

L'utilizzo di frontend basati su Microsoft Office utilizzando OOXML (Office Open XML)



Alberto Baroni

Raffaele Rialdi

Chi siamo

Raffaele Rialdi

- Developer dal 1987
- Microsoft MVP dal 2003
- Speaker, Trainer, Consulente
- Guida le attività di development di Vevy Europe
- Membro commissione ISO JTC1
 - (include Office Open XML)



Twitter: @raffaeler

Email: raffaeler@vevy.com

Website: <http://iamraf.net>

Alberto Baroni

- co-fondatore Sinergetica Srl
- Per 10 anni si è occupato di soluzioni IT in ambito energetico.
 - Ha collaborato alla progettazione e realizzazione di alcune delle più importanti soluzioni informatiche utilizzate dagli Operatori Energetici

Twitter: @Alkaloon

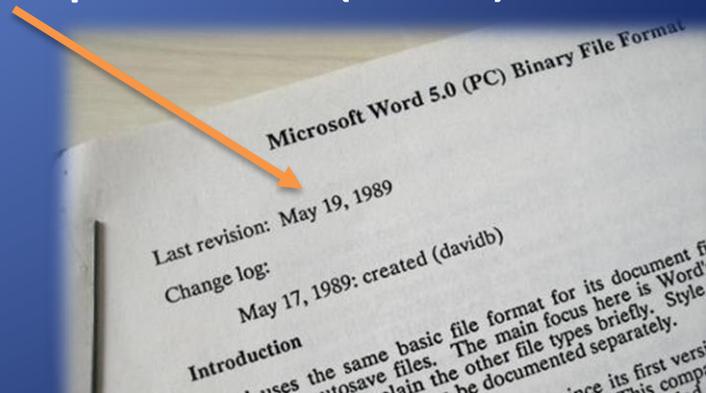
Email: alberto.baroni@outlook.com

Website: <http://www.sinergetica.biz>

OOXML

Office Open XML

- OOXML è standard ISO/IEC DIS 29500
- OOXML è standard ECMA-376:
 - <http://www.ecma-international.org/publications/standards/Ecma-376.htm>
 - Il pdf del solo reference è composto da 5000 pagine!
- Permette di generare documenti senza MS Office
 - I formati OOXML sono: docx, xlsx, pptx
- Il formato binario è sempre stato disponibile (NDA)
 - Oggi liberamente scaricabili:
<http://www.microsoft.com/interop/docs/OfficeBinaryFormats.mspx>



Microsoft e "openness"

- Dal 2006 Iniziativa "Open Specification Promise"
 - <http://www.microsoft.com/openspecifications/>
 - Un impegno a pubblicare gratuitamente le specifiche di tutti i protocolli interni usati in OS e prodotti

- Dal 2014 Una svolta in direzione open source
 - .NET Framework (BCL e CLR su GitHub)
 - .NET Core su Linux e Mac
 - <http://www.dotnetfoundation.org/>
 - Asp.net 5, EF, Owin, .NET Core, SignalR, OpenXml SDK, etc.

LA STRUTTURA DEL PACKAGE OPC

Il Package OPC è un file ZIP

- Open Packaging Conventions (standard ISO)
 - Il formato file è un normale ZIP
 - La struttura interna è definita nello standard

The image illustrates the internal structure of an OPC package (a ZIP file) for a Word document. It shows a WinRAR window titled "Raf.docx - WinRAR (evaluation copy)" displaying a ZIP archive with the following internal structure:

Name	Size	Pac...	Type
..			Folder
_rels			Folder
media			Folder
word			Folder
[Content_Types].xml	920	390	XML Document

Red arrows point from the WinRAR interface to detailed views of the internal file lists and XML content:

- Left View:** A file list showing image files:

Name	Size	Pac...
..		
image.jpg	28.575	42.591
image2.jpg	30.478	45.473
image3.jpg	34.275	50.418
- Right View:** A file list showing the internal structure of the _rels folder:

Name	Size	Pac...
..		
.rels	298	298
- Bottom View:** A file list showing the internal structure of the word folder:

Name	Size	Pac...
..		
_rels		
document.xml	9.013	1.906
footer.xml	2.000	734
header.xml	2.057	758
numbering.xml	1.008	505
settings.xml	544	353

