Mark Logic Corporation

MarkLogic Toolkit for PowerPoint®
Developer's Guide

Release 1.1
September, 2010
Last Revised: 1.1-1, September, 2010

1.0 Overview and Requirements

This chapter introduces the MarkLogic Toolkit for PowerPoint® and lists the product requirements and supported platforms. It includes the following sections:

- Overview of the MarkLogic Toolkit for PowerPoint
- System Requirements

1.1 Overview of the MarkLogic Toolkit for PowerPoint

The MarkLogic Toolkit for PowerPoint allows you to integrate Microsoft PowerPoint 2007/2010 directly with content stored in a MarkLogic Server database. Key features of the MarkLogic Toolkit for PowerPoint include:

- Ability to Access Content from MarkLogic Server in Microsoft PowerPoint 2007/2010
- Custom Browser-Based Add-In
- No C# Coding Required for Add-In
- Configurable Installer to Deploy to PowerPoint Client

1.1.1 Access Content from MarkLogic Server in Microsoft PowerPoint 2007/2010

An *add-in* in Microsoft PowerPoint is a panel that appears in the PowerPoint interface that allows you to perform custom actions. The MarkLogic Add-in for PowerPoint® uses PowerPoint's add-in facility to create a panel in PowerPoint that allows you to run a MarkLogic Server application in the MarkLogic Add-in for PowerPoint. You can use the MarkLogic Add-in for PowerPoint to retrieve content from MarkLogic Server and use it in a PowerPoint document, and you can use the Add-In to take content in a PowerPoint document and use it in a MarkLogic Server application.

1.1.2 Custom Browser-Based Add-In

The MarkLogic Add-in for PowerPoint exposes a browser in the Microsoft PowerPoint interface, and the browser has access both to Microsoft PowerPoint and to anything over HTTP. The Add-In communicates with an HTTP Server application running in MarkLogic Server. The MarkLogic Server application can perform whatever tasks you want, and therefore allows for completely custom behavior of the MarkLogic Add-in for PowerPoint. For example, the application can provide a search interface to content stored in MarkLogic Server, and can allow you to use content from that search in PowerPoint. The application can have any interface that you can run in a browser, and it can do things like allow users to insert content into PowerPoint, extract content from PowerPoint, and so on.

1.1.3 No C# Coding Required for Add-In

The MarkLogic Add-in for PowerPoint uses a JavaScript library to communicate with Microsoft PowerPoint 2007/2010, and it uses HTTP to communicate with MarkLogic Server. Consequently, you can create full-featured applications to run in the MarkLogic Add-in for PowerPoint, and those applications run as a MarkLogic Server HTTP Server application.

The usual way to create a MarkLogic Add-in for PowerPoint is to use the PowerPoint software development tools and create a C# program that communicates with PowerPoint. The MarkLogic Add-in for PowerPoint requires no developer-written C# code; all of the communication with PowerPoint is done via a JavaScript library. The JavaScript library communicates with PowerPoint on the client side and with MarkLogic Server on the server side, requiring no C# coding. You can make server-side code changes to your MarkLogic Server application which will change the behavior of your MarkLogic Add-in for PowerPoint.

1.1.4 Configurable Installer to Deploy to PowerPoint Client

The MarkLogic Toolkit for PowerPoint includes a customizable Windows installer file (`.msi`). You can use this to create a customized installer which points to your application. Once you have created the custom installer, you can deploy it on any number of Microsoft PowerPoint instances, either by having individual users install the `.msi` file or by having a network administrator do a network installation.

1.2 System Requirements

The MarkLogic Toolkit for PowerPoint has the following system requirements:

For the Microsoft Windows client:

- Microsoft Office 2007/2010
- .NET Framework 3.5
- Microsoft Visual Studio Tools for Office Runtime 3.0
- Windows Installer 3.1

   (Note that the MarkLogic Add-in for PowerPoint installer checks for the above, and provides a facility to install them).

- The 2007 Microsoft Office Primary Interop Assemblies, which are installed with Office 2007 and also available separately
- Administrative privileges on the Windows machine

For the Windows environment in which you customize the MarkLogic Add-in for PowerPoint installer (`.msi`) file:

- All of the prerequisites for the Microsoft Windows client (above)
- Microsoft Windows SDK v6.0 or greater

For the MarkLogic Server instance with which the MarkLogic Add-in for PowerPoint communicates:

- MarkLogic Server version 4.0-5 or later, on any supported platform. See the *Release Notes* and the *Installation Guide* in the MarkLogic Server documentation for details on the requirements for MarkLogic Server.
- Administrative privileges for the MarkLogic Server environment are required.

2.0 Preparing the Installer and Installing the MarkLogic Add-in for PowerPoint

This chapter describes the installation process for the MarkLogic Add-in for PowerPoint. The procedures involve installing an application in an instance of MarkLogic Server, creating a custom installer for the Add-in, and installing the MarkLogic Add-in for PowerPoint. The following sections are included:

- MarkLogic Toolkit for PowerPoint Installation Package
- Installation Procedure

2.1 MarkLogic Toolkit for PowerPoint Installation Package

You can download the latest MarkLogic Add-in for PowerPoint installation package from developer.marklogic.com. The installation package is a zip file that expands to a folder with the name and version number of the MarkLogic Toolkit for PowerPoint, and that folder contains the following:

- `addin.deploy` folder: Contains the MarkLogic Add-in for PowerPoint installation packages (`.msi` and `.exe` files).

- `config` folder: Contains the `config.idt` file, which is used when customizing the installation package.

