The Open Document Format and its Impact on Accessibility for Persons with a Reading Impairment

Jan Engelen; Christophe Strobbe

Kath. Univ. Leuven, Research Group on Document Architectures
Kasteelpark Arenberg 10, B-3001 Leuven (Belgium)
e-mail: Jan.Engelen@esat.kuleuven.be; Christophe.Strobbe@esat.kuleuven.be

Abstract

It has become very common in the current information society to talk about "open" and to use this term as a quality mark. Open standards, open source software, open archives, open formats etc. are all very much promoted. In this contribution, we would like to focus on the file structure of documents such as texts, spreadsheets and presentations, and more specifically on the Open Document Format. ODF is becoming increasingly popular for many reasons, but it also is the first document format for use in office suites that has unique features built in (as of version ODF 1.1) for persons with a reading impairment such as low vision, blindness or dyslexia.

Keywords: open document format; accessibility; ODF; OOXML

1   What is a Document Format?

When we produce computer documents, be it a text document, a spreadsheet, a presentation etc., the result of our work has to be stored in computer files. Sometimes this is just one file (a Microsoft Office Word document, a slideshow presentation…), sometimes the contents of a document can only be restored from several files (a well known example are HTML webpages where text and images are stored in separate files).

When we look at the availability of documentation for document formats, it is possible to categorize them as "open" or "closed". Over the years, commercial software makers have come up with document formats that were tied to their own products. Examples of proprietary formats for text processing include Microsoft Word's binary format and WordPerfect's WP format. These formats often changed when new versions of the software products were released, and forced users to migrate to a newer version of the product. This upgrade cycle has to do with several things. On the one hand, a software maker may want to incorporate new features into the next version of their product, and this may necessitate changes to the format (commercial argument). On the other hand, users feel forced to buy the newer version of the product because they fear that they may no longer be able to exchange documents with users who have migrated to the newer version. Users may also feel frustrated because the changes to the document format and the document format itself are not well documented, so they cannot judge the impact of the changes.

At the other end of the spectrum are so-called "open" formats. These formats may have been created by a de facto or official standardization organization and through an open process, for example the HyperText Markup Language (HTML) of the World Wide Web Consortium (W3C). The availability of the W3C's Extensible Markup Language (XML), which is not a document format but a generic syntax for document formats and other data, has led to the creation of hundreds of formats, by standardization organizations as well as companies, individuals and online communities. With the advent of XML, and with the availability of many free or open-source XML parsers, creating, reading and implementing document formats came within the reach of many more users. Document formats passed through standardization or were based on standards, or for which documentation was available to everyone (free or for a fee) gained popularity for several reasons, one being that these characteristics appear to guarantee long-term readability and usability.

It would not be correct, however, to equate proprietary with "closed". The specification for Microsoft's Rich Text Format, for example, is available on Microsoft's web site[1]. Corel offers software development kits (SDKs) for the WordPerfect file format[2]. Portable Document Format (PDF) was created by Adobe Systems but its specification is publicly available, and subsets of the format have undergone or are undergoing standardization by the International Organization for Standardization (ISO). Open formats may also change, and may force users to get new versions of software products, just like changes to proprietary formats.

2   The Open Document format

The development of the Open Document Format (ODF) was initiated by developers of word processing and other software who wanted to make their product available in the public domain (open source software). In this particular case the major stimulus came from OpenOffice.org, an office suite that can be used freely by everyone. But as we will see later, there are nowadays many more products supporting this format.
The ODF format has the following characteristics:

- several XML files are produced to describe one document; minimally one has four XML files for any document (cf. table 1);
- the content description of these files is well documented and was the result of a public ("open") standardization procedure (more details about this later);
- the different files are usually (but not necessary) bundled together in one single ZIP file. Zipping files is nowadays a de facto standardized procedure for compressing and joining files and folders together. This action is always transparent for the user.

meta.xml	information about the document (author, time of last save, ...)
styles.xml	styles that are used in the document
content.xml	main document content (text, tables, graphical elements)
settings.xml	document and view settings (such as magnification level and selected printer); these are usually application specific

Table 1: The four basic building blocks of an ODF document

More details can be found in the relevant Wikipedia page[3] or in chapter 17 (Packages) of the complete ODF standard[4].

