INTRODUCTION TO DATA MIGRATION

Trish Rose-Sandler
VRA Conference Kansas City, MO
March 28 2007

Definition

- "Data migration is the *transferring of data* between storage types, formats, or computer systems. Data migration is usually performed programmatically to achieve an *automated migration*, freeing up human resources from tedious tasks. It is required when organizations or individuals change computer systems or upgrade to new systems."

  Wikipedia 2/21/07

3 Stages of Data Migration

- **Pre Migration**
- **Migration**
- **Post Migration**

3 Stages of Data Migration

- **Pre Migration – Analyzing, Mapping, Normalizing/Transforming, Testing, Backup**
- **Migration**
- **Post Migration – Quality Control, Cleanup, Update Cataloging Guidelines**

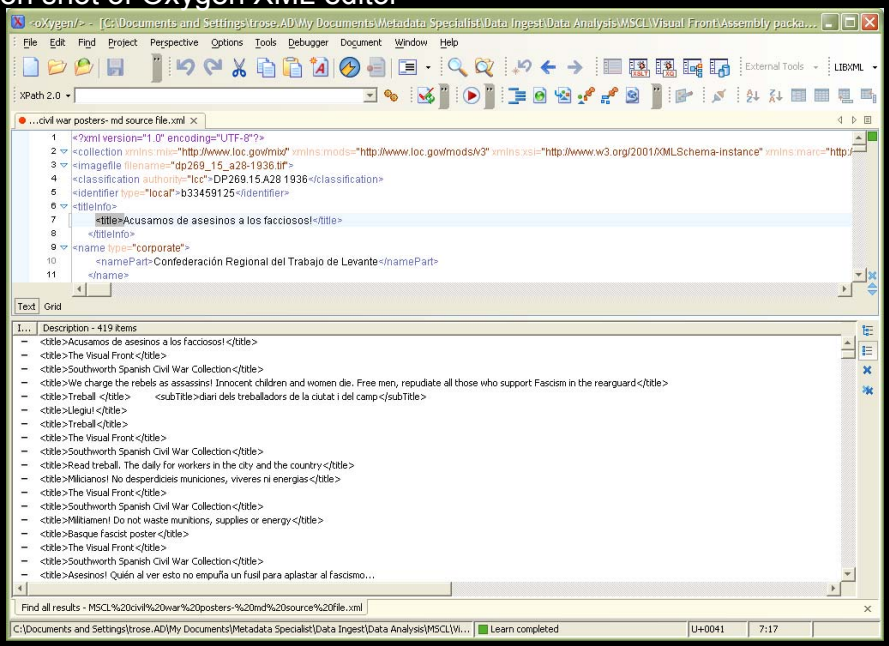
Pre-Migration: Data Analysis

Evaluate data for:

- **Consistency**
- **Unnecessary redundancy across records**
- **Identify errors**
- **Re-evaluate relationships & structures**

