purpose and Goals

Purpose: To improve the state of access and use of controlled vocabularies in the digital environment. The specific **goals** of HIVE are to:

1. Provide an affordable approach for generating subject metadata using automatic metadata generation techniques and pulling concepts from multiple subject thesauri.
2. Create an interoperable vocabulary server using the Simple Knowledge Organisation Systems (SKOS).
3. Create a usable and functional system that will aid resource catalogers and resource authors creating subject metadata.
4. Develop and deliver a series of workshops on SKOS and the HIVE model.
5. Evaluate the effectiveness and usability of HIVE for the Dryad repository and the larger library, museum, and archival community.

Interdisciplinary Sample

To demonstrate an effective system, we will begin with a sample of **controlled vocabularies**, and **a sample of documents/data objects** representing interdisciplinary topics found in the Dryad repository. Our vocabulary sample includes the *Library of Congress Subject Headings* (LCSH), the *National Biological Information Infrastructure Biocomplexity Thesaurus* (NBII Thesaurus), and the Getty's *Thesaurus of Geographical Names* (TGN). Representatives from each of these controlled vocabularies have agreed to provide us with the vocabulary source, and to also serve on HIVE's advisory board. Support letters from vocabulary contributors are provided in the attachment section, as supporting documents 1, 2, and 3.

The document sample is drawn from the Dryad repository. SILS/MRC and NESCent have been working on the Dryad repository for the last year, and our relationship provides an optimal partnership for the development and evaluation of HIVE. The sample will reflect research topics addressed in NESCent's **Primate Life Histories Working Group**[6] and the **Wood Anatomy and Wood Density Working Group**[7]. NESCent hosts these and other working groups to bring together researchers exploring topics in the field of evolutionary biology. The Primate Life Histories Working Group brings together researchers interested in comparing mortality and fertility schedules across primates, and better understanding the ecological and behavioral factors that impact mortality and fertility patterns in these taxa. The Wood Anatomy and Wood Density Working Group brings together scientists studying the evolution of wood anatomy in relation to phylogenetic and environmental factors. These working groups provide an excellent basis for our sample because of their interdisciplinary nature and because they represent both historic and active areas of publication in key journals associated with the Dryad repository.

I. Building HIVE

A. Vocabulary Preparation

Objective: To convert our sample vocabularies to SKOS (or vocabulary subsections, representing the topics reflected in the working group topics of "primate life histories" and "wood anatomy and wood density").

Tasks*:*

1. **Create an algorithm for converting vocabulary following NISO Z39.19 to SKOS**. Our process will convert LCSH MARC21 or MARC-XML records, TGN XML records, and NBII XML records to SKOS. In 2004, the USGS developed a SKOS Thesaurus prototype, with a segment of the NBII Thesaurus, as part of the Semantic Web Advanced Development for Europe (SWAD-Europe) initiative. Our methods will incorporate this effort, and we will keep in close dialog with the Library of Congress, sharing approaches and methods as they explore SKOS for LCSH during the latter part of 2008.

2. **Run our initial SKOS vocabulary conversion algorithm against our sample of controlled vocabularies**. We will run an initial algorithm against LCSH, TGN, and NBII Thesaurus as a first pass, to determine the percentage of each vocabulary that can be automatically converted to SKOS, evaluate the results, and identify problems that can be addressed via additional programming, and problems requiring human expertise. The evaluation will consider SKOS syntax and encoding, and consider structural, formative, observational, and comparative evaluation techniques reviewed by Owens and Cochrane (2004). An example of SKOS results for the heading "estrogen" from the *NBII Thesaurus* is presented in Figure 1.

