Pre-publication draft only. Final publication may be cited as:

An Agency Perspective on Digital Library Evaluation

Michael Khoo
The iSchool at Drexel University
3141 Chestnut Street
Philadelphia, PA 19104 USA
michael.khoo@ischool.drexel.edu

David McArthur
GoH LLC
4201 Wilson Blvd.
Arlington, VA 22230 USA
dmcarthur1@nc.rr.com

Lee Zia
National Science Foundation
4201 Wilson Blvd.
Arlington, VA 22230 USA
lzia@nsf.gov
An Agency Perspective on Digital Library Evaluation

INTRODUCTION: FUNDING AGENCIES, DIGITAL LIBRARIES, AND EVALUATION

Funding agencies play key roles in digital library projects. They provide financial support, guidelines for research and development, and requirements for reporting project outcomes. In the case of publicly supported agencies, governments in turn require such bodies to account for their own activities and expenditures. A common concern of agencies is therefore to ensure that they are obtaining good value from the projects that they fund.

Different agencies have different evaluation models and strategies. Key evaluation requirements are usually published in program solicitations, where agencies may provide specific evaluation models and advice. In answering a solicitation, a project agrees to abide by the evaluation requirements of that solicitation. This ‘contractual model’ – in which an agency provides funds to a project, and that project then reports on results to the agency – constitutes a simple model of evaluation requirements, as represented in Figure 1a.

[INSERT FIGURES 1A AND 1B ABOUT HERE]

In practice this simple relationship is often complicated by local conditions. To illustrate this complexity, we provide a case study of a United States funding agency, the National Science Foundation (NSF), and of one of the programs that it funds, the National Science, Technology, Engineering, and Mathematics Education Digital Library (NSDL), focusing on the example of web metrics.

The case study is in three parts:

(a) The national context of digital library development in the United States;
Chapter x: Chapter Title

(b) NSF organization and evaluation; and

(c) NSDL organization and evaluation, including a web metrics case study.

Citing a range of reports from NSF and the NSDL, we will show how NSDL evaluation is shaped by a range of national, program, and project factors, and how NSF and NSDL are linked by multiple organizational and evaluation channels, some of which arise from the position of the program in the NSF agency structure, and others, often less clear, which stem from the collective status of the various NSDL projects as a virtual organization that is situated within the wider NSDL program. We argue that these multiple channels can be beneficial, because they permit multiple views on the same organization, which can be triangulated with both internal project data and wider digital library research data. But we also argue that multiple channels can be challenging to manage, particularly in widely distributed organizations. Some agency evaluation requirements may be clear and compelling to projects, while others may be vague and not backed by compliance mechanisms; some requirements can be complementary, while others may be inconsistent; and balancing multiple requirements may be difficult for projects on limited resources.

We conclude with some recommendations for organizational sense-making activities that programs and individual projects can engage in, in order to develop successful evaluation plans. They emphasize the importance of careful planning of project-level evaluations, fully supported by the necessary resources; and also the more deliberate organizational articulation of the important connections between project and program evaluation, taking into consideration the wider context of multiple, potentially conflicting, and changing evaluation requirements and channels.

THE UNITED STATES CONTEXT

Digital library development in the United States takes place within a range of national policy contexts, including the ongoing strategic importance of science education, and the creation of
a national cyberinfrastructure to support science and engineering research and education (e.g. Atkins, 2003).

**Science Education in the United States**

Various reports (e.g. “Committee,” 2007) have noted the importance of science, technology, engineering and mathematics (STEM) education skills in a 21st century global economy, and the stagnation of these skills in the United States workforce. In response, the United States federal government, universities, educational foundations, and others, are applying innovations in computer and information technologies to the design of educational technologies such as digital libraries, in order to increase STEM educational and workforce capacity. Despite the investment of billions of dollars in educational technologies, there is ongoing debate in the United States as to their effectiveness. Metastudies of the use of a range of technologies in schools (e.g. Waxman et. al. 2003) have also shown improved educational outcomes in a number of areas. On the other hand, a recent report to the United States Congress (Dynarski et al. 2007) on the effectiveness of reading and mathematics software found that test scores were not significantly higher in classrooms using selected reading and mathematics software products.

