

# ready for a new day

for Developers

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Developing with the OpenXML File Format

Chewy Chong

Developer Evangelist – Singapore

[chewyc@microsoft.com](mailto:chewyc@microsoft.com)

# Agenda

- Introducing OpenXML
- Inside the Formats
- Separating “Data” from “View”
- Using .NET to program against OpenXML

# Quick Vote

Who here really cares?

Mass. to use Microsoft Office in ODF plan | CNET News.com - Windows Internet Explorer

http://news.com/Mass.+to+use+Microsoft+Office+in+ODF+... government chooses OpenDocument

In memoriam: James Kim, CNET senior editor, 1971 - 2006

**cnet NEWS.com**

Search:

Today on CNET | Reviews | News | Downloads | Tips & Tricks | CNET TV | Compare Prices

Today on News | Business Tech | Cutting Edge | Access | Threats | Media 2.0 | Markets | Digital Life | Blogs | Extra

## Mass. to use Microsoft Office in ODF plan

By Martin LaMonica  
Staff Writer, CNET News.com  
Published: August 24, 2006, 7:22 AM PDT

TalkBack | E-mail | Print | del.icio.us | Digg this

Massachusetts will begin using OpenDocument as the document format later this year as planned, but it will be with Microsoft Office in the near term, the state's top technology executive said.

As expected, Louis Gutierrez, chief information officer of Massachusetts' Information Technology Division, on Wednesday sent a letter seen by CNET News.com to advocates of people with disabilities. The letter was in response to their concerns about the commonwealth's plan to move to the OpenDocument format standard.

In addition, Gutierrez last week wrote to the state's Information Technology Advisory Board with an update on the OpenDocument format implementation plan, as had been planned.

Last year, Massachusetts caught international attention for its decision to standardize by January 2007 on ODF, a document format standard not supported in Microsoft Office.

**Disability unfriendly?**  
The move was criticized by disability-rights groups, which said that going to ODF-compliant products, such as the open-source OpenOffice suite, would not adequately address their needs. In general, Microsoft Office has better assistive technologies,

Open document man resigns from Mass CTO post | Channel Register - Windows Internet Explorer

http://www.channelregister.co.uk/2005/12/29/mass\_odef\_cio/ Boston Globe Office Mass

**CHANNEL REGISTER**

HOME | DISTRIBUTION | HARDWARE | SERVERS/STORAGE | PC BUILDER | MARKET NEWS | RESELLER | SECURITY | SERVICES | SOFTWARE

By Ashlee Vance in Mountain View 29 Dec 2005 21:53  
More by this author

## Open document man resigns from Mass CTO post

Tired of being a lightning rod

Download Desktop News Alerts | Get Channel News on your mobile | Channel Newsletter Sign up here

RTFM: sound thinking for the modern masses

Click here to download first episode

Internet | Protected Mode: Off | 100%

Corel to support Microsoft Office, ODF formats | Tech News on ZDNet - Windows Internet Explorer

http://news.zdnet.com/2100-3513\_22-6139228.html Office Document Formats

**ZDNet** Where Technology Means Business

HOME | NEWS | BLOGS | WHITE PAPERS | DOWNLOADS | REVIEWS

Security | Networking | Hardware | Software | Web Technology | IT Management | OSes | All News

home / ZDNet news

## Software

SEARCH

## Corel to support Microsoft Office, ODF formats

By Martin LaMonica, CNET News.com  
Published on ZDNet News: November 29, 2006, 6:40 AM PT

TALKBACK ADD YOUR OPINION

Worthwhile?   + 0 0 VOTES

ZDNet Tags: Microsoft Office, XML,

Corel, the maker of the WordPerfect word processor, said that it intends to support Microsoft's latest Office document formats and its rival, OpenDocument.

The company intends to add both formats to its WordPerfect Office suite in the middle of next year.

On Thursday, Microsoft is set to release to businesses Office 2007, an upgrade of its productivity suite which introduces an XML-based file format called [Office Open XML](#).

Because so many more documents will be created in that format, Corel has decided to make opening and editing those document types an option in the WordPerfect word processor as well as in the company's presentation application and Quattro Pro spreadsheet, said Richard Carriere, general manager of office productivity for Corel.

In addition, Corel in the middle of next year will allow people to open and view word processor documents stored in the OpenDocument format, or ODF.

INTRODUCING THE HP BLADESYSTEM

The HP ProLiant BL460c server blade features Dual Core Intel® Xeon® Processors

Take a peek under the hood

hp » Learn more

### LATEST ENTERPRISE CONTENT

- Hillary Clinton, Joe Lieberman help launch game-rating ads
- Oracle ups bid for India's i-flex
- Microsoft sails through document standard vote



# Why You Should Care...

- Practically..
  - Do I have all the tools required to access it?
  - How much data do I have?
- Automation..
  - Can I program against it?

