Office X Roadmap & Current Progress

Technical Note

16th June 2017

Version

This document contains proprietary information that is the property of Deep-Secure Ltd.

It is made available to third parties under specific non-disclosure agreements.

Use of the information outside of any such agreement is prohibited.

**Table of Contents**

[1. Introduction 3](#_Toc485374145)

[1.1 Overview 3](#_Toc485374146)

[1.2 Scope 3](#_Toc485374147)

[1.3 Web Applications 3](#_Toc485374148)

[1.4 Structure of Document 4](#_Toc485374149)

[2. Word 5](#_Toc485374150)

[2.1 Overview 5](#_Toc485374151)

[2.2 Further Information 6](#_Toc485374152)

[3. Excel 8](#_Toc485374153)

[3.1 Overview 8](#_Toc485374154)

[3.2 Further Information 9](#_Toc485374155)

[4. PowerPoint 11](#_Toc485374156)

[4.1 Overview 11](#_Toc485374157)

[4.2 Further Information 11](#_Toc485374158)

[5. References 13](#_Toc485374159)

# Introduction

## Overview

* + 1. This document describes the current implementation of the parsing, verifying and rendering support for the Microsoft “Office X” formats (Office Open XML, [1]), within the iX appliance, together with the projected future roadmap.
		2. The Office X format is complex, with a detailed specification running to many thousands of pages. It supports legacy applications dating back over 20 years, and provides various hooks to which proprietary binary blobs of data can be attached.
		3. The goal is to produce a transformed document that looks and behaves identically to the original source document. We are using a phased approach for implementation of Office X features. The intent is to support all Office X documents as input for parsing, stripping out unsupported features, and rendering with as much fidelity as possible.

## Scope

* + 1. This implementation is restricted to Word, Excel and PowerPoint documents, and excludes any other formats supported by the Office Open XML specification.
		2. Please note that the feature lists for each of the Office applications are almost certainly not exhaustive. The Office suite is complex with many features, some of which do not manifest themselves in the files saved to disk.

## Web Applications

* + 1. The Office 365 web applications provide an Office-like experience for editing Office X documents within a web browser. They provide the facility to open documents from OneDrive/SharePoint or to export/round-trip documents to a PC, in addition to authoring from scratch.
		2. The web applications provide a cut-down set of functionality, and some additional online collaborative tools, but still export in the standard Office Open XML format. A reduced set of parsing and rendering functionality is required to support documents edited entirely online. However, Office 365 mostly supports round-tripping of documents authored using the full desktop applications, even when they can’t be created online.[[1]](#footnote-1) The complexity of the web applications will inevitably increase over time, adding support for more features as they go.

## Structure of Document

* + 1. The remainder of this document describes the current and planned functionality levels for the Office X applications. The features for each application are listed in a table, with columns as follows:
* Feature – a brief description of the application feature.
* C – a rough complexity measure (Low/Medium/High), also indicative of the risk level associated with the implementation.
	+ Low complexity features require the parsing of few XML tags, probably in a single part of the XML zip file.
	+ Medium complexity features require the parsing/rendering of many tags, perhaps in several parts of the XML zip file.
	+ High complexity features require the parsing/rendering of many tags, with complex relationships between parts of the XML zip file.
* O – whether the feature is currently available in the Office 365 online applications. Y means it is, P means it’s partially supported.
* R1 – current release (iX 1.3). Y indicates whether the feature is supported (Y) or partially supported (P).
* R2 – next release of Office X functionality (currently scheduled for end of June 2017). Y indicates whether the feature will be supported.
* R3 – following release of Office X functionality (currently scheduled for end of July 2017). Y indicates whether the feature will be supported.
* R4 – following release of Office X functionality (currently on product roadmap). Y indicates whether the feature will be supported.
	+ 1. Note that although the complexity measure is based on a rough estimate of the complexity of the XML which must be parsing and rendered to support that feature, implementing a similar feature across all applications (e.g. MathML support) will almost certainly utilise common code.

# Word

## Overview

* + 1. Word probably requires the most in-depth implementation to provide a sufficiently accurate transformed document.