- `cpf` folder: Contains the PresentationML Process and PresentationML Process Tags pipelines. The Process pipeline sets document properties for slide package (.pptx) parts and their associated images. The Process Tags pipeline sets the tags for any presentations, slides, or slide components in the document-properties for slide package (.pptx) parts.

- `docs` folder: Contains the documentation for the MarkLogic Add-in for PowerPoint, including the *PowerPoint Add-In Developer's Guide* (this document), the JavaScript API documentation, and the XQuery API documentation.

- `image-util` folder: Contains .exe and .bat for converting directory of PowerPoint presentations to folders of .PNG images for associated slides.

- `js` folder: Contains the JavaScript API source file to use in your applications.

- `Samples` folder: Contains a sample MarkLogic Server application that communicates with the MarkLogic Add-in for PowerPoint.

- `xquery` folder: Contains the XQuery API library modules for use with your applications.

There are three main parts to the installation process: installing the application in MarkLogic Server, creating the custom installation package for the Add-In, and deploying the Add-In in Microsoft PowerPoint. The remainder of this chapter describes the installation process for the three main parts of the MarkLogic Add-in for PowerPoint installation.

2.2 Installation Procedure

This section describes the various parts of preparing a MarkLogic Add-in for PowerPoint installer for deployment. For information about system requirements, see "System Requirements" on page 4.

This section contains the following parts:

- Installing Your Application in MarkLogic Server
- Creating a Custom Installer for the MarkLogic Add-In for PowerPoint
- Installing the MarkLogic Add-in for PowerPoint on Clients with Microsoft PowerPoint
- Uninstalling the MarkLogic Add-in for PowerPoint
- Modifying the Registry to Change the URL during Development

2.2.1 Installing Your Application in MarkLogic Server

To use the MarkLogic Add-in for PowerPoint, you need an application that runs on MarkLogic Server to which the Add-In communicates. The application can be any application that runs on a MarkLogic Server HTTP Server. The application uses the JavaScript API as its interface to communicate with Microsoft PowerPoint. For a description of a simple application, see "Getting Started With the MarkLogic Add-in for PowerPoint" on page 12.

After you install your application, make a note of the URL to that application, as you need to enter that in your `.msi` file before you deploy the application to your users. A sample URL might be `http://marklogic.myserver.com:8123/`, where the HTTP Server is running on the machine `marklogic.myserver.com` and is running on port 8123.

2.2.2 Creating a Custom Installer for the MarkLogic Add-In for PowerPoint

You can create a custom `.msi` file for the MarkLogic Add-in for PowerPoint. The customization modifies the following properties for the MarkLogic Add-in for PowerPoint, and those properties end up stored in the Windows registry under
`HKEY_CURRENT_USER\MarkLogicAddinConfiguration\PowerPoint`:

- `URL`: The URL of a MarkLogic Server application to which the MarkLogic Add-in for PowerPoint connects. The default value is `http://www.marklogic.com`.

- `RbnBtnLbl`: The label that appears on the custom ribbon button created by the Add-In. The default value is `My Button`.

- RbnGrpLbl: The label that specifies the Group in which the button appears in the ribbon (the ribbon group label). This appears in the ribbon underneath the button. The default value is My Group.

- RbnTabLbl: The label for the ribbon tab. The default value is My Tab.

- CTPTitle: The title for the Custom Task Pane that has the browser embedded. The default value is My Task Pane.

- CTPEnabled: A boolean that, when true, indicates that the MarkLogic Add-in for PowerPoint pane is opened when PowerPoint starts, or when false, indicates that the user must enable the Add-In using the button. The default value is true.

- User: The username for the MarkLogic Server the Add-in is configured with. This is required for the Button Save functionality.

- Auth: The password for the MarkLogic Server the Add-in is configured with. This is required for the Button Save functionality.

- RibbonXqy: default value: /utils/upload.xqy?uid=
This value, concatenated with the value from URL, is the location of the upload.xqy file which is required for Button Save functionality. File is found in Samples/utils.

The following parts outline several different ways you can customize the .msi file. You can use whichever way makes is the most convenient in your development environment.

- Edit the config.idt and use MsiDb to customize the .msi file
- Use Orca.exe to customize the .msi file

2.2.2.1 Edit the config.idt and use MsiDb to customize the .msi file

Perform the following steps to edit the .msi file using the config.idt file supplied in the installation zip package.

1. Open the config/config.idt file supplied in the zip installation package.

Note: If you do not want to use the config.idt file, you can export the idt information from the .msi file by running the following command:

```
MsiDb -f "<directory where idt is to be exported>" –d
"C:\MyAddin\MarkLogic_PowerPointAddin_Setup.msi" -e Registry
```

2. Use a text editor to edit the URL, CTPTitle, and other values in the idt file. Make sure to save your changes to the idt file.

3. Use the `MsiDb` in the Windows SDK to run the following command to update the `.msi` file with your new values.

```
MsiDb -f "<directory where idt is located>" -d
"MarkLogic_PowerPointAddin_Setup.msi" -i config.idt
```

For example: `C:\> cd "C:\Program Files\Microsoft SDKs\Windows\v6.0A\Bin"`

```
C:\Program Files\Microsoft SDKs\Windows\v6.0A\Bin> MsiDb -f
"C:\MyAddin\MyConfig" -d
"C:\MyAddin\MarkLogic_PowerPointAddin_Setup.msi" -i
config.idt
```

This updates the `.msi` file with your new values.

2.2.2.2 Use Orca.exe to Customize the .msi File

You can use the `Orca.exe` utility to customize the .msi file. `Orca.exe` is part of the Windows SDK, but is not always installed by default. To modify the .msi file using Orca, perform the following steps:

1. Launch Orca.exe (for example, from the start menu). If you do not have Orca installed, install it. For details on Orca, see the Microsoft SDK documetation.

