

Moderní formáty dokumentů pro kancelářské aplikace

Milan Tomeš

Ústav systémového inženýrství a informatiky, FES, Univerzita Pardubice

Abstrakt: *V současné době se objevují nové možnosti pro ukládání dokumentů, dochází k zásadnímu posunu od klasických formátů z kancelářského balíku MS OFFICE směrem k otevřeným standardům, založených na XML a Opensource. Jedním z nich je ODF - formát přijatý v nedávné době jako ISO standard pro dokumenty a podporovaný i Ministerstvem Informatiky ČR. Microsoft v nové verzi OFFICE 2007 přichází s vlastním formátem OpenXML. Tato problematika se úzce dotýká nejen občanů, ale i institucí veřejné zprávy, proto je třeba znát důvody zavádění těchto moderních formátů stejně tak i z toho vyplývající důsledky.*

Abstract: *At present time is appearing new possibilities of saving documents, there is an essential shift from the classic formats of MS OFFICE package towards open standards based on XML and Opensource. One of them is ODF – format, that has recently been acknowledged as an ISO standard for documents and supported by Ministry of informatics Czech Republic. Microsoft is coming in new version of OFFICE 2007 with his own format OpenXml. This problems is nearly concerning not only citizens, but also establishment of public administration, hence is necessary to known reason of implementing this modern formats as well as inferential implications.*

Klíčová slova: *ODF, dokument, formát, MS OFFICE, OpenOffice, XML.*

Úvod:

Dokumenty obecně jsou jednou z nejdůležitějších částí podnikových procesů, neboť jsou v nich obsaženy téměř všechny informace, které jakýkoliv subjekt při svém rozhodování potřebuje. V současnosti je zcela zřejmý a cílený posun od klasických papírových dokumentů směrem k elektronickým ve formě textových dokumentů, tabulek, grafů, databází, prezentací atd. To s sebou však přináší problematiku kompatibility, neboť v žádném případě nepoužívají všichni uživatelé stejné softwarové prostředky. V oblasti státní správy je toto také velmi závažný problém, neboť je občan při své elektronické komunikaci s úřadem nucen používat takové softwarové vybavení, jaké má úřad a dostává se tak tím do závislosti na dané softwarové platformě, což sebou navíc přináší další finanční náklady v podobě nutnosti takový software nakoupit.

Drtivá většina kancelářských dokumentů je dnes ve světě vytvářena ve vlastních formátech jednotlivých kancelářských aplikací. Kdo chce s dokumentem pracovat, musí si pořídit aplikaci (a často i verzi), v níž byl dokument vytvořen. Filtry mezi různými vlastními formáty jsou nedokonalé, protože většina komerčních výrobců si specifikace svých formátů chrání jako cenné know-how.

Pokud tedy firma nebo instituce začne používat určitý kancelářský software, je nucena i nadále kupovat pouze produkty daného výrobce. Kromě toho k používání stejných aplikací nutí i každého, kdo chce dále pracovat s jejími kancelářskými dokumenty. V případě státních institucí je dnes běžně rozšířené nucení uživatelů k používání výrobků konkrétních soukromých firem, což by bylo v jiném oboru naprosto nemyslitelné.

Stále více se využívá komunikace prostřednictvím e-mailu a elektronických dokumentů mezi občany či společnostmi a vládou. Je nutno předejít situaci, kdy by uživatelé byli závislí na jednotlivém dodavateli, proto je třeba obrátit se směrem otevřeným standardům.

Tato problematika se netýká jen státní správy nebo České republiky, ale v dnešní globalizované společnosti se dotýká celého světa. Díky internetu spolu mohou komunikovat subjekty kdekoliv a kdykoliv a je nezbytně nutné, aby informace, které si poskytují byly v pro obě strany v přijatelných formátech a mohly je tak efektivně zpracovávat.

Je proto paradoxem, že velmi dlouhou dobu, prakticky až do letošního roku neexistoval žádný mezinárodní standard, který by definoval jednotný formát elektronických dokumentů.