Figure 1: *NBII Thesaurus* Public Display and SKOS Encoding for Entry "Estrogens"

Public display	SKOS Encoding corresponding to the public display
Estrogens BT **Sex hormones** NT **Phytoestrogens** RT **Estrus**	<!-- **NBII starts here** --> ... <!-- **Authorized heading** --> <skos:**Concept** rdf:resource="http://www.nbii.gov/... Biocomplexity_Thesaurus/foo/**Estrogens**"/> <!-- **Hierarchical relationships start here** --> <skos:**Broader** rdf:resource="http://www.nbii.gov/... Biocomplexity_Thesaurus/foo/**Sex hormones**"/> <skos:**Narrower** rdf:resource="http://www.nbii.gov/... Biocomplexity_Thesaurus/foo/**Phytoestrogens**"/> <!-- **Reciprocal relationships start here** --> <skos:**Related** rdf:resource="http://www.nbii.gov/... Biocomplexity_Thesaurus/foo/**Estrus**"/>

[6] http://www.nescent.org/science/awards_summary.php?id=2053
[7] http://www.vegfunction.net/wg/27/27_wood_anatomy.htm

3. **Refine the SKOS conversion algorithm**. Modify the SKOS conversion algorithm based on the results of task two, execute the revised algorithm, and evaluate the results following metrics established for task two. As part of this second evaluation, we will revise the algorithm to include the following determinations: 1. Algorithm modifications that will further aid the large scale SKOS conversion of our sample vocabularies, and 2. Any SKOS conversion vocabulary problems requiring manual review that can be automatically detected and flagged. For example, we anticipate that inverted headings or string headings in LCSH (e.g., LCSH headings: "New World monkeys--Ecology--Peru", and "Trees--Africa, Central Anatomy") may not directly map to SKOS as single term concepts, but an algorithm that locates commas or dashes between concepts may be able to automatically flag these types of problems. Based on our evaluations, we may elect to SKOS sub-sections of one or more vocabularies sufficient to let us proceed with our demonstration. Professor Robert Losee, a senior staff member on the HIVE team, is an expert in research methods, information retrieval, and thesauri development, and will develop an appropriate sampling method corresponding with the interdisciplinary topics of our document sample.

B. Vocabulary Server Development

Objective: To build a system that can search SKOS vocabularies, and a database that can store and present SKOS relationships for selection during cataloging.

Tasks:

1. **Build a database system using that supports search and vocabulary enhancement functions.** The database should support the following functions: 1. Given a keyword, find term, 2. Given a term, find all related terms and their relationships, and 3. Add term. The database will be accessible via standard web services interfaces, including both SRU/ZTHES[8] and SPARQL[9] to bridge between library systems and semantic web systems, and take advantage of the SKOS/RDF relationship of our SKOS vocabularies. NESCent will store the SKOS vocabularies locally for duration of this project. The long-term plan is to update our vocabulary during selected intervals, by obtaining them from remote locations, managed by the organizations responsible for maintains the vocabularies.

2. **Implement an algorithm that will automatically extract keywords from a document's abstract.** Term frequencies for stemmed terms will be used as the basis for automatic indexing. Terms will be stemmed using a Porter Stemmer. Our methods will integrate work by Losee (2007) which examines controlled and uncontrolled terms in relation to inclusion in thesauri. We will experiment with a corpus of terms extracted from controlled vocabularies related to our sample topics. We will explore several automatic indexing algorithms and methods, including those in the Lucene toolkit[10] (Gospodnetic & Hatcher, 2004). In particular, an information theoretic method will be tested to determine which terms function as "signals" and which as "noise" based on their term frequencies across abstract collections, as well as a method that selects terms with the highest ability to separate relevant and non-relevant documents.

3. **Develop a module linking to Dryad's deposition of datasets for articles** that will 1. extract keywords from document abstracts, 2. combine the automatically extracted keywords with the author-assigned keywords, 3. search the SKOS vocabularies accessible via the server, and 4. display the search results to the author or cataloger initiating the cataloging sequence. This work will align with the controlled vocabulary module in DSpace, including the recent addition developed by the University of Minho, which supports additional semantic relationships encoded in standardized thesauri following Z39.19. Figure two illustrates the architecture of the HIVE vocabulary server, and Figure 3 illustrates a possible end user interface for HIVE's semi-automatic metadata assignment feature.

[8] http://zthes.z3950.org/srw/zthes-srw-1.0.html.
[9] http://www.w3.org/TR/rdf-sparql-query/.
[10] http://lucene.apache.org/.

