

UPGRADE is the European Journal for the Informatics Professional, published bimonthly at <http://www.upgrade-cepis.org/>

Publisher

UPGRADE is published on behalf of CEPIS (Council of European Professional Informatics Societies, <http://www.cepis.org/>) by **Novática** (<http://www.ati.es/novatica/>), journal of the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática*, <http://www.ati.es/>)

UPGRADE monographs are also published in Spanish (full version printed; summary, abstracts and some articles online) by **Novática**

UPGRADE was created in October 2000 by CEPIS and was first published by **Novática** and **INFORMATIK/INFORMATIQUE**, bimonthly journal of SVI/FSI (Swiss Federation of Professional Informatics Societies, <http://www.svifi.ch/>)

UPGRADE is the anchor point for UPENET (UPGRADE European NETWORK), the network of CEPIS member societies' publications, that currently includes the following ones:

- **Informatik-Spektrum**, journal published by Springer Verlag on behalf of the CEPIS societies GI, Germany, and SI, Switzerland
- **ITNOW**, magazine published by Oxford University Press on behalf of the British CEPIS society BCS
- **Mondo Digitale**, digital journal from the Italian CEPIS society AICA
- **Novática**, journal from the Spanish CEPIS society ATI
- **OCG Journal**, journal from the Austrian CEPIS society OCG
- **Pliroforiki**, journal from the Cyprus CEPIS society CCS
- **Pro Dialog**, journal from the Polish CEPIS society PTI-PIPS

Editorial Team

Chief Editor: Llorenç Pagés-Casas, Spain, pages@ati.es

Associate Editors:

François Louis Nicolet, Switzerland, nicolet@acm.org
Roberto Carniel, Italy, carniel@dgf.uniud.it
Zakaria Maamar, Arab Emirates, Zakaria.Maamar@zu.ac.ae
Soraya Kouadri Mostéfaoui, Switzerland, soraya.kouadrimostefaoui@gmail.com
Rafael Fernández Calvo, Spain, rfcvalvo@ati.es

Editorial Board

Prof. Wolfried Stucky, CEPIS Former President
Prof. Nello Scarabottolo, CEPIS Vice President
Fernando Piera Gómez and Llorenç Pagés-Casas, ATI (Spain)
François Louis Nicolet, SI (Switzerland)
Roberto Carniel, ALSI - Tecnoteca (Italy)

UPENET Advisory Board

Hermann Engesser (Informatik-Spektrum, Germany and Switzerland)
Brian Runciman (ITNOW, United Kingdom)
Franco Filippazzi (Mondo Digitale, Italy)
Llorenç Pagés-Casas (Novática, Spain)
Veith Risak (OCG Journal, Austria)
Panicos Masouras (Pliroforiki, Cyprus)
Andrzej Marciniak (Pro Dialog, Poland)
Rafael Fernández Calvo (Coordination)

English Language Editors: Mike Andersson, Richard Butchart, David Cash, Arthur Cook, Tracey Darch, Laura Davies, Nick Dunn, Rodney Fennemore, Hilary Green, Roger Harris, Michael Hird, Jim Holder, Alasdair MacLeod, Pat Moody, Adam David Moss, Phil Parkin, Brian Robson

Cover page designed by RFCalvo, © Rafael Fernández Calvo 2006

Layout Design: François Louis Nicolet

Composition: Jorge Llácer-Gil de Ramales

Editorial correspondence: Llorenç.Pagés-Casas@ati.es

Advertising correspondence: novatica@ati.es

UPGRADE Newsletter available at

<http://www.upgrade-cepis.org/pages/editinfo.html#newsletter>

Copyright

© Novática 2006 (for the monograph)

© CEPIS 2006 (for the sections UPENET and CEPIS News)

All rights reserved under otherwise stated. Abstracting is permitted with credit to the source. For copying, reprint, or republication permission, contact the Editorial Team

The opinions expressed by the authors are their exclusive responsibility

ISSN 1684-5285

Monograph of next issue (February 2007)

"Next Generation Web Search"

(The full schedule of UPGRADE is available at our website)



The European Journal for the Informatics Professional
<http://www.upgrade-cepis.org>

