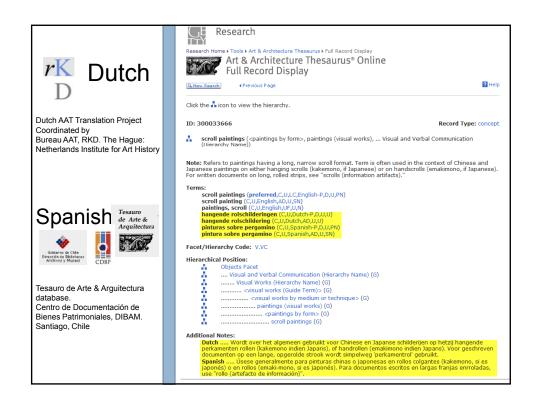
# THE GETTY VOCABULARIES TECHNICAL UPDATE

Joan Cobb Gregg Garcia Information Technology Services J. Paul Getty Trust

November 8, 2011

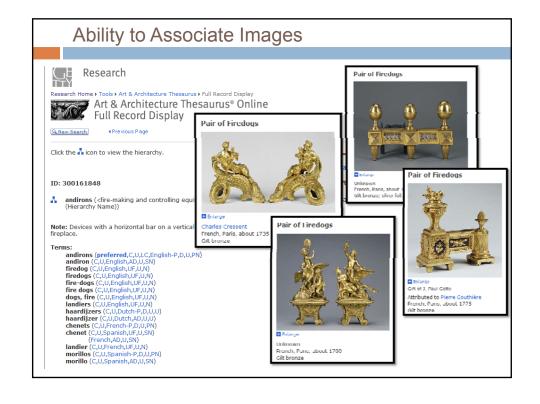
#### International Terminology Working Group August 23 – 26, 2010