## Feature Type:

## E - Essential FeaturesI – Intermediate FeaturesA – Advanced Features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Feature** | **C** | **O** | **R1** | **R2** | **R3** | **R4** |
| E | Basic Document Support with Formatted Text | M | Y | Y |  |  |  |
| E | Simple Text Styles | M | Y | Y |  |  |  |
| I | Advanced Text Styles | M | Y |  |  | Y |  |
| E | Simple Headers & Footers | M | Y | Y |  |  |  |
| I | Advanced Headers & Footers | H | Y |  | Y |  |  |
| I | Footnotes and Endnotes | M | Y | Y |  |  |  |
| E | Simple Paragraph Formatting | M | Y | Y |  |  |  |
| I | Advanced Paragraph Formatting | M | P |  |  | Y |  |
| E | Page Breaks | L | Y | Y |  |  |  |
| E | Section Properties | L | P | Y |  |  |  |
| I | Paragraph Numbering | M |  | Y |  |  |  |
| I | Table of Contents (Fields) | H |  |  | Y |  |  |
| A | Embedded Objects | M |  |  |  |  | Y |
| E | Simple Embedded Pictures | M | Y | Y |  |  |  |
| I | Advanced Embedded Pictures | M |  |  |  | Y |  |
| I | Shapes and SmartArt | H |  |  |  |  | Y |
| A | Charts | H |  |  |  |  | Y |
| E | Simple Tables | H | Y |  | Y |  |  |
| I | Advanced Tables | H | P |  |  | Y |  |
| A | Cross-references | H |  |  |  |  | Y |
| A | Bookmarks | H |  |  |  |  | Y |
| A | Equations (MathML support) | H |  |  |  |  | Y |
| A | Outline Levels | H |  |  |  |  | Y |
| A | Document Properties | L |  |  |  |  | Y |
| I | Comments | L | Y |  |  |  | Y |
| I | Hyperlinks | L | Y |  |  |  | Y |
| I | Revisions | L | Y |  |  |  | Y |
| A | Master & Subdocuments | L |  |  |  |  | Y |
| A | Document Templates | L |  |  |  |  | Y |

## Further Information

* + 1. The features listed in the table are explained in more detail here:
* Basic Document Support with Formatted Text - a simple single page document containing one or more paragraphs with text with simple inline formatting (e.g. bold or italic).
* Simple Text Styles – Basic text styles, fonts and font effects.
* Advanced Text Styles – full text style support, with all styles, style sheets and themes.
* Simple Headers & Footers – simple text header & footer, with support for different first, odd and even pages.
* Advanced Headers & Footers – headers and footers including advanced formatting and pictures etc.
* Footnotes & Endnotes – footnotes and endnotes, with same level of formatting support as paragraphs and runs of text.
* Simple Paragraph Formatting – indentation, justification, line spacing etc.
* Advanced Paragraph Formatting – widow and orphan control and other pagination options.
* Page Breaks – user-inserted page breaks.
* Section Properties – section formatting (page size/orientations, columns, margins, headers & footers).
* Paragraph Numbering – paragraph number, with various number formats.
* Table of Contents (Fields) – support for inserting a table of contents (requires support for fields in general).
* Embedded Objects – embedding of arbitrary binary objects (e.g. COM objects).
* Simple Embedded Pictures – inline JPEG/BMP/PNG/GIF images with scaling.
* Advanced Embedded Pictures – JPG/BMP/PNG/GIF images, with anchor, rotation, text wrap.
* Shapes & SmartArt – the support of a sufficient subset of DrawingML mark-up to support the creation of diagrams as provided by SmartArt (which could be most of it).
* Charts – Excel charts (includes embedding the Excel spreadsheet).
* Simple Tables – simple grid with textual content.
* Advanced Tables – all content allowed, with table formatting.
* Cross-references – field code support for cross-referencing.
* Bookmarks - hyperlinked document markers (these are relatively difficult to support, because they span multiple
* Equations (MathML support) – support for content created in the maths-specific mark-up language.
* Outline Levels – outline levels, as per the outline view (and used for outline paragraph numbering etc).
* Document Properties – document properties, including custom properties. This is probably mostly shared code with Excel/PowerPoint.
* Comments – comments on the document, with author attribution.
* Hyperlinks – links to external content.
* Revisions – change tracking.
* Master & Subdocuments – the compilation of subdocuments into a master document.
* Document Templates – support for document templates (dotx) in addition to docx.

