I. Purpose

DLISTConnection will develop and evaluate an information and technology literacy (ITL) service in support of science and health literacy by 1) federating training materials, software documentation, and similar learning objects not systematically collected and described in the NSDL and 2) designing, implementing, and assessing a controlled vocabulary for existing ITL standards by aligning them with science and health literacy benchmarks. Further, DLISTConnection will develop rights management policies to facilitate harvesting and use of diverse learning objects by applying selected rights elements Evaluation will include NSDL testbeds and an informetric analysis of the effectiveness of the metadata for standards and rights. Two new communities, ITL professionals and Native Americans will be involved. DLISTConnection thus builds a foundation for the NSDL goal of science literacy by providing current and new audiences of end-users and collections providers with four innovative yet essential services:

1. addition of health sciences-specific ITL learning objects to the NSDL;
2. availability of crosswalks connecting ITL standards to science and health literacy benchmarks and the mapping of those standards and benchmarks to the learning objects;
3. access to intellectual property rights metadata to facilitate re-use and re-purposing of learning objects; and
4. application of citation indexing and analysis to learning objects.

II. Objectives and Methods to Be Employed

During the two-year period of the project, investigators will work together to:

1. Assemble and catalog an OAI compliant collection of 3000 learning objects supporting 87 information and technology literacy competency outcomes
2. Develop rights metadata for resources that transition from being proprietary or copyrighted to those shared under Open Content licenses and identify rights associated with harvesting of ITL materials
3. Correlate and align the information and technology literacy outcomes in the standards from the American College & Research Libraries (ACRL), American Association of School Librarians (AASL) & Association for Education and Communication Technology (AECT), and International Society for Technology Education (ISTE) National Educational Technology Standards (NETS), with the American Association for the Advancement of Science (AAAS) Project 2061 Benchmarks for Science Literacy and Health Literacy outcomes and create a controlled vocabulary
4. Build an interoperable and heterogeneous digital library by integrating Eprints with Dspace communities, and experimenting with harvesting, searching, and citation indexing
5. Evaluate using NSDL testbeds (HEAL and AskNSDL VRD) and informetrics methods to assess how items are being used and understand the effectiveness of the standards and rights metadata.

III. Intellectual Merit

DLISTConnection provides a new form of intellectual access to unique materials that are currently not collected by NSDL; by harvesting content from at least three types of publishers, commercial database publisher DIALOG, public domain materials available from NLM and works for hire shared through a membership clearinghouse, LOEX, and aligning information and technology literacy with science and health, and cataloging them to the collected resources, information discovery and access of uncollected ITL objects are not only tailored to the unique needs of K-20+ audiences, science educators, librarians, technologists, and end-users can begin to understand how competencies fit across disciplinary boundaries.

IV. Broader Impacts

DLISTConnection brings new kinds of resources (training and help) and new communities systematically (instructional librarians and Native Americans) into the NSDL. This proposal goes beyond the voluntary, sporadic, and hard-to-sustain efforts of organizations that maintain web pages with links to information literacy resources, often known only to the respective organizations' members to lay the foundation for how interdisciplinarity (complex learning across disciplines), collaboration, and disintermediation, key characteristics of future scholarship, may be modeled in a Digital Library service protocol.

I. Background and Significance

DLISTConnection will develop and evaluate an information and technology literacy (ITL) service in support of science and health literacy by 1) federating training materials, software documentation, and similar learning objects not systematically collected and described in the NSDL and 2) designing, implementing, and assessing a controlled vocabulary for existing ITL standards by aligning them with science and health literacy benchmarks. Further, DLISTConnection will develop rights management policies to facilitate harvesting and use of diverse learning objects by applying selected rights elements Evaluation will include NSDL testbeds and an informetric analysis of the effectiveness of the metadata for standards and rights. Two new communities, ITL professionals and Native Americans will be involved. DLISTConnection thus builds a foundation for the NSDL goal of science literacy by providing current and new audiences of end-users and collections providers with four innovative yet essential services: the availability of crosswalks connecting ITL standards to science and health literacy benchmarks and the mapping of those standards and benchmarks to the learning objects; access to intellectual property rights metadata to facilitate re-use and re-purposing of learning objects; and application of citation indexing and analysis to learning objects.

DLISTConnection represents an advance of DLIST, the *Digital Library of Information Science and Technology* available at http://dlist.sir.arizona.edu, a project of the School of Information Resources and Library Science (SIRLS) at the University of Arizona in partnership with the Arizona Health Sciences Library (AHSL). DLIST aims to create a Web-accessible, open, repository of instructional and research materials in Library & Information Science and Technology, in forms (types of materials) and formats (media type) that have not been collected by the traditional or digital library. [Coleman and Bracke, 2003]

DLIST is based on open source software. At its core is the Open Archives compliant Eprints 2.2 package developed at the University of Southampton, UK [Open Archives, Eprints]. Eprints requires the Linux operating system, the Apache web server with mod_perl, the Perl programming language with a small number of extra modules, and the MySQL database system. Webalizer, another open source software, is used to analyze and customize DLIST usage reports [Webalizer]. Figure 1 shows a screenshot of the DLIST home page.

Figure 1: DLIST home page

Eprints software is generally used to build institutional or discipline repositories of scholarship, the outputs of research such as journal articles, technical reports, conference proceedings, theses, dissertations, and books. Whole journals, books, and conference proceedings or their components such as individual chapters, and articles can be deposited into a Web-accessible digital storage system and described using a database form. Examples and well-known Eprint archives exist for disciplines such as CogPrints for Cognitive Science and ArXiv for Physics [Cogprints, ArXiv]. In its simplest sense DLIST is a discipline repository for at least three subjects: Library Science, Information Science (IS), and Information Technology (IT). However, DLIST is unique among Eprint repositories because right from its inception the vision was to accommodate learning objects as well as research articles and reports. The impetus for such integration came from several sources. First, there is an increasingly strong trend that views teaching materials along with academic research as part of the intellectual output of scholarship [Scholarly Communication Editorial Board]. To accommodate this, DLIST extends eprint types of refereed research articles and preprints to learning objects. DLIST learning objects include various types of learning materials such as tutorials, syllabi, course outlines, and lesson plans, informal materials such as presentations, and bibliographic instruction (ITL) materials such as pathfinders, database search guides, and webliographies. Second, we found that librarians and technologists in academic libraries and in commercial settings such as database publishing, create a wealth of instructional content including tutorials for software, electronic resources, libraries and databases, bibliographies, pathfinders, subject guides, and software documentation that are neither managed nor harnessed efficiently for use. These materials, however, are extensively re-used and re-purposed in library instruction, face-to-face and remote reference services, and classroom teaching in all levels of education from K-20+ to impart ITL skills (and hence referred to as ITL materials). Additionally, easy availability of ITL materials along with research articles facilitates their own lifelong learning and professional development. Sundaram, in her dissertation investigating the nature of information work at the health sciences libraries' reference desk found the re-use of library instructional materials teaching/learning to be significant job activities [Sundaram, 1996]. Thus, Information literacy became an early collection development focus for DLIST.

