

EPUB Use in Digital Libraries: Developing an Online EPUB Creator Application

Zeki Çelikbaş

Istanbul Technical University. 34469 Maslak, İstanbul, Turkey. Email: celikbas@itu.edu.tr

Abstract: Libraries have a unique role as a bridge between users and information containers. Currently, information containers usually consist of books, journals and electronic files. While the format of information containers changes over the years, the library's role stays the same. As digital libraries proliferate, the classical library has a new role as information producer. In this paper, we discuss the EPUB format, developed by the International Digital Publishing Forum (IDPF) as an open standard for e-books, and its use as a new information container format in libraries and digital libraries. Several different formats (e.g., .doc, .odf, .txt, .html, .DjVu) are used in digital libraries, but the EPUB format stands out as it has open file structure and is easy to produce and use. Because of its ease of use, mobile devices and e-book readers such as Kindle immediately started to support it. As EPUB is an open and free format, it enables publishers and digital libraries to publish license-free e-books. We review the EPUB format in this paper and show how to produce an EPUB file along with its use with digital library applications such as DSpace, EPrints and Greenstone.

Keywords: Digital libraries; eBooks; EPUB.

Introduction

This work is divided into two parts. First, we discuss the basic features of EPUB format for librarians on the basis of the literature and describe its structure. Second, we describe an EPUB application that we developed to create .epub files and use it in digital libraries and institutional repositories.

Why EPUB Format?

We have been working on two separate projects: Creating an institutional archive and developing an interlibrary loan (ILL) application. To accomplish both tasks, electronic documents should first be created. Currently, Adobe's portable document format (PDF) is used extensively as an e-book format. The EPUB document format as an open structure is being increasingly used because it has no license restrictions and is easy to use. We decided to use EPUB format for electronic documents to exchange theses and chapters of books among different Turkish institutions as we needed a quick and simple tool. The proliferation of and widespread use of electronic readers and the rise of EPUB format use played an important role in choosing the EPUB file format. Table 1 compares the two formats, PDF and EPUB.

Table 1. Comparison of PDF and EPUB

| Product Format | Filename extension | Image support | Interactivity support | Reflowable content | Open standard | Bookmarking | DRM support |
|----------------|--------------------|---------------|-----------------------|--------------------|---------------|-------------|-------------|
| PDF | .pdf | Yes | Yes | No | Yes | Yes | Yes |
| EPUB | .EPUB | Yes | No (Yes*) | Yes | Yes | No | Yes |

* Interactivity is supported in new version.

Source: <http://www.anypdftools.com/ebook-reading/pdf-vs-epub.html>

What is EPUB?

EPUB is a free and open standard for eBooks maintained by the International Digital Publishing Forum (IDPF) (EPUB 3 Overview, 2011). The EPUB file format was designed to be open to anyone to

use or to develop it and is reflowable (the text can be resized and rearranged to suit whatever display it is being read on).

The EPUB version 2.0.1 consists of three specifications:

- *Open Publication Structure* (OPS) contains the formatting of its contentX
- *Open Packaging Format* (OPF) describes the structure of the .epub file in XMLX uze
- *Open Container Format* (OCF) collects all files as a zip archive,

EPUB internally uses XHTML or DTBook (an XML standard provided by the DAISY Consortium) to represent the text and structure of the content of documents, and a subset of CSS to provide layout and formatting. XML is used to create the document manifest, table of contents, and EPUB metadata. Finally, the files are bundled in a zip file as a packaging format. For more information, see EPUB (2011) and EPUB 3 Overview (2011).

Table 2. A short list of EPUB viewers and software for EPUB files.

| Software | Platform | DRM formats supported | Notes |
|------------------------------|---------------------------------|-----------------------|------------------------------------------------------------------------------------------------------------------------|
| Adobe Digital Editions | Windows, Mac OS X | Adobe Content Server | Requires online activation. |
| Aldiko | Android | Adobe Content Server | Supports EPUB for Android phones. |
| Bluefire Reader | iOS | Adobe Content Server | Supports EPUB and PDF. |
| Bookworm | Web | ? | Free open source online EPUB reader. |
| Calibre | Windows, Mac OS X, GNU/Linux | None | More often used for library management, conversion, and transferring to devices than reading. |
| CoolReader | Windows, GNU/Linux, Android | None | XML/CSS based E-Book reader for desktops and handheld devices. Supported formats: FB2, TXT, RTF, TCR, HTML, EPUB, CHM. |
| EPUBReader | Firefox add-on | None | Enables reading EPUB-files from within Firefox. |
| FBReader | Windows, GNU/Linux, PDAs | ? | Incomplete EPUB support. |
| FBReaderJ | Android | ? | Open source. |
| Freda | Windows Mobile, Windows Phone 7 | None | |
| Google Books | Web | ? | Supports EPUB and PDF. |
| ICE Book Reader Professional | Windows | None | |

