

Öppna dokumentformat ger mervärden

Predrag Mitrovic, CTO Microsoft Sverige
predrag.mitrovic@microsoft.com
<http://blogs.msdn.com/swesecurity>

Tid	Skriftspråk
Ca. 30.000 år f. kr.	Grottmålningar som kan ses som ett berättande av en historia eller en ritual.
Ca. 4.100 – 3.800 år f. kr.	Sumeriska symboler som utvecklades från att vara bildsymboler till att vara symboler som representerar ljud.
Ca. 3.200 år f. kr.	Egyptiska hieroglyfer som symbolspråk
Ca. 1.700 år f. kr.	Förtegen till dagens alfabet hittas i gruvor i Sinai. Ett "alfabete" med 30 tecken som representerar konsonanter.
Ca. 1.200 år f. kr.	Fenicianska alfabetet. Utifrån det bildades grekiska alfabetet (idag vårt latín), hebreiska och aramaiska (arabiska/persiska språk) språkrädet.
1922	DIN 476 (Deutsches Institut für Normung) specificerar en internationell standard för pappersstorlekar som blir ISO-standard ISO 216. Här anges pappersstorleksstandarderna A och B (A4, B5 till exempel). Standarden grundas på tankar och idéer från tiden för franska revolutionen.
1950-talet	IBM's Binary-Coded Decimal (BCD) omvandlar decimala siffror till binärkod genom 4 bitar.
1962/1963	IBM's Extended Binary Coded Decimal Interchange Code (EBCDIC) är ett teckenomvandlingspråk som omvandlar tecken till binärkod genom 8 bitar.
1967	American Standard Code for Information Interchange (ASCII) är ett teckenkodningssystem baserat på skrivna engelska genom 7 bitar. Ur ASCII skedde en explosionsartad utveckling för att tillhandahålla tecken för icke-engelska språk.
1986	Standard Generalized Markup Language (SGML) som ISO-standard. SGML är ett så kallat metaspråk som kan användas för att beskriva dokument.
1991	Unicode blir industristandard som gör att text i alla format och språk kan kodas för att användas av datorer.
1993	Hyper-Text Markup Language (HTML) presenteras som ett arbetsdokument av IETF. HTML blev standarden för att beskriva webbsidor på Internet.
1998	XML är ett beskrivningsspråk som kan beskriva mängder av olika data. XML är utmärkt att dela data mellan olika system på Internet, därav dess lönsamhet som dokumentformat. Om HTML beskriver webbsidor kan XML ses som beskrivande språk för data som flödar över Internet.

Filetyp	Enkel Googlesökning under .SE-domänen sep-2006*	
htm + html	737,000,000	
	Dokumentformat	Procent
PDF	99,400,000	81% Adobe
DOC	13,400,000	10.8% MS Word
txt	4,260,000	3.4% Textformat
RTF	2,200,000	1.8% Rich Text Format
ppt	2,200,000	1.8% MS Powerpoint
xls	1,590,000	1.3% MS Excel
PS	689,000	0.5% Postscript
wpd	144,000	0.1% Corel Wordperfect
wp	37,800	0.0% Wordperfect
sxw	16,300	0.0% Star/Openoffice Writer
tex	892	0.0% Tex
sxi	816	0.0% Star/Openoffice Presentation
wp5	204	0.0% WordPerfect 5
sxc	98	0.0% Star/OpenOffice Calc
.latex	4	0.0% LaTeX
Total Dok	123,939,114	
Totalt	860,939,114	

*.se filetype: htm

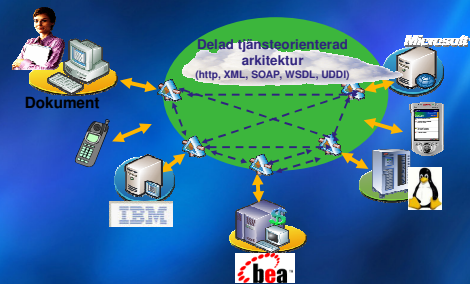
Enkel analys - utfall

- Mycket information har skapats/fortsätter att skapas. IT en möjliggörare i processen.
- Ett dokumentbibliotek (webbplats) kan använda, och använder uppenbarligen, många olika format → komplexitet att dela, utväxla och bevara information.
- Mängder av olika applikationer används för informationsarbete.

Kravlista från verkligheten

- Valfrihet:**
 - Organisationen väljer format, schema, programvaror och plattformar.
- Kompatibilitet:**
 - Interoperabilitet eller prestanda = fel fråga! Interoperabilitet med bibehållen prestanda, kvalitet, data och återgivning
- Gemensamhet:**
 - Interoperabilitet med vad jag har, utan att behöva riva upp och ersätta!
- Helhetssyn:**
 - Interoperabilitet över system, applikationer, användare, data och program.

Vision realiseras med XML

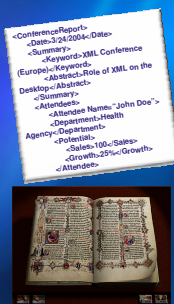


Vad är ett dokumentformat?

