
Excerpt From *The Seybold Report*, December 22, 2004, Vol 4., No. 18.

DOI: Slow But Steady Wins the Race?

Nothing is as depressing as an idea whose time has come but few appreciate.

This is the fate of the Digital Object Identifier, or DOI, a persistent form of ID for a chunk of digital data (as opposed to a url, which can change).

This chunk could be a manuscript (or piece of one), an image, a film, a musical score (or single sound), a block of programming code - anything that exists as a digital entity. The DOI standard is the brainchild of the non-profit International DOI Foundation (www.doi.org), which has been laboring for several years to expand the acceptance and use of the standard.

It's been a slow climb.

One reason for the lack of progress is that this seemingly simple concept has some very complicated consequences, which many publishers just don't seem to want to deal with (or can't figure out how to profit from).

The industry already has the ISBN standard for identifying publications, but this number is really little more than the equivalent of a stick-on lapel badge that says, "Hi, I'm _____" in a language (numbers) that everyone can read. A DOI, though, is meant not only as an ID tag, but also as a direct signpost to the metadata held within a digital entity. A click on a DOI in a publisher's Web site, then, could take you right into a profile of the document, making DOIs not just a search tool but a powerful way to aggregate and sift through vast seas of content. And because DOIs can identify small units of data, they're valuable for aggregating data and tracking and protecting copyrights.

The first two steps of a DOI experience are "identify" and "describe," with the third being "resolve," and here things can get quite complicated. Having located the data, and possibly identified its nature through its metadata, what do you do with it? Read it? Buy it? Query its publisher for reprint rights? Any commercial, computational or intellectual outcome is a possibility, but the most basic resolution is simply finding what you're after. DOI is not yet a Web standard, though, so for the time being, these resolutions have to be mediated through a special Web site built just for this purpose. Eventually, though, they could work just like a url, only better.

To make more complex resolutions possible, especially those in which commercial or intellectual rights have to be respected, the International DOI Foundation has had to compile a digital rights-definition dictionary through which transactions are mediated. This has become a massive and sprawling thing, encompassing and making coherent several diverse dictionaries already in use, such as those in the journal-publishing world, as well as dictionaries compiled by the movie industry's MPEG standards organization. This is perhaps the most underappreciated task in publishing and computing history.

For the mostly scientific, technical and medical journal publishers who use the system, it works very well. The foundation licenses DOI-registration agencies, which issue the DOI numbers. Though there are only 10 of these, they include some giants, including CrossRef in the journal-distribution market and the European Union publishing house, the Office for Official Publications of the European Communities, which assigns DOIs for the myriad documents churned out by the EU's bureaucracy and its diverse branches.

The numbering system itself is very simple, consisting of a prescribed prefix followed by any alphanumeric sequence a publisher likes. (The DOI syntax is a National Information Standards Organization standard, and it's hoped that it will made an International Standards Organization standard sometime this winter.) That following sequence could be a publisher's already existing ID scheme (ISBN numbers, for example), but typically they will be finer grained, allowing small bits of data to have their own ID numbers, so they can be re-used in various permutations and combinations or licensed or sold apart from any larger work of which they might once have been a part. It's a very powerful rights management and marketing tool, not to mention a key to organizing, managing, and re-using vast quantities of data.

Nevertheless, "only" 14 million DOIs have been issued to date, the large majority of which have been applied in the STM journal market. There are more than 600 "prefix holders," or publishing bodies that use their own distinguishing DOI prefix numbers. But this community is growing steadily and DOIs now generate some 5 million "resolutions" per month. Two of the newest registration agencies are the U.K.'s Nielsen BookData, a leading bibliographic service, and the American periodical distribution giant R.R. Bowker.