Uniform Resource Names: Handles, PURLs, and Digital Object Identifiers

William Y. Arms

Identifying objects in digital libraries seems simple but proves to be surprisingly complex. Uniform Resource Locators (URLs) are fine for locating digital objects, but digital libraries need names that identify the actual content, not merely the location where it is stored, just as we know a colleague by his name, not by the number of his office.

An early articulation of this need was an Internet RFC by Sollins and Masinter in 1994 [5]. They called for a system of Uniform Resource Names (URNs) to identify objects by what they are. The RFC listed several criteria for URNs, including that they should be globally unique and they should persist for all time. The RFC described the need for a resolution system to hold huge numbers of URNs, and return information about the corresponding objects—typically the locations where copies are stored.

Also in 1994, the Corporation for National Research Initiatives (CNRI) began development of a system to store and resolve identifiers, known as Handles [1]. Handles can identify any type of resource and can resolve to many copies of an object. The system provides decentralized administration and is designed to manage huge numbers of Handles. Handles have one serious disadvantage. Effective use requires the user's Web browser to incorporate special software. CNRI provides this software, but digital libraries have been reluctant to require their users to install it. Therefore, most applications of Handles use proxy servers that do not support much of the power of the system.

To avoid special browser software, the Online Computer Library Center developed the Persistent URLs system (PURL) using entirely standard Web technology [3]. A PURL is a URL that provides indirect addressing of Web resources. Thus, it is possible to change the URL of a resource, and the corresponding data in the PURL server, without changing the PURL itself. Both Handles and PURLs have been used in a number of digital library applications.

Systems of identifiers need administrative structures to manage the name space, to provide guidelines for use and to ensure compliance. Digital Object Identifiers (DOIs) are used by publishers to identify online resources, notably journal articles. They are administered by the International Digital Object Identifier Foundation (IDF) and use the technology of the Handle system. Here is a typical DOI:

DOI: 10.1045/july2000-arms

The part before the slash is assigned by the IDF: "10" identifies it as a DOI and "1045" identifies the publisher as *D-Lib Magazine*. The part of the string after the slash is assigned locally, in this case by CNRI, the publisher of *D-Lib Magazine*.

In a series of papers in 1998 and 1999, the director of IDF, Norman Paskin, explored the fascinating question what should a DOI refer to [4]. For example, should the DOI for a journal article refer to the underlying work, a particular manifestation perhaps in print or in a digital format, or a specific copy? Cataloguers have long been sensitive to such distinctions. Their expertise has been embodied in the IFLA reference model [2], which provides guidelines for distinguishing between the variations of a work. In the interest of simplicity, Paskin eventually recommended that DOIs should refer to the work, not to a manifestation of that work. Thus, the printed version of an article and a digital version have the same DOI.

Identifiers are an area where the needs of libraries and publishing are not well supported by the commercial development of the Web. Around 1995, it seemed likely the major browsers would support some version of URN. Unfortunately, this opportunity was lost in the rivalry between Netscape and Microsoft. Handles, PURL, and DOIs are partial ways to fill this gap, but it is sad that nothing generally-purpose is supported by the major browsers. **C**

REFERENCES
1. Corporation for National Research Initiatives. The Handle System; www.handle.net.
2. International Federation of Library Associations. IFLA Study Group on the Functional Requirements for Bibliographic Records. Deutsche Bibliotek, Frankfurt-am-Main (1997); www.ifla.org/VII/s13/frbr/frbr.pdf.
3. Online Computer Library Center. Persistent URL (PURL); purl.oclc.org.
4. Paskin, N. DOI: Current status and outlook. *D-Lib Mag.* (May 1999); www.dlib.org/dlib/may99/04paskin.html.
5. Sollins, K., and Masinter, L. Functional Requirements for Uniform Resource Names. Request for Comments, 1737. Internet Engineering Task Force. (Dec. 1994); www.ietf.org/rfc/rfc1737.txt.

WILLIAM Y. ARMS (way@cs.cornell.edu) is a professor in the computer science department of Cornell University, Ithaca, NY.