Microsoft

SpreadsheetML Basics

Disclaimer

The information contained in this slide deck represents the current view of Microsoft Corporation on the issues discussed as of the date of publication. 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 presented after the date of publication.

This slide deck is for informational purposes only. MICROSOFT MAKES NO WARRANTIES, EXPRESS, IMPLIED OR STATUTORY, AS TO THE INFORMATION IN THIS DOCUMENT.

Complying with all applicable copyright laws is the responsibility of the user. Without limiting the rights under copyright, no part of this slide deck may be reproduced, stored in or introduced into a retrieval system, or transmitted in any form or by any means (electronic, mechanical, photocopying, recording, or otherwise), or for any purpose, without the express written permission of Microsoft Corporation.

Microsoft may have patents, patent applications, trademarks, copyrights, or other intellectual property rights covering subject matter in this slide deck. Except as expressly provided in any written license agreement from Microsoft, the furnishing of this slide deck does not give you any license to these patents, trademarks, copyrights, or other intellectual property.

Unless otherwise noted, the example companies, organizations, products, domain names, e-mail addresses, logos, people, places and events depicted herein are fictitious, and no association with any real company, organization, product, domain name, email address, logo, person, place or event is intended or should be inferred.

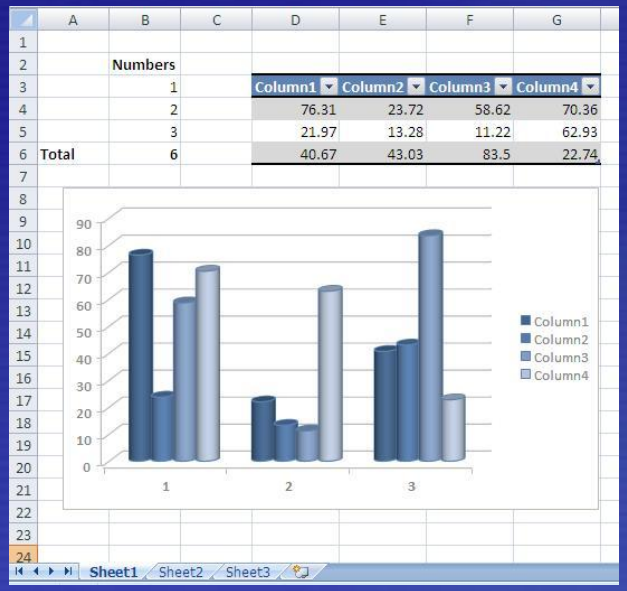
Microsoft, 2007 Microsoft Office System, .NET Framework 3.0, Visual Studio, and Windows Vista are either registered trademarks or trademarks of Microsoft Corporation in the United States and/or other countries.

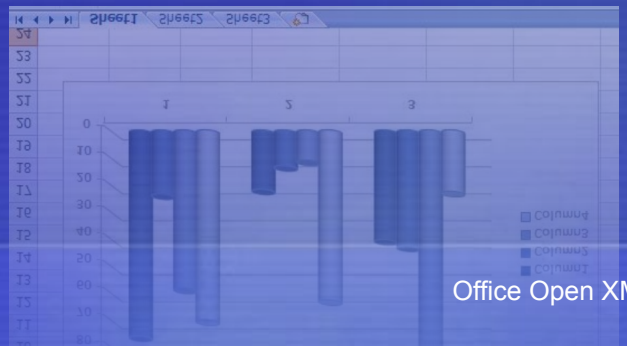
The names of actual companies and products mentioned herein may be the trademarks of their respective owners.

