Standards for Electronic Documents

APDIP e-Note 18 / 2007

Inside

APDIP e-Notes present an analytical overview of specific issues related to information and communication technologies for sustainable human development in the Asia-Pacific region. APDIP e-Notes are developed by the United Nations Development Programme's Asia-Pacific Development Information Programme (UNDP-APDIP) based at the UNDP Regional Centre in Bangkok, Thailand. For more information, visit Hhttp://www.apdip.netH or contact Hinfo@apdip.netH

Summary

The office productivity software industry has had a tumultuous history in the last two decades, much due to harsh competitions. In order to 'lock' users to their software by making it difficult for end users to easily read, edit and save their documents in other office productivity software, corporations have developed electronic document formats that are closed, proprietary and lacked adequate documentation.

In response to this scenario, the OpenDocument Format for Office Applications (ODF) open standard emerged to offer users full control of their documents. Users can change their office productivity software and their documents will still retain their full fidelity. ODF is an important milestone and has achieved widespread use especially in the public sector. Other open standards discussed in this APDIP e-Note include the Microsoft-released Office Open XML and Adobe Systems' Portable Document Format.

In summary, this APDIP e-Note provides a brief introduction to the history of document standards, explores the different standards for electronic documents, and details the development of ODF. It also looks at how governments worldwide have started to adopt ODF in public administration.

Introduction

Office productivity software is extensively used to create electronic documents, spreadsheets and presentation files. These documents are widely shared within and across government agencies, commercial industries, educational institutions, and across countries, cultures and time zones. With millions of users of office productivity software, computer literacy is now equated by many to literacy in operating a word processor, spreadsheet or presentation application.

The office productivity software industry has had a tumultuous history in the last two decades, much due to harsh competitions. As a result, corporations have attempted to 'lock' users to their software, making it difficult for end users to easily read, edit and save their files in other office productivity software.

One method that corporations use to lock users on to their platform is to make it difficult for existing users of their software to migrate their documents to a different application. This can be achieved by using closed, proprietary and undocumented office document formats structured in undecipherable binary form that competitors will find difficult, if not impossible, to 'reverse engineer'.

As every suite of office software use its own closed standards, it becomes impossible for users to easily migrate their documents to other applications developed by different corporations. Even if software developers try to reverse engineer, this process almost never ensured perfect conversion. Thus, converted documents could never achieve 100 percent document fidelity.

Another method of locking users is to continuously develop newer upgraded versions of the software applications. By having document formats and software applications that cannot interoperate smoothly between different versions of the same software, and sometimes by refusing to recognize older versions, users are often forced to upgrade their office productivity software. Entire organizations have been known to allocate resources to upgrade their office productivity software at the expense of upgrading individual computers and other hardware on a need-be basis.

Necessity is the Mother of All Inventions

The problems with closed document formats are not just affecting software developers and competing companies; they are affecting end users as well. The computer industry is hardly three decades old and it is already facing problems of electronic archeology:

documents created by end users 10 years ago or less cannot be opened with 100 percent fidelity in modern office software.

The closed nature of the document formats also means that the technical strength of the formats is comparatively inferior due to lack of peer review. Consequently, documents have become the vehicle for malicious activities via viruses and worms. These problems affect not just casual computer users, but users in government organizations, trade industries and educational institutions.

Necessity is the mother of all inventions, and this is also the case with the need for a truly open and unencumbered specification for office documents. Open standards help to solve many of the problems noted above.

Open standards are standards that are developed and maintained in a transparent, collaborative process, which is open for participation to all interested parties. A truly open standard is also free of patent or licensing restrictions. This is important for two reasons. Firstly, it ensures a level playing field for competitors and thus, promotes innovation and choice. Secondly, it ensures that users are able to use the standard without worrying about being sued for infringing patents or violating licensing terms.

Open standards should be implemented by multiple vendors on multiple platforms. To achieve this, open standards must be free of proprietary dependencies[1] and single-vendor functionality.[2] Open standards exist to ensure interoperability and any particular focus on a specific vendor platform will derail the efforts of good interoperability. Thus, open standards are generally developed within a consortium of interested parties that agree to work together in the interest of interoperability.

