ARTICLE EXCERPTED FROM:

INFORMATION STANDARDS QUARTERLY

FALL 2009 | VOL 21 | ISSUE 4 | ISSN 1041-0031

SPECIAL ANNIVERSARY EDITION: PART FOUR

LIBRARY AUTOMATION IN INDIA: PAST, PRESENT, AND FUTURE

NISO CELEBRATES 70 YEARS, PART 4: WHERE TO FROM HERE?

RATING ILS INTEROPERABILITY

ERMS WORKFLOW: TWO VIEWS FROM THE TRENCHES

INSTITUTIONAL IDENTIFIERS IN REPOSITORIES: A SURVEY

NISO
How the information world CONNECTS

Michael Giarlo

MICHAEL GIARLO

Institutional Identifiers in Repositories: A Survey Report

The National Information Standards Organization (NISO) established a working group in July 2008 to recommend an identifier standard, with associated metadata and implementation strategy, for identifying institutions involved in information creation, sharing, and management. An institutional identifier is defined as a symbol or code that will uniquely identify institutions and that will describe relationships between entities within institutions.

The Institutional Identifiers (I^2) Working Group, co-chaired by Grace Agnew (Rutgers University Libraries) and Tina Feick (Harrassowitz), is also charged with defining what minimum set of data is required for unique identification and what other data may be used to support the business models of respective organizations. As a first step, the I^2 Working Group identified three compelling scenarios for usage of the I^2 identifier: the commercial information supply chain, library workflow, and institutional repositories (IRs). The subgroup charged with the IR scenario surveyed institutional repository managers and developers to determine the current practices and needs of the IR community regarding an institutional identifier. This article is a summary of the survey report. The complete report is available on the NISO website. The I^2 IR scenario subgroup is incorporating the survey findings and the group's conclusions into their final scenario.

Audience and Distribution

The intended audience of the survey was repository managers and developers. In order to increase the diversity of respondents, the group decided to take two approaches.

First, the group nominated a number of repositories considered prominent and augmented this short list with repositories identified via OpenDOAR, a directory of open access repositories. The directory allowed the group to associate potential survey respondents with repositories, and to choose repositories that are diverse with regard to geography, type of repository, software platform, and industry. The group decided that one hundred was a good number of potential respondents.

Second, acknowledging that any such list would be incomplete, the group identified a number of mailing lists that were likely to be followed by the repository community. These lists are enumerated in Appendix A of the full report.

The survey was distributed via the Survey Monkey website on June 18th, 2009 to the one hundred individually-chosen repository contacts and via the group to the identified mailing lists, as well as from group members' personal blogs.

The survey was distributed via the Survey Monkey website on June 18th, 2009 to the one hundred individually-chosen repository contacts and via the group to the identified mailing lists, as well as from group members' personal blogs. Survey Monkey generated one link for each of these purposes so that results from individually-chosen contacts and those from listservs and blogs could be kept distinct, which was useful for group members to gauge the success of each approach. The survey remained open until Monday, July 6th, 2009, a period of seventeen days.

It is likely that repositories from academic and research libraries may have been overrepresented in the survey results. The IR scenario group intends to include repository communities from public libraries, archives, and other less well-represented sectors in future work.

Response Analysis

29 of the 100 identified repository contacts responded to the survey, with 21 of these completing the full survey. 136 persons responded to the survey sent out to mailing lists and blogs, with 81 of these completing the survey. In total, the survey had 165 responses, of which 102 respondents answered every question.

CONTINUED »

SELECTED FINDINGS

A detailed summary of all the questions and responses is available in the full report. Here are some selected findings of interest:

INSTITUTIONAL IDENTIFIER USAGE

58.1% of repositories include identifiers for themselves, 49.7% of which are public. 41.9% do not include identifiers for themselves.

46.1% of repositories include identifiers for their organizations, 35.6% of which are public. 62.9% do not include identifiers for themselves.

74.2% of repositories that include institution identifiers also include identifiers for institutional subdivisions. 26.9% are used only internally.

ASSIGNMENT OF INSTITUTIONAL IDENTIFIERS

37.5% use systems to assign institutional identifiers:

- Handle.net
- DSpace
- DNS
- OCLC
- ISIL
- ePrints
- EDINA
- California Digital Library

41.7% use manual processes to assign institutional identifiers:

- By the repository team
- By a single individual
- By an outside department

9.7% use a combination of manual processes and systems to assign institutional identifiers.

READ THE FULL REPORT AT:
www.niso.org/apps/group_public/document.php?document_id=2855

ISSUES POTENTIALLY SOLVED BY A STANDARDIZED INSTITUTIONAL IDENTIFIER

31.9% have yet to encounter any issues they would consider potentially solvable by standardized institutional identifiers.

14.9% state a standardized institutional identifier would have helped track institutions across name changes, disambiguate similarly-named institutions, and tie collections to institutions.

10.6% state a standardized institutional identifier would have helped identify and enumerate organizational units, especially in multi-lingual environments.

8.5% state a standardized institutional identifier would have helped tie authors to institutions.

