This product is part of the RAND Corporation technical report series. Reports may include research findings on a specific topic that is limited in scope; present discussions of the methodology employed in research; provide literature reviews, survey instruments, modeling exercises, guidelines for practitioners and research professionals, and supporting documentation; or deliver preliminary findings. All RAND reports undergo rigorous peer review to ensure that they meet high standards for research quality and objectivity.

TECHNICAL REPORT

Steps Toward a Formative Evaluation of NSDL

Tora K. Bikson • Nidhi Kalra • Lionel A. Galway • Grace Agnew

Sponsored by the National Science Foundation

Transportation, Space, and Technology

A RAND INFRASTRUCTURE, SAFETY, AND ENVIRONMENT PROGRAM

The research in this report was sponsored by the National Science Foundation and was conducted under the auspices of the Transportation, Space, and Technology (TST) program within RAND Infrastructure, Safety, and Environment.

Published 2011 by the RAND Corporation
1776 Main Street, P.O. Box 2138, Santa Monica, CA 90407-2138
1200 South Hayes Street, Arlington, VA 22202-5050
4570 Fifth Avenue, Suite 600, Pittsburgh, PA 15213-2665
RAND URL: http://www.rand.org
To order RAND documents or to obtain additional information, contact
Distribution Services: Telephone: (310) 451-7002;
Fax: (310) 451-6915; Email: order@rand.org