# Excel

## Overview

* + 1. Unlike PowerPoint (and to some extent Word), the most fundamental part of Excel is the cell data and formulae, rather than the presentational aspects. Other features (e.g. charts, pivot tables) can be reconstructed once the document has been transformed through the guard. Hence, many features are relegated to later phases.

## Feature Type:

## E - Essential FeaturesI – Intermediate FeaturesA – Advanced Features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Feature** | **C** | **O** | **R1** | **R2** | **R3** | **R4** |
| E | Workbook with Multiple Sheets | L | Y | Y |  |  |  |
| E | Text Cells | M | Y | Y |  |  |  |
| E | Formulae (passed through unparsed) | L | Y | Y |  |  |  |
| I | Parsed Formulae (common functions only) | M | Y |  |  |  | Y |
| I | Parsed Formulae (complete set) | H | Y |  |  |  | Y |
| A | Headers & Footers | L |  |  |  |  | Y |
| E | Text & Cell Formatting (simple) | L | Y |  | Y |  |  |
| I | Text & Cell Formatting (advanced) | L | Y |  |  | Y |  |
| E | Number Formatting | L | Y |  | Y |  |  |
| A | Conditional Formatting | L |  |  |  |  | Y |
| E | Tables | H | Y |  |  | Y |  |
| A | Pivot Tables | H |  |  |  |  | Y |
| E | Charts (bar chart) | H | Y |  | Y |  |  |
| E | Charts (simple) | H | Y |  |  | Y |  |
| I | Charts (advanced) | H | Y |  |  |  | Y |
| A | Sparklines | L |  |  |  |  | Y |
| I | Print Area | M |  |  |  |  | Y |
| I | Page Layouts | M |  |  |  |  | Y |
| I | Embedded Pictures | M |  | Y |  |  |  |
| A | Embedded Objects | H |  |  |  |  | Y |
| A | Shapes and SmartArt | H |  |  |  |  | Y |
| A | Equations (MathML support) | H |  |  |  |  | Y |
| A | External Data Sources | M |  |  |  |  | Y |
| A | Document Properties | L |  |  |  |  | Y |
| I | Comments | L | Y |  |  |  | Y |
| I | Hyperlinks | L | Y |  |  |  | Y |
| A | Revisions | M | Y |  |  |  | Y |

## Further Information

* + 1. The features listed in the table are explained in more detail here:
* Workbook with Multiple Sheets – support for excel documents containing one or more worksheets in the workbook.
* Text Cells – cells containing text, with very basic formatting.
* Formulae (passed through unparsed) – parsing formulae through unprocessed.
* Parsed Formulae (common functions only) – support for parsing the common formulae (with full parsing/validation against a schema.
* Parsed Formulae (complete set) – support for parsing the full set of formulae (approximately 350) with full parsing/validation against a schema.
* Headers & Footers – spreadsheet headers and footers (relatively simple compared to Word).
* Text & Cell Formatting (simple) – simple formatting and colour options.
* Text & Cell Formatting (advanced) – full formatting including support for styles and themes.
* Number Formatting – formatting of numbers.
* Conditional Formatting – formatting based on a set of formula-based criteria.
* Tables – grouping of data to allow filtering and sorting.
* Pivot Tables – grouping of data across multiple ranges or sheets, to allow complete filtering/sorting/joining of the data.
* Charts (bar chart) – basic bar chart with axes etc.
* Charts (simple) – support for a further selection of common charts.
* Charts (advanced) – all chart types with all options.
* Sparklines – support for sparklines (simple chart in single worksheet cell).
* Print Area – definition of the area of a worksheet to be printed.
* Page Layouts – page formatting for printing, including orientation, margins and so on.
* Embedded Pictures – insertion of JPEG/BMP/PNG/GIF images with scaling and rotation.
* Embedded Objects – embedding of arbitrary binary objects (e.g. COM objects).
* Shapes & SmartArt – the support for a sufficient subset of DrawingML mark-up to support the creation of diagrams as provided by SmartArt (which could be most of it).
* Equations (MathML support) – support for text content created in the maths-specific mark-up language.
* External Data Sources – pulling of data from external sources, such as database or web sites.
* Document Properties – document properties, including custom properties. This is probably mostly shared code with Word/PowerPoint.
* Comments – comments on the document, with author attribution.
* Hyperlinks – links to external content.
* Revisions – change tracking.