But this debate is not just about improvements in standardized test results. Constructivist educational theorists argue that the significance of educational technologies lies not just in their ability to promote test scores – which may in the end may just measure rote learning skills – but also in their promise to promote greater of depth of critical thinking, increase student engagement and retention, and a range of other educational outcomes (Honey, 2001; Viadero, 2007). These more complex and qualitative learning outcomes are as significant in debates over the effectiveness of educational technologies as are test scores.

**Cyberinfrastructure**
A second US policy context is that of the creation of a national cyberinfrastructure supporting STEM research and education. Stimulated by the blue-ribbon report on revolutionizing science and engineering (Atkins, 2003), over the past several years NSF has begun to develop programs that to explore the potential of cyberinfrastructure, and this research has slowly extended from the physical sciences to the social sciences (Berman, 2005) and to education (Bernat et al., 2006). These recent reports have also suggested that cyberinfrastructure applications might transform science education as dramatically as they are now changing practice in the sciences.

The increasing emphasis on cyberinfrastructure for learning, now often referred to at NSF as “cyberlearning”, has two broad consequences for the development and evaluation of digital learning technologies in general and digital libraries in particular. First, it places a premium on the design of innovative teaching and learning environments that extend the boundaries of what students can learn, explore the new roles teachers can play, and invigorate the processes through which learning happens. Early work in these programs is as much about breaking existing moulds in education, as it is about using cyberlearning to make minor enhancements in traditional pedagogies aimed at incremental improvements on standardized tests. Evaluation of cyberlearning programs and projects, in turn, will need to reflect these priorities. Such evaluations may be very different than ones, such as those noted above, that examine the near-term effectiveness of educational technologies in helping to achieve well-defined student learning outcomes. Early cyberlearning evaluation goals are likely to emphasize the formative assessment of risky yet highly innovative educational tools and learning environments.

The second consequence of cyberlearning for digital library evaluation relates to “virtual organizations”: geographically distributed teams whose members share a common interest or goal, and work primarily or exclusively at a distance using cyberinfrastructure technologies.
Early work in virtual organizations at NSF is examining how teams come together – and split apart – to work more effectively, and how cyberinfrastructure can support such dynamic organization synthesis and the advance work of virtual teams. In terms of evaluation, fluid virtual organizations imply that teams working together may be connected through multiple management channels, each of which may have its own evaluation priorities. Because virtual organizations may reconfigure “on the fly”, team members may not be aware of all the evaluation requirements or expectations that are imposed on them at a given time. All this is potentially far more complicated than the simple organizational connections depicted in Figure 1a.

THE NATIONAL SCIENCE FOUNDATION

Having described the cyberinfrastructure and virtual organization contexts within which NSF is set, we now turn to the organizational structure of NSF, and the NSDL program within this organization.

NSF is an independent federal agency created by Congress in 1950 “to promote the progress of science; to advance the national health, prosperity, and welfare; to secure the national defense” (NSF, 2006). NSF has an annual budget of about $6 billion, and funds basic research in many, including mathematics, computer science and the social sciences. The agency is composed of a hierarchy of directorates and offices, divisions, and programs within each division responsible for funding individual projects. For instance, the NSF Directorate for Education and Human Resources (EHR) administers the Division of Undergraduate Education (DUE); DUE in turn funds the NSDL program; and the NSDL program funds individual NSDL projects (Figure 1b). Each NSDL project subscribes to the research and development aims expressed in the program-level solicitations.

---

1 NSDL encompasses learners of all ages; the program is in DUE for historical reasons.
Each organizational layer at NSF embraces and focuses the mission of the organizational level above it. NSF’s mission as a whole is to promote US science capacity in all areas. EHR narrows that focus to formal and informal education at all levels; DUE works specifically with undergraduate education; and finally, the NSDL program implements DUE’s mission in the specific context of digital libraries. Working back up the organizational hierarchy, we can see that NSDL projects report to the NSDL program, which reports to DUE, which reports to EHR, which reports to the NSF Director.

