If I knew then what I know now….

Building the new database for your migrated data

Susan Jane Williams,
Data Specialist, Scholars Resource; Independent Consultant and Developer

Building the target for your data

- In 2007, Many VR professionals will still be using off-the-shelf products to facilitate data entry to meet local needs
  - Lack of access to larger institutional computing resources
  - Specialized needs, including using specific standards (VRA Core, CCO)

The cataloging utility as a bridge

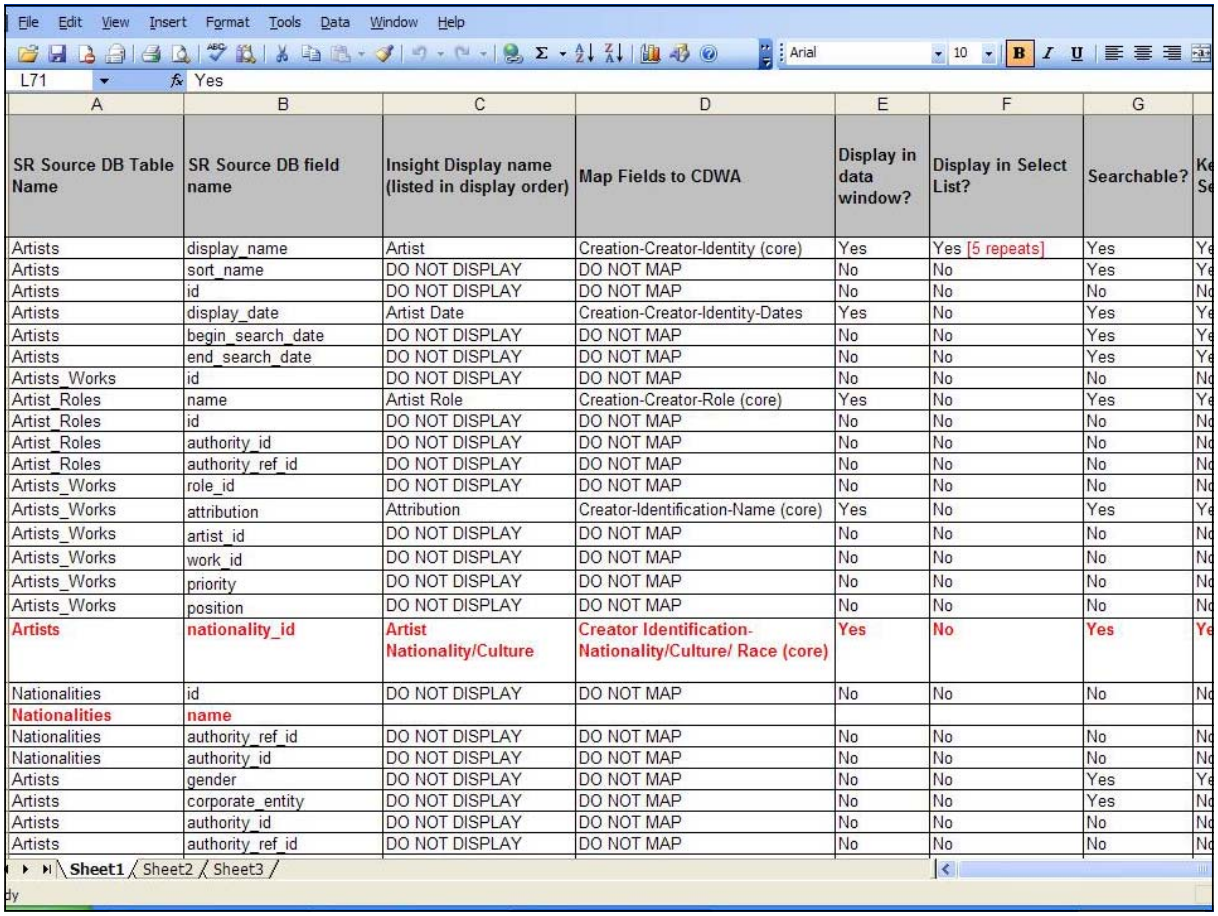
- Need to create local specialized data, but also the need to hook that to teaching tools
- The local institutional choice of DAM or DAP might not support the standard that you wish to use and may require exporting data from a cataloging utility to that system *at this time*

A “cataloging utility”

- That is, not only the table structure that holds the data….
- And not just the data structure that can be used in other applications….
- But, an understanding of the user interface that facilitates and guides data entry

So, you have done the first steps….

- Planning documents involving partners across your institution
- Data dictionaries
- Crosswalks of local collections

