Reference Accuracy
Best Practices for Making the Links

By **Carol Anne Meyer**, Maxwell Publishing Consultants

Reference citations are critical, providing measurability and discoverability. The problem is that reference quality and accuracy, as provided by the manuscript author, have been a known issue in publishing for many years.

Reference citations in scholarly publishing have never been more important than in the age of electronic distribution. Not only do references cite relevant previous work, they also provide one of the most valuable functions in electronic publishing—an actionable, clickable link to more information about the citation. The link may be to additional metadata in a secondary source or to the full text of the cited reference from a library collection or a publisher Web site.

Actionable reference links are just the tip of an interconnected scholarly communication system. Almost from the earliest days of electronic publishing, publishers have worked toward an interlinked body of knowledge and have envisioned links to gene sequences, chemical structures, multimedia, and Web site data in addition to the journal literature.[1] Increasingly, scholars are making their data available through databases such as GenBank. In some fields, journals are beginning to require authors to submit data sets to these data depositories. Links between the literature and the data will enhance the value of each.

The measurable impact of new publishing models, such as institutional repositories, will also increasingly depend on linking the content to other sources.[2] Studies of open access journals have compared the impact of journals accessible under traditional subscription models with those that are free to readers.[3] The measurements of impact frequently rely on citation analysis; thus, any problems with reference accuracy can underestimate the impact of an author's work or of a journal

ISI's impact factor was the original measure of citation analysis to evaluate journal quality. The impact factor and more recently developed citation measures, such as the *h*-index, provide input to a host of decisions—from library journal selection, to tenure, to appointments, to grants. Citation analysis remains a growing and important part of the environment,[4] available not just through Thomson's Web of Science and Citation Indexes, but also through comprehensive tools such as Scopus from Elsevier, Google Scholar,[5,6] and discipline-specific secondary services such as the Smithsonian/NASA Astrophysics Data System (ADS).[7]

Clearly, reference citations are critical, enabling measurability and discoverability. The problem, of course, is that reference quality and accuracy have been a known issue in journal publishing for many years. Even before the dawn of electronic distribution, periodic reference studies exposed sloppy reference work and exhorted authors and journal publishers to do better.[8]

Such studies continue in a number of fields and report reference errors in published articles that range from 4% to 48%, depending on the community and journal being studied.[9,10] As University of Chicago Press (UCP) Journals Division Editorial Manager Diane Berneath Lang puts it, "Difficulties in reference accuracy didn't come to light until everything became electronic."

WHAT READERS AND AUTHORS WANT

Studies of journal readers consistently show that reference linking is among the most highly valued features of electronic publications. An NRC Research Press reader survey reported that 59% wanted linking from citations to cited articles—second in rank only to backfile availability.[11] Likewise, studies of academic and family physicians have shown that reference linking is among the most highly valued electronic journal features.[12,13]

Authors, who, we must remember, are readers too, also appreciate the convenience of linking. As an example, the journal *Peptide Science* includes in its list of author benefits a variety of linking services—including reference linking to and from Chemical Abstracts Service (CAS) and Web of Science, as well as database linking to Cambridge X-Ray Group, The Genome Database (GDB), The Protein Data Bank (PDB), etc.[14]

Researchers are also increasingly relying on linking variations such as forward linking, or the ability to identify papers that have cited a particular article after it has been published, and multiple resolution links, which give the user the choice of linking targets for the same article.

VALUE ADDED

The debate about open access models acknowledges the value added by publishers in providing copyediting and reference linking. Indeed, open access publishers themselves are covering the costs of copyediting and reference linking.[15]

Publishers often berate themselves for not adequately communicating the value that they add to the scholarly communication process. A report on the recent Web forum on access to primary literature, sponsored by *Nature*, provides a succinct list of added value, which includes a number of steps, such as the importance of editing, copyediting, proofreading, reference checking, linking, and electronic display.[16]

Typically, journal author guidelines clearly state that the accuracy of the references is the responsibility of the author. As one example among many, the author instructions for the journal *Radiology* state, "It is the responsibility of the author(s) to ensure the accuracy of all references. This accuracy is essential for one of the features of Radiology Online: In the references of the online article, the hyperlinks to the abstracts of the referenced articles will not function unless the bibliographic information in the two matches exactly."[17]

Many journals have their own reference styles. Some use numbered styles, while some use alphabetical references. Some are based on standard reference styles, such as the American Medical Association (AMA) or American Psychological Association (APA), or other reference manuals and style guides. Most authors submit to journals with differing reference styles, and the tedium of changing from one style to another may not be the most pressing activity on a researcher's to-do list. And, as we shall see, even if the style is correct, references may contain inaccurate data.