Figure 2: HIVE Vocabulary Server

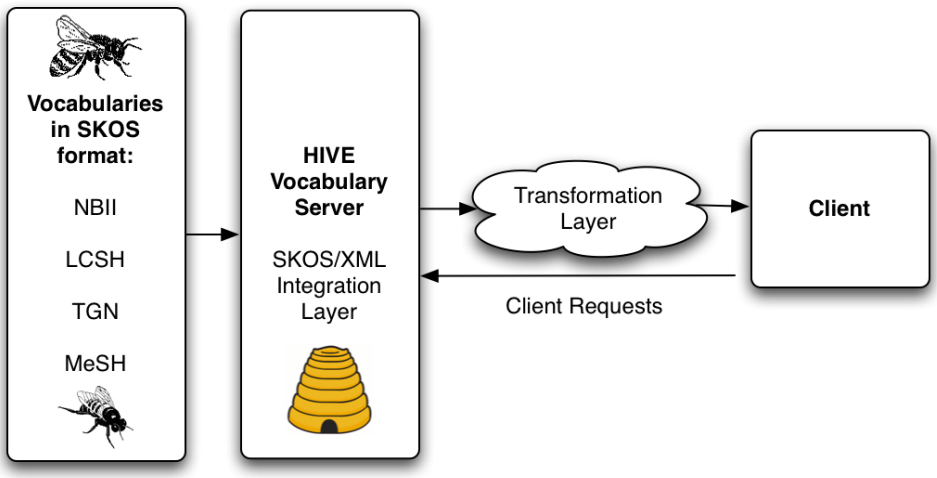
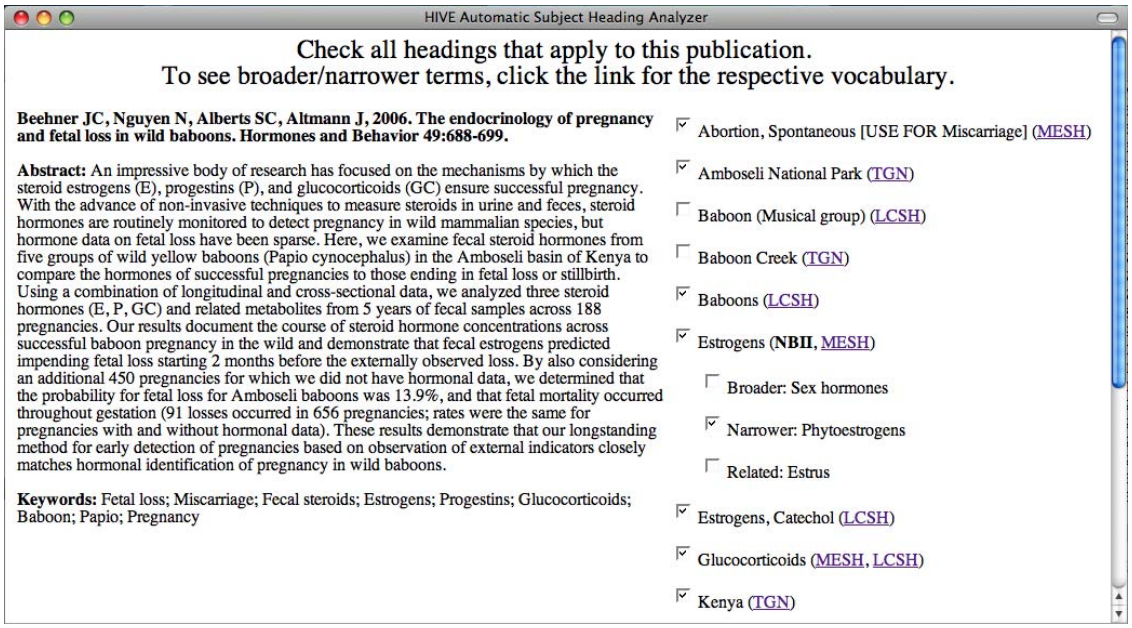


Figure 3: HIVE interface example



II. Sharing HIVE

A. Workshop Development

Objective: Develop workshop content for introducing and teaching both SKOS and the HIVE approach to LIS professionals**.**

Workshop preparation, goals, and activities:

1. We will develop a **SKOS/HIVE workshop wiki**, which participants will access in order to complete background readings on the SKOS standard and controlled vocabulary issues prior to workshop attendance.
2. We will develop four modules for our workshop:
    - The introductory module will review the importance of controlled vocabularies and document historical problems with their use; introduce SKOS and how can apply it to different vocabularies; and introduce the HIVE model.