Attempts were made to distinguish DLIST from other repositories such as Eisenhower National Clearinghouse (ENC), Multimedia Educational Repositories for Learning and Online Teaching (MERLOT), The Gateway to Educational Materials (GEM), Library Orientation Exchange (LOEX), American Library Association's (ALA) Government Documents Round Table (GODORT), California Clearinghouse for Library Instruction (CCLI), Educational Resources Information Center (ERIC), National Science Digital Library (NSDL), CITIDEL, NCSTRL and Networked Digital Library of Theses and Dissertations (NDLTD) [ENC Online, MERLOT, GEM, LOEX, ALA GODORT, CCLI, ERIC, NSDL, CITIDEL, NCSTRL, NDLTD]. MERLOT, GEM, ENC focus on educational materials in all disciplines for K-16, while the NDLTD collects digital theses and dissertations in all academic disciplines. CITIDEL and NCSTRL collect in the areas of Computer Science. None of these collect the unique genre of ITL materials – learning objects that have been created to provide instruction in the use of digital libraries, electronic resources, databases, and software. The NSDL portal features the *Librarians' Index to the Internet*, a database of useful materials organized according to traditional disciplines, and offering resources about information literacy [op. cit.]. LOEX, CCLI, and GODORT gather instructional materials to facilitate exchange among their members/participants and collect information and technology literacy materials. However, several aspects limit the effectiveness and influence of these:

1. They are known only among librarians who participate in the respective organizations.
2. They do not create repositories with metadata but rather web pages of links organized in a directory.
3. It generally is not clear who holds the copyright or how materials may be used.
4. It is impossible to discern whether others are using the linked resources because citation indexing, which allows one to identify the works an author cites, has not been adapted and applied.
5. Most importantly, none of these map to information literacy and technology competency standards in alignment with science and health literacy benchmarks.

The overarching goal of DLISTConnection is to overcome these limitations.

II. Conceptual Approach

Our project supports the vision of the NSDL as "a learning environments and resources network for science, mathematics, engineering and technology education" by taking into account the need for services that lie outside the mentioned content domains [Zia, 2001]. One such content area that is ripe for development is **Information Literacy**. Information literacy and the technological literacy that bolsters it underlie the acquisition of science literacy, an NSDL goal, by making it possible for individuals to discover efficiently and apply effectively the scientific information they need and want. Descriptions of the science-literate and the information-literate person underscore the relationship between the two areas of knowledge:

> "[T]he science-literate person is aware that science, mathematics, and technology are interdependent human enterprises with strengths and limitations; understands key concepts and principles of science; is familiar with the natural world and recognizes both its diversity and unity; and uses scientific knowledge and scientific ways of thinking for individual and social purposes." [Nelson, 1999].
>
> "[I]nformation literate people are those who have learned how to learn. They know how to learn because they know how knowledge is organized, how to find information, and how to use information in such a way that others can learn from them. They are people prepared for lifelong learning, because they can always find the information needed for any task or decision at hand." [ALA Presidential Committee on Information Literacy, 1989]

In health sciences education, the relationship is understood and its significance appreciated. For example, medical school faculty and medical librarians collaborate to construct problem-based or evidence-based learning experiences for students who must be able to identify and locate relevant medical literature in order to make a diagnosis when presented with a hypothetical case [Kaplowitz and Yamamoto, 2001]. Consequently, cultivating information literacy in conjunction with science literacy by developing and aligning the vocabulary for them is a key focus of our project.

Information literacy is important for science learning because traditional and digital library use builds skills in seeking, gathering, evaluating and using information. The relationship between undergraduate education and the library is acknowledged in the 1986 Carnegie report:

> "The quality of a college is measured by the resources for learning on the campus and the extent to which students become independent, self-directed learners. And yet we found that today, about one out of every four undergraduates spends no time in the library during a normal week, and 65 percent use the library four hours or less each week. The gap between the classroom and the library, reported on almost a half-century ago, still exists today." [Prologue, 1986]

A year later, the ALA Presidential Committee on Information Literacy was appointed and it released its own report three years later. This report highlighted the need for research into the use of information resources in various disciplines and the information management skills that affect student performance and retention. It also called for "an increased emphasis on information literacy and resource-based learning."[op. cit.] In the 1990s, states, librarians, professional societies, and schools have all engaged in developing standards for content areas and generic skills which cut across all subjects such as information and technology literacy.

There is no doubt that there has been an increasing trend in educational reform efforts and businesses towards competency or skills and standards based education; many educators see the 1983 report *A Nation At Risk* as the start of the education standards movement [National Commission on Excellence in Education, 1983]. As part of this trend, stakeholders in various professional communities have developed standards, further broken down into benchmarks or performance indicators and reflecting learning goals or outcomes, for competencies in: information literacy, educational technology (includes information technology and computing) competency, and science literacy. Identified with science literacy are tool literacy, socio-structural literacy, and resource literacy [Wellborn and Kahan, 2000]. Competency is

another term that is used for literacy; the California State University system for example uses the phrase Information Competency rather than Information Literacy, though nationally Information Literacy appears to be the preferred one. [Brose, 2002]. The proliferation of standards can be confusing; there are state standards and national standards. Professional societies often develop and manage the national standards. A distinction is also made between content standards and generic ones. The *McREL Compendium* is a database of standards and benchmarks that is maintained by Mid-Continent Research and Learning [McREL]. It contains more than 260 benchmarks for science literacy in grades K-12 and 4149 benchmarks for all standards in all disciplines, grades K-12. Table 1 highlights the important literacy standards accepted at the national level, the literacy promoted, the professional society (community) or group that publishes/manages the standard, the audience group for the standard, and the number of standards together with number of benchmarks or performance indicators or outcomes.