Vol. VII, issue No. 6, December 2006

- 2 Editorial: UPGRADE in Top Position of Google PageRank for ICT Journals — *Geoff McMullen*

Monograph: Open Document Format (ODF) (published jointly with Novática*)

Guest Editors: *Jesús Tramullas-Saz, Piedad Garrido-Picazo, Marco Fioretti*

- 3 Presentation: OpenDocument Standard for Digital Documents — *Jesús Tramullas-Saz and Piedad Garrido-Picazo*
- 5 Open by Design: The OpenDocument Format Standard for Office Applications — *Erwin Tenhumberg, Donald Harbison, and Rob Weir*
- 14 Is OpenDocument an Open Standard? Yes! — *David A. Wheeler*
- 25 OpenDocument Hidden Traps and their Side Effects on Free/Open Source Software — *Marco Fioretti*
- 29 ISO-26300 (OpenDocument) vs. MS-Office Open XML — *Alberto Barrionuevo-García*
- 38 Interoperability: Will the Real Universal File Format please Stand Up? — *Sam Hiser and Gary Edwards*
- 47 ODF: The Emerging Document Format of Choice for Governments — *Marino Marcich*
- 50 Promotion of the Use of Open Document Formats by the IDA and IDABC Programmes — *Miguel A. Amutio-Gómez*
- 53 A Brief History of Open Standards in Denmark — *John Gøtze*
- 57 Standard Open Formats and Libre Software in the Extremadura Public Administration — *Luis Millán-Vázquez de Miguel*

UPENET (UPGRADE European NETWORK)

- 60 From **Mondo Digitale** (From AICA, Italy)
Programming Languages
Programming Languages: An Introduction — *Carlo Ghezzi*
- 64 From **Pro Dialog** (PTI-PIPS, Poland)
E-Commerce
Organization and Economics of Entertainment Services Networks Exchanging Virtual Goods — *Andrzej P. Urbański*

CEPIS NEWS

- 70 Harmonise Project: Assigning Responsibilities among the Partners and Planning Next Steps — *François-Philippe Draguet*
- 71 News & Events: European Funded Projects and News Updates

* This monograph will be also published in Spanish (full version printed; summary, abstracts, and some articles online) by **Novática**, journal of the Spanish CEPIS society ATI (*Asociación de Técnicos de Informática*) at <http://www.ati.es/novatica/>.

Presentation

OpenDocument Standard for Digital Documents

Jesús Tramullas-Saz and Piedad Garrido-Picazo

The boom experienced by the libre software movement and its unlimited possibilities, together with the growing success enjoyed by software libre tools in the commercial and business world, has overshadowed a number of key issues that are inherent to the movement. We are of course referring to open standards, which are core elements to achieve the interoperability of software applications. Standards are what set the rules of the game for many aspects and functionalities of software tools. Compliance with standards, whether *de facto* or *de iure*, is what can make or break an application. Much of the software development going on today is guided by standards, as is the much vaunted issue of "software quality".

For this reason the scant interest aroused by standards in recent years is even more surprising. Obviously, an application is more exciting (and useful) than a technical standard, especially if it is an open standard. However, the former are impossible without the latter. Most of the protocols on which the communication and transfer of information over the Internet is based are open (or nearly open) standards. The *World Wide Web Consortium* directs its efforts towards formulating and accepting standards in the knowledge that without standards it

would be impossible to carry on developing the Web and its advanced information services. But this seems to be contradicted by the lack of standards for what is the most widely created and disseminated information worldwide: the information stored in office automation documents.

Whoever controls users' desktop controls their applications. And 80% of all computer users work with classic desktop applications, with a word processor or a spreadsheet. Office work in companies, the work done in public administrations or in educational environment, are examples of the intensive use of office automation applications. The information they use, generate, and transform is stored in office automation documents. The formats of these documents have become *de facto* standards which have cunningly imposed their 'rules' on all users. These rules, which are often abusive, limit freedom of choice, compatibility, and interoperability, and oblige users to consider as normal a series of costs that, in any other context, would immediately be denounced as monopolistic abuse and contrary to the principles of fair competition.