A good copyeditor alone, or in combination with an automated reference extraction and checking process, can catch a majority of reference errors—resulting in higher link matching rates. As a result, the linked articles will see an increase in usage, and the metrics such as impact factor will increase for the target journals.

Ten years ago, Clifford Lynch, Director of the Coalition for Networked Information (CNI), speculated that journals not available electronically would become second-rate—not by their lack of quality, but by their lack of accessibility.[18] Similarly, references to resources that are available electronically may still be limited in accessibility if data errors prohibit seamless linking access.

THE IMPORTANCE OF COPYEDITING TO ACCURATE REFERENCE LINKING

In the pre-electronic world, faculty members would frequently send research assistants to the physical library to copy journal articles they discovered in an article's references. If the references contained major errors, retrieving the original article became difficult.

Reference errors still interfere with a researcher's ability to retrieve a work, but in a more visible way. Despite the near-universal policy of charging authors with reference accuracy, publishers spend a substantial amount of their production time in providing quality control for references. *Some publishers acknowledge the reality that authors might not be able to perform this task adequately. The American Geophysical Union (AGU) maintains that shifting the responsibility of reference accuracy from authors to publishers can allow authors to concentrate on scholarship, while publishers undertake the burden of creating accurate reference links.*[19] *In fact, the responsibility is likely to continue as a shared load.*

Petit Ferrer, Manager of Editorial Services for SPi's Publishing Division, reports that for one publisher customer, 16% of the time in copyediting and post-editing was spent on the references. Similarly, in a recent study of manuscripts at Blackwell Publishing, Wates and Campbell found that a high percentage of authors' changes in response to copyediting queries concerned reference clarifications.[20]

When we discuss citation reference linking, we are primarily concerned with links from the reference section of a scholarly article or book to the full text of the cited articles—whether in journals or proceedings. Many other links can be found in reference sections: links to secondary records (such as PubMed); links to books or theses; links to Web resources such as technical reports, author home pages, company Web sites, and so forth; and increasingly, links to data.

Citation references to full-text articles can typically be linked in several ways:

1) Links to full text at the publisher's (or aggregator's) own site;
2) Links to full text at another publisher's site, typically implemented with digital object identifiers (DOIs) managed through the CrossRef linking service;
3) Links to link resolvers or other OpenURL systems available from library systems vendors;
4) Links to bibliographic records, which may then lead to additional full-text services; and
5) Links to author preprints posted on individual home pages or collected in institutional repositories.

All of these types of linking rely on the accuracy of the reference data to provide the best results.

CROSSREF MATCHING ISSUES

One of the metrics tracked closely at CrossRef is the query matching rate, which measures the percentage of queries that find a matching DOI. CrossRef's overall matching rate has been steadily increasing over the years. For example, CrossRef Executive Director Ed Pentz reported that the match rate for 2006 was up 31% over 2005, to 34%.[21] According to data provided by Director of Technology Chuck Koscher, the matching rate for the period January through May 2007 was 43%.

According to Amy Brand, Director of Business Development, the query matching rate is dependent on a number of factors, including the content (breadth and depth) of material in the CrossRef database, CrossRef's matching technology, and, of course, data quality. Data quality issues can be caused by both errors in the metadata of target articles deposited by publishers and errors of the references submitted as part of the CrossRef query process.

CrossRef staff has made improvements in all of the components of the equation, although the data quality issues are somewhat beyond their control, as they accept deposits and queries prepared by a wide variety of publishers and agents, such as journal hosting platforms. Still, CrossRef has added staff to support metadata quality initiatives. The plan is to give more feedback to publishers on data issues so that they may either correct data or improve their processes for higher-quality submissions in the future. Brand says, "I still find it surprising and frustrating that there is no truly reliable source of metadata in our industry. There will eventually be more machine-driven techniques to help create what we call trustworthy information." This sentiment echoes publishers' long-standing calls for metadata repositories.[22]

Typical Data Problems

Data problems that cause CrossRef queries to fail to match include inaccuracies in the year, problems with author names, and variations in journal titles. Koscher estimates that 50% of all failed queries may be caused by problems in journal titles. Another type of problem that can affect either author names or titles includes variations in character representations. As an example, a dash in a manuscript can be represented by many variations: a short dash, an ASCII dash, an en-dash, or a dash from a particular font set. Alphanumeric page numbers may also cause problems. In preparing the references, an author might inadvertently leave off the alpha character. Even the best copyeditor wouldn't necessarily catch such a mistake without an automated reference validation process.

