

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

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?	Ke
Artists	display_name	Artist	Creation-Creator-Identity (core)	Yes	Yes [5 repeats]	Yes	Ye
Artists	sort_name	DO NOT DISPLAY	DO NOT MAP	No	No	Yes	Ye
Artists	id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists	display_date	Artist Date	Creation-Creator-Identity-Dates	Yes	No	Yes	Ye
Artists	begin_search_date	DO NOT DISPLAY	DO NOT MAP	No	No	Yes	Ye
Artists	end_search_date	DO NOT DISPLAY	DO NOT MAP	No	No	Yes	Ye
Artists_Works	id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artist_Roles	name	Artist Role	Creation-Creator-Role (core)	Yes	No	Yes	Ye
Artist_Roles	id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artist_Roles	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artist_Roles	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists_Works	role_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists_Works	attribution	Attribution	Creator-Identification-Name (core)	Yes	No	Yes	Ye
Artists_Works	artist_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists_Works	work_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists_Works	priority	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists_Works	position	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists	nationality_id	Artist Nationality/Culture	Creator Identification-Nationality/Culture/ Race (core)	Yes	No	Yes	Ye
Nationalities	id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Nationalities	name						
Nationalities	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Nationalities	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists	gender	DO NOT DISPLAY	DO NOT MAP	No	No	Yes	Ye
Artists	corporate_entity	DO NOT DISPLAY	DO NOT MAP	No	No	Yes	Nd
Artists	authority_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd
Artists	authority_ref_id	DO NOT DISPLAY	DO NOT MAP	No	No	No	Nd

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: vitalDatesCreator> birthdate; <cdwalite: vitalDatesCreator> deathdate
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: nationalityCreator>
Artists	gender				Person/Corporate Body Authority - Gender	<cdwalite: genderCreator>
Artists	corporate_entity	AGENT name (type) (work record)				
Artists	authority_id	AGENT name. vocab				
Artists	authority_ref_id	AGENT name. refid (work record)				
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: attributionQualifierCreator>
Artists_Works	artist_id					
Artists_Works	work_id					
Artists_Works	priority					

[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	G
Title.Variant.Work	LOCATION	Classification	Artist_Sortname1	Artist_Sortname2
Rotonde de Chartres	Paris, France	Architecture	Ledoux, Claude-Nicolas	
Ducal Palace	Dijon, France	Architecture	Mansart, Jules Hardouin	Gabriel, Ange-Jacques
Ducal Palace	Dijon, France	Architecture	Mansart, Jules Hardouin	Gabriel, Ange-Jacques