Advantages of Using EPUB

As can be seen from the structure of EPUB files, the EPUB format has an open structure, is easy to create and can be integrated into long term preservation plans. For long term preservation, the text, metadata and multimedia content need to be converted successfully (Fig. 1). EPUB files can easily be exported to other file formats.

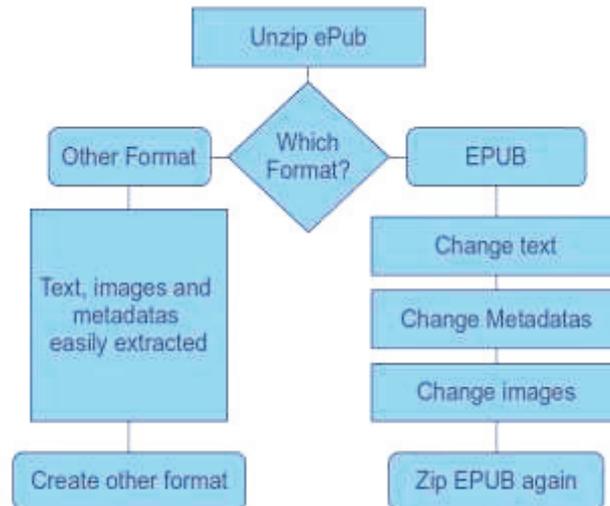


Figure 1. EPUB flow chart for long term preservation

How to Create an EPUB File

To create an EPUB file, one of the following methods can be used (Castro, 2011):

- Coding with any text editor;
- Using a special application (Calibre, Adobe InDesign, eCub, etc.); and
- Using conversion tools (ePub Tools, eScape - ePub Creator, etc.).

Aaron DeMott's tutorial on “How to Make an EPUB eBook by Hand” is a very informative document to understand how to create an EPUB file (DeMott, 2011). An EPUB file is created in three main steps:

1. Prepare the content;
2. Prepare the XML files (container.xml, book.opf, toc.ncx and all eBook content in one or more content.xml files); and
3. Put in the container.

The zip file layout should look something like this:

```

- mimetype
META-INF/
  - container.xml
OEBPS/
  images/
  - book.opf
  - toc.ncx
  - stylesheet.css
  - content.xhtml
  
```

EPUB Use in Libraries

Many libraries have started to lend e-books and e-book readers to their users. Most e-book readers support the EPUB format (see, for instance, http://wiki.mobilerread.com/wiki/EBook_Lending_Libraries; <http://sonysearch.overdrive.com/>; http://reed.ru/en_libraries_epub.php). All one needs to borrow an e-book from a library is an e-reader and a library card with an Adobe ID. Most libraries that lend e-books use the Adobe Digital Rights Management (DRM) protection system. This means that most dedicated e-readers (e.g., Sony,

Kobo, and so on) and several e-reader apps (e.g., Bluefire, OverDrive) can be used to read these DRM-protected library e-books (Reading, 2011).

Currently, the official EPUB standard does not include any specifications for DRM. However, most EPUB distributors at this time use the Adobe ADEPT DRM system, which means that EPUB files should be readable in most DRM-enabled devices (excluding Apple's iPad) (What is ePub, 2011).

Because of the XML standard, it is very easy to integrate DRM into EPUB. An EPUB file can optionally contain DRM, but it is not required by the specifications. Today, many books are sold with EPUB-integrated DRM support.

Adobe's DRM is a proprietary system, but Open Digital Rights Language (ODRL) as an open standard for rights management is on its way. The ODRL Initiative is an international effort aimed at developing and promoting an open standard for policy expressions. ODRL provides flexible and interoperable mechanisms to support transparent and innovative use of digital content in publishing, distribution and consumption of digital media across all sectors and communities (ODRL, 2011).