# PowerPoint

## Overview

* + 1. The initial implementation only supports textual content of slides. This is because there is a lot of upfront effort required to render an initial, compliant PowerPoint document.

## Feature Type:

## E - Essential FeaturesI – Intermediate FeaturesA – Advanced Features

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Type** | **Feature** | **C** | **O** | **R1** | **R2** | **R3** | **R4** |
| E | Basic Text Slides | H | Y | Y |  |  |  |
| E | Multiple Masters and Layouts | H | P | Y |  |  |  |
| E | Formatted Text (simple) | M | Y |  | Y |  |  |
| E | Formatted Text (advanced) | M | Y |  |  | Y |  |
| E | Embedded Pictures | M | Y | P | Y |  |  |
| A | Embedded Objects | M |  |  |  |  | Y |
| E | Shapes  | H | Y |  | Y |  |  |
| I | SmartArt | H | Y |  |  | Y |  |
| A | Charts | H |  |  |  |  | Y |
| A | Text Effects | M |  |  |  |  | Y |
| I | Slide Transitions | M | P |  |  | Y |  |
| A | Slide Display Timings | L |  |  |  |  | Y |
| I | Animations | M | P |  |  | Y |  |
| A | Advanced Animations | H |  |  |  |  | Y |
| A | Object Actions | M |  |  |  |  | Y |
| A | Online and External Content | L | P |  |  |  | Y |
| I | Handouts | M |  |  |  |  | Y |
| E | Notes | M | Y |  |  | Y |  |
| A | Custom Slideshows | L |  |  |  |  | Y |
| A | Slideshow Properties | M |  |  |  |  | Y |
| A | Storyboarding | H |  |  |  |  | Y |
| A | Document Properties | L |  |  |  |  | Y |
| I | Comments | L | Y |  |  |  | Y |
| I | Hyperlinks | L | Y |  |  |  | Y |

## Further Information

* + 1. The features listed in the table are explained in more detail here:
* Basic Text Slides – slides containing simple text.
* Multiple Masters and Layouts – support for presentations containing multiple slide masters with multiple layouts per master.
* Formatted Text (simple) – text formatting including size, font and alignment.
* Formatted Text (advanced) – includes full text formatting with colour, orientation, spacing etc.
* Embedded Pictures –insertion of JPEG/BMP/PNG/GIF images with scaling and rotation.
* Embedded Objects – embedding of arbitrary binary objects (e.g. COM objects).
* Shapes– support for basic set of shapes.
* SmartArt – the support for a sufficient subset of DrawingML mark-up to support the creation of diagrams and shapes as provided by SmartArt (which could be most of it).
* Charts – Excel charts (includes embedding the Excel spreadsheet).
* Text Effects – effects applied to text and bounding boxes
* Slide Transitions – transition effects, including transition properties such as speed.
* Slide Display Timings – the amount of time for which to display a slide.
* Animations – basic animations
* Advanced Animations – full set of animations, including motion paths
* Object Actions – things that happen when you click on a shape or some text (e.g. branch to another slide).
* Online and External Content – video and audio content either on local machine, network or the Internet.
* Handouts – formatted layouts of multiple slides. David.Robinson@deep-secure.com
* Notes – formatted layouts of notes associated with each slide.
* Custom Slideshows – slideshows containing subsets and/or reordering of slides.
* Slideshow Properties – for example, whether to show audio controls, or to disable animation and so on.
* Storyboarding – complex combination of shapes, objects and rich drawing mark-up.
* Document Properties – document properties, including custom properties. This is probably mostly shared code with Word/PowerPoint.
* Comments – comments on the document, with author attribution.
* Hyperlinks – links to external content.

# References

1. Information technology — Document description and processing languages — Office Open XML File Formats, ISO/IEC 29500
1. For example, Word documents containing multiple sections are handled by the web application, but cannot be created in the web application itself. Some features however, such as charts in PowerPoint Online, do not roundtrip. Importing from the desktop application results in the chart being replaced with a simple graphic image of the chart, losing the ability to edit it subsequently. [↑](#footnote-ref-1)