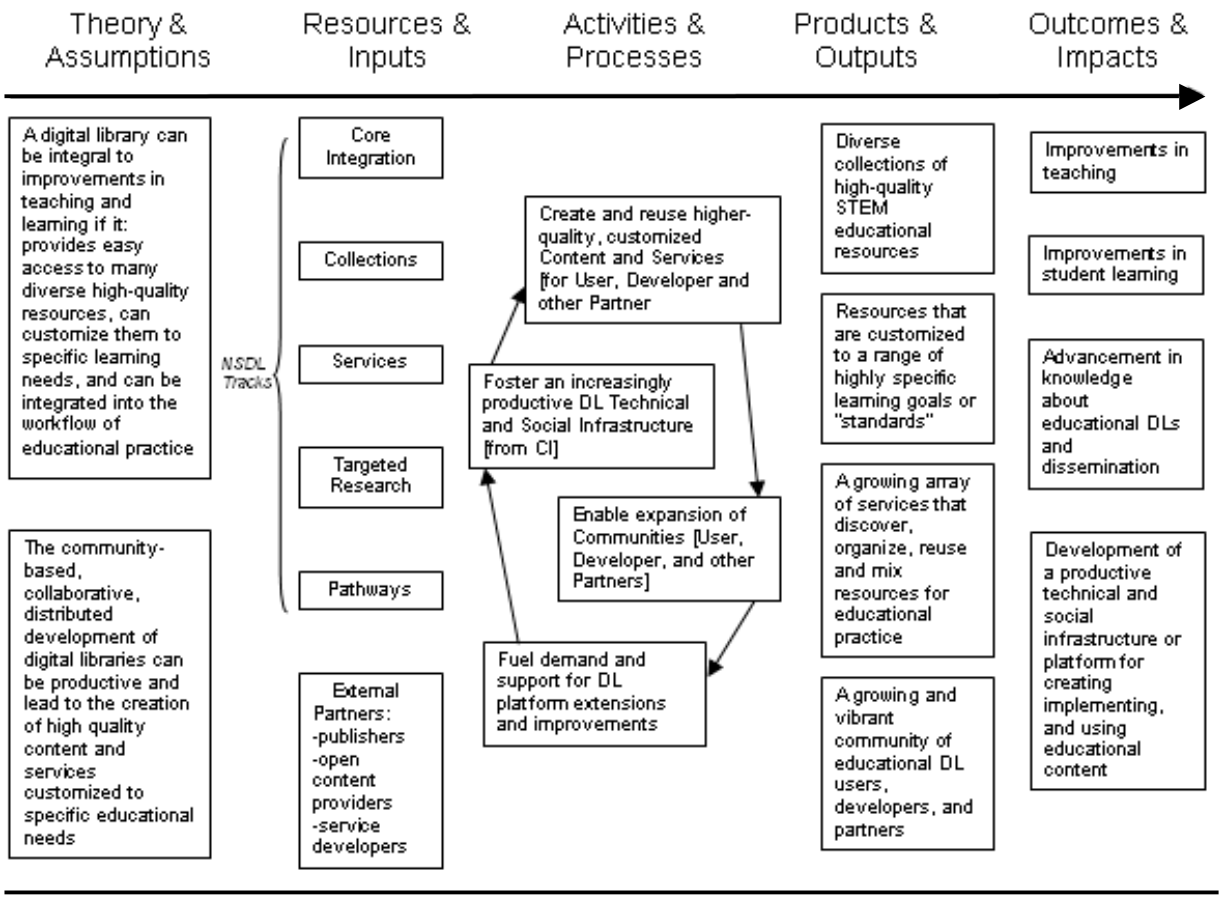
SUMMARY

The NSF's NSDL program turned ten years old in 2010. This technical report presents results of a preliminary program evaluation carried out by RAND in response to a request for proposals developed by NSF. These results represent the second phase of NSF's multistage evaluation plan. We organized our work around three principal evaluation goals:

- Provide an initial evaluation of the NSDL based on existing information resources, including results of a prior phase of evaluation.
- Conduct limited primary data collection studies to pilot test instruments and approaches and complement findings from extant information.
- Design a robust formative evaluation for a future phase of effort, refining research questions and methods.

BACKGROUND AND APPROACH

The NSDL program started in 2000 in NSF's Directorate for Education and Human Resources (EHR). Between then and 2010, it awarded more than 250 grants in several program tracks: Collections of digital contents (this track was discontinued in 2004); Pathways to assist in locating and using extant digital resources and tools (this track was initiated in 2004); Services; Targeted Research; and a Core Integration component, funded through a long-term cooperative agreement to establish and maintain an organizational and technical infrastructure to coordinate and support interactions among the projects in the other tracks. Early in Fiscal Year 2009, coinciding with the emergence of the program's emphasis on distributed learning, the Core Integration effort was divided into two separate projects, one a Resource Center and the other providing Technical Network Services.



SOURCE: McArthur, D., NSDL Evaluation: Phase 1, Washington, D.C.: National Science Foundation, 2008.
NOTE: DL = digital library. CI = Core Integration.

Figure S.1 NSDL Program Logic Model

Over this time period, the NSDL program logic model (see Figure S.1) changed little, except for the division of the Core Integration component (under resources & inputs) into the separate projects noted above. However, they share the same set of responsibilities as were formerly carried out by the Core Integration team. We therefore took that model as the starting point for our research.

RESEARCH PROCEDURES

Research procedures for this preliminary study were organized to follow its main evaluation goals. We began by addressing what can be learned from existing information in three key areas: the *digital resources developed*, the *avenues for accessing*

them, and the *expected outcomes for their use*. Specifically, we addressed these main research questions in response to the first aim:

- What is the state of health of the resource collections?
- What evidence do current web metrics provide about activities undertaken by users to find and access those resources?
- Are there early signs of positive effects from use of NSDL resources among science teachers or learners?

For this purpose, we assessed all current collections, examined the websites of the NSDL portal and several pathways, as well as support for moving between and among them, and reviewed project abstracts, reports, and research papers.

When we had initial answers to these questions from extant information, we then designed and carried out some primary data collection efforts to supplement the findings and to provide a better base for a future evaluation. In particular, we conducted pilot trials aimed at *assessing the metadata* intended to assist users in finding, identifying, selecting, and obtaining sought resources and at *exploring the usability* of available resources for accomplishing selected routine tasks in science teaching. Finally, we conducted a *minicase study of one full instantiation of the NSDL model*—a demonstration of curriculum customization to meet the needs of earth science teachers using resources from the Digital Library for Earth System Education (DLESE).