It was in this context that the Organization for the Advancement of Structured Information Standards (OASIS) started its work in 2002 to define a complete open standard for office documents. The members involved in defining the specification came from a wide range of companies and organizations such as, Adobe Systems, IBM, Intel, Novell and Sun Microsystems, as well as developers from widely-used free and open source software projects such as AbiWord, Gnumeric, KOffice, OpenOffice.org and others.

The work at OASIS culminated with the OpenDocument Format for Office Applications (ODF) standard. ODF is designed to store traditional office documents such as reports, books, spreadsheets, charts, presentations and word processing documents. As an open standard, ODF ensures easy implementation by software developers in their own software, as well as interoperability with any application that supports ODF. It is based on XML[3] and as its specifications are well documented and available publicly, it is fairly easy for any developer to add ODF support to an existing application or build new software that uses ODF. As such, ODF was immediately popular upon its publication as a standard, with many competing office applications adopting it as their native format. In 2006, ODF was accepted as an International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC) standard and published as ISO/IEC 26300.

Users in Control

The development of ODF is an important milestone for many reasons. For the very first time in the modern history of personal computers, users have full control of their documents. Users can change their office productivity software and their documents will still retain their full fidelity.

Global Adoption in the Public Sector

Governments and administrative bodies have been quick to recognize the merits of ODF and have started to integrate ODF as national policies for document use and exchange:

1. **Australia:** The National Archives of Australia (NAA) is moving its digital archives system to support ODF. The move to ODF comes as the NAA receives information in disparate file types. Converting it to ODF *ensures the longevity of the information*.[a]

2. **Belgium:** The Belgian government announced its intentions to fully use ODF from September 2008,[b] stating that: *Until recently users of one of these suites experienced difficulties with exchanging documents with users of other software. These last few years have, however, seen an important standardization effort, which has led to the definition of a new standard. … The federal Council of Ministers therefore proposes establishing ODF as the standard for exchanging office documents such as texts, spreadsheets, presentations within the civil service as soon as the format is definitively approved by ISO.*

3. **Brazil:** Brazil's government has recommended ODF as its preferred document format.[c]

4. **Denmark:** Folketinget (the Danish Parliament) unanimously adopted a decision on 2 June 2006 to impose on the Danish government a duty to ensure that before January 2008, the public sector's use of IT is based on open standards and that all digital information and data that the authorities exchange with citizens, companies and institutions are available in formats that are based on open standards.[d]

[1] Proprietary dependencies refer to technical dependencies on external proprietary platforms.
[2] Single-vendor functionality refers to functionality which may only be implemented on a particular vendor platform due to technical limitations.
[3] XML or Extensible Markup Language is a flexible way to create common information formats and share both the format and the data on the World Wide Web, intranets and elsewhere. XML is a formal recommendation from the World Wide Web Consortium.
http://www.netproject.com/docs/migoss/v1.0/glossary.html

5. **France:** A report commissioned by the former French Prime Minister, Dominique de Villepin, recommended that all French government publications should be made available in ODF.[e]

6. **Hong Kong Special Administrative Region:** In March 2006, ODF was added to the Hong Kong SAR's Government Interoperability Framework as a recommended standard.[f]

7. **India:** The Government of India strongly supports ODF, with the Ministry of Information and Technology stating that[g]: *...considering the huge potential of eGovernance in the nation as well as the need to adopt open standards to make our data systems more inter-operable and independent of any limiting proprietary tools, we feel that ODF is a great technological leap and a big boon to further propel IT right to India's grass root levels. I congratulate this initiative of leading private and public organizations...*

   There have been other moves to adopt ODF in India:

   - The Allahabad High Court of India made a policy decision to use ODF.[h]
   - The Election Commission of the Government of India is working on adopting ODF using open source software.[i]
   - The government of Delhi moves to ODF-based office suites for cost reasons.[i]
   - The state of Kerala in India suggests the use of ODF in its draft IT policy,[j] stating that: *Open standards such as Unicode and Open Document Format and Open Architectures will be followed in e-governance projects to avoid total dependence on select vendors.*

8. **Italy:** ODF has been adopted as an Italian standard for document exchange.[k]

9. **Norway:** A proposal on: *'The Public Sector Use of Open IT Standards and Open Source Software,*[l] from a working group established by the Ministry of Modernization of Norway stated that: *ODF constitutes a significant contribution toward establishing a future common standard for document exchange…and should be evaluated as a future administrative standard.*

   Recently, the Minister of Renewal of Norway revealed the Norwegian Standards Council recommendation for mandatory government use of ODF and PDF.[m] The recommendation, if adopted, would require all government agencies and services to use ODF and PDF in a primary capacity, while other formats are only used in a redundant capacity.