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# INTRODUCING OPENXML

for Developers

 Office Microsoft®

 Windows Vista™

Microsoft®  
Exchange Server 2007

# Office Open XML Formats

- New XML file formats for Microsoft Word, Excel and PowerPoint
  - New formats will be the **default** file formats, w/ new file type extensions (.docx; .pptx; .xlsx)
  - Fully 100% compatible with existing formats
- Open, transparent format improves interoperability
  - XML - Transparent, XML format enables new integration scenarios for documents and LOB systems
  - ZIP container - allows for standard compression on all files without user effort
  - Licensing - Removed need for license by providing a Covenant that says we won't enforce IP against folks implementing the format (100% royalty free)
- Standardization
  - Ecma International - created TC45 to fully document the Open XML formats
    - Members include: Apple, Barclays Capital, BP, the British Library, Essilor, Intel Corporation, NextPage Inc., Statoil ASA and Toshiba
    - Current spec is already over 2000 pages



# Office Open XML Formats

- Added Benefits: compact and robust
  - ZIP container allows for standard compression on all files without user effort (Dramatic file size improvements)
  - Significantly more robust files to help minimize data loss
- Backward Compatible: Microsoft Office 2000, Microsoft Office XP, Microsoft Office 2003
  - Patches for compatibility available by launch
  - Open, edit and save new formats
- Legacy support: Current Microsoft Office 97-2003 binary file formats supported
  - Support for XML formats from Microsoft Office 2003, Microsoft Office XP continued
- Developers: Endless potential for developers
  - Build solutions to read, write, and modify Microsoft Office files (without the need to run Microsoft Office APIs)

# The Role of XML with Documents

Scenario	Example
<b>Document Assembly</b> Server-based or user-assisted construction of documents from archived content or database content	Create sales reports from financial and forecast data stored in a CRM system
<b>Content Reuse</b> Much easier to move content between documents, including different document types	Apply content stored in Word documents to Web pages quickly and efficiently
<b>Content Tagging</b> Add domain-specific metadata to document content to enable custom solutions	Tag presentations using a specific taxonomy to improve knowledge management efficiency
<b>Document Interrogation</b> Query document repositories based on custom data, content types or document metadata	Search for all documents containing a specific company name or sales contact
<b>Document Sanitization</b> Remove unwanted content like comments or embedded code from your document when appropriate	Remove all tracked changes and comments from a Word document before it is published

# Open XML Formats Architecture



User view:  
Single file

Questionnaire  
.docx



Developer view:  
Modular file

## File container

### Document Parts

- Most parts are XML
- Each XML part is a discreet, compressed component
- Can add, extract and modify individual parts without using Office programs
- Corruption or absence of any part would not prohibit the file from being opened

Document properties

Comments

WordML/SpreadsheetML, etc.

Custom-defined XML

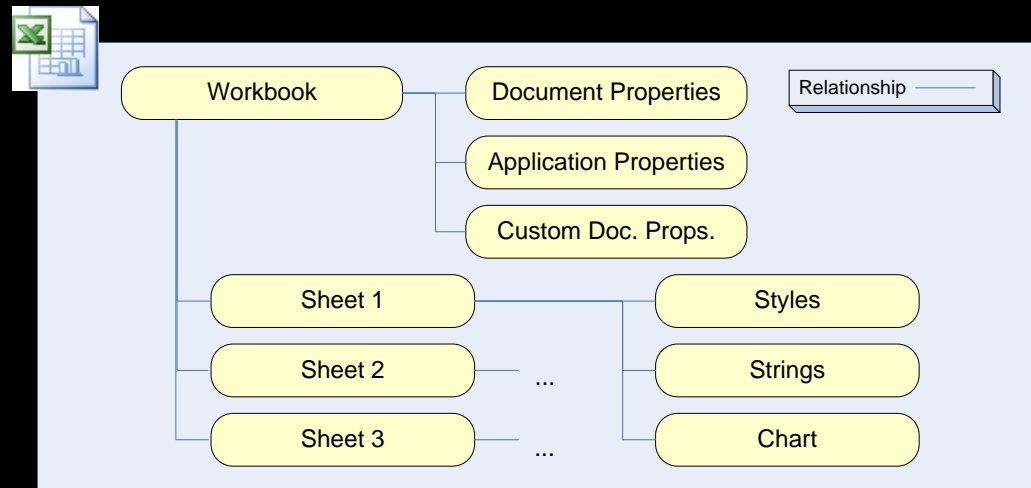
Images, video, sound

Embedded code/macros

Charts

# Components of the New Formats

- Package – ZIP Container
- Part – The “files” inside the ZIP
- Content Types – Each part has a content type that is enforced on open
- Relationships – Any part that references another part must do so via a relationship