- The SKOS module will provide exercises on how to create a SKOS vocabulary, working with selected software (e.g., Protégé), and review processes and issues with automatic SKOS conversions.
- The HIVE module will give participants the opportunity to interact with a working version of HIVE. Participants be able to log into the Dryad repository HIVE system remotely and will be given sample articles to catalog within the system. We will develop instructor led exercises, and exercises for the participants to complete in individually and in groups.
- The evaluation module will share HIVE evaluation methods and results with participants and discuss the importance of evaluation in the workplace. We will also present them with approaches which can be used to assist them with their own HIVE implementation.

The wrap-up/conclusion will revisit key points covered during the day, discuss the future of SKOS and HIVE' (potential short term and long term implications), and provide time for questions and an open discussion.

B. Workshop Delivery

Objective: To deliver a well formed and insightful workshop, for LIS professionals, on SKOS and HIVE at five select locations**.**

Workshop logistics, location, and contacts:

The workshops will be delivered on site by two project members at five select locations. The workshop will be a single full-day event (approximately 9:00 a.m. to 5:00 p.m.). A registration fee of $125.00 per participant covers various workshop expenses, discussed in the budget justification, including workshop packets, networked connectivity to SKOS and HIVE exercises and Dryad's HIVE server; and lunch, coffee, and snacks. Workshop institutions and the primary contact/consultants for each workshop are as follows:

1. **Columbia University, New York City** (Robert Wolven, Director of Library Systems & Bibliographic Control)
2. **University of Carlos, Madrid, Spain** (Dr. Eva Mendez, Associate Professor, Dpto. Biblioteconomía y Documentación)
3. **University of California, San Diego** (Martha Hruska, Associate University Librarian, Collection Services)
4. **University of North Texas, Texas Center for Digital Knowledge** (TxCDK) (William E. Moen, Associate Professor, Director of TxCDK, and Director of Research, School of Library and Information Sciences)
5. **University of North Carolina at Chapel Hill** (Kristin Martin, Electronic Resources Cataloger)

III. Evaluating HIVE

Objective: To evaluate HIVE within the Dryad context, with resource authors/scientists and in the larger LIS context with workshop participants.

HIVE/Dryad Evaluation:

We will use a **multi-method approach** to evaluate the usefulness and effectiveness of HIVE for the Dryad repository using automatic information retrieval approaches, a transaction log analysis, a content analysis, through survey approaches and structured interviews. We will develop a sample of articles with our test topics of "primate life histories" and "wood anatomy and density" drawing from journals that are connected to Dryad. The sample will be conducted in three separate evaluations:

1. **Automatic subject metadata generation/information retrieval evaluation**: We will use tools based on Lucene to extract author-assigned keywords and keywords in the document abstract. These terms will be processed via HIVE to guide the automatic assignment of subject metadata. Queries will be generated from document titles, and relevant documents will be based on citation frequencies. Performance will be computed using common retrieval evaluation measures such as precision and recall.
2. **Transaction log analysis and content analysis**: We will conduct a **transaction log analysis** to study the interaction of resource authors/scientists with HIVE. Scientists who are members of our test working groups will be provided with test documents and data objects for the Dryad repository and will work with

HIVE to generate subject metadata for these objects. We will analyze factors, such as time, toggling, keystrokes, and other usability factors (Dumas & Redish, 1993; Rubin & Hudson, 1994) to assess the value of HIVE. Professional catalogers on the HIVE team will perform a **content analysis** of the results, following the methodology of Krippendorff (2004). The School of Information and Library Science maintains an Interactive Design Lab, which we will explore to gather addition types of data for evaluation.

3. **Survey and structured interview**: We will survey scientists about their experience working with HIVE and conduct several targeted structured interviews to gather additional data.