Dalším problémem je dlouhodobá archivace. Jestliže uživatel používá formát, který není standardizovaný a otevřený, vystavuje se velkému riziku, že svůj dokument za několik let neotevře. Protože výrobci software přicházejí se stále novými verzemi svých produktů a protože vývoj v oblasti IT je a stále bude více než hektický, morální životnost jak hardware tak software se pohybuje zhruba kolem pouhých dvou let, je dnes bohužel celkem běžnou praxí, že dokumenty vytvořené v nových verzích kancelářského balíku není možné otvírat ve verzích starších. Vzhledem k tomu, že každý právní subjekt má povinnost své dokumenty několik let zpětně archivovat, je toto skutečně závažným problémem.

V letošním roce se však situace změnila a v následujících letech dojde v této oblasti ke zcela zásadním změnám. Za prvé došlo u mezinárodní organizace pro standardizaci ISO ke schválení jednotného formátu pro publikaci dokumentů jménem „OpenDocument format“, zkráceně ODF a za druhé nejsilnější hráč na poli kancelářských balíků firma Microsoft přichází s novou verzí OFFICE 2007, ve které dochází kromě jiných k zásadní změně právě ve formátu dokumentů. Již se nebude používat známý formát *.doc, *.xls ani *.ppt, ale Microsoft přichází s vlastním novým formátem Open XML. Ten má být také otevřený, není však mezinárodním standardem. Nicméně v jeho prospěch samozřejmě hovoří míra penetrace mezi uživateli.

Kromě těchto dvou formátů jsou samozřejmě více či méně rozšířeny i jiné formáty, ať již PDF, Metro, DJVU, či jiné, ale dá se předpokládat, že se rozhodne mezi dvěma prvními zmíněnými formáty. Proto se tento článek vzhledem k rozsahu bude týkat také pouze nich.

Je velmi důležité se s těmito novými formáty seznámit, aplikovat je do praxe a využít je v co nejširší míře pro svůj prospěch, neboť s elektronickými dokumenty pracujeme všichni, dotýkají každého z nás bez ohledu na jeho pozici. Je nezbytné, aby nejenom státní správa, ale všichni uživatelé tento trend zachytili a co nejrychleji se mu přizpůsobili.

Společné znaky formátů Open XML a ODF

Oba formáty jsou v mnohém stejné, nicméně jsou však vzájemně nekompatibilní.

Společně rysy obou formátů jsou:

- Multiplatformní, tedy nezávislé na platformě Windows, Linux, apod.
- Otevřený a bezplatný – specifikace formátu jsou volně k dispozici a každá firma či vývojář jej může zdarma využít ve svém software.
- Založeny na moderní technologii XML a komprimovány známou metodou ZIP.

Výhody těchto nových formátů:

- Uživatel nemusí platit za to, že si chce přečíst své dokumenty.
- Uživatel není závislý na žádném dodavateli, může si vybírat.
- S dokumenty může pracovat každá aplikace či informační systém.
- Dokumenty půjdou otevřít i za mnoho let - specifikace formátu jsou veřejně dostupné a vždy budou.
- Otevřené formáty umožňují oprostít se od závislosti na jednom dodavateli a připravit se na svobodnější jednání v budoucnosti.

Jak pozná změny uživatel při práci? Nijak – práce probíhá stejným způsobem, a ukládání do souboru taktéž. Soubor má však jinou koncovku. Uvnitř souboru je struktura dat zcela jiná, ale to uživatele nemusí zajímat.

XML

Tento standard umožňuje definovat strukturu dokumentu a umožnit tak inteligentní zpracování jednotlivých jeho částí v jiných aplikacích nebo výměnu dat mezi aplikacemi. Současné verze dokumentů se zaměřují především na popis vzhledu a už méně podporují jejich další zpracování v informačních systémech. Nové verze dokumentů tak umožní jejich mnohem efektivnější využití především v informačních systémech a podnikových aplikacích.