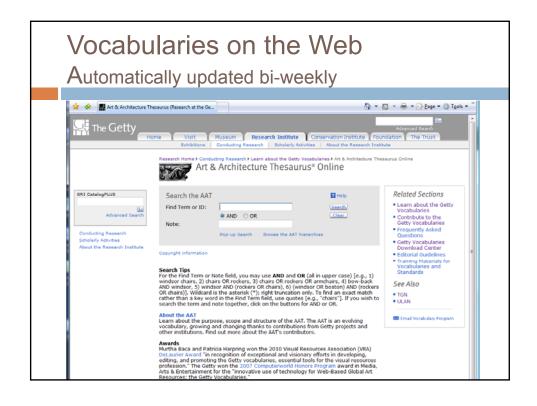


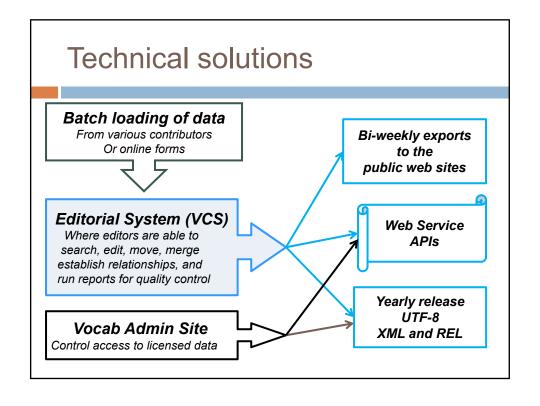


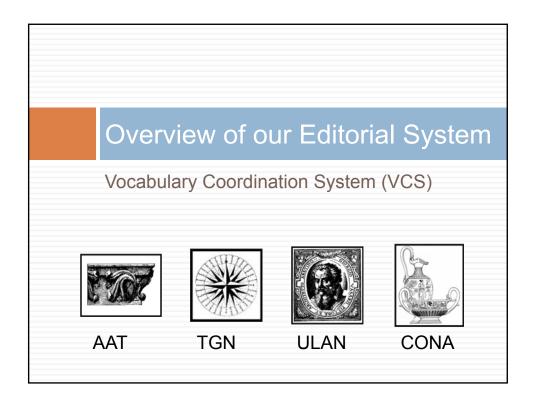


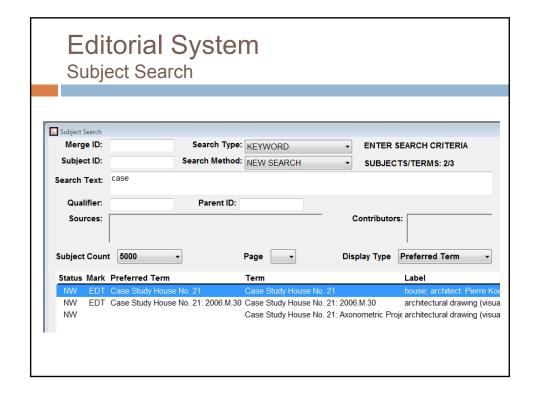


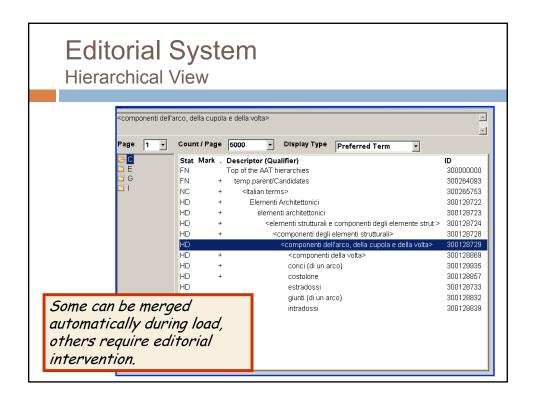


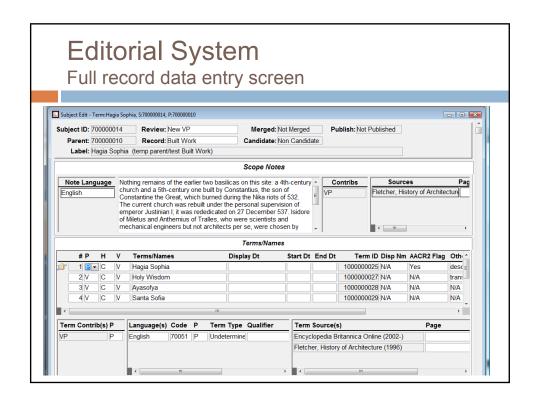


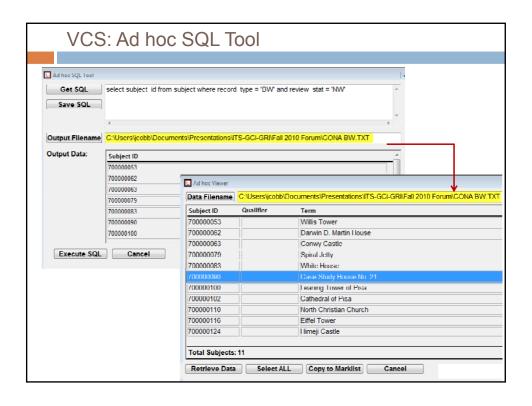


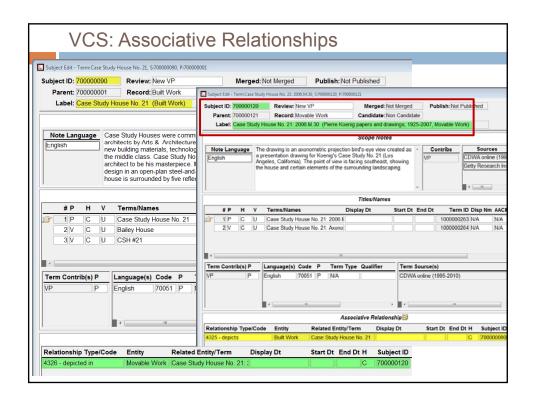


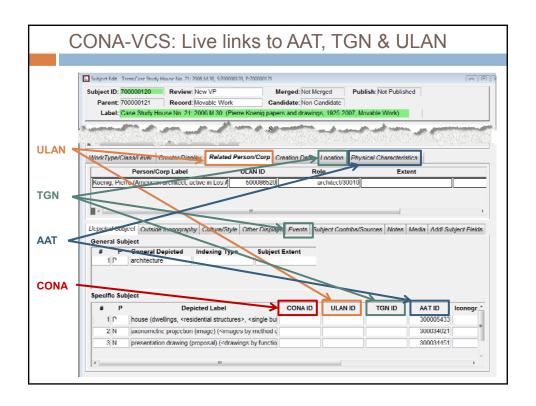


