iConference 2014 Proceedings Documentation

Maxi Kindling, Heinz-Alexander Fütterer, Stephanie van de Sandt and Alexander Petrus
Berlin School of Library and Information Science, Humboldt-Universität zu Berlin

Abstract
This documentation describes the iConference 2014 proceedings compilation workflow, publication process, lessons learned, and links to the relevant materials such as templates and sourcecode.

Keywords: iConference, conference proceedings, workflow, publishing, EPUB, Mobipocket, pdf, research data, author guidelines, templates, metadata


License: CC-BY 3.0

Research Data: Petrus, Alexander (2014) Java Tool for DOI registration via CrossRef XML. https://github.com/allexpetrus/iConference2014; See urn:nbn:de:koeb:11-100215746 for all other supplementary materials such as Perl script, templates, and guidelines

Contact: maxi.kindling@ibi.hhu-berlin.de

0 Executive Summary
The iConference 2014 proceedings compilation workflow consisted of the following tasks:

- provision of templates and author guidelines,
- revision of the conference articles with regard to format and layout aspects,
- processing and format conversion of separate articles and the proceedings compilation into PDF/PDF/A, EPUB and Mobipocket format
- metadata creation and enrichment (persistent identifiers, citation, document properties)
- upload of all files and metadata to the IDEALS repository (batch upload with support of the IDEALS team)
- DOI registration via CrossRef
- workflow documentation

Lessons learned can be summarized as follows:

- Plan to have enough manpower for the final processing and upload tasks about at least 4 weeks prior to the conference.
- Put particular emphasis on the creation and communication of the template, with regard to format conversion.
- For the processing of documents it is most important that all participating team members use the same operating system and office software versions.
- Try to get as much feedback as possible concerning the functionality of the provided mobile formats.
- Establish a culture of research data sharing in LIS, but also plan more staff to manage this task.
- Be prepared for some authors being later than late in submitting the final versions and in answering mails, but also look forward to have many motivating contacts.
1 Introduction

The iConference 2014 “Breaking Down Walls: Culture, Context, Computing”\(^1\) took place at Humboldt-Universität zu Berlin (HU Berlin) and was organized by the Berlin School of Library and Information Science at HU Berlin (BLSIS)\(^2\) in cooperation with the Royal School of Library and Information Science at the University of Copenhagen (RSLIS)\(^3\).

In the following we describe the process of compiling and publishing the iConference 2014 proceedings\(^4\). The motivation for this documentation is to share the knowledge and lessons learned during this process with the succeeding editors of iConference’s proceedings, with colleagues that have similar tasks, and to preserve this knowledge for ourselves being actively involved in the electronic publishing and open access community.

1.1 Acknowledgements

The proceedings team consisted of the Proceedings Chair Maxi Kindling and her student assistants Heinz-Alexander Fütterer, Stephanie van de Sandt, and Alexander Petrus from BLSIS. The team was supported by Program Chair Elke Greifeneder and her student assistant Hazel Engelsmann from RSLIS. The proceedings are published by the iSchools on the IDEALS Open Access repository at the University of Urbana-Champaign in the US (Illinois Digital Environment for Access to Learning and Digital Scholarship) with the help of Sarah Shreeves and Howard Adrian Ding. The proceedings’ work is based on the previous iConference 2013 proceedings with thanks to Linda Schamber and her team (University of North Texas). The cover layout is from Violetta Sekulovic (BLSIS). Hazel Engelsmann and Toine Bogers (RSLIS) tested the 2014 template and gave useful recommendations.

1.2 Content

The iConference 2014 Proceedings is a compilation of 168 peer-reviewed Papers, Notes, Poster descriptions, Sessions for Interaction and Engagement (SIE) and Workshop descriptions. It consists of single articles published on IDEALS as well as a proceedings compilation that also includes an additional preface and the lists of conference chairs, program committee and reviewers of the iConference 2014.

The IDEALS repository offers statistics reports for collections and single items. Each iConference is presented in an IDEALS collection. Ten weeks after the iConference 2014 there are 11,059 downloads in total for the 2014 collection (https://www.ideals.illinois.edu/handle/2142/45869, 15.05.2014).