A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Demo

Poking around a sample  
OpenXML document

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Demo

## Quick Sanitization

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# INSIDE THE FORMATS

for Developers

 Office Microsoft®

 Windows Vista™

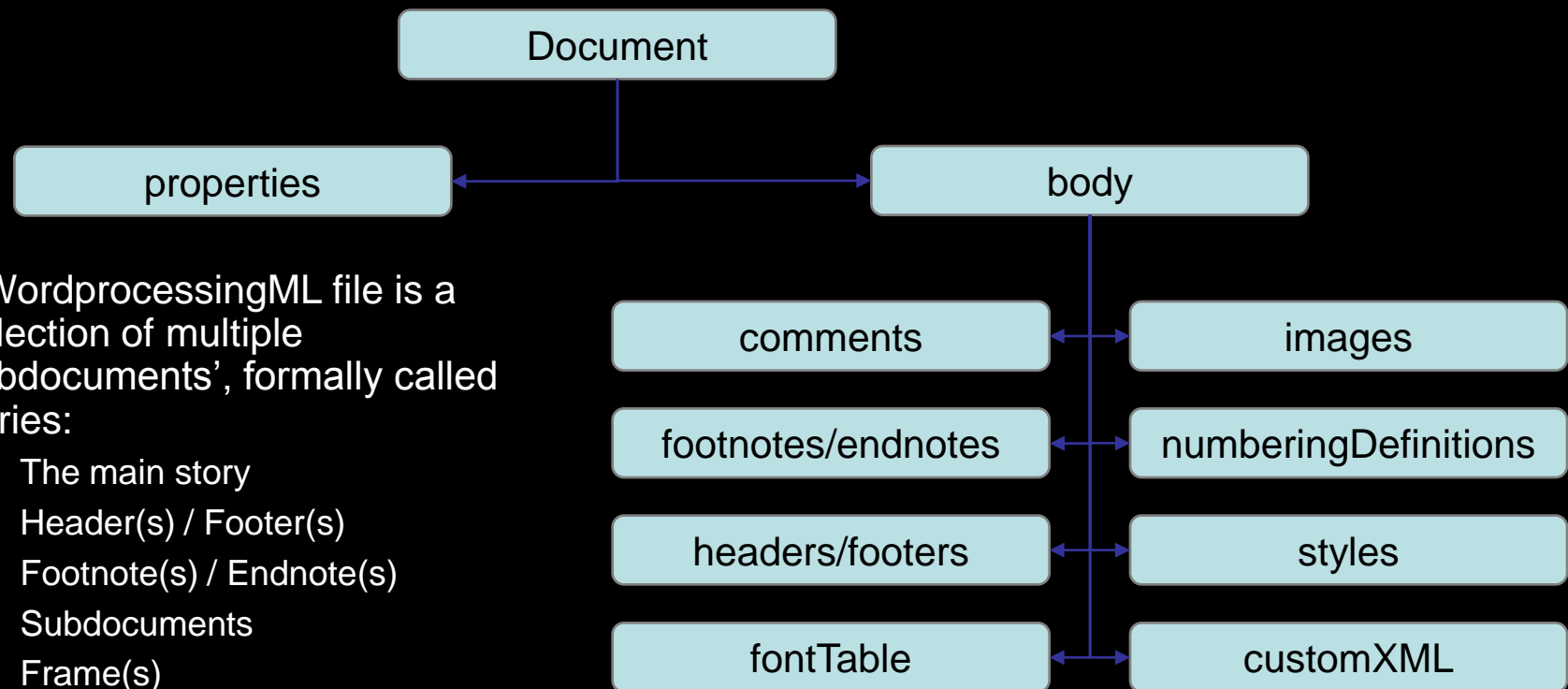
Microsoft®  
Exchange Server 2007

# Design Strategies

- Don't reinvent the wheel
  - Utilizes the WordprocessingML 2003 file format constructs
  - Developers familiar with WordprocessingML 2003 will find the formats very similar
- XML everywhere
  - Any place where we don't need binary (e.g. image storage), don't have it
  - Binary only where it's an object we don't 'own'
- Break out parts where it benefits developers and end users
  - Example: Move comments into separate part (clear at the firewall)
  - Example: Move style definitions into separate part (change the style sheet easily)
  - Example: If one area of the document is corrupted, the rest of the document can still be recovered



