

SCHOLARLY PUBLISHING OFFICE WHITEPAPER*

A comparison of persistent identifier systems for SPO

Kevin S. Hawkins

Executive Summary

This paper compares SPO's current use of the CNRI Handle System to possible uses of *DOIs* (digital object identifiers), both those registered by U-M and those deposited through CrossRef. Joining CrossRef is worth considering because of the benefits gained through an economy of scale, but SPO would need to commit itself to fulfilling the membership requirements and could face conflicts of interest.

Background

Computer scientists have long recognized that a fundamental design flaw in the World Wide Web is that the lack of persistent identifiers for Web resources. Various remedies to this flaw have been developed, all of which work on top of the existing Web infrastructure. At U-M, Core Services chose the CNRI Handle System® over ARK and PURL for University Library content, and SPO uses them for nearly all of its publications. However, another solution (also turned down by Core Services) has gained traction in both the publishing and library communities: the DOI® System, which is “managed by an open membership collaborative development body, the International DOI Foundation (IDF), founded in 1998.”¹

Handles

SPO could continue creating handles and using the CNRI Handle System infrastructure implemented by Core Services.²

A handle is a persistent identifier for an object (such as 2027/spo.3336451.0010.202) which can be resolved into a URL using a resolution service (such as the Handle System Proxy Server, <http://hdl.handle.net/>). The handle itself must be a string composed of UTF-8 characters beginning with a *prefix* representing a naming authority (in this case, 2027) followed by a slash and then a *suffix* representing an object identifier. The Handle System does not specify what exactly a handle resolves to, allowing for a variety of applications, including resolution to more than one object or to metadata about the object.

The University Library is a *resolution service provider*, having paid a one-time \$100 to receive its own prefix (2027) and run a Handle System identifier/resolution system. It can therefore create its own handles that begin with this prefix. The University Library currently uses only ASCII characters in its handles, and they resolve to a URL for the digital object's current location.

* This work is licensed under a Creative Commons Attribution 3.0 License. To request permission to use this content in a way not allowed by the Creative Commons license, contact copyright@umich.edu. © The Regents of the University of Michigan, 2007.

¹ http://www.doi.org/overview/sys_overview_021601.html

² For more on the Core Services handle infrastructure, see <http://www.umdl.umich.edu/docs/arch/general/handle-conventions.html>.

DOIs: the University Library as registration agency

In addition to, or instead of, creating handles through the University Library's identifier/resolution system, SPO could create DOIs for its digital objects by seeking to become a DOI registration agency.

A DOI is a persistent identifier following the same syntax as a handle (having the form “prefix/suffix”), except that the prefix always begins with 10 . and ends with a number assigned to an organization that wishes to register DOIs (a DOI *registration agency*). That is, there is a sort of compound prefix composed of two parts, followed by a suffix that identifies a particular object within that namespace.

A DOI can be resolved into a URL using a resolution service (such as the DOI System Proxy Server, <http://dx.doi.org/>). The DOI itself currently must be a string composed of ASCII characters, though expanded capability of including Unicode (presumably UTF-8) characters, as allowed in the Handle System, is envisioned.³ Currently DOIs resolve to one or more digital objects,⁴ or possibly to metadata about an object in a standard format.⁵

The DOI System's resolution service is an application of the CNRI Handle System.⁶ However, the “implementation [of the Handle System] in the DOI System has been supplemented by expanded technical infrastructure and features specific to DOI System applications,”⁷ so essentially the DOI System's resolution service is a special case of the Handle System.

The IDF charges an organization \$35,000 per year to be a DOI registration agency, but these organizations are expected to contribute to the governance of IDF before and after becoming a registration agency. Most registration agencies allocate DOIs to other organizations, similar to how domain registrars allocate domain names within a TLD, though some are themselves large publishers or libraries with many digital documents.

In short, the DOI System has the following advantages compared with the Handle System:

- DOIs can be used with various tools⁸ for end-users and programs.
- The DOI System includes a data model with specified semantics, plus a social infrastructure, that encourages interoperability and broad adoption by the publishing community.

CrossRef DOIs

A third option—again in addition to or instead of using handles—is for SPO to join CrossRef and deposit DOIs according to the requirements of this organization.

CrossRef, a service run by Publishers International Linking Association, Inc. (PILA), claims to be “the official DOI(R) link registration agency for scholarly and professional publications. It operates a cross-publisher citation linking system that allows a researcher to click on a reference citation on one publisher’s platform and link directly to the cited content on

³ <http://www.doi.org/factsheets/DOIIdentifierSpecs.html>

⁴ http://www.doi.org/overview/sys_overview_021601.html

⁵ http://en.wikipedia.org/wiki/Digital_object_identifier . For more information, see http://doi.org/handbook_2000/metadata.html

⁶ Therefore, DOIs can be resolved through the global handle resolver at <http://hdl.handle.net/> in addition to other resolution services. (See http://en.wikipedia.org/wiki/Digital_object_identifier .)