To have a standard is, by definition, recommendable. A standard establishes requirements and the rules of the game.

The Guest editors

Jesús Tramullas-Saz is an associate professor in the Dept. of Documentary Sciences, *Universidad de Zaragoza*. He is a member of the GRIO (*Gestión de Recursos de Información en las Organizaciones*, Information Resource Management in Organizations) research group and a coordinator of the Thematic Network on Digital Documentation (National R&D&I Plan, 2004–2005 and 2006–2007). He is the principal researcher on the project "Semantic Web and digital libraries: the development information services based on RDF and Topic Maps" (2006–2007). His main research interests are digital libraries and digital information services, content management, and libre software tools for information management. He is responsible for maintaining the Spanish translation of the libre software *Greenstone Digital Library Software*. <jesus@tramullas.com>.

Piedad Garrido-Picazo is an assistant professor in the Dept. of Computing and Systems Engineering, *Universidad de Zaragoza*. She is a member of the GRIO (*Gestión de Recursos de Información en las Organizaciones*, Information Resource Management in Organizations) research group. She belongs to the Thematic Networks on Information Retrieval from Digital Texts and Libraries, Digital Documentation, and Soft-Computing-based Web Information Access Systems. Her main research interests are xml databases, libre software for information management, digital libraries, RDF and Topic Maps (XTM) in the context of the semantic Web. <piedad@unizar.es>.

Jesús and Piedad have coordinated together the book *Software libre para servicios de información digital*, Prentice Hall, Madrid 2006.

Marco Fioretti is a hardware systems engineer and a part time freelance writer for several Linux and other IT magazines. He is the co-founder (in 2002) and current coordinator of the Run Up to date Linux Everywhere Project <www.rule-project.org>, which makes modern Free Software easily installable on older computers. Marco Fioretti has always been interested in truly Open file formats and protocols, from e-books to portable databases and the Universal Business Language, and in their impacts on economy, education and civil rights. In 2004, for example, he investigated the philosophical links between Free Software and Scouting. In the same spirit, in 2006 he co-founded Elèutheros <www.eleutheros.it> which aims to promote a wider, official adoption of Free Software and non proprietary formats within the Catholic Church. More recently, Fioretti has also started to follow the attempts of the Free Software and disabled users communities to communicate more effectively. His more recent initiative to promote the importance of Free Standards and Software among all citizens is the Family Guide to Digital Freedom <<http://digifreedom.net>>. Marco Fioretti is also member and contact for Italy of the OpenDocument Fellowship <<http://opendocumentfellowship.org>>, a volunteer organization that promotes the use and development of the OpenDocument format, and the author of Everybody's Guide to OpenDocument <www.linuxjournal.com/article/8616>. <marco.fioretti@opendocumentfellowship.org>.

However, a standard may be used in such a way as to create undesirable results, especially when patents and legal constraints that favour one party over another are involved. For this reason it is essential for standards to be open, developed in a spirit of collaboration among equals, and for their specification not to contain hidden constraints that make it hard to use. If these standards also ensure all citizens the right to access, store, and transform digital information, regardless of the platform used, they immediately acquire an incalculable economic, social, and political value.

OpenDocument is a standard (*the only standard*) for office automation documents which sets out all the desirable characteristics that should be found in such documents. Its version 1.0 has been endorsed by the *International Organization for Standardization* (ISO), as ISO/IEC standard 26300:2006. It is the result of the open, collaborative work of all the main players; software developers, solution implementers, and end-users. It is public and free of charge, and the legal requirements included in the standard prevent any partial or abusive use. It also makes use of other open standards within its own specification, such as XML, SVG, and Dublin Core. OpenDocument had already been developed and approved as a standard by the *Organization for the Advancement of Structured Information Standards*, OASIS, in 2005, which ensures its continuing support and development by the main industry players.