# WordprocessingML



- A WordprocessingML file is a collection of multiple 'subdocuments', formally called stories:

- The main story
- Header(s) / Footer(s)
- Footnote(s) / Endnote(s)
- Subdocuments
- Frame(s)
- Comment(s)

# SpreadsheetML

Workbook

properties

styles

sharedStrings

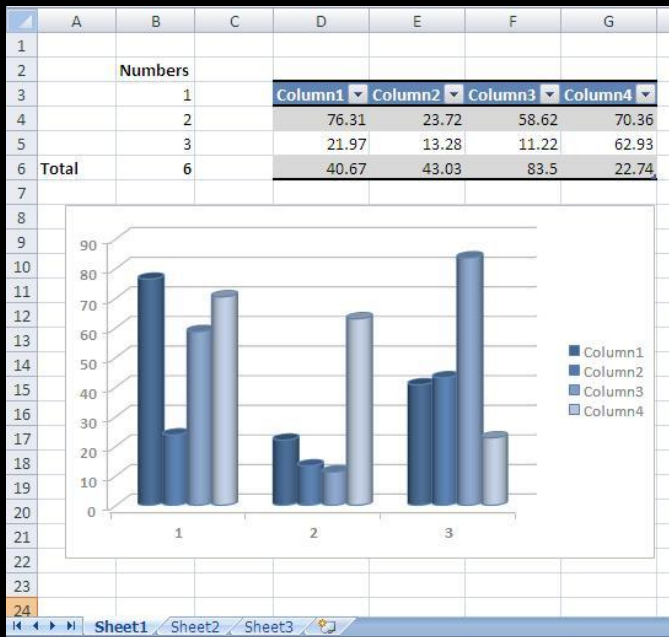
calcChain

sheet1..N

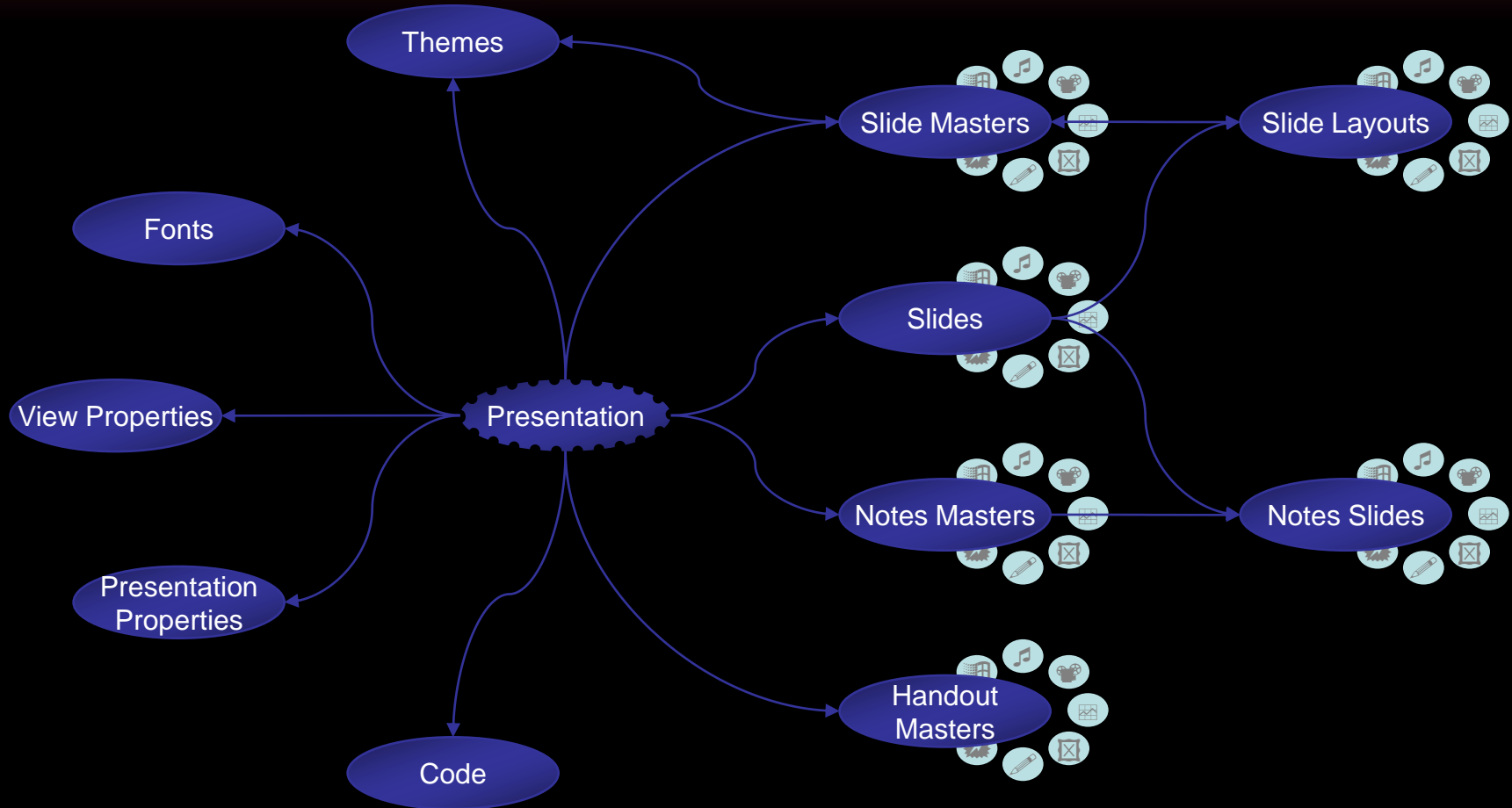
table

chart

drawing



# PresentationML



A sunset scene with a bright sun low on the horizon, casting a warm orange glow over a range of dark, silhouetted mountains.

