NISO Thought Leader Meeting on Institutional Repositories

Background and Goals
The National Information Standards Organization (NISO) has recently been awarded a substantial Mellon grant to pro-actively foster new standards activities. As part of this effort, NISO is convening a series of "thought leader" meetings focusing on specific scholarly information issues, including the focus of this report – institutional repositories. There are a number of large issues hindering IR's progress and broad adoption, including: simplifying the deposit of information into the repository; the types and formats of content to be deposited; adequate metadata, authority control, and controlled vocabulary strategies; unified access to compound digital objects; document version control; development of full-text searching, text mining and analysis tools; linking between references, text (including tables and graphs), data, and digital media within and across repositories; preservation and long-term data migration planning; copyright and intellectual property issues; and cross-disciplinary or cross-institutional repository authentication, deposit, coordination, and interoperability. The focus of the IR thought leader was these barriers and their potential solutions. At the end of the day, the group (see below for participant list) was tasked with determining which one or two specific barriers might be reduced or eliminated by standards-based solutions, and how NISO might approach the creation and implementations of these solutions. In this way, the thought leader meeting was meant to set an agenda for future standards development and will assist NISO's mission of facilitating the library, publisher, and vendor communities in problem solving for scholarly communications questions.

Pre-Meeting Activities
The thought leaders participated in a pre-meeting conference call to discuss the session’s agenda. During this call, the group contemplated the pain points that are blocking deeper and wider adoption of institutional repositories. The result was the identification of four general themes:

- **Authoring/Workflow Solutions.** Are there processes that can be developed to better engage would-be depositors? Can the steps by which an author contributes work to an IR be streamlined and/or reimagined to address expressed pain points among researchers? Can IRs emerge as workflow tools that are fundamentally ingrained in the development of scholarship? Can IRs serve not just a place to store scholarship, but as a place to work on and work with scholarship?

- **Distribution Solutions.** Can IR materials be hooked into other distribution channels in an efficient manner? Can crosswalks be created to allow for a simpler and more beneficial sharing of materials among IRs and other content repositories, search engines, databases, etc.? Are current metadata standards sufficient to promote this type of sharing?

---

Greg Tananbaum  
Consulting Services at the Intersection of Technology, Content, and Academia  
www.scholarnext.com  
(510) 295-6504  gtananbaum@gmail.com
• Legal Considerations. Can author agreements be simplified so that the copyright confusion 
suppressing IR participation is reduced or eliminated? Can mechanisms be developed to 
automate the copyright clearance function?

• Redefining the Notion of IRs. Can we develop a better standard for what an IR is, 
what its mission should be, and how success is defined. Given the theoretical 
nature of this theme, and NISO's limited ability to develop a "solution" here, this final theme was 
 omitted from the next stages of the project.

Meeting Schedule
The thought leaders convened in Washington, D.C., on February 12, 2008. The morning 
session examined the various impediments that were identified by the group in the pre- 
meeting activities. The list of pain points was subsequently winnowed down to a handful 
of key issues that might successfully be addressed by a standards-based solution. The 
afternoon session was spent crafting a charge around which a technical working group 
could be created to explore and develop a standard, recommended practice, schema, or 
tool that would mitigate or eliminate the identified pain points.

Meeting Summary
The group began by discussing the scope of institutional repositories, their constituents, 
and their content. It was agreed that NISO would be best served to take an expansive 
notion of IRs as a place where a diverse range of scholars deposit and access objects. 
From this broad view, the group defined the critical functions of institutional repositories 
as ingestion, discovery, use/reuse, interoperability, and preservation/management. The 
focus of the group thus turned to what NISO might be able to do in regard to one or more 
of these functions that would result in a better IR experience, broader adoption, and the 
reduction or removal of perceived roadblocks.

There was a collective belief that scholars have resisted institutional repositories at least 
in part due to the duplication of effort that depositing one’s objects typically involves. A 
scholar’s work often has life in multiple domains, including:

• Journals
• Subject Matter Repositories
• Course Management Systems
• Personal Pages/ePortfolios
• Data Repositories
• Learning Object Repositories
• General Web
• Private Work Groups
• Institutional Repositories

Greg Tananbaum
Consulting Services at the Intersection of Technology, 
Content, and Academia 
www.scholarnext.com 
(510) 295-6504       gtananbaum@gmail.com
Scholars are likely to want their objects discoverable in several of these domains. At present it is a significant hassle to get the object into multiple domains, owing to unique uploading mechanisms, duplication of effort, and so forth. As a result, creators pick and choose which domains to pursue. The group strongly advocated investigation of a solution that renders the incremental cost of depositing across multiple domains virtually zero. It would touch upon the Authoring/Workflow Solutions, the Distribution Solutions, and, tangentially, the Legal Considerations contemplated in the pre-meeting discuss. The group felt confident that this approach would dramatically encourage IR deposits.

A key component in the attainment of this solution will be to capture content much closer to the point of creation. Eventually, the “boxes” around IRs and the other domains outlined above may give way to an environment in which work is created, revised, reviewed, disseminated, and discussed within a single “pool”. The barriers between desktop and web, creation and discovery, upload and download would be blurred. It may well be that NISO’s work in the development of this IR solution will help accelerate this reality. In the short term, however, the goals are more modest. Capturing content closer to the point of creation will render IRs a more organic part of a scholar’s workflow. It will, for example, reduce the need for an author to double back after a paper is written and reenter its metadata into a dedicated IR form. It will provide a significant enticement to scholars by allowing them to deposit objects across multiple domains, including IRs, at a convenient time in the object’s life cycle.