3   Why is this Format Important for Persons with Disabilities?

The ODF format is based on XML technology, which is promoted through the World Wide Web Consortium (W3C), and reuses formats whose accessibility has been verified through W3C's Web Access Initiative. Also when the ODF standard was developed, under the umbrella of the OASIS consortium, accessibility requirements were taken on board.

However, serious accessibility-related problems showed up when the US State of Massachusetts adopted the use of ODF as the only admissible interchange format for official documents in 2005. That decision has provoked a lot of criticism by groups that do not believe in open source solutions but also by organizations of handicapped persons fearing that they would be forced to use software with less accessibility provisions than their current, Microsoft-based, tools. That is why OASIS set up a special ODF accessibility subgroup in 2005.
Anyhow, there remains still a lot of confusion on the topic of accessibility to information. It is much more important for users with a visual or other impairment that the software they are using is accessible and usable then that the resulting document formats are accessible. In practice these files will never be read by human beings but only by machines. Despite this, the file format **is** important because it may or may not contain the data needed for an accessible reproduction.

At this point in time, general computer accessibility to Microsoft Windows-based software is quite good, especially for persons with a visual impairment. They can efficiently use their special hardware and a special computer program, called a *screenreader*, gives them access to the information and the commands on the computer screen. This is not because of Microsoft cared for this but because a whole group of external companies is building screen enlargement and screenreader software to be used on Windows platforms.
The software packages that support the ODF format currently are less accessible than the Microsoft office products although they are catching up rapidly. Promoters of Unix/Linux systems are especially convinced that this is only a matter of time because, for example, the Gnome Unix desktop is at the same time a so-called Recommended Engineering Accessibility framework[5]. These are frameworks that permit intimate interaction between general applications and accessibility software.

4   ODF Accessibility Guidelines

Current status
The ODF document format has, right from the beginning, been developed with accessibility in mind.
For this process one could rely heavily on the long term experience gathered around web accessibility via the WAI guidelines.
The details can be found in "Accessibility Guidelines for Implementations of Open Document Format v1.1. Draft 19, 14 March 2007" [6] and the major items are:

a) About the ODF format itself:
- Descriptive texts should be used for anything that is not text (graphs, pictures, sound inserts etc.). All necessary tagging is available.
- Tables and especially column and row headers must be marked up as such. This permits screenreaders to speak out table information together with the cell location information.
- A strict scheme of document divisions and corresponding headers should be maintained, using named stylesheets.
- There is a provision for logical description of navigation inside drawing layers.

b) About the software used for ODF production or conversion:
- The software must check the use of the accessibility features and stimulate authors to use them as much as possible.
- When converting a document in another format, all the accessible information must be kept and must remain available for further conversion, e.g. back into the original format.
- Users must be able to have the layout following general rules (e.g. on font size or color schemes) set up at the level of the operating system. Personal layout wishes (e.g. for persons with low vision) must always adhered to (stylesheet priority management).

Most of the above items have been incorporated into ODF 1.1

Future work
The ODF accessibility sub-committee has a number of work items planned for the next release of ODF. These are:
1. Background images. Access to any information contained in images used as backgrounds.
2. Navigation. The way in which people with disabilities can navigate round an individual slide in a presentation. Improving access to tabular data as is found in spreadsheets.
3. Multi-modalities. The provision of access in alternate modalities. For example, improving access to charts and graphs.
4. Reviewing ODF support for a wider range of disabilities.
5. More detailed support for spreadsheets. Easier access to header information, cell labels and formulas.