Other issues:

- Uniqueness
- Interoperability
- De-duplication
- Persistence
- Statistics
- Indexing
- Workflow

IDENTIFIERS AND CONTEXTS

56.6% report that institutional identifiers used in the repository are not used for other library activities (e.g., electronic resource sharing, ILL, etc.)

22.6% report that these identifiers are used in other contexts.

60.3% consider it important to have a single identifier that serves all organizational purposes. 25.4% do not consider it important.

Clear Trends

The survey showed that standardized institutional identifiers are seen as important and it was agreed there is a need for them in the repository community. The need for identifiers is underscored by the ways in which repository content is shared. A clear majority of repositories include identifiers for the repository itself and many include institutional identifiers. Those that include the latter generally also include identifiers for subordinate units within the identified institution. Most of these identifiers are not used in other usage contexts—e.g., Inter-Library Loan, electronic resource management systems, etc.—but there is some agreement that it would be important for a single identifier to be used for all organizational purposes. The majority of respondents would be willing to participate in a registry of institutional identifiers provided that participation is voluntary and cost free.

Institutional identifiers already in use are largely based upon the Uniform Resource Identifier (URI) standard, whether they take the form of Hypertext Transfer Protocol (HTTP) URIs, Uniform Resource Names (URNs), CNRI Handles, or OCLC PURLs. An overwhelming majority of respondents consider resolvability of institutional identifiers important.

Metadata Elements

The core required metadata associated with an institutional identifier should be the Institution Name element, the Parent Institution element, and the Uniform Resource Locator (URL) element. A Region element is largely considered unnecessary, and pluralities consider Address and State/Province unnecessary. Most repositories are already collecting some or all of the core metadata elements considered required or preferred. There is little agreement on the necessity of the following core metadata elements: Related Institution, Variant Name, City, and Country.

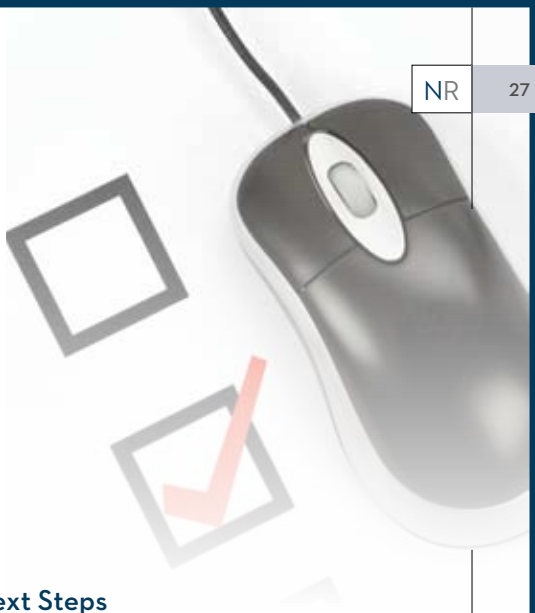
Areas with Little Agreement

Institutional identifiers are assigned in various ways: some are handled manually, others via automated processes, and others via a combination of manual and automated processes. A third of respondents would prefer to reflect institutional hierarchy in the identifiers, with nearly as many preferring to have non-hierarchical identifiers. There were a range of answers to the question of which organization would be best-suited to manage a registry of institutional identifiers.

Conclusions

After analyzing the survey results, the IR Scenario sub-group summarized their conclusions as follows:

- Participation in a registry of managed institutional identifiers should be voluntary and cost free.
- Institutional identifiers should be resolvable.
- Assignment of identifiers should be possible via both manual and automated processes.
- Each participating organization may or may not have a primary institution identifier.
- The relationship and provenance of the institution governed by the identifier should be captured in the identifier metadata, as the hierarchy may not be durable.
- Thus, an institution may use only a single identifier or may have multiple identifiers assigned to whatever division they find useful locally. Said division may be by research units, departments, institutional repositories, projects, or other division as needed by the institution.
- An institution has the right to use the primary institution identifier to represent its institutional repository or other processes as needed, if they prefer not to manage multiple identifiers.



Next Steps

All of the I² scenario work is nearing completion. The working group has already begun using the scenarios to define a set of required and optional metadata elements and to position the I² identifier with other existing identifiers. Also under discussion are the issues of registry and a maintenance/registration agency. You can follow the work of the I² working group on their public workroom page or by signing up for the I² Info interest group mailing list.

| NR | doi: 10.3789/isqv21n4.200907

MICHAEL GIARLO <leftwing@alumni.rutgers.edu> is an Information Technology Specialist at the Library of Congress and co-chair of the I2 IR Scenario Group.

OTHER MEMBERS of the I2 IR Scenario Group are: Jessica Colati (Colorado Alliance of Research Libraries), Jody L. DeRidder (University of Alabama), Robert Harris (NJVid and William Paterson University), Amanda Hill (JISC Names project, University of Manchester), John Kunze, (California Digital Library), Lisa Macklin, co-chair (Emory University), and Linda Tadic (Audiovisual Archive Network).

RELEVANT LINKS

I² Working Group Workroom
www.niso.org/workrooms/i2

Institutional Identifiers in Repositories Survey Report
www.niso.org/apps/group_public/document.php?document_id=2855

I² Info Mailing List
www.niso.org/lists/i2info/

OpenDOAR
www.opendoar.org/