We then synthesized the findings from these exercises to yield results and recommendations, both for what NSF and the NSDL should do now, as well as for the design of a full formative evaluation. In the next section, we summarize key findings and recommendations and suggest a design for the anticipated future evaluation of the NSDL.

FINDINGS AND RECOMMENDATIONS

State of Health of the Resource Collections

The project relied on existing data derivable from NSDL sites to evaluate the authoritativeness, usability, and integration of the resources they make available. For this purpose, we cataloged and analyzed measurable characteristics representing attributes of interest for assessing their state of health. The three-tiered sample included 50 still-viable collections (of 71 funded by NSF), the 14 functioning pathways, and

nsdl.org as a one-stop shop for finding NSDL resources. Here, we highlight key findings and recommendations for future improvements.

Authoritativeness. Currency of the collections is questionable. Of the 50 collections, 14 of them (over a quarter) are not current, yet 13 of these are still available for search and retrieval via the NSDL. However, stability of the collections themselves is noteworthy, with 95 percent or more of records sampled from collections and pathways being still available. Evidence of scientific authoritativeness of the resources is weaker than would be desired. Just over half of the Pathways and just under half of all collections have published peer review practices. Further, only 40 percent of collections have evidence of professional oversight, compared with 86 percent of Pathways.

Metadata Usability. Metadata are generally problematic across resources—just over a third of pathways and just over a tenth of collections have good metadata. Sample searches often turn up heterogeneous resources (e.g., activities, articles, historic images) in a single display because results are subject matter–based and not use-based. When resources are found, metadata do not provide users with clear information about what can be done with them. The majority of pathways and collections provide some kind of copyright guidance, but it is inconsistent and usually hard to interpret. Privacy policies are available in the majority of pathways but in less than half of the collections.

Integration. Nsdl.org is not an integrated one-stop shop—it does not appear to be substantially greater than the sum of its parts. Further, there is no real evidence of pathway collaboration across the NSDL—no common elements of look, feel, and service characterize them, while the NSDL is mainly acknowledged as a funding source. Moving from pathway to pathway creates a sense of dislocation and of starting over.

Recommendations to NSF and the NSDL. Noncurrent resource collections should be identified and removed. For viable collections, it is important to promote regular checking to find and fix broken links. Additionally, policies and practices that evidence authoritativeness (peer reviews, professional oversight) should be encouraged and promulgated. With respect to usability, we recommend that the NSDL resource team and pathways researchers work with educator audiences to develop common metadata elements and vocabularies for core filtering elements that would address the workflow needs of STEM teachers.

NSF and the NSDL, in consultation with pathways leaders, should consider focusing nsdl.org on a social networking service model. For instance, the NSDL resource team could provide centralized statistics and surveys; users might “join” pathways and

manage their memberships; centralized folders and workspaces might be offered for each user; and users should have ways to share the resources they have created or customized, as well as to learn which resources are most frequently accessed by other users like themselves. More generally, multiple avenues should be pursued to realize the aim of having the NSDL be more than just the sum of its parts.

Recommendations for a Future Phase of Evaluation. In our judgment, evaluation of the state of health of the resource collections is largely complete on the basis of work done for the second phase effort. For a third phase of evaluation, we recommend two types of initiatives.

First, it would be desirable to revisit the pathways and collections periodically (e.g., at six-month intervals) to assess the extent to which selected noteworthy shortfalls are being remedied. On a more ambitious note, metadata evaluation should be undertaken to identify and address educators' needs for more context-friendly and useful core filtering elements. For this purpose, we would recommend an extension of the metadata pilot evaluation reported here. Extensions could include cognitive walk-throughs with representative users and experimental trials with added "educator schema" filtering elements.