**Program- and project-level evaluation in NSDL**

Program-level NSDL evaluation is approached as a holistic and integrative process that focuses on understanding the collective utility of the program as a whole. Project-level evaluation focuses on the individual projects within the program, to determine whether or not they are proceeding according to the schedules provided in the original project proposals, and whether they are meeting their stated goals (c.f. Frechtling, 2002). These two activities may appear to be mismatched, but they are actually related in complex ways: for instance, project-level evaluation data are fed back to NSF, where they may also offer the potential to be repurposed and used to answer different questions at the program level than at the project level. In NSF’s view, projects do not just succeed or fail. Rather, since many NSF projects are experimental in nature, there is an interest in data as a description of what happened, as much as an overall judgment of success or failure. In highly experimental conditions, all data are potentially useful.

This wider perspective may not necessarily be shared by individual NSDL projects, where perceptions of success and failure may be more keenly felt, not least because the grant funding process is ultimately a competitive one, in which future success with an agency is often seen, rightly or wrongly, as being partly related to current success. To avoid this perception, program evaluation requirements should be explained carefully to individual
projects, and stress “the inherent relationship between evaluation and program implementation. Evaluation is not separate from, or added to, a project, but rather is part of it from the beginning. Planning, evaluation and implementation are all parts of a whole, and they work best when they work together” (Frechtling, p. 3).

THE NATIONAL SCIENCE DIGITAL LIBRARY (NSDL)
The NSDL is a multi-year NSF program that is developing a wide range of STEM educational resources, tools and services, within a distributed digital library structure. Its origins lie in discussions in the academic, pedagogical, and computing science communities in the 1990s on the building of an undergraduate digital library for science, mathematics, engineering and technology education (“NSF SMETE-Lib Study,” 1998; Wattenberg, 1998). NSF subsequently developed plans for a National Science, Technology, Engineering, and Mathematics Education Digital Library Program (e.g. Zia, 2000). Officially launched in 2000, by 2006 the NSDL program had disbursed over $150,000,000 to over 200 projects. The mission of the library “to increase literacy and interest in science, mathematics, and technology; ready tomorrow’s workforce for STEM-related careers; support the advance of knowledge and the solving of real-world problems; and promote continued excellence in the nation’s scientific pursuits” (NSDL, 2007), resonates both with NSF’s overall mission, and also with the United States’ current strategic concerns with increasing STEM capacity.

Over time the NSDL program has moved away from funding a larger number of specific individual digital library components (tools, services, collections, etc.), and towards the support of a smaller number of integrated portals, known as ‘Pathways’ projects. Pathways projects are

- grouped by grade level, discipline, resource or data type … [to] best support efficient resource discovery for broad categories of users. Pathway portals are developed and managed in partnership with organizations and institutions that have a history and
expertise in serving their portal's target audiences – in effect, they act as reference librarians for their communities. (NSDL n.d.a.; “What are NSDL Pathways”) NSDL Pathway projects develop their own websites and catalogs. In addition, Pathway catalogs are federated at the main NSDL website at nsdl.org. Users of NSDL therefore can either go straight to the relevant Pathway portal, and also search the Pathways portals from the central site at nsdl.org.

**NSDL Program Coordination**

A distinctive characteristic of the NSDL program is that individual NSDL projects are expected to collaborate amongst themselves in areas of program development such as technological standards, information and library standards, program-wide communication tools, etc. As such, NSDL projects, when taken together, constitute a virtual organization or entity, ‘NSDL,’ which is distinct from the NSDL program that is administered by NSF (represented by the dotted line in Figure 1b). This virtual structure arises out of wider NSF research concerns with the management of large-scale, distributed virtual organizations. Project level ‘virtual organization’ coordination activities in NSDL are concentrated in two areas, the NSDL ‘Core Integration’ project, and the NSDL community (Figure 2).