# SEPARATING “VIEW” FROM “DATA”

for Developers

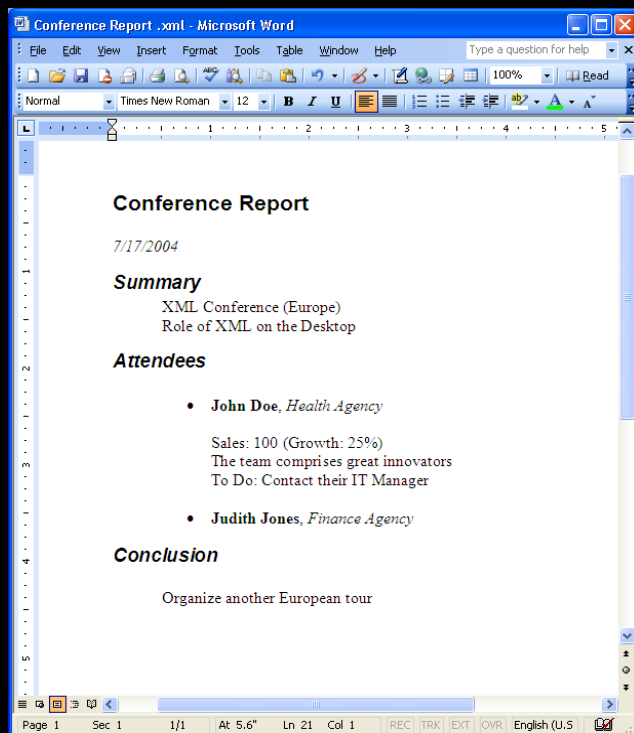
 Office Microsoft®

 Windows Vista™

Microsoft®  
Exchange Server 2007

# The Role of XML

## Reference and Custom-defined Schemas



### XML Reference Schemas

- Display-oriented (for example, Bold, Italics, Tables, Paragraphs, Styles)
- Open Document Format
- Enable Archival and File Formats Interoperability

### Custom-defined Schemas

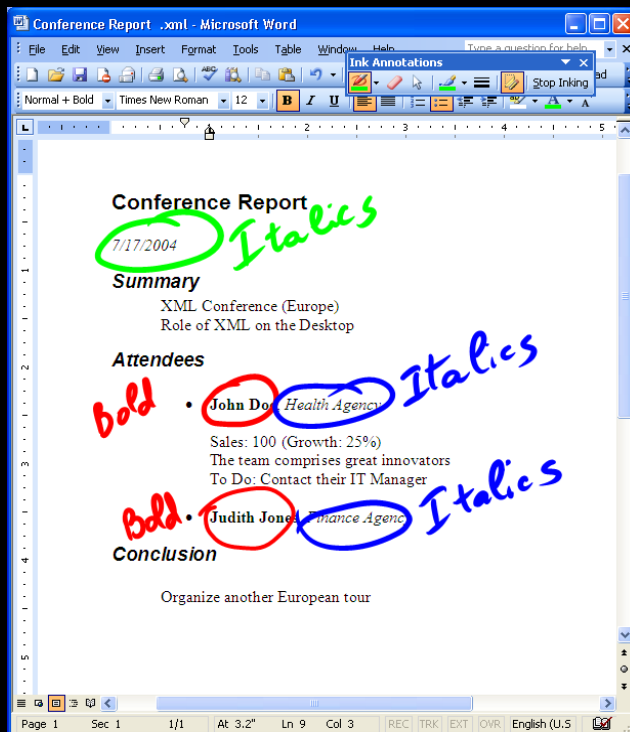
- Data-oriented (for example, Price, Invoice)
- Represents the business information stored in the document
- Enable System Integration