Evaluative Evidence from Web Metrics

A second evaluation effort turned on exploring web metrics to see what evidence they might provide for addressing questions about the composition and growth of user communities, users becoming reusers and resource contributors, "cross-pollination" among pathways, and other aims envisioned in the NSDL program's logic model. For this purpose, we interviewed researchers responsible for nsdl.org web metrics. Additionally, we interviewed web metric staff at five pathways selected for their reputation using web metrics. We supplemented these interviews with a review of web metrics literature. Main findings and recommendations follow.

Nsdl.org and Web Metrics. The primary uses of web metrics at the top-level site are for overall traffic monitoring and for site modification and redesign. It is not presently feasible for nsdl.org to develop a clear understanding of its users because it lacks effective ways of getting data about them. The only avenue for learning more about its user community appears to be responses to brief pop-up online surveys, of unknown quality and representativeness. Presently, there is also not an ability to classify and characterize users by general site usage patterns, although this is a desired future goal for web metrics. Additionally lacking are comprehensive data about the pattern of

traffic into and out of nsdl.org. Thus, nsdl.org now has little knowledge about the demographic characteristics of its user community, nor a general understanding of its usage patterns.

Pathways and Web Metrics. Like nsdl.org, pathways chiefly use web metrics for monitoring overall traffic and for site redesign purposes. Like nsdl.org, all but one rely on pop-up surveys. Similarly, pathways undertake only very limited assessments of users' site usage patterns. Additionally, they do not know whether specific other pathways, or nsdl.org, or still other sites are the main sources or the main destinations of their users.

There is little communication about web metrics between pathways and nsdl.org, and no mandated reporting of web statistics either to the Resource Center or to Technical Network Services (TNS). There had been some attempt at standardization of web data using Omniture, but, presently, some pathways have moved to Google Analytics because it is available at no cost. In any case, the net result is that no summaries are available from pathways related to overall traffic flow, sources of users, and the like to help inform the present evaluation.

Recommendations to NSF and the NSDL. Standardized, basic web metrics should be selected and shared by nsdl.org and the pathways. Additionally, there should be mandated reporting of some web metrics (e.g., overall traffic flow, immediate sources and destinations of users) at designated intervals. Further, NSF and the NSDL should plan for future usage data collection when new technologies (e.g., social media) are likely to play a greater role in nsdl.org and the pathways. Some current web metrics might not be applicable to these technologies, although the new media are likely to become important in the growth and vitality of user communities.

Recommendations for a Future Evaluation. Recommendations for a future evaluation are dependent, in part, on the ability to obtain standardized web metrics from nsdl.org and the pathways as recommended earlier. To examine whether emergent user communities are exhibiting vitality, growth, and connectivity, as anticipated by the program's logic model, it would be desirable at minimum to measure basic traffic volumes over time, plus the sources and destinations of site visitors. These data would be useful for constructing a map of traffic relationships between nsdl.org, the pathways, and other key sites.

It is also important to measure the extent of use of private spaces by users to retain and tailor materials to their own needs and the rate or volume of contributions by users to shared spaces, for sites that offer these services. Such measures would shed light on

whether customized or customizable resources were being developed to meet particular teaching or learning goals, as the logic model predicts.

A last recommendation would be to pursue user classification by site usage patterns to provide a richer picture of users. Specifically, in given sites, it might be possible to carry out experiments where user computers—with users' consent—are instrumented for more detailed data collection and analysis.

Signs of Teaching or Learning Effects

The prior evaluation phase sought early signs that use of NSDL resources might improve teaching or learning in the STEM disciplines. We started by examining abstracts for the 250+ NSDL-funded projects and then reviewing, in detail, project reports and research papers from those that directly involved teachers or learners. Such a case survey was expected to yield a taxonomy of project features, as well as emerging themes or promising trends related to teaching or learning. Additionally, the review could be suggestive of hypotheses for further testing in a future evaluation while identifying exemplary cases for site visits. Chief findings and recommendations follow.

Body of Teaching or Learning Research. Among all the projects funded by the NSDL during its decade-long lifetime, at most about a dozen touch in some way on teaching or learning. Each of the studies addresses a different tool or resource, and each uses a different context and different procedures. These great variations in a very small body of work preclude categorizing project features or approaches.