1.3 Student assistants profile

We identified six skills student assistants need to have in order to support the team:

- Office software LaTex (especially templates, reference lists, citations, figures, tables)
- Interest (and at least basic understanding) in electronic publishing technologies, especially:
  - XML creation and XSD validation
  - Format conversion with PDF creating software
  - Format conversion and usage experiences with formats for mobile reading devices
- Scripting with Perl (see below)
- Programming with Java (see below)
- An understanding of reference management with tools such as Zotero (see below)
- Soft skills such as patience, communication, team work and organizing skills

---

\(^1\) http://ischools.org/the-icconference/
\(^2\) http://www.ibi.hu-berlin.de/
\(^3\) http://vsa.ku.dk/english/
\(^4\) https://www.ideals.illinois.edu/handle/2142/45869
2 Workflow

2.1 Templates and author guidelines
We revised the 2013 proceedings workflow adding some points. The main modification was that we took
the step of processing the formats for mobile devices, EPUB and Mobipocket files, by ourselves. In 2013
this was done by the software company Newgen\textsuperscript{5}.

At the first step we discussed and revised the proceedings template as well as the author guidelines
from 2013. The guidelines summarize the formal framework for the authors who contribute to the
proceedings as well as supportive information. The 2014 template used the identical letter size and the same
page margins. At the top of each document we added a box for the document’s metadata, taking FLOS
ONE as an example.\textsuperscript{6} We also decided to use numbered headings. The text itself is formatted in a single
column at 10 pt font size. Used fonts are Latin Modern Roman and Latin Modern Sans.

We decided to again provide a LaTeX template for the benefit of LaTeX authors, even though all
but six authors used the Microsoft Word template. Unfortunately, we needed to rebuild LaTeX files for the
purpose of continuing page numbers with regard to the separate format conversions. Both templates and
the documentation were offered on the iConference 2014 website.\textsuperscript{7}

We recommend to check the template details in particular with regard to the (mobile) formats that are planned to be produced (e.g. difficulties with tables and figures as can be seen below). Furthermore it
should be seriously considered to offer only one template for one text processing software. LaTeX is actually
the preferred option, although this is not realistic (yet). The testing of software that can automatically
convert Office formats to LaTeX (e.g. writer2latex for Open Office and AbiWord for MS Office) might be
helpful.

2.2 Final submissions
After revision of their articles regarding the reviewers statements, the authors submitted the final versions
through the conference management software ConfTool.\textsuperscript{8} Using the template was compulsory as stated in
the author guidelines. For further communication with the authors, ConfTool provides all necessary contact
information. It also provided the articles’ metadata that had been submitted by the authors with their
proposals.

In case of unclear formatting or difficulties with the article’s structure we contacted the authors.
 Most formatting problems arose from captions, tables of figures, and tables of tables as well as not properly
applied heading styles, e.g. independently formatted “bold” headings without a style. Most of the documents
only needed little reformatting, but a few caused many e-mails. Certainly more effort on the communication
and advocacy of the template would be beneficial. We did not do any copy editing with very few exceptions
(missing letters in headings etc.). Authors were responsible for spell-checking and copy editing on their own
as well as for the accuracy of their articles. With this decision we followed the 2013 proceedings guidelines.
In addition, for a native German team, copy editing of all English articles e.g. with respect to grammar
would have been a large effort.

\textsuperscript{5} http://www.newgensoft.com/
\textsuperscript{6} e.g. http://www.plosone.org/
\textsuperscript{7} Please find all supplementary material such as templates via urn:nbn:de:kobv:11-100215746.
\textsuperscript{8} http://www.conftool.net/
2.3 Documents processing and metadata creation

The structure of the proceedings compilation resulted from the order of the conference tracks: Papers, Notes, Poster Descriptions, Workshop Descriptions and Sessions for Interaction and Engagement Descriptions. The order of the articles within the tracks is based on the submission identifiers assigned by ConfTool.

Documents were submitted in the file formats doc, docx or tex. All documents were closely examined and, as stated before, reformatted, and restructured if necessary. All documents were saved in docx format. LaTeX files were converted to docx format.

Metadata for all articles were extracted from the documents into a metadata spreadsheet using a Perl script. We added a Digital Object Identifier (DOI) for each article. The DOI range for the iSchools publications had been previously arranged. The editors task is to mint each DOI by the registration of metadata including a URI where the publication is available. Once the DOIs had been assigned, we were able to add the article’s citation in APA style into the metadata box. For this, a Perl script created a BibTeX file from the metadata spreadsheet. The BibTeX file was imported into the Zotero bibliography. The Zotero bibliography is available online.\(^9\)

The metadata spreadsheet was also useful for checking the presentations’ metadata in ConfTool. ConfTool automatically generates a conference schedule, but not all author and title information were consistent with the final submission of the articles. Several minor changes were necessary. In addition, the metadata was used to create document properties with the help of the tool Word Metadata Changer (see below).