Objectives

- This module covers the core concepts underlying all SpreadsheetML documents:
  - Workbook Architecture
  - Anatomy of an XLSX
  - Rows, columns, values, formulas
  - Strings: inline plain text, rich text, shared strings
  - Formatting Options
  - Calculation Chain

SpreadsheetML





Workbook

- properties
- styles
- sharedStrings
- calcChain
- sheet1..N
  - table
  - chart
    - drawing

SpreadsheetML Design Goal: Performance

SpreadsheetML has been optimized in many ways, based on analysis of real-world spreadsheet usage patterns:

- Small tag size (often a single character)
- Shared strings
- Shared formulas
- Sparse table markup allowed
- Optional r="A1" attribute for faster loading

The minimal XLSX

- Required: **workbook.xml**, the document "start part"
- Required: at least one sheet, **worksheet.xml**
- Required: one relationship part (**.rels**)
  - Must be in a **_rels** folder
- Required: **[Content_Types].xml**
  - Required part for all Open XML documents
  - *Three* content types must be defined:
    - SpreadsheetML main document (for the start part)
    - Worksheet
    - Package relationships (for the required relationships)
- Everything else is optional
  - Worksheet **<sheetdata>** is required, but may be empty

Minimal Workbook/Worksheet

workbook.xml:

```
<workbook>
   <sheets>
    <sheet name="Sheet1" sheetId="1" r:id="rId1"/>
   </sheets>
</workbook>
```

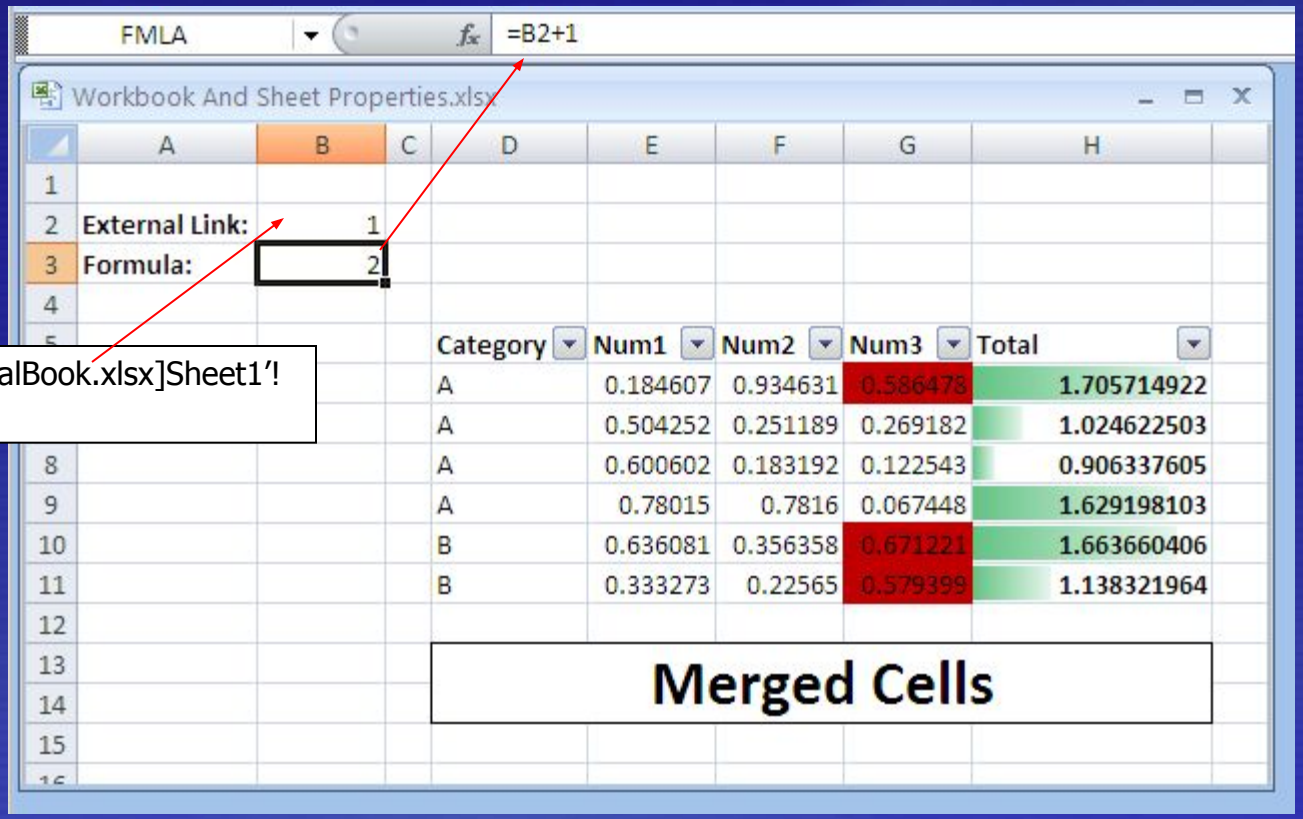
sheet1.xml:


```
<worksheet>
   <sheetData/>
</worksheet>
```