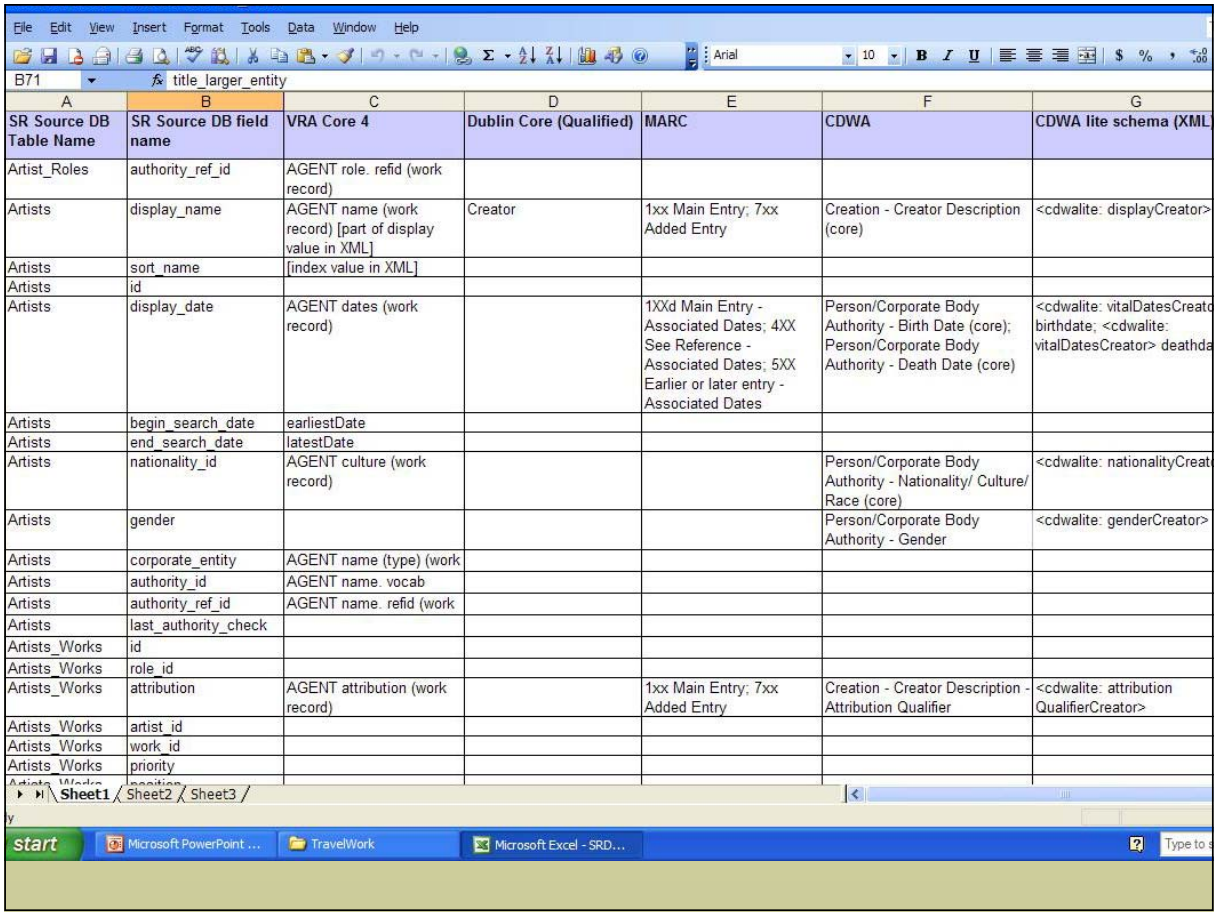


File Edit View Insert Format Tools Data Window Help

L71 Yes

A	B	C	D	E	F	G
SR Source DB Table Name	SR Source DB field name	Insight Display name (listed in display order)	Map Fields to CDWA	Display in data window?	Display in Select List?	Searchable?
Artists	display_name	Artist	Creation-Creator-Identity (core)	Yes	Yes [5 repeats]	Yes
Artists	sort_name	DO NOT DISPLAY	DO NOT MAP	No	No	Yes
Artists	id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists	display_date	Artist Date	Creation-Creator-Identity-Dates	Yes	No	Yes
Artists	begin_search_date	DO NOT DISPLAY	DO NOT MAP	No	No	Yes
Artists	end_search_date	DO NOT DISPLAY	DO NOT MAP	No	No	Yes
Artists_Works	id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artist_Roles	name	Artist Role	Creation-Creator-Role (core)	Yes	No	Yes
Artist_Roles	id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artist_Roles	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artist_Roles	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists_Works	role_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists_Works	attribution	Attribution	Creator-Identification-Name (core)	Yes	No	Yes
Artists_Works	artist_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists_Works	work_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists_Works	priority	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists_Works	position	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists	**nationality_id**	**Artist Nationality/Culture**	**Creator Identification-Nationality/Culture/ Race (core)**	**Yes**	**No**	**Yes**
Nationalities	id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Nationalities	**name**					
Nationalities	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Nationalities	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists	gender	DO NOT DISPLAY	DO NOT MAP	No	No	Yes
Artists	corporate_entity	DO NOT DISPLAY	DO NOT MAP	No	No	Yes
Artists	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No
Artists	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No

Sheet1 Sheet2 Sheet3



File Edit View Insert Format Tools Data Window Help

B71 title_larger_entity

A	B	C	D	E	F	G
SR Source DB Table Name	**SR Source DB field name**	**VRA Core 4**	**Dublin Core (Qualified)**	**MARC**	**CDWA**	**CDWA lite schema (XML)**
Artist_Roles	authority_ref_id	AGENT role. refid (work record)				
Artists	display_name	AGENT name (work record) [part of display value in XML]	Creator	1xx Main Entry; 7xx Added Entry	Creation - Creator Description (core)	<cdwalite: displayCreator>
Artists	sort_name	[index value in XML]				
Artists	id					
Artists	display_date	AGENT dates (work record)		1XXd Main Entry - Associated Dates; 4XX See Reference - Associated Dates; 5XX Earlier or later entry - Associated Dates	Person/Corporate Body Authority - Birth Date (core); Person/Corporate Body Authority - Death Date (core)	<cdwalite: vitalDatesCreato birthdate; <cdwalite: vitalDatesCreator> deathda
Artists	begin_search_date	earliestDate				
Artists	end_search_date	latestDate				
Artists	nationality_id	AGENT culture (work record)			Person/Corporate Body Authority - Nationality/ Culture/ Race (core)	<cdwalite: nationalityCreat
Artists	gender				Person/Corporate Body Authority - Gender	<cdwalite: genderCreator>
Artists	corporate_entity	AGENT name (type) (work				
Artists	authority_id	AGENT name. vocab				
Artists	authority_ref_id	AGENT name. refid (work				
Artists	last_authority_check					
Artists_Works	id					
Artists_Works	role_id					
Artists_Works	attribution	AGENT attribution (work record)		1xx Main Entry; 7xx Added Entry	Creation - Creator Description - Attribution Qualifier	<cdwalite: attribution QualifierCreator>
Artists_Works	artist_id					
Artists_Works	work_id					
Artists_Works	priority					

Sheet1 Sheet2 Sheet3

start Microsoft PowerPoint ... TravelWork Microsoft Excel - SRD...

The “art” of database design

- Cataloging is the single most expensive component of creating the digital asset
- A balance must sometimes be struck between choices one might make in guiding efficient and accurate data entry and the degree of standards adherence

What *you* bring to database design

- The understanding of your own workflow, local needs, patron concerns, level of expertise of your cataloging staff (professionals or students) will be key to building the right user interface for your cataloging utility—this is the next step beyond data dictionaries and other planning documents.

Flat versus Relational Databases

- “Flatfile” data is what we are used to seeing in spreadsheets
  - Multiple values are either expressed in separate columns: “Subject 1”, “Subject 2” or are run together in the same column with punctuation or other dividers: “Subject 1; Subject 2; Subject 3”

Excel sample

C	D	E	F	
Title.Variant.Work	**LOCATION**	**Classification**	**Artist_Sortname1**	**Artist_Sortnam**
Rotonde de Chartres	Paris, France	Architecture	Ledoux, Claude-Nicolas	
Ducal Palace	Dijon, France	Architecture	Mansart, Jules Hardouin	Gabriel, Ange-J
Ducal Palace	Dijon, France	Architecture	Mansart, Jules Hardouin	Gabriel, Ange-J

Relational Databases

- Relate information stored in multiple tables
- Ideally, there is no redundancy of data entry—each value that might be reused in data entry is only entered once and stored in one table that is *related* for use everywhere else in the database (made available anywhere needed in the data entry workflow)
- Numeric keys are normally used in this process

Sample of a table of related data

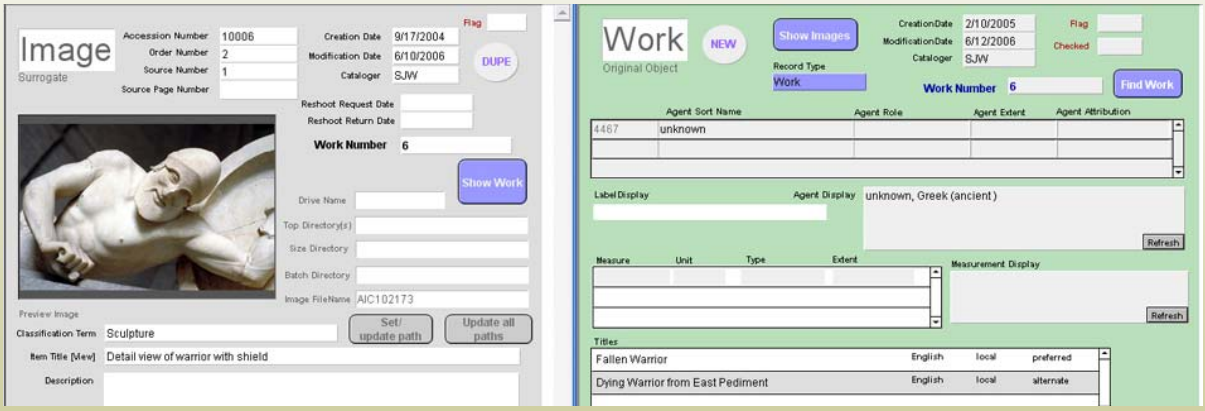
fk_AgentID	fk_WorkID	AgentRoles
4467	1	
	1	
4467	2	
4467	6	
4470	8	director
2906	9	designer
4315	9	designer
4471	11	designer
3876	8	costume designer
4468	9	designer
4469	7	designer
3562	14	architect
2257	15	architect
4467	10	
4266	16	architect
4266	17	architect
2727	3	painter
2727	5	painter

GUI interfaces to data tables

- Obviously, looking at rows and columns of related numeric keys is not user friendly, so most commercial databases allow you to build graphic user interfaces (GUI)—forms—for data entry

