

Please send comments to John M. Saylor, NSDL Director of Collection Development (jms1@cornell.edu).

1.0 INTRODUCTION

1.1 MISSION OF THE NATIONAL SCIENCE DIGITAL LIBRARY (NSDL)

The mission of the NSDL is to enhance science, technology, engineering and mathematics education through a partnership of digital libraries joined by common technical and organizational frameworks. Individually and collectively these partners engage and inform multiple clienteles, using shared resources to serve many communities of users, each with its own level of knowledge and learning model. The NSDL embodies long-standing library traditions of service, longevity, equal access, fair use, and privacy, as well as innovations that foster a spirit of inquiry and the accessibility of science to all.

1.2 DEFINITION AND OBJECTIVES OF THE NSDL COLLECTION

1.2.1 DEFINITION

The **NSDL Collection** is *collection of sets of resources*. These *sets of resources* are also referred to as *collections*.

We shall use the term *collection* for the remainder of this document to mean an organized arrangement (i.e. set) of *resources*. Such *collections of resources* may be formally or informally published, suggested or contributed by NSDL users, or assembled automatically and dynamically. A *collection* in the NSDL may have been organized by a person or organization, or may be organized automatically. As a general rule, *collections* that are considered to be part of the **NSDL Collection** are not actually held within NSDL-owned computers or storage systems. Instead, individual *collections* typically are held and managed by their owners or providers.

The term *resource* is used in a general sense to describe anything that might be included in the **NSDL Collection**. In the **NSDL Collection**, *resources* are usually managed as “sets of *resources*” known as *collections*. *Resources* might be static or dynamic materials, formally or informally published, contributed by users, developed by information or domain specialists, or generated dynamically. The term *resource* makes no distinction between items and *collections* of items.

1.2.2 OBJECTIVES

The **NSDL Collection** supports the mission of the NSDL. The **NSDL Collection** covers all aspects and all levels of science, technology, engineering, mathematics and medicine (STEM) information that could potentially be used for educational purposes.

1.3 PURPOSE OF THIS DOCUMENT

This document is intended for users of and contributors to the **NSDL Collection**. This document describes a Collection Development Policy and strategy for the **NSDL Collection** (accessible at <http://nsdl.org>.)

It covers the:

- NSDL mission and objectives of the **NSDL Collection** and the purpose and maintenance of this document (Section 1.0 Introduction),
- Intended users of the **NSDL Collection** (Section 2.0 Clientele),

- Scope of the **NSDL Collection** – subjects, formats, etc. covered (Section 3.0 Scope),
- Components of the **NSDL Collection** – *resources*, *resource* organization and description, persistent naming (Section 4.0 **NSDL Collection** Components)
- *Resource* identification and selection responsibility (Section 5.0 *Resource* Identification and Selection Responsibility),
- Selection criteria, scientific and/or educational quality, accessibility, items not selected (Section 6.0 Selection Criteria and *Resource* Quality),
- **NSDL Collection** maintenance, evaluation and archiving (Section 7.0),
- **NSDL Collection** History (Section 8.0),
- Strategies and priorities for building the **NSDL Collection** – 2003-2005 (Section 9.0), and
- Technical options for *resource* ingest (Section 10.0).
- Recommended Practices and References (Section 11.0)

This document does not cover tools and services for using the **NSDL Collection**. For information about tools and services for using the NSDL Collection see <http://tobeadded>

1.4 MAINTENANCE OF THIS DOCUMENT

This document was organized and compiled by the NSDL Director of Collection Development in consultation with the Core Integration Team and the Content Standing Committee of the NSDL. Others are welcome to provide input. The NSDL Director of Collection Development is responsible for ongoing maintenance of this document.

2.0 CLIENTELE

The **NSDL Collection** is intended to meet the educational and informational needs of a wide variety of users, from pre-school children to self-directed adults, in both formal and informal education settings. Both teachers and learners are important users of the **NSDL Collection**. The **NSDL Collection** also supports the general public's need for information in all areas of STEM.