2. In the Orca window, select File > Open and navigate to the location of the MarkLogic Add-in for PowerPoint `.msi` file (for example, `c:/tmp/MarkLogic_PowerPointAddin-1.0-1/addin.deploy`).

3. Select the `MarkLogic_PowerPointAddin_Setup.msi` file and click Open.

4. Select the Registry table on the left pane.

5. Edit the values for the `URL`, `CTPEnabled`, `RbnGrpLbl`, `RbnTabLbl`, `CTPTilte`, and/or `RbnBtnLbl` names.

Warning: Do not edit any other values than these, otherwise the installation might not work correctly.

6. Use File > Save to save the `.msi` file, or use File > Save As... to save the file to a different `.msi` file.

This updates the `.msi` file with your new values.

2.2.3 Installing the MarkLogic Add-in for PowerPoint on Clients with Microsoft PowerPoint

After you have prepared the `.msi` as described in the previous section, it is ready to install on your Microsoft Windows clients for PowerPoint. You can use the standard Windows techniques to deploy the installer across an enterprise, or you can make the `.msi` file and the `setup.exe` file available to your users and have them run the installer on their own machines using the steps below.

To run the installation, perform the following steps:

1. Update the `.msi` file as described in "Creating a Custom Installer for the MarkLogic Add-In for PowerPoint" on page 7.

2. Run the `addin.deploy/setup.exe` file. This installer checks for all of the dependencies needed by the MarkLogic Add-in for PowerPoint, and then installs them if they are not already installed.

   Note: If the setup program determines that you need to install .NET 3.5, the .NET installation will take some time and requires a reboot.

3. When it is done checking for dependencies, and if all of the dependencies are met, it automatically launches the MarkLogic Add-in for PowerPoint installation (the `MarkLogic_PowerPointAddin_Setup.msi` file).

4. Follow the prompts for the installation, clicking Next until it installs the MarkLogic Add-in for PowerPoint.

5. Click Close when the installer completes.

6. Launch Microsoft PowerPoint to test the installation. You should see a new tab at the top of the PowerPoint interface with the name you added in your customization procedure earlier, and the My Task Pane to the right. If you made no modifications to the .msi file, the Add-In looks like the following in Microsoft PowerPoint:
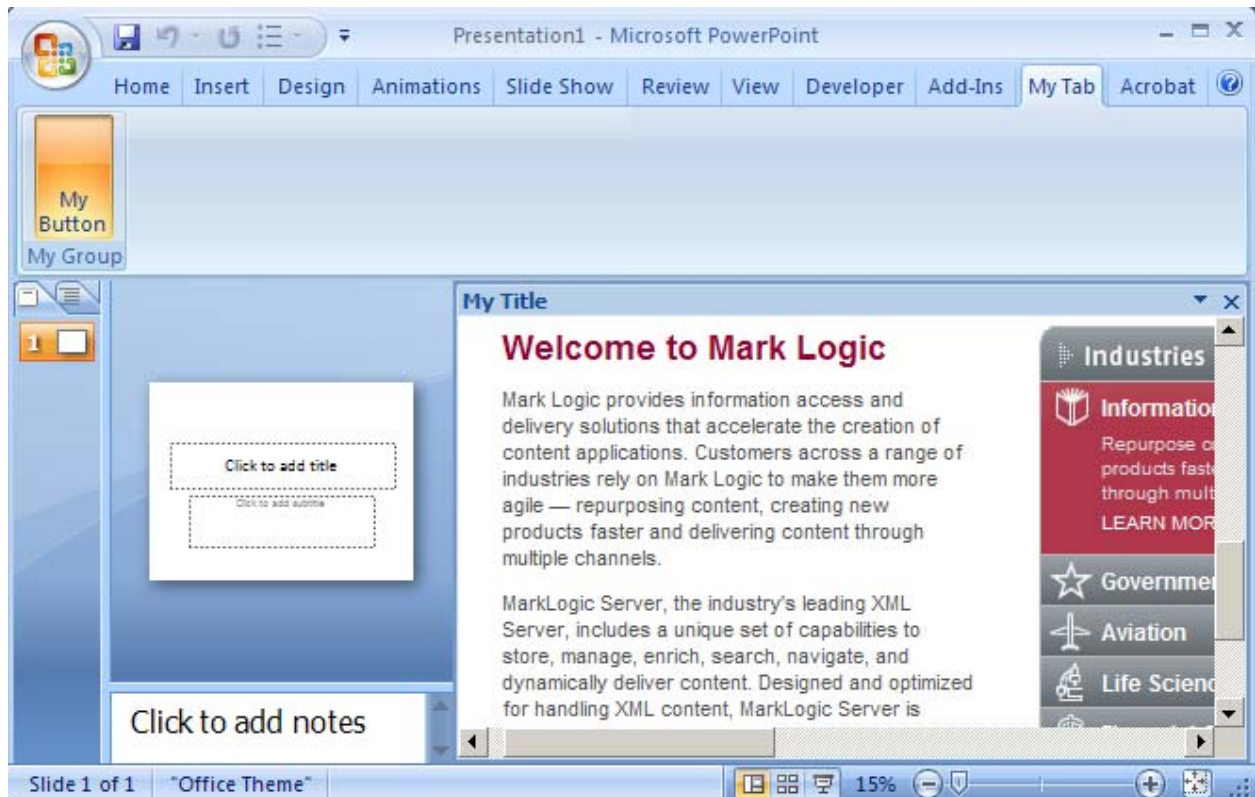
2.2.4 Uninstalling the MarkLogic Add-in for PowerPoint

If you need to change any of the parameters of the MarkLogic Add-in for PowerPoint (the URL or the names of the customizable fields, for example), then you must uninstall the current version of the Add-In before installing a new version.