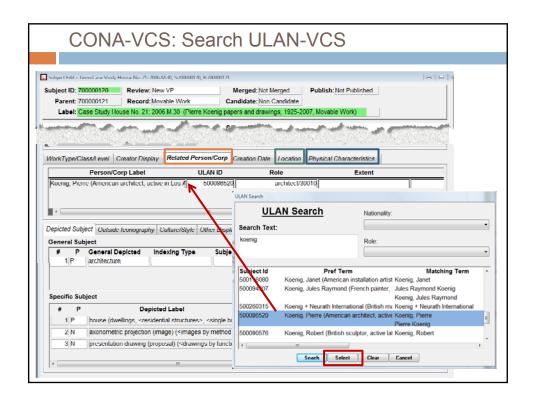


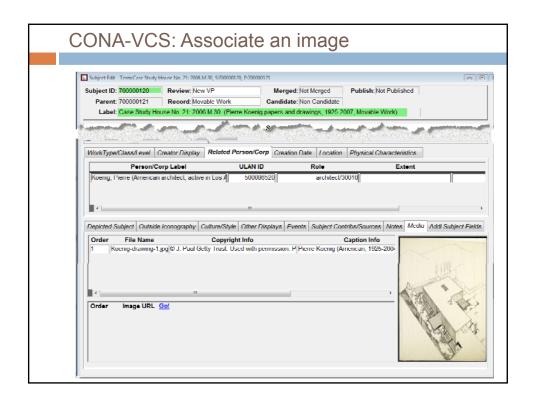


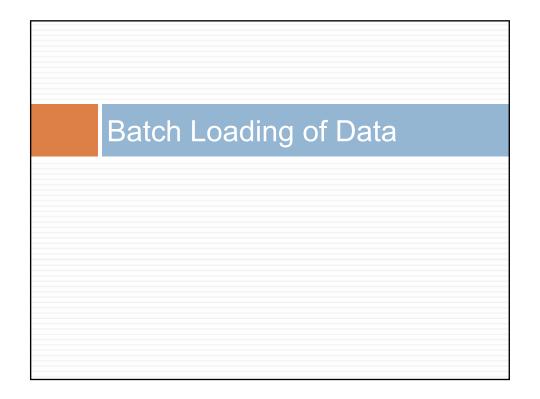












### Batch loading of data

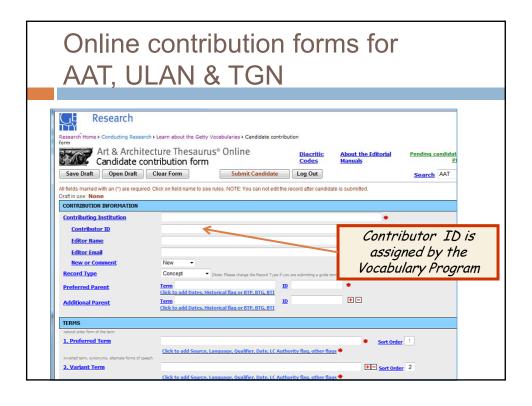
 Information about how to contribute is available on our Web site

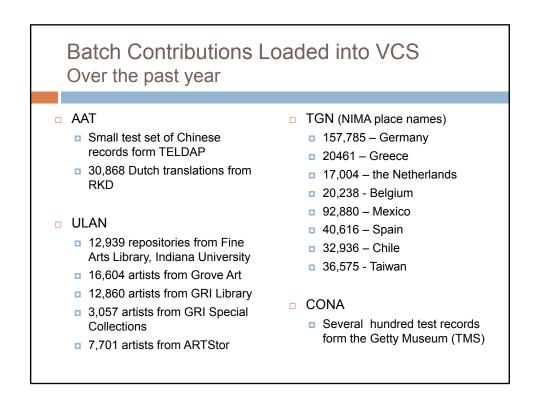
http://www.getty.edu/research/tools/vocabularies/contribute.html

