NISO Work Item:
Recommended Practice for Tracking Link Origins in a Networked Information Environment
(Short Title: Link Origin Tracking)

The following proposed work item is submitted by:
Scott Bernier and Oliver Pesch, EBSCO Information Services

Proposal Last Modified [Copyedit]: December 10, 2015
Approved by NISO Discovery to Delivery Topic Committee
September 21, 2015

Work Item Title: Develop a recommended practice that would allow libraries, publishers and other content providers to accurately track the sites/platforms from which incoming links originate when they pass through a link resolver.

Background and Problem Statement:
Libraries strive to improve the ways in which users access their collections. Gaining a definitive understanding of where a user began his or her library experience/search before ultimately arriving at the content licensed by a library is an important factor in determining the value of a platform, how to allocate resources, etc. Publishers, recognizing that users have a number of options for discovering content, supply metadata to various discovery and abstracting and indexing vendors in the hopes of increasing the discoverability and use of the content they create and to which their customers subscribe. Publishers who wish to measure the success of their metadata programs will often turn to web log analysis to track where their users are coming from.

Web log analysis typically analyzes HTTP referrer information; however, in today’s networked information environment, simple HTTP referrer analysis may misrepresent the results. Figure 1 provides a simple model of the paths through which a user may link to a publisher’s content.

Figure 1 depicts three paths to full text: using a DOI to link through CrossRef, using an OpenURL link resolver, and a direct link to the content host. If a content host is using HTTP referrer analysis to determine the site where the user started his or her research, the links coming through a link resolver will represent the domain of the link resolver and not that of the platform where the user originated the search.
Initial studies from EBSCO extrapolate that upwards of 50% of links pass through link resolver channels, thus “losing” the data relative to search origin. Critical data for publishers and libraries are skewed because the origin of the users request to retrieve the document appears to come from the provider of the link resolver. When a link resolver provider is also a discovery provider or other platform provider, the co-mingling of these data points further skews usage rates and value propositions.

The goals of this proposed work item are to:

- Investigate effective and scalable options to allow the content host and individual libraries to determine the original source of the link;
- Provide all parties in the information supply chain with guidance on how to implement the recommended solutions; and
- Provide content hosts with information on how to leverage the proposed solution to obtain more accurate statistics on the source of traffic to their site so that they can more accurately assess the value and effectiveness of their various discovery partners.

*As a by-product of this work, publishers will be in a position to provide their customers with reporting on how users are discovering their content.*

**Statement of Work:**

**Project Goals:**

1. Investigate options for passing the link origin information of the discovery or A&I service to the publisher. The investigation will focus on link resolver pathways, while considering other link pathways, including those making use of DOI-handle servers.
2. Produce a report on this technology that will serve as a recommended practice for link resolvers, content hosts and potentially discovery/A&I vendors to implement and take advantage of the passing of link origin information.
3. Implement one or more proof-of-concept projects that demonstrate the proposed technique.

**Specific Deliverables:**

1. A NISO Recommended Practice outlining the recommended approach to passing and using link origin information.
2. A promotion and education plan.
3. One or more proof-of-concept services that can exchange the link origin information using the proposed approach.

**Process:**
Create a working group that will meet via telephone and WebEx to create the deliverables outlined above. Test versions of the proposed solution will be developed in concert with the specifications so that proposed concepts can be tested and refined.

**Partners and Participation:**

The committee members should be drawn from the community that would benefit from these new features and who have the skill set to produce a specification that is effective in using current approaches to web services. The organization types to be represented are as follows:

- Publishers – the primary consumers of the link origin information used in reporting
- Link Resolver vendors – who may be required to implement technology changes
- Discovery or A&I vendors – who are interested in ensuring their sites get attribution for the link-outs they generate
- Administrators of DOI handle servers
- Librarians – who are interested in understanding how their users are accessing various content provider web sites.
The following skill sets from working group participants are needed; these may come from any of the participating organizations and a given individual may possess more than one of these skills:
- Individuals with experience in network protocols who could propose or evaluate proposals based on HTTP headers, etc.
- Developers of OpenURL link resolvers to assist in assessing the effort of implementing various proposals
- Publisher representatives that understand the statistical needs of their organizations relative to tracking user origins and using that information to assess the effectiveness of their metadata programs
- Writing and communication skills for preparing and communicating the proposed recommended practice.

Timeline:
- Appointment of working group: February 2016
- Approval of initial work plan: April 2016
- Completion of information gathering: July 2016
- Creation of prototype: December 2016
- Completion of initial draft Recommended Practice for public comment: February 2017
- Completion of final Recommended Practice for NISO publication: May 2017