To remove the MarkLogic Add-in for PowerPoint from a Microsoft PowerPoint environment, perform the following steps:

1. On the Windows machine in which the MarkLogic Add-in for PowerPoint is installed, open the Control Panel (for example, from the Start menu).

2. In the Control Panel, click the Uninstall a program link (Windows Vista) or use the Add/Remove Programs Control Panel (Windows XP or Server 2003) to uninstall the MarkLogic_PowerPoint program.

2.2.5 Modifying the Registry to Change the URL during Development

The URL with which the MarkLogic Add-in for PowerPoint communicates with MarkLogic Server is stored in the Windows Registry, and the installer updates that registry key during installation. During development, it is sometimes convenient to change the URL (or some of the other parameters) without creating a new installer, uninstalling the old one, and reinstalling the new one. You can modify the registry directly to change these key values. For example, run the following command from a Windows command prompt to change the URL key to `http://localhost:8888`:

```
REG ADD HKEY_CURRENT_USER\MarkLogicAddinConfiguration\PowerPoint /v URL /d
http://localhost:8888/
```

3.0 Install the CPF Pipelines

You can think of a .pptx document as essentially a bag of XML. It's a .zip file, which holds various XML documents that are the serialization of a PowerPoint Presentation in XML format. When you save to MarkLogic Server, you can extract the individual pieces of any Open XML package (.dotx, .pptx, .xlsx) by enabling the 'Open XML Extract' pipeline in Content Processing. (Note: Open XML Extract requires the Status Change Handling pipeline be enabled as well.)

Included in the toolkit are two additional pipelines, PresentationML Tags Process and PresentationML Process. PresentationML Tags Process will dereference the tags for tagged presentations, slides, and slide components, setting the tag names and values found in associated tag#.xml parts n the .pptx package within the document properties for slide#.xml parts.

Additionally, the PresentationML Process pipeline sets document properties on slide#.xml from extracted presentations that refer to associated images. Likewise, the pipeline sets properties on associated images that reference the original slide#.xml from the extracted presentation.   It doesn't matter which is saved first, slide or image, but the location and naming of images must adhere to certain conventions for properties to be set for the related images and slides.

Searches for content will most likely occur on the slide#.xml part. By setting associated image and tag information in the document properties, we can quickly dereference the parts we want from the extracted .pptx in MarkLogic.

This chapter covers the following topics:

- Installing the Pipelines
- Naming Conventions
- Properties Set
- Relation to Add-In and Opportunities for New Applications

3.1 Install the Pipelines

You'll find the required files in the /cpf directory of the Toolkit .zip. To install the pipeline:

1. Copy pptx-set-tags-action.xqy to
   <MarkLogicInstallDir>/ Modules/MarkLogic/conversion/actions/pptx-set-tags-action.xqy on your MarkLogic Server.

2. Copy map-slide-action.xqy to
   <MarkLogicInstallDir>/ Modules/MarkLogic/conversion/actions/map-slide-action.xqy on your MarkLogic Server.

3. Copy presentation-ml-support.xqy and presentation-ml-support-content-types.xqy to:
<MarkLogicInstallDir>/ Modules/MarkLogic/openxml
on your MarkLogic Server.

4. Install the PresentationML Tags Process and Presentation ML Process Pipelines

    a. Open the Admin Interface of MarkLogic Server

    b. Navigate to Databases -> <database-name> -> Content Processing -> Pipelines

    c. Click the "load" tab

    d. In the "directory" field, enter the path to the /cpf directory of the Toolkit .zip

    e. Click Ok

    The pipelines will load. You now just have to enable them for your database.

5. Enable the Pipelines

    f. Navigate to Databases -> <database-name> -> Content Processing -> Domains -> Default Documents -> Pipelines

    g. Check the checkboxes for the PresentationML Tags Process and Presentation ML Process Pipeline

    h. Click "ok"

    By default, the scope is root "/", with a depth of infinity. This is just a quick guide to get you up and running. For more information on working with and configuring CPF (the Content Processing Framework), please see the CPF documentation on our developer zone: http://developer.marklogic.com/pubs/4.0/default.xqy.

- The PresentationML Tags Process pipeline will pick up documents in state "initial" and once processed, set their state as "tagged".
- The PresentationML Process pipeline will act on documents in state "tagged", and once processed, will set their state as "slidemapped"

You may choose to just use the PresentationML Process pipeline and not the Tags pipeline. To do this, the following must be completed:
- A) Disable/Delete the PresentationML Tags Process pipeline in the Admin UI
- B) Delete the PresentationML Process pipeline in the Admin UI
- C) Update the PresentationML Process pipeline initial state to "initial" in presentationml-pipeline.xml and reload it through the AdminUI

3.2 Naming Conventions

A presentation, foo.pptx, extracted using the Office Open XML Extract pipeline, will have the presentation parts extracted to a directory /foo_pptx_parts/. This directory location is a sibling to where the original.pptx is saved in MarkLogic. The contents of the directory are the unzipped presentation, maintaining the naming and structure the files had in the .pptx.

For the PresentationML Process pipeline to be useful to you, and for properties to be set on slides and associated images, images must be saved in a sibling directory to the extracted parts directory for the presentation. The name of the directory should be the name of the presentation, with the file extension replaced with _PNG. An example directory for foo.pptx and /foo_pptx_parts/ is: /foo_PNG/.