10. **South Africa:** The country's largest science and technology research organization, the Council for Scientific Industrial Research (CSIR), is adopting ODF throughout the organization.[n] CSIR President and CEO Dr. Sibusiso Sibisi states that: *Open document standards are of prime importance for allowing open access to information, now and in the future. By using open document standards to store our data, the CSIR is not locked into a specific vendor that developed and implemented a proprietary standard, thus eliminating the risk of not being able to access current data in future when such a standard may cease to be supported.*

11. **United Kingdom:** The British Education Communication Technology Agency (BECTA) is the UK agency in charge of defining IT policy for all schools in the UK. BECTA specifies the use of ODF alongside rich text format and plain text for office documents in its technical specification for institutional infrastructure.[o]

12. **United States of America:** The American National Standards Institute recently balloted and approved ODF as an American National Standard.[p] In the public sector, the state of Massachusetts specifies the use of ODF as the open standard for desktop application interoperability in its Enterprise Technical Reference Model.[q] In addition, many states and government agencies in the United States of America are moving to office productivity software that uses ODF.[r]

There are many other public policy actions taken by governments worldwide and the adoption of ODF is growing very rapidly in the public sector.[s] Figure 1 below indicates the use of ODF in the public sector[4] (areas colored in red indicate regions with significant ODF adoption activity in the public sector).

Ecma 376

In 2004, the Telematics between Administrations Committee (TAC) of the European Commission noted that[t]: *Because of its specific role in society, the public sector must avoid that a specific product is forced on anyone interacting with it electronically. Conversely, any document format that does not discriminate against market actors and that can be implemented across platforms should be encouraged.*

TAC issued the following recommendations to Microsoft[t]:

- *Consider issuing a public commitment to publish and provide non-discriminatory access to future versions of its WordML specifications.*
- *Consider the merits of submitting XML formats to an international standards body of their choice.*

In December 2006, Microsoft released Office Open XML (OOXML) which was approved by Ecma International as an Ecma Standard (Ecma 376). Ecma 376 is an XML-based electronic document format providing essentially the same functionality as ODF. Ecma 376 was subsequently submitted for fast track approval to ISO.[u]

[4] It should be noted that public sector policies on electronic documents may change over time and as such the regions noted in Figure 1 should not be regarded as a definitive source of ODF adoption in the public sector.

