


Object Ingest Tool

 There is new documentation that explains the Ingest Processing

General set up

The difference between this tool and the [Ingest Tool](#) is that Ingest Tool uses an elaborate mix of config file, EAD collection and transformations and file names to ingest items correctly. This tool uses a simple configuration file, directories and file names to perform ingests.

Running the ingest tool

The program requires three parameters:

- Path to collection directory: collDir
- Path to item directory: itemDir
- Path to config directory: configDir
 - Repository.properties for fedora configuration

Image Collection Ingests

In the structure below, the Hohenberger directory is the collDir and photos is the itemDir.

```
-- Hohenberger
|
|- IngestConfig.properties
|- ATM-MC2-7-1-1-10
  |- ATM-MC2-7-1-1-10.tif
  |- ATM-MC2-7-1-1-10-full.jpg
  |- ATM-MC2-7-1-1-10-screen.jpg
  |- ATM-MC2-7-1-1-10-thumb.jpg
  |- ATM-MC2-7-1-1-10-j2k.jp2 // Optional scalable image
  |- ATM-MC2-7-1-1-10.txt // Optional OCR'ed page text
  |- ATM-MC2-7-1-1-10-mods.xml
  |- ATM-MC2-7-1-1-10-dc.xml
  |- mets-properties.xml // optional
  |- policy.xml // optional
|- ATM-MC2-7-1-1-11
|- ...
```

Paged Document Ingests

This directory structure shows how each book item should be laid out in the file system. Files belonging to each paged document is in a directory where the directory name is the assigned ID.

```

-- MassDigitization
|- IngestConfig.properties
|- VAA4276
  |- VAA4276-0001
    |- VAA4276-0001.tif
    |- VAA4276-0001-full.jpg
    |- VAA4276-0001-screen.jpg
    |- VAA4276-0001-thumb.jpg
    |- VAA4276-0001.txt           // OCR'ed page text
    |- mets-properties.xml       // optional
    |- policy.xml                // optional
  |- VAA4276-0002
  |- ...
|-metadata
  |- VAA4276-marc.xml
|-pdf
  |- VAA4276.pdf
|-text
  |- VAA4276.xml                 // e.g. TEI
|-mets-properties.xml           // optional
|- policy.xml                   // optional
|- VAA4592
|- ...

```

Multi Copy Paged Document Ingests

In this case, the object hierarchy has three levels: manifest->paged doc->page image. In the structure below, isl-aad-8761 is the manifestation level object and defines its own metadata. Sheet books are at the book level (isl-aad-8761-01 and isl-aad-8761-03). Page/Image level objects are under these.

```

-- isl
|- IngestConfig.properties
|- isl-aad-8761
  |-metadata
    |- isl-aad-8761-mods.xml
    |- isl-aad-8761-dc.xml
  |-isl-aad-8761-01
    |-isl-aad-8761-01-01
      |- isl-aad-8761-01-01.tif
      |- isl-aad-8761-01-01-full.jpg
      |- isl-aad-8761-01-01-screen.jpg
      |- isl-aad-8761-01-01-thumb.jpg
      |- isl-aad-8761-01-01.txt           // OCR'ed page text
      |- mets-properties.xml             // optional
      |- policy.xml                     // optional
      |- ...
    |-isl-aad-8761-01-01
    |- ...
  |-pdf
    |- isl-aad-8761-01.pdf
    |- mets-properties.xml               // optional
    |- policy.xml                       // optional
  |-isl-aad-8761-03
  |- ...
  |- mets-properties.xml                 // optional
  |- policy.xml                         // optional
|- isl-aad-8765
|- ...

```

Journal Ingests

There are four levels: volume, issue, article and page image. In the structure below, VAA4025-060 is the volume identifier and VAA4025-060-4 is the issue identifier.

```

-- imh
|- IngestConfig.properties
|- VAA4025-060 // volume level
  |- metadata
    |- VAA4025-060-tei.xml // TEI header
  |- VAA4025-060-4 // issue level
    |-VAA4025-060-4-001 // these are pages
      |- VAA4025-060-4-001.tif
      |- VAA4025-060-4-001-full.tif
      |- VAA4025-060-4-001-screen.tif
      |- VAA4025-060-4-001-thumb.tif
      |- ...
      |- mets-properties.xml // optional
      |- policy.xml // optional
    |-VAA4025-060-4-002
      |- ...
    |-articles
      |-VAA4025-060-4-a01
        |-page-list.txt // flat file with 1 page-id per line
        |-pdf
          |-VAA4025-060-4-a01.pdf
        |-metadata
          |-VAA4025-060-4-a01-mods.xml
          |-VAA4025-060-4-a01-dc.xml
          |-VAA4025-060-4-a01-tei.xml // TEI Independent header.
        |- mets-properties.xml // optional
        |- policy.xml // optional
      |-metsnav
        |- VAA4025-060-4-metsnav.xml
      |-text
        |- VAA4025-060-4.xml
      |- mets-properties.xml // optional
      |- policy.xml // optional

```

Multi Volume Paged Ingests

```

--inauthors
|- IngestConfig.properties
|- VAA3765
  |- metadata
    (formats to be defined)
  |- VAA3765-1 // Volume 1
    |-metadata
      (volume level metadata, tbd)
    |-pdf
      |-VAA3765-1.pdf
    |-VAA3765-1-001 // Page 001
      |- VAA3765-1-001.tif
      |- VAA3765-1-001-full.tif
      |- VAA3765-1-001-thumb.tif
      |- VAA3765-1-001-screen.tif
      |- VAA3765-1-001.txt // OCR text
    |-text
      |- VAA3765-1.xml // TEI
  |- mets-properties.xml
  |- policy.xml

```

Configuration

Removed the outdated config attachment.

See the attached IngestConfig.properties file for a list of configuration items.

File

Modified 

Microsoft Powerpoint Presentation Ingest Processing Explained.pptx
Ingest processing documentation

Nov 04, 2008 by Muzaffer
Ozakca