Slides within the _PNG directory, must be name Slide1.PNG, Slide2.PNG …SlideN.PNG, where the slide image number corresponds to the number of the slide.xml found within the extracted parts. (/foo_pptx_parts/ppt/slides/slide1.xml, etc.)

When slides are extracted and saved to the server, the PresentationML process pipeline looks for an associated image, following these conventions. If an image is found, document-properties are set on both the slide#.xml and Slide#.PNG. Likewise, when images are saved following these conventions, the PresentationML Process pipeline checks for an associated slide#.xml for the image. If the slide#.xml is found, document properties are set on both the slide#.xml and Slide#.PNG.

3.2 Properties

Properties set by the PresentationML Process Pipeline are set in the ppt module namespace, "http://marklogic.com/openxml/powerpoint".

slide#.xml

```
 <ppt:pptx>/Foo.pptx</ppt:pptx>                          //URI of source .pptx
<ppt:pptxdir>/Foo_pptx_parts/</ppt:pptx>                 //URI of directory of extracted .pptx
<ppt:slideimg>/Foo_PNG/Slide10.PNG</ppt:slideimg>  //URI of associated slide image
<ppt:index>10</ppt:index>                                //Index of Slide
```

Slide#.PNG

```
<ppt:pptx>/Foo.pptx</ppt:pptx>                      //URI of related source .pptx
<ppt:pptxdir>/Foo_pptx_parts/</ppt:pptx>            //URI of directory of extracted .pptx
<ppt:slide>                                         //URI of related source slide#.xml
     /Foo_pptx_parts/ppt/slides/slide10.xml
</ppt:slide> //URI of related source slide#.xml
<ppt:index>10</ppt:index>                           //Index of slide
```

Properties set by the PresentationML Tags Process Pipeline are also set in the ppt module namespace, "http://marklogic.com/openxml/powerpoint".

presentation.xml: only set for presentation tags

```
<ppt:presentationtags>                       // a list of presentation tags
  <ppt:tag ppt:rid="rId9">                   // @rid = relationship id for associated tagN.xml file
    <ppt:tagname>LESSON</ppt:tagname>                         //tagname
    <ppt:tagval>PPT128398027882810419</ppt:tagval>          //tagvalue
    <ppt:tagfile>/FOO_pptx_parts/ppt/tags/tag1.xml</ppt:tagfile>  //associated tag file
    <ppt:custompart>/FOO_pptx_parts/customXml/item2.xml</ppt:custompart>
                                                         //associated metadata part
  </ppt:tag>
</ppt:presentationtags>
```

slide#.xml: set for any slidetags, shapetags, or presentationtags

```
<ppt:slidetags>                          // a list of slidetags
  <ppt:tag ppt:rid="rId1">
    <ppt:tagname>MASTER</ppt:tagname>
    <ppt:tagval>PPT128398030720335821</ppt:tagval>
    <ppt:tagfile>/FOO_pptx_parts/ppt/tags/tag2.xml</ppt:tagfile>
    <ppt:custompart>/FOO_pptx_parts/customXml/item1.xml</ppt:custompart>
  </ppt:tag>
</ppt:slidetags>
```

```
<ppt:shapetags>                                    // a list of shapetags
  <ppt:shape>
    <ppt:tag ppt:rid="rId2">
        <ppt:tagname>LOCALIZATION</ppt:tagname>
        <ppt:tagval>PPT128398031606240610</ppt:tagval>
        <ppt:tagfile>/FOO_pptx_parts/ppt/tags/tag3.xml</ppt:tagfile>
        <ppt:custompart>/FOO_pptx_parts/customXml/item6.xml</ppt:custompart>
    </ppt:tag>
  </ppt:shape>
</ppt:shapetags>

<ppt:presentationtags>                             // a list of presentation tags
  <ppt:tag ppt:rid="rId9">
    <ppt:tagname>LESSON</ppt:tagname>
    <ppt:tagval>PPT128398027882810419</ppt:tagval>
    <ppt:tagfile>/FOO_pptx_parts/ppt/tags/tag1.xml</ppt:tagfile>
    <ppt:custompart>/FOO_pptx_parts/customXml/item2.xml</ppt:custompart>
  </ppt:tag>
</ppt:presentationtags>
```

3.3 Relation to Add-In and Opportunities for New Web Applications

The Add-In enables the ability to save presentations to MarkLogic from the Button menu. This save actually saves both the presentation as .pptx and the associated _PNG images using PowerPoint. If the Pipeline is installed and configured, you can use the properties on slide#.xml returned from searches to quickly fetch and display the images in any web based application. Likewise, the properties on images can be used in a content application to quickly grab the associated slide#.xml.

By saving presentations to MarkLogic from the Button menu, and with the PresentationML Process pipeline enabled, you now have a way using document properties to help you quickly build webapps for search and re-use of slides. This search and reuse is not limited to use within the PowerPoint client. Using the XQuery API, you can mix and match presentations, merging slides, and generating new presentations on the Server, without ever opening PowerPoint.

4.0 Getting Started With the MarkLogic Add-in for PowerPoint

This chapter describes how to quickly get started using the MarkLogic Add-in for PowerPoint. It is a good starting place to understand what the MarkLogic Add-in for PowerPoint does, and includes the following sections:

- Download and Unzip the Installation Package
- Create an HTTP App Server in MarkLogic Server
- Customize the MarkLogic Add-in for PowerPoint Installer
- Install the MarkLogic Add-in for PowerPoint
- Install the PresentationML Process CPF Pipeline
- Sample Application
- Button Menu

4.1 Download and Unzip the Installation Package

If you have not already done so, download the MarkLogic Add-in for PowerPoint installation package from developer.marklogic.com. After you download it, unzip it into a convenient location (for example, c:/tmp).