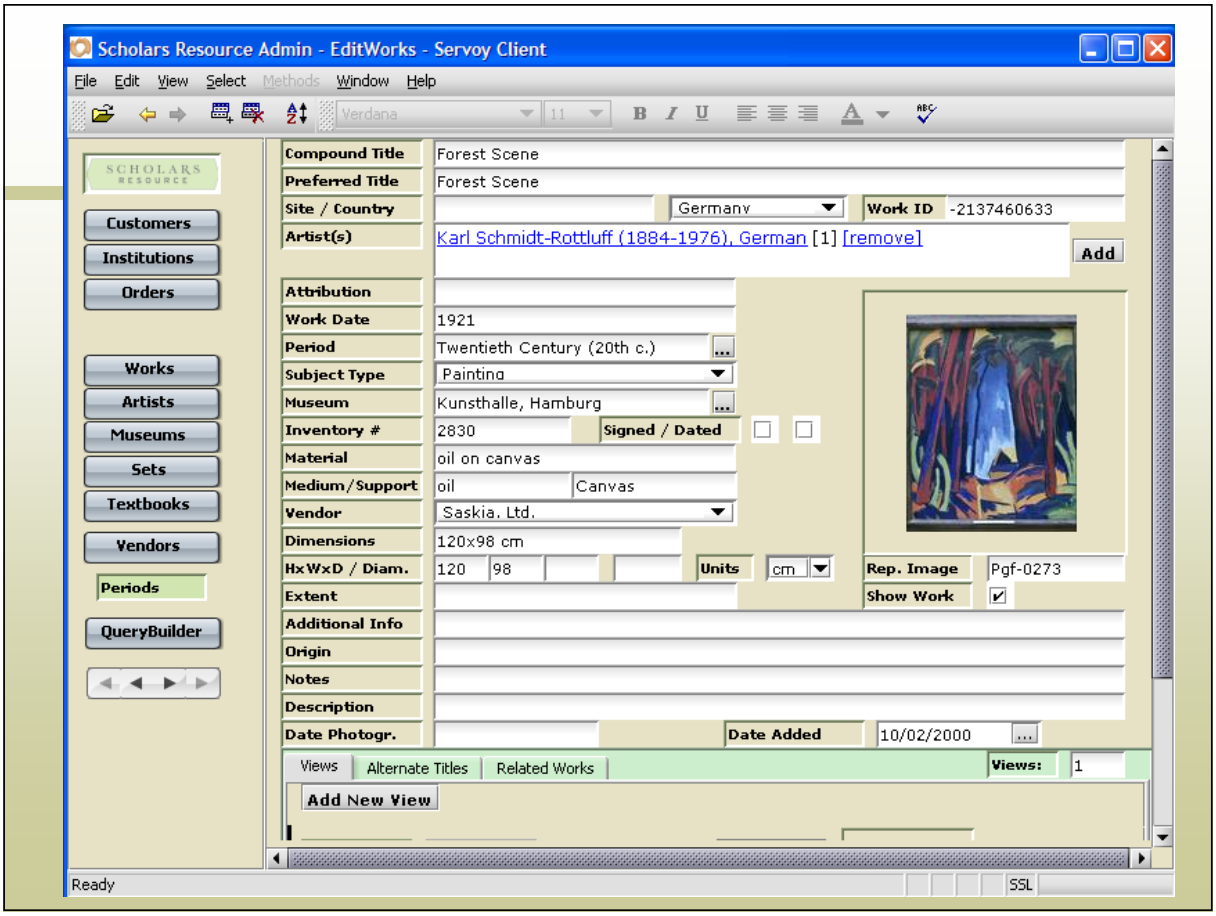
Work and Image entry layouts

- A clipped sample of 2 forms:



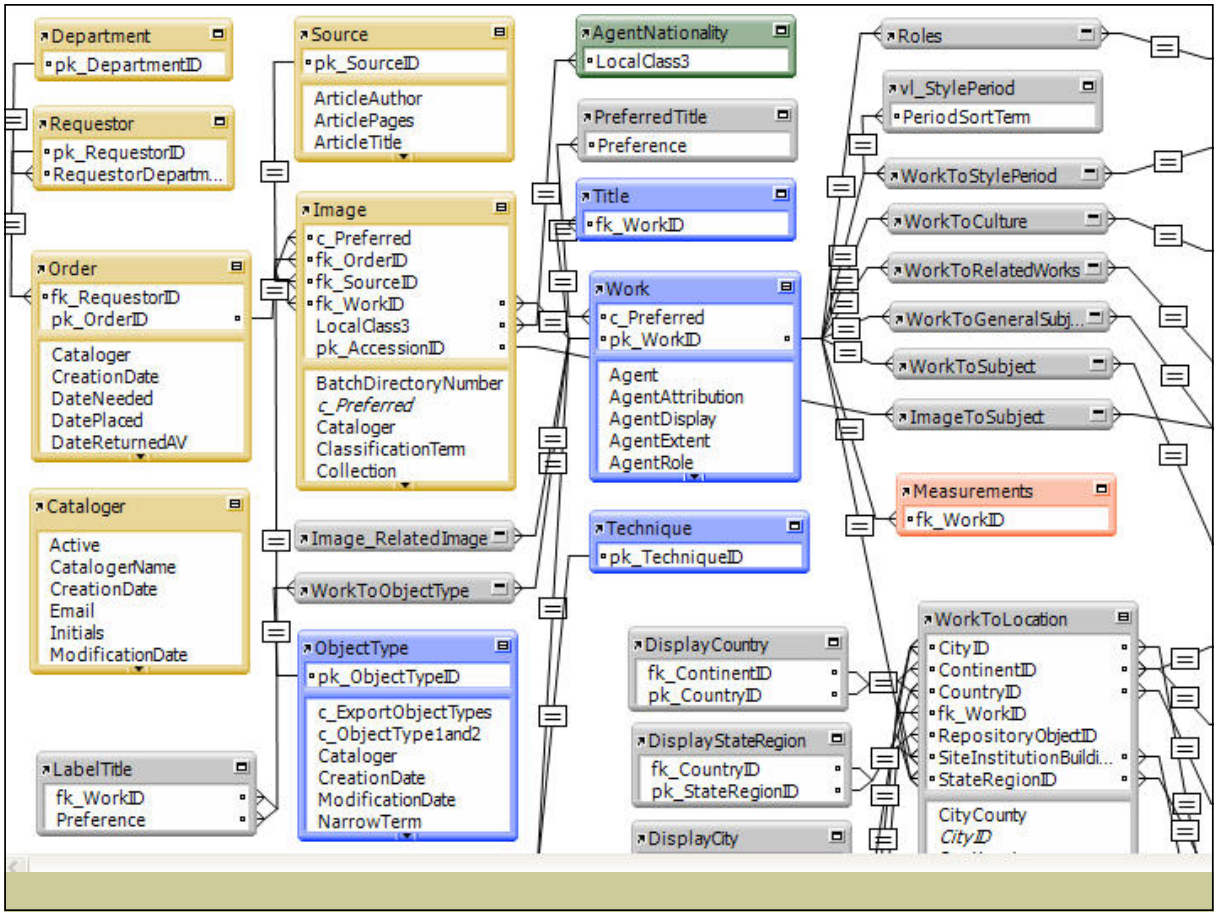
Splitting the interface design from the database (table) design