DEMO

SHEETS

Sample Sheet



Worksheet Part – Main Sections

1. Sheet properties (everything before sheetData)
   - Viewing: selected tab, active cell, etc.
   - Print options: orientation, resolution, page margins, etc.
   - Miscellaneous: default row height, sheet protection, etc.

2. The cell table (sheetData, empty if not a worksheet)
   - Row, cells, values, strings (shared-strings indexes), formulas

Sheet Properties

```
<worksheet>
    <sheetPr filterMode="1"/>
    <dimension ref="A2:H14"/>
    <sheetViews>
        <sheetView tabSelected="1" workbookViewId="0">
            <selection activeCell="B3" sqref="B3"/>
        </sheetView>
    </sheetViews>
    <sheetFormatPr defaultRowHeight="15"/>
    <cols>
        <col min="1" max="1" width="12.85546875" bestFit="1" customWidth="1"/>
        <col min="3" max="3" width="3.28515625" customWidth="1"/>
        <col min="4" max="4" width="11.140625" bestFit="1" customWidth="1"/>
        <col min="8" max="8" width="17.140625" style="1" customWidth="1"/>
    </cols>
    <sheetData/>
    <sheetProtection objects="0" scenarios="0"/>
    <printOptions/>
    <pageMargins left="0.7" right="0.7" top="0.75" bottom="0.75" header="0.3"
footer="0.3"/>
    <pageSetup orientation="portrait" horizontalDpi="300" verticalDpi="300"/>
    <headerFooter/>
</worksheet>
```