Díky této interoperabilitě je možné vytvářet dokumenty rychle a automaticky z celé řady nejrůznějších datových zdrojů, ať již s použitím vlastních kancelářských aplikací nebo bez nich. Získávání dat z existujících systémů lze provádět rychleji a efektivněji.

Otevřený formát

Formáty kancelářského balíku MS Office (a nejen ty) patří do skupiny tzv. uzavřených formátů. Tzn., že společnost Microsoft neposkytl veřejnosti specifikace svých formátů, aby tak mohla donutit uživatele používat právě jejich kancelářský balík a ne jiný. Uživatel je jednoduše vázán na jednoho jediného dodavatele kancelářského balíku či dalších částí vašeho IT, ostatní výrobci nemohou do svých produktů implementovat podporu pro dokumenty MS Office, která by fungovala na 100 % - nemají k dispozici specifikace těchto formátů.

Naproti tomu otevřený formát má své specifikace zcela veřejné a každý kdo bude chtít, jej může využít, ať již pro implementaci do svého produktu, nebo pro jiné účely.

Open Document format - ODF

Co je to ODF

OpenDocument formát (ODF) neboli správněji celým názvem OASIS Open Document Format for Office Applications (OASIS otevřený formát dokumentu pro kancelářské aplikace) je dle [2] otevřený souborový formát určený pro ukládání a výměnu dokumentů vytvořených kancelářskými aplikacemi. ODF zahrnuje textové dokumenty (jako např. poznámky, knihy, dopisy aj.), prezentace, tabulky, grafy a databáze. Standard ODF byl vyvinut sdružením OASIS a vychází ze staršího souborového formátu používaného aplikacemi OpenOffice.org. Formát ODF (i OpenOffice.org) je založen na XML. Skutečnost, že je OpenDocument aplikací XML, je pro většinu uživatelů nepodstatným technickým detailem. Vyplývají však z ní některé výhody, např. rozšiřitelnost, čitelnost i bez speciálních nástrojů a interoperabilita s dalšími aplikacemi XML.

Formát OpenDocument je od 3. května 2006 standardizován Mezinárodní organizací pro normalizaci jako standard ISO/IEC 26300.

Formát ODF byl vyvinut veřejně množstvím organizací, jeho specifikace je veřejně přístupná a formát může být použit libovolnou aplikací bez omezení, jeho účelem je nabídnout otevřenou alternativu uzavřeným formátům, především od společnosti Microsoft (DOC, XLS a PPT využívané kancelářským balíkem Microsoft Office, ale také licencovanému formátu Microsoft Office Open XML).

Výhody ODF

Formát OpenDocument má potenciál nevýhodnou a konkurenci omezující závislost na výrobci softwaru od základu změnit. Každý uživatel si bude moci svobodně vybrat kancelářský software nejlépe vyhovující jeho potřebám, v případě potřeby přejít na jiný produkt a bez obav vyměňovat dokumenty se svým okolím. Důsledkem používání formátu

OpenDocument by tak mělo být snížení nákladů a usnadnění komunikace.

Společnostem také přináší využívání ODF značné úspory na softwarovém vybavení i na dalším rozšiřování jejich infrastruktury. Řada dnes dostupných kancelářských aplikací s podporou ODF, je totiž k dispozici zcela zdarma (např. kancelářský balík OpenOffice)

Nevýhody ODF

Největší nevýhodou je, že software pracující s tímto formátem není tolik rozšířen a tudíž se jeho uživatel stále musí přizpůsobovat jiným formátům. Důvodem tohoto stavu je jednoznačně konzervativnost uživatelů, kteří nechtějí nic měnit, pokud to funguje a dále silný marketingový vliv společnosti Microsoft.