Additionally, the methods typically employed by the projects are not ones that support inferences about teaching or learning effects. Experimental or quasi-experimental methods are largely absent. Assessments rely in the main on teachers' self-reported behavior changes or intentions to change. Consequently, a systematic case survey was not feasible for the second phase of evaluation. However, examination of this body of work surfaced some research sites that might be good candidates for case study. It also contributed to the design of a pilot usability study with a small sample of current and prospective STEM teachers.

Recommendations to NSF and the NSDL. The NSDL program should be strengthened by allocating more funding for targeted research to explore and evaluate teaching and learning effects associated with the use of NSDL resources. Teaching and learning support have been key aspects of the program's logic model from the start, even at the "products & outputs" stage, as well as part of the expected longer-term outcomes and broader impacts. In particular, it is critical to direct more research toward

understanding and facilitating the transition from NSDL "activities & processes" to education-specific "products & outputs."

Recommendations for a Future Evaluation. A future phase of evaluation should routinely examine all forthcoming project reports (those produced since the review done for this report) to look for indicators of teaching or learning effects. However, because of the length of time that would likely be needed to elicit such effects, we would not expect to find major new contributions to the existing body of work. Consequently, we recommend continuing in two evaluative directions related to teaching and learning effects; both approaches have been piloted in the present study.

First, we recommend extending and replicating a resource usability study with a sample of current or future science educators large enough to yield statistically significant results. The study should investigate the relative advantages of orienting pathways by grade level versus subject matter or remaining generic. Results of such an experiment should valuably complement the suggested metadata evaluation and help inform efforts to shape and enhance the NSDL to better meet the needs of STEM teachers and learners.