- Bulk contributions
  - Can range from a few thousand to several million records
  - Every new contribution requires close coordination
- Online Web candidate contribution forms
  - URL: vocabcontrib.getty.edu
  - Uses the same XML schema as bulk contributions
  - Harvested on the 1st of every month

#### Contribution XML schemas

- Basic Schemas
  - Getty\_Vocabulary\_AAT.xsd
  - Getty\_Vocabulary\_TGN.xsd
  - Getty\_Vocabulary\_ULAN.xsd
  - Getty\_Vocabulary\_CONA.xsd
- Schemas containing specific enumeration values are updated every month
  - AAT : associative\_type, contrib, language, lookup\_lists, source
  - TGN: associative\_type, contrib, language, lookup\_lists, source plus place\_type
  - ULAN: associative\_type, contrib, language, lookup\_lists, source plus event, nationality, place, and role
  - CONA: associative\_type, contrib, language, lookup\_lists, source plus event, nationality, role, and class





## Loading the data

- Programs to load the data are very complex
  - Several rounds of testing are necessary
  - Actual load takes from several hours to several days

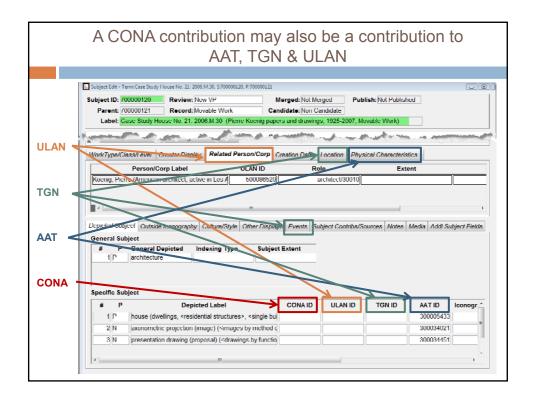


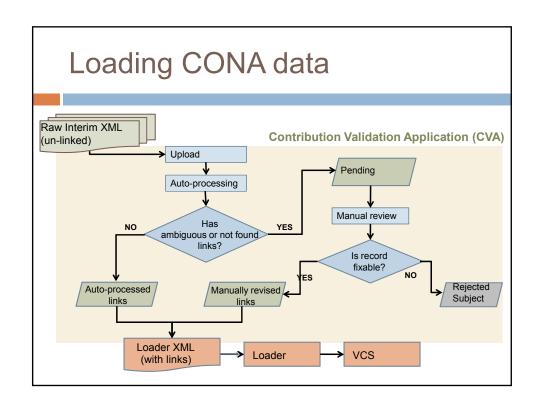
- Merging
  - Rules are different for each vocabulary
  - Need to accommodate differing standards
  - Some of the merge routines can be skipped with translations because the matching records are already known
- Unicode
  - Contributions can be submitted in any encoding