2.1 EDUCATIONAL LEVELS

The following educational levels are served:

- Pre K
- K-12
 - Elementary
 - Middle school
 - High school
- Undergraduate
- Graduate and research
- Continuing education

2.2 EDUCATIONAL ROLES

The following educational roles are served:

- Student
- Instructor
- Curriculum Developer
- Librarian
- Administrator

General public

3.0 SCOPE

The **NSDL Collection** covers any topic within STEM, including interdisciplinary education and research areas, topics that are at the boundaries of disciplines, and topics that draw from the social sciences in addition to the STEM areas. The *collections* that make up the **NSDL Collection** vary greatly in size, depth, scope, type of materials, level of review and educational level.

3.1 BREADTH

The **NSDL Collection** is intended to be as broad as possible in all STEM areas. However, there are fundamental differences in the breadth of *resources* needed and suitable for different educational levels.

For example, in the K-12 education setting, and particularly in K-8, students work within a more narrowly defined curricula directed by an instructor. For this educational level, the **NSDL Collection** will emphasize materials that are tied to the curriculum, with an emphasis on material of known quality. As a consequence, the **NSDL Collection** will be narrower in breadth than the **NSDL Collection** geared to the undergraduate educational setting.

In the undergraduate educational setting, and often in 9-12 grades, students are expected to explore widely and to use many information sources, including some primary materials. They are also expected to learn to evaluate different information resources, and become more critical users of information. Therefore, materials for secondary and undergraduate education in the **NSDL Collection** will be broader in scope, less focused on particular curricula, and less likely to have had strict quality controls.

The consequence of these kinds of considerations of needs at different educational levels is that the scope, level of review, and quality control will be highly variable, with most *resources* not explicitly reviewed by the NSDL.

.

3.2 SUBJECT COVERAGE

The subject scope of the **NSDL Collection** includes all areas of science, technology, engineering, mathematics, and medicine (STEM). The subjects of the collections in the **NSDL Collection** are currently expressed by a subset of Gateway to Education Materials (GEM) categories. (See appendix A). Therefore, any *collections* and *resources* of potential use (or re-use) in education and learning that relate to these broad areas are relevant to the **NSDL Collection**.

The goal is to cover all these GEM categories. For pragmatic reasons, in the short run the **NSDL Collection** will be stronger in some subjects, but the long term objective is to cover all in equal proportion.

3.3 GEOGRAPHIC AND LANGUAGE COVERAGE

The **NSDL Collection** is initially intended to support science education in the United States. Science is international, and many *collections* and *resources* located in the **NSDL Collection** will be of value to students and instructors outside the USA.

The **NSDL Collection** includes relevant *resources* from around the world, as well as *resources* that describe scientific phenomena anywhere in the world. Some subjects, e.g., earth sciences, are intrinsically global.

The **NSDL Collection** currently tends to have materials in English as that is the dominant language in science, but *resources* in other languages commonly used in schools in the United States are certainly in scope, and will be added as they become available.

3.4 CULTURAL SCOPE

The NSDL program supports several projects that aim to broaden the cultural scope of the NSDL, e.g., ethnomathematics. The **NSDL Collection** development strategy welcomes and encourages the addition of such *resources* to the **NSDL Collection**.

3.5 FORMATS OF RESOURCES

The **NSDL Collection** includes *resources* in a wide variety of formats, including dynamic, interactive and executable *resources*. The **NSDL Collection** may also include information about items that are not available digitally such as descriptions of objects in museum *collections*. *Resources* that may require a proprietary or special piece of software or hardware are within scope. It is important that the format and access or user requirements are accurately described in the metadata provided about the *resource* by the individual *collection* provider.

4.0 COMPONENTS OF THE NSDL COLLECTION –

4.1 RESOURCES

Resources are defined in Section 1.2.1.

4.2 RESOURCE ORGANIZATION AND DESCRIPTION

Much of the power of the NSDL is manifest in services built upon structured information about the entire contents of the **NSDL Collection**, eventually including well-defined but highly diverse relationships among the entities in the **NSDL Collection**. In the short term, most of this structured information takes the form of metadata records, characterizing *resources* individually or in sets of *resources* called *collections*.