# The Role of XML

## Reference and Custom-defined Schemas

### XML Reference Schemas

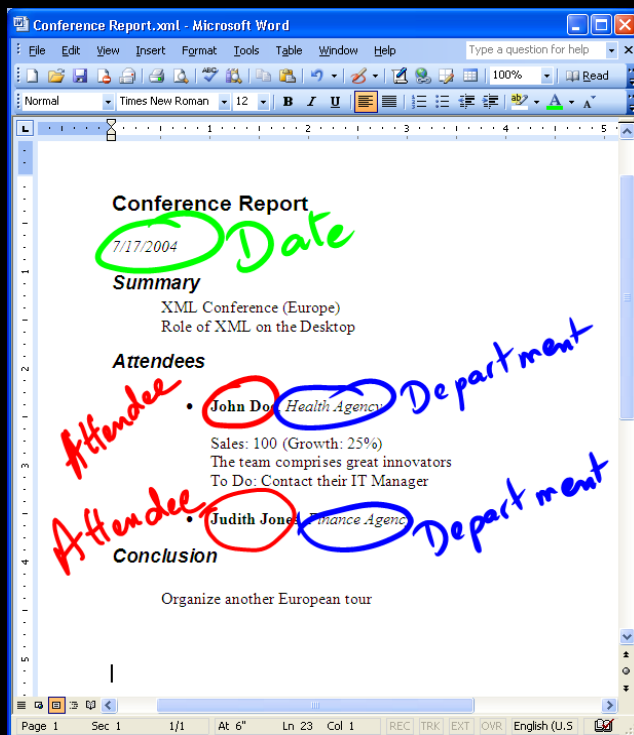
- Display-oriented (for example, Bold, Italics, Tables, Paragraphs, Styles)
- Open Document Format
- Enable Archival and File Formats Interoperability



```
<w:p>
  <w:r>
    <w:rPr><w:b /></w:rPr>
    <w:t>John Doe</w:t>
  </w:r>
  <w:r>
    <w:rPr><w:i /></w:rPr>
    <w:t>Health Agency</w:t>
  </w:r>
</w:p>
```

# The Role of XML

## Reference and Custom-defined Schemas



```
<ConferenceReport>
  <Date>3/24/2004</Date>
  <Attendees>
    <Attendee Name="John Doe">
      <Department>
        Health Agency
      </Department>
      <Potential>
        <Sales>100</Sales>
        <Growth>25%</Growth>
        ...
      </Attendee>
  </Attendees>
</ConferenceReport>
```

## Custom-defined Schemas

- Data-oriented (for example, Price, Invoice)
- Represents the business information stored in the document
- Enable System Integration

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Demo

## Leveraging “Custom Schema” to Customize Document



A scenic background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# USING .NET TO ACCESS OPENXML

# Tools for Accessing Data In Office Open XML files

- XML Editing
  - System.XML
- ZIP Manipulation
  - WinFX System.IO.Packaging
- Office Open XML Resource Kit
  - Code Snippets - Beta 2
    - C# and VB.NET
  - Validation Library
    - Parses a file and reports on schema, relationship errors and warnings
  - Serialization/Deserialization Library
    - Flattens package into a single file for ease of development in simple construction scenarios
- WinFX – System.IO.Packaging