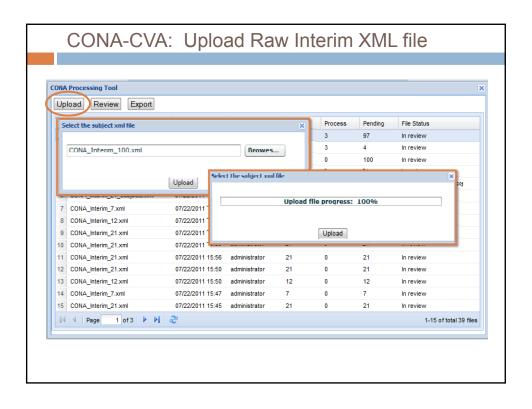
#### Loading the data

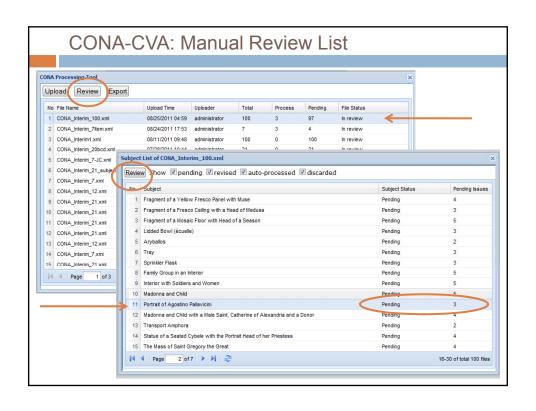
Multi-lingual data

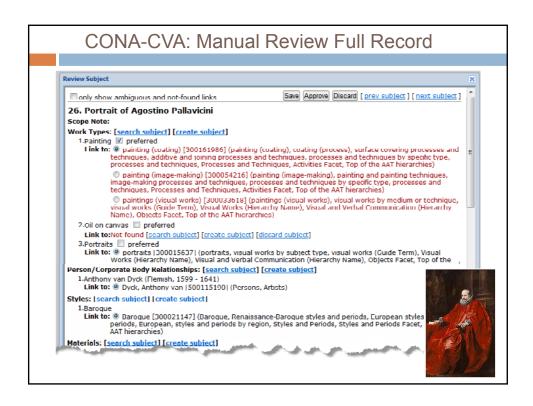
- Worked through a number of issues related to ordering and classifying terms in different languages
- Adjusted the data structure to make the term qualifier, term type and part of speech language specific
- Modified the rules combining terms
- Updated the way the information is presented to users
- Changes to the core editorial system require changes to everything else