**NSDL Core Integration Projects**

NSF funds a single NSDL ‘Core Integration’ project (‘NSDL-CI’) tasked with developing program-wide technical, logistical and managerial infrastructure. For instance, NSDL-CI maintains a portal for the NSDL at nsdl.org, and is building a federated repository, which together provide access to and search across all Pathways’ collections. The NSDL-CI project
has broader responsibilities, and deeper resources (including some full-time staff), than other NSDL projects.

The NSDL-CI project has developed a quasi-official managerial relationship with Pathways projects that includes some of the oversight roles more traditionally associated with NSF program officers. The relationship is described as ‘quasi-official’ because while it replicates some of the typical patron-client relationship between NSF programs and individual NSF projects, it is also viewed by NSF as an experimental form of infrastructure for a large-scale virtual organization, and is therefore open to interpretation. The relationship is described in a number of formal and informal agreements, including the ‘Memorandum of Understanding’ (MoU: ref), which sets out the NSDL-CI requirements for Pathways projects.

The NSDL Community

Virtual management of a variety of day-to-day NSDL affairs is further coordinated by a ‘community’ whose members are drawn from NSDL projects, and who participate in committees, email lists, conference calls, and an NSDL Annual Meeting. NSDL committees work to develop common NSDL policies in areas such as collection development, governance, sustainability, technical standards, and evaluation (via the Educational Impact and Evaluation Standing Committee, or EIESC (NSDL, n.d.a)).

NSDL evaluation practices

NSDL committees and the Core Integration project constitute a ‘middle layer’ between NSF and individual NSDL projects (Figure 2). In this complex organizational space, evaluation involves more than the simple “channel” model illustrated in Figure 1a, or even the hierarchical model of Figure 1b. In reality, there are a number of evaluation ‘channels’ between NSF and NSDL that are of varying degrees of formality.

The most formal channel (arrow 1 in the Figure) is between NSF and individual Pathways projects. It is laid out in NSDL program solicitations, which ask potential projects:
What evidence will be sought to inform the progress towards project goals and why is this of value? As a component of NSDL, how will usage of the services offered by the project be ascertained? What evidence of impact on users will be gathered and why? (NSF, 2007b)

Successful project applicants for NSDL program funding have to provide a detailed account of how they would address these questions and make the data available to NSF.

A second evaluation channel is that between NSF and individual NSDL projects, mediated by NSDL-Core Integration, and described by the MoU (see above) (arrow 2 in Figure). Pathways projects have to sign the MoU before NSF will release funds to them. In return for receiving funds, the MoU requires Pathways to participate in the larger NSDL program evaluation effort by defining yearly benchmarks and metrics and by providing web metrics data to NSDL-CI. However, as was noted above, NSDL-CI has only a quasi-official status between NSF and NSDL projects, and there are therefore no real sanctions if a Pathways project does not fulfil its evaluation requirements.

The third evaluation channel depicted in Figure 2 is that between NSF and NSDL-CI. This channel reports the results of NSDL-CI’s evaluation on its own activities. These activities include the development of evaluation capacity within the NSDL program as a whole, and so this channel also reports Pathways evaluation data that have been collected and analysed by NSDL-CI. This relationship is a formal one – NSDL-CI’s responsibilities are defined by NSDL solicitations.

Finally, there is a network of informal evaluation channels in the form of the work done by the Educational Impact and Evaluation Standing Committee (EIESC; see above). This work includes the coordination and collection of cross-project data of interest to NSF, such as organizational structure, site usage, and collection development policies, and so on (Giersch, 2005). EIESC results are presented in EIESC workshops and reports, and reported to both
NSDL-CI and NSF. A significant channel here is the NSDL Annual Meeting (AM), where projects can report evaluation findings; and satisfaction surveys indicate that the AM is very favourably rated as a venue for face-to-face interaction with colleagues in other projects. All this activity is quasi-official.