A key responsibility of NSDL partners is to provide ample characterization of the *collections* for which they are responsible. In some cases (such as for *resources* that are largely textual) *open-access* to content may be sufficient, because that will allow an increasing variety of automated processes to create the required characterizations. In most cases, however, partners will be responsible for providing metadata that embody key characteristics of their *resource* sets, sufficient to support effective usage by NSDL service providers and, in the final analysis, by NSDL audiences of educators and learners.

The precise means by which partners may fulfill these obligations will evolve as the NSDL infrastructure and services advance, embracing richer forms of metadata harvesting, data mining, and other technologies for organizing and discovering information.

Currently, the NSDL maintains a metadata repository that lists all the *resources* in the **NSDL Collection**, with as much or as little metadata as is known about each *resource*. This metadata may be full records created manually, minimal metadata captured automatically (e.g., by Web crawling) or a combination of the two.

The plans for future development of the metadata repository include an annotation service by which recommendations, reviews and other external information can be linked to *resources*. Eventually the metadata repository will be one key component of a much larger data warehouse that will serve as the logical aggregation point for the multiple entities in the NSDL context and their relationships.

4.3 PERSISTENT NAMING

An important characteristic of science *resources* listed in the **NSDL Collection** is that they have unique, persistent names. For *resources* for which metadata is provided by the [Open Archives Protocol for Metadata Harvesting](#), this is a requirement of the protocol. For *collections* assembled by web crawling, the simplest unique and persistent identifier for the *resource* will probably be the URL. Ensuring persistent URLs for these *resources* will ensure that advanced NSDL features (such as annotations, embedding of science resources in learning objects, and association of science *resources* with state standards) will function reliably. There are potential alternatives to persistent URLs (e.g. embedded metatags with unique names), but these are still in development.

5.0 RESOURCE IDENTIFICATION AND SELECTION RESPONSIBILITY

The **Core Integration Collection Development Team (CICDT)** and the partner libraries share responsibility for the strategic development of the **NSDL Collection**. Selection of *resources* for the **NSDL Collection** may be by members of the NSDL program, by partner libraries, by recommendations from users of the library, and any of these may employ well-documented forms of automated *resource* selection. The term “partner” and “partner library” may eventually take on formal meanings; for now, they apply to organizations that share NSDL commitment to excellence in education and offer *resources* that comply with this policy document, as determined by the CICDT in consultation with the Content Standing Committee.

5.1 PLANNING RESPONSIBILITY

- Planning the **NSDL Collection** development strategy draws from the National Visiting Committee, the Content Standing Committee, and the Core Integration Executive Committee.
- Within the core integration team, the CICDT consists of the Director of Collection Development (John M. Saylor) who has overall responsibility for *collection* development for the **NSDL Collection**. The Director of Publisher Engagement (Michael Luby) is responsible for outreach to commercial publishers

and the Director of Project Relations (Susan Jesuroga) is responsible for outreach to NSDL funded collections.

5.2 SELECTION

- Users of the NSDL are actively solicited to suggest *resources* for inclusion in the **NSDL Collection** via web forms and other methods.
- Partner libraries identify and select *resources* for the NSDL within their scope, subject to the **NSDL Collection** development policy. The CICDT may identify and select *resources* outside the scope of the partner libraries, including broad-based *collections*.

6.0 SELECTION CRITERIA AND RESOURCE QUALITY

Selection of *resources* for the **NSDL Collection** typically occurs by selecting *collections* that are known to be within the NSDL Scope (Section 3), are likely to be of value to the NSDL clientele (Section 2), and that are under the care of an NSDL partner or other reputable organization. A fundamental criterion is the trade-off between the value of the *collection* to the mission of the NSDL and the effort required to bring it into the **NSDL Collection**. Additional factors and criteria are detailed below.

6.1 OPEN OR CLOSED ACCESS

Educational *resources* can be open-access, or access can be restricted, requiring some form of authentication. A common situation is that access to a *resource* is restricted, but there is another *resource* that describes it (e.g., a metadata record) that is open-access.

With most open-access *resources*, particularly text, the NSDL can build some services by automatic processing of content. In general, the **NSDL Collection** will include restricted access *resources* only when they are described by open-access metadata records, provided by the owner of the *resources*, with explicit permission to include the metadata in the metadata repository and use them in NSDL services.