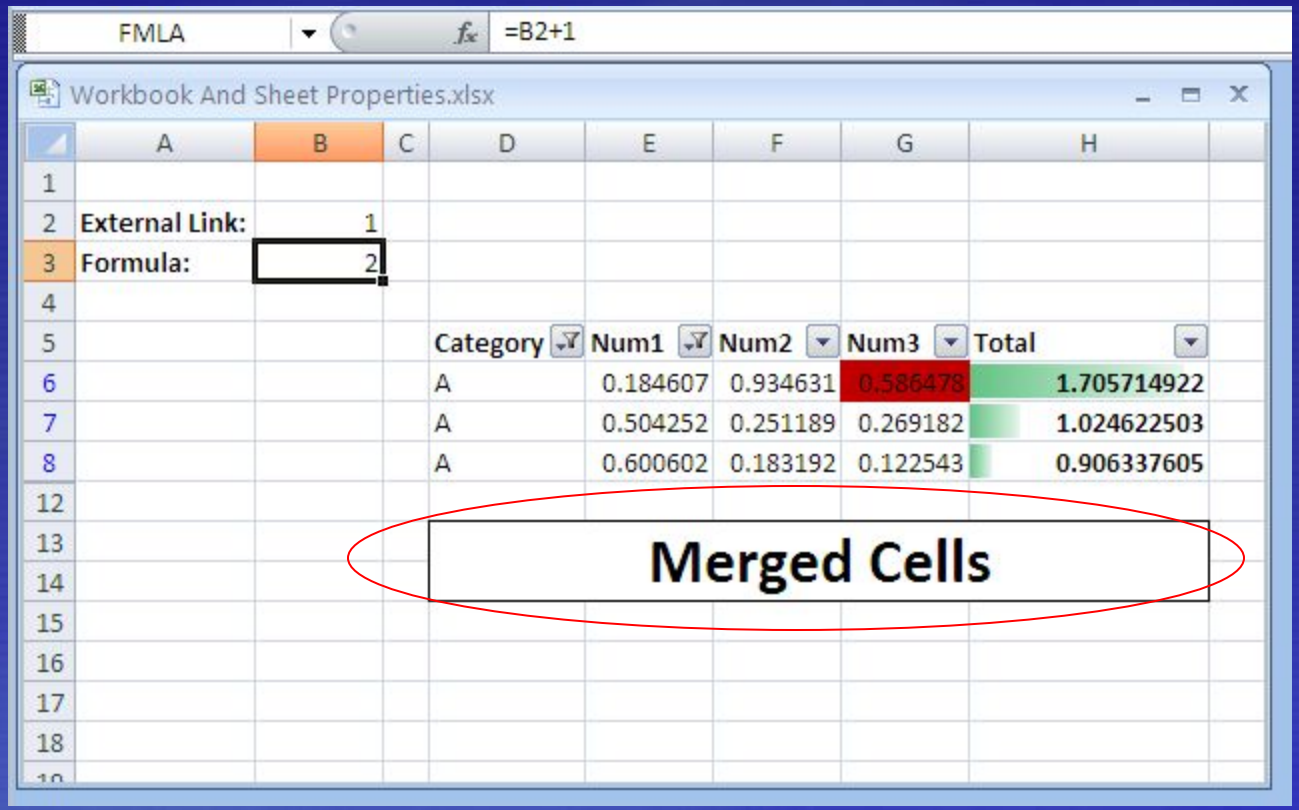
Cell Table: <sheetData> element

```
<row r="7" spans="4:8">
    <c r="D7" t="s">
        <v>2</v>
    </c>
    <c r="E7">
        <v>0.50425224796279555</v>
    </c>
    <c r="F7">
        <v>0.25118866081991786</v>
    </c>
    <c r="G7">
        <v>0.26918159410869791</v>
    </c>
    <c r="H7" s="1">
        <f t="shared" ref="H7:H11" ce="1" si="0">SUM(
        <v>1.0246225028914113</v>
    </c>
```

```
<row r="8" spans="4:8">
    <c r="D8" t="s">
        <v>2</v>
    </c>
    <c r="E8">
        <v>0.6006019062877066</v>
    </c>
    <c r="F8">
        <v>0.18319235857964333</v>
    </c>
    <c r="G8">
        <v>0.12254334000604317</v>
    </c>
    <c r="H8" s="1">
        <f t="shared" ce="1" si="0">SUM(E8:G8)</f>
        <v>0.9063376048733931</v>
    </c>
```

mergeCells

```
<worksheet>
    <sheetData/>
        <mergeCells>
            <mergeCell ref="D13:H14"/>
        </mergeCells>
```



The Sheet-Level Pieces

- Comments
- Formulas & References & Defined Names
- Tables
- AutoFilter
- External Links
  - General
  - Special Directory Relationships
- PivotTable
  - PivotTable
  - PivotCache
- QueryTable
- Metadata

WORKBOOK PROPERTIES

Workbook Properties: Elements

- <fileVersion>
- <workbookPr>
- <calcPr>
- <bookViews>
- <sheets>

```
<workbook>
    <fileVersion lastEdited="4" lowestEdited="4" rupBuild="3814"/>
    <workbookPr backupFile="1" saveExternalLinkValues="0" updateLinks="never"/>
    <calcPr calcId="122211" calcMode="manual" iterate="1"/>
    <bookViews>
        <workbookView showHorizontalScroll="0" showVerticalScroll="0"
showSheetTabs="0" xWindow="45" yWindow="15" windowWidth="9420"
windowHeight="5460" tabRatio="701"/>
    </bookViews>
    <sheets>
        <sheet name="Sheet1" sheetId="1" sh:id="rId1"/>
        <sheet name="Sheet2" sheetId="2" sh:id="rId2"/>
        <sheet name="Sheet3" sheetId="3" sh:id="rId3"/>
    </sheets>
</workbook>
```