4.2 Create an HTTP App Server in MarkLogic Server

Create an HTTP Server in MarkLogic Server to use as an App Server for the MarkLogic Add-in for PowerPoint. For example, log into the Admin Interface, go to Groups > Default > App Servers and click the Create HTTP button. Note the server and port number, as you will need that to modify the installer with the URL of this server.

4.3 Customize the MarkLogic Add-in for PowerPoint Installer

Use the procedure in "Creating a Custom Installer for the MarkLogic Add-In for PowerPoint" on page 7 to customize your MarkLogic Add-in for PowerPoint installation. You must customize the URL with the one to your MarkLogic Server App Server, and you can modify any of the other settings, too, if you want.

4.4 Install the MarkLogic Add-in for PowerPoint

Install your customized MarkLogic Add-in for PowerPoint as described in "Installing the MarkLogic Add-in for PowerPoint on Clients with Microsoft PowerPoint" on page 8.

4.5 Install the PresentationML Process and PresentationML Tags Process CPF Pipelines

Install the PresentationML Process and PresentationML Tags Process pipelines, as described in section 3.0 "Install the PresentationML Process and PresentationML Tags Process CPF Pipelines" on pg. 10.

4.6 Sample Applications

The sample applications are included in the `Samples/` directory of the zip file. These applications allow you to search within PowerPoint 2007/2010 documents expanded and stored in MarkLogic Server as well as Word 2007/2010 and Excel 2007/2010 documents. Sample code is provided on an as-is basis; the sample code is not intended for production applications and is not supported. For details, including setup instructions, see the `README.txt` file and the `samples-license.txt` file in the `Samples` directory of the zip file.

The sample application allows you to search Word and Excel documents (WordprocessingML, SpreadsheetML) as well as PowerPoint presentations unzipped and expanded in PresentationML format.

NOTE: *The Toolkit for PowerPoint Samples leverage solutions developed previously for the Toolkits for Word and Excel. For the Search samples to work properly, the following pipelines are required. These should be installed, configured, and attached to your database, similar to the PresentationML Process pipeline we installed and configured in Section 3.0.*

Required pipelines and descriptions:

Status Change Handling:
Included with MarkLogic Server. This pipeline handles CPF transitions.

Office Open XML Extract:
Included with MarkLogic Server. This pipeline unzips Office 2007/2010 documents on save to the database.

WordprocessingML Process:
Included with MarkLogic Server. This pipeline merges split runs of text in paragraphs to improve search and resuse of text within document.xml, the main body of content in a .docx package.

SpreadsheetML Process:
Included with MarkLogic Toolkit for Excel®. This pipeline saves text from the SharedStrings table into its associated worksheet. This is done so search can be performed against worksheets, without having to join and dereference the SharedStrings.xml part in the .xlsx package.

PresentationML Process:
Included with MarkLogic Toolkit for PowerPoint® . This pipeline is detailed thoroughly in this document in Section 3.0. Document properties are set on slides and associated images.

PresentationML Tags Process:
Included with MarkLogic Toolkit for PowerPoint® . This pipeline is detailed thoroughly in this document in Section 3.0. Document properties are set on slides for tagged presentations, slides, and shapes.

The samples also demonstrate how to save metadata in a custom XML piece to a PowerPoint Presentation, as well as the ability to save Presentations directly to MarkLogic Server from the Button menu.

4.6.1 Setup

The following is a quick-start guide to installing and configuring the Add-in With PowerPoint.

Assuming Office 2007/2010 is already installed on your system, to install the Add-in for PowerPoint, just double-click setup.exe. As with any good Windows app, click Next, Next, Next, and you'll be installed and ready to use the Add-in. If any prerequisites are missing, you'll be prompted to download and install them. Once the prerequisites are installed, the Add-in for PowerPoint will install with a default configuration. A couple of simple updates (described below), and we'll be up and running with the Samples.

The following will walk you through setting up an HTTP server and configuring the sample code to run within the task pane in PowerPoint.

1. Copy presentation-ml-support.xqy and presentation-ml-support-content-types.xqy to <ServerInstallationDir>/Modules/MarkLogic/openxml

2. Copy the Sample directory to a location on the Server that you will serve using your HTTP Server. example: C:\Program Files\MarkLogic\Docs\Samples

3. Create an HTTP Server in MarkLogic Server, and set its root directory to the Samples dir included in the .zip

    a. On MarkLogic Server, In the Admin interface, Navigate to "Groups" -> "Default" -> "App Servers"

    b. Click the "Create HTTP" tab

    c. Enter a servername: example: "PowerPoint-samples"

    d. Enter the root: this is the path of where the /Samples directory is found. example: " Program Files\MarkLogic\Docs\Samples"

    e. Enter a port: example: 9001

f. Set database to the database where your expanded PowerPoint files are located

g. Click "OK" to save the above information

4. Set the URL for the Add-in to the server you just created

a. In Regedit, navigate to "HKEY_CURRENT_USER" -> "MarkLogicAddinConfiguration" -> "PowerPoint"

b. set the URL property value to the url where the samples are located: example: http://localhost:9001

c. Set the User, and Auth values for the username andpassword of the server you're connecting with. These are required for Button Save functionality.

d. The default RibbonXqy property value should be sufficient as upload.xqy is found in Samples/utils. If you move or edit the file or change the name, be sure to update this value appropriately.

5. You're almost ready to start using the Samples. Just make the quick updates required in the next sections **(4.6.1.2)** and you'll be up and running. They're quick, simple updates, but warrant a little detail.