Pre-Migration: Data Analysis

Screen shot of Oxygen XML editor

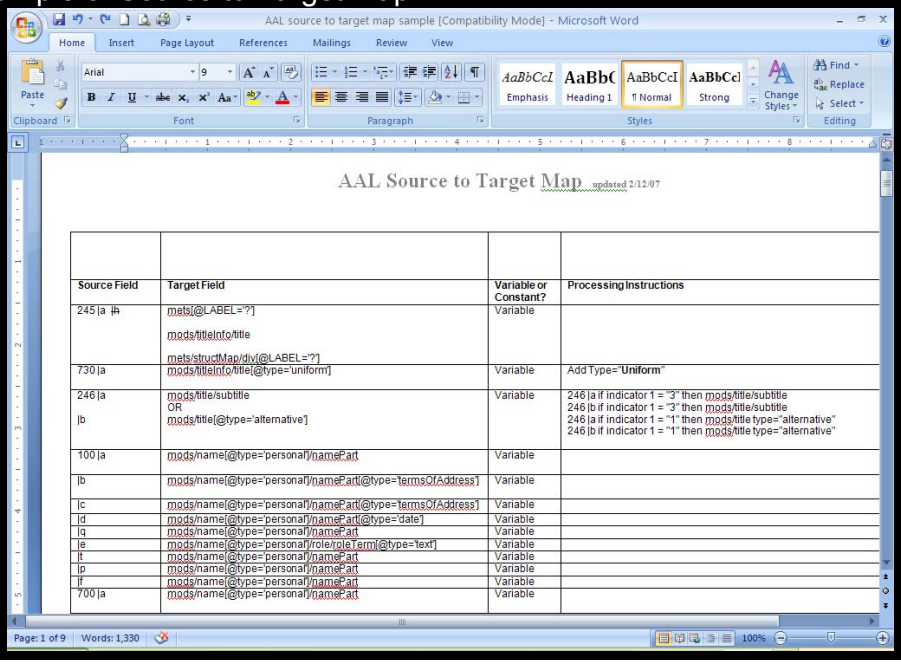


Pre-Migration: Mapping

Source		Target
CreatorName	→	vra:agentName dc:creator
Date	→	vra:date dc:date
Site	→	vra:locationName dc:coverage
General Notes	→	vra:description dc: description

Pre-Migration: Mapping

Example of Source to Target map



AAL source to target map sample [Compatibility Mode] - Microsoft Word

AAL Source to Target Map updated 2/12/07

Source Field	Target Field	Variable or Constant?	Processing Instructions
245 |a |h	mets[@LABEL='?'] mods/titleInfo/title mets/structMap/div[@LABEL='?']	Variable	
730 |a	mods/titleInfo/title[@type='uniform']	Variable	Add Type="**Uniform**"
246 |a |b	mods/title/subtitle OR mods/title[@type='alternative']	Variable	246 |a if indicator 1 = "3" then mods/title/subtitle 246 |b if indicator 1 = "3" then mods/title/subtitle 246 |a if indicator 1 = "1" then mods/title type="alternative" 246 |b if indicator 1 = "1" then mods/title type="alternative"
100 |a	mods/name[@type='personal']/namePart	Variable	
|b	mods/name[@type='personal']/namePart[@type='termsOfAddress']	Variable	
|c	mods/name[@type='personal']/namePart[@type='termsOfAddress']	Variable	
|d	mods/name[@type='personal']/namePart[@type='date']	Variable	
|q	mods/name[@type='personal']/namePart	Variable	
|e	mods/name[@type='personal']/role/roleTerm[@type='text']	Variable	
|t	mods/name[@type='personal']/namePart	Variable	
|p	mods/name[@type='personal']/namePart	Variable	
|f	mods/name[@type='personal']/namePart	Variable	
700 |a	mods/name[@type='personal']/namePart	Variable	

Page: 1 of 9 Words: 1,330

Pre-Migration: Mapping

Source Set

- Examples of parsing
  - Source data stores all agent names in creator field w/o a qualifying role
    - Creator
  - Target data stores all agent names separate from but linked to roles
    - Agent.name
    - Agent.role