Podpora formátu ODF

Dle [1] podporují ODF například tyto instituce:

- Úřady na celém světě se k formátu OpenDocument staví pozitivně. Např. Belgická vláda rozhodla, že všechny belgické úřady si musí od září 2008 předávat dokumenty jen v otevřených standardech (což zatím splňuje jen formát OpenDocument, bude zvaženo i OpenXML, pokud se stane standardem). Dánská vláda spustí v září čtyřměsíční pilotní program, který zahrnuje publikování dokumentů v OpenDocument formátu (ODF).
 - V ČR jej podporuje Ministerstvo informatiky. Celý projekt podporuje a rozvíjí OSS aliance (<http://www.oss.cz>), nad kterou převzalo záštitu právě Ministerstvo Informatiky.
 - Evropská unie jej prostřednictvím svého programu IDABC aktivně propaguje, viz [8]
 - Stát Massachusetts ve své nové informační politice doporučil, aby všechny jeho úřední dokumenty byly do ledna 2007 převedeny na OpenDocument.
- Dle [1] ODF plně implementují tyto produkty a podporují tyto firmy:
- OpenOffice.org 2, IBM Workplace, 602Office 2, StarOffice, KOffice a řada výrobců podporu pro ODF plánuje pro další verze svých produktů.
 - IBM, Novell, Sun Microsystems, Corel, Opera Software, Oracle, Red Hat a další.

Microsoft Office Open XML:

Nový formát Open XML bude standardním formátem v aplikacích nové řady Microsoft Office 2007. Dle [3] se jedná o nový plně veřejně dokumentovaný formát, který zohledňuje vlastnosti produktů MS Office a zároveň podporuje zpětnou i budoucí kompatibilitu. Právě kompatibilita formátu je důležitá pro uživatele po celém světě. Ti si tak mohou být jistí, že v nových verzích Office otevrou dokumenty uložené ve starších verzích. Dokumenty uložené v nových verzích otevrou ve starších verzích s pomocí zdarma distribuovaného konvertoru. Obrovskou výhodou je, že dokumenty uložené v současných i budoucích verzích budou moci zpracovat, i když nebudou mít MS Office (buď pomocí vlastních nástrojů nebo nástrojů třetích stran). Výhodou formátu OpenXML je také připravenost na provázání/integraci s uživatelskými aplikacemi či vylepšený výkon ve srovnání s jinými XML formáty.

Microsoft podporuje tento standard a jeho technologii již od verze Microsoft Office 2000. V té době byl jazyk XML využit v rámci souborů ve formátu HTML, podporovaných aplikacemi Word, Excel a PowerPoint. Ještě více byl formát XML podporován a využíván v následujících verzích Microsoft Office XP a Microsoft Office 2003, a především pak v novějším produktu Microsoft Office InfoPath 2003.

Celý tento nový formát ukládání dat bude, stejně jako formát Office 2003, k dispozici pod „royalty-free license“, tedy bezplatnou licenci, která je volně šiřitelná – může ji používat kdokoli, je doživotní a je zdarma k dispozici. Pro Office 2000, XP i 2003 (se kterými bude

nový Office zpětně kompatibilní) vyjdou rozšíření (konvertory) pro podporu tohoto formátu.

Tento nový formát souborů bude výchozí pro aplikace Word, Excel a PowerPoint a budou mít koncovku *.docx, *.xlsx, *.pptx

Společnost Microsoft s firmami Apple, Barclays Capital, BP, the British Library, Essilor, Intel Corporation, NextPage Inc, Statoil ASA a Toshiba jako spolusponzory požádala o spolupráci na standardu mezinárodní organizaci Ecma. Členové organizace Ecma se mohou podílet na technickém vývoji standardu. Dalším krokem pak bude žádost o schválení standardu ISO u organizace International Organization for Standardization.

Výhody Open XML

Menší velikosti souborů – Open XML umožňují podstatně menší velikosti souborů než stávající formáty Office 2003. Toho bylo dosaženo kombinací standardní kompresní technologie ZIP, která automaticky komprimuje každý komponent v souboru a sníženou potřebou redundancí dat ve formátu XML.

Lepší obnova dat – Bude možné otevřít a použít neporušené části souboru tam, kde je poškozena pouze jedna komponenta, například graf nebo obrázek. Open XML je navíc vybaveno technologií detekce poškození a pokusu o opravu poškozených souborů v okamžiku, kdy se uživatel pokusí takový soubor otevřít.