5   Accessibility Testing

At the CSUN 2007 conference, Jonathan Whiting and Aaron Anderson (WebAIM) gave a presentation on "Creating Accessible Content in OpenOffice.org"[7]. Another recent evolution is the development of software packages that audit the accessibility features of ODF documents. In 2006 IBM and the U.S. Department of Education organized a contest to produce such testing software. The winning solution ("ODF accessibility validation tool'") was developed by two American students (from Capitol College and Oklahoma University) and a Chinese student from Tsinghua University (Bejing). This was also announced at the well-known Technology & Persons with Disabilities Conference (CSUN 2007). The winning application and several others are given to the open source community via Sourceforge.org[8]. An online Open Document Format (ODF) Accessibility Evaluator is also available on the website of the Illinois Center for Information Technology Accessibility[9].

6   Why is ODF Readily Accepted by so Many Authorities and Companies?

One of the major goals of the Open Document Format is to guarantee access to content on very long time scales and this without technical legal barriers. In other words, efficient archiving with guaranteed future retrieval possibilities and the wish to become independent of Microsoft's business strategy are among the main drivers of adoption.

The fact that the early adopters are mainly public authorities has definitely increased the visibility of accessibility aspects in ODF as these authorities nowadays often have the legal obligation to consider the needs of all the citizens.

It is expected that ODF will slowly gain momentum mainly through acceptance by authorities.

The Massachusetts case had shown the weak point: very few software packages that natively use ODF were available in late 2005. And the incident also lead to the creation of the OASIS subgroup on Accessibility.

One of the early adopters is the Belgian government that has decreed that only open formats are acceptable as an exchange format between the Belgian authorities, and this from 2008 onwards. If realized in time, Belgium would be the first country to prohibit the use of closed document formats. As could be expected, Microsoft has reacted strongly.

7   How Popular is ODF in Reality?

By the end of 2006 there were eleven word processors, six spreadsheet programs and 5 presentation managers (Powerpoint-like programs) with support for ODF available. Furthermore, three groups are active in the development of conversions from Microsoft Word into the ODF format and vice versa. They are SUN Microsystems, the Open Document Foundation and the public domain Sourceforge.net project, "ODF Add-in for Microsoft Word"[10]. Within the UK Royal National Institute of the Blind, a project has been set up to turn ODF documents into the Daisy format, the new, worldwide accepted standard for talking books and multimedia documents. The ODF format is also used in the online text processing facilities of **docs.google.com**. Documents produced online can be stored in Microsoft Word, Microsoft RTF, and OpenDocument formats. As it is possible to upload and to download the online documents, the **docs.google.com** facility can in fact be used for file format changes too: upload in one format, download in another. PDF can be used as output, not as input.

8   Standardization

The ODF format v1.0 became an international ISO standard in 2006. After having been developed within a working group of the OASIS foundation, it was passed through the International Organization for Standardization, ISO, where it became ISO standard ISO/IEC 26300. The proponents of ODF have created several organizations for discussion and exchange of information. Two of them are very well known:

Figure 1: Logo of the ODF Alliance[11]

Figure 2: Logo of Opendocument-xml.org[12]

Meanwhile Microsoft has launched a counterattack by creating and promoting XML version of its proprietary office formats, and called them Office Open XML format (OOXML)[13]. This format will be used in Microsoft Office 2007.

In the beginning of 2007, this lead to a very controversial issue on the standardization of XML-based open document standards. The Open Document Standard became ISO 26300 (700 pages) through a very formal process typical for ISO work. Microsoft's alternative, OOXML (Office Open XML) was produced in one year by a technical committee[14] chaired by two Microsoft persons, contained many references to specific behavior of Microsoft software that were not documented and counted 6000 pages. It was ratified as ECMA-376 by ECMA International[15], which was consequently described as "a private association that drafts standards on demand"[16].