EPUB Application Development

There have been special applications for the interlibrary electronic document exchange for a long time. ILLiad, Ariel and OCLC's applications and even librarians' corporate email addresses have been used for document exchange. Librarians used various file formats for this purpose such as multipage TIFF, PDF, original office document files and image files. When we developed KITS, an ILL module used by many university libraries in Turkey, we had to tackle the issue of exchanging different ILL transaction files among libraries. We chose the EPUB open standard because of its simple structure and developed the Online EPUB Creator.

The Online EPUB Creator

This application has been created to produce a file from scanned images for librarians who will be exchanging electronic documents among different libraries.

The Online EPUB Creator runs on a web server. The short description of this application is that it produces an EPUB file from the images that are transferred by means of web forms. A working copy of the Online EPUB Creator was installed on a public server. Another version is still serving KITS members. Both applications have their own features. The following web addresses can be used to access the application: <http://epub.ankos.gen.tr/> and <http://kits.ankos.gen.tr/onlinepub>.

The files generated are held on the system for a certain period of time (24 hours to 15 days). At the end of this period, files are automatically deleted, although users who have the delete link activated can delete files any time they wish. This sounds a bit like a public file sharing system, but there are several public document creation services working in the same manner.

Application Infrastructure

The application we developed consists of the following components:

- Web server (Apache);
- PHP scripting language;
- Asbjorn Grandt's zip php class (Grandt, 2011); File download class (A php class sends the requested files to the clients and prevents unwanted downloads) (Ashraf, 2011); and
- EPUBCheck (to confirm compliance of generated EPUB with standards described by the International Digital Publishing Forum (IDPF)) (EPUBCheck, 2011).

An open source Java tool, called EPUBCheck, validates and detects errors in the structural markup (OPS, OPF, and OCF) as well as the XHTML and image files. The tool can be run from the command line, or used in webapps and applications as a library. Fig. 2 shows how this check can be made on any EPUB file. Thus, no structure or file errors are encountered in an EPUB file. This file can then be used as an e-book.

```

$ java -jar epubcheck-1.2.jar
uploads/49a5e86d5cde/ebook.epub

Epubcheck Version 1.2

No errors or warnings detected

```

Figure 2. EPUBCheck validation and error detection tool

Creating EPUB files using our application is very easy. All one needs to do is to enter images into the form elements in the web page and then press the "Create Epub" button. Figures 3 and 4 show the input and confirmation screens. The latter shows the URL link of the newly-created EPUB file along with the delete link.