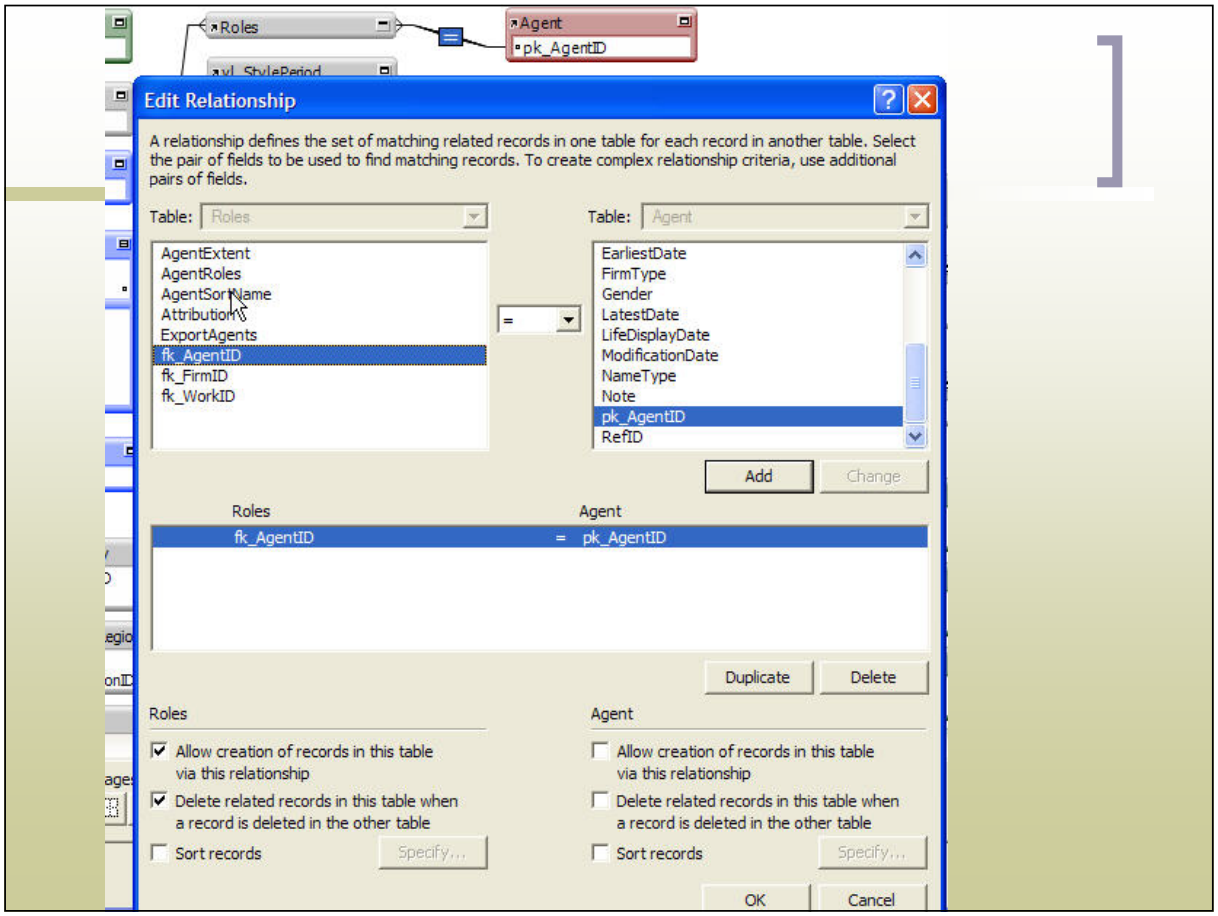
- A more sophisticated approach that may be more flexible
- Possible in either Filemaker 7+ or MS Access
- Possible with other tools such as Servoy (provides the functional interface to SQL or Oracle tables) (Sample screenshot, next slide)



ER (Entity Relationship) Diagrams

- Relational databases such as Filemaker Pro (FMP) or MS Access also use graphic tools to show specific fields in tables and tables related to each other in the entire database


Department
pk_DepartmentID
Requestor
pk_RequestorID
RequestorDepartm...
Order
fk_RequestorID
pk_OrderID
Cataloger
CreationDate
DateNeeded
DatePlaced
DateReturnedAV
Cataloger
Active
CatalogerName
CreationDate
Email
Initials
ModificationDate
LabelTitle
fk_WorkID
Preference
Source
pk_SourceID
ArticleAuthor
ArticlePages
ArticleTitle
Image
c_Preferred
fk_OrderID
fk_SourceID
fk_WorkID
LocalClass3
pk_AccessionID
BatchDirectoryNumber
c_Preferred
Cataloger
ClassificationTerm
Collection
Image_RelatedImage
WorkToObjectType
ObjectType
pk_ObjectTypeID
c_ExportObjectTypes
c_ObjectType1and2
Cataloger
CreationDate
ModificationDate
NarrowTerm
AgentNationality
LocalClass3
PreferredTitle
Preference
Title
fk_WorkID
Work
c_Preferred
pk_WorkID
Agent
AgentAttribution
AgentDisplay
AgentExtent
AgentRole
Technique
pk_TechniqueID
DisplayCountry
fk_ContinentID
pk_CountryID
DisplayStateRegion
fk_CountryID
pk_StateRegionID
DisplayCity
Roles
vl_StylePeriod
PeriodSortTerm
WorkToStylePeriod
WorkToCulture
WorkToRelatedWorks
WorkToGeneralSubj...
WorkToSubject
ImageToSubject
Measurements
fk_WorkID
WorkToLocation
CityID
ContinentID
CountryID
fk_WorkID
RepositoryObjectID
SiteInstitutionBuildi...
StateRegionID
CityCounty
CityID


Roles
Agent
pk_AgentID
Edit Relationship
A relationship defines the set of matching related records in one table for each record in another table. Select the pair of fields to be used to find matching records. To create complex relationship criteria, use additional pairs of fields.
Table: Roles
Table: Agent
AgentExtent
AgentRoles
AgentSortName
Attribution
ExportAgents
fk_AgentID
fk_FirmID
fk_WorkID
=
EarliestDate
FirmType
Gender
LatestDate
LifeDisplayDate
ModificationDate
NameType
Note
pk_AgentID
RefID
Add
Change
Roles
Agent
fk_AgentID = pk_AgentID
Duplicate
Delete
Roles
Allow creation of records in this table via this relationship
Delete related records in this table when a record is deleted in the other table
Sort records
Specify...
Agent
Allow creation of records in this table via this relationship
Delete related records in this table when a record is deleted in the other table
Sort records
Specify...
OK
Cancel

Portals and subforms

- Using forms/layouts, you can create “windows” looking into tables of related data (tables with relationships established between them) and showing multiple data values—more than one “answer” per record (multiple locations, agents, titles, etc. for each work). In FMP these are known as portals, in Access, subforms.

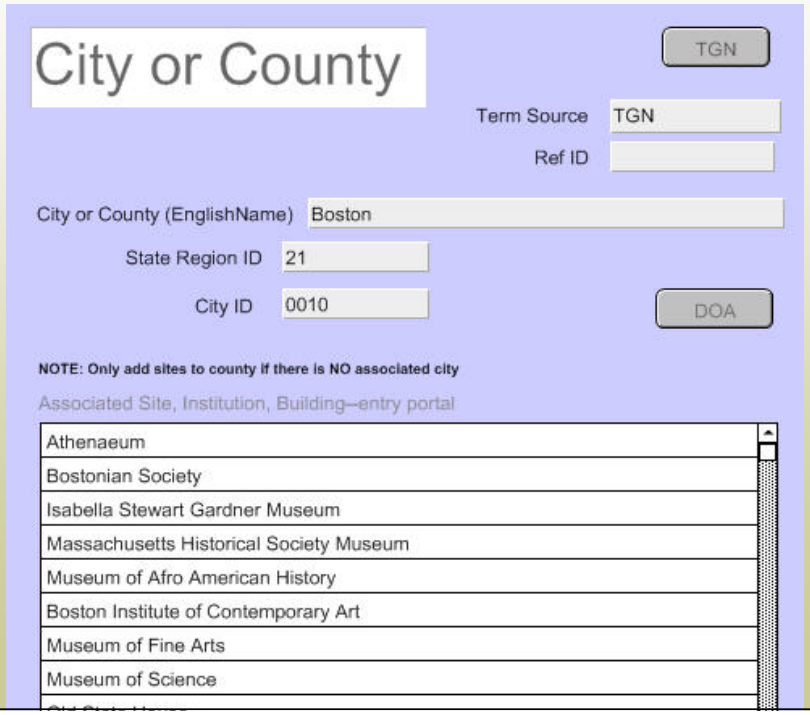
VireoCat Locations Portal (on Works form)

Location Entry

Click on a value to see its Authority Record...