Koscher notes that CrossRef's matching technology is binary—either it finds a DOI or it doesn't. Although fuzzy matching has been added to the technology, he notes that machines don't deal as well with ambiguities as humans. One possible solution is to return multiple hits, where an editor can choose the appropriate answer. "The critical point," Koscher says, "is that once the article is out of human hands, the world is very rigid. The dash variations we discussed are insignificant to humans, but they will stop a machine that doesn't know about them. We can accommodate some problems by making technological improvements in the matching algorithms, but it's a lot like playing 'Whack-a-Mole'—several new problems pop up each time you fix one. Editorial processes are a far more effective way of dealing with the problem." In order to make a change to the technology, each problem, he explains, requires a time-consuming cycle of discovery, implementation, testing, and resolution. On the other hand, reference changes in an editorial process can be made instantly.

HIGHWIRE LINKING EXPERIENCE

A journal article or other scholarly content often resides in interlinked publisher, subject, or other aggregated collections. One such collection, HighWire Press, analyzes references in each article and inserts toll-free links to other HighWire-hosted content, according to Journal Manager Helen Atkins. Any unmatched references that pass through this first filter go through optional subsequent matching processes to create links to additional services determined by the publisher: CrossRef, bibliographic databases, and Web of Science. Optional OpenURL outbound links are being implemented, and additional link types are under development.

HighWire does not publish its matching rates, but its publishers report that where time and effort are spent proofing references, higher link rates result. Some HighWire publishers include DOIs, PubMed IDs, and other identifiers in their tagged references to increase reference linking rates.

PUBMED LINKING

Many electronic refereed articles in the biological and medical sciences are linked to PubMed, the US National Library of Medicine's (NLM) MEDLINE service. In addition, as we shall see, some publishers in the disciplines covered by PubMed also use its citations to validate references during production, or even as early in the process as peer review. PubMed provides a host of linking services that include full-text links to PubMed Central, publisher or aggregator sites, related articles, sequences in molecular biology databases, sections of textbooks, library holdings, and nonbibliographic resources. The combination of inbound and outbound links means that PubMed can act as a "hub through which you can link to-and-from the Web-based version of a journal, and much other data..."[23]

Seventy-five percent of the metadata in PubMed is created from data supplied by publishers, and the remainder is processed using scanning and optical character recognition (OCR).[24] Although a quality control process is in place to verify the accuracy of the records, any data collection activity of this magnitude will inevitably contain data errors. Despite this issue, PubMed citations are often used as a *de facto* authority file for bibliographic records in the fields of life sciences and medicine.

WORKFLOW OPTIONS

So how are publishers incorporating the important step of reference checking and validation into their workflows? We will look in detail at examples from several organizations.

Publisher Workflow

All publishers struggle with the problem of improving reference quality. Many of those we spoke with for this paper have instituted automated processes to supplement the critical hands-on copyediting step.

For example, the current process at the **University of Chicago Press (UCP)** is for accepted manuscripts to be converted to a structured format using Epic. Then, proprietary scripts clean them up and check the references against the CrossRef and PubMed databases for author, journal, and page number information. At this stage, the process does not retrieve the DOI, as UCP does not print the DOI in its journals. Echoing the problems CrossRef's Koscher identified, Lang notes that the automated checking process with the NLM and databases doesn't handle Greek letters, italics, or special characters very well, so matches of items that are present in the database are often missing. If the check against the databases results in a link, the copyeditor only edits the reference for conformance to the appropriate journal style.

Then, manuscript editors put references into the correct journal style. The tasks include renumbering references and adding missing citation information. The copyeditor looks at each reference to determine if there *is* a link and if there *should be* a link. He or she will be able to recognize odd citations like legal references or letters or transcripts. UCP is in the process of replacing proprietary tools with off-the-shelf packages and converting to an XML workflow.

Director of Publishing and Communications Karen Schools Colson and Managing Editor Alice O'Donnell from the **Association for Research in Vision and Ophthalmology (ARVO)** describe two distinct workflows—one for the electronic-only open access *Journal of Vision (JOV)*, and the other for the more traditional *Investigative Ophthalmology & Visual Science (IOVS)*. JOV authors are required to provide their own links, which must lead to freely available versions of the articles. A production editor randomly checks the links for accuracy. Ninety percent of the references are linked to full text or to PubMed, from which users can generally link to further sources.

IOVS authors provide traditional reference sections that are cleaned up in the copyediting process. ARVO is in the process of migrating to the NLM DTD for transferring files to HighWire, which then inserts links to PubMed, other HighWire-hosted articles, and CrossRef. Colson estimates that 95% or more references are linked.

Computer science publisher **Association for Computing Machinery (ACM)** has a slightly different twist on reference processing, according to Deputy Director of Publications Bernard Rous. They match vendor-supplied XML files—where the references are tagged as references but not yet parsed into reference components—with records from an internal metadata database using technology from a third-party source. The reference links that users see in the online version of an ACM journal article are actually dynamically generated from this database rather than displayed from the author-submitted references.