[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
2225	4	architect

[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

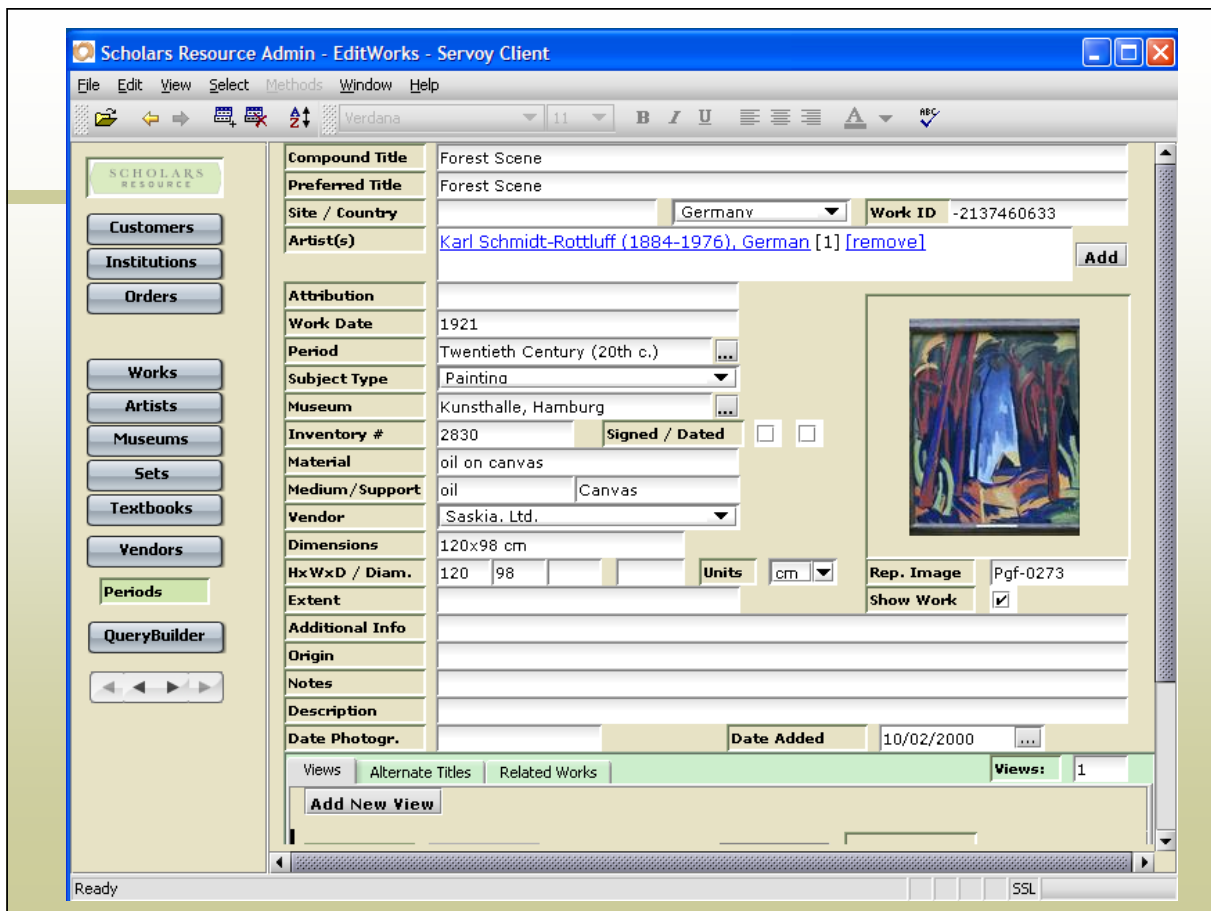
[Work and Image entry layouts]

- A clipped sample of 2 forms:

The image displays two screenshots of a database interface. The left screenshot shows an 'Image' entry form with fields for Accession Number (10006), Order Number (2), Source Number (1), Source Page Number, Creation Date (9/17/2004), Modification Date (6/10/2006), Cataloger (S.J.W.), Reshoot Request Date, Reshoot Return Date, Work Number (6), Drive Name, Top Directory(s), Size Directory, Batch Directory, Image FileName (AIC102173), Classification Term (Sculpture), Item Title (Detail view of warrior with shield), and Description. The right screenshot shows a 'Work' entry form with fields for Creation Date (2/10/2005), Modification Date (6/12/2006), Cataloger (S.J.W.), Record Type (Work), Work Number (6), Agent Sort Name (4467), Agent Role (unknown), Agent Extent, Agent Attribution, Label Display (Agent Display: unknown, Greek (ancient)), Measure, Unit, Type, Extent, Measurement Display, and Titles (Fallen Warrior, Dying Warrior from East Pediment).