The XML content shown in the bottom view is:

```
<?xml version="1.0" encoding="utf-8"?>
<Types xmlns="http://schemas.openxmlformats.org/package/2006/content-types">
  <Default Extension="xml" ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.document.main+xml" />
  <Default Extension="rels" ContentType="application/vnd.openxmlformats-officedocument.relationship+xml" />
  <Default Extension="jpg" ContentType="image/jpeg" />
  <Override PartName="/word/header.xml" ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.header+xml" />
  <Override PartName="/word/footer.xml" ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.footer+xml" />
  <Override PartName="/word/settings.xml" ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.settings+xml" />
  <Override PartName="/word/numbering.xml" ContentType="application/vnd.openxmlformats-officedocument.wordprocessingml.numbering+xml" />
</Types>
```

Il Package

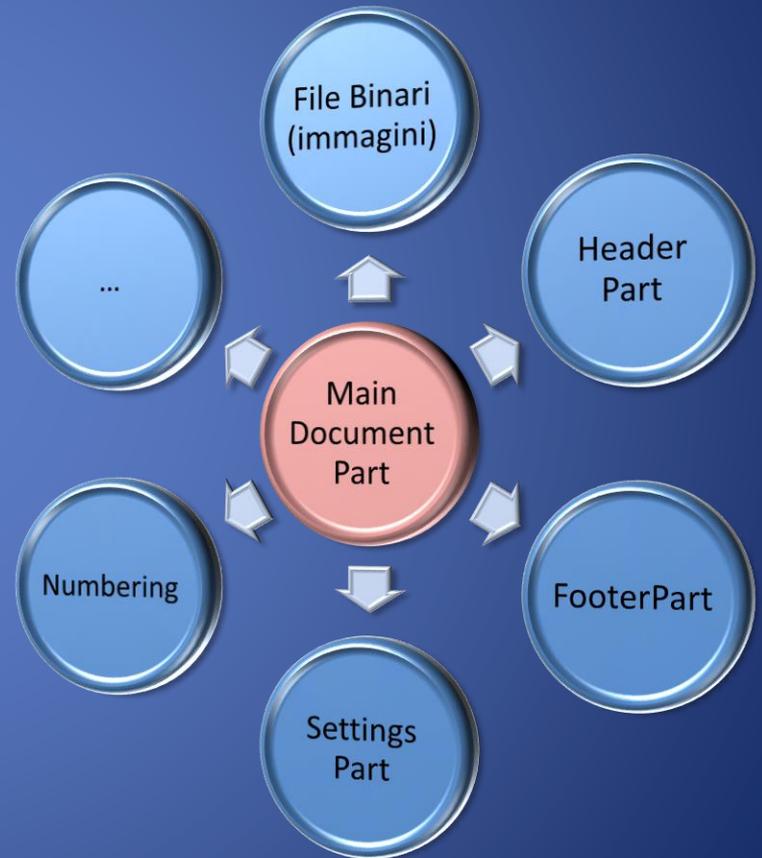
- È uno zip ma con una struttura definita nello standard
 - Lo stesso package usato per i file XPS
- Si può costruire a mano ...
- Il Framework.NET conosce il formato OPC
 - `System.IO.Packaging.Package`
- Office Open XML SDK è ancora più specifico
 - Conosce i contenuti 'speciali' di Office

WordProcessingML (Word Processing Markup Language)

DOCX

Struttura base di un DOCX

- Ogni "part" è un file XML nel package OPC
- La relazione tra part è definita tramite file di relazioni (XML)
- Immagini e altri file sono inclusi nello zip
- Ogni part ha un ID univoco



MainDocumentPart minimale

```
mainPart.Document = new Document(  
    new Body(  
        new Paragraph(  
            new Run(  
                new Text("Hello, world")  
            ))));
```



```
<w:document  
    xmlns:w= (namespace)  
<w:body>  
    <w:p>  
        <w:r>  
            <w:t>Hello, world</w:t>  
        </w:r>  
    </w:p>  
</w:body>  
</w:document>
```

Formattare il carattere

```
<w:r>  
    <w:rPr>  
        <w:b/>  
    </w:rPr>  
    <w:t>Grassetto</w:t>  
</w:r>
```

Bold

```
new Run(  
    new RunProperties(new Bold()),  
    new Text("Grassetto") );
```

```
<w:r>  
    <w:rPr>  
        <w:i/>  
    </w:rPr>  
    <w:t>Corsivo</w:t>  
</w:r>
```