#	Name	For What?	By Whom & When?	For Whom?	How Many?
1.	Information Literacy Standards for Student Learning	Information Literacy	[AASL & AECT, 1998]	K-12 students	9 standards; 29 performance indicators
2.	National Content Standards (including National Science Education Standards, National Geography Standards, etc.)	Literacy in Subject Content areas: Health, History, Geography, Physical Education, Art, Math Economics, Social Studies, English Language Arts, Science, Foreign Language	Various professional societies in the different 'content' areas manage these standards	K-12 students	12 standards Within each discipline ('content' area), there may be more subjects; for example, the Science content standards B, C, D are standards for Physical, Life & Earth & Space Sciences
3.	Information Literacy Competency Standards for Higher Education	Information Literacy	[ACRL, 2000]	Higher education (undergrad)	5 standards 87 outcomes
4.	National Educational Technology Standards	Information Technology Literacy	]ISTE, 1998]	K-12 Students, Teachers, Admins.	18 areas, 165 indicators
5.	Project 2061: Benchmarks	Science Literacy	[AAAS, 2000]	K-12 students	859 benchmarks

Table 1: Educational standards (accepted at the national level) for various competencies

Facilitating standards-based discovery of materials is a NSDL need [Fulker and Janee, 2002]. The specific need for 'aligning competencies' across multiple disciplinary standards surfaced more recently at the DLESE/NSDL Developer's Workshop in February [Sumner, 2003]. There have been a few attempts to correlate information literacy standards with selected national subject matter standards; similarly, attempts

to correlate information literacy with technology skills have also been made [Spitzer et al, 1998], but none of these use a vocabulary that transcends disciplines or catalogs the learning resource with the literacy outcome. For example, the *McREL Compendium* offers correlations between benchmarks but none between information literacy and science, health literacy. A search for "information technology" retrieved 9 benchmarks (across grades and disciplines) but none for information literacy. *Align to Achieve* has a standards database alignment system that uses the *McREL Compendium* and provides a searchable database. Again, ITL is not aligned/mentioned here either [A2A]. Another NSDL project, *StandardConnection* is investigating automatic generation of standards metadata using these databases [Sutton et al, 2001] but it is also not aligning standards across disciplines. Table 2 shows textual correlations at a broad level between information literacy and science standards [Laherty, 2000].

ACRL Information Literacy Competency Standard for Higher Education	**National Science Education Standard (Content Standard)**
Standard 1: The information literate student determines the nature and extent of information needed. Standard 3: The information literate student evaluates information and its sources critically and incorporates selected information into his knowledge base or value system.	Standard A Science as inquiry
Standard 2: The information literate student accesses needed information effectively and efficiently. Standard 4: The information literate student, individually or a member of a group, uses information effectively to accomplish a specific purpose.	Standard E Science and technology
Standard 5: The information literate student understands many of the economic, legal, and social issues surrounding the use of information and accesses and uses information ethically and legally.	Standard F Science in Personal and Social Perspectives

Table 2: Correlating Information Literacy with Science Literacy (general – topmost level)

A vocabulary that maps across ITL and science, health literacy outcomes, and the assignment of these terms to resources, are not sufficient to fulfill the needs of an information and technology literacy service for NSDL. Experience with DLIST and the literature show that manual deposit processes and metadata generation are bottlenecks in building collections; collection building and use needs to be centered on communities that already exist and digital library technologies that promote interoperablity. Thus three other factors, **Intellectual Property**, for understanding the rights characteristic of ITL materials, **Interoperability** (technical, content and organizational) for facilitating the federation and/or harvesting of materials from diverse communities of creators, and **Heterogeneity,** through service to professional and underrepresented communities, such as Native Americans, are important aspects of our approach to designing and developing DLISTConnection.

While the information age may be characterized by the "visible instability of the public domain" it is nevertheless true that the production of **Intellectual Property** (IP) is dependent on the production of a public domain, as are the returns IP generates [McSherry, 2001]. The intellectual commons is not only a source of new, "propertizable" ideas; it also legitimates the activities of information-based industries. There is some support for this argument when one considers how often commercial database publishers make their training materials freely available (for example, see letter of support from major database publisher DIALOG which permits DLIST to harvest training materials from their site). Yet 'data' which cannot assert IP protection has increasingly been relocated to the private domain in recent years and it certainly does not mean that a genre of commercially produced materials exists that may be 'gifted' into

the commons. Many of the materials supporting the acquisition of information and technology literacy may fall into the public domain (for example, all material created at U.S. government expense by U.S. government employees) or be covered under the Open Content Licenses (certainly documentation fits here) [Open Content]. Others are created as works-for-hire or by academic faculty who may or may not own the IP to it (their institutions may own it). Copyrights and their management are complex, and will be key to the success of any future enterprise, such as ours, where the knowledge products are to be shared and re-used. Advice found on the Creative Commons website, which makes all sorts of license templates available for all sorts of materials including courseware sums the complexity quite simply: "Offering your work under a Creative Commons license does not mean giving up your copyright. It means offering *some* of your rights to any taker, and only on certain conditions." [Creative Commons]. The mission of the university, to educate and make knowledge openly available, and the incentive for commercal publishers to reach new markets (for example Netscape demonstrated how new markets can be gained when the product is given away for free) underscore that rights issues surrounding ITL materials are different from research articles (MIT, for example, makes course materials openly available). But, the availablity of these materials in the Internet Commons does not mean that they have been gifted into the public domain or that digital libraries and repositories can indiscriminately link to them much less borrow the content; rights associated with harvesting of content and metadata, must be explored with commerical and professional society publishers, faculty, attorneys, and authors and rights elements made available in metadata.

The architectural goal of NSDL is to accommodate "heterogeneous participants, content, and users through a *spectrum of interoperability*' [Lagoze et al, 2002]. DLISTConnection will be an Open Archives compliant data and service provider by integrating Eprints and DSpace collections; we will experiment with automatic content harvesting methods of non-OAI compliant databases/websites, and implement reference linking of learning objects using methods documented by [Open Archives, DSpace, Bergmark et al, 2001; Chakraborti et al, 1999; Chen et al, 1998; Open Citation; Hitchcock et al, 2002; Liu, 2002]

To further NSDL's **Heterogeneity** objectives and along the lines of building upon existing communities, DLISTConnection will involve Native American graduate students in internship positions as outreach specialists working with American Indian teachers, students, and librarians. Cajete [1999] has documented the effectiveness of teaching science and enhancing science literacy among Native American students through a culturally sensitive approach. Roy & Larsen [2002] have explored the application of Cajete's model to the creation of a virtual library for a tribal college. Their work will inform the design of the internships. Interns will be selected from the SIRLS Knowledge River program; this award-winning program focuses on library and information issues relevant to Hispanics and Native Americans. Loriene Roy of the University of Texas at Austin or Gregory Cajete of the University of New Mexico will help design the internships and coordinate the work with Native American nations and tribes.

