

ISO/IEC 29500 – Office OpenXML – formát pre Office Applications

**Normy pre riadenie a prevádzku IT,
Bratislava, 16.2.2011**

Juraj Šitina, SK 34

SK 34

- **Popis dokumentov a tvorba jazykov**
 - Členovia: Ivan Vazan, Igor Hladík, Juraj Šitina
 - **Rámec pôsobnosti:**
 - Jazyky pre opis logickej štruktúry dokumentov
 - Jazyky pre opis objektov ako dokumentov v prostredí webu
 - Architektúra spracovania dokumentov a formátovanie pre logické dokumenty
 - Jazyky opisujúce interaktívne dokumenty
 - Štrukturovaný jazyk pre hypermedia dokumenty a aplikačné prostriedky
 - API's pre spracovanie dokumentov
-

ISO/IEC 29500 – dokumentový formát

- Norma má štyri časti a dve úrovne:
 - **ISO/IEC 29500-1:2008 - Základy a Markup language Reference**
 - **ISO/IEC 29500-2:2008 - Otvorená konvencia na kompresiu súborov**
 - **ISO/IEC 29500-3:2008 - Markup kompatibilita a rozšíriteľnosť**
 - **ISO/IEC 29500-4:2008 - Prechodové migračné vlastnosti**
 - **Strict a Transitional**
 - Strict neobsahuje elementy/atribúty zaisťujúce verné zachytenie informácií zo starších dokumentov
-

História

- December 2006 - ECMA (European Computer Manufacturer Association) štandard - ECMA-376



NEXTpage®



TOSHIBA
Leading Innovation >>>



PUBLIC AFFAIRS OFFICE

- April 2008 – ISO/IEC 29500

Prečo ISO/IEC29500

- Otvorený štandard pre dokumenty:
 - word-processing
 - Tabuľkové procesory
 - Prezentácie
 - Interoperabilita
 - Platformovo nezávislé
 - SW, HW
 - Internacionalizácia
 - Dlhodobá uchovateľnosť dokumentu
 - Nové možnosti pre aplikácie
 - Integrácia údajov z obchodných procesov
 - Podpora Custom XML schém
 - High Fidelity Migrácia zo starších binárnych formátov (*.doc, *.xls, *.ppt)
 - 400+ miliónov používateľov
 - Miliardy dokumentov za posledných 20 rokov
-

Štruktúra ISO/IEC29500 formátu

Základné jazyky pre jednotlivé typy dokumentov

- WordprocessingML (prípona .docx)
 - SpreadsheetML (prípona .xlsx)
 - PresentationML (prípona .pptx)
-

Štruktúra ISO/IEC29500



dokument.
docx

Užívateľ vidí jeden súbor



Súborový kontajner

Vývojár/softvér vidí modulárny súbor

Dokument Open XML

- Väčšinu tvorí XML
- Každá XML časť je diskretný, komprimovaný komponent
- Možnosť pridávania, extrahovania a úprav jednotlivých častí bez použitia Office programov
- Narušenie alebo absencia ľubovoľnej časti nebráni otvoreniu súboru

Vlastnosti dokumentu

Komentáre

WordML, SpreadsheetML ...