At **Nature Publishing Group**, XML-aware editorial software is used to clean up references and run queries against PubMed and CrossRef databases at the copyediting stage. A copyeditor puts unmatched references through an individual quality assurance step. *Nature*'s process builds reference sections from an internal RefMaster metadata database, which queries a range of citation databases in addition to PubMed and CrossRef. A prequery filter normalizes titles to increase the match rate, which is 44% according to Web Production Manager Amanda Ward.

Vendor Workflows

So what goes on inside the "black box" of vendor-supplied copyediting services, and how does the process effect reference linking accuracy? According to SPi's Ferrer, who manages the company's copyediting operations for journal publishing in the Philippines, the workflow often depends on the publishers' requirements. SPi's Philippine operation has about 100 copyeditors in its Manila and Dumaguete facilities. Another group of copyeditors that focuses on books is based in Pondicherry and Chennai, India. A smaller number of freelance editors, also based in Manila and India, who are working on web-based presubmission editing, supplement these teams.

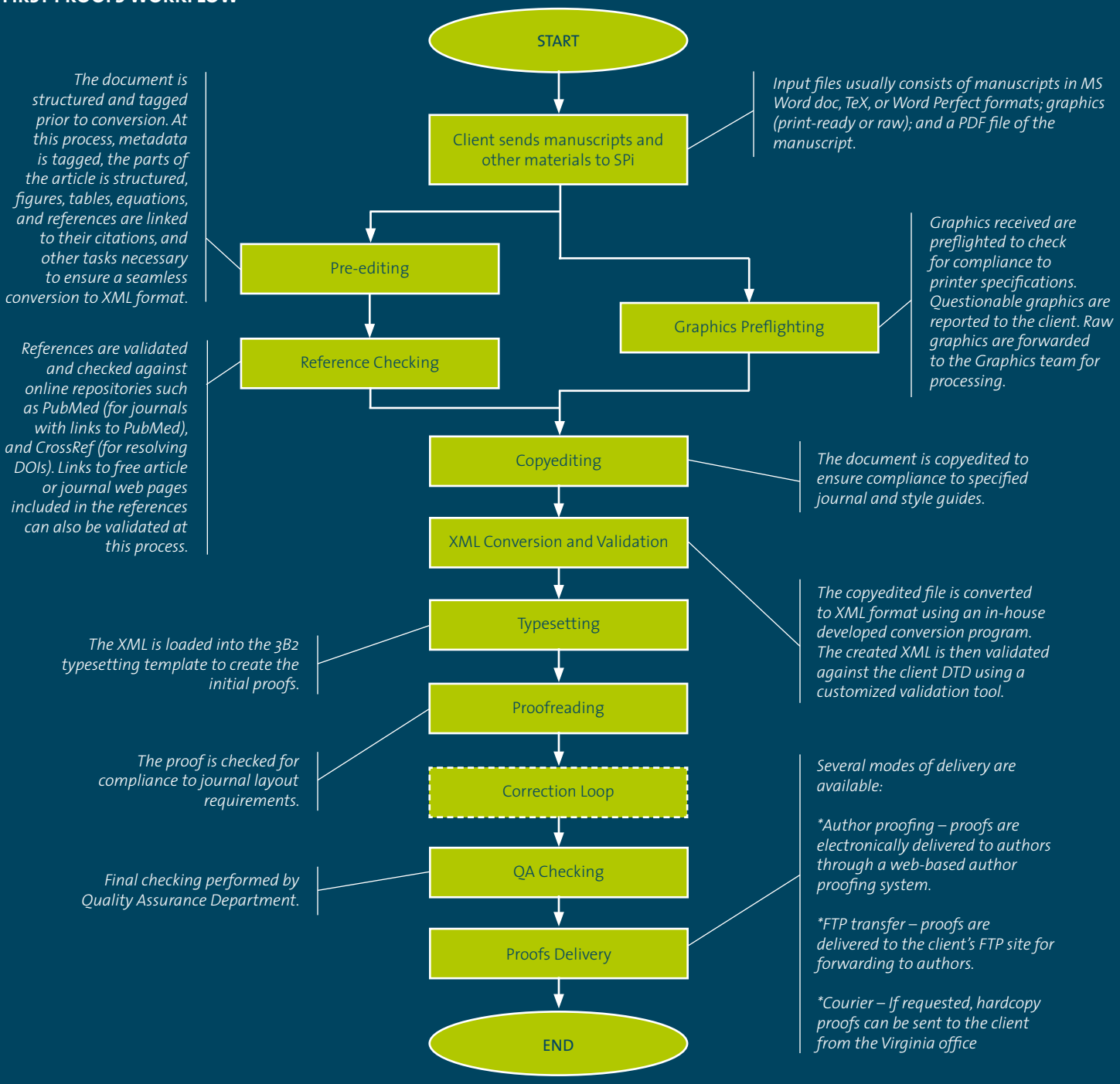
Figure 1 illustrates a typical customer workflow. The input is frequently a Microsoft Word file but can be whatever an author submits. The pre-editing or file structuring step results in a formatted and tagged manuscript, in preparation for copyediting and transformation to XML or SGML (as required by the publisher). At this stage, references are tagged to a granular level. SPiCE, SPi's proprietary Word-based editing tool, allows editors to interactively use macros and menus to tag and clean the document.

