

Automatic Exposure Capturing Technical Metadata for Digital Still Images

Günter Waibel/RLG




Where museums, libraries, and archives intersect

The Digital Preservation Challenge

- Why digitize collections?
 - Access beyond geographical limitations
 - Added functionality
- The price
 - The more sophisticated, the more volatile
 - Engraved stone slab vs. a book vs. a digital text
- The challenge
 - Extending the life-span of digital information
- The gain (long-term)
 - Return on investment
 - Retention of unique born-digital information

Digital Preservation Strategies

- The more you know about a file, the better your chances for preserving it!
 - Digital Preservation = capturing massive amounts of Metadata
- Digital Preservation Standards
 - PREservation Metadata: Implementation Strategies (PREMIS)
 - File-format neutral preservation metadata element set
 - Data Dictionary to be issued December 2004
 - NISO Z39.87: Technical Metadata for Digital Still Images
 - File-format specific technical metadata element set
 - Draft Standard issued June 1, 2002
 - Balloted Standard to be issued December 2004  RLG

NISO Z39.87 Technical Metadata for Digital Still Images

- NISO Z39.87-2002 (AIIM 20-2002) [Draft Standard for Trial Use]
- In XML – MIX (NISO Metadata for Image in XML Schema)
- “First line of defense” against obsolescence
- Sections
 - **Basic image parameters** record information crucial to displaying a viewable image
 - **Image creation** metadata records information crucial to understanding the technical environment in which a digital image file was captured
 - **Imaging performance assessment** metadata records information that allows evaluation of the digital image’s quality, or output accuracy
 - **Change history** metadata records information about the processes applied to an image over its life cycle
- NISO Z39.87 [revisions as of September 16th 2004]
 - 124 data elements
 - 38 Mandatory / Mandatory if Applicable

Automatic Exposure – An RLG initiative

- Overarching goal: Economic implementation of NISO Z39.87
 - Minimize the cost of technical metadata acquisition
 - Maximize the ability to ensure long-term access to digital images
- Initiative supported by
 - Digital Library Federation (DLF)
 - Museum Computer Network (MCN)
- For documents & updates
 - <http://www.rlg.org/> > Quick Links > Current Projects

Automatic Exposure investigations

- What do we want?
 - Capture mechanism: a host for the metadata
 - Editing mechanism: a place to add metadata
 - Export mechanism: a way to transfer metadata from the file to a preservation database
- Leveraging existing specifications
 - Available Metadata
 - What technical metadata do we currently have access to?
 - Mapping Z39.87 to TIFF, EXIF, JPEG 2000 (JPX)
 - Extraction Tools
 - How can technical metadata be extracted for transfer into preservation databases?
- Expanding to NISO Z39.87
 - What mechanisms can we identify which could give us access to the full NISO Z39.87 element set?

Leveraging Existing Specifications – Available Metadata

- TIFF (the present)
 - Option 1: Technical metadata in fileheader tags
 - Option 2: EXIF 2.2
 - Extension of TIFF fileheader tags (JPEG/TIFF files)
- JPEG 2000 (the future)
 - Metadata based on DIG35

Quantitative Mapping, Complete

	NISO Z39.87 <i>Complete</i>	TIFF 6.0 /TIFF EP Mapping Elements	TIFF w/ EXIF 2.2 Mapping Elements	JPEG 2000 Mapping Elements
Basic Image Parameters	40	16	13	8
Image Creation	41	9	19	34
Imaging Performance Assessment	36	22	16	7
Change History	7	2	1	6
Total	124	49	49	55