Multiple evaluation channels in NSDL

The existence of plural official and quasi-official evaluation channels in a single program context, as represented in Figure 2, is a novel experience for many NSDL projects. Usually, an NSF project has a one-on-one relationship with NSF. In the case of the NSDL, however, NSDL-CI, the MoU with the Pathways projects, and the NSDL community, constitute an additional web of organizational layers between individual projects and the NSF. While these layers do not have the authority of the traditional NSF-project relationship, they do demand attention from projects. At the same time, these multiple evaluation requirements may not necessarily be clear or useful to individual NSDL projects, for a number of reasons. The requirements may be novel; they may be in addition to the goals stated in NSDL program solicitations; they may differ from individual project goals; and they may require a learning curve.

Individual NSDL Pathways projects have therefore to identify, translate and adapt these complex NSDL evaluation channels objectives into their own strategic objectives. In cases where NSDL program goals do not align with individual project goals, and given the limited evaluation resources sometimes available to projects, there is a danger that the program goals will take a back seat to project goals or worse, remain unrecognized or ignored. There is the possibility that the evaluation data which NSF sees as being useful for the purpose of synthesis and program evaluation, may not be the same data that individual projects see as being useful for their own projects. Individual projects may even see some of the data useful
to NSF as being harmful to their participation in future NSF solicitations, in the sense that they could hurt a project’s standing in the eyes of NSF.

**Web Metrics – An NSDL Evaluation Case Study**

In this section, we illustrate this multi-channel evaluation environment through a discussion of the use of web metrics in NSDL. Web metrics are methods for collecting and analyzing traffic to and through a website. Apparently objective, they can yield widely different statistics. Here are the May 2006 figures for the NSDL portal at nsdl.org, as measured by two different web metrics tools used by NSDL-CI, and normalized to a nominal 1000 visits to nsdl.org as recorded by Tool B:

<table>
<thead>
<tr>
<th></th>
<th>Tool A</th>
<th>Tool B</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Hits</strong></td>
<td>37,317</td>
<td>n/a</td>
</tr>
<tr>
<td><strong>Page views</strong></td>
<td>14,401</td>
<td>4,165</td>
</tr>
<tr>
<td><strong>Visits</strong></td>
<td>1,875</td>
<td>1,000</td>
</tr>
</tbody>
</table>

The 37:1 ratio between hits recorded by Tool A and visits recorded by Tool B for the same time period underscores that data from individual NSDL projects are to be treated with caution, unless it is known how they were recorded and calculated. It follows that implementing useful NSDL program-wide web metrics requires standardization in tools and units of measurement.

NSDL-CI has worked for a number of years to standardize web metrics across the projects (Jones et al., 2004). However, as of 2006, NSDL projects still used a range of tools, with the attendant issues of lack of standardization. Beginning in 2006 NSDL-CI began requiring, through a clause in the MoU, that Pathways projects implement a common javascript-based page tagging web metrics tool on each of their sites. The implementation was supported by NSDL-CI, which paid the costs of the implementation, provided specialist technical support, and organized web metrics workshops. It was hoped that this would lead to consistency in
overall program data aggregation. However, the implementation was slower than had been expected, for several reasons.

First, there is often a lack of web metrics resources at the individual Pathway level. Pathways project managers are typically faculty who work part-time on their project, supported by part-time student assistants, and who face a learning curve with respect to implementing, configuring, and reporting web metrics. Second, implementation also requires access to individual site code (to add the javascript), and to the server (to upload the code); and some Pathways with external web developers have had trouble accessing the project servers in order to implement the code. Third, while the Pathways were required under the MoU to implement the web metrics tool, there was no formal provision for NSDL-CI to enforce either the MoU or the specific web metrics requirements that it contained. Combined with the fact that the Pathways projects often had more immediate priorities than implementing web metrics, such as getting their site up and running in the first place, it was perhaps not surprising that the web metrics implementation proceeded slowly, with the best results being obtained with Pathways that were more established or that had full-time staff with the necessary expertise.

These outcomes can be mapped onto the organizational structure in Figure 2, as follows:

- Arrow 1. There was no official requirement in NSDL solicitations for individual projects to report web metrics direct to NSF; consequently, little reporting took place through this channel.

