





TGN



ULAN



CONA