STRINGS

Strings in SpreadsheetML

Two ways a string can be stored:

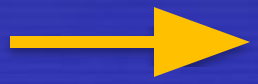
1. Inline strings
   - Provided for ease of translation/conversion
   - Useful in XSLT scenarios
   - Excel and other consumers may convert to shared strings
2. An entry in the shared-strings table
   - May be either a simple string or formatted text

- These approaches may be mixed/combined

Inline Strings

- Inline string support provides a very simple mechanism for programmatically populating a worksheet
- Especially useful in XSLT scenarios
- Excel 2007 converts to shared strings on save
  - If you're consuming Open XML documents, you must handle both cases: inline strings and/or shared strings
- To convert our shared-strings example to inline strings, just replace sheetdata:

```
<sheetData>
  <row><c t="inlineStr"><is><t>Paris</t></is></c></row>
  <row><c t="inlineStr"><is><t>Seattle</t></is></c></row>
  <row><c t="inlineStr"><is><t>London</t></is></c></row>
  <row><c t="inlineStr"><is><t>Copenhagen</t></is></c></row>
  <row><c t="inlineStr"><is><t>Paris</t></is></c></row>
  <row><c t="inlineStr"><is><t>London</t></is></c></row>
</sheetData>
```



	A
1	Paris
2	Seattle
3	London
4	Copenhagen
5	Paris
6	London

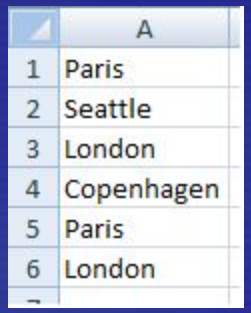
Shared Strings

- By default, strings are stored in a shared-strings part:
  - Each unique string is stored once
  - Cells store the index (0-based) of the string
- This design is based on analysis of typical spreadsheet contents: highly repetitive strings are very common

- Benefits:
  - Users: reduced file size, improved performance
  - Developers: all strings are in one part, simplifying search, localization, and other common string-handling objectives

Shared Strings: example

Worksheet contents:



	A
1	Paris
2	Seattle
3	London
4	Copenhagen
5	Paris
6	London

sharedStrings.xml contents:

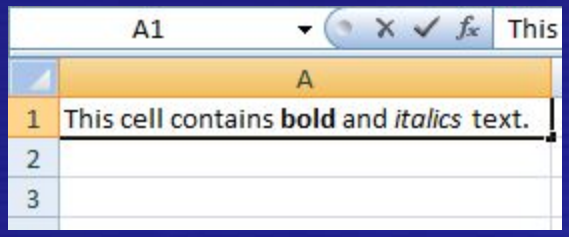
```
<sst xmlns="..." count="6" uniqueCount="4">
  <si>
    <t>Paris</t>
  </si>
  <si>
    <t>Seattle</t>
  </si>
  <si>
    <t>London</t>
  </si>
  <si>
    <t>Copenhagen</t>
  </si>
</sst>
```

6 string references,
4 unique strings

Paris = string 0

```
<row r="1" spans="1:1">
  <c r="A1" t="s">
    <v>0</v>
  </c>
</row>
```

Rich Text Strings



- Stored in sharedStrings.xml
- *One* entry for the entire cell
- Note run properties **<rPr>**