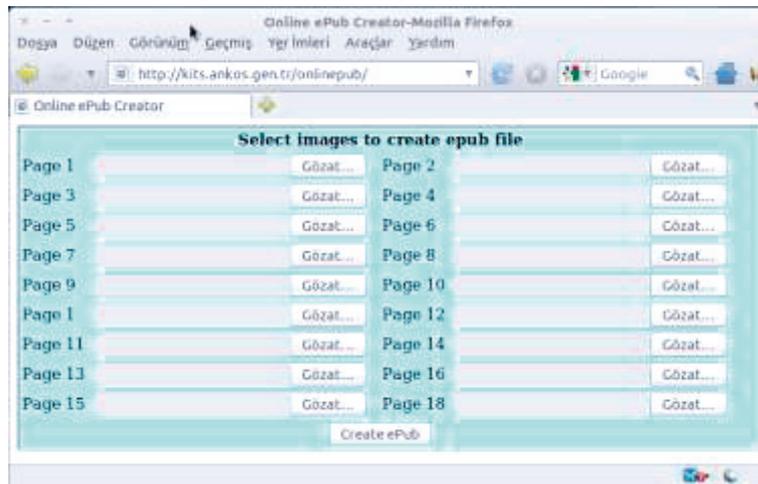


Figure 3. Online EPUB Creator application

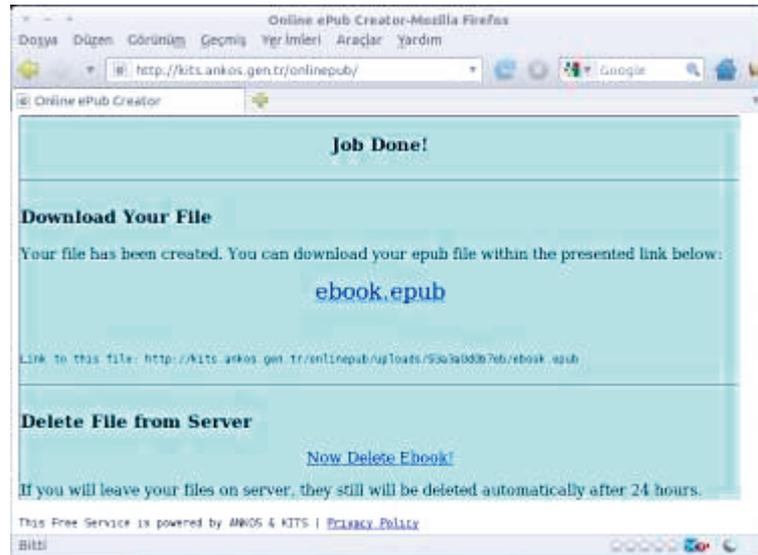


Figure 4. Online EPUB Creator's confirmation screen

All the processed files are located under a randomly created directory. The "ebook.epub" created with the name of the electronic document is also available. The OpenURL link for the distribution of this information is also created. At the bottom of Figure 4 a special link contains information on deleting this file. When this link is clicked, the created directory and all of its contents get deleted

permanently. A link to a license file containing terms of use of this service is also given. While this application is concentrated on creating public EPUB files, the ILL application (KITS) has more restrictions such as the number of times a specific EPUB file can be downloaded before it gets deleted. Restrictions can be defined by the creator. The next version of the system will include more features such as the ability to define the number of times image files can be replicated, requiring more complex Javascript libraries and AJAX capabilities.

EPUB and Digital Libraries

The most important factor in serving EPUB files with digital library applications is that of compatibility with librarianship standards. When selecting a digital library application to be used in an institution, the following three criteria should be considered:

1. EPUB support;
2. Compliance with standards; and
3. Ease of installation and usage.

Three digital library applications (DSpace, EPrints, and Greenstone) were tested for this purpose. As EPUB is quite a new standard, these three applications do not support EPUB files, although they are compatible with standards. All three are easy to use although DSpace and EPrints are harder to implement. Table 3 lists major characteristics of these digital applications.

Table 3. Characteristics of digital library applications

| | Bulk Data Input | Bulk Metadata Input | Browsing and Listing | Plug-in Support for EPUB Files |
|------------|-----------------|---------------------|----------------------|--------------------------------|
| DSpace | No | No | Yes | No |
| EPrints | No | No | Yes | No |
| Greenstone | Yes | Yes | Yes | No |

As Greenstone offers the opportunity to add bulk file and metadata input, we decided to use it in our project. None of them has plug-in support for EPUB files, although we plan to develop one for Greenstone. We configured Greenstone to process EPUB files using an unknown plug-in (Fig. 5). We added an EPUB mimetype to the Dublin Core format to run this plug-in.