Quantitative Mapping, M/MA

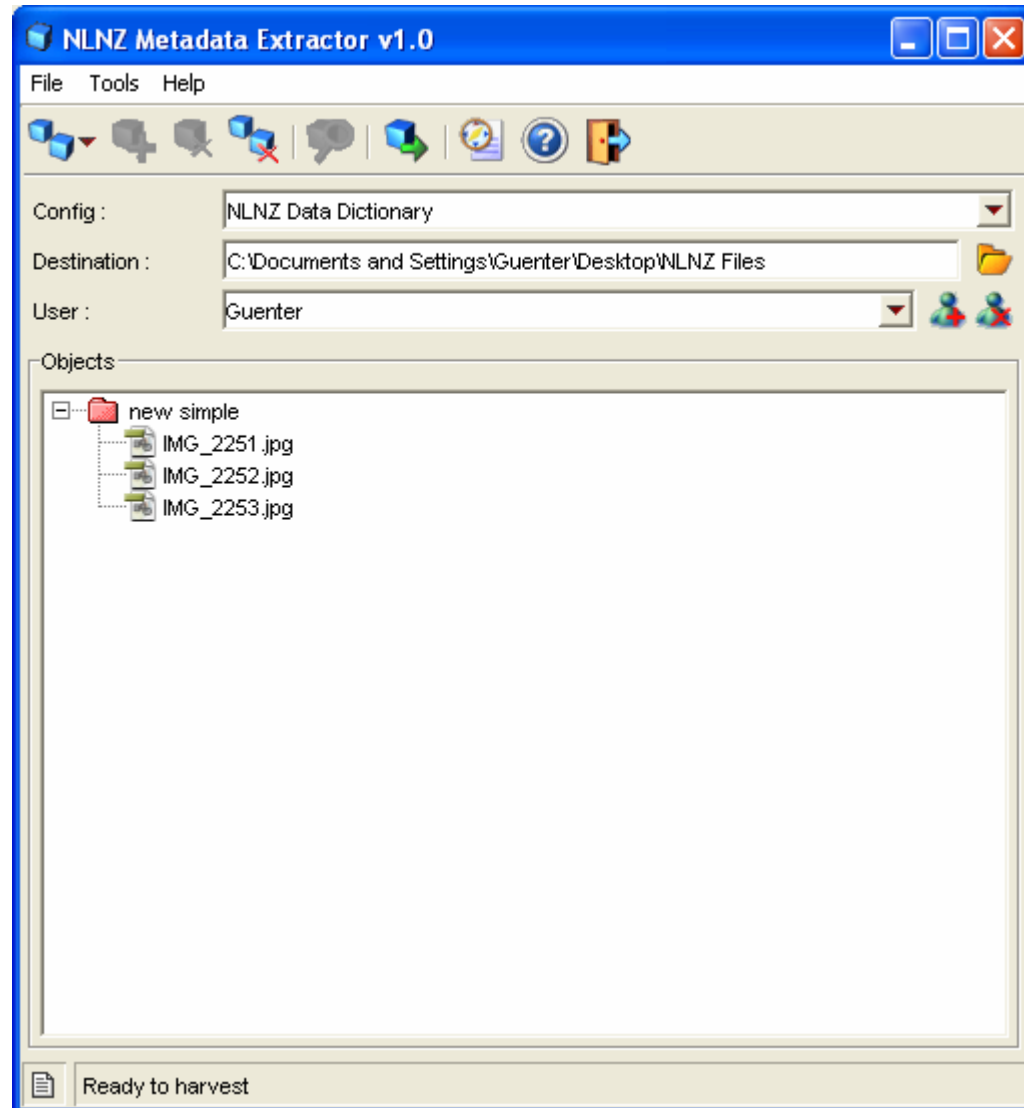
	NISO Z39.87 M/MA	TIFF 6.0 /TIFF EP Mapping Elements	TIFF w/ EXIF 2.2 Mapping Elements	JPEG 2000 Mapping Elements
Basic Image Parameters	19	7	9	7
Image Creation	2	1	1	1
Imaging Performance Assessment	15	12	8	3
Change History	2	2	1	3
Total	38	22	19	17