However, OpenDocument is not just a format standard for office automation documents. The underlying philosophy is different. Some may argue that a standard is nothing more than a technical issue, and in purely formal terms they may be right. But the technical excellence of OpenDocument, which is undoubtedly superior to that of any existing office document format, is a product of the underlying philosophical approach and the way in which this philosophy has been applied to working methods and techniques, and to the resultant end-product. The adoption of an open and democratic collaborative working method has made it possible to involve all interested parties, thereby ensuring that all their many and varied demands are addressed. The involvement of industry interests ensures the presence of standard-compliant products on the market, which in turn ensures inter-platform portability and interoperability.

Any citizen can access the document containing the technical specification, which means that the information contained in OpenDocument format documents are not subject to arbitrary decisions by third parties. This open approach also encourages competition between products and, given a range of options, users are free to choose the one they think is best suited to their needs, or the one that is most technologically advanced. This independence is not only desirable; it is essential and inevitable. The OpenDocument format is prepared for the Semantic Web. All tagging complies with the XML standard and documents may include the *Dublin Core* metadata standard (ISO 15:836:2003). Various output can be obtained using XSLT. As they are XML-tagged text documents, the tools and libraries for search engines such as Lucene or Xapian are able to process them with a minimum load. For example, by combining with other standards it is possible to generate *Topic Maps* (ISO/IEC 13250:2003) from document content, with all that can mean for the development of information extraction and representation systems.

One issue that is of paramount importance, which is often completely neglected, is the preservation of digital information. Although many organizations seem still to be unaware of it, the medium- and long term retention of digital assets, most of which are stored in office automation documents, is a growing concern, both due to knowledge management issues affecting the organization itself, and due to legal issues related to their activity. While there is a standard for the long-term preservation of electronic documents (ISO 19005-1:2005) using PDF, the fact is that it is only valid for final versions of a document. OpenDocument is able to keep an activity log of the content of a document, and can "remember", say, the formulas used to reach a mathematical result. Add to this the growing demand for records and document management in the field of public administrations, companies, etc., which also have their own set of standards (ISO 15489-1/2:2001 *Records Management*; UNE/ISO 15489-1/2:2006 *Documents Management*), and it would seem clear that OpenDocument has a long future ahead of it and plenty of ground to cover.

Translation by Steve Turpin

Useful References on ODF

- OpenDocument version 1.0 – specification: <<http://www.oasis-open.org/committees/download.php/12572/OpenDocument-v1.0-os.pdf>>.
- OASIS OpenDocument Essentials: <<http://books.evc-cit.info/>>.
- Opportunities for Innovation with OpenDocument Format XML: <<http://opendocument.xml.org/files/LOW10771-USEN-00.pdf>>.
- OpenDocument - Formula TC: <http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office-formula>.
- OASIS OpenDocument Format for Office Applications: <http://www.oasis-open.org/committees/tc_home.php?wg_abbrev=office>.
- The OpenDocument Foundation: <<http://opendocumentfoundation.us/>>.
- OpenDocument Format Alliance: <<http://www.odfalliance.org/>>.
- OpenDocument Fellowship: <<http://opendocumentfellowship.org/>>.
- Technorati: OpenDocument: <<http://technorati.com/posts/tag/OpenDocument>>.
- OpenDocument xml.com: <<http://opendocument.xml.org/>>.
- Open Interoperative Document Initiative: <<http://www.oidi.org/tiki-index.php>>.
- OpenDocument Format Viewer: <<http://opendocumentfellowship.org/odfviewer>>.
- OpenOffice: <<http://www.openoffice.org/>>.
- Koffice: <<http://www.koffice.org/>>.
- O Reilly ONLamp: What Is OpenDocument: <<http://www.onlamp.com/pub/a/onlamp/2006/07/27/what-is-opendocument.html>>.
- Sam Hiser's blog: <<http://fussnotes.typepad.com/plexnex/>>.
- Andy Updegrave's blog: <<http://www.consortiuminfo.org/standardsblog/>>.
- Bob Sutor's blog: <<http://www.sutor.com/newsite/blog-open/>>.
- Charles H. Schulz's blog: <<http://www.libervis.com/blogs/5/charles>>.
- David A. Wheeler's blog: <<http://www.dwheeler.com/>>.
- Erwin Tenhumberg's blog: <<http://blogs.sun.com/dancer/>>.
- Ron Weir's blog: <<http://www.robweir.com/blog/index.html>>.