Continent	Country	State/Region	City OR County	Site OR Building	RepositoryObjectID	Type
1 Africa	0074 Egypt	81 Upper Egypt	0022 Pyramids of Giza	0354 Pyramid complex of Mycerinus		formerGeogr aphic
5 North and	0252 United States	21 Massachusetts	0010 Boston	0165 Museum of Fine Arts	0003 11.1738	repository

Actual City table (in form view)



Understanding the new role of XML in data mobility

- XML facilitates pushing data between all sorts of applications
- CCO and VRA Core 4 were both formed with an eye to XML

What is XML?

- Extensible Markup Language (XML) is a universal language for sharing data between applications. XML is most appropriate for situations where the volume of data is generally small, as the data is transmitted as text, and controlling the structure of the data is important.

How does XML work?

- It “tags” data—identifies what that data is (what meaning it holds).

MARC tags by using numeric designators:

for instance a “245” field is always a title, a “700” or “7xx” field is a personal name (creator)

MARC example

245 00 |**a** [Asher Brown Durand, half-length portrait, three-quarters to the right, head three-quarters to the left, with both hands resting on cane, at right] [graphic].
260 __ |**c** [between 1845 and 1850]
300 __ |**a** 1 photograph : |**b** half plate daguerreotype, gold toned.
500 __ |**a** Identification from reproduction of daguerreotype in Newhall, History of Photography, 1949, p. 25.
500 __ |**a** Scratched on back of plate: 157; [undeciphered canceled inscription] [...eion]
500 __ |**a** Corners rounded.
520 __ |**a** Artist (engraver, painter), President of the National Academy of Design.
510 4_ |**a** Facing the light / H. Pfister. Washington : Smithsonian Institution Press, 1978, |**c** p. 313.
506 __ |**a** Original served by appointment only.
540 __ |**a** No known restrictions on publication.
541 __ |**c** Transfer; |**a** U.S. War College; |**d** 1920; |**e** (DLC/PP-1920:46153).
580 __ |**a** Forms part of: Daguerreotype collection (Library of Congress).
500 __ |**a** Produced by Mathew Brady's studio.
600 10 |**a** Durand, Asher Brown, |**d** 1796-1886.
655 _7 |**a** Portrait photographs |**y** 1840-1850. |**2** gmgpc
655 _7 |**a** Daguerreotypes |**y** 1840-1850. |**2** gmgpc
700 1_ |**a** Brady, Mathew B., |**d** 1823 (ca.)-1896.
773 0_ |**t** Daguerreotype collection (Library of Congress) |**w** (DLC) 95861318
852 __ |**a** Library of Congress |**b** Prints and Photographs Division |**e** Washington, D.C. 20540 USA |**n** dcu
856 41 |**3** b&w film copy neg. post-1992 |**d** cph |**f** 3c09970 |**q** p |**u** http://hdl.loc.gov/loc.pnp/cph.3c09970
856 41 |**3** b&w film copy neg. pre-1992 |**d** cph |**f** 3a13388 |**q** p |**u** http://hdl.loc.gov/loc.pnp/cph.3a13388

XML tags

- XML tags with natural language—easy to see what the information (the data value) is within the “chicken lips”

><

XML example

```
<!-- AGENT   -->
<set>
<display>Jasper Francis Cropsey (American painter, 1823-1900)</display>
<index>
<agent>
<name type="personal" vocab="ULAN" refid="500012491">Cropsey, Jasper
      Francis</name>
<dates type="life">
<earliestDate>1823</earliestDate>
<latestDate>1900</latestDate>
</dates>
<culture>American</culture>
<role vocab="AAT" refid="300025136">painter</role>
</agent>
</index>
</set>
```

Schema: Where the data standard and XML meet

Once a data standard like VRA Core 4.0 is devised, with all the elements and qualifiers laid out, the standard can then be expressed in one XML document called the schema—a road map to then apply to a specific XSLT style sheet that tells a database how to export data into XML

VRA Core 4.0 XML schema (a small sample)

```
<!-- Agent   -->
<xsd:complexType name="agentType">
<xsd:annotation><xsd:documentation>VRA Agent element.
    Subelements are used for different types of data (names, roles,
    dates, etc.). At least one subelement must be
    provided.</xsd:documentation>
</xsd:annotation>
<xsd:sequence minOccurs="1" maxOccurs="unbounded">
<xsd:element name="attribution" type="basicString" minOccurs="0" />
<xsd:element name="culture" type="basicString" minOccurs="0" />
<xsd:element name="dates" type="agentDateType" minOccurs="0" />
<xsd:element name="name" type="agentNameType" minOccurs="0" />
<xsd:element name="role" type="basicString" minOccurs="0" />
</xsd:sequence>
<xsd:attributeGroup ref="vraAttributes" />
```

What is XSLT?

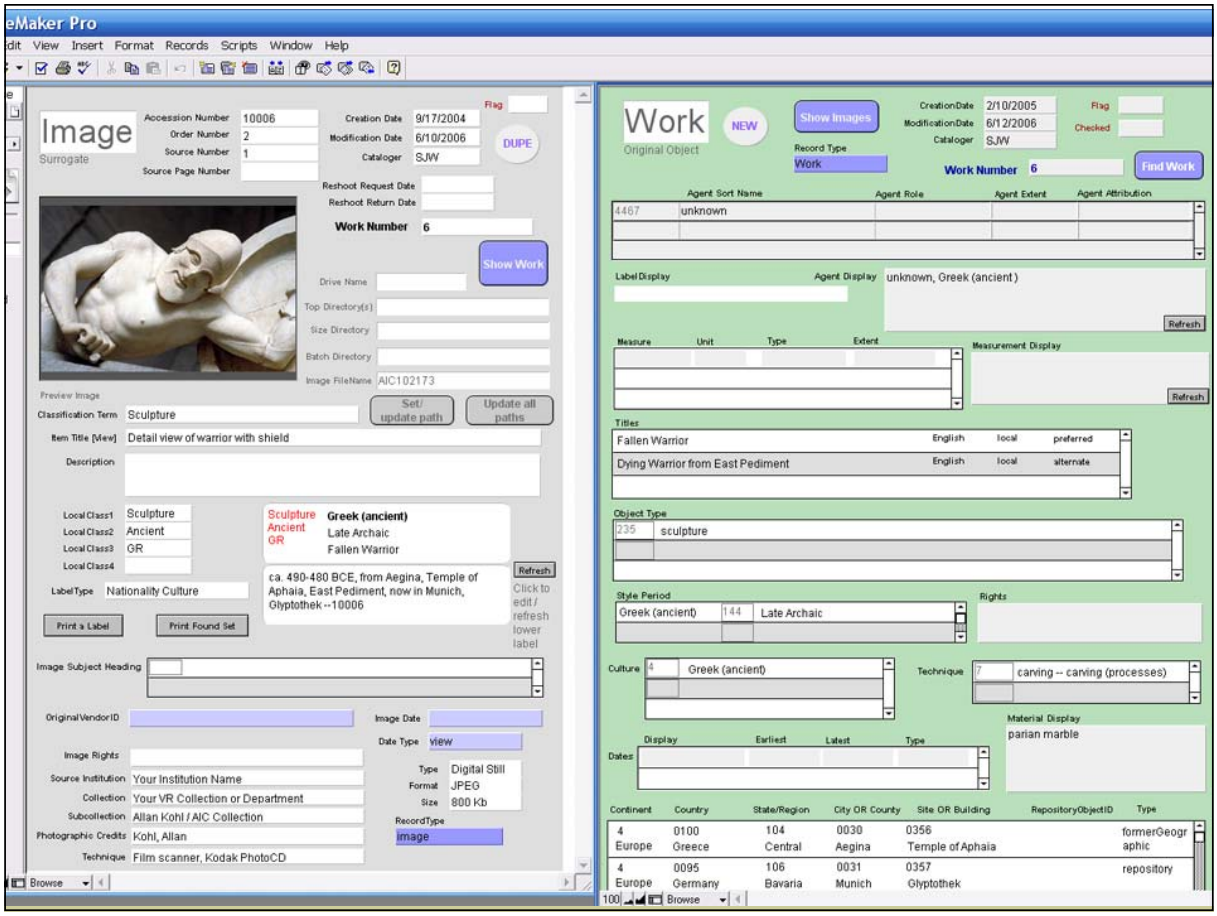
- You can export XML data from FileMaker or Access (and many other programs) to an assortment of applications simply by applying the appropriate Extensible Stylesheet Language Transformation (XSLT) style sheet.