An optional step is reference validation against metadata databases, such as CrossRef or PubMed. References that don't link can be examined by editors more closely to identify any problems. Ferrer says, "Databases for reference validation add a useful step. It would speed up the production process if all the references were complete and accurate during pre-editing. We find that even though the author bears responsibility to ensure that references are correct, in production, there is still a lot of work to do, especially in terms of following the journal styles. References generate a lot of queries to authors. The earlier in the process the references are checked, the better it is. Typesetters shouldn't have to work on them. Authors prefer to concentrate on their research and not worry about the references, but they are also very important."

The next stage is for the tagged Word file to be copyedited. Because of the pre-editing and post-editing automation routines, copyeditors can concentrate on the core language editing tasks instead of spending a lot of time checking references. One of the reference tasks still undertaken by the copyeditor is checking on the completeness of the article title. In the post-editing stage, author queries are compiled into a document, and a manual inspection of the reference list takes place. Any incomplete references generate an author query.

A similar workflow takes place in book production, according to Rukhshad Banaji, SPi's Group Head of Editorial Services in Pondicherry, India, which provides end-to-end editorial and production services for STM and scholarly books. The

Figure 1:
FIRST PROOFS WORKFLOW



workflow may vary from publisher to publisher, and even from book to book, but typically, a final manuscript first goes to a project manager, who acts as the publisher's primary representative and who will assign it to editorial and graphics workflows. The editorial process, like the process for journal production in Manila, consists of both pre-editing and copyediting phases. The pre-editing phase includes checking references. Reference callouts are checked against the citation list to make sure every callout corresponds to a reference and that every reference is cited in the text. References are also reordered to conform to the required style for the project, which may include reordering surnames and first names, replacing names with initials, and punctuating properly. The pre-editing phase uses a combination of fully automated tools and those that rely on operator interaction. As with SPi's Journals team, SPiCE plays a major role in both the pre-editing phase and the subsequent copyediting of the manuscript.

After the pre-editing stage, when the copyeditor discovers any missing citations or citations with errors through a manual inspection, the editor uses additional resources such

as Internet searches, CrossRef, PubMed, or Google Scholar to verify missing reference information. The copyeditor resolves as many issues as possible within the time and budget constraints of the project. After copyediting, the manuscript goes through either an XML-native or a more conventional composition and quality control process. PDF proofs are sent to authors with the queries compiled in the editing process. Banaji estimates that as many as 50% of the queries involve corrections to references, noting that large reference books have many more references, and thus more reference problems than other kinds of books.

Final PDF and XML files that conform to the DTD specified by the publisher include granularly tagged references. Banaji notes that references in books tend to represent a wider variety of content types than those in journals. References to proceedings, reports, and theses are common, in addition to references to journal articles. Tagging these differently structured references adds complexity to the problem.

AUTOMATION TOOLS

A number of tools are used by publishers and copyediting vendors, ranging from proprietary programs to licensed software.[25] A few of the more prevalent options are discussed in this section.

CrossRef Simple Text Query

Available as a free tool for individual authors on the Web at http://www.crossref.org/freeTextQuery/, Simple Text Query is based on a customized version of eXtyles refXpress, which takes an unparsed text string, marks it up with XML, and passes it to the CrossRef system, which returns the DOI. Users paste in the references, which are capped at 10,000 characters. The DOIs that are returned are live hyperlinks. It works on the references regardless of their editorial style. At least one publisher has asked its authors to use Simple Text Query to incorporate DOIs in their submissions.[26]

Peer Review Systems

Both Manuscript Central and Editorial Manager have incorporated reference checking into pre-acceptance stages of manuscript processing. Aries Systems has partnered with Inera to incorporate its eXtyles technology into Editorial Manager.[27] Thomson, which now owns ScholarOne, has announced a pilot at Blackwell Publishing that integrates Web of Science with Manuscript Central to facilitate reference checking during the review process.[28]

Microsoft Word and Extensions

As we have seen, a number of publishers feel that copyeditors are more productive working in a familiar environment such as Word. Several tools, including Inera's **eXtyles** (http://www.inera.com/extylesinfo.shtml) and SPi's **SPiCE**, make the most of both automation and human judgment by creating macros and menus that will automate reference processing and mark references in a way that makes it easier for a person to identify the components that are present and accurate in the familiar Word environment.