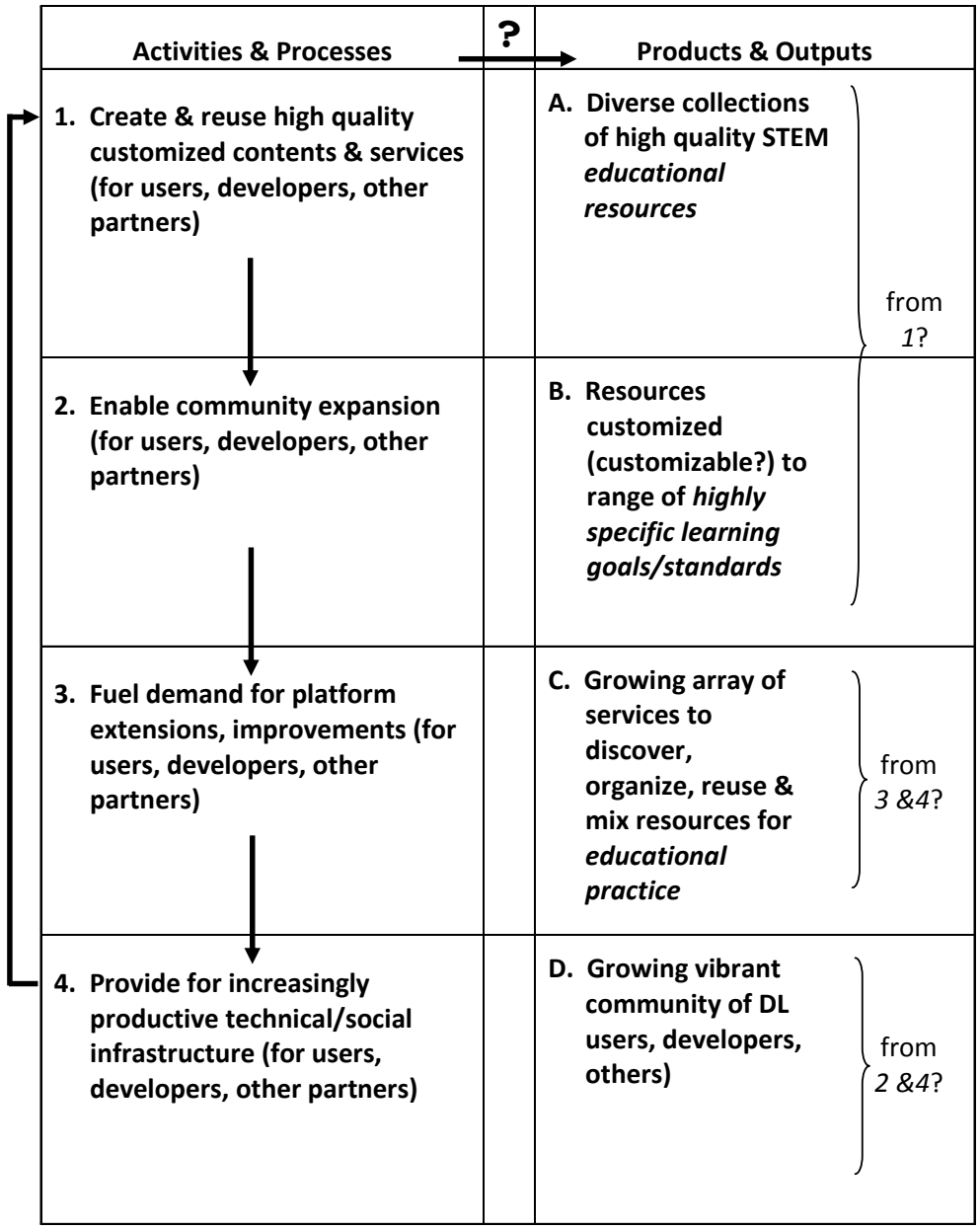
A second direction involves full case studies of projects that are currently putting NSDL resources into teachers' hands for ongoing classroom use. One such case is the Denver Public Schools Curriculum Customization Service project, which was the focus of the minicase study reported here. Another attempt to instantiate the NSDL program model is represented by the Robert Noyce Scholarship project, which takes a quite different approach to attempting to introduce digital resources into STEM teaching and learning. A future evaluation should document and compare alternative avenues for bridging the gap between digital libraries and distributed learning.

Suggested Overall Design for a Future Evaluation

Earlier, we identified some specific research recommendations for inclusion in a further, more complete formative evaluation of the NSDL program. Here, we suggest an overall design for the third-phase program evaluation.

Conceptual Framework for a Full Formative Evaluation. We suggest beginning again with a focus on the intermediate components of the logic model—activities & processes, and products & outputs—for two principal reasons. First, the present evaluation makes clear that it is still too early to look for longer-term summative outcomes and impacts. NSDL resources are only just beginning to find their way into contexts of use, so significant effects of the NSDL on STEM teaching and learning in U.S.

education still lie in the future. Second, it is appropriate to look more closely at whether and how the activities & processes specified in the logic model are leading to envisioned products & outputs, especially as they are expected to facilitate the transition from digital library resources to distributed learning uses (see Figure S.2).



NOTE: Emphases are ours.

Figure S.2 Proposed Logic Model Components for Future Evaluation

In Figure S.2, arrows among the four types of activities & processes called out in the logic model indicate that they are supposed to be mutually influential; they are intended to give rise to a virtuous circle in which valued NSDL contents and services engender growing use, which generates demand for improvements, which, in turn, leads to an increasingly productive social and technical infrastructure in which initial resources are reused, customized, and supplemented. We have added hypothetical links, with question marks, between the two intermediate components of the logic model. The initial account of the logic model assumes, but does not explain how, the activities & processes set out in the left half of the figure will result in the four types of products & outputs on the right—especially the products & outputs that will constitute high quality educational resources (A) customized to a range of highly specific learning goals (B) and will yield a growing array of services for finding and organizing these resources to meet educational practice needs (C). A future evaluation, although still formative in nature, should investigate, assess, and, ideally, make recommendations for strengthening transitions between activities & processes and the educationally relevant products & outputs set out in the logic model.