6.2 SCIENTIFIC AND/OR EDUCATIONAL QUALITY OF THE RESOURCE

In the context of the **NSDL Collection**, the scientific and/or educational quality of a *resource* or *collection* is the suitability for use in some part of science education. Scientific and/or educational quality has many aspects – one *resource* may be scientifically accurate but not useful for education; another may be suitable for middle school students, but too narrow for undergraduates. Scientific and/or educational quality is also subjective; it depends on the judgment of individuals.

Ideally, every aspect of the scientific and/or educational quality of *resource* would be carefully assessed from a variety of viewpoints and that information would be available with the *resource*. (For example, information that the Eisenhower National Clearinghouse has reviewed an article and judged it suitable for high school education.) In practice, such information is known for only a small fraction of the *resources* in the **NSDL Collection**.

Commitment to scientific or educational quality—as determined by the CICDT, in consultation with the Content Standing Committee—is a primary criterion for selecting NSDL partners and, by implication, the associated *collections*. Nonetheless, the

“quality” of individual *resources* (especially as perceived by various users of the NSDL) will be variable, due largely to differences among partners in their approaches to *collection* building. Methods will range from individually peer-reviewed *resources* to *collections* developed via automated methods, such as natural language processing.

The NSDL is developing a framework for recording information about the scientific and/or educational quality of *resources*. This framework is likely to include:

- metadata that can be associated with *resources* or *collections*,
- annotations that allow comments and reviews on *resources*,
- information on how the *resource* came into the **NSDL Collection**, including descriptions of the algorithms used in automated methods.

When available, this information will be stored in the metadata repository and available to NSDL services. For instance, a browse service might use this information to restrict its scope to *resources* that have been recommended for high school students.

6.3 SELECTION CRITERIA

The NSDL program encourages broad variety in the *resources* that individual *collections* manage, but *resources* are most valuable to the NSDL and integrated most easily if they satisfy the following criteria:

1. The content of *resource* is appropriate to fulfilling the mission of the NSDL
2. The content of the *resource* matches the subject scope of **NSDL Collection**. Since almost all scientific materials have the potential to be used in some aspect of education, this scope is very broad.
3. The *resource* fills a gap identified via user feedback or evaluation studies.
4. *Resources* that have been developed with funding from the United States Government and especially from the NSDL program have special importance. Other government funding is also a criterion, e.g., other NSF educational projects, the Digital Libraries Initiative, the Institute of Museum and Library Services, science.gov, etc.
5. The *resources* are available with *open-access* or there is reasonably full metadata available for ingest to the metadata repository and use in NSDL services. (See Section 6.1.
6. Quality information is provided especially for *resources* intended for K-12 education. It is important that individual *collections* have clearly stated their criteria for selection in their own *collection* development policy.
7. The *collection* or *resources* satisfies one of the technical methods listed in Section 10.
8. *Resources* are most useful if they are well managed by a reliable source. This includes: good metadata, reliable delivery, expectations of permanence, etc.

6.4 ACCESSIBILITY

Priority will be given to those *resources* that follow federal accessibility guidelines.

6.5 RESOURCES CONSIDERED UNSUITABLE FOR THE NSDL COLLECTION

The **NSDL Collection** will strive not to include *resources* that are:

- “bad science,” e.g., inaccurate information or pseudoscience, except when appropriately tagged or embedded in educational materials that raise awareness on these matters.
- socially undesirable, e.g., pornography.

In a *collection* of this size, no guarantees can be made that all such items are eliminated.

7.0 NSDL COLLECTION MAINTENANCE, EVALUATION, AND ARCHIVING

7.1 REMOVAL OF ITEMS FROM THE NSDL COLLECTION

Though *resources* that are clearly out-of-scope or that become inaccessible may be removed from the NSDL, de-accessioning is not a primary means for maintaining the quality of the **NSDL Collection**. Instead, various means are to be provided for marking content as to quality and for presenting views of the NSDL that adhere to (user-specific) quality thresholds. If a more detailed policy is developed on the removal of materials from the **NSDL Collection**, it will appear here.

7.2 NSDL COLLECTION EVALUATION

Evaluation of the **NSDL Collection** is an ongoing and evolving process that is the responsibility of the Core Integration Team working with the NSDL Educational Impact Standing Committee.