M/MA = Mandatory/Mandatory if applicable

Leveraging Existing Specifications – Extraction Tools

- Community based
 - JHOVE
 - The JSTOR-Harvard Object Validation Environment
 - National Library of New Zealand
 - “Metadata Extractor”
- Industry based
 - Eastman Kodak Picture Metadata Toolkit
 - Adobe Extensible Metadata Platform (XMP)

Example 1: National Library of New Zealand



Extracted XML record

```
<Object>
  <Name>new simple</Name>
  <ID>0</ID>
  <ReferenceNumber/>
  <GroupIdentifier/>
  <PersistentIdentifier/>
  <MasterCreationDate locale="PDT">
    <Date format="yyyyMMdd">20040412</Date>
    <Time format="HHmmssSSS">160939254</Time>
  </MasterCreationDate>
  <ObjectComposition>simple</ObjectComposition>
  <StructuralType>
    <Name/>
    <Extension/>
  </StructuralType>
  <HardwareEnvironment>x86</HardwareEnvironment>
  <SoftwareEnvironment>OS: Windows XP 5.1, JVM: Sun Microsystems Inc. 1.4.1_02</SoftwareEnvironment>
  <InstallationRequirements/>
  <AccessInhibitors/>
  <AccessFacilitators/>
  <Quirks/>
  <MetadataRecordCreator>Gunter</MetadataRecordCreator>
  <MetadataCreationDate locale="PDT">
    <Date format="yyyyMMdd">20040412</Date>
    <Time format="HHmmssSSS">160939264</Time>
  </MetadataCreationDate>
  <Comments/>
  <Files>
    <File xmlns:nz_govt_natlib_xsl_XSLTFunctions="nz.govt.natlib.xsl.XSLTFunctions">
      <FileIdentifier/>
      <Filename>
        <Name>IMG_2253.jpg</Name>
        <Extension>jpg</Extension>
      </Filename>
      <FormerFilename>
        <Name/>
        <Extension>jpg</Extension>
      </FormerFilename>
      <Size>2912354</Size>
      <FileDateTime>
        <Date format="yyyyMMdd">20040412</Date>
        <Time format="HHmmssSSS">154354353</Time>
      </FileDateTime>
    </File>
  </Files>
</Object>
```

Example 2: Adobe XMP

- Open-Source specification for sharing metadata across applications
- Extracts existing metadata (TIFF, EXIF, DIG35)
- Embeds metadata as an XMP packet (XML)
- Access for viewing / editing metadata
 - Adobe Photoshop File Info
- Option to customize metadata set
 - Adding fields through a custom panel
 - RLG will create & share NISO Z39.87 custom panel
- Export metadata
 - Individual file: “Save” on File Info – Advanced Screen
 - Batch: Script and Droplet
 - Creates XML file

Automatic Exposure - Evaluation

- Glass half full / half empty
 - 45-60% overlap on M/MA
 - Some extraction tools available
- Theory vs Practice
 - Not all possible fields of a specification used by capture devices

Automatic Exposure - Progress

- International Imaging Industry Association (I3A) IT10 Meeting (September)
 - TIFF – no hope for aligning header structure with NISO Z39.87
 - JPEG 2000 – promise of the future
 - additional Technical Metadata in codestream
 - RLG working with Kodak to update mapping
- American Museums Digital Imaging Survey Benchmarking Conference @ RIT (September)
 - Sinar, Digital Transitions (PhaseOne), Xerox show interest
 - Since then: Creo
 - Old friends: BetterLight, Kodak

Automatic Exposure – Next Steps

- Leveraging
 - “Scorecard” for Tools
 - Communicate to community options for technical metadata extraction
 - Adobe XMP custom panel
 - JPEG 2000 revised mapping
- Expanding to NISO Z39.87
 - Keep advocating with industry
 - Rally the cultural heritage community to exert pressure
 - Example: Barbara Bridgers (MET)

Thanks!

- More Info
 - <http://www.rlg.org/> > Quick Links > Current Projects
 - Guenter.Waibel@notes.rlg.org
 - Robin.Dale@notes.rlg.org