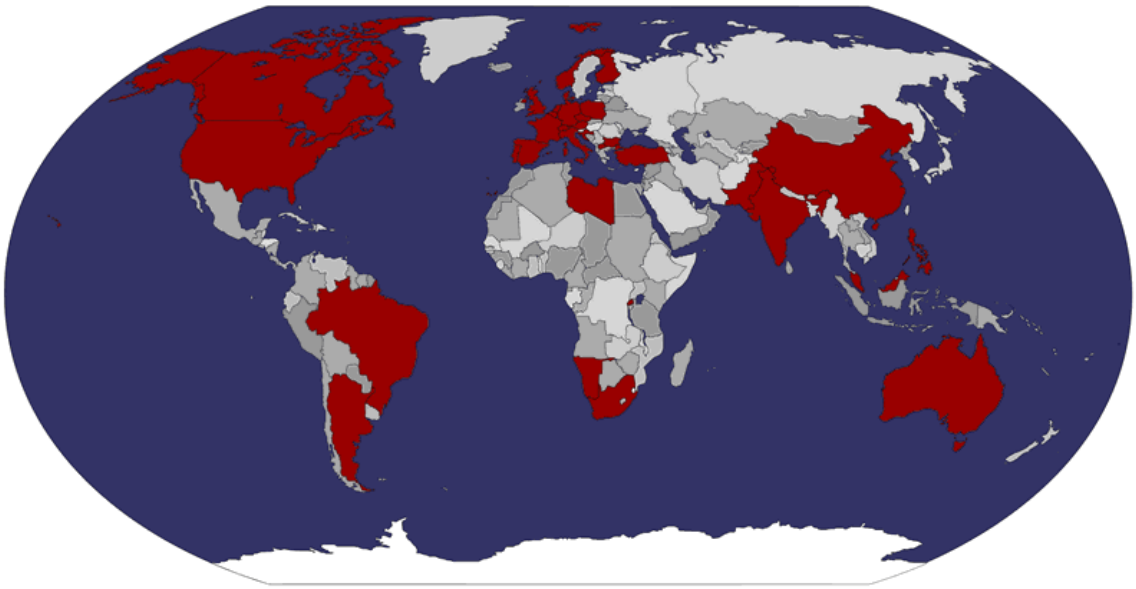

Figure 1: Worldwide ODF Adoption Activity in the Public Sector

Since Ecma 376 serves a similar purpose as ODF, national bodies participating at ISO and other interested parties have criticized its submission for fast-tracked approval as an ISO standard. The negative and divided responses received by ISO from participating national bodies and interested observers with regard to Ecma 376 highlighted a number of technical problems with the proposed standard, contradictions with other ISO standards, and possible licensing issues.[v]

Ecma 376 as a proposed ISO standard is currently in the five-month review period and it remains to be seen if the issues noted would be sufficiently addressed and resolved at the end of the review period.[w]

Portable Document Format

It is important to note that in addition to ODF and Ecma 376, the Portable Document Format (PDF) by Adobe Systems is considered by some organizations as well as national standards bodies to be suitable for document exchange. PDF files are in widespread use and are well supported by many applications running on multiple platforms. The PDF file format is flexible enough to describe complex documents and it is considered the de-facto standard for electronic archiving purposes.[m] In addition, specialized subsets of the PDF file format have been approved as ISO standards for specific industries and functions.[5] Adobe Systems recently announced that the entire PDF file format specifications will be submitted to ISO for consideration as an international open standard.

It must be noted that while the PDF specification is suitable for document exchange and archiving, it is not suitable for common office productivity functionality. The PDF file format has no specific support for word processing, spreadsheets and multimedia presentations. It is strictly structured for the capture, distribution and printing of electronic documents across multiple platforms while retaining full document fidelity. From that perspective, PDF fulfills an important niche but it is not a replacement for either ODF or Ecma 376. With the possible introduction of the PDF specification as an international ISO standard, government policy makers may consider the use of PDF files in the area of document archival and document distribution only.

Conclusion

Open standards in the office document space will foster competition and facilitate innovation. The effort to migrate to established open standards will be rewarded with the millions of documents that will be accessible for many years to come. Extensive studies performed by various analysts have indicated substantial cost savings when embracing open standards such as ODF due to increased choices in the marketplace. As such, governments, industries and educational institutions should actively explore and pursue strategies in this area.

~ Ditesh Kumar, Malaysian National Computer Confederation – ODF Special Interest Group

References

a. H. Dahdah, 'National Archives readies move to OpenDocument'. http://www.computerworld.com.au/index.php?id=954149621&eid=-6787

b. 'Open Standards'. http://presscenter.org/archive/20060623/432d0130470a88df1105dda38d1282b0/?lang=nl&prLang=en

[5] Examples of specialized subsets of PDF that have been accepted by ISO include PDF/Archive (PDFA) and PDF/Exchange (PDFE).