# System.IO.Packaging API Functionality

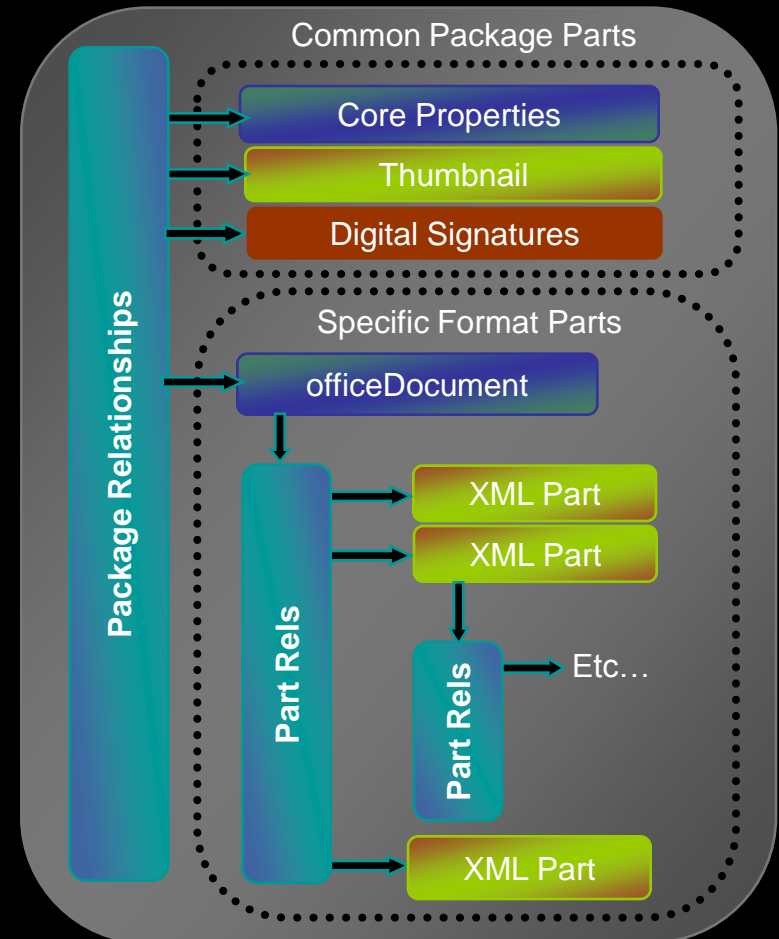
- Create/Open packages
- Create and delete parts and relationships
- Read and write part streams
- Iterate through collections of parts and relationships

Package  
PackagePart  
PackagePartCollection  
PackageRelationship  
PackageRelationshipCollection  
PackUriHelper

# System.IO.Packaging.Package

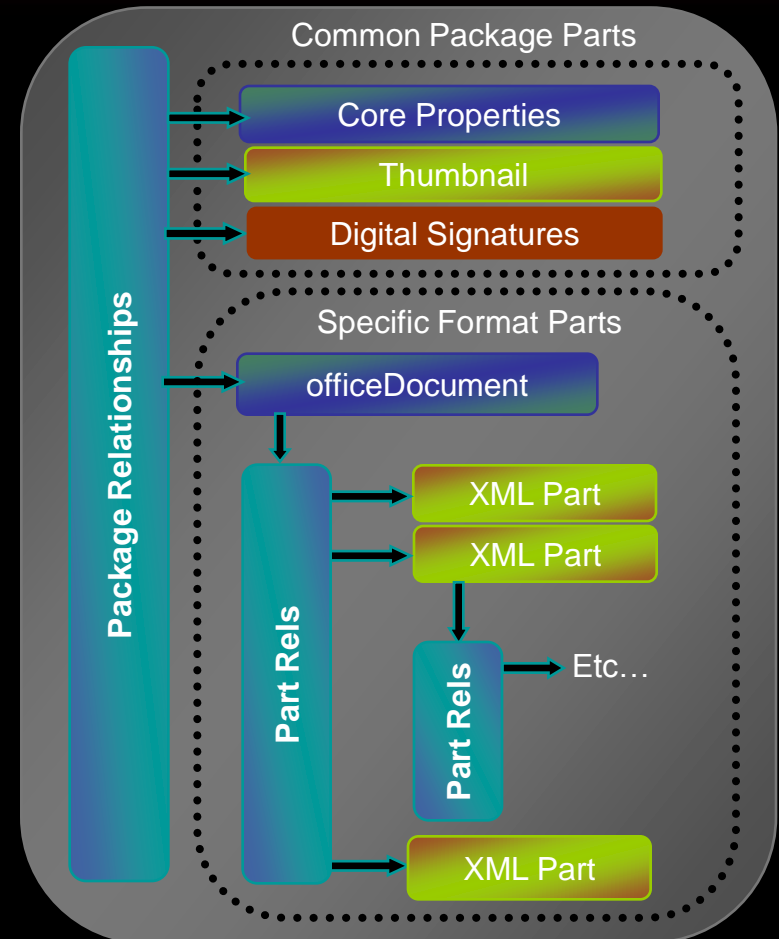
- Package class provides methods to create, enumerate and delete the following entities:

- Package
- Package Relationships
- PackageProperties
- Parts



# System.IO.Packaging.Relationship

- Relationships tie the parts together
- Required to find parts (part names are not guaranteed)
- Iterate through RelationshipCollection by Type or ID
- Relationship Properties:
  - ID
  - Package
  - RelationshipType
  - SourceUri
  - TargetMode
  - TargetUri