Adding continuous page numbers to each article and to the article’s citations, and adding document properties (metadata) to the files were the pre-final steps in processing the documents after they were „camera ready”. This step required a complete compilation document. The challenge of continuous page numbers is that they require a stable version of the document – after this step, the text length of each page must not change anymore to have a citeable article version with stable page numbers.

2.4 Publication formats

We decided to publish the proceedings in PDF/A format as well as document formats for reading with mobile devices. We used EPUB (version 2.0), it being a free and open standard for which many converting and validation tools are available. The EPUB format is compatible with a wide range of devices. We decided to provide each separate document in epub format as well as the whole compilation. This also applies to Mobipocket format. While the EPUB standard is supported by most of the tablet and e-reader devices, it is not supported by Amazon’s very popular and commonly used Kindle-product line, which is why it is important to offer another format beside EPUB. We decided not to offer the PRC format as previously in 2013, because the converting software Calibre that we used does not offer this format. Mobipocket as a further development of the PRC format was used in an encrypted form (azw) by the Amazon Kindle. Though the Kindle uses kf8 now, it is still capable to read mobi files, which was one of the reasons we opted for this standard.

For the processing of documents we used Microsoft Word 2013 that worked well with single documents. In collaboration with Adobe Acrobat Professional XI the docx files were converted to PDF/A format for the purpose of long term preservation. By using the PDF/A-1b standard we ensured that all fonts, annotations, color management and figures were embedded and preserved.

\(^9\) https://www.zotero.org/
\(^10\) https://www.zotero.org/groups/iConference2014/bibliography/items
Irritatingly we encountered major performance problems (system and software crashes) while converting the compilation as a whole, both to PDF format as to PDF/A. It took us many hours of testing several settings on different operating systems e.g. MacOS and Windows, and several PDF conversion tools, but without success. The conversion worked fine with Microsoft Word 2010, but resulted in a completely different page numbering in the compilation due to format differences, which would have caused citation difficulties for individual articles. It was therefore not an option to switch to MS Word 2010, since we had already finalised all articles including the page numbers (already in the citations), using MS Word 2013. In the end it was literally impossible to convert the entire compilation to PDF/A. EPUB files and Mobipocket files were converted using the software Calibre.

2.5 Repository upload

For the batch upload to the IDEALS repository we prepared the metadata spreadsheet including DOIs and additional information e.g. filenames, handles, and copyright information. The majority of article files and the spreadsheet had been made available via Dropbox to the IDEALS team.

Metadata landing pages for each article were created, attaching the appropriately assigned files. The assignment was based on the IDEALS handle range (having a collection ID for the proceedings) plus the ConfTool ID of each article. The remaining files had to be uploaded manually. For the publication of the separate articles an embargo was set until 28 February 2014.

Finally all articles were compiled in one document adding a proceedings cover, an imprint page with a compilation DOI, ISSN, and ISBN (following 2013 proceedings), lists of chairs, program committee, and reviewers, a preface and the table of contents. Due to conversion difficulties (see above) to PDF/A format we created the table of contents automatically, but had to link the bookmarks manually. We converted the compilation into PDF, EPUB and Mobipocket format. The proceedings compilation files were the last files to be uploaded to IDEALS by ourselves.

The IDEALS handle identifier serves as a stable URI for the DOI referring. Once we had all Handle Identifiers we converted all required metadata into CrossRef XML by using a tool. The XML metadata that is needed for the DOI registration was uploaded through an API from CrossRef, the DOI service agency for the US. DOIs were resolvable after the iConference 2014.

All in all the main work started with the template and author guidelines from the mid of September 2013 and ended with this documentation by the end of April. A timeline for all activities is given below.

---

12 http://calibre-ebook.com/
2.6 Tools

For documentation purposes we used a spreadsheet with continuously updated information concerning each article in columns, e.g. „final submission available“, „ready formatted“ and „PDF/A generated“ etc. Dropbox was used for sharing all documents, and a Google Document for the documentation of todos and lessons learned, as well as writing this documentation.

2.7 Adobe Acrobat

The conversion from docx format to PDF/PDF/A format was done via Word 2013 and Adobe Acrobat XI. Both software suites are licensed by Humboldt-Universität, so this decision was obvious. Acrobat XI installs an Add-on, the Acrobat PDFMaker, to Word 2013, which enables a direct export to PDF format from an existing docx file. The desired output format PDF/A was created with the following settings, see Figures 1–4.
These settings created a single PDF/A-file for each article. Fonts were embedded, footnotes were properly linked, headings were in the PDF’s table of contents.