Italic

```
new Run(  
    new RunProperties(new Italic()),  
    new Text("Corsivo") );
```

```
<w:r>  
    <w:rPr>  
        <w:highlight w:val="yellow"/>  
    </w:rPr>  
    <w:t>Evidenziato</w:t>  
</w:r>
```

Highlighted

```
new Run(  
    new RunProperties(  
        new Highlight(  
            { Val = HighlightColorValues.Yellow } ) ,  
        new Text("Evidenziato") );
```

Formattazione di paragrafo

- Giustificazione

- `<w:jc w:val="right"/>`
 - (left, center, both, ...)

```
new Paragraph(  
    new ParagraphProperties(  
        new Justification() { Val = JustificationValues.Right } ) )
```

- Spaziatura

- `<w:spacing w:after="0" w:line="240" w:lineRule="auto"/>`

```
new Paragraph(  
    new ParagraphProperties(  
        new SpacingBetweenLines() { After = 0, Line = 240, LineRule = LineSpacingRuleValues.Auto } )
```

- Conversioni

- Per i twips la conversione è `(int)(val * 1440)`
- La spaziatura si converte con `(int)(NumLines * 240.0m)`
- Le specifiche spiegano come, dove, quando

- Cambio pagina

- `<w:r> ... <w:br w:type="page"/> ... </w:r>`

```
run.Append(new Break() { Type = BreakValues.Page });
```

L'importanza degli Style

- Devono essere creati in una Part di tipo StyleDefinitionsPart
- In questa Part si crea una collection di Style

```
Styles styles = new Styles(  
    new Style(  
        new StyleName() { Val = "Normal" },  
        new PrimaryStyle()  
    ) { Type = StyleValues.Paragraph, StyleId = "Normal", Default = true } ,  
  
    new Style( ...
```

- Il nome dello Style ("Normal") viene specificato nella formattazione del paragrafo

```
        new ParagraphProperties(  
            new ParagraphStyleId() { Val = StyleName } )
```



**COSTRUIRE UN DOCUMENTO CON
LINQ O DOCUMENTREFLECTOR**

Document properties

```
WordprocessingDocument doc = ...
doc.PackageProperties.Creator = "Raf";
doc.PackageProperties.Category = "Sample";
doc.PackageProperties.Subject = "Office OpenXML";
...
```

Document Properties Location: C:\dev.net\OpenXML\SDK2Word\03-DocumentsProperties\bin

Author:	Title:	Subject:
Raf	DocumentProperties Sample	Office OpenXML
Keywords:	Category:	Status:
Caesar De Bello Gallico Latin	Sample	First draft by Raf
Comments:		
This is a sample for TechDays/WPC 2009 conference		