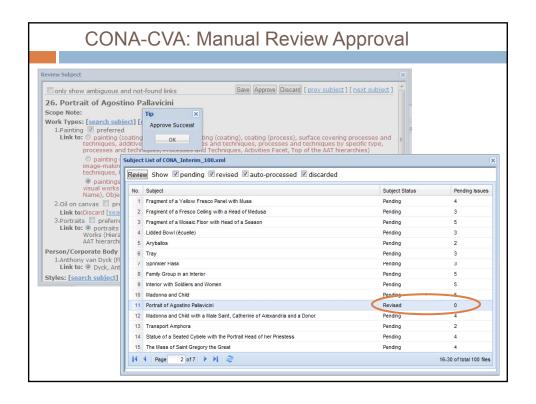


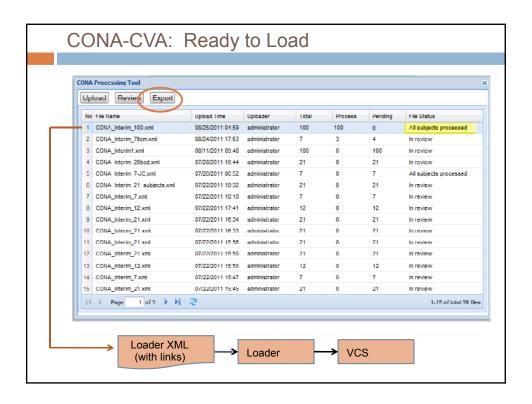








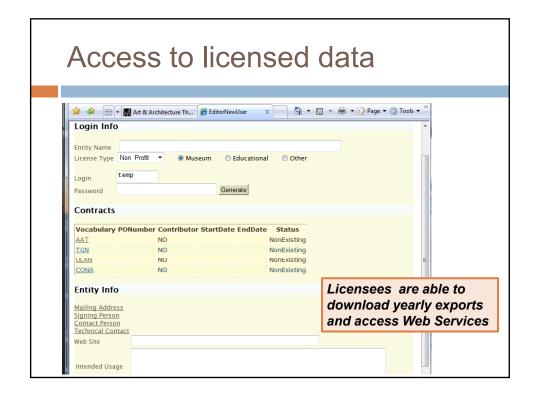


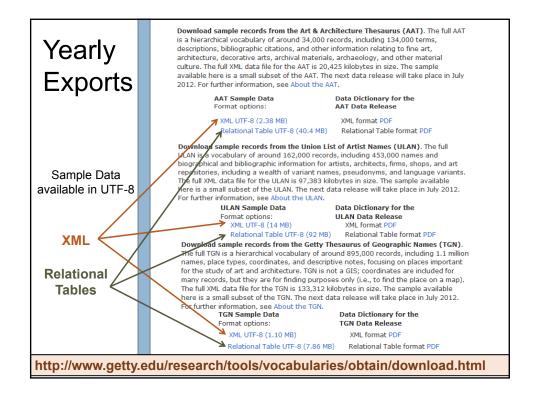


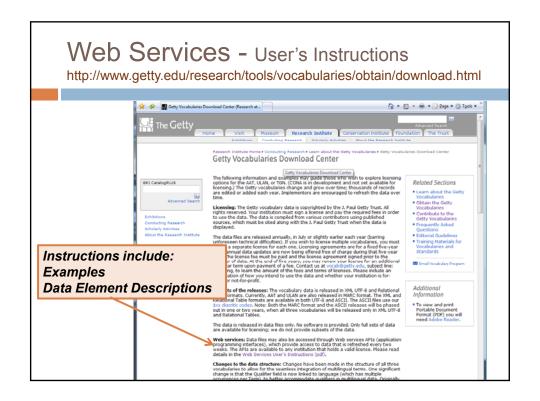
#### **CONA Development Schedule**

- Part 1 Editorial system (VCS-CONA) went live in October 2010
- Part 2 We are ready to accept contributions NOW
- Part 3 Public Web Site from <u>www.getty.edu</u>
  - Delayed to accommodate loading to reach critical mass. Some possibilities are:
    - Getty Museum and other GRI objects
    - ARTStor (GRI is currently working on contract)
    - Built Works Registry (BWR)
    - The Met
    - Rijksmuseum
  - Earliest target date for CONA on the Web is January 2013
- Part 4 Yearly Export and Web Services APIs
  - Earliest target date is July 1, 2013

# Exports Vocabularies on the Web (AAT, ULAN, & TGN Yearly publications (UTF-8 XML & REL) Web Services







#### Web Services

#### Access to basic information

- Information that is available on our public web site is also available via our Web Services
  - Term Match uses the same criteria
    - AAT Term and Note
    - ■TGN Name, Place Type, Nation
    - ULAN Name, Role, Nationality
  - Full record for a particular subject ID
  - Preferred and non-preferred parent hierarchies for a particular subject ID
  - Immediate preferred and non-preferred children for a particular subject ID

#### Web Service

#### Ability to synchronize

- List of records that have been merged along with their old and new IDs
- Current ID for a given subject ID, accounting for whether or not a record has been merged
- List of records that have been modified based on a date range and one of the following parameters
  - Overall subject record edits
  - Added, deleted, modified terms
  - Scope note edits
  - Moved records
  - New records

# Examples of New Web Services for In-house Reports and CONA CVA

- GetChildCount returns an indexing alpha character along with the count of the number of subject records with a
  preferred term starting with that letter
- GetChildrenReports returns basic child record information given parent ID, alpha index character, language code and candidate status
- GetMergedSubjectIDsReports returns the count, subject IDs and merge dates of all records merged into the given current subject ID
- GetMergedSubjectReports returns full record information for a merged subject record
- □ **GetParentHierarchy** returns total parent levels along with pertinent parent/child information for each level
- GetSourcesReports returns full source information based on the date of edits or creation, alpha index, and editor
- GetSubjectBriefReports returns a limited set of full record data for a current subject ID
- GetSubjectIDsDiacritics returns a list of subject IDs given a comma delimited list of Getty diacritic codes; also allows user to input a hierarchy parent ID and candidate status
- GetSubjectIDsReports returns a list of subject IDs given a broad range of criteria: creation/edit date, alpha
  indexing character, editor, candidate status, parent hierarchy ID, and contributor
- GetSubjectReports returns full record information given a current subject ID
- CreateSubject creates a candidate record for a given preferred term and parent ID
- TGNGetSubjectIdsCoordinates returns a list of subject IDs given a set of coordinate points, parent key and candidate status
- CONAlconographyMatch term search matching service for the Iconography module within CONA
- CONACreateIconography creates a candidate Iconography record given a preferred term