Our plan will allow us to evaluate HIVE from a number of different angles and help us identify strengths and limitations, high priority areas requiring modification, and plan for future HIVE development. We will share our approaches and outcomes with all workshop participants and make our evaluation documentation and reports accessible on our project wiki.

HIVE evaluation by LIS professionals:

Similar to the Dryad-specific HIVE evaluation involving scientists, we will use a muti-method approach to assess the value and usefulness of HIVE for the larger library, museum, and archival community. We will develop a sample of articles from items represented/or to-be-represented in Dryad that are understandable to the non-evolutionary biologists for the transaction log/content analysis. We will recruit participants from our workshop for this evaluation, as well as a survey and a focus group discussion, as described here:

1. A **pre-workshop survey** will gather data about LIS professionals' opinions on controlled vocabularies (i.e., what they currently use in their workplace, how they use them, their interest in SKOS and new enabling technologies). We will also gather general demographic and professional information.
2. A **post-workshop survey** will gather data from the LIS community on the value of the HIVE approach (i.e., HIVE's interface and usability, and recommendations for vocabularies that should be converted to SKOS) as well other next steps.
3. A **focus group** will gather additional data similar to the post-workshop survey, but allow for dialog and discussion of these HIVE issues and its potential.

We will follow UNC's Institutional Review Board procedures for all testing and recruitment activities in accordance with University policy, to ensure high standard and ethical behavior in our research approach.

4. PROJECT RESOURCES: BUDGET, PERSONNEL, AND MANAGEMENT

Budget

The bulk of HIVE's budget is dedicated to supporting project staff, much of which is cost-shared. In particular, the programmer (1.0 FTE), the P.I. (Jane Greenberg, .15 FTE), and co-P.I. (Ryan Scherle, .10 FTE) are essential components of the overall workplan. Funding for two graduate student assistants is also included, as they are key members of the HIVE team. Money has been requested to support travel to important scholarly and scientific conferences in order to disseminate project findings and keep up with new discoveries and advancements that are important to meeting project goals. Also included in the budget are computer/technological support costs, generic office supplies, and funding to support full-day dissemination workshops in five locations worldwide. Annual advisory board meeting costs have also been budgeted.

Personnel

Project Staff

• **Jane Greenberg** is the Francis Carroll McColl Term Professor and Director of the Metadata Research Center in the School of Information and Library Science at UNC Chapel Hill. She has worked as professional cataloger in a number significant institutions, including the Schomburg Center for Research in Black Culture, a research division of the New York Public Library, Research Libraries, where she was head of Special Collections Cataloging and Acting Head of Cataloging. She is the recipient of the 2008 Frederick G. Kilgour Award for Research in Library and Information Technology by LITA, a division of the American Library

Association. Dr. Greenberg will serve as the P.I. and overall project manager, functioning as the main contact between SILS and NESCent, as well as for all advisory board activities.

• **Ryan Scherle** is the Digital Data Repository Architect for the National Evolutionary Synthesis Center at Duke University. He served as a lead programmer for the Variations project at Indiana University and is an accomplished programmer and data architect. He will act as Co-P.I. for the project, working closely with the project programmer and involving himself with all aspects of project evaluation.

• **Robert Losee** and **Todd Vision**, UNC Professors, will function as senior staff members, helping to coordinate HIVE evaluation. **Kristin Martin**, a cataloger for UNC Libraries, will also assist with HIVE evaluation activities. **Aaron Brubaker**, Director of Instructional Technology at SILS, will provide technical support and guidance for the project. **Jeffrey Tibbs** and **Wanda Monroe**, SILS staff members, will aid with budgetary and public relations guidance.