Zvýšená bezpečnost - Soubory s potenciálně nebezpečným obsahem lze totiž snáze identifikovat a zneškodnit. Rovněž lze ze souborů snadno odstranit osobní údaje a důvěrný obsah (například poznámky) ještě předtím, než jsou takové soubory odeslány.

Dokumenty vytvořené v současných binárních formátech Microsoft Office budou rovněž plně kompatibilní s aplikacemi Office 2007. Uživatelé tak budou moci ukládat dokumenty v jejich původních formátech a nadále je sdílet s lidmi, kteří budou používat Office 2007. I po přechodu na nový Office budou schopni používat své původní binární dokumenty. To také zajišťuje, že řešení třetích stran založená na stávajících verzích Microsoft Office budou i nadále pracovat.

Nevýhody Open XML

Mezi odborníky v oblasti IT panují velké pochybnosti o otevřenosti tohoto formátu, vzhledem k dominantní pozici firmy Microsoft na trhu a její samozřejmé snahy maximalizovat svůj zisk, nicméně Microsoft tvrdí, že je licence otevřená a používat ji může kdokoli.

Slabou stránkou tohoto formátu také je, že není standardizován a nikdo jiný jej nevyužívá.

Podpora formátu Open XML

Vlivem dominantního postavení firmy Microsoft je velkou výhodou tohoto formátu, že s novou verzí OFFICE jej automaticky přijme a bude používat velké procento uživatelů.

Dále Microsoft založil uskupení s názvem Open XML Formats Developer Group, které má za cíl Open XML podporovat. V současné době má skupina přibližně 40 členů, mezi kterými jsou mimo jiné Apple, Intel, Toshiba a další.

Pohled na situaci:

V současné době se nacházíme v předělu mezi dvěma zásadními etapami. První bylo dominantní postavení MS OFFICE na trhu a nyní došlo k obrovskému posunu směrem k otevřenosti a standardizaci. Je velmi důležité, aby se státní správa tuto situaci uvědomovala a byla připravena se jí přizpůsobit, respektive ji využít k prospěchu svému a svých zákazníků – občanů.

ODF byl přijat jako standard ISO, lze jej běžně používat, nicméně jeho rozšíření je podstatně menší než u Office. Naproti tomu u Office dochází k podobnému vývoji a prosazování vlastního otevřeného formátu, ale se značným zpožděním, neboť MS OFFICE 2007, která mají tento formát používat budou na trh uvedeny až začátkem roku 2007 a podle doporučení analytické společnosti Gartner [5] má Microsoft jen malou šanci získat ISO standard na svůj nový formát Open XML. "Budoucnost navrhovaného formátu je nejistá. Microsoft jej zatím na konci roku 2005 odeslal na schválení ECMA," říká analýza společnosti Gartner. "Microsoft plánuje po schválení ECMA svůj formát předložit ISO. Ale je pravděpodobné, že ISO neschválí více podobných XML formátů."

Tím, že se ODF stalo standardem ISO, bude mnohem snadnější situace pro složky veřejné správy. Samotná Evropská unie i ministerstva po celém světě čekala jaký formát získá označení ISO standard – nyní by tyto složky měly začít tento formát podporovat a dodavatelé softwaru integrovat podporu do všech svých aplikací. Samozřejmě však nejde o žádné nařízení, nelze tedy očekávat, že ODF přijme i konkurence.

Microsoft versus ODF

V odborném tisku [9] se objevila zpráva, že pod tlakem mnoha státních správ se společnost Microsoft rozhodla sponzorovat open source projekt, jehož cílem bude vývoj nástroje pro konverzi Open XML formátu Office 2007 do formátu ODF. Nová verze Microsoft Office tak bude rovněž podporovat konvertor, který jí umožní ukládat přímo do ODF.

Tento projekt se jmenuje „Open XML Translator“. Finální verze pro Word 2007 bude uvolněna koncem roku 2006, překladače pro Excel a PowerPoint 2007 potom počátkem roku 2007. Pro starší verze Microsoft Office bude podpora OpenDocument rovněž k dispozici, stejně jako podpora Open XML.