7.3 PERSISTENCE AND ARCHIVING

This is a placeholder to describe policy and strategy for archiving the **NSDL Collection**.

8.0 HISTORY OF THE NSDL COLLECTION

8.1 INITIAL RELEASE, DECEMBER 2002

For the initial release in December 2002, the only method of ingest available was OAI harvesting of metadata with manual review. Priority was given to *collections* that provided metadata for all *resources*, either in or easily mapped to Dublin Core format, and supported the Open Archives Initiative protocol.

The initial release included the individual *resources* from almost all NSDL-funded *collections* that met the technical requirements and a small number of other *collections*, chosen for their diversity of content and low effort needed to collect them. The quality of the metadata was checked manually and transformed to a common format.

In addition, the initial **NSDL Collection** included *collections* identified only by the *collection*, with no information about individual *resources*.

9.0 NSDL COLLECTION DEVELOPMENT STRATEGY AND PRIORITIES 2003-2005

9.1 CORE INTEGRATION COLLECTION DEVELOPMENT TEAM

To date all selection for the **NSDL Collection** has been by the Core Integration Collection Development Team (CICDT). This effort needs to be expanded.

The Core Integration Team has three tasks related to the development of the **NSDL Collection**:

- a. Coordination,
- b. Ingest into the metadata repository,
- c. Selection of *resources* and *collections* not developed by NSDL funded projects or partner libraries.

During 2003-2005 the CICDT focus will be:

- In conjunction with the NSF, the NSDL National Visiting Committee (NVC), and the NSDL Content Standing Committee to create a group of partner libraries that will be responsible for comprehensive *collection* development in focused areas of interest.
- With the team of the iVia Open Source Virtual Library System to implement a system of machine-assisted *collection* development.
(<http://www.dlib.org/dlib/january03/mitchell/01mitchell.html>)
- To publish and refine more detailed **NSDL Collection** development guidelines and priorities.
- To organize a group of volunteer selectors to create *collections* of high-value resources from the Web. This group will contain science librarians, NSF Program Directors and science teachers. Plans for continuation of this group will have to be developed to sustain it beyond the first two years if it is successful.

9.2 PARTNER LIBRARIES

Priority will be given to identifying a pilot group of “partner libraries” committed to scientific and educational quality as discussed in Section 6.0.

Each partner library will take the lead in servicing a portion of the NSDL clientele or a subject area. Each partner library will select and collect *resources* within some part of the total NSDL scope. A partner library has two major responsibilities in developing the **NSDL Collection**. These are:

- Identify and assemble *resources* and *collections* within the identified scope consistent with this policy.
- Provide digital library services to an associated clientele, thereby gaining continual feedback on the quality and usability of the *resources*. .
- Provide information about the *resources* to the **NSDL Collection**. The partner library will expose its *resources* for ingest by the **NSDL Collection** using some or all of methods described in Section 10.

To meet these objectives, partner libraries should follow the selection criteria in Section 6.3 or employ equivalently selective *collection* development practices.

9.3 PRIORITIES 2003-2005

This section proposes the priorities that should guide *collection* providers, partner libraries and the core integration team in selecting *resources*.

It is **proposed** that the CICDT be guided by the following priorities. It should be understood that this is a rough hierarchy of priorities, but the balance between items is a matter of judgment.

- Building the **NSDL Collection** is a phased process with some areas given more attention initially than others. These areas will be middle school science, science for non-science majors, and mathematics.
- *Resources* from partner libraries and NSDL-funded project that have *collections* ready for ingest are a high priority.
- The metadata repository already contains records that describe many valuable *collections*, but without records of the individual *resources*. Many of these should be crawled to index content especially if they contain textual materials.
- Make agreements to have a number of proprietary information *resources* from groups such as society publishers or commercial publishers available for discovery in the NSDL. These groups would be required to provide metadata to the **NSDL Collection** by exposing their metadata using OAI with automatic checking.
- An increasingly wide range of materials have been placed on the Web that are within the NSDL scope and satisfy the criteria for OAI harvesting of metadata with automated checking. They provide an opportunity to increase the size of the **NSDL Collection**, especially for undergraduate education. All *resources* identified by this method will be assembled into one or more *collections* and evaluated using the **NSDL Collection** Selection Criteria. (See Section 6.3)
- CICDT should make a yearly recommendation to the NSF NSDL program regarding areas that need to be supported for *collection* development in order to increase the breadth and depth of the **NSDL Collection**.