- Cell refers to string 0:

```
<row r="1" spans="1:1">
  <c r="A1" t="s">
    <v>0</v>
  </c>
</row>
```

```
<sst xmlns="..." count="1" uniqueCount="1">
 <si>
  <r>
   <t xml:space="preserve">This cell contains </t>
  </r>
  <r>
   <rPr>
    <b/>
    <sz val="11"/>
    <color theme="1"/>
    <rFont val="Calibri"/>
    <family val="2"/>
    <scheme val="minor"/>
   </rPr>
   <t>bold</t>
  </r>
  <r>
   <t xml:space="preserve"> and </t>
  </r>
  <r>
   <rPr>
    <i/>
    <sz val="11"/><color theme="1"/>
    <rFont val="Calibri"/>
    <family val="2"/>
    <scheme val="minor"/>
   </rPr>
   <t>italics</t>
  </r>
  <r>
   <t xml:space="preserve"> text.</t>
  </r>
 </si>
</sst>
```

FORMATTING

SpreadsheetML Formatting Options

- Direct Cell Formatting (XF)
  - Fonts
  - Fills
  - Borders
  - Numeric Formatting
- Cell Styles
- Table Styles
- PivotTable Styles

Direct Formatting

```
<styleSheet>
    <fonts count="2">
        <font>
            <sz val="11"/>
            <color theme="1"/>
            <name val="Calibri"/>
            <scheme val="minor"/>
        </font>
        <font>
            <b/>
            <sz val="11"/>
            <color theme="1"/>
            <name val="Calibri"/>
            <family val="2"/>
        </font>
    </fonts>
    <fills count="2">
        <fill>
            <patternFill patternType="none"/>
        </fill>
        <fill>
            <patternFill patternType="gray125"/>
        </fill>
    </fills>
    <borders count="1">
        <border>
            <left/>
            <right/>
            <top/>
            <bottom/>
            <diagonal/>
        </border>
    </borders>
    <cellStyleXfs count="1">
        <xf numFmtId="0" fontId="0" fillId="0" borderId="0"/>
    </cellStyleXfs>
    <cellXfs count="2">
        <xf numFmtId="0" fontId="0" fillId="0" borderId="0" xfId="0"/>
        <xf numFmtId="0" fontId="1" fillId="0" borderId="0" xfId="0" applyFont="1"/>
    </cellXfs>
    <cellStyles count="1">
        <cellStyle name="Normal" xfId="0" builtinId="0"/>
    </cellStyles>
</styleSheet>
```

	A	B
1		Numbers
2		1
3		2
4		3
5	Total	6

```
<sheetData>
    <row r="1" spans="1:2">
        <c r="B1" s="1" t="s">
            <v>0</v>
        </c>
    </row>
    <row r="2" spans="1:2">
        <c r="B2">
            <v>1</v>
        </c>
    </row>
    <row r="3" spans="1:2">
        <c r="B3">
            <v>2</v>
        </c>
    </row>
    <row r="4" spans="1:2">
        <c r="B4">
            <v>3</v>
        </c>
    </row>
    <row r="5" spans="1:2">
        <c r="A5" s="1" t="s">
            <v>1</v>
        </c>
        <c r="B5">
            <f>SUM(B2:B4)</f>
            <v>6</v>
        </c>
    </row>
</sheetData>
```

DEMO

Applying Cell, Table, PivotTable Styles

- Referenced by Name
- Explicit formatting is described using formatting records (xf)

FORMULAS AND CALC CHAIN

Formulas, References, Defined Names

- Excel saves out exactly what you see in the cell at runtime.
- Implication: Excel re-parses the formula on load, and serializes it on save

- Formula links to external workbooks:
  - Abstract file path to relationships part
  - Excel caches snapshot of external workbook structure (sheets & cell tables)

Formulas: example

```
<row>
  <c>
    <v>1</v>
  </c>
</row>
<row>
  <c>
    <v>2</v>
  </c>
</row>
<row>
  <c>
    <v>3</v>
  </c>
</row>
<row>
  <c>
    <f>SUM(A1:A3)</f>
  </c>
</row>
```

DEMO

Microsoft | developer & platform **evangelism**