⁷ http://www.doi.org/overview/sys_overview_021601.html

⁸ <http://doi.org/tools.html>

another publisher's platform, subject to the target publisher's access control practices."⁹ CrossRef originally allowed only DOIs that resolve to one URL but now allows multiple resolution, using either an interim page with a choice of links or a pop-up menu when mousing over a DOI.¹⁰ CrossRef specifies that a single DOI shall be shared among all manifestations of a given work¹¹ and is therefore "owned" (managed) by the publisher which owns the copyright in a given work.

CrossRef charges publishers at least \$250 per year (on a sliding scale depending on revenue) to receive one or more Handle prefixes (beginning with 10.), plus a fee for each DOI deposited and annual fines (per article) for not linking "a majority of its current journal content" to existing DOIs.¹² This creation of "outbound links" can be automated by querying the CrossRef Resolver.¹³

In addition to the general benefits of DOIs given above, there are additional benefits to participation in CrossRef created by its economy of scale:

- SPO would be coerced into providing links to content from other publishers using DOIs, making its own content more useful.
- Other publishers would discover SPO content and create links to it using DOIs, increasing exposure of our content.
- The DOI directory is OpenURL-enabled, so users querying a citation would be directed to copies of the item available through institutional subscriptions, when applicable.¹⁴
- Users (and SPO staff, when troubleshooting) could search for SPO articles directly from the CrossRef website by using a form¹⁵ or pasting a free-text citation.¹⁶

However, there are also concerns:

- SPO's existing content lacks structured markup for almost all citations, preventing use of automated querying of the CrossRef Resolver. The free-text citation analyzer might work; otherwise, SPO would need to enhance its markup before querying the Resolver.
- CrossRef members are required to register DOIs for all *definitive works*: journal articles, conference papers, and book titles¹⁷ but not preprints, postprints, or self-archived versions of works.¹⁸ In cases where SPO publishes only one of a number of equally "definitive" (in our opinion) manifestations of a work, SPO would need to coordinate multiple resolution with other publishers. This could be especially problematic in cases where SPO believes it has the right to publish a manifestation of the work while the rightsowner does not wish to cooperate. (For example, the publisher of the Garo books might put them online and charge for access but not want to indicate that the content is freely available for SPO.)

⁹ <http://www.crossref.org/>

¹⁰ http://www.crossref.org/mr/mr_main.html

¹¹ <http://www.crossref.org/02publishers/doi-guidelines.pdf>

¹² http://www.crossref.org/02publishers/20pub_fees.html

¹³ See http://www.crossref.org/02publishers/25query_spec.html for various methods of querying the Resolver.

¹⁴ <http://www.crossref.org/03libraries/16openurl.html>

¹⁵ <http://www.crossref.org/guestquery/>

¹⁶ <http://www.crossref.org/freeTextQuery/>

¹⁷ http://www.crossref.org/02publishers/24upload_spec.html

¹⁸

Privacy policy of IDF and CrossRef

There is an additional concern with using DOIs with the University Library as a registration agency or when depositing through CrossRef. IDF's privacy policy states that "Our logs collect and store only domain names or IP addresses, dates and times of visits, and the pages visited. Data from the logs may be used to measure the number of visitors to the site."¹⁹ The policy of collecting IP addresses was question on a discussion page on Wikipedia,²⁰ with a user citing email communication "with CrossRef and DOI" saying that they use IP addresses to check for whether the system is abused.²¹ CNRI, on the other hand, does not mention collecting IP addresses.[<http://www.handle.net/privacy.html>]

If implementing DOIs, we will need to investigate whether use of DOIs would violate the Library's privacy policy²² or would otherwise be inappropriate for SPO.

Recommendation

Becoming a DOI registration agency would be quite expensive, and it's unclear that SPO could gain approval as a registration agency. Furthermore, there are no clear benefits to doing so rather than continuing to use the Handle System. While the Core Services implementation of the Handle System is currently simple, SPO could work with Core Services to add features available to DOI registration agencies.

While joining CrossRef is worth considering because of the benefits gained through an economy of scale, there is a possibility that SPO would encounter a conflict of interest over intellectual property rights, forcing it to choose between supporting our philosophy of access to content and our obligations as a CrossRef member. SPO would need to commit itself to fulfilling the membership requirements or risk needing to pay penalties, and it seems unlikely that SPO would ever be able to do this while still maintaining its production-oriented publishing workflow, which requires a scalable, cost-efficient, and somewhat generic model for SPO's publishing. Citations in Level 1 and Level 2 collections would be especially problematic since the citations themselves can be in inaccurate OCR text.

In short, since the Handle System already provides the persistence that SPO most needs in its persistent identifiers, becoming a DOI registration agency or joining CrossRef would provide SPO with few additional benefits but many additional costs. In addition, no users or content providers have ever requested DOIs as a feature of our publications, and it seems that few cited sources in our content have DOIs themselves. Therefore, there's no external motivation for SPO to implement them. SPO should continue using the Handle System implementation run by Core Services.

¹⁹ <http://www.doi.org/privacy.html>

²⁰ http://en.wikipedia.org/wiki/Talk:Digital_object_identifier

²¹ http://en.wikipedia.org/wiki/Talk:Digital_object_identifier#Privacy_protection.3F

²² <http://www.lib.umich.edu/policies/privacy.html>