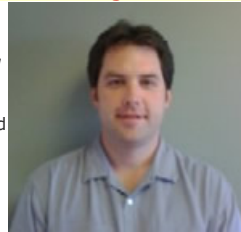


Evolution and revolution: pragmatism versus dogmatism

Ed Pentz
Executive Director

Publishers International Linking Association



The Internet and World Wide Web have already profoundly changed scholarly journal publishing, but we are only at the beginning of what will be an even deeper transformation over many years. Scientific publishing is now in a turbulent phase of innovation and experimentation, and diversity and flexibility will be critical if major improvements in both practices and economic models are to be achieved. The Public Library of Science (PLS) and PubMed Central (PMC) propose dogmatic solutions that confuse both economic and technological issues. Although these forces are related, understanding the central issues requires considering them separately.

I propose the CrossRef collaborative reference linking service as a superior and pragmatic alternative model to PLS and PMC. CrossRef is a collaborative, non-profit venture of 71 scholarly journal publishers, created last year specifically to integrate information sources across the Internet.

In contrast to centralized proposals, CrossRef has been designed to underpin the linking of information sources wherever they reside on the Internet. It was therefore conceived from the outset as a distributed system, and as a result does not straightjacket individual initiatives into complying with the restrictive conditions inherent in centralized systems. By federating diverse participants, CrossRef is also an excellent insurance policy against the dangers of especially intense in the fast-moving technological change of the Internet world of putting all of one's eggs in one basket, in this case the basket of PLS and PMC.

The essence of the CrossRef approach is that by developing and implementing consensus standards we deliver exactly the same interoperability and functionality as those aspired to by centralized archives. But the CrossRef model avoids the operational, financial and strategic limitations of central archives any organization can freely join the CrossRef system to take advantage of the added functionality that is gained by collaborating at a pre-competitive level, without any additional burdensome requirements being made on the ways they choose to run their existing operations.

CrossRef's membership is international and represents all areas of scholarly publishing; 60% of members are non-profit organizations. It was established to make broad-based reference linking efficient and scalable, and to be an inclusive, distributed (or minimally centralized), standards-based technical infrastructure that allows content providers to add value that will benefit scholars.

The PLS is demanding that publishers grant unrestricted free distribution rights to articles after six months and deposit these in a central repository. PMC, in an announcement made in this forum, relaxed its requirement for deposition, instead letting publishers stipulate that full text indexed by PMC may only be viewed at publishers' own sites. But PMC insists that articles must be available free within one year of publication.

The technological strategy issue here is a centralized versus a distributed approach, and the economic issue is whether or not to impose a particular business model. CrossRef favours a distributed approach, collecting only a minimal amount of metadata on a central server, while the abstracts and full text of articles remain at publishers' sites. This is a process sometimes referred to as distributed aggregation. Full text can be limited to subscribers of the journals or it can be available free to any reader. CrossRef takes advantage of the distributed nature of the Internet and is business-model neutral publishers, libraries and scientists can work out new economic models collaboratively.

Most journals are now available online and are moving away from being mere replications of their print versions to being true electronic documents that take full advantage of the online environment. Publishers are investing significant amounts of money to innovate, and competition plays an important role in stimulating this. There is pressure on publishers to enhance their systems continuously and match the features and functionality of other publishers. If full-text articles were available in one central government-run database, there would be little incentive for publishers to invest resources in their online systems and innovation would be hindered.

Another important issue is that of digitizing older content. With a centralized, free archive, there would be little incentive to digitize older articles. Marty Blume of the [American Physical Society](#) points out in his [contribution](#) to this debate that there are considerable costs in scanning all the APS content back to 1893 and that they need to recover these costs with a modest fee. The oldest articles for which CrossRef has metadata are from 1849 issues of the *Astronomical Journal*. The [Astrophysics Data System](#) scanned the full text with funding from NASA, but most journals will not enjoy this type of funding. Publishers have invested, and continue to invest, large resources in online systems that benefit scholars.