## Conservation Thesaurus

A project of the Getty Conservation Institute, in collaboration with GRI and ITS

#### What we learned over the past 2 years

- There is a greater overlap between the subject headings in AATA and AAT than we thought
- We could NOT find any conservation concepts that were out of scope for AAT
- We could NOT find any AAT concepts that would never be used in conservation
- AAT was not the only vocabulary that would be useful to the conservation community

#### As a result the focus has shifted

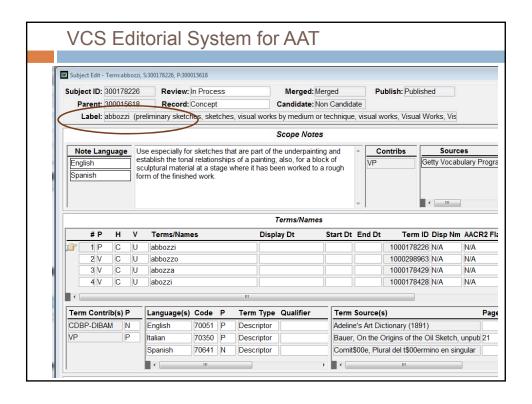
- □ From
  - Building a separate thesaurus specifically for conservation
- □ To
  - Educating the conservation community on how to use AAT, TGN & ULAN
  - Facilitating contributions from the conservation community to AAT and CONA

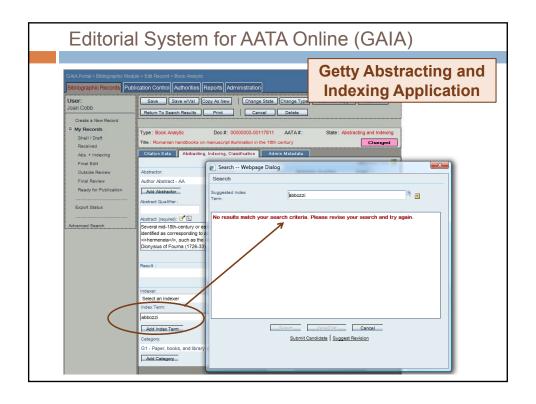
# Some Current Getty Conservation Institute Goals

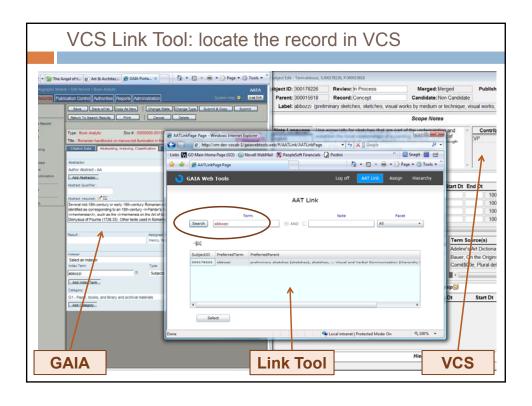
- Use all of the Getty Vocabularies as much as possible with GCI resources
- Contribute GCI concepts and terms to AAT and CONA
- Use the tools we built in Part 2 to keep AATA and AAT synchronized
- Facilitate batch contributions from targeted institutions
  - International Center for the Study of the Preservation and Restoration of Cultural Property (ICCROM)
  - Conservation Department, The Museum of Fine Arts, Houston (MAFH) – Art Conservation Database Project (ACD)
- Present a presentation at the annual meeting of the American Institute for Conservation (AIC) in May, 2012