c. Brazillian ePING Framework, Version 2.0.
http://www.apdip.net/projects/gif/country/BR-GIF.pdf

d. 'Openize Denmark, Parliament Orders'.
http://gotze.eu/2006/06/openize-denmark-parliament-orders.html

e. C. Bernard, 'On Equal Terms'.
http://www.ladocumentationfrancaise.fr/rapports-publics/064000728/index.shtml

f. The Government of the Hong Kong Special Administrative Region, 'The HKSARG Interoperability Framework'.
http://www.ogcio.gov.hk/eng/infra/download/s18.pdf

g. 'Roadmap to change the way we document'.
http://www.odfalliance.org/press/Release%2020060623.pdf

h. Allahabad High Court ODF Usage.
http://www.allahabadhighcourt.in/faq.htm

i. 'Delhi Government Phases Out MS Office, Adopts Free ODF-Based Office Suites To Save Licence Fee; LIC, Delhi HC & Nirvachan Sadan Set To Follow'.
http://epaper.timesofindia.com/Repository/ml.asp?Ref=RVRELzIwMDYvMDYvMjQjQXIwMDEwNQ==&Mode=HTML&Locale=english-skin-custom

j. 'Kerala's draft IT policy released'.
http://www.hindu.com/2007/01/18/stories/2007011801800700.htm

k. 'ODF as an Italian standard'.
http://webstore.uni.com/unistore/public/productdetails?productId=UNICI2630000!EEN

l. 'Public Sector Use of Open IT Standards and Open Source Software'.
http://www.regjeringen.no/en/dep/fad/Documents/Horinger/Horingsdokumenter/2005/Horing-Bruk-av-apne-standarder-og-apen-k.html?id=97645

m. G. Updegrove, 'Norwegian Standards Council Recommends Mandatory Use of ODF and PDF'. http://www.consortiuminfo.org/standardsblog/article.php?story=20070513180219689

n. 'CSIR opts for open document format in its migration to an open source environment'.
http://www.csir.co.za/plsql/ptl0002/PTL0002_PGE013_MEDIA_REL?MEDIA_RELEASE_NO=7454764

o. 'Technical specification: institutional infrastructure'.
http://events.becta.org.uk/display.cfm?resID=14615&CFID=10796395&CFTOKEN=cbbe0beb4f8cf01f-CCEA5383-D89B-4E08-6DCAC662AD5191D2

p. E. Montalbano, IDG News Service, 'Microsoft backs adding ODF to ANSI standards'.
http://www.networkworld.com/news/2007/051707-microsoft-backs-adding-odf-to.html

q. Commonwealth of Massachusetts, 'Enterprise Technical Reference Model – Version 3.5'.
http://www.mass.gov/Aitd/docs/policies_standards/etrm3dot5/etrmv3dot5intro.pdf

r. ODF Migration in the United States of America.
http://opendocumentfellowship.org/government/precedent#S38

s. ODF Adoption in the Public Sector.
http://opendocumentfellowship.org/government/precedent

t. 'TAC approval on conclusions and recommendations on open document formats'.
http://europa.eu.int/idabc/en/document/2592/5588

u. 'Ecma International approves Office Open XML standard.
http://www.ecma-international.org/news/PressReleases/PR_TC45_Dec2006.htm

v. M. Bennett, IT Week, 'Open XML hits further problems'.
http://www.computing.co.uk/itweek/news/2174488/open-xml-hits-further-problems

w. 'Estimates of the Costs of Implementing Office Open XML and ODF in the Central Government', Denmark: Ramboll Management.
http://www.odfalliance.org/resources/RamboellReport.pdf

Additional Reading

J. Kaplan, 'APDIP e-Note 7 – Open Technologies: Catalyst for transformation'. http://www.apdip.net/news/apdipenote7

S. H. Nah, 'Free/Open Source Software: Open Standards', Bangkok: UNDP-APDIP
http://www.iosn.net/open-standards/foss-open-standards-primer/ and
http://en.wikibooks.org/wiki/FOSS_Open_Standards

UNDP-APDIP Government Interoperability Framework (GIF) Project (with links to existing GIFs).
http://www.apdip.net/projects/gif

'A Global View of ODF Policy'.
http://www.odfalliance.org/resources/GlobalViewODFPolicy.pdf

ConsortiumInfo.org. http://www.consortiuminfo.org

'Office Open XML'.
http://en.wikipedia.org/wiki/OpenXML

'OpenOffice.org Statistics'. http://stats.openoffice.org

'Portable Document Format'.
http://en.wikipedia.org/wiki/Pdf

'Q&A Microsoft Co-Sponsors Submission of Office Open XML Document Formats to Ecma International for Standardization'.
http://www.microsoft.com/presspass/features/2005/nov05/11-21Ecma.mspx