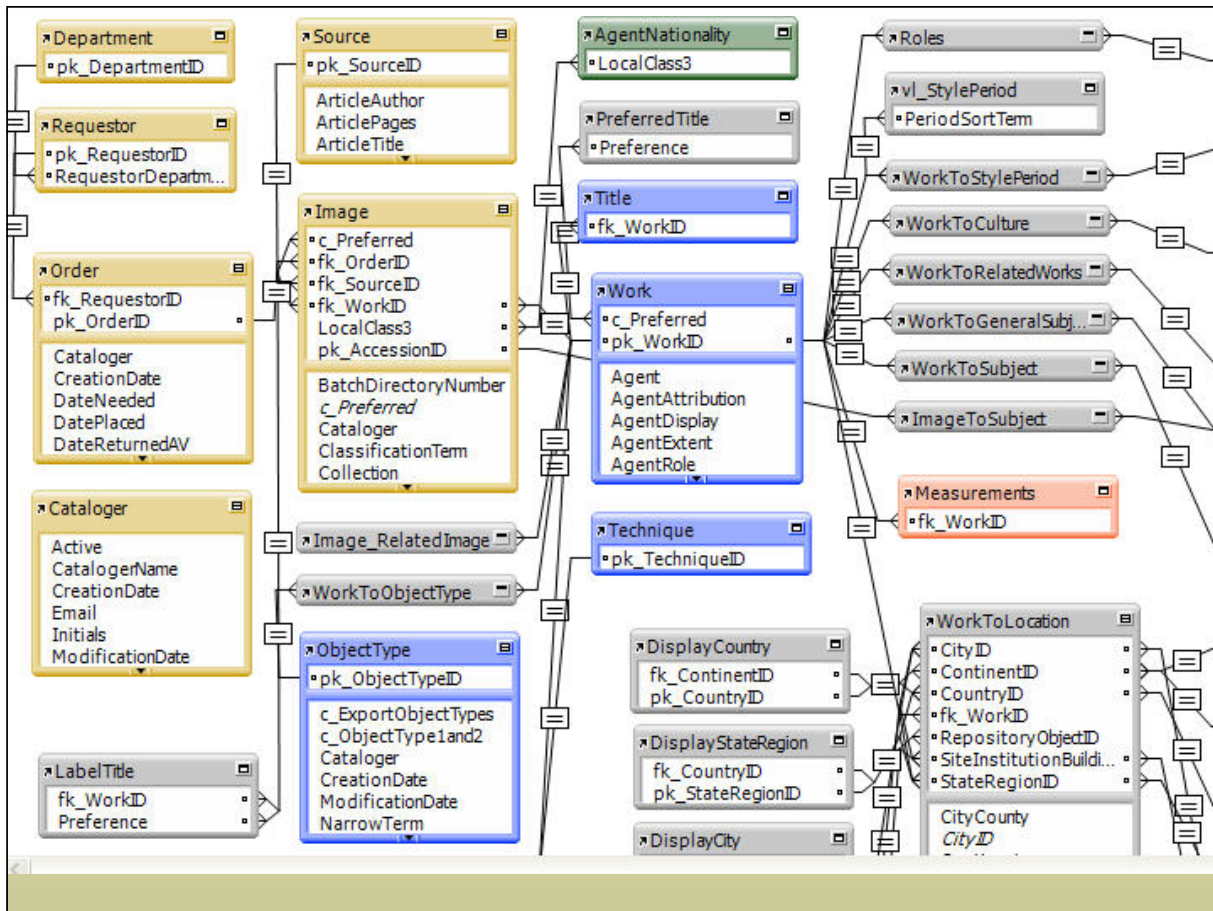
[Splitting the interface design from the database (table) design]

- 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



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

- AgentExtent
- AgentRoles
- AgentSortName
- Attribution
- ExportAgents
- fk_AgentID**
- fk_FirmID
- fk_WorkID

Table: Agent

- EarliestDate
- FirmType
- Gender
- LatestDate
- LifeDisplayDate
- ModificationDate
- NameType
- Note
- pk_AgentID**
- RefID

Roles = Agent

fk_AgentID = pk_AgentID

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

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

Buttons: Add, Change, Duplicate, Delete, 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 Eqypt	0022 Pyramids of Giza	0354 Pyramid complex of Mycerinus		formerGeographic
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)]

The screenshot shows a web form titled "City or County" with a light blue background. At the top left is a large text input field containing "City or County". To its right is a "TGN" button. Below this, there are two more input fields: "Term Source" with "TGN" and "Ref ID" which is empty. A "DOA" button is located to the right of the "Ref ID" field. The main section contains three input fields: "City or County (EnglishName)" with "Boston", "State Region ID" with "21", and "City ID" with "0010". Below these is a "NOTE: Only add sites to county if there is NO associated city" and a label "Associated Site, Institution, Building—entry portal". A table with a scroll bar on the right lists several sites:

Athenæum
Bostonian Society
Isabella Stewart Gardner Museum
Massachusetts Historical Society Museum
Museum of Afro American History
Boston Institute of Contemporary Art
Museum of Fine Arts
Museum of Science

[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 |u 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?]