Vlastné definované XML

Obrázky, video, zvuk

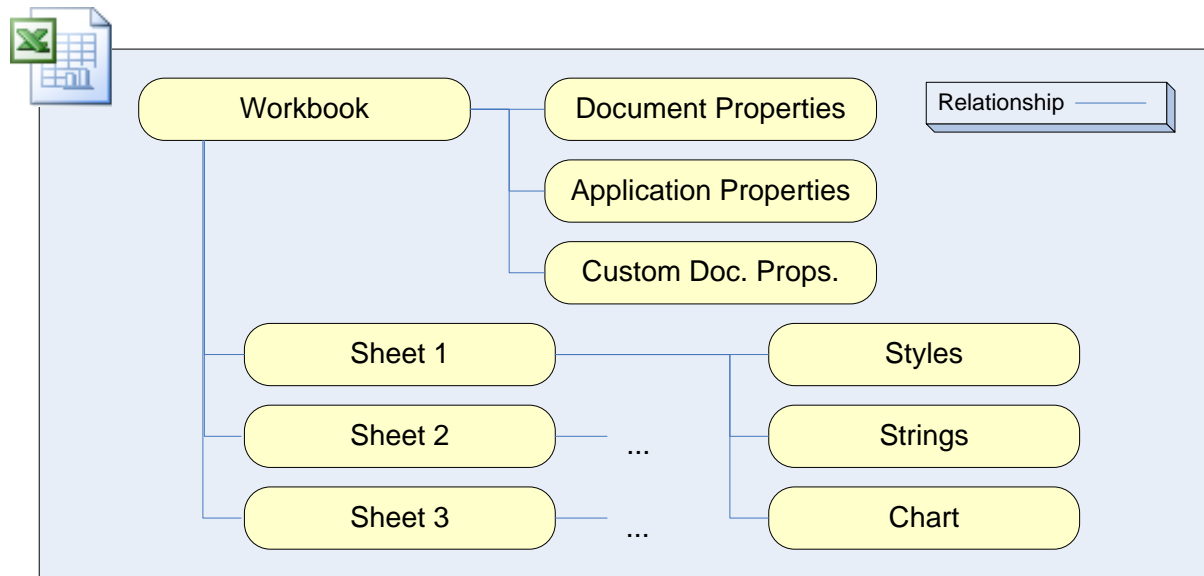
Vložený kód / makrá

Grafy

Štruktúra ISO/IEC29500

Open Packaging Convention - OPC

- Balíček – ZIP kontajner
- Časti dokumentu – súbory v ZIP balíčku
- „Content Types.xml“ – definícia a vynútenie typov obsahu
- Vzťahy – povinné vzájomné referencovanie častí dokumentu



Štruktúra ISO/IEC29500

- Balíček by mal obsahovať aspoň tri časti:
 - Časť dokumentu
 - Časť definície typu obsahu
 - Časť definujúca vzťahy v balíčku

Najjednoduchší dokument by mohol vyzerat' takto:

```
<?xml version="1.0" encoding="UTF-8"?>
<w:document xmlns:w="http://schemas.openxmlformats.org/wordprocessingml/2006/main">
  <w:body>
    <w:p>
      <w:r>
        <w:t>Hello world</w:t>
      </w:r>
    </w:p>
  </w:body>
</w:document>
```

Norma ISO/IEC29500 v praxi

- **Generovanie dokumentov**
 - tvorba dokumentov na strane servera, napr. z databázy
 - Nahradenie Office Automation
 - **Znovupoužitie obsahu**
 - Jednoduchší presun obsahu medzi dokumentami
 - **Rozšírenia obsahu**
 - Pridávanie špecifických metadát do obsahu dokumentov pre vlastné riešenia
 - **Vyhľadávanie dokumentov**
 - Požiadavky na úložištia dokumentov založené na určení dát, typov obsahu alebo metadát
 - **Ozdravovanie dokumentov**
 - Odstránenie nechceného obsahu (komentáre, vložený kód...)
-

Aplikácie s použitím OpenXML

• Windows:

- Altsoft XML2PDF server 2007
- AltViewer documents preview
- Altova XML Spy
- Corel Office
- Create Word 2007 documents without Word installed
- docx to RTF
- Madcap Flare
- Microsoft Office 2000, Office XP and Office 2003
- Microsoft Office 2007
- Mindjet's MindManager
- Monarch V.9.0 from Datawatch
- ODF-Converter
- OOX-UOF Converter
- Open ERP Software
- Open XML Translator
- Open XML translator for OpenOffice (for Linux and Windows Versions of OpenOffice.org)
- Open XML Writer
- PythonOffice (Python API to read and write Excel XML documents from within Python programming language)
- Sourceforge Project to allow .NET(C#) developer to have component that will interact with Open Xml file
- Special Templates for Master Thesis
- Word 2007 Map Editor for Mindjet MindManager
- Xpertdoc Studio 2007 reporting solution

• Mac OS X:

- DOCX convertor for the Mac
- docx to html Konverter
- docx to RTF Konverter
- MacLinkPlus Deluxe version 16 by DataViz
- Microsoft Office 2008
- Neo Office 2.1
- Sun Open XML import filter for spreadsheets
- Word Counter 2.2.1

• Linux:

- Gnumeric – open source Spreadsheet
- Open Office –Novell edition
- Open XML translator for OpenOffice (for Linux and Windows Versions of OpenOffice.org)

• Aplikácie nezávislé na OS:

- Docx2Doc Web Service
 - DOCX convertor on Palm handheld devices
 - OpenXML4J - Open XML framework for Java
 - OpenXMLDeveloper.org (hundred of developers, multiple platforms)
 - PHPEXCEL - Web Development (PHP)
 - Webservice with mailing of Document, docx to HTML
 - Word 2007 file(docx) construction using Java
-

ISO/IEC29500 - Záver

- <http://openxmldeveloper.org>
 - **Apache POI OpenXML Java API**
 - <http://www.interoperabilitybridges.com/projects/apache-poi-openxml-java-api>
 - **PHP OpenXML API**
 - <http://openxmlapi.codeplex.com/>
 - **Open XML SDK 2.0**
 - <http://www.microsoft.com/downloads/en/details.aspx?FamilyID=c6e744e5-36e9-45f5-8d8c-331df206e0d0&DisplayLang=en>
 - Správa a ďalší vývoj formátu má teraz v rukách medzinárodné spoločenstvo a nie je v rukách jednej veľkej firmy
-

Ďakujem za pozornosť.

jurajs@microsoft.com