

Embedded metadata add value to scientific publications

John W. Miescher
 Bizgraphic Co, Geneva, Switzerland; miescher@bizgraphic.ch

Introduction

Metadata are data about data¹ that help to describe the content or characteristics of an electronic or physical object. Scientific publishing depends on such data for searching, classifying, and bibliographic referencing. Information about title, subject, keywords, authors and publisher of a scientific paper helps readers to retrieve online sources that are relevant to their research. This article is about embedding such metadata directly in manuscripts to facilitate researchers' and science editors' work, to form their knowledge base, and to properly share and cite scholarly information with others.

Well-behaved documents are both search- and user-friendly. 'Search-friendly' refers to embedding useful metadata that facilitate online management of references in individual libraries or collection of documents. 'User-friendly' means that users can easily locate passages of interest.

Search-friendly documents

Search-friendliness, or machine-readability, is increasingly important in view of the global influence of digitization and open access in the changing publishing and archiving environment.² Most scholarly publications are becoming available on the Internet, which makes their processing and systematic archiving a real challenge. To organize a bulk of the Internet content, scholarly papers should be easily classifiable with little or no manual intervention, which requires properly embedding metadata. Explicit metadata facilitate the work of librarians, digital asset managers and non-expert users' because:

- sources are automatically classified and indexed for searching across a collection of documents
- journal submissions and publications are easier to locate and cite
- interdisciplinary networking and sharing of information is facilitated
- authors get more and better exposure
- it is no longer necessary to hunt for citation relevant metadata on the Internet, particularly beneficial for students who lack patience or the wherewithal to locate relevant repositories can easily find the documents they are looking for
- citations and bibliographic references can be generated off-line, a must for self-published articles, work-in-progress and editorial content.

Metadata standards

Dozens of metadata standards are currently available,⁴ each being linked to its own vocabulary.⁵ Unfortunately, none of the standards is universally applicable. A student seeks data to generate citations while experts search their collection of papers, employing certain technical criteria. Information about book publishers, image or painting

Ideally, they should generate a list and submit it along with their manuscripts. XMP sidecar files are probably the best option to ensure integrity of their metadata

- reviewers may suggest changes to the titles and descriptions in addition to factual adjustments
- editors and translators may include different data and add keywords for optimal searches through search engines
- publishers can adapt metadata and add specifics such as Creative Commons licenses, copyrights details, dates of submission and acceptance, and ISSN/DOI identifiers
- libraries⁶ and content providers, who gather metadata for their catalogues, should ensure that useful metadata accompany each document for automatic classification, indexing, and retrievability.

Adding metadata to PDF files

Adobe's XMP⁷ technology¹ is well suited for embedding metadata. This is the format implemented in PDF documents. It has placeholders for Dublin Core⁸ elements and other standard meta types such as Dicom⁹ for medical applications and IPTC, which is used by the international press community and professional photographers to secure their copyrights. It also allows to define proprietary sets with their own namespaces as well as an unlimited number of custom attribute/variable pairs which can be used to describe anything. To view, to edit and to export metadata, a suitable (free or low cost) software is required.³ To embed these in a PDF document, Acrobat[®] or another PDF tool that can import XMP files, is used.

User-friendly documents

Online scientific papers, monographs, manuals, and all other multi-page documents should be easy-to-read and easy-to-navigate on any reading device. Style and multimedia content of open-access documents should not depend on proprietary (paid) software. They should be searchable, have bookmarks (in applications that allow for it, such as PDF files in Acrobat or Adobe Reader), an interactive table of contents, and an interactive index, cross references, and links to external resources. Even copyrighted, password-protected, or encrypted material should allow users to print, copy portions of the text, and add bookmarks and comments. Unencrypted metadata should be available for easy cataloguing.

Bookmarks and interactive tables of contents require no extra work and no extra cost. Authors and graphic designers working on final layouts of the documents simply have to employ consistent styles for titles and subtitles. All professional layout programmes offer this facility. When exporting or saving documents as PDF files it is necessary to check the box *Create bookmarks under Options in the Save as ... dialogue*; under *InDesign*[®] (or similar), first create a table of contents under menu *Layout > Table of Contents* and check the appropriate boxes in the export dialogue.

Conclusion

Search- and user-friendly documents can facilitate scholarly communication. They can be beneficial for document producers and distributors. The purpose of this article is to raise awareness among all parties concerned and to

encourage them to embed meaningful metadata (including citation specifics) in all items they process and publish. If that is not possible due to some reasons, eg if authors submit online or librarians who are not allowed to modify a document, collective influence should be exercised to change the current publishing environment.

References

- 1 Metadata. Available online at <http://en.wikipedia.org/wiki/Metadata> (accessed 27 June 2014).
- 2 Gasparyan AX, Avayazyan L, Kitas GD. Open access: changing global science publishing. *Croatian Medical Journal* 2013;54(4):403-406. doi: 10.3325/cmj.2013.54.403
- 3 digi-libris software for collectors of ideas and things. Available online at <http://digi-libris.com/> (accessed 27 June 2014).
- 4 Metadata standards. Available online at http://en.wikipedia.org/wiki/Metadata_standards#Available_metadata_standards (accessed 27 June 2014).
- 5 Namespace. Available online at http://en.wikipedia.org/wiki/Namespaces#Naming_system (accessed 27 June 2014).
- 6 Enhancing the QUALITY and Transparency Of health Research. Available online at <http://www.equator-network.org/> (accessed 27 June 2014).
- 7 Sidecar file. Available online at http://en.wikipedia.org/wiki/Sidecar_file (accessed 27 June 2014).
- 8 Submit Guide. Available online at <https://cds.cern.ch/ftp/submit-guide/> (accessed 27 June 2014).
- 9 Digi-libris Reader – Citation Help (extract). Available online at http://digi-libris.com/html/citation_variables.html (accessed 27 June 2014).
- 10 Dublin Core[®] Metadata Initiative. Available online at <http://dublincore.org/> (accessed 27 June 2014).
- 11 Extensible Metadata Platform (XMP). Available online at <http://www.adobe.com/products/xmp/> (accessed 27 June 2014).
- 12 Why do we need well-behaved documents? <http://digi-libris.com/html/flowchart.html> (accessed 27 June 2014).

13th EASE General Assembly and Conference School of Medicine, Strasbourg, June 2016

"Scientific integrity: editors on the front line"

Misconduct in science is increasing and editors are often the first to detect this. They receive information from different sources, have to assess it, then liaise with the authors' institutions to make decisions (corrections or retractions). The conference will address the challenges faced by editors in this process.

Organising committee:
 Hervé Maisonneuve
 Liz Wager
 Evelyne Decullier
 Alison Clayton and
 Chris Sterken



Photo: Sébastien Hanssens - Office du tourisme de Strasbourg et sa région

Proposals for sessions should be sent to:
 Hervé Maisonneuve
 herve.maisonneuve@gmail.com