III. Work Objectives

This proposal has five work objectives and under each objective we briefly describe the prior work that has been done and the advances we plan. Project activities are to aggregate a collection of learning resources such as training, help, documentation, guides, and other instructional materials created for instruction in the use of digital libraries, electronic resources, and information technologies by automatic and manual harvesting of content (and metadata, when available) from professional communities, commercial and government publishers, and specific to health sciences, correlate and align information and technology literacy with science and health literacy and create a controlled vocabulary, organize a workshop to foster understanding the nature of rights for ITL materials produced commercially and as works-for-hire, but by custom presumed and used as if it is a part of the intellectual commons, and catalog the resources with rights and other metadata facilitating discovery by literacy outcomes. We will investigate reference linking, citation indexing in learning materials and conduct a study to assess the use of these materials and the effectiveness of their metadata. Finally, NSDL projects, Health Education Assets Library (HEAL) and AskNSDL Virtual Reference Desk (VRD) will serve as testbeds.

1. Assemble and catalog an OAI-compliant collection of 3000 learning objects supporting approximately 87 information and technology literacy outcomes and performance indicators

1. A. Prior Work: We have implemented DLIST, an OAI-compliant eprints server. Initiated officially in October 2002, it has more than 170 registered users and author self-deposited materials of various types (journal articles, conference proceedings, presentations, user guides, and tutorials). Figure 2 shows sample Eprints metadata for a resource, the PubMed Help page, http://www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhelp.html#PubMedSearching. Eprints metadata is not intended for instructional materials; there is no information about which literacy this resource can meet but some rights metadata may be given using Element #5, Public Domain.

1. Eprint Type: Guide
2. Later Version: N/A
3. Commentary On: -
4. Status: Unpublished
5. Public Domain: Yes
6. Editors Names: -
7. Title: PubMed Help
8. Subjects: Information Literacy
9. Year: 2003
10. Abstract:
11. Conference: N/A
12. Series: -
13. Conference Date: -
14. Location:
15. Volume: -
16. Number: -
17. Publisher: NLM
18. Commentary on: -
19. Alternative Locations: http://www.ncbi.nlm.nih.gov/entrez/query/static/help/pmhelp.html#PubMedSearching
20. Keywords: Search guide
21. Additional Information:
22. References: -
23. Comments and Suggestions:
24. Format: HTML
25. You may offer an additional brief description of the format.
26. Security level (of this document. Who is allowed to download it?) Anyone

Figure 2: Eprints metadata for a public domain material

1. B. Planned Advances:

1. B. 1. Harvest the ITL materials

The goal of this task is to characterize the nature of the ITL resources and repositories that exist in order to prepare the groundwork for developing an ITL service protocol. In building collections of ITL materials we face several significant problems that will be resolved through collaboration and policies:

1. Lack of metadata: We know the existence and URLs of specialized portals and professional communities creating/aggregating high quality materials in this area; however, most of them do not generate metadata for the content they themselves aggregate and may never do so.
2. Multiple versions of documents: Most ITL materials are versions (software and interfaces are released thus), and also are updated often; the frequency of update may always be unpredictable.
3. Proliferation of relevant links: Often, ITL materials link to other materials that are elsewhere on the Internet commons.
4. Heterogeneity of communities: In terms of both content and organizational structure ITL materials are created and used by distinct communities of professionals (see letters of support for a sample).

Focused crawling, mining communities, and utilities such as *Wget* are some of the automatic methods to build collections and we will experiment with all three methods as documented by [Bergmark, et al, 2002; Chakraborti et al, 1999; Chen et al, 1998; Wget] to determine which work best with the different communities of ITL materials creators and materials. Table 3 shows a breakdown of a few of the communities from whom we will automatically harvest materials, the content harvesting method(s) we plan, and the approximate number of items estimated available for transfer into the collection.

Community	Type of repository	Harvesting method	Number of items
ACRL	Content management	Manual, Human	~300
LOEX	Website	Wget	~2500
DIALOG	Website	Wget	~1250
MedlinePlus	Website/Database	Focused crawling	~500
PubMed	Website/Database	TBD	~100

Table 3: Content harvesting of ITL materials

1. B. 2. Describe the ITL materials

The sample Fig. 1 showing the metadata for the PubMed search guide is woefully inadequate in meeting the information discovery needs of users. Since Eprints software was not intended for learning objects, there is no real element for adding standards/outcomes metadata. Our interest in fostering communities and standards-based practices means that we will work with members of ACRL Instruction Section and Science & Technology Section, UA Libraries, and AHSL instructional librarians to build OAI-compliant DSpace communities to catalog and store the metadata (and content, when necessary). We will use Dublin Core-Education (DC-Ed) standards framework and GEMCAT4 [DC, GEMCAT] for manual metadata creation. Self-archiving by experts in the area of ITL such as Bonnie Gratch-Lindauer will be encouraged (see letter of commitment). Metadata will be created for all ITL materials in a two-step process. In the first step, metadata will be completed for elements Title, Creator, Description, Contributor, Publisher, Date, Format, Identifier (system supplied), Language, and Coverage. In the second step (2nd year of project) metadata, with the appropriate controlled vocabularies, will be completed for the education-specific elements (audience, etc.) and for Subject, Type, Source, Relation, Coverage, and Rights.

2. Develop rights metadata for resources that transition from being proprietary or copyrighted to those shared under Open Content licenses and identify rights associated with harvesting

2. A. Prior Work: DLIST collections will include materials that fall under a number of different types of intellectual property (IP) agreements; they may be materials that have been created as work for hire or be items subject to previous copyright transfer agreements. Experience with DLIST suggests that it is time to broaden our understanding of rights management beyond the usual focus on protecting copyright holders with encryption, licensing, and other methods designed to thwart unauthorized use. An initial step toward this understanding occurred in the fall of 2002, when SIRLS organized a Copyright Colloquium co-sponsored with the Arizona Chapter of the American Society for Information Science and Technology and the UA Library. Professor Siva Vaidhyanathan, author of *Copyrights and Copywrongs: The Rise of Intellectual Property and How it Threatens Creativity* spoke on "The Future of Fair Use and the Corruption of Copyright" [Vaidhyanathan, 2001; Vaidhyanathan, 2003]. In spring 2003, a SIRLS graduate student is completing a DLIST IP Internship. The template created by Project Rights Metadata for Open Archiving [Project RoMEO, 2002] informing users of copyright policies is being extended with information about specific IS & IT publishers' copyright agreements and their current practices of rights managements. DLIST has in place policies for collection development, metadata, and user privacy.