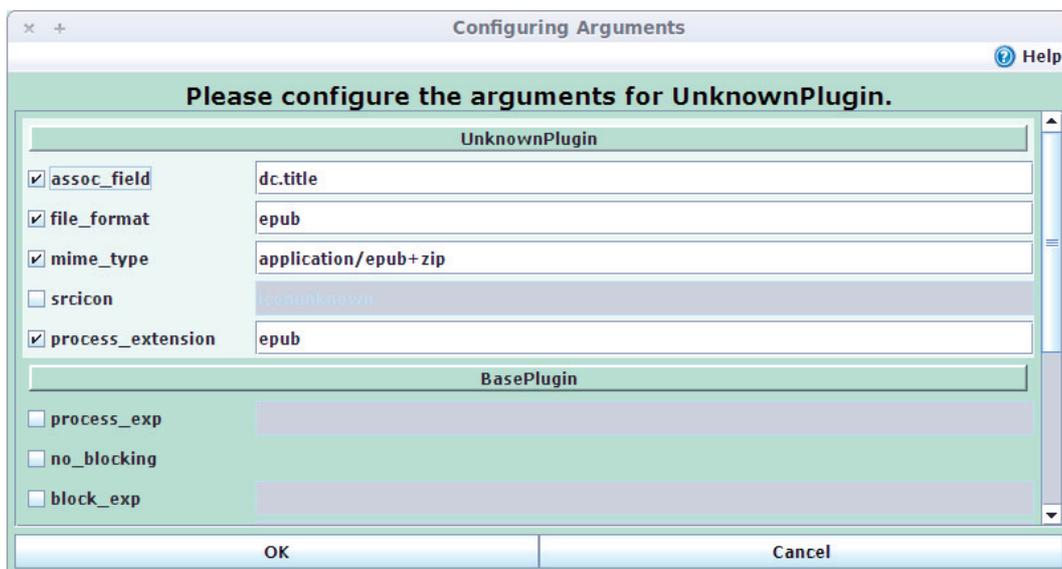


Figure 5. EPUB mimetype

Digital library applications do not yet have plug-in support to display EPUB files. EPUB will benefit from such third party applications. EPUB display applications were listed earlier in this paper in the "What is Epub?" section. We selected EPUBReader (<http://www.epubread.com/>) as our EPUB viewer. This application works as a Firefox browser plug-in, which means that the application we developed will be accessible through Firefox and offers platform-independent access to users. Needless to say, those wishing to see EPUB files using a different viewer can do so.

Conclusion

Although EPUB is not a new standard, it has not been fully supported by libraries and digital library applications for several reasons. Among them are:

- EPUB creation tools are insufficient;
- Adequate tools do not exist for Digital Rights Management; and
- Publications on EPUB are scarce.

Some libraries have started to buy e-book readers for their users, yet not all of them have EPUB support. Librarians must consider having this feature available in their e-readers.

The Open Digital Rights Language (ODRL) is an XML-based language for expressing, providing and agreeing to rights information of any kind of content and can be easily integrated into EPUB files, thereby increasing their potential use in the near future.

References

- Ashraf, Q. (2011). PHP File Download Class. *PHP Classes Repository - PHP Classes*. Retrieved, 4 April 2011, from <http://www.phpclasses.org/package/6687-PHP-Serve-files-for-download.html>.
- Castro, E. (2010). EPUB straight to the point: Creating ebooks for the Apple iPad and other ereaders. Berkeley, CA: Peachpit Press, 2011. Retrieved, 14 May 2011, from <http://www.ipdfconverter.com/blog/iphone-ipad/epub-straight-to-the-point-creating-ebooks-for-the-apple-ipad-and-other-ereaders-free-pdf-ebook-download.html>.
- DeMott, A. (2011). epub eBooks Tutorial. Retrieved, 4 April 2011, from <http://www.jedisaber.com/ebooks/tutorial.asp>.
- EPUB - *Wikipedia, the free encyclopedia*. Retrieved, 03 April 2011, from <http://en.wikipedia.org/wiki/EPUB>.
- EPUB 3 Overview. (2011, February 15). International Digital Publishing Forum | Trade and Standards Organization for the Digital Publishing Industry. Retrieved, 03 April 2011, from <http://idpf.org/epub/30/spec/epub30-overview.html>.
- EPUBCheck - Validation tool for EPUB - Google Project Hosting. (2011). *Google Code*. Retrieved, 4 April 2011, from <http://code.google.com/p/epubcheck>.
- Grandt, A (2011). Zip (archive, create zip file, zip class). *PHP Classes Repository - PHP Classes*. Retrieved, 4 April 2011, from <http://www.phpclasses.org/package/6110-PHP-Create-archives-of-compressed-files-in-ZIP-format.html>.
- ODRL Initiative. (2011). Retrieved, 4 April 2011, from <http://odrl.net/>.
- Reading Library eBooks on your iPad, Android, eReader and iPhone - epubBlog: News and information on the EPUB Standard. (2011). *Free EPUB eBooks for your iPad, Android, Kobo, Nook and Sony eReaders | Unleash Your Books*. Retrieved, 4 April 2011, from <http://www.epubbooks.com/blog/668/library-books-on-your-ereader>.
- What is ePub? (2011). epubBlog: News and information on the EPUB Standard." *Free EPUB eBooks for your iPad, Android, Kobo, Nook and Sony eReaders | Unleash Your Books*. Retrieved, 4 April 2011, from <http://www.epubbooks.com/blog/what-is-epub>.