The key components of the CrossRef system are: unique persistent identifiers (CrossRef uses digital object identifiers (DOIs)), standardized metadata (XML), a metadata database and the CrossRef Reference Resolver, a DOI look-up service. Publishers deposit with CrossRef the metadata and a unique DOI associated with each article. As the first registration agency of the [International DOI Foundation](#), CrossRef deposits the DOIs and URLs of the articles in a DOI directory. Publishers (primary and secondary), agents, libraries and others, can then query the DOI database to add a DOI link to all the references at the end of an article, identifying its location on the Web. As each article has a unique identifier, it ends the famous "HTTP 404 Page not Found": if publishers move sites, and URLs change, they simply add the new URLs to the DOI database.

Scientists reading an article can as a result click seamlessly from a reference to the original article at the publisher's site. When a user clicks on a link, the DOI automatically directs them to the URL deposited by the publisher. In most cases a user arrives at the abstract page for the article where there are links to the full-text article. CrossRef does not collect abstracts or full text. Hence, each publisher must show users at least a full bibliographic citation and information on acquiring the article; the vast majority also show the abstract for free. Subscribers to the journal can go straight to the full text, whereas non-subscribers may be able to purchase the article or get information on a subscription. If the full text is available at no charge, then all users can get access to it directly.

References have always been a key part of scholarly articles and they are the means by which authors can point readers to material that provides the foundation and background for their work. Being able to link references so that readers can get to the cited article in one or two clicks is a major benefit for scholars. Some online features such as video or online commentaries are interesting but not vital to the article. Reference links are vital journals that do not have them are less valuable to the research community.

By taking a pragmatic, evolutionary approach, CrossRef has a system already in place that benefits scholars. With 71 publishers onboard, metadata for 3 million articles from 3,800 journals and 400,000 reference links a month being clicked, CrossRef is a reality. With the basic infrastructure of identifiers

and metadata in place, CrossRef will enable new technical advances and business models.

We believe that a problem with both the PLS and PMC initiatives is that they do not sufficiently address the real needs of scientists. The Web's potential can only be fully realized when scientists can obtain all the information they need in a convenient, comprehensive and efficient manner, through complete indexing, linking and searching. This should include all the relevant literature and information sources, not just journals. Any system needs to be sufficiently flexible to handle, and link in seamlessly, other sorts of publications, such as monographs, books, reference works, news and other media, as well as databases of grey literature such as research projects and biological databases.

To attempt to force all these sources into one, centralized, platform is not only naive but impossible, and defeats the very purpose and philosophy of the Web, which relies on diverse initiatives made interoperable by consensus standards across distributed platforms. Were such public centralized literature archives to happen, new initiatives would only come from government. The latter have generally a poorer track record in picking winners than the private sector where competition forces inferior technologies and services into extinction.

In the future, CrossRef will include more diverse types of content as outlined above. It will also introduce new metrics, for example on linking patterns, and automatic citation analysis. The CrossRef platform is also ideally suited to evolving towards a CrossSearch allowing one-stop-shop searching across the thousands of journals participating in the initiative. The next steps will depend on what CrossRef members and the research community judge to be the most valuable.

The PLS and PMC are increasing friction in scholarly publishing by taking a revolutionary and aggressive stance. The alternative is that publishers, scientists and librarians can work together to exploit technological advances, and invent new business models that decrease friction. We are in the midst of a very exciting time in scholarly journal publishing, and are really just at the beginning. In five to ten years, scholarly journals will be very different from how they now appear, and in retrospect it may be the evolutionary approach of CrossRef that will be recognized as having been revolutionary.

Ed Pentz is Executive Director of Publishers International Linking Association, Inc (PILA), the not-for-profit membership organization set up in January 2000 to run the CrossRef service. Previously Pentz was Electronic Business Development Manager at Academic Press and worked on IDEAL, the online journal system, linking strategy and DOI implementation. Before that he held editorial and electronic publishing positions at Academic Press and Harcourt in the United Kingdom. Pentz is on the Board of Directors of the International DOI Foundation and has also been Chair of the NISO DOI Syntax Committee.