2. B. Planned Advances:

2. B. 1. Organize a Copyright Workshop. We will convene a two-day workshop on the University of Arizona campus regarding IP rights management for ILT materials in repositories. The workshop will be scheduled toward the end of the first year. The topics to be addressed include:

- First North American Serial Rights vs. transfer of copyright from author to publisher--the impact on DLIST submissions
- The processes of self-archiving, harvesting and their application to ITL materials

- The role of DLIST in educating authors and publishers regarding open sharing of copyrighted material
- The advantages of and limits to open sharing from publishers' perspectives
- Learning objects as works for hire and the implications for reusability
- Metadata elements for rights and permissions information

We will invite publishers, authors, librarians, attorneys, and other interested parties. Featured speakers will include experts in the area of digital copyright such as Georgia Harper of the University of Texas, Siva Vaidyanathan of New York University, Arnold P. Lutzker, Lutzker & Lutzker LLP and ADL SCORM. Representatives from a consortium like CNI, professional societies such as LOEX, LRACCC, ACRL, SLA, and publishers/editors such as DIALOG and Elsevier have expressed their willingness to work with us (see letters of support). The workshop will be organized into a series of panels, lectures, and discussions. Selected invited lectures will be web cast to a wider audience and proceedings of the conference will be archived in DLIST. An immediate outcome will be the development of a clear statement of DLIST policies, positions, and procedures regarding copyright of ITL materials. This report will include at a minimum, the specific rights metadata that need to be created for ITL resources. A larger outcome will be a model for handling rights and permissions associated with harvesting/federation.

2. B. 2. Articulate Rights Associated with Harvesting of ITL materials:
Copyright issues surrounding digital libraries often utilize different points of view. At the simplest level and from a resource point of view, three types of resources exist [Decomate II, 1998]: public-domain/free of cost materials, materials available under national, campus or consortium licenses between the library and publishers, and materials available on a pay-per-view basis under license. For ITL materials more categories than these exist and other views to elicit rights and permissions are needed. Project RoMEO, currently underway is investigating the rights issues surrounding the 'self-archiving' of research in the UK academic community under OAI-PMH [op. cit.]. Through a series of stakeholder surveys, the processes associated with self-archiving and rights issues surrounding them are being identified. Six processes have been identified in self-archiving and these include: authorship, archiving or preprint, archiving of refereed paper, metadata repository as data providers (DPs), metadata repository as service providers (SPs), and access of paper by end-user. Rights issues of authorship, identified thus far, include: who owns copyright (academic writing paper or institution), awareness of copyright ownership (does the academic know who owns copyright), how is copyright practiced (by custom or contract), etc. DLISTConnection will be a DP and SP; and, processes surrrounding ITL materials will be significantly different from those of research articles and preprints. While these materials may need to be openly available to stimulate use, assignment of re-use and distribution rights may be complex and need to be communicated via metadata.

There are many digital rights management metadata languages such as ODRL, xRML, and PRISM. The Publishing Requirements for Industry Standard Metadata (PRISM), version 1.1, is an XML metadata (draft) standard that articulates clear relationships among rights and the DC metadata frameworks. It is being developed by IDEAlliance, a consortium of publishers whose members include technology vendors such as Adobe, database publishers such as Lexis-Nexis and magazine publishers such as Time Inc. and McGraw-Hill. PRISM can be defined as an RDF/XML document that uses the Dublin Core vocabulary and a major application is the tracking of rights related to content through PRL [PRISM, 2003].

3. Correlate and align the information and technology literacy outcomes from ALA, AASL & AECT, NETS, and Science Literacy Benchmarks and Health Literacy and develop a controlled vocabulary

3. A. Prior Work: DLIST uses only the subject Information Literacy to identify that a resource is about Information Literacy. ERIC thesaurus [2001] and GEM have less than a handful terms for this topic.

3. B. Planned Advances:

Create crosswalks and a controlled vocabulary for ITL indicators

ITL indicators are the benchmarks, outcomes, which map across standards. Crosswalks are tools that map "the relationships and equivalencies between two or more metadata formats. Crosswalks support the ability of search engines to search effectively across heterogeneous databases, i.e. crosswalks help promote interoperability" [DC, 2001]. A crosswalk can also be used to correlate and align competencies in various standards. We will first create a crosswalk between the performance indicators/benchmarks in the following standards: *ACRL Information Literacy Competency Standards for Higher Education*, ISTE's *National Educational Technology Standards (*NETS), AASL & AECT *Information Literacy Standards for Student Learning* and AAAS Project 2061's *Benchmarks for Science Literacy* and *Health Literacy*. We will next create a controlled vocabulary for the aligned indicators. Table 1 shows the number of outcomes in each standard; correlation and alignment may distill that number to about 100.

Figure 3 shows some different types of controlled vocabularies. A controlled vocabulary may be defined as a list of equivalent terms in the form of a synonym ring or a list of controlled terms in the form of Authority Files. We have permission to use the *McREL Compendium* (copy of letter attached) and we will use identify, develop, and test synonym rings for ITL indicators that have benchmark correlations with the science and health standards before we create an authority file (list of subject headings) based on this.

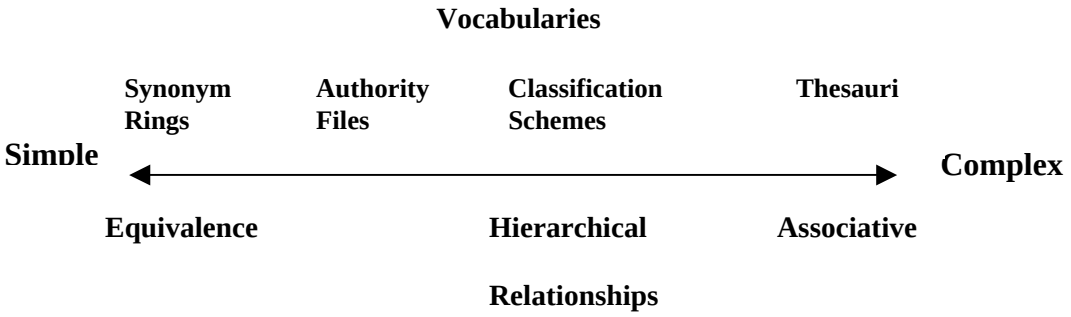

Fig. 3 Types of controlled vocabularies (Source: [Rosenfeld and Morville, 2002])

4. Build an interoperable and heterogeneous digital library by integrating Eprints with DSpace communities, and experimenting with harvesting, searching, and citation indexing

4. A. Prior Work: Three areas of DLIST architecture development are relevant: 1) configuration of Eprints to fit learning materials, 2) citation indexing studies; 3) initial experiments towards a federated library system with ARC.