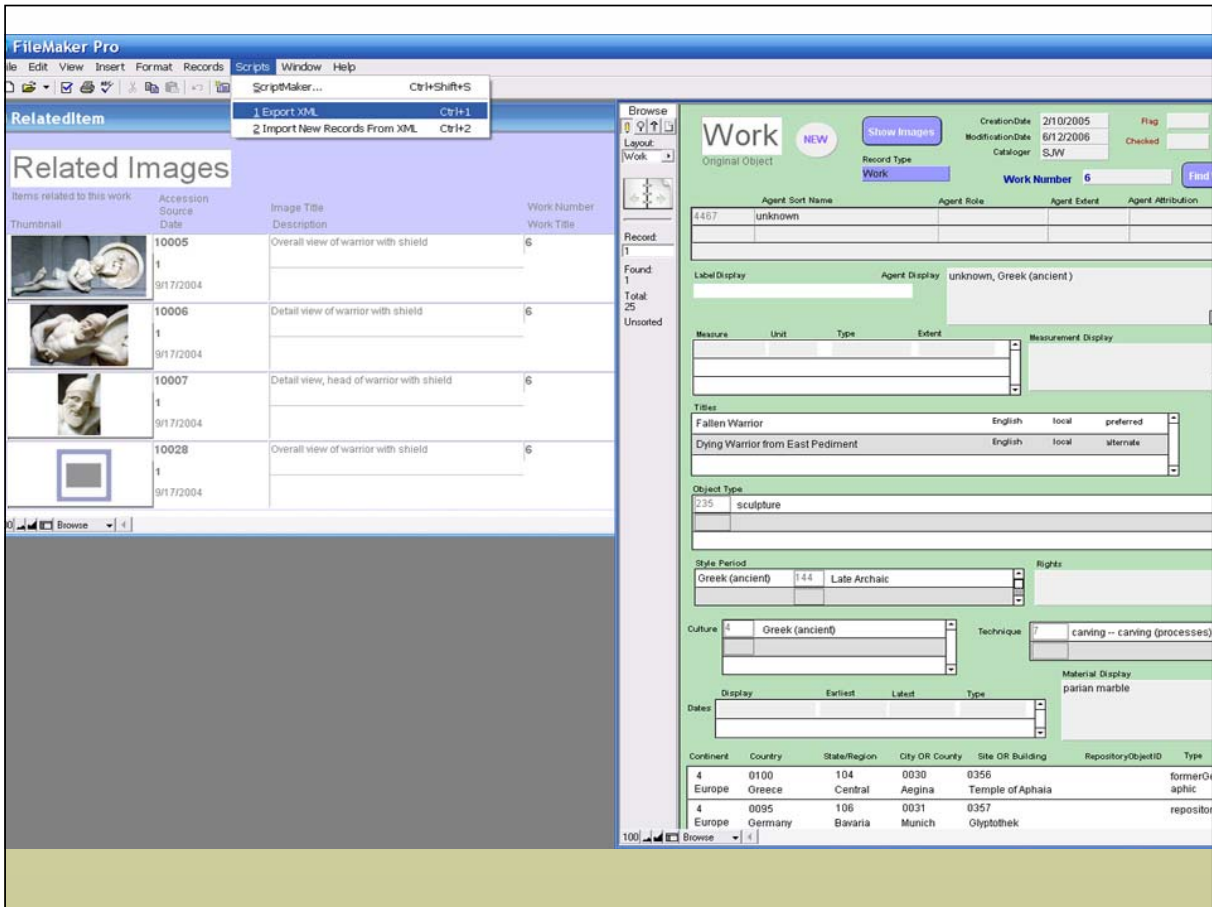
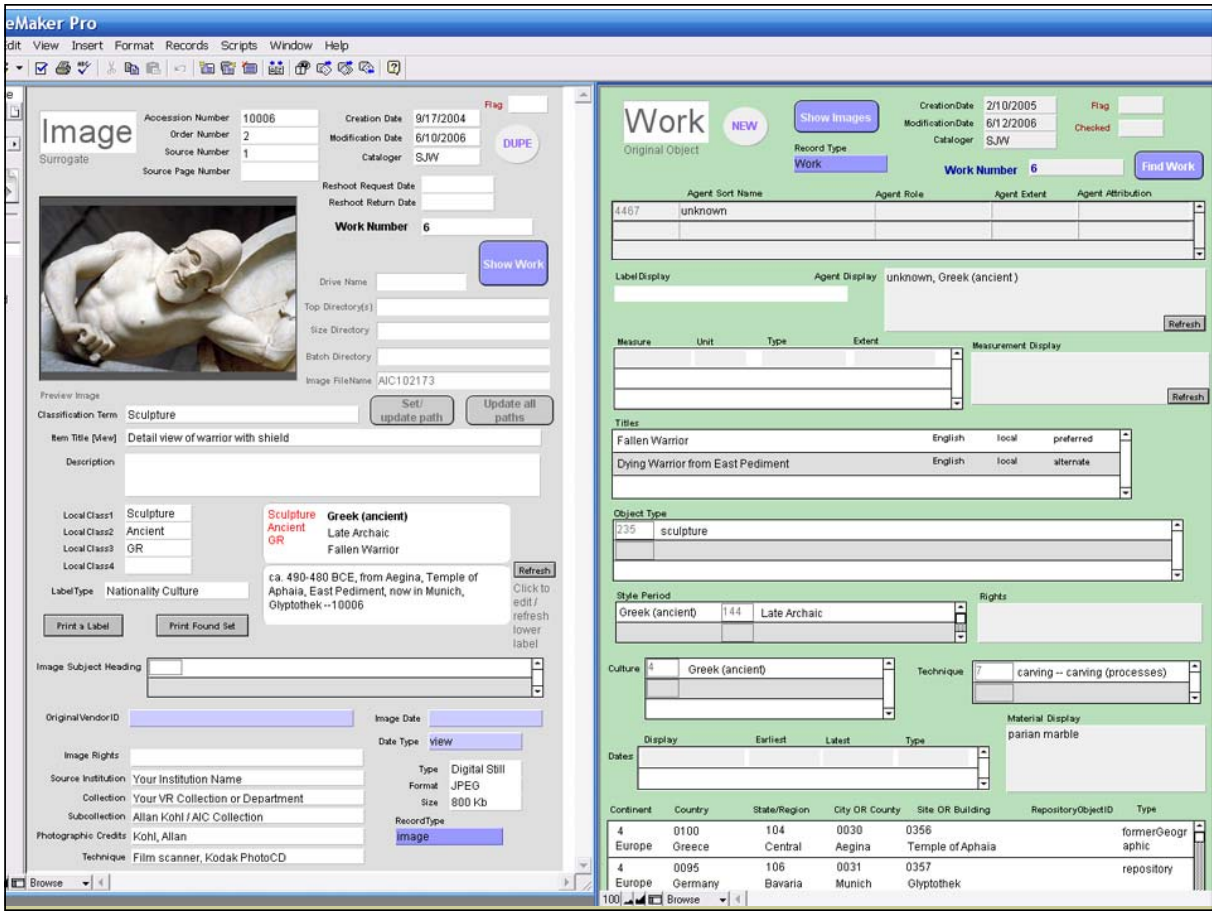
- You can export XML data from FileMaker or Access (and many other programs) to an assortment of applications simply by applying the appropriate Extensible Stylesheet Language Transformation (XSLT) style sheet.