XML Editors

The most widely used XML editors among scholarly publishers include:

- XMetal (http://na.justsystems.com/content.php?page=xmetal), and
- Arbortext's Epic (http://www.arbortext.com/html/epic_editor_overview.html)

Additional editors are available, including some open source options, and are discussed in a comprehensive review by van den Broek.[29]

Reference Processing

Automated tools to parse, tag, and match references can be built from text mining technologies. Some organizations build their own tools. A third-party example is Parity Computing's Reference Extractor technology, which automatically parses untagged references and normalizes them against an internal or public metadata database.[30]

This list is a sampling of tools. Editing services available from vendors like SPi, which specialize in electronic editing, incorporate a number of these and other tools in ensuring reference accuracy.

BEST PRACTICES

If reference data accuracy is key to providing a better experience for readers and for authors, what are the best practices for improving it?

After reviewing the literature and talking with a variety of publishers and service providers, we have found a number of common practices used to improve reference quality and reference linking rates. We note that the parties that have adopted these practices range from small-society publishers to huge commercial operations. Some publishers, whether small or large, have the technical resources to implement these practices in-house. Others have outsourced the solution to organizations whose technical competencies complement their own internal strengths.

Structure Documents

In order to accomplish automated transformations for every stage in production where references are important—from manuscript submission through acceptance, copyediting, proofing, and online and print publication and distribution—a structured workflow based on XML will allow the automation of workflow tasks. XML allows transformations among different applications of the data. Well-formed XML can drive author proofs; create aggregator metadata deposits;

be transformed for CrossRef deposits and reference queries; create print and online output; and finally provide files to archive services such as Portico.

In order to maximize the benefit for reference citation applications, references must be parsed and tagged in a detailed and granular fashion. This tagging may occur through automation, computer-assisted, or manual methods. An XML workflow also allows the use of XSLT transformations for creating additional outputs or to check quality.[31,32]

Check Reference Quality Earlier in the Process

Some organizations are incorporating reference validation and accuracy checks as early as the peer review stage. The ability of peer review systems to do reference checking can improve the quality of the peer review process, even for papers that may never be accepted for publication.[33] The process allows editors and reviewers to catch some of the problems of poor reference citation practices, as documented in the literature discussed previously.

CrossRef is also working with organizations to move reference matching into earlier parts of the production process. This change in workflow allows time to correct data errors that may be causing reference mismatches. According to Koscher, reference checking is much more productive in the editorial phase than as a postproduction step, "The further upstream references get linked, the more powerful the process."

Separate Routine Cleanup from the Art of Editing

A number of required editorial changes to manuscripts, including references, represent technical changes for style, font substitution, adequate abbreviations, or order. Regardless of whether the more routine steps are automated (see below), separating the routine (and sometimes mind-numbing) tasks from those requiring the judgment, discretion, and art of a skilled editor allows both greater accuracy and greater creativity for more complex problems.

Automate Style Changes

The most successful reference processing methods incorporate some type of automation, whether Word macros, third-party tools, or batch processes, to normalize references for style. The degree of automation can vary from a fully automated batch process to a computer-assisted process where the software flags anomalies for a human to resolve. Publishers must, of course, adopt a consistent reference style.

A style change process will typically perform the following functions:

- Identify the document type of the reference;
- Normalize punctuation;
- Replace journal names or abbreviations with standard journal title abbreviations;
- Mark missing or suspicious reference elements;
- Reorder references (alphabetical or numbered); and
- Make author names consistent and accurate.

Incorporate Authority Control

A number of publishers incorporate an automated matching process to metadata databases. Some go as far as replacing author-supplied references with records from internal metadata databases; others supplement reference data in the manuscript with data from matching records in internal databases such as PubMed or other MEDLINE variations, CrossRef, ADS, or other bibliographic citation resources.

Take Advantage of Standards

In choosing reference styles, DTD, and reference policies, publishers are well-advised to adhere to existing standards. A variety of metadata standards exist, and an exhaustive survey is beyond the scope of this paper but can be found elsewhere in the literature.[34] In addition, organizations such as the Council of Science Editors (CSE) publish policies on editorial practices that can be of great use. An example is a policy paper on citing references with group (sometimes called corporate) authors.[35]

Provide Training

Some publishers prefer to have copyeditors work in the native author-supplied file format (such as Word) that may or may not be enhanced with macros to help with reference editing, while others choose to adopt XML-based editing tools or typesetting software. Either choice can work well. The key to success is ultimately the editor's familiarity with the editing environment. The training investment is likely to be greater for XML tools.[36]

ACCURACY NOW LEADS TO MORE OPTIONS IN THE FUTURE

The world of scholarly communication has rapidly embraced linking technology for references, and it is poised for "richer mutual linking between journals and databases, and in a 10- or 15-year time frame... the rise of new kinds of publications that offer the best of both of these worlds," according to *Nature*'s Timo Hanney.[37]

The resurgence of interest in evaluation metrics for journals and other scholarly communication vehicles like institutional repositories makes the accuracy of linking citations more critical than ever before for authors, for journal publishers, and for the academic community.