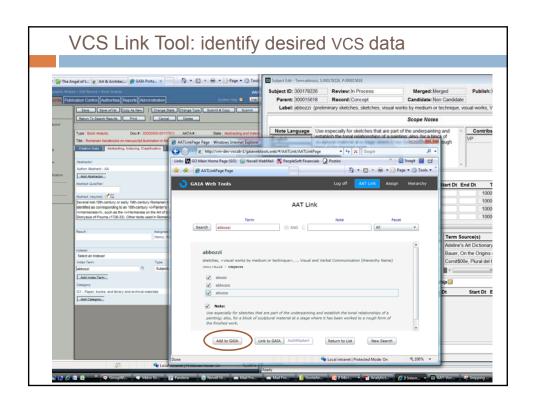
## VCS Link Tool

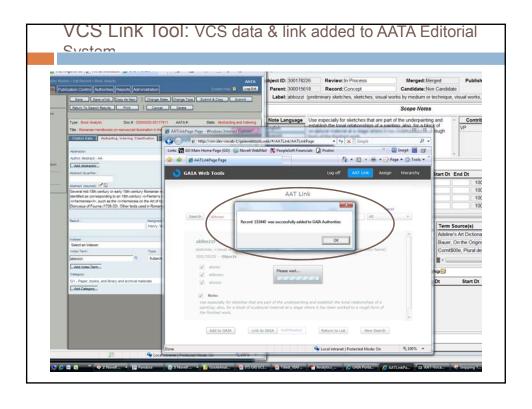
A way to keeping AATA Online indexing synchronized with AAT

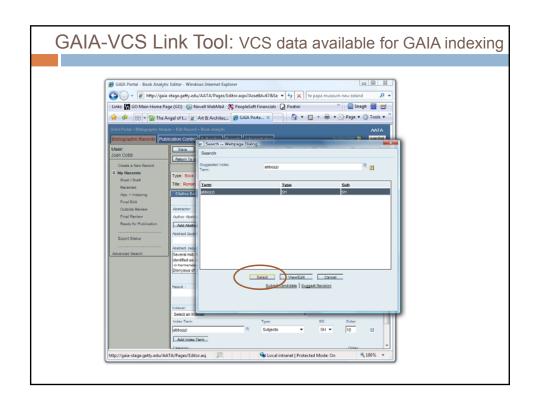


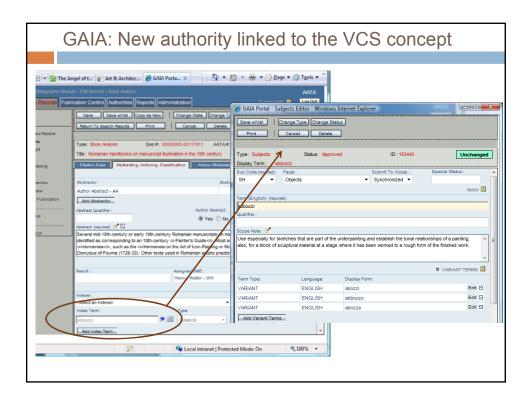


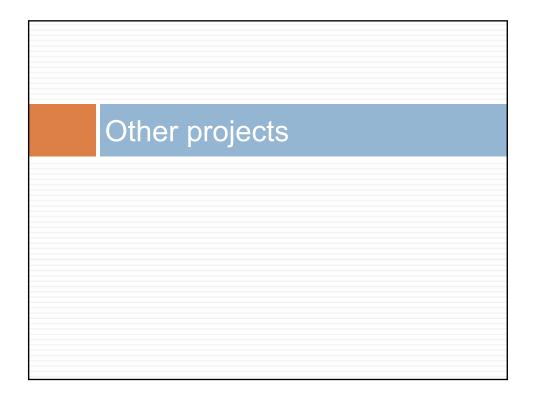












# AATA Online and GCI Project Bibliographies (PBO)

- Site will be redesigned with new branding
- Changes to the site will be rolled out in phases
- Thesaurus assisted access will be included ASAP
- Current Technical Upgrade Go-Live: February 2012
  - Will include only those features that will be carried forward to the redesigned site (e.g., Browse Categories, Flexible Results Lists, Multiple Search Forms, Print, Download)

#### Getty Resource Discovery Project

- A Trust-wide initiative to create integrated access to the Getty's collections and resources.
- Phases of development
  - Prototype
  - Usability Study
  - □ Phase One (Museum, Library, special collections)
  - Phase Two (include additional resources)
  - Phase Three (integrate the Getty vocabularies)

## Contact Information

Joan Cobb: <a href="mailto:jcobb@getty.edu">jcobb@getty.edu</a>
Gregg Garcia: <a href="mailto:ggarcia@getty.edu">ggarcia@getty.edu</a>

Information Technology Services
J. Paul Getty Trust