10.0 TECHNICAL OPTIONS FOR *RESOURCE* INGEST INTO THE NSDL COLLECTION

10.1 INGEST OPTIONS

- Method A (OAI harvesting of metadata with automated checking) will be the standard method for ingest of *collections* that meet the requirements of metadata records for all *resources* and support of an OAI server.
- Method B (OAI static repository) will be an alternative to method B, for small *collections*.
- Method C (crawled *collections* of textual materials with some level of manual selection) will be the standard method for ingest of *collections* that satisfy key

criteria (open-access, crawlable, and primarily textual), but do not meet the criteria for method A or B.

Method A (OAI harvesting of metadata with automated checking) is under development and should be available in summer 2003. Method B(OAI static repository) depends on development of the associated OAI standard; a prototyping and demonstration study has been published (spring 2003). Method C (crawled *collections* of textual materials with some level of manual selection) depends on partnering and/or staffing arrangements that have yet to be determined.

Priority will be given to gearing up for method C; this requires both effort to automate the ingest process and modification of services, such as searching and browsing to reflect the differences in the knowledge of the *resources* as recorded in the metadata repository.

11.0 RECOMMENDED PRACTICES AND REFERENCES

11.1 RECOMMENDED PRACTICES

NSDL Metadata Primer (<http://metamanagement.comm.nsdlib.org/outline.html>)
Institute for Museum and Library Services. *A Framework of Guidance for Building Good Digital Collections*.
(<http://www.imls.gov/pubs/forumframework.htm>)

Open Archives Protocol for Metadata Harvesting (<http://www.openarchives.org>)

11.2 REFERENCES

APPENDIX A: GEM TOPICS CURRENTLY USED BY NSDL

A.1.0 – ARRANGED ALPHABETICALLY

Aging

☐Agriculture

☐Algebra

☐Applied mathematics

☐Arithmetic

☐Astronomy

☐Biological and life sciences

☐Biology

☐Body systems and senses

☐Botany

☐Calculus

☐Careers

☐Chemistry

☐Computer science

☐Discrete mathematics

☐Earth science

☐Ecology

☐Education (General)

☐Educational media

☐Educational technology

☐Embryology

☐Engineering

☐Entomology

☐Environmental health

☐Functions

☐General science

☐Geography

☐Geology

☐Geometry

☐Health

☐Histology

☐History of science

☐Human sexuality

☐Instructional issues

☐Integrating technology in the classroom

☐Mathematics

☐Measurement

☐Metallurgy

☐Meteorology

☐Multimedia education

☐Natural history

☐Number sense

☐Number theory

- ☐ Nutrition
- ☐ Oceanography
- ☐ Paleontology
- ☐ Patterns
- ☐ Pharmacology
- ☐ Physical sciences
- ☐ Physics
- ☐ Probability
- ☐ Science
- ☐ Space sciences
- ☐ Statistics
- ☐ Technology
- ☐ Technology planning
- ☐ Trigonometry

APPENDIX A: GEM TOPICS CURRENTLY USED BY NSDL

A.2.0 – ARRANGED HIERARCHICALLY

-Education (General)

Instructional issues

Careers

-Technology

Educational technology

Educational media

**Integrating technology in the
classroom**

Multimedia education

Technology planning

-Health

Aging

Body systems and senses

Environmental health

Human sexuality

Nutrition

-Mathematics

Algebra

Applied mathematics

Arithmetic

Calculus

Discrete mathematics

Functions

Geometry

Measurement

Number sense

Number theory

Patterns

Probability

Statistics

Trigonometry

-Science

Agriculture

Astronomy

Biological and life sciences

Biology

Botany

Chemistry

Computer science
Earth science
Ecology
Embryology
Engineering
Entomology
General science
Geology
Histology
History of science
Metallurgy
Meteorology
Natural history
Oceanography
Paleontology
Pharmacology
Physical sciences
Physics
Space sciences