Podpora ODF je ze strany Microsoftu překvapivá, neboť dlouho tvrdil, že ji neplánuje kvůli nezájmu zákazníků a nutnosti zpětné kompatibility se staršími verzemi Office. Microsoft stále věří, že Open XML se stane světovým standardem, ale rozhodl se dát zákazníkům možnost volby.

Závěr:

Otázka, který formát nakonec vyhraje, je však stále otevřená. Rozhodující je samozřejmě množství uživatelů, kteří daný formát přijmou a budou používat. Zde má jistě Microsoft silnou převahu.

Ať už totiž zvítězí pouze jediný z konkurentů (málo pravděpodobné) nebo budou k dispozici možnosti využívat více formátů daného typu (mnohem pravděpodobnější), dá se jednoznačně předpokládat, že výrobci aplikací budou podporovat všechny vážné aktivity daného typu, neboť tak přináší konkurenční výhodu ve formě podpory technologie, která dokáže vylepšit aplikaci.

Je pravděpodobné, že nedojde k jednoznačné emigraci uživatelů od MS Office k jiným kancelářským balíkům, ale spíše o postupný a pomalý vývoj, který ovlivní jak Open Source aktivity, tak i samotný Microsoft a na trhu si budou oba dál vzájemně konkurovat.

Z hlediska státní správy se dá se však předpokládat, že pokud by došlo k rozhodnutí o sjednocení formátů, byl by zřejmě zvolen formát ODF a pro toto použití ho lze rozhodně doporučit.

Ať již získá převahu kterýkoliv z daných formátů, zaměstnanci státní správy, firmy, či soukromé uživatele v každém případě může těšit, že tento proces je prospěšný a rozhodně přináší vyšší standardizaci v dané oblasti a tedy lepší možnosti výměny dat mezi aplikacemi

i institucemi.

Literatura:

- [1] Molčan F. Fakta k formátu Open Document. OSS Alliance, 02.07.06
- [2] Wikipedie, otevřená encyklopedie, <http://cs.wikipedia.org/wiki/ODF>, [10.08.06]
- [3] Microsoft, Tisková zpráva „XML bude výchozím formátem souborů v Microsoft Office 12“, 03.06.05, http://www.microsoft.com/cze/presspass/msg/20050603_news1.mspx, [10.08.06]
- [4] Office Development Team, Microsoft® Office „12“ Beta 1, listopad 2005
- [5] Gartner group, <http://www.gartner.com>, [10.08.2006]
- [6] OASIS, Specifikace formátu ODF, Open Document Format for Office Applications v1.0, OASIS Standard, 1 May 2005, <http://docs.oasis-open.org/office/v1.0>
- [7] Microsoft, Tisková zpráva „Microsoft podporuje otevřené formáty dokumentů Office“, 30.11.05, http://www.microsoft.com/cze/presspass/msg/20051130_news1.mspx, [10.08.06]
- [8] IDABC - European eGovernment services, Promotion of Open Document Exchange Format, <http://europa.eu.int/idabc/en/document/3428/556>, [10.08.06]
- [9] Ecma International, specifikace formátu Open XML, červen 2006, http://www.ecma-international.org/news/TC45_current_work/TC45_available_docs.htm, [10.08.06]
- [10] Nate Mook, Microsoft to Support OpenDocument, 06.05.06, <http://www.betanews.com>, [08.07.06]

Poděkování:

Tento článek vzniknul s využitím materiálů a souhlasem OSS Aliance, která se prosazuje o šíření Open Source Software v České republice a je pod záštitou Ministerstva informatiky ČR a FJFI ČVUT, viz <http://www.oss.cz>

Kontaktní adresa:

Ing. Milan Tomeš

Ústav systémového inženýrství a informatiky, Fakulta ekonomicko-správní,

Univerzita Pardubice, Studentská 84, 532 10 Pardubice

tel.: 466 036 147

e-mail: milan.tomes@upce.cz