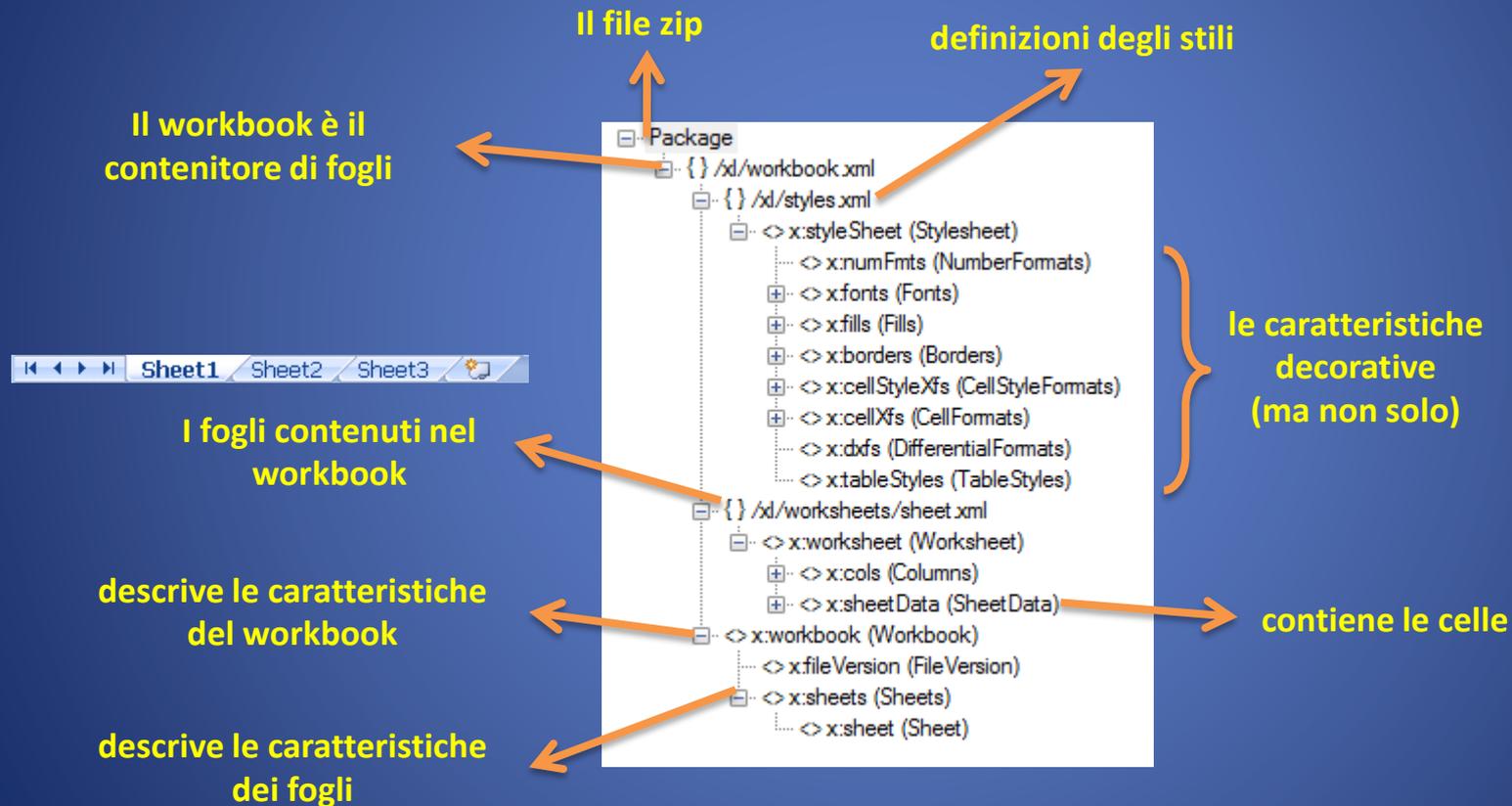
Property	Value
Description	
Title	DocumentProperties Sample
Subject	Office OpenXML
Tags	Caesar De Bello Gallico Latin
Categories	Sample
Comments	This is a sample for TechDays/AWP...
Origin	
Authors	Raf
Last saved by	Raffaele Rialdi
Revision number	3
Version number	
Program name	Microsoft Office Word

- Sono fondamentali per indicizzare i file
- MS Word ne mappa alcune su quelle della Shell
 - Ma solo dopo che Word le ha aperte
 - <http://www.iamraf.net/Samples/Read-shell-file-properties>
- Il resto sono parte del package OPC
 - Lo stesso vale quindi per xlsx e pptx

SpreadSheetML (SpreadSheet Markup Language)

XLSX

Gli elementi base del package di SpreadsheetML

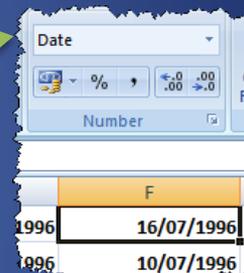


Gli StyleSheet

- Servono a definire:

- caratteristiche decorative delle celle (font, colore, bordi, ...)
- il tipo del contenuto (numero e sui formati, date e suoi formati, testo)

OrderID	CustomerID
10248.00	VINET



- Esempio:

- B2 usa il primo stile in `<cellXfs ... />`
- C2 usa il quarto stile in `<cellXfs ... />`

```
<x:c r="B2" s="0" t="inlineStr">
  <x:is>
    <x:t>VINET</x:t>
  </x:is>
</x:c>
<x:c r="C2" s="3" t="n">
  <x:v>5</x:v>
</x:c>
```

La definizione degli stili (cellXfs)

Tag	SDK2
cellXfs	CellFormats
cellStyleXfs	CellStyleFormats
xf	CellFormat



```
<x:cellXfs count="4" xmlns:x="...">
  <x:xf numFmtId="0" fontId="0" fillId="0" borderId="0" xfid="0">
    <x:alignment horizontal="general" ... />
  </x:xf>
  <x:xf numFmtId="15" fontId="1" fillId="0" borderId="0" xfid="0">
    <x:alignment horizontal="general" ... />
  </x:xf>
  <x:xf numFmtId="0" fontId="2" fillId="0" borderId="0" xfid="0">
    <x:alignment horizontal="center" ... />
  </x:xf>
  <x:xf numFmtId="165" fontId="3" fillId="0" borderId="0" xfid="0">
    <x:alignment horizontal="general" ... />
  </x:xf>
</x:cellXfs>
```