Eprints Configuration to Fit DLIST User Needs: The eprint types that come with the default eprint configuration have been extended to include teaching materials. Similarly, the eprint subjects have been modified to reflect DLIST goal to be a discipline repository. These and other changes to Eprints have been described in the literature [Coleman and Bracke, 2003].

Citation Indexing Investigations: Three open source software were compared, *Research Index, Reference Linking APIs* and *ParaTools* that are currently available for reference linking and citation indexing. All three software are in varying stages of development [Bergmark; Citeseer; Jewell]. *ResearchIndex* or *CiteSeer* offers the most functionality and can be used to index Postscript and PDF research articles on the Web, but it has the least documentation for integration with Eprints. *ParaTools*, or the *ParaCite ToolKit*, provides a set of Perl modules that extract the references from bibliographies in documents and then parses these references into their component parts (such as author, year, volume, and

title). Para Tools were the easiest to integrate with Eprints and this is currently implemented in DLIST with searches for references being retrieved by Google. The *Java Reference Linking APIs*, made available to us by Donna Bergmark, are not a citation tool but a set of APIs to build one; requiring significant programming work and a CS graduate student is currently working with it [Bergmark].
Initial Experiments with Federation: We were able to successfully harvest the DLIST collection metadata, complete a focused metadata harvest of ArXiv, and search both collections using ARC, a harvester and search interface developed at Old Dominion University [ARC; Liu, 2002].

4. B. Planned Advances:
4. B. 1. DLISTConnection:
In building collections with distributed contents, such as NSDL, the benefits of using a component-based, distributed architecture have become clear. Suleman and Fox describe a framework for developing open digital libraries in which "digital libraries are modeled as networks of extended Open Archives, with each extended Open Archive being a source of data and/or a provider of services" [2001]. Thus, DLIST is envisioned as a Federated Digital Library System. Some of the characteristics of such a system are

- Distributed collections and services that are administered by independent parties.
- Heterogeneity throughout the system in content and technical implementation, unified by defined protocols for interoperation. This will include variety in the formats and genres of objects stored, variety in rights management schema, and variety in the types of user interfaces to the services both for information discovery and information deposit.

DLISTConnection will be the metadata repository that end-users will use to search from the ITL collection proposed, in addition to other existing and future collections. ARC will be used for metadata harvesting and for providing a unified search and browse interface to the user. Metadata will be harvested via OAI-PMH, and as noted in Work Objective 1, we will explore specific techniques for harvesting from non-OAI compliant data providers and topic relevant communities. DLISTConnection presents a simple, quick interface for the user although the problem of consistency of the metadata is possible. Metadata will be updated on a regular basis, with the exact frequency to be determined once the service is under way. Figure 4 shows how the metadata harvesting process works and the logical view of DLISTConnection.

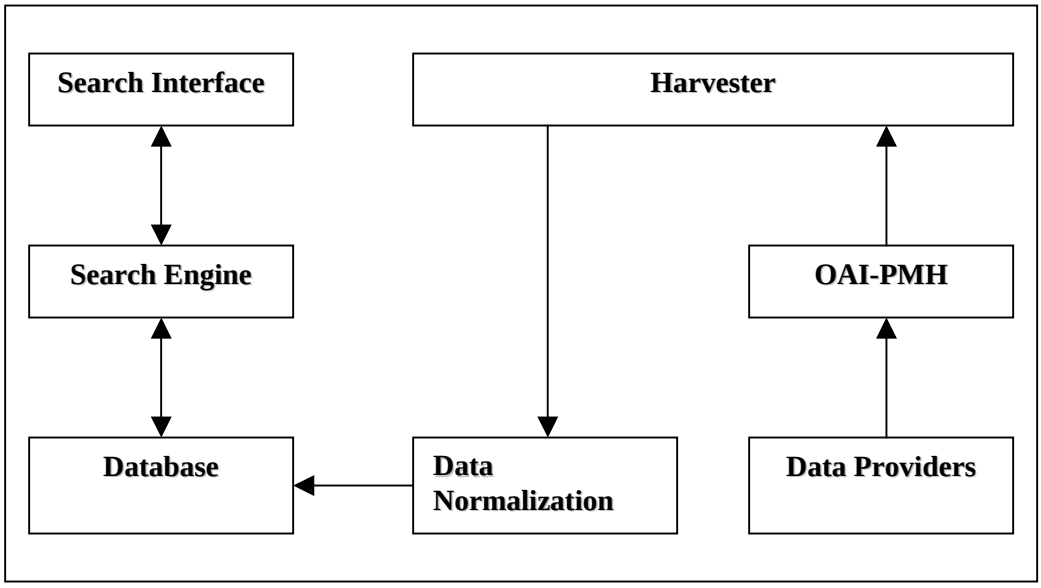

Figure 4 DLISTConnection

DLISTConnection will be designed in such a manner that it can take advantage of services and software modules (components) provided by the NSDL Core Integration group or other DL initiatives. This could include user profiles, annotations and reviews, or other services (Lagoze et al, 2002). OAI interfaces allow

for harvesting of DLISTConnection content by NSDL and other service providers. Steps for moving DLIST towards a federated model are to: implement DSpace for housing the ITL collections; experiment, streamline, and document the process for automatic harvesting of ITL collections; experiment, streamline, and document the process of metadata generation for ITL materials; implement ARC for unified browse and search through DLISTConnection; and, experiment, streamline, and document the process for reference linking and citation indexing of learning objects. This is shown schematically in Figure 5.