ABOUT THE AUTHOR

Maxwell Publishing Consultants' Principal, Carol Anne Meyer, has 25 years of predominantly digital publishing experience in trade, scholarly, and professional markets. Prior to forming her own consulting business, Carol served as journals publisher at the Association for Computing Machinery (ACM) and Director of New Media at Little, Brown & Company. Maxwell Publishing Consultants provides marketing, research, and business strategy services to organizations involved in scholarly publishing. See www.maxpubcon.com.

ABOUT SPi

SPi's Publishing Division provides conversion, editorial, content production, and business process outsourcing solutions for journal, book, and database publishers, as well as content aggregators. A key capability is reference tagging to facilitate linking from references in a paper to the full text of that article on a publisher's Web site. To learn more about our reference tagging capabilities, drop us an email at bizdev@spi-bpo.com or visit us at www.spi-bpo.com.

REFERENCES

1 Hunter, Karen, "Adding Value by Adding Links," *Journal of Electronic Publishing*, March 1998, Vol. 3, Issue 3, http://www.press.umich.edu/jep/03-03/hunter.html.

2 Scholze, Frank, and Dobratz, Susanne, "International Workshop on Institutional Repositories and Enhanced and Alternative Metrics of Publication Impact, 20–21 February 2006, Humboldt University Berlin Report," *High Energy Physics Libraries Webzine*, Issue 13, October 2006, http://library.cern.ch/HEPLW/13/papers/2/.

3 McVeigh, Marie E., *Open Access Journals in the ISI Citation Databases: Analysis of Impact Factors and Citation Patterns: A Citation Study from Thomson Scientific*, Thomson Scientific, October 2004, http://scientific.thomson.com/media/presentrep/essayspdf/openaccesscitations2.pdf.

4 Garfield, Eugene, "The Agony and the Ecstasy—The History and Meaning of the Journal Impact Factor," International Congress on Peer Review and Biomedical Publication, Chicago, September 16, 2005, http://garfield.library.upenn.edu/papers/jifchicago2005.pdf.

5 Meho, Kokman I. and Yang, Kiduk, "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science vs. Scopus and Google Scholar," *Journal of the American Society for Information Science and Technology*, in press, http://www.slis.indiana.edu/faculty/meho/meho-yang-03.pdf (author preprint).

6 Bakkabalbasi, Nisa, et al., "Three Options for Citation Tracking: Google Scholar, Scopus, and Web of Science," *Biomedical Digital Libraries*, 2006, Vol. 3, Issue 7, http://dx.doi.org/doi:10.1186/1742-5581-3-7.

7 Accomazzi, Alberto, et al., "Creation and Use of Citations in the ADS," *Library and Information Services in Astronomy V (LISA V)*, Sandra Ricketts, Christina Birdie and Eva Isaksson, eds., Astronomical Society of the Pacific, in press, http://arxiv.org/abs/cs/0610011 (author preprint).

8 De Lacey, Gerald, Record Carol, and Wade, Jenny. "How Accurate are Quotations and References in Medical Journals?" *British Medical Journal*, 1985, Vol. 291, pp 884–886, http://www.pubmedcentral.nih.gov/picrender.fcgi?artid=1416756&blobtype=pdf.

9 Lukic, Ivan Kresimir, Lukic, Anita, Gluncic, Vicko, et. al, "Citation and Quotation Accuracy in Three Anatomy Journals," *Clinical Anatomy*, 2004, Vol. 17, Issue 7, pp 534-539, http://dx.doi.org/10.1002/ca.10255.

10 Siebers, Robert, and Holt, Shaun, "Accuracy of References in Five Leading Medical Journals," *The Lancet*, October 21, 2000, Vol. 356, Issue 9239, page 1445, http://dx.doi.org/10.1016/S0140-6736(05)74090-3.

11 Holmes, Aldyth, "Publishing Trends and Practices in the Scientific Community," *Canadian Journal of Communication*, 2004, Vol. 29, No. 3, http://www.cjc-online.ca/viewarticle.php?id=839&layout=html.

12 Torre, Dario, et al., "What Do Academic Primary Care Physicians Want in an Electronic Journal?" *Journal of General Internal Medicine*, March 2003, Vol. 18, Issue 3, 209–212, http://dx.doi.org/10.1046/j.1525-1497.2003.20529.x.