Advisory Board

The HIVE Advisory Board includes experts in metadata, data standards, and controlled vocabulary development and usability. The role of the advisory board is to provide direction, support, and expert guidance throughout the duration of the project. We have planned two annual meetings (for year one and year two), and a third meeting near the completion of our project to share findings and plan for next steps. We will work with our advisory board members through virtual technology, so that they can provide feedback on our test instruments and participate in our HIVE evaluation. Our advisory board is strengthened by two international members, Alistair Miles, the lead author of the SKOS standard, and Dr. Eva Mendez, who has translated RDF and SKOS standards into Spanish for the Spanish speaking library community. The advisory board members include:

• **Barbara Tillett**, Acting Chief of the Cataloging Distribution Service and Chief of the Cataloging Policy and Support Office

• **Joseph Shubitowski**, Head, Information Systems, Getty Research Institute

• **Mike Frame**, Director of Research & Technology for the USGS NBII Program

• **Lisa Zolly**, Chief Knowledge Manager, NBII Program

• **William Moen,** Associate Professor, School, Director of TxCDK, and Director of Research, School of Library and Information Sciences

• **Eva Méndez Rodríguez,** Associate Professor, University Carlos III of Madrid

• **Kathy Wisser,** NC ECHO Metadata Coordinator

• **Kristin Martin,** Electronic Resources Cataloger, UNC Libraries

• **Alistair Miles,** Researcher, CCLRC Rutherford Appleton Laboratory

• **Jim Balhoff,** Research Programmer, National Evolutionary Synthesis Center (NESCent)

Management

The project will be administered by the UNC/SILS/Metadata Research Center. Dr. Jane Greenberg, will be responsible for all managerial arrangements, including:

- Coordinate team and advisory board meetings
- Establish project deliverables following the predetermined schedule
- Ensure integration of work performed by the different organizations
- Oversee the implementation of the project listserv and wiki in order to enable effective team communication

5. DISSEMINATION

We will disseminate the HIVE model and our results through communication technologies and scholarly, professional, and educational venues. We will share all project findings and documentation via our wiki throughout the durations of the project. We will broadcast project updates via appropriate listservs. We will present the HIVE

model and evaluation findings at relevant conferences and meetings, such as the IMLS Web-WISE, American Library Association, Association of Library of Information Science Education, Dublin Core, American Society of Information Science & Technology, Open Repositories, and Society of American Archivists. We will submit our findings to peer-reviewed journal publications so as to ensure wide dissemination to all interested audiences. Targeted journals include *College & Research Libraries*; *International Journal of Metadata, Semantics & Ontologies*; *Information Technology & Libraries*; *JALISE*; *JASIS&T*; *CCQ*; and *American Archivist*. As part of our publication plan, we will work with Wanda Monroe, our public relations staff person, to also publish in widely read literature, such as *American Libraries*, *Chronicle of Higher Education*, and *Nature*.

Our workshops will reach out to library, museum, and archival professionals in the national and international community to teach them the practical and important skills that are involved in HIVE dynamic vocabulary integration.

6. SUSTAINABILITY

The Metadata Research Center (MRC) and National Evolutionary Synthesis Center (NESCent) are committed to maintaining and growing HIVE for the larger library environment as well for the Dryad repository. The MRC is part of a larger collaborative environment, involving a number of partnerships with SILS faculty and library professionals within the University community and the Research Triangle Park area of North Carolina. These partnerships provide ample opportunities to extend and sustain HIVE in a variety of interdisciplinary settings. The SILS Metadata Research Center and NESCent are well positioned to continue to develop HIVE and extend the HIVE model to other vocabularies relevant to evolutionary biology and related scientific disciplines. NESCent currently hosts work in ontology integration that could benefit HIVE if linked with SKOS.

The interest and excitement from significant partners in the library and museum community, such as the Library of Congress, United States Geological Survey (USGS), and the Getty Research Institute, has made it evident that the HIVE model has the potential to be widely adopted. Several individuals and organizations, beyond those mentioned above in this proposal, are aware of our proposal and have offered additional opportunities for workshops and presentations on HIVE. For example, Ralph Swick, Technology and Society Domain Leader, **World Wide Web Consortium**, has inquired about a HIVE workshop linking the **W3C and MIT Libraries**. On an international level, Dr. Thomas Baker, Director of Specifications and Documentation, Dublin **Core Metadata Initiative**, has also expressed interest in our plans and inquired about workshop delivery at **Humboldt University, Germany**, and a Dublin Core pre-conference workshop.

*** References included in "Supportingdoc4.pdf".**