[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)]

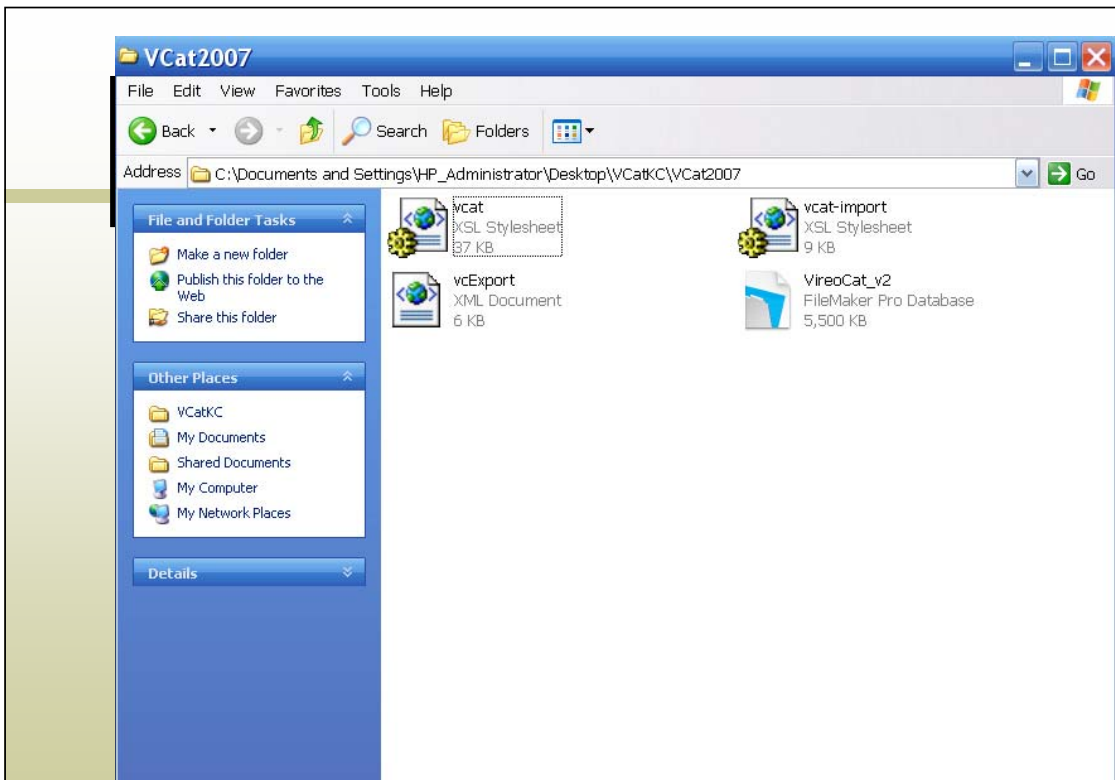
- First slide—work and image screens, perform find
- Second slide showing the work record and 4 related image records and export selected from menu bar
- Third slide, resultant XML document (excerpt)
- Fourth slide, showing the database folder containing database, 2 stylesheets and XML document



```

<?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/programming
- 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