Formattazioni standard predefinite

```
<x:fills count="1" xmlns:x="...">
  <x:fill>
    <x:patternFill patternType="none" />
  </x:fill>
</x:fills>
```

```
<x:borders count="1" xmlns:x="...">
  <x:border>
    <x:left />
    <x:right />
    <x:top />
    <x:bottom />
    <x:diagonal />
  </x:border>
</x:borders>
```

```
<x:numFmts count="1" xmlns:x="...">
  <x:numFmt numFmtId="165"
    formatCode="[Blue][&lt;=100]0.00;[Red][&gt;100]0.00" />
</x:numFmts>
```

Basato su ...

```
<x:cellStyleXfs count="1" xmlns:x="...">
  <x:xf numFmtId="0" fontId="0" fillId="0" borderId="0" xfid="0">
    <x:alignment horizontal="general" ... />
  </x:xf>
</x:cellStyleXfs>
```

Ignorato perché dentro cellStyleXfs

Il contenuto delle celle

```
<x:row r="2" xmlns:x="...">
```

```
<x:c r="A2" s="2" t="n">  
<x:v>10248</x:v>  
</x:c>
```

```
<x:c r="B2" s="3" t="s">  
<x:v>14</x:v>  
</x:c>
```

```
<x:c r="C2" s="4">  
<x:v>35250</x:v>  
</x:c>
```

```
<x:c r="D2" s="0" t="inlineStr">  
<x:is>  
<x:t>VINET</x:t>  
</x:is>  
</x:c>
```

```
<x:c r="E2" s="0">  
<x:f>H2*10</x:f>  
</x:c>
```

```
</x:row>
```

- La tabella è svolta per righe (row)
- Ogni cella è identificata con un reference
 - r="A2" , r="B2" , r="C2" , ...
- Ogni cella ha uno stile
 - s="2"
- I contenuti possono essere differenti:
 - Numerico → t="n"
 - Shared string → t="s"
 - Inline string → t="inlineStr"
 - Il tag child è di tipo <is><t>stringa</t></is>
 - Formula (senza attributo t)
 - Il tag child è di tipo <f>...</f>
 - Excel salva il valore calcolato delle formule nel tag <v />
 - Data (senza attributo t)
 - Nello stile è scritto se il numero rappresenta una data
 - Boolean → t="b"
 - Errore

Shared String Part

- Una lookup table definisce la lista delle shared string
- Dov'è la stringa?

– `<c r="A3" s="1" t="s"> <v>0</v> </c>`



	A	
1	Rosso	
2	Bianco	
3	Rosso	
4		

- L'editor può fornire suggerimenti quando si digitano le prime lettere di una parola già presente nella tabella

- Le shared string sono in una part separata
- In alternativa si può usare inlineStr

– `<c r="C1" t="inlineStr">
 <is>
 <t>inline string</t>
 </is>
</c>`

`<v>0</v>`

`<t>Rosso</t>`

`<v>1</v>`

`<t>Bianco</t>`

```
<sst xmlns="...." count="3"  
  uniqueCount="2">  
  <si>  
    <t>Rosso</t>  
  </si>  
  <si>  
    <t>Bianco</t>  
  </si>  
</sst>
```

Alberto Baroni

«THE EASY WAY»

ClosedXML

The easy way to OpenXML

- ClosedXml è una libreria per agevolare gli sviluppatori a creare documenti Excel 2007/2010/2013.
- Permette di manipolare i documenti senza dover gestire le complessità di Open Xml.
- Project website:
 - <https://closedxml.codeplex.com/>
- NUGET website:
 - <https://www.nuget.org/packages/ClosedXML>

Link utili

- Power tools per OOXML (powershell)
 - <https://powertools.codeplex.com/>
- Altova Tool per OOXML
 - <http://www.altova.com/xmlspy/office-open-xml.html>
- ClosedXML project
 - <https://closedxml.codeplex.com/>
- OOXML <-> ODF
 - <http://odf-converter.sourceforge.net/>
- Binario <-> OOXML
 - <http://b2xtranslator.sourceforge.net/>
- Editor integrato in Visual Studio
 - Cercare estensione "Open XML Package Editor for Visual Studio"

Domande ?

