

Connecting Authors and Repositories Through SWORD

By incorporating SWORD support into an add-in for Microsoft Word, it is now possible for authors to deposit articles to Information Repositories directly from their word processor. Furthermore, in order to simplify and make the submission process as transparent and simple as possible, the SWORD related information can be incorporated into template files, so that all that is required from authors is to click a button.

Additionally, since templates can incorporate semantic information, articles can be validated against the template as part of the submission process, enabling authors to correct errors prior to submission, which should result in a higher level of metadata and compliance of the content submitted to repositories. Also, through the add-in, author metadata can be gathered in a largely automated fashion, reducing duplication in data entry and author aggravation.

Enabling Document Submissions and Archiving As Part of the Authoring Process

Thought the extensibility in Microsoft Word, a word processing application popular in many scientific and technical disciplines, it is possible to enable authors to upload and deposit documents as part of the authoring process to repositories which incorporate support for the SWORD protocol. The SWORD support is incorporated as part of the freely available Article Authoring Add-in for Word 2007.

In fact, the usefulness of the SWORD functionality is not limited just to deposits into institutional or subject repositories, but it extends to the submissions of articles to the intake phase of peer review and publishing systems, or to other online document processing and approval workflows.

The use of SWORD as the underlying protocol opens up the possibilities of interoperability with diverse systems across multiple operating systems and implementations. To simplify the author interaction, the SWORD related information can be incorporated into Word template files. Authors can download template files from repositories, journals, or their institution's web sites, and then press a button in the user interface presented by the add-in to upload their articles to the designated SWORD location from the template.

Validating Semantics and Metadata at the Source

In order to ensure a minimum level of metadata and content consistency, templates may incorporate some requirements, such as allowed and required sections (Introduction, Abstract, Methods, Conclusion, for example), minimum and maximum length for sections, keyword tagging, and the presence of author information. Documents can be checked for consistency against the template as part of the submission process. Non compliant documents can then be corrected by the author, before they are allowed to be uploaded into the repository.

Having to enter information about themselves and their collaborators on web based forms is a common source of irritation to authors. This manual process is also open to errors introduced while entering the metadata. In order to simplify this aspect of the submission process, and to promote more reliable metadata into repositories, the add-in enables author metadata to be gathered automatically and interactively from a popular email application's address book, Microsoft Outlook. Any new author data entered by the author is stored in the address book enabling re-use.

The extensibility in the XML based Office OpenXML format used by Microsoft Word 2007, formally referred to as ECMA-376 and ISO /IEC 29500, enables the semantic information and author metadata to be packaged with the articles and preserved through the submission process, enabling it to be harvested within repositories for indexing. Templates can also indicate that documents should be automatically converted to other formats, for example the National Library of Medicine's Article XML format, as part of the upload process.

Optimizing for New Workflows Focused on Digital Consumption and Archival

The intended result of incorporating SWORD support into the Article Authoring Add-in is that of providing greater integration between the authoring process and publishing and archiving workflows, while making the submission process simple for authors. Additionally, as the use of repositories increases and content consumption migrates away from print, the preservation of semantics and metadata as part of XML based documents becomes increasingly important both for providing alternative presentations of the content, through transformations, and for providing richer data for semantic analysis and search.