- Format innehåller och beskriver information
- Metod att hålla och presentera det intellektuella innehållet
- Information om innehållspresentation

Applikation & information – relationen

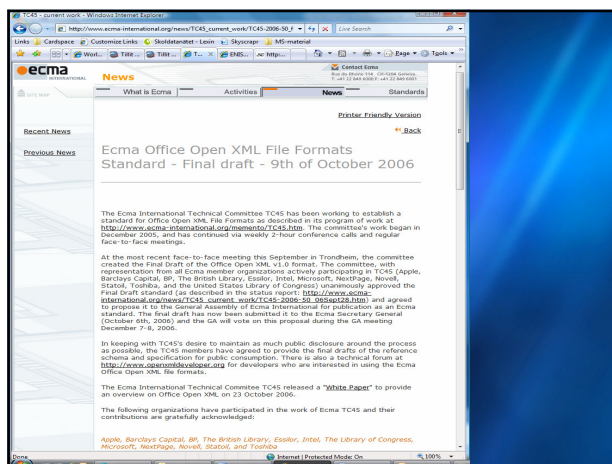
- Intelligensen att förenkla skapande
- "Artists" verktyg för att presentera innehåll
- Applikationer kopplas hårt, löst eller inte alls mot format



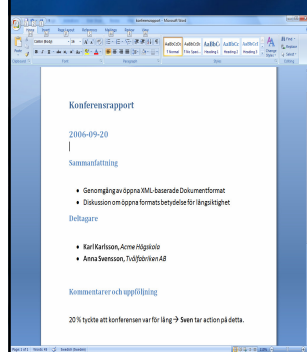
Open XML – vad är det?

Öppet, royaltyfritt och interoperabelt:

- ECMA (European Computer Manufacturers Association) ansvarar för underhåll och vidareutveckling.
- Standardisering i ECMA under principen "en organisation, en röst" = seriös standardisering.
- ECMA's mål med Open XML = bred implementation av verktyg på många plattformar för högsta interoperabilitet.



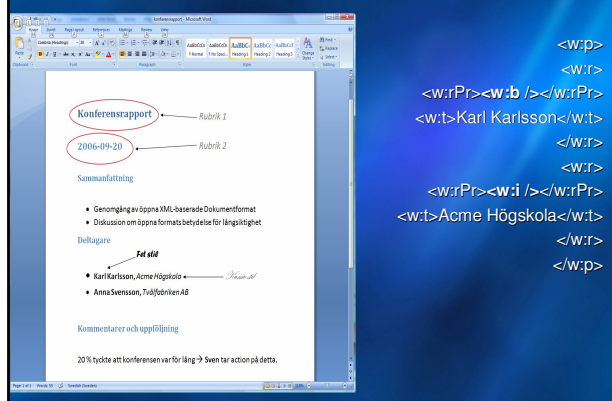
XML-schema definierat



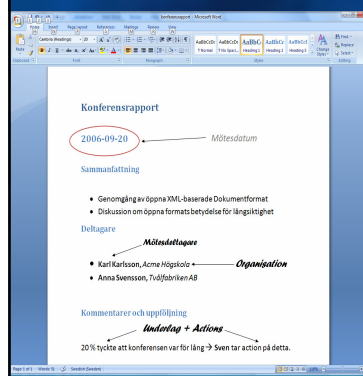
XML Referensschema

- Presentationsorienterat (fetstil, kursiv, tabell, paragrafer)
- Öppet dokumentformat
- Medger interoperabilitet för arkivering och filformat
- Kunddefinierat schema
 - Dataorienterat (pris, faktura)
 - Representerar affärsinformationen som lagras i dokumentet
 - Medger systemintegration

XML referensschema



Kunddefinierat schema



```
<Konferensrapport>
<Date>2006-09-20</Date>
<Summary>
  <Keywords>XML-konferens (Europe)</Keywords>
  <Abstract>Öppna dokumentets betydelse</Abstract>
</Summary>
<Attendees>
  <Attendee Name="Karl Karlsson">
    <Department>Acme Högskola</Department>
    <Evaluation>
      <Value>20%</Value>
    </Evaluation>
    <Action>Tilldelat till Sven tar action på detta.
  </Attendee>
```

Förändring = utmaning



- Miljoner användare har skapat miljarder av dokument under de senaste 20 åren
- Användare kräver bakåtkompatibilitet med existerande dokumentformat och produkter
- Migrering till XML bör inte kompromissa prestanda eller potentiell funktionalitet

ODF och Open XML

- Sun's format, designat för OpenOffice (Open Office XML Format)
 - Sun, IBM, Novell med flera pushar för obligatoriska/bindande ODF-policybeslut för att saluföra deras produkter och tjänster
- Trots ovan opponerar sig inte Microsoft mot ODF på något sätt
 - ODF är ett av många format för många användningsområden
- Microsoft stöttar insatser för datainteroperabilitet med ODF
 - Via världens bästa tredjepartsutvecklare av XML-format i Frankrike (topp-projekt på sourceforge.net)
- Val mellan format och interoperabilitetsaspekten driver produktkonkurrensen; användare undviker inläsning med öppna format, förutsatt att schemat inte läser in dig!



- www.OpenXMLDeveloper.org
- Formerat för att tillhandahålla ett forum där utvecklare fritt kan arbeta med informationsutbyte, bolla idéer, teknisk support, tips och verktyg kring Open XML-format.
- Öppet för alla utan kostnad, allt för att möjliggöra en bred uppslutning och samverkan så att lösningar som använder Open XML kan skapas för alla plattformar.
- Över 500 medlemmar är med!

Summering

- Öppna XML-baserade dokumentformat är framtiden
- MS ansatser är öppna, och blir en industristandard
 - Öppet för branschmöjligheter
 - Öppet för konkurrerande alternativ
 - Öppet till community- och standardiseringsgrupp att förädla
- MS Open XML-format ger all funktionalitet utan "inläsning"
- Microsoft som företag står för branschövergripande interoperabilitet och tillväxt genom XML-format

Tack för ordet så här långt.

Predrag Mitrovic
predrag.mitrovic@microsoft.com
<http://blogs.msdn.com/swesecurity>