- Arrow 2. There was a quasi-official requirement, described in the MoU, for projects to implement and report standardized web metrics to NSDL-CI. As described above, this implementation proceeded patchily.

- Arrow 3. There was an official requirement for NSDL-CI to report what web metrics data it did collect to NSF. This reporting occurred on a monthly basis (NSDL n.d.c.)
Chapter x: Chapter Title

Arrow 4. There was a quasi-official requirement for NSDL projects to work with the EIESC to develop and report web metrics. Once again, this work proceeded patchily.

In sum, there was no formal organizational channel that encouraged the consistent reporting of program-wide web metrics data to NSF; however, there was a series of informal channels, which although they yielded more heterogeneous data, still provided data that were of use at the NSDL program level.

NSF REFLECTIONS ON EVALUATION IN VIRTUAL ORGANIZATIONS

The NSDL web metrics evaluation work just described attempted to integrate a range of projects, tools, and reporting channels, with varying degrees of success. The work highlights several important lessons related to project and program evaluation, and the relationship between these two activities, especially in virtual organizations.

First, since projects need considerable expertise and resources to properly adopt an evaluation activity such as web metrics, NSF programs must ensure at the outset that projects have adequate budget and staffing to support the appropriate implementation of the software. Early in the NSDL program that turned out not to be the case. Only with this foundation in place will it be realistic for the program to impose a requirement that individual projects implement a common set of web metrics and make the data available for shared cross-project analysis.

More broadly, agencies such as NSF must understand that the use of such tools for evaluation come with real costs as well as potential benefits. These costs not only include technical ones, but social and strategic ones as well. Different projects, even within a single program like NSDL, can have distinct goals, and the use of a common set of web metric tools and measures will not necessarily meet their evaluation requirements. A new NSDL project, for example, may be uninterested in gathering web metrics data about the volume or geographical diversity of its users; they may more interested in developing a web site in the first place. The program may therefore want to accommodate these diverse needs by, for example, encouraging
projects to use web metric tools—or other ones—in different ways to address their specific usage and project-level evaluation questions. However, at the same time the program could also insist that they gather additional data using common web metrics adopted by other projects, in order to address broad program-level evaluation questions. This, in turn, requires the program to have a solid understanding of its key evaluation questions, and, in the case of NSDL, how innovative web metrics can be used to address them.

**Supporting evaluation in virtual organizations**

Ideally a program would establish these questions and their associated standard metrics early in the life cycle of projects, providing them with clear guidelines on what to measure, when, and how. However, in dynamic programs like NSDL that explore emerging information technologies, foster growing virtual organizations, and include multiple evaluation channels, such detailed advanced planning can be impractical. Nevertheless, there may be several ways of dealing effectively with this challenge in the future.

*Clarity.* It would be helpful if both projects and programs established, at least tentatively, their major evaluation goals early on, and—since such goals can change, sometimes rather radically, over time—also engaged in a continuing dialogue on evaluation goals and methods. Conversations at this level could, for example, help NSDL program managers and the NSDL community to understand how the adoption of web metrics might address (or not) the needs of specific projects. Conversely, the discussions would also provide projects with insights about program-level evaluation goals that web metrics might speak to, even if they do not address project-level evaluation issues.

*Flexibility.* Many projects in programs as diverse as NSDL will have distinct evaluation goals and will need to consider the deployment of a range of evaluation web metrics, as well as other measures of progress—as will the NSDL program as a whole. If web metrics technologies are to be used on a broad basis to help evaluate digital library projects and
programs, they will need to be implemented in standardized ways across sites; for example, projects should adhere to common definitions of terms like ‘visits,’ ‘unique visits,’ and ‘page views.’ The tools, however, will also need to be supple enough to be tailored to the evaluation needs of specific projects and programs.