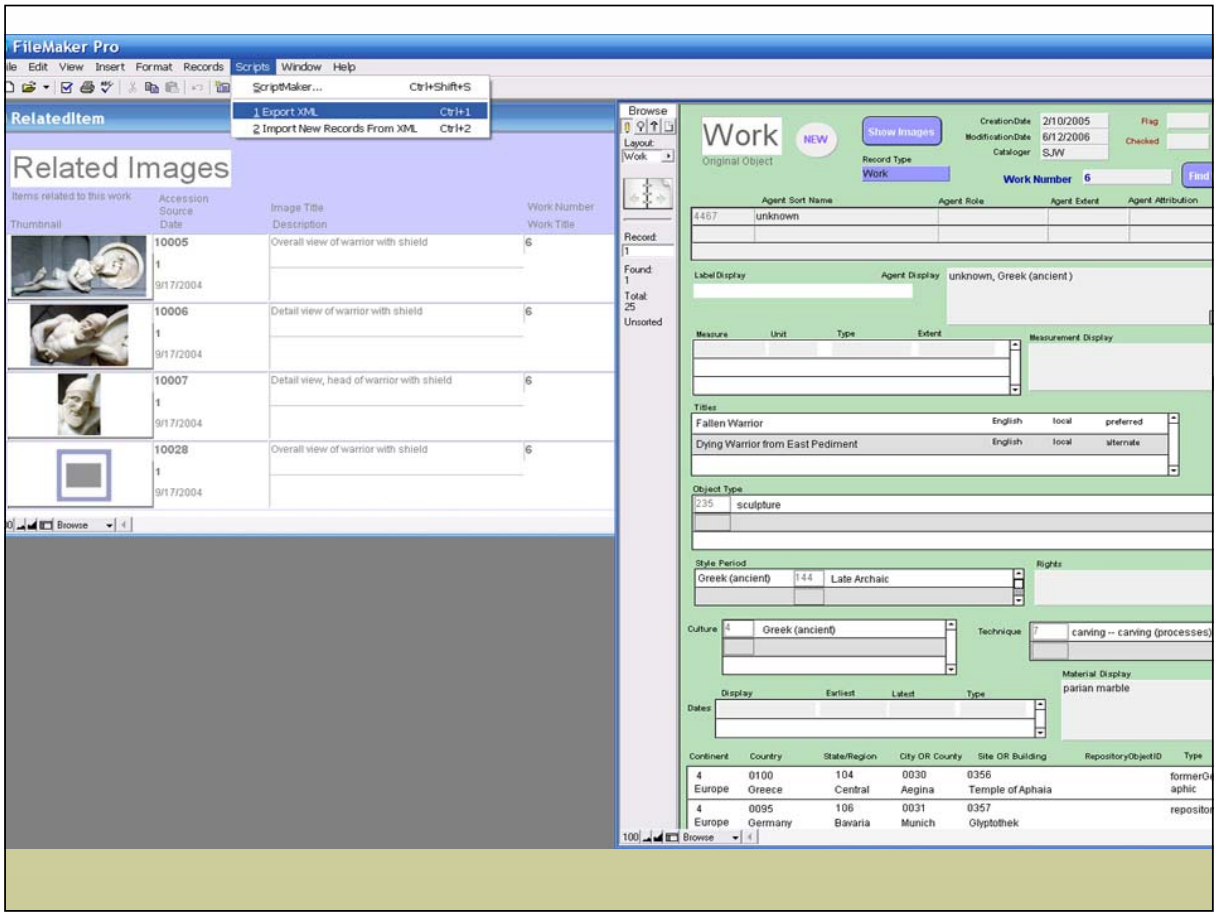
XLST Sample—how the XML is actually exported from a database

```
<!-- Agent   -->
<set>
<display>
<xsl:value-of select="fm:AgentDisplay" />
</display>
<index>
<xsl:for-each select="fm:AgentSortName/fm:DATA">
<xsl:variable name="i">
<xsl:value-of select="position()" /> </xsl:variable>
<agent>
```

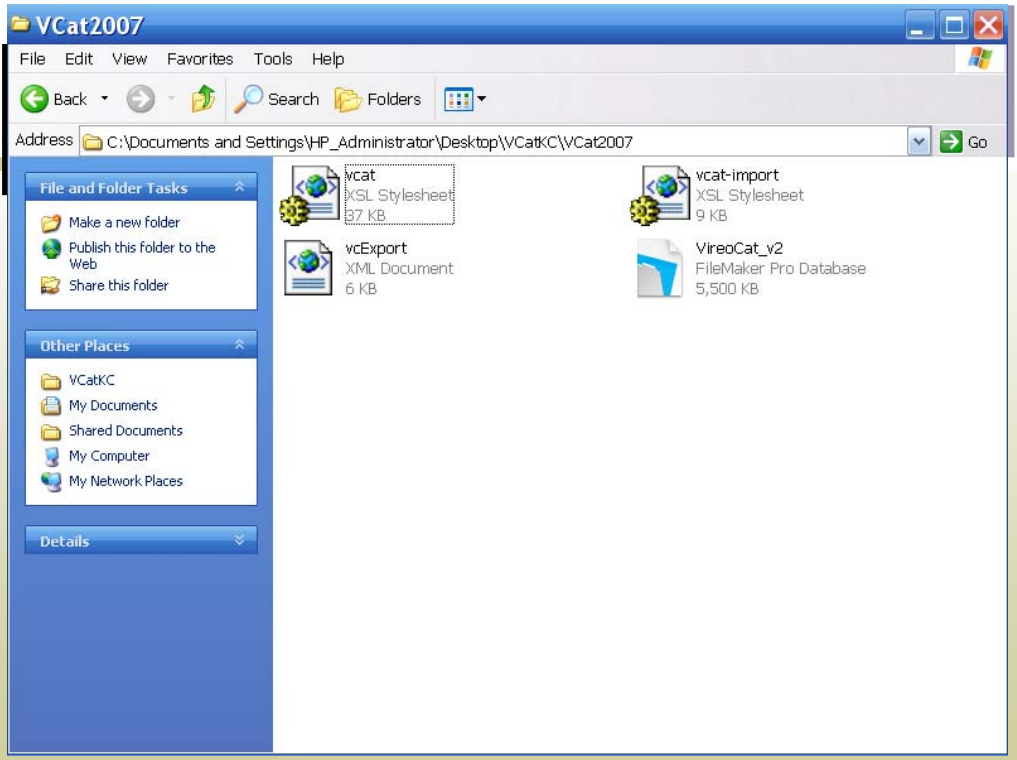
4 Screenshots of an XML export (VireoCat)