Pre-Migration: Mapping

Target Set

Mix of community standards and local elements

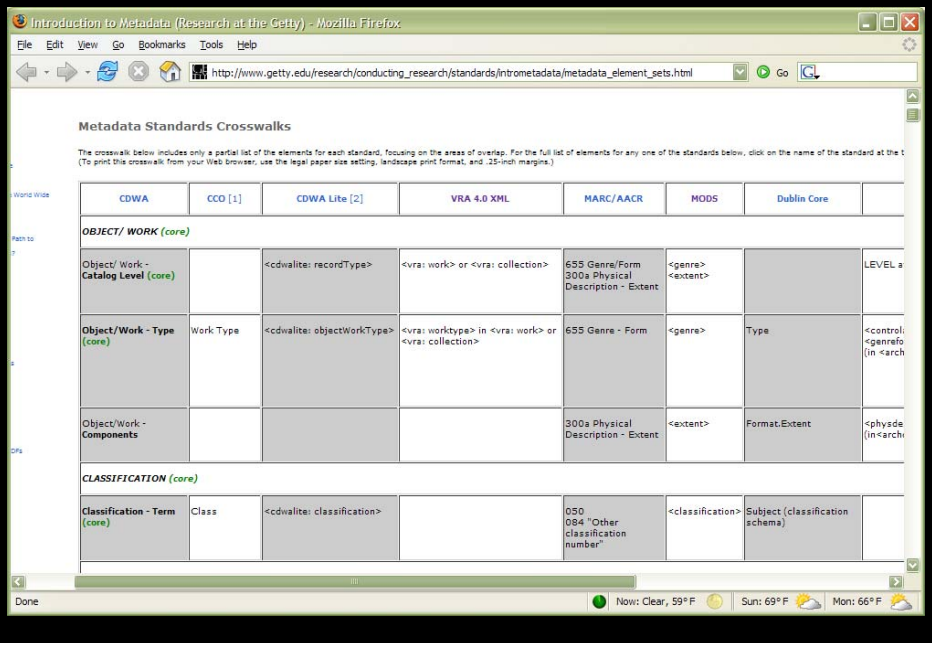
- Descriptive – e.g. Core 4.0
- Technical – e.g. Core 4.0, MIX
- Rights – e.g. Core 4.0, METS Rights
- Local – e.g. dateImagePurchased

Pre-Migration: Mapping

- Familiarize yourself with standards' documentation
  - Examples
  - VRA Core 4.0 standard
    - Core 4.0 documentation
    - Cataloging Cultural Objects (CCO) content standard
  - MODS standard
    - MODS documentation guidelines <http://www.loc.gov/standards/mods/>
    - DLF/Aquifer Implementation Guidelines for Shareable MODS Records <http://www.diglib.org/aquifer/dlfmodsimplementationguidelines_

Pre-Migration: Mapping

Crosswalks may be useful e.g. Getty crosswalk



Pre-Migration: Mapping

The goal of normalization is to transform or clean up your data values so they conform to accepted standards, are more consistent, and can be understood by any user of your images

Can be done during

- Pre Migration
- Migration
- Post Migration

Pre-Migration: Mapping

Data to Normalize

Abbreviations

e.g. Material= ol on cv.; source = DOA; attribution=sch of

Inconsistencies

e.g. expressions of circa: Date= c1947; c. 1988; c 500 AD;
ca. 15th Century

Formatting conventions

e.g. title=[Grapevines at Mission de San Ignacio]

Pre-Migration: Mapping

Other types of normalization

Source system uses codes for values

e.g. Name=Johnson, Ben type=1 (1=personal)

Create indexed versions of dates

e.g. 2/9/01, mid 15th century, October 1945
YYYY-MM-DD, YYYY-MM, YYYY (ISO 8601 standard)

Pre-Migration: Mapping

Assess metadata granularity

Examples of narrow and broad worktypes

Narrow: **City planning, urbanism, landscape design, garden design, environmental design**

Broad: Architecture

Narrow: **costume design, fashion design, clothing, jewelry, ornament, body decoration**

Broad: **Costume and Jewelry**

Pre-Migration: Testing

Hand pick records to evaluate mapping decisions and to test for normalization and diacritics problems

Diacritics- best to encode in UTF-8 Unicode or use Unicode decimal or hexadecimal character references

Display

Karls□tejn (Str□edoc□eský kraj, Czech Republic)--Castle

Exported using Unicode decimal character references

Karlštejn (Středočeský kraj, Czech Republic)--Castle

Pre-Migration: Final Word

BACK-UP, BACK-UP, BACK-UP

Migration: A few words

Enlist the help of a programmer or database administrator

Need for assistance will depend on your source and target systems (what tools they provide for migrating data), how much data normalization you'll need, restructuring of the data, etc.

A db administrator can help with target system setup (forms, reports, security, backup, etc)

Post Migration

Quality Control (QC)

Data Cleanup

Update Cataloging Guidelines