6. Now start the PowerPoint application, the Samples default page should be located in the pane on the right side of PowerPoint. This default page includes links to all the samples, and provides brief descriptions.

Samples include: Search, Office Search, and Metadata. Saving is built into the Button menu (and can be accessed through the javascript api for the pane). Click the link to navigate to the individual Sample app.

4.6.1.2 Update search.js, officesearch.js

The Samples include demonstrations of how to open a PowerPoint document in MarkLogic Server into PowerPoint, and likewise the ability to save a document directly to MarkLogic Server from PowerPoint.

Note: There are a few ways to open documents from MarkLogic Server and save documents to MarkLogic Server by using the pane and/or Microsoft Office. In the sample, we've provided one example from the Button menu. This way saves associated images at the time the .pptx is saved to MarkLogic using PowerPoint. If you check the api documentation, you'll find functions for opening/closing presentations through the Add-in to/from WebDAV. Also, if you setup a WebDAV client, you can open and save presentations directly from WebDAV using PowerPoint's native open/save/close functionality, and no access through a pane is required.

To use the included Samples:

1. update search.js

    a. Change the uname/pwd for the following lines to the credentials of the HTTP Server you configured above.

       ```
       MLA.insertImage(picuri,"uname","pwd");
       MLA.insertSlide(tmpPath, filename,slideidx, url, "uname","pwd",retain);
       MLA.openPPTX(tmpPath, filename, url, "uname","pwd");
       ```

2. update officesearch.js

    a. Change the uname/pwd for the following lines to the credentials of the HTTP Server you just configured above.

       ```
       MLA.insertImage(picuri,"uname","pwd");
       MLA.insertSlide(tmpPath, filename,slideidx, url, "uname","pwd",retain);
       MLA.openPPTX(tmpPath, filename, url, "uname","pwd");
       MLA.embedOLE(tmpPath, title, url, "uname","pwd");
       ```

Remember, these are Samples, and we aim to keep it simple. There are ways you can obfuscate the login information, but we leave that as a task/choice for the Add-in developer based on their requirements, and which functions they choose to use.

4.6.2 Search Sample

You are provided with a search box. Underneath the search box are radio buttons for the options Slides, Images, and Presentations. Enter some text and hit "enter" or click "search" to perform a search against documents in MarkLogic Server.

A. A search will be performed in the Server for any Slides that contain the text you've entered. The results returned depend on the option selected.

    a. Slides
       The images for saved slides that meet the search criteria are displayed in the pane. Double-click the image and the corresponding slide will be inserted to the current slide position in the presentation. If you check 'Retain Format', the slide will maintain the look from its original source presentation. If left unchecked, the slide will be inserted and will take on the look for the layout defined in the current presentation.

Though images are displayed, the actual slide is inserted into the presentation and is editable.

b. Images

This searches the document properties of image files. Double-Click to insert the image into the slide currently displayed in PowerPoint.

c. Presentations

A list of Presentations which contain slides that meet the search criteria is displayed. Underneath the URI for the presentation in MarkLogic, is some metadata about the presentation that is selected from the properties part of the extracted Presentation package. Click the URI link to have the Presentation open in PowerPoint.

4.6.2 Office Search Sample

You are provided with a search box. Enter some text and hit "enter" or click "search" to perform a search against documents in MarkLogic Server. A search will be performed for any PowerPoint Slides, Excel Worksheets, or Word paragraphs that meet the search criteria. Results all come with an option to insert, open, or embed into the active presentation. (Embed for PowerPoint is disallowed so the option is grayed out.)

A. ***PowerPoint***: If a slide matches the search criteria, you'll see the slide image in the search results. Underneath are the options to Insert Slide or Open Document. Select an option and click the slide image. There is no option to Retain Format here, slides inserted here will take on the appearance of the target presentation theme.

B. ***Excel***: If a worksheet matches the search criteria, you'll be provided a preview of the row that matched the criteria. Underneath this row are the options to Insert Table, Open Document, or Embed Document. Select an option and click the table preview to get the desired results.

Insert Table: The preview row will be inserted into the presentation using a PowerPoint table.

Open Document: The Workbook that is returned for the search will open into Excel.

Embed Document: The Workbook matching the search criteria will be embedded into the active slide in PowerPoint.

C. ***Word***: If a paragraph matches the search criteria, you will see a snippet of the paragraph, with the options to Insert Text, Open Document, or Embed Document. Select an option and click the snippet to get the desired results.

Insert Text: The text will be inserted into the slide at the cursor position. You must have the cursor selecting an area for text insert.

Open Document: The document that has the paragraph that matches the search criteria will open into Word.

Embed Document: The document that contains the paragraph that matched the search criteria will be embedded into the active slide in PowerPoint.

4.6.3 Metadata Sample

You can add a custom metadata document as a custom XML piece to the active presentation and save it within the .pptx package.

1. Enter values in the fields, and click "Add Metadata" to add the metadata document to the package.

2. If you edit the metadata, and click "Add Metadata", the metadata will be updated for the document.

3. Click "Remove", to remove the custom piece entirely from the package.

The sample saves the information as dublin core metadata. But you can add any well-formed XML as a custom piece. You don't have to use dublin core. To do this, you just have to create your own form for editing/saving the metadata to the document. See the code for the example.

While the metadata is saved with the package, it is only accessible to users using the Add-in who are able to expose the values using the pane. Once saved to MarkLogic, you can now search on that custom metadata to find your presentations.