Sociotechnical Systems Research Foundation. A sociotechnical systems-theoretic foundation guided our approach to the current evaluation and should continue to provide a research base for a future evaluation phase. The NSDL is envisioned as a complex, interconnected sociotechnical community (see, for example, the fourth activities & processes box in Figure S.2).

From the standpoint of sociotechnical systems theory as it applies to understanding the adoption of new technologies, successful outcomes can be viewed as a function of properties of the technologies, the contexts of use, and the implementation processes by which the former are integrated into the latter. Implementation is construed as the series of decisions made and actions taken once an organization makes the go-ahead decision to adopt a new technology. Sociotechnical systems theory sees implementation processes as aiming at the joint optimization of the capabilities of the social and technical constituents of the system. In this process, the technologies are modified to better fit the aims and capacities of the user context; concurrently, users' behaviors and environments will need to change (via, e.g., training, revisions of practice) to take advantage of the affordances of the new technologies. The guiding implementation assumption is that, like the terms of a simultaneous equation, social and technical challenges cannot be separately resolved. Implementation process characteristics are expected to be the strongest predictors of successful technological

innovation in social contexts (e.g., see the review in Bikson and Eveland, 1991). This theoretic lens guided the minicase study reported in Chapter Six.

A future evaluation should look for evidence that NSDL resources can be or are being customized for use in educational contexts (see Figure S.2, elements B and C under "Products & Outputs"). User involvement (or *demand-side input*) is a strong factor in optimizing a technology to suit the needs of a particular adopting context.

"Nested" Design and Methods

We suggest a four-tiered evaluation design, but we have put "nested" in quotes because, although the envisioned levels are interdependent in varied ways, the proposed design is not truly nested. What follows is a brief account of each level of the hierarchical sociotechnical system design approach (see Figure S.3).

NSDL: the program and the site (N=1)

PATHWAYS: exhaustive sample (N=14-17), plus selected tools and services

COLLECTIONS: extensive sample of collections hosted by pathways, plus a sample of resources made accessible via pathways or nsdl.org

USERS: survey of teachers who have had NSDL-related professional development; full-case studies of sites where teachers are known to be using NSDL resources (e.g., in classrooms); experimental trials with prospective users

Figure S.3 Proposed Hierarchical Approach for Future Evaluation Research

The top level is the NSDL itself, including both the program and the nsdl.org site, the top social and technical dimensions of the sociotechnical system. At this level, there is an N = 1, susceptible only to a case study approach. Interviews with key stakeholders within and outside NSF and the NSDL would be the main method of choice, supplemented by a review of current integration efforts and an examination of web metrics focused on overall volumes, as well as source and destination traffic patterns.

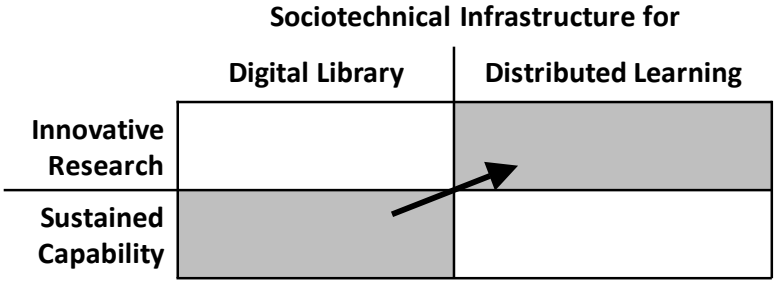
Next is the level of pathways, for which an exhaustive sample is appropriate, given their small number. Here, we suggest semistructured interviews with pathway project leaders and (if different) those in charge of web metrics. Additionally, we propose an examination of web metrics, including volumes and source/destination patterns. We also suggest an examination of the level of use of personal spaces, where

offered. At both this and the top levels, it should be possible to learn whether there are growing and contributing user communities.