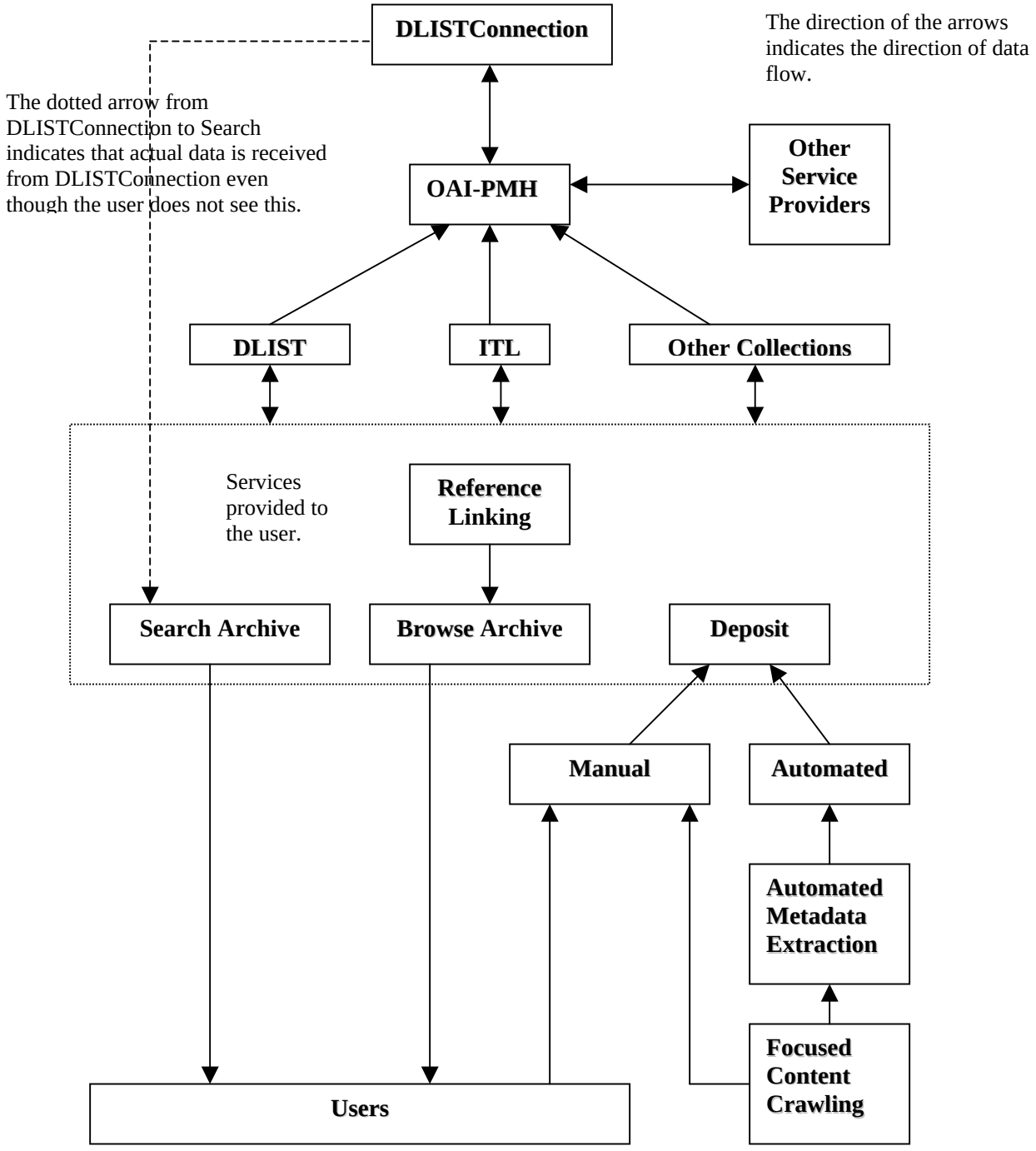

Figure 5 System Architecture

5. Evaluate using NSDL test beds (HEAL and AskNSDL VRD) and informetrics methods to assess how items are being used and understand the effectiveness of the standards and rights metadata

5. A. Prior Work: Citation analysis is who cites whom, how many times, in which journal, etc. and includes techniques such as co-citation analysis, bibliographic coupling, citation content and co-word analysis. It provides statistical ways to map knowledge structure and communities in a discipline, identify research fronts, etc. It is also known as informetrics. Informetrics incorporates bibliometrics, newer areas of cybermetrics, webmetrics, linkmetrics, and e-metrics where methods focus on measures, analysis and use of online documents and/or networked information services. Citation analysis is also closely tied to Scientometrics, the study of science scholarship. While citations in research literature have been studied extensively, use of citations in learning materials is not well documented. But, the importance of citations can be seen in the emergence of search engines such as Teoma and database services such as GetCited [Teoma, GetCited]. Thelwall and his team have done extensive work with links on academic websites but they do not focus online learning materials [Thelwall, 2003]. An informetric study, in progress, is investigating the creation and use of citations and links in web learning materials in undergraduate Geographic Information Science (GIS) courses [Coleman and Karthik, 2003]. Preliminary data shows that citation motivations do fit what has been developed earlier by [Moravcik and Murugesan, 1975] such as conceptual, operational, organic, evolutionary, confirmational, but more types are needed.

5. B. Planned Advances:

5. B. 1. Assess the use and usefulness of ITL materials, standards and rights metadata

Using the categories of citation types and uses derived from the GIS informetric study we will design and conduct two types of assessment, a citation content analysis and a small user test. A random selection of ITL materials will be analyzed and categories of ITL citations described. The citation content analysis will adapt the methods documented in Thelwall [op cit.]. The user test will be modeled along the lines of a standard usability test, where a set of tasks ranging from easy to impossible given and users are observed with the goal of understanding how standards and rights based information seeking behaviors are satisfied.

5. B. 2. Evaluate DLISTConnection with NSDL test beds

Two projects, NSDL HEAL and AskNSDL VRD, have agreed to collaborate. HEAL, the Health Education Assets Library, a collaborative NSDL research project, has a multimedia collection of 3000 items and continues to grow. HEAL is already planning to implement software bridges using OAI and this makes it a natural fit with DLISTConnection. AskNSDL VRD, the Virtual Reference Desk, is developing a service for virtual reference and plans to integrate our service. VRD can use the literacy-mapped materials as a knowledge base to explore tiered level virtual reference desk services. Both of these collections will be used to determine the efficiency of DLISTConnection in being a service provider.

IV. Project Timeline

	Months								
Work Objective and Name of Lead	**0**	**3**	**6**	**9**	**12**	**15**	**18**	**21**	**24**
Objective 1: Collection building – Malone	✓	✓	✓	✓	✓	✓	✓	✓	
Objective 2: Rights – Malone		✓	✓	⊛	✓	✓			
Objective 3: Crosswalks, Vocabulary – Coleman			✓	✓	✓	✓			
Objective 4: DLISTConnection – Bracke	✓	✓	✓	✓	✓	✓	✓	✓	✓
Objective 5: Evaluation – Coleman					⊛	✓	✓	✓	✓

Figure 6 Timeline for work objectives.