```

<captionSet>
</captionSet>
<play>Repository: Glyptothek, Munich, Bavaria, Germany (85); Former: Temple of Aphaia, Aegina, Central Greece
</play>
<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>
</location>
<location 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>
</location>

```

Locations in IRIS

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

Orders Images Image Views Authorities Geography VRMS Summary Web Resources IRIS 2005
 Work Location Repositories List Publications Brown University Art Slide Library

Work No. 109641 Record Type work Label Title f.84, Pepper Harvest in Coilum

RELATED WORKS 1 RELATED IMAGES

Search Title Agents Location Date/Period Mat./Tech. Measures Texts Subjects Description Notes

LOCATIONS

Location Type	Location Qualifier	Extent
Repository		
P 1060	Geog. Place Name Paris (France)	Qualifier Historic Place Name H
Repository P 19118	Repository Name Bibliotheque Nationale de France	Repository Label Paris.BN Repository No. FR.2810
Built Work P	Built Work Repository Name	Built Work Label Name

CCO Location Display B Repository: Bibliotheque Nationale de France (Paris, France) ID: FR.2810

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

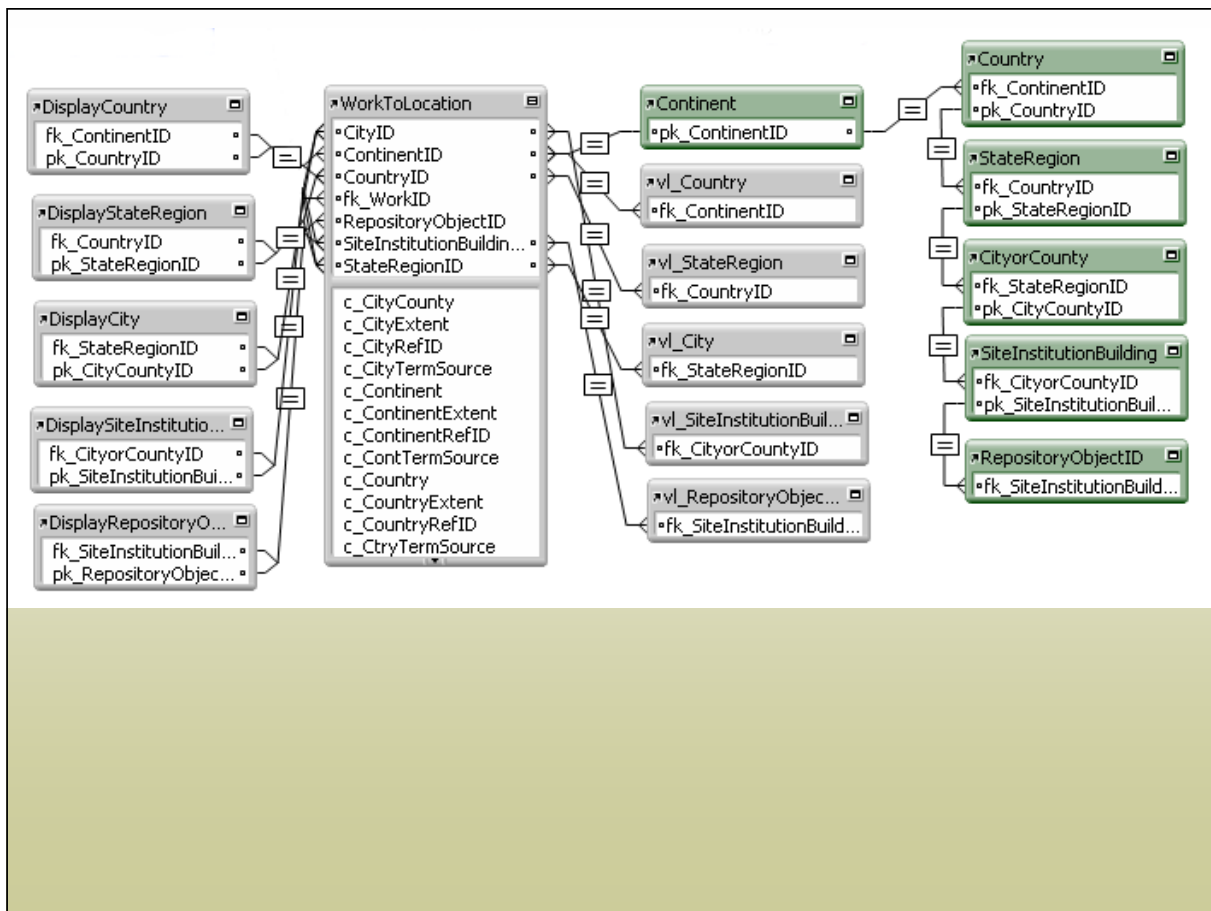
Constraining/ guiding entry

Location Entry

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

Continent	Country	State/Region	City OR County	Site OR Building	RepositoryObjectID	Type
4	0125	65	0016	0313		repository
Europe	Italy	Lazio	Rome	Museo dei Conservatori		
4						
Europe	0003 Albania					
	0005 Andorra					
	0015 Austria					
	0022 Belarus					
	0023 Belgium					
	0029 Bosnia and Herzegovina					
	0031 Bouvet Island					
Text Ref	0033 British Indian Ocean Territory					
Text Ref Name	0036 Bulgaria					
	0040 Byzantine Empire (H)					

Find Locations



[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!

;-)