Resource collections constitute the third level, where current NSF-funded collections are expected to number about 50. The evaluation of collections is already largely complete. The future evaluation effort should therefore be limited to follow-up checks to determine the extent to which policy recommendations from the this preliminary evaluation have been or are being implemented.

The lowest level, comprising users, merits the greatest attention in a third phase of evaluation. Here, the methods should attempt to elicit evaluative information about the transition to educational uses of NSDL resources for teaching and learning. For this purpose, we recommend a survey of teachers who have completed courses on how to use the available resources for STEM teaching. Additionally, for the user level, we would suggest building on the trials piloted and discussed in this report. In particular, a more in-depth metadata assessment and an expanded resource usability study would help determine whether high-quality STEM educational resources could be discovered, organized, used, mixed, and reused to meet educators' needs. Further, it would be desirable to carry out full-size case studies of projects that are making concentrated efforts to put NSDL-based STEM resources into teachers' hands for classroom use. For this purpose, the Denver Public Schools Curriculum Customization Service effort and the Noyce scholars program would merit study, along with any other such projects that emerge in the coming year.

Evaluation Focus. As an evolving, complex, distributed sociotechnical system, the NSDL program has multiple "personas." It is an interconnected library of digital resources that should provide a platform for enabling advances in STEM teaching and learning. Like other NSF programs, it is charged with stimulating innovative research, but it must also be concerned about the sustainability of the systems that can support such an endeavor. Figure S.4 displays these varied personas in a matrix.



NOTE: The shaded areas represent the evaluation domains we suggest for future evaluation emphasis.

Figure S.4 Evaluation Foci: Alternative Options

Each quadrant of the matrix represents a plausible domain within which formative evaluation results might be sought. For instance, when the *L* in *NSDL* means *library* (as in previous proposal solicitations and implicit in the activities & processes component of the logic model), evaluation might focus on the extent to which the program yields cutting-edge methods and tools for understanding and advancing library science for STEM resources in the ever-changing digital era. Or evaluation might focus on how well efforts are working to sustain and improve the resources the distributed library makes available for future research and development, much like a distributed "facility"—comparable, say, to geoscience facilities funded by NSF.

On the other hand, when *L* primarily means *learning*, as seems implicit in the products & outputs component of the logic model, as well as in the changed title of the 2010 NSDL call for proposals, the target of evaluation might shift to education-related results. If so, a future evaluation might focus on advances in research on how and why digital resources can markedly improve teaching and learning in the STEM disciplines. Alternatively, it could emphasize how sustainable those instructional improvements are after the initial period of research funding ends.

These options are not mutually exclusive in principle. In practice, they might not all be feasible to include in a future formative evaluation design, given limited resources. The shaded areas in Figure S.4 represent the evaluation domains we suggest for emphasis in a future evaluation phase. The Innovative research in digital library science (top left quadrant, unshaded) could reasonably be supported by other funding programs within and outside NSF. We would encourage NSF to make available and publicize its NSDL resources for research purposes beyond the NSDL program itself

while not including this domain as a focus of the proposed next phase of evaluation. However, both to support digital library science research and to enable advances in research on educational uses of digital STEM resources, it is critical to sustain the developed contents and services. We therefore recommend that a future formative evaluation examine progress toward sustaining the distributed digital library infrastructure (bottom left quadrant, shaded).

On the other hand, the study reported here indicates that efforts to put digital STEM education resources into classroom contexts have only just begun and are too few in number to justify an evaluation of their sustainability (in Figure S.4, bottom right quadrant, unshaded); rather, that should be the aim of a future summative evaluation. In the meantime, a future formative evaluation should pursue and examine factors that affect or could affect the adoption, implementation, use, and usability of the developed digital library infrastructure to meet the needs of STEM educators (top right quadrant, shaded). In these ways, the future evaluation efforts can inform and assist the anticipated transition from digital libraries to distributed learning. The NSDL should move aggressively to orient its program resources toward understanding how STEM education can be improved given a supportive high-quality digital teaching and learning infrastructure.