The symbol ⊛ indicates tentative schedule for workshop and preparatory work for evaluations such as submitting for human subjects approval.

V. Collaboration and Sustainability

Several potential sources for sustainability exist. Commercial database publisher DIALOG, professional society publishers such as LRACCC, ALA ACRL, and membership-sustained, not-for-profit clearinghouses, namely LOEX, have indicated a willingness to explore how DLISTConnection can be developed and used. Since they are also content developers with well-developed business goals and communities of service, preliminary forays resulting in a 'hand-over' plan are likely. Second, the business plan drafted as a result of the internship will be pursued further with SLA, CNI, and McREL.

VI. Significance to NSDL

The significance of DLISTConnection lies in:

1. Aggregation of key genres of ITL material – help, training, documentation, database search guides - currently not represented in the NSDL collections.
2. Application of innovative metadata elements to support discovery of materials linked to information literacy standards that have been mapped to science literacy for -K-20
3. Application of innovative metadata elements to support the transition of proprietary and/or copyrighted materials to open content and creative commons licenses.
4. Application of informetrics techniques to a new body of knowledge to support discovery and to document use of assistance systems such as software documentation, corporate training, database searching instructions, and information literacy tutorials.
5. Support for collaboration between and among librarians and faculty in the STEM disciplines on learning strategies that integrate science literacy and information literacy.

VII. Project Team and Management

The team consists of Anita Sundaram Coleman, an Assistant Professor at SIRLS, University of Arizona, Tucson, which she joined in 2001. Coleman's doctoral dissertation investigated information work at health science libraries; findings from user studies were used to develop prototype tools that may be embedded in scholarly journals to support practitioner-based information work (publishers called them, value-added services to online publications). Coleman brings experience in the development of technological tools to support information work and the ability to forge interdisciplinary and community connections.

Cheryl Malone is an Associate Professor at SIRLS since 2001 and prior to that was a member of the faculty at the University of Illinois, at Urbana-Champaign from 1996 to 2001. Her background includes professional positions at the University of Michigan and the University of Texas at Austin libraries. She has published widely in library history, specifically the history of library collections and services for African Americans. Her research projects include a study of Internet access in public libraries supported by the ALA's Carroll Preston Baber Research award and an investigation into the North American Industry Classification System. She serves as the Faculty Advisor for the Electronic Depository Pilot Project, a partnership between the UA Library, the US Government Printing Office, and SIRLS. She has extensive practice with instructional assistance systems, information and technology literacy instruction, and under-represented groups such as African-Americans and Native Americans.

Paul Bracke is the Head of the Systems and Networking department at AHSL and is involved in the development of systems-architecture models to enable effective planning of integrated library systems. He has previous systems experience working for University of Arizona Main Library and University of Texas Medical Branch at Galveston. He brings systems expertise and keen interest in fostering the development of distributed digital libraries and standards for interoperability.

Three graduate students each year, a total of 6 during the two years of project duration, and 5 graduate interns in the second year of the project will complete the DLISTConnection team. The three graduate students will meet disciplinary requirements and fulfill specific roles: 1 graduate student each year from Computer Science will serve as the Administrator and Systems Developer, 1 graduate student each year

from LIS will serve as the Metadata Specialist/Technical Writer (responsible for writing documentation of automatic harvesting and metadata generation processes), and 1 graduate student in the first year, preferably a SIRLS student with a Law degree or working towards one, will serve as the Intellectual Copyright Specialist with another graduate student in the second year helping in the Evaluation tasks. Internships are proposed only for the second year of the project. Three of the five graduate interns will be drawn from the SIRLS Knowledge River program and they will work as DLIST Outreach Specialists Two other interns will be drawn from the graduate program in the UA Eller College of Business and SIRLS and they will write a business plan for DLIST besides serving as proof readers for the documentation of DLISTConnection.

The DLIST group meets weekly (Coleman as Project Director) and we will continue with this arrangement. Other smaller meetings for focused tasks will be scheduled on an as needed basis.

VIII. Impact and Expected Results

Deliverables include:

1. The ITL collection(s) of help, training, search guides, and other instructional materials supporting 87 information and technology literacy competencies and performance indicators in the health sciences
2. ITL materials mapped with literacy outcomes and rights metadata
3. A crosswalk aligning benchmarks, performance indicators, and outcomes of the ACRL *Information Literacy Competency Standards for Higher Education*, AASL & AECT *Information Literacy Standards for Student Learning*, ISTE *National Educational Technology Standards for Students*, with AAAS Project 2061 *Benchmarks for Science Literacy* and *Health Literacy*
4. A subject Authority List for Information and Technology Literacy outcomes
5. DLISTConnection, one-stop shopping center for quality ITL resources and communities
6. Assessment of user behavior with ITL and rights metadata
7. Testbed results of DLISTConnection with NSDL HEAL and AskNSDL VRD
8. Workshop and reports on intellectual property issues surrounding ITL materials
9. Documentation for use of digital library components for integrating core repository software (like Eprints with DSpace), reference linking, focused crawling module for content harvesting), automated metadata extraction module, and automated deposit module, and metadata harvesting module
10. Outreach activities for new communities of Native Americans integrating ITL, science, health literacy
11. Reports & plans: Rights associated with harvesting, DLISTConnection business plan

DLISTConnection provides for a new form of intellectual access to unique materials that are currently not collected by NSDL. In service to K-20+ audiences, science educators, librarians, technologists, and end-users, DLISTConnection harvests and federates learning resources from diverse locations. It creates a vocabulary and provides crosswalks connecting the outcomes in information and technology literacy standards to health and science literacy benchmarks and maps them to resources that support the acquisition of these skills. It facilitates instructional information discovery and use through its federating and mapping functions as well as re-use and re-purposing through the application of rights metadata.

DLISTConnection brings new kinds of resources (training and help materials created for use of digital libraries and electronic resources) and new communities of creators and users (instructional librarians who are mostly women and Native American students) from diverse environments (professional societies, commercial and government publishers) into the NSDL, thereby increasing the effectiveness of NSDL as a "learning environments and resources network." Rights associated with harvesting and linking of learning objects are made clearer. This proposal goes beyond the voluntary, sporadic, and hard-to-sustain efforts of various librarians' organizations that maintain Web pages with links to information literacy resources, resources often known only to the respective organizations' members to lay the foundation for how the key characteristics of future scholarship--interdisciplinarity (managing complex learning of tools and concepts across disciplines), collaboration, and disintermediation--can be modeled in a Digital Library service.