In late 2006, Microsoft wanted to put the OOXML specification on a fast ratification track within ISO because the document had already been ratified by ECMA. This created a lot of dismay within organizations that had been developing and promoting ODF. People were even asked to lobby with national ISO delegates to cancel the fast-track procedure. Websites listed arguments against the fast-track process, for example "EOOXML objections" on the Grokdoc website[17].

In spite of this, it was recently (April 2, 2007) announced[18] that ISO-Joint Technical Committee 1 has started the voting period for ISO/IEC standard DIS 29500…

9   And now?

ODF still seems to attract quite a lot of organizations. It is public domain and vendor independent, it is well defined (and not too complex) and it is accessible. However, what is even more important for reading impaired users is the fact that the software producing open office documents is made accessible. The existence of different types of converter plug-ins has taken away the major objection against the use of ODF as people can, for example, stay with the more traditional Windows-based platforms.

Acknowledgment

The authors wish to acknowledge the support of Kris Van Hees (PhD student at K.U.Leuven) and Peter Korn (SUN Microsystems) during the compilation of this contribution.

Notes and References

[1]   Rich Text Format (RTF) Specification 1.6: http://msdn2.microsoft.com/en-us/library/aa140277(office.10).aspx.
Word 2003: Rich Text Format (RTF) Specification:
http://www.microsoft.com/downloads/details.aspx?familyid=AC57DE32-17F0-4B46-9E4E-467EF9BC5540&displaylang=en.

[2]   http://apps.corel.com/partners_developers/csp/wordperfect_fileformatsdk.htm

[3]   http://en.wikipedia.org/wiki/OpenDocument

[4]   ODF standards:
The ISO version:
http://www.iso.org/iso/en/CatalogueDetailPage.CatalogueDetail?CSNUMBER=43485&scopelist=PROGRAMME
The freely available OASIS version can be downloaded from:
http://www.oasis-open.org/committees/download.php/19274/OpenDocument-v1.0ed2-cs1.pdf
(OpenDocument 1.0, Second Edition ; 722 pages) or
http://docs.oasis-open.org/office/v1.1/OS/OpenDocument-v1.1.pdf (OpenDocument 1.1).

[5]   Other examples are the Java Accessibility API, the Apple Accessibility API (with limited possibilities) and WAI ARIA for web based applications.
Info from Korn & Schwerdtfeger: *Accessiblity guidelines for ODF* (cf. ref. #7 below)

[6]   KORN, P.; SCHWERDTFEGER, R. eds. *Accessibility Guidelines for Implementations of Open Document Format v1.1.* Draft 19, 14 March 2007. http://www.oasis-open.org/committees/download.php/22957/ODF_Accessibility_Guidelines_19_14Mar2007.odt
(only available in Open Document Format).

[7]   http://webaim.org/presentations/2007/CSUN/ooo.htm

[8]   Meanwhile there are already a number of ODF validator projects on Sourceforge.net. On April 8, 2007 we found: ODF_Accessibility_Tester, ValidODF, Accessibi Add-on component of OpenOffice, ODF Accessibility Validator, Check the accessibility of ODF file, **ODF accessibility validation tool**, odfav, av4odf, Validator for ODF Accessibility, SalihiODF and ODFable.
More details by filling in their name in the search project page of www.sourceforge.net.

[9]   http://odf.cita.uiuc.edu/

[10]   http://sourceforge.net/project/showfiles.php?group_id=169337

[11]   http://www.odfalliance.org/

[12]   http://opendocument.xml.org

[13]   http://en.wikipedia.org/wiki/Office_Open_XML

[14]   ECMA International TC45: http://www.ecma-international.org/memento/TC45.htm

[15]   Formerly known as European Computer Manufacturers Association; http://www.ecma-international.org/.

[16]   http://press.ffii.org/Press_releases/FFII_opposes_Fasttrack_adoption_of_Microsoft_OOXML_format_as_ISO_standard

[17]   http://www.grokdoc.net/index.php/EOOXML_objections

[18]   http://www.ecma-international.org/news/PressReleases/PR_TC45_April2007.htm