

# 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**

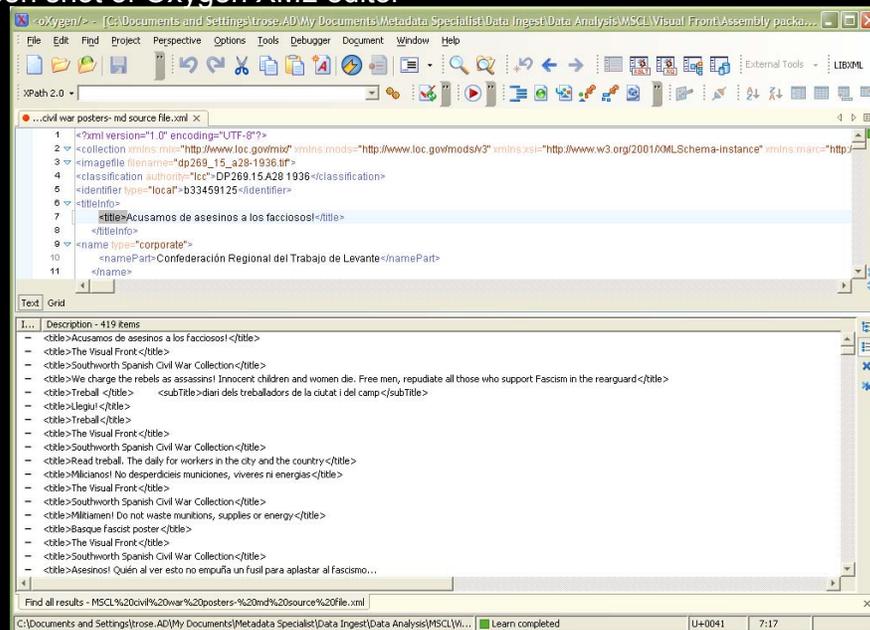
# Pre-Migration: Data Analysis

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

Source Field	Target Field	Variable or Constant?	Processing Instructions
245 [a] [b]	<code>mats{@LABEL=?}</code> <code>mods:titleInfo/title</code> <code>mats:structMap/div[@LABEL=?]</code>	Variable	
730 [a]	<code>mods:titleInfo/title[@type=uniform]</code>	Variable	Add Type="Uniform"
246 [a] OR [b]	<code>mods:title/subtitle</code> <code>mods:title[@type=alternative]</code>	Variable	246 [a] if indicator 1 = "3" then <code>mods:title/subtitle</code> 246 [b] if indicator 1 = "3" then <code>mods:title/subtitle</code> 246 [a] if indicator 1 = "1" then <code>mods:title type="alternative"</code> 246 [b] if indicator 1 = "1" then <code>mods:title type="alternative"</code>
100 [a]	<code>mods:name[@type=personal]/name:art</code>	Variable	
[b]	<code>mods:name[@type=personal]/name:art[@type=termsOfAddress]</code>	Variable	
[c]	<code>mods:name[@type=personal]/name:art[@type=termsOfAddress]</code>	Variable	
[d]	<code>mods:name[@type=personal]/name:art[@type=date]</code>	Variable	
[q]	<code>mods:name[@type=personal]/name:art</code>	Variable	
[e]	<code>mods:name[@type=personal]/role/terms[@type=text]</code>	Variable	
[f]	<code>mods:name[@type=personal]/name:art</code>	Variable	
[p]	<code>mods:name[@type=personal]/name:art</code>	Variable	
[r]	<code>mods:name[@type=personal]/name:art</code>	Variable	
700 [a]	<code>mods:name[@type=personal]/name:art</code>	Variable	

# 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/dlfformatsimpleimplementationguidelines/>

# Pre-Migration: Mapping

Crosswalks may be useful e.g. Getty crosswalk

Introduction to Metadata (Research at the Getty) - Mozilla Firefox

http://www.getty.edu/research/conducting\_research/standards/intrometadata/metadata\_element\_sets.html

Metadata Standards Crosswalks

The crosswalk below includes only a partial list of the elements for each standard, focusing on the areas of overlap. For the full list of elements for any one of the standards below, click on the name of the standard at the top of the table. (To print this crosswalk from your Web browser, use the legal paper size setting, landscape print format, and .25-inch margins.)

	CDWA	CCO [1]	CDWA Lite [2]	VRA 4.0 XML	MARC/AACR	MODS	Dublin Core
<b>OBJECT/ WORK (core)</b>							
Object/Work - Catalog Level (core)			<cdwalite: recordType>	<vra: work> or <vra: collection>	655 Genre/Form 300a Physical Description - Extent	<genre> <extent>	LEVEL a
Object/Work - Type (core)	Work Type		<cdwalite: objectWorkType>	<vra: worktype> in <vra: work> or <vra: collection>	655 Genre - Form	<genre>	Type <control: <genrefo (in <arch
Object/Work - Components					300a Physical Description - Extent	<extent>	Format: Extent <physde (in <arch
<b>CLASSIFICATION (core)</b>							
Classification - Term (core)	Class		<cdwalite: classification>		050 084 Other classification number <sup>1</sup>	<classification>	Subject (classification schema)

Done

Now: Clear, 59°F Sun: 69°F Mon: 66°F

# 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. 15<sup>th</sup> 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 15<sup>th</sup> 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

Karlsřtejn (Strředocřeský kraj, Czech Republic)--Castle

Exported using Unicode decimal character references

Karls&#780;tejn (Str&#780;edoc&#780;esky&#769; 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