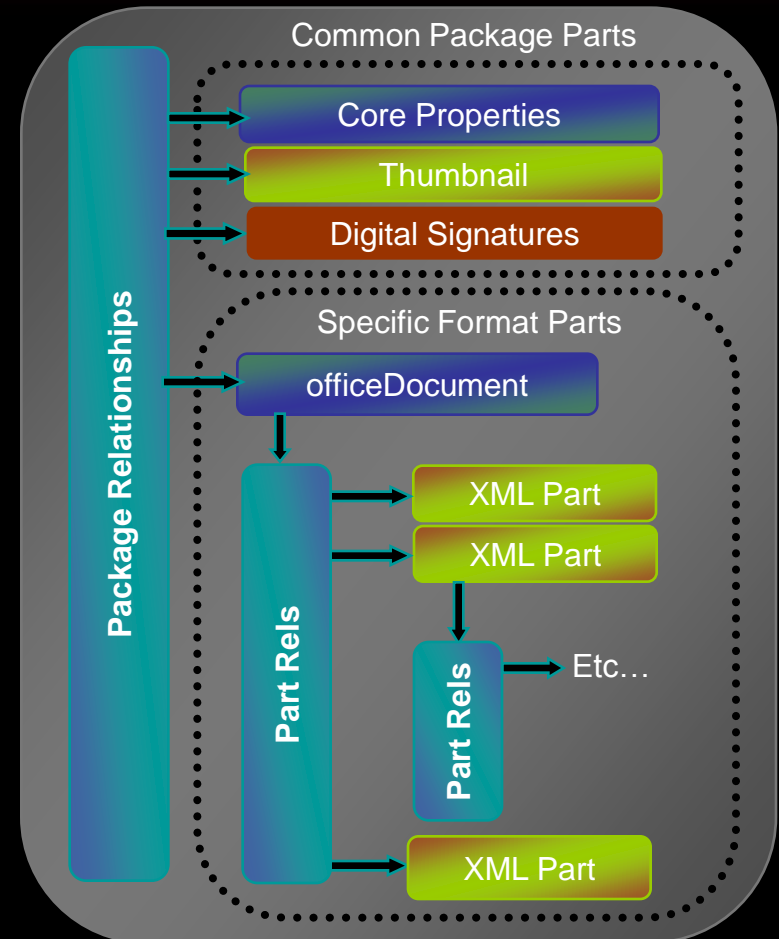
# System.IO.Packaging.PackagePart

- Parts are the objects of data within the

```
<w:body>  
  <w:p w:rsidR="001B7EF4"  
    w:rsidRDefault="001B7EF4">  
    <w:r>  
      <w:t>The Cow jumped over the  
moon.</w:t>  
    </w:r>  
  </w:p>...
```

- PackagePart Properties:

- CompressionOption
- ContentType
- Package
- Uri



A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Demo

## Using the API to Automate Boring Things

# Resources

- Office Preview Site: <http://www.microsoft.com/office/preview/>
- Brian Jones's Blog: [http://blogs.msdn.com/Brian\\_Jones/](http://blogs.msdn.com/Brian_Jones/)
- Kevin Boske's Blog: <http://blogs.msdn.com/KevinBoske/>
- Office 2003 Reference Schema Information:  
<http://www.microsoft.com/office/xml/>







- New community formed to help bring developers together
  - Currently sponsored by almost 40 institutions from around the world
- Community and Web site for information exchange
  - Code Snippets; Tools; Discussions
- Free of charge. Available to everyone that wants to participate and/or encourage development on all platforms.
- Be one of the first to join the community!
- <http://openxmldeveloper.org>

A background image showing a bright sun setting behind a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# Q&A

for Developers

 Office Microsoft®

 Windows Vista™

Microsoft®  
Exchange Server 2007

A background image showing a bright sun setting or rising over a range of dark, silhouetted mountains. The sky is a gradient of orange and yellow.

# ***Microsoft***<sup>®</sup>

***Your potential. Our passion.***<sup>™</sup>

© 2006 Microsoft Corporation. All rights reserved. Microsoft, Windows, Windows Vista and other product names are or may be registered trademarks and/or trademarks in the U.S. and/or other countries.

The information herein is for informational purposes only and represents the current view of Microsoft Corporation as of the date of this presentation. Because Microsoft must respond to changing market conditions, it should not be interpreted to be a commitment on the part of Microsoft, and Microsoft cannot guarantee the accuracy of any information provided after the date of this presentation.

MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS PRESENTATION.

**for Developers**

 **Office** Microsoft®

 **Windows Vista** ™

Microsoft® **Exchange Server 2007**