13 Torre, Dario, et al., "Family Physicians' Interests in Special Features of Electronic Publication," *Journal of the Medical Library Association*, July 2003, Vol. 91, Issue 3, pp 337–340, http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=164396&rendertype=abstract.

14 *Peptide Science* Web Page, http://www.americanpeptidesociety.org/publications/journal.asp.

15 "Hindawi Adds Ten New Titles to its Open Access Collection," News Release, February 2007, LibLicense, www.library.yale.edu/~llicense/ListArchives/0702/msg00031.html.

16 "Access to the Literature: The Debate Continues," *Nature Web Focus*, http://www.nature.com/nature/focus/accessdebate/35.html.

17 Radiology, Publication Information for Authors, Radiological Society of North America, http://www.rsna.org/publications/rad/pdf/pia.pdf.

18 "Networked Information: Finding What's Out There: Clifford A. Lynch Interview," *Educom Review*, 1997, Vol. 32, No. 6, http://www.educause.edu/apps/er/review/reviewArticles/32642.html.

19 Schwarzman, Alexander B., et al., "XML-Centric Workflow Offers Benefits to Scholarly Publishers," XML 2004 Conference and Exhibition Proceedings, November 15-19, 2004, p. 1, http://www.idealliance.org/proceedings/xml04/abstracts/paper71.html

20 Wates, Edward and Campbell, Robert, "Author's Version vs. Publisher's Version: An Analysis of the Copy-Editing Function," *Learned Publishing*, April 2007, Vol. 20, No. 2, pp 121-129, http://dx.doi.org/10.1087/174148507X185090.

21 Pentz, Ed, *CrossRef Newsletter*, March 6, 2007, http://crossref.org/01company/newsletter/newsletter030607.pdf.

22 Doyle, Mark, "Pragmatic Citing and Linking in Electronic Scholarly Publishing," *Learned Publishing*, January 2000, Vol. 13, No. 1, pp 5–14, http://alpsp.publisher.ingentaconnect.com/content/alpsp/lp/2000/00000013/00000001/art00002.

23 Smith, Kent A., and Sequeira, Ed, "Linking at the US National Library of Medicine," *Learned Publishing*, January 2001, Vol. 14, No. 1, pp 23-28, http://dx.doi.org/10.1087/09531510125100232.

24 "Keyboarding Ceases as a Data Creation Method for MEDLINE® Citations." *NLM Tech Bulletin*, July–August 2004, p 339, e5, http://www.nlm.nih.gov/pubs/techbull/ja04/ja04_keyboard.html.

25 "Editing Tools that Help to Streamline the Publishing Process," *Science Editor*, September-October 2004, Vol. 27, No. 5, p 155, http://www.councilscienceeditors.org/members/securedDocuments/v27n5p155.pdf.

26 Meyer, Carol Anne, "CrossRef Annual Meeting Summary," December 8, 2006, p 5, http://www.crossref.org/10meetings/2006_mtg_summary.pdf, available to CrossRef members.

27 "Aries Systems and Inera Collaborate to Reduce the Cost and Time of Manuscript Publication," *Aries Systems News Release*, November 17, 2004 http://www.editorialmanager.com/homepage/press-releases/200411.pdf.

28 "Thomson Scientific, ScholarOne and Blackwell Publishing Collaborate to Offer a Simple Manuscript Review Process," *Thomson Scientific News Release*, July 10, 2006 http://www.scholarone.com/06-07-10_press1.shtml.

29 van den Broek, Thijs, "Choosing an XML Editor," 2004, (updated 2005), http://ahds.ac.uk/creating/information-papers/xml-editors/.

30 Parity Computing Web site, http://www.paritycomputing.com/web/solutions/reference_processing_linking.html.

31 Piez, Wendy, "XSLT for Quality Checking in a Publication Workflow," XML Conference & Exposition 2003, http://www.idealliance.org/papers/dx_xml03/papers/

32 Schwartzman et al., p 8.

33 Robinson, Majied, "Beta, Collaboration and Workflow Tools: the STM Publishing Survival Kit," *EPS Insights*, February 8, 2006

34 Morgan, Cliff, "Metadata for STM Journal Publishers: A Review of the Current Scene," *Learned Publishing*, January 2004, Vol. 17, No. 1, pp 31-37, http://dx.doi.org/10.1087/095315104322710232.

35 *CSE Recommendations for Group-Author Articles in Scientific Journals and Bibliometric Databases* http://www.councilscienceeditors.org/editorial_policies/groupauthorarticles.cfm.

36 Schwarzman et al., p 9.

37 Hannay, Timo, *Towards 2020 Science*, Microsoft 2005, p 19, http://research.microsoft.com/towards2020science/downloads/T2020S_ReportA4.pdf.