4.7 Button Menu

From the Button menu you are provided the options to select either 'Save To MarkLogic' or 'Save To MarkLogic As'. If the presentation has not been previously saved anywhere (it only exists in memory) you'll be provided a simple form to enter a name. Enter a name for your presentation in the form and click Save. The presentation will be saved as .pptx to the MarkLogic Server you've configured with the Add-in. A set of associated images for the slides will be saved at the same time. You can then use Search to find your documents and slides. Slides are made available immediately for re-use in new presentations.

You don't need to add the .pptx extension to the filename before saving, it will automatically be appended. If you add it, that's ok, it won't be added twice.

Save As always prompts the user with the simple form to name the file. The simple form and save functionality are all accessible through the javascript API as well.

If the presentation already has a name (it's been saved locally, and is opened in PowerPoint from a file on the client system) , Save To MarkLogic will not prompt for a new document name. The name of the document will be used as the name for the presentation in MarkLogic.

5.0 Generating Images for Slides

We realize you may have presentations and slides you'd like to start searching and re-using immediately. Having image representations of each slide is helpful for displaying search results. Rather than opening each presentation individually in PowerPoint and saving to MarkLogic from the Button menu options, (which not only saves the presentation to MarkLogic but also generates the slide images), a simple utility is provided for generating presentation slide images. This chapter covers:

- OMPM (office migration planning manager)
- Toolkit image generator for .pptx

5.1 OMPM

To search and reuse PowerPoint using the Toolkit, the presentations need to be in the PowerPoint 2007/2010 .pptx format. Microsoft provides a utility, the Office Migration Planning Manager (OMPM) that is a bulk utility for converting pre-2007 office documents to the latest 2007 formats. We suggest running this on legacy PowerPoint presentations to generate presentations in the newer .pptx format. It's a smart and helpful utility. It's also simple to use and you can run reports first to see which presentations might fail for conversion, prior to converting so you can take any necessary action to help them convert properly.

To learn more about OMPM and download, visit:

http://technet.microsoft.com/en-us/library/cc179179.aspx

5.2 Toolkit image generator for .pptx

The toolkit image generator will convert a directory (and sub-directories) of PowerPoint 2007/2010 files into the respective .PNG format in a *_PNG directory. This is a very simple utility, provided for convenience.

The utility opens presentations in PowerPoint in a background process. It's using PowerPoint to save the slide images as .PNG files. So you need to run this on a machine with PowerPoint installed.

There are two files in /image-util: MarkLogic_PowerPoint_Images.exe and save-images.bat.

Update save-images.bat by replacing C:\PresentationsToConvert with the path and directory name where you're .pptx for conversion are located. The directory path and name must not have any spaces.

The second parameter is true/false. This is just to enable debug messages in the console as the images are saved.

You can use save-images.bat to execute the .exe, either from the command line, or by double clicking the .bat.

The utility will open any .pptx, .ppsx, .pptm, .ppsm, .potx, or .potm files found in the specified directory, and its subdirectories. The presentation will have a _PNG dir of associated image files saved as a sibling to the original file. You can then load these into MarkLogic using WebDAV, the server's own load functionality in the AdminUI, or any other load utility.

Note: As this utility uses PowerPoint, make sure you execute on a system that has PowerPoint 2007/2010 installed. Also, the utility works best if you disable any Add-Ins enabled in PowerPoint before use. You can disable Add-Ins by navigating to :

Button -> PowerPoint Options -> Add-Ins -> Select Com Add-Ins and click "Go…".

In the dialog box that appears, uncheck any Add-Ins that don't need to be loaded while PowerPoint is used to generate the images. After using the utility, be sure to re-enable the Add-Ins that are important to you.

There are other ways to generate images for slides as well. And then there is the size of images, quality, etc. This utility is provided to help jump start your PowerPoint development. The toolkit is open source and the code is available on the developer site. If you modify the utility code or roll your own image generator, just remember to take into account the naming conventions used by the PresentationML Process pipeline, the Samples, etc. to retain functionality.

6.0 Developing Applications for the MarkLogic Add-in for PowerPoint

This chapter describes some techniques for building XQuery applications that communicate with the MarkLogic Add-in for PowerPoint, and includes the following section:

- JavaScript Library
- XQuery Libraries

6.1 JavaScript Library

The JavaScript library is used to communicate between a MarkLogic Server application and Microsoft PowerPoint. It has functions that provide interaction with a Microsoft PowerPoint document from within the embedded browser in the Add-in pane. The library is included in the installation zip package under `js/MarkLogicPowerPointAddin.js`. For the function signatures of the APIs in the JavaScript library, see the JavaScriptDoc in the `docs/jsdocs` directory of the installation zip package.

6.2 XQuery Libraries

When you create a MarkLogic Server application that communicates with PowerPoint, you can use any of the functions from the MarkLogic Server XQuery library. For example, there are two XQuery libraries that have functions directly useful for Office 2007/2010-based applications:

- The ZIP Package library: `package.xqy`
- The WordProcessingML library: `word-processing-ml.xqy`

These libraries are installed under the `/Modules/MarkLogic/openxml` directory in the MarkLogic Server installation directory.

Additionally, there is a library of XQuery functions designed to support the Microsoft PresentationML format. This library is included in the Toolkit for PowerPoint installation zip package as `xquery/presentation-ml-support.xqy` and `presentation-ml-support-content-types.xqy`. Likewise, copy these to `/Modules/MarkLogic/openxml`.

The documentation for this library is included in the zip package as `docs/xquery-apidoc/presentation-ml-support.html`.
There are a small number of functions in this library, but they do a lot of heavy lifting for you. You can generate presentations on the server from existing presentations using this library. PowerPoint is not required to leverage these tools until you go to actually open the presentation.