- First slide—work and image screens, perform find
- Second slide showing the work record and 4 related image records and export selected from menu bar
- Third slide, resultant XML document (excerpt)
- Fourth slide, showing the database folder containing database, 2 stylesheets and XML document


eMaker Pro
dit View Insert Format Records Scripts Window Help
Image
Surrogate
Accession Number 10006
Order Number 2
Source Number 1
Source Page Number
Creation Date 9/17/2004
Modification Date 6/10/2006
Cataloger SJW
Flag
DUPE
Reshoot Request Date
Reshoot Return Date
Work Number 6
Show Work
Drive Name
Top Directory(s)
Size Directory
Batch Directory
Image FileName AIC102173
Preview Image
Set/ update path
Update all paths
Classification Term Sculpture
Item Title [New] Detail view of warrior with shield
Description
Local Class1 Sculpture
Local Class2 Ancient
Local Class3 GR
Local Class4
Sculpture Greek (ancient)
Ancient Late Archaic
GR Fallen Warrior
ca. 490-480 BCE, from Aegina, Temple of Aphaia, East Pediment, now in Munich, Glyptothek --10006
Refresh
Click to edit / refresh lower label
LabelType Nationality Culture
Print a Label
Print Found Set
Image Subject Heading
Original Vendor ID
Image Date
Date Type view
Image Rights
Source Institution Your Institution Name
Collection Your VR Collection or Department
Subcollection Allan Kohl / AIC Collection
Photographic Credits Kohl, Allan
Technique Film scanner, Kodak PhotoCD
Type Digital Still
Format JPEG
Size 800 Kb
RecordType image
Browse
Work
Original Object
NEW
Show Images
Record Type Work
CreationDate 2/10/2005
ModificationDate 6/12/2006
Cataloger SJW
Flag
Checked
Work Number 6
Find Work
Agent Sort Name
Agent Role
Agent Extent
Agent Attribution
4467 unknown
Label Display
Agent Display unknown, Greek (ancient)
Refresh
Measure
Unit
Type
Extent
Measurement Display
Refresh
Titles
Fallen Warrior English local preferred
Dying Warrior from East Pediment English local alternate
Object Type
235 sculpture
Style Period
Greek (ancient) 144 Late Archaic
Rights
Culture 4 Greek (ancient)
Technique 7 carving -- carving (processes)
Material Display
parian marble
Dates
Display
Earliest
Latest
Type
Continent Country State/Region City OR County Site OR Building RepositoryObjectID Type
4 Europe 0100 Greece 104 Central 0030 Aegina 0356 Temple of Aphaia former Geogr aphic
4 Europe 0095 Germany 106 Bavaria 0031 Munich 0357 Glyptothek repository
100
Browse


FileMaker Pro
File Edit View Insert Format Records Scripts Window Help
ScriptMaker... Ctrl+Shift+S
1 Export XML Ctrl+1
2 Import New Records From XML Ctrl+2
RelatedItem
Related Images
Items related to this work
Thumbnail
Accession Source Date
Image Title Description
Work Number Work Title
10005 1 9/17/2004 Overall view of warrior with shield 6
10006 1 9/17/2004 Detail view of warrior with shield 6
10007 1 9/17/2004 Detail view, head of warrior with shield 6
10028 1 9/17/2004 Overall view of warrior with shield 6
Browse
Layout: Work
Record: 1
Found: 1
Total: 25
Unsorted
Work
Original Object
NEW
Show Images
Record Type Work
CreationDate 2/10/2005
ModificationDate 6/12/2006
Cataloger SJW
Flag
Checked
Work Number 6
Agent Sort Name
Agent Role
Agent Extent
Agent Attribution
4467 unknown
Label Display
Agent Display unknown, Greek (ancient)
Measure
Unit
Type
Extent
Measurement Display
Titles
Fallen Warrior English local preferred
Dying Warrior from East Pediment English local alternate
Object Type
235 sculpture
Style Period
Greek (ancient) 144 Late Archaic
Rights
Culture 4 Greek (ancient)
Technique 7 carving -- carving (processes)
Material Display
parian marble
Dates
Display
Earliest
Latest
Type
Continent Country State/Region City OR County Site OR Building RepositoryObjectID Type
4 Europe 0100 Greece 104 Central 0030 Aegina 0356 Temple of Aphaia
4 Europe 0095 Germany 106 Bavaria 0031 Munich 0357 Glyptothek
100
Browse

```
<?xml version="1.0" encoding="UTF-8"?>
- <vra xmlns="http://www.vraweb.org/vracore4.htm">
 - <work xmlns="" id="W6" source="VireoCat Test Database, Susan Jane Williams (consultant)" refid="6">
  - <agentSet>
     <display>unknown (sculptor, Greek, ancient )</display>
   - <agent>
      <name type="Personal">unknown</name>
     - <dates type="">
        <earliestDate/>
        <latestDate/>
      </dates>
      <culture/>
      <role vocab="AAT" refid="300025181">sculptor</role>
      <attribution/>
     </agent>
  </agentSet>
  - <dateSet>
     <display>ca. 490-480 BCE</display>
   - <date type="creation">
      <earliestDate>-0490</earliestDate>
      <latestDate>-0480</latestDate>
     </date>
  </dateSet>
  - <descriptionSet>
     <description/>
  </descriptionSet>
  - <locationSet>
     <display>Repository: Glyptothek, Munich, Bavaria, Germany (85); Former: Temple of Aphaia,
      Aegina, Central Greece and Euboea, Greece ()</display>
   - <location type="formerGeographic">
      <name type="geographic" vocab="DOA" refid="" extent="building">Temple of Aphaia</name>
      <name type="geographic" vocab="TGN" refid="7011087" extent="inhabited
       place">Aegina</name>
      <name type="geographic" vocab="TGN" refid="7001395" extent="region">Central Greece and
       Euboea</name>
      <name type="geographic" vocab="TGN" refid="1000074" extent="nation">Greece</name>
      <name type="geographic" vocab="TGN" refid="1000003" extent="continent">Europe</name>
```



My better late than never epiphany….

- Appreciating the symmetry of the elements in the work, image and collection records (work title, image title and so forth)
- Appreciating the distinction between indexed and display values, and how that can give you flexibility in data entry

Creating Display values

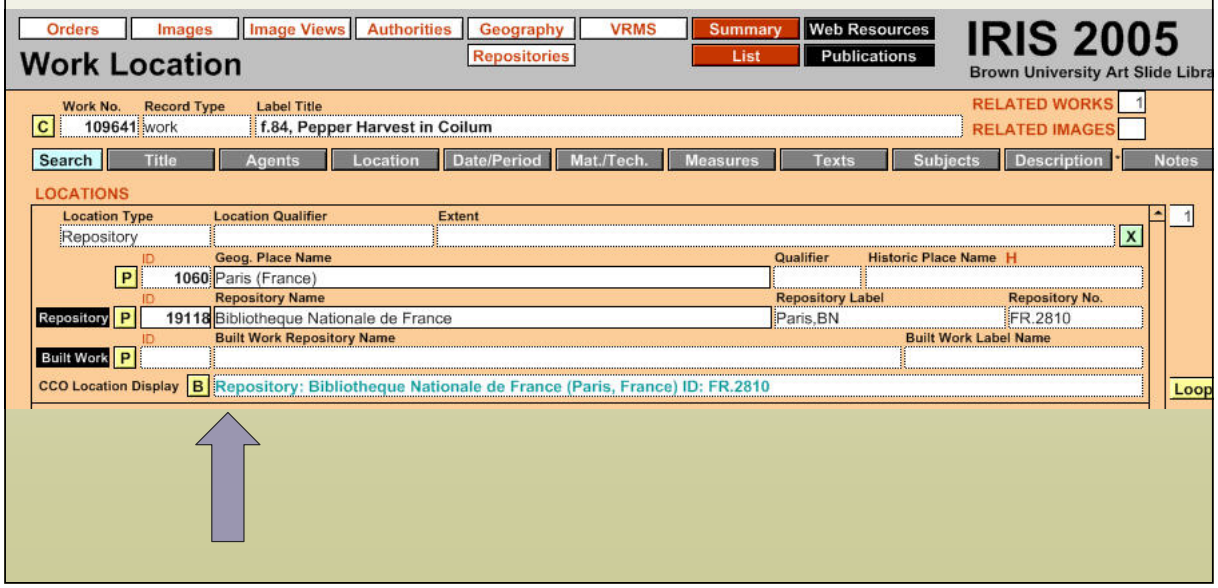
- Can be created from the indexed values “on the fly” in the xml via the stylesheet, or
- Can be “pre-created” within the database by scripting/programing
- Both approaches are likely to be used in different fields–gives flexibility in design choices