The suggested tool recommended for pursuit by the group can be succinctly described as follows. A common deposit mechanism should be developed to allow institutional repositories to capture objects as close to their creation point as possible (Authoring/Workflow Solutions). The capture of these objects will be part of a larger context that will allow for their exposure across a variety of domains such as journals, subject matter repositories, and course management systems (Distribution Solutions). How might this be executed in practical terms? The tool in question would likely be a desktop widget that exposed scholarly objects to their creator’s choice of the multiple domains outlined above. A handful of specific domains would need to be part of the beta process (i.e., selected test journals, content management systems, institutional repository systems, subject matter repositories, and learning object repositories). The widget would capture or allow the author to enter appropriate metadata for the selected systems; when applicable, determine publisher e-print policies; optionally select author addendum or Creative Commons licenses for the work; and submit the work to both the selected repositories and, when applicable, to the target journal using appropriate authentication mechanisms. Once the prototype was developed and tested, NISO and its partners would be responsible for analyzing results and extrapolating standards off of the experience. Looking past the testing phase, the actual mechanism/widget/tool by which common depositing was accomplished need not be tightly managed by NISO, or anyone else for that matter. Among those that might earn the NISO “seal of approval” through adherence

Greg Tananbaum
Consulting Services at the Intersection of Technology, Content, and Academia
www.scholarnext.com
(510) 295-6504 gtananbaum@gmail.com
to the standards and best practices identified in the prototype stage: commercial versions
of the widget that offered advanced functionality such as deposit destination suggestions
and rights management guidance (the Legal Considerations theme identified in the pre-
meeting call); and tools that incorporated the common deposit mechanism into an
existing scholarly communication workflow process (for example, journal submission
systems). Similarly, content destinations from the domains outlined above could make
themselves eligible to receive the exposed objects by adhering to the developed
standards. In essence, NISO would facilitate a common deposit pilot project in order to
develop a protocol. Subsequently, the deposit process would be open to any other entities
that adhered to the NISO-developed standard.

The specific next steps as envisioned by the group are as follows. NISO should first
determine the mandatory and optional elements that must be captured by the tool. It
should explore the panoply of use cases that might shape the tool, including the specific
idiosyncrasies of various academic disciplines, the range of possible deposit scenarios,
the wide variety of content types that might be suitable, and so forth. NISO then should
inventory the standards that might be repurposed or leveraged to address these use cases,
content types, and data elements. This should not be done unilaterally. Rather, NISO
should identify a group of representative stakeholders who would benefit from the
creation of such a tool. This group might be small and nimble or a grand coalition of
prospective beneficiaries, at NISO’s discretion. Together, NISO and its partners could
facilitate and coordinate the development of a rapid widget prototype, then develop
standards/best practices off of that experience. A successful execution of this project
would be judged by two criteria: (1) the extent to which the deposit population views this
tool as a markedly easier process for uploading content into IRs and other domains, and
(2) the extent to which the standards that emerged from the prototype stage facilitated
wide adoption of the common deposit protocol by a variety of actors (software providers,
publishers, information providers, universities, etc.) within the scholarly communication
chain.

The group contemplated a number of specific considerations in the pursuit and
implementation of this goal. Among the questions a NISO-led effort might expect to
address are the following:

- To what extent can the common deposit mechanism facilitate standardization of
  metadata elements at the time of creation (e.g., institutional affiliation, subject
  area)?
- How can the common deposit mechanism account for the version control of
  objects?
- How do objects that utilize the common deposit mechanism move from the
  desktop to the participating domains? Is this a push or pull process?
• How best can the fields that a content creator completes as part of the common deposit process be contextualized based on the domains he/she selects as exposure paths?
• Can the common deposit mechanism incorporate data as well as text, and, if so, can the data be captured at the instrumentation level?
• Can the common deposit mechanism move beyond desktop widget and also become an integrated component of existing deposit protocols (PubMed Central, Blackboard, journals, etc.), so that submitters to those systems have the opportunity to expose their content to these other domains?
• How can the ingest process incorporate a rights component that renders use terms portable with the object?

Participants

<table>
<thead>
<tr>
<th>Name</th>
<th>Organization</th>
<th>Title</th>
</tr>
</thead>
<tbody>
<tr>
<td>Michele Kimpton</td>
<td>DSpace Federation</td>
<td>Executive Director</td>
</tr>
<tr>
<td>Sayeed Choudhury</td>
<td>Johns Hopkins</td>
<td>Associate Director for Library Digital Programs</td>
</tr>
<tr>
<td>Teresa Ehling</td>
<td>Cornell</td>
<td>Director, Center for Innovative Publishing</td>
</tr>
<tr>
<td>Catherine Candee</td>
<td>CDL</td>
<td>Exec Director, Strategic Publishing &amp; Broadcast Services, UCOP</td>
</tr>
<tr>
<td>James Hilton</td>
<td>University of Virginia</td>
<td>CIO and Professor of Psychology</td>
</tr>
<tr>
<td>Paul Curran</td>
<td>Bournemouth University</td>
<td>Vice Chancellor</td>
</tr>
<tr>
<td>Larry Lannom</td>
<td>CNRI</td>
<td>Director, Information Management Technology</td>
</tr>
<tr>
<td>Rob Tansley</td>
<td>Google</td>
<td>Software Engineer</td>
</tr>
<tr>
<td>Charles W. Bailey, Jr.</td>
<td>Digital Scholarship</td>
<td>Publisher</td>
</tr>
</tbody>
</table>