Research. Finally, a significant challenge associated with the effective use of web metrics in evaluation of NSDL projects and the program as a whole, is that not only are the metrics new and somewhat unstable, but the science of making inferences from them to users’ intentions is still in its infancy; we can see the results of users’ interactions with a web site, but we do not have access to the thought processes that gave rise to those interactions.

In order to make highly effective use of the promising new ideas behind web metrics, NSDL and related programs at NSF will therefore need to continue to support research on web metrics, not just fund deployment of them. NSDL is an ideal testbed for a many interesting research studies that could, for example, track a range of metrics longitudinally, and triangulate web metrics data and other sources of information on user behaviour (Khoo et. al. 2008). This sort of research could lead to a new generation of web metric tools and protocols for digital library evaluation and improvement.

RECOMMENDATIONS

This web metrics case study suggests several broad recommendations that agencies, their programs, and their projects might follow to improve the success of their evaluation plans.

First, projects should have well-developed individual evaluation plans, and the programs and agencies that support them should provide sufficient resources and requirements to ensure the project evaluations yield useful data. This seems straightforward, but project budgets are sometimes cut and often evaluation plans are compromised. In other cases, promising educational projects are still generating useful data when the project ends; absent an extension and supplementary funding such projects may fail to deliver their most significant results.
Second, at a broader level, programs and program directors who understand program-level evaluation needs should, as Frechtling notes, make clear to individual projects what the requirements of this wider context are, and how they impact the projects. Ideally, directors would articulate a logic model (Trochim, et. al. 2008) for the program that encompasses all project inputs, resources, outputs, and outcomes. The connections between program goals, and hence its evaluation priorities, and project evaluation data may be multiple and nuanced.

For example, as the web metrics case study illustrates, NSDL has a program level evaluation goal to track the growth of its collections. NSDL projects that manage collections would clearly provide evaluation data relevant to this program goal; however projects that develop specific tools and services, which rarely maintain collections, would not -- although their evaluation data could inform other NSDL program-level goals. This suggests that programs need not only to communicate their overall evaluation goals to projects, but should also translate these into different requirements for distinct types of projects. Additional resources may be required to ensure that projects can fulfil these evaluation tasks as well as their individual project evaluation plans.

Finally, agencies and programs will need to confront the evaluation complexities inherent in the growing trend towards programs, like NSDL, that connect individual projects into interdependent virtual organizations. Articulating beforehand how program-level evaluation needs translate into project-level evaluation requirements will not be possible in programs whose components reconfigure from time to time. As new teams and committees arise, which they have done through the lifetime of NSDL, agency program directors – as well as the virtual organization groups that hold governing authority – will need to interpret these changes in terms of transformed program-level goals, and translate them into new project-level evaluation needs. These same decision-makers will also have to establish the appropriate
levels of compliance for new evaluation mandates, and, if they are strong requirements, program directors will have to guarantee projects have adequate resources to comply.

CONCLUSION

Virtual organizations like NSDL naturally result in dynamic new relationships among teams and projects, and these connections are often less official and more informal than typical project-program relationships. These multiple evaluation channels are useful (as the web metrics case study demonstrates), and they will become more common as virtual organizations like the NSDL proliferate.

An important theme to emerge from this complexity concerns the nature and role clear of organizational communication in a virtual organization. While sophisticated information and communication technologies promise the possibility of rapid and clear virtual organizational communication between funding agencies and projects, the potential of these communication tools sometimes serves as much to reveal the knowledge differences and communication difficulties between groups, as to support seamless collaboration and knowledge building. From the point of view of NSF, this suggests the importance, when planning any evaluation of a virtual organization that goes beyond aggregating individual project results and considers broader issues, such as the value of the digital library as a whole, of ensuring that all stakeholders in the evaluation process, including agencies and projects, are on the same page with regard to aims, requirements, and resources. This observation will only become more important as NSF continues to fund and evaluate large-scale virtual organizations, including digital libraries.

REFERENCES


Figure 1a (left): Simple evaluation model, showing flow of resources and results

Figure 1b (right): Actual evaluation model in the case of NSF and the NSDL
Figure 2: Resource and results flows in the NSDL.