Location in VireoCat:
multiple table design requires “extent” in XML;
Display created on the fly in the XML

```
cionSet>
ionSet>
play>Repository: Glyptothek, Munich, Bavaria, Germany (85); Former: Temple of Aphaia, Aegina, Central Gree
cation type="formerGeographic">
name type="geographic" vocab="DOA" refid="" extent="building">Temple of Aphaia</name>
name type="geographic" vocab="TGN" refid="7011087" extent="inhabited place">Aegina</name>
name type="geographic" vocab="TGN" refid="7001395" extent="region">Central Greece and Euboea</name>
name type="geographic" vocab="TGN" refid="1000074" extent="nation">Greece</name>
name type="geographic" vocab="TGN" refid="1000003" extent="continent">Europe</name>
cation>
cation type="repository">
name type="geographic" vocab="DOA" refid="" extent="building">Glyptothek</name>
name type="geographic" vocab="TGN" refid="7004333" extent="inhabited place">Munich</name>
name type="geographic" vocab="TGN" refid="7003669" extent="region">Bavaria</name>
name type="geographic" vocab="TGN" refid="7000084" extent="nation">Germany</name>
name type="geographic" vocab="TGN" refid="1000003" extent="continent">Europe</name>
cation>
```

Locations in IRIS

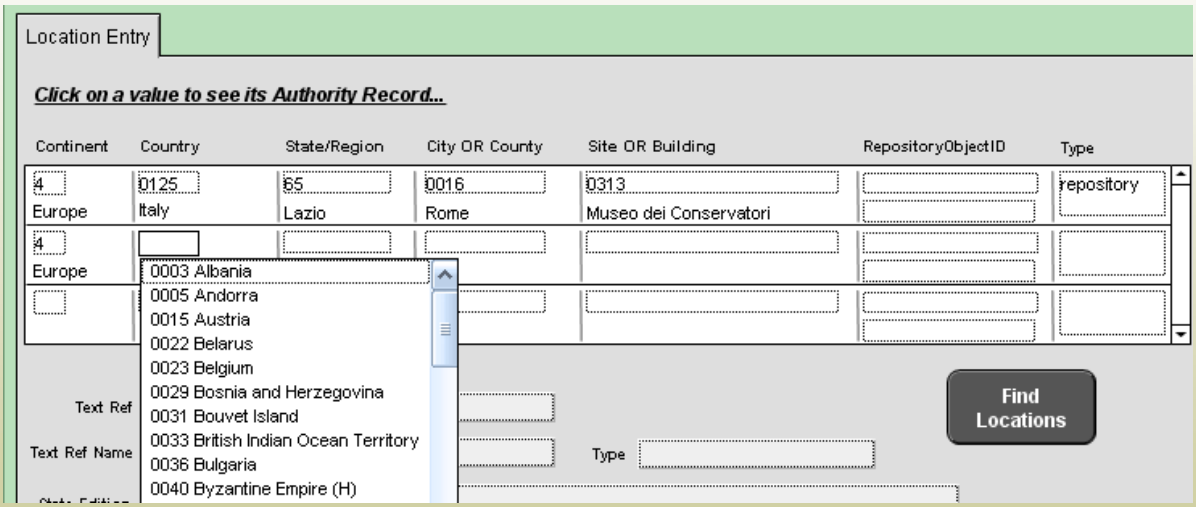
Showing CCO display form “pre-created” in the database from parsed values

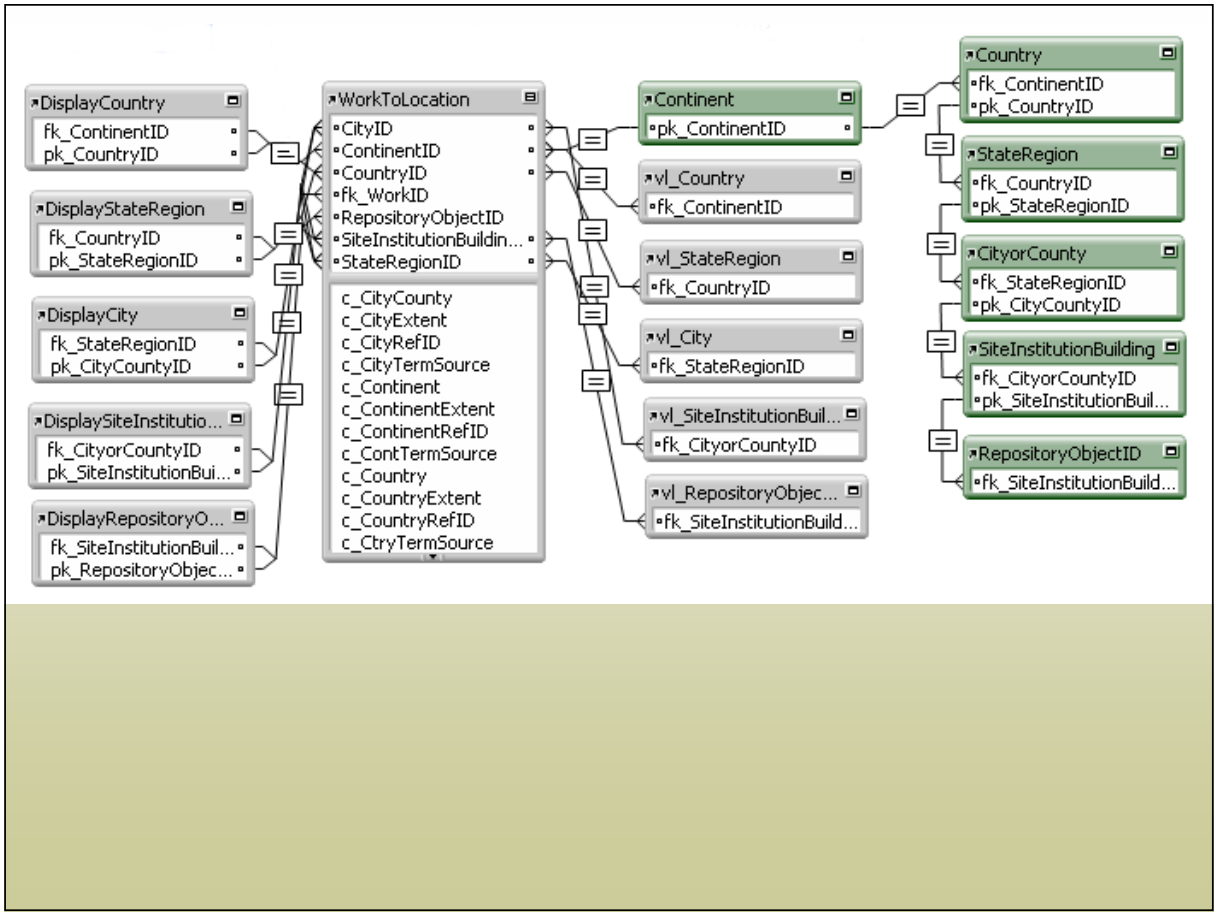


Using scripting and design to guide cataloging

- First slide shows set-up in VireoCat—as cataloger chooses continent, then next level (nation) is constrained to only the choices in that continent, and so forth down to building/site level
- Second slide shows multi-table design necessary to do this, with 2 sets of additional linking tables to constrain and display parent/child lists

Constraining/ guiding entry





Hobbesian choices?

- The multi-table design guides entry and may be a good solution, especially for student workers, but….
- It complicated the design of XML export and import
- Good to look at all these issues holistically, which means acquainting yourself with XML and scripting and design, at least conceptually

And what is my next stylesheet?

- Taking relational XML export and running it through a stylesheet so that it becomes flattened CSV to use in DAPs such as CONTENTdm that are not yet XML ready!

;-)