However, this did not work for the compilation. We wanted to create headings from each document’s title. All other settings were identical (see “Publication formats”).

2.8 Calibre

We tested several software options for converting docx into EPUB and Mobipocket format. We tried Pages, iBooks Author, KindleGen and Sigil, but for our needs Calibre\textsuperscript{13} was the best choice.

Calibre is a free and open source e-book library management application, which is able to convert e.g. the docx format or PDF format into several e-book formats. It was necessary to have a software that

\textsuperscript{13} http://calibre-ebook.com/
runs on every operation system because our team has differing machines. It was also important that it could convert a file instead of creating an e-book within the program. Moreover, we were looking for a software that works locally on a computer instead of an online tool, because for copyright reasons we would not have been allowed to upload our files. Calibre fulfilled all these points and was the tool with least post-editing.

We used Calibre Version 1.13. After uploading all the files we linked the cover of the proceedings to the first item and cloned it to the other items. In the bulk convert, we used the following settings and converted the files into “EPUB” and mobi file type “old” format:

- **Look and feel**: Front-size key: 12, 14, 16, 18, 20, 22, 24; embed referenced fonts; x insert blank lines between paragraphs; x smarten punctuation
- **Page setup**: output profile: iPad 3
- **Table of Contents**: x Force use of auto-generated Table of Contents; Level 1 TOC (XPath expression: //h:1); Level 2 TOC (XPath expression: //h:2); Level 3 TOC (XPath expression: //h:3)

![Figure 6: Settings in Calibre](image)

As both the template (see above) and docx-files are based on XML the structure of the text file was mostly well implemented. There were just a few problems which appeared during the converting process:

- Enumerated headlines would not appear in the Table of Contents (ToC). We had to change our template for the converting process to ensure that the ToC worked automatically.
- Some illustrations disappeared in the e-book. These were mostly defragmented graphs or figures that contained text fields. We had to make screenshots and replace them. Figures that were flown around by text had to stand alone. Figures that did not appear were cut out and pasted back in as a figure (some were formatted in a wrong way).
• Captions must not be written in a MS Word text field. Captions in text fields got lost in the e-book and re-appeared in arbitrary locations.
• Tables were a huge problem for the Mobipocket documents. They were too big and could not be represented in a reasonable way. All rows of the table vanished, which caused problems with our infobox. We were not able to solve this problem appropriately.

2.9 Perl script
A Perl script was used to extract the metadata from the documents automatically. A script by idiotkid converted docx to plain text files\textsuperscript{14}. Another script then extracted the metadata and exported both a spreadsheet and a bibtex file.\textsuperscript{15}

2.10 Word Metadata Changer
Word Metadata Changer is a program for reading/writing metadata from/into Microsoft Word files. It has a 10 day free use license which was enough for our purpose. We used Word Metadata Changer version 2.7.3, it is available under http://www.metadatachanger.com.

2.11 Java
The metadata extraction from the spreadsheet to create an xml file for the registration of DOIs via CrossRef was done by a Java tool. The xml validates an xsd from CrossRef.\textsuperscript{16} The tool’s code is available from https://github.com/alexpeterus/iConference2014.

3 Research data
As stated in the proceedings preface, the iConference 2014 promoted the provision of research data upon which the proceedings articles are based to promote re-use and transparency in LIS research. To all authors we offered to publish research data either in the IDEALS repository or another appropriate archive, assigning persistent identifiers e.g. DOI and to link the research data set(s) in the proceedings article.

Digital research data management and archiving has become a major research field for the scholarly community of Library and Information Science (LIS) especially in the last decade. Thus it was expected that LIS researchers not only developed a consciousness for the relevance of research data management, but according to the motto “practice what you preach” would share their research data.

In the end we linked just a single research data set to a conference note that was co-authored by the proceedings editor. Nevertheless it must be granted that during the conference we had heterogeneous feedback from conference participants. Some of them mentioned their research data in the article and partly provided a link. The majority agreed to raise attention to LIS research data.

In sum, a culture of research data sharing and archiving in the LIS is still in the early stages. We strongly recommend to consider research data linking and publication in the upcoming iConferences. Much more effort must be put into LIS research data sharing, and into building best practices in research and in teaching.\textsuperscript{17}

\textsuperscript{14} http://sourceforge.net/projects/docx2txt/
\textsuperscript{15} http://help.crossref.org/#tools
\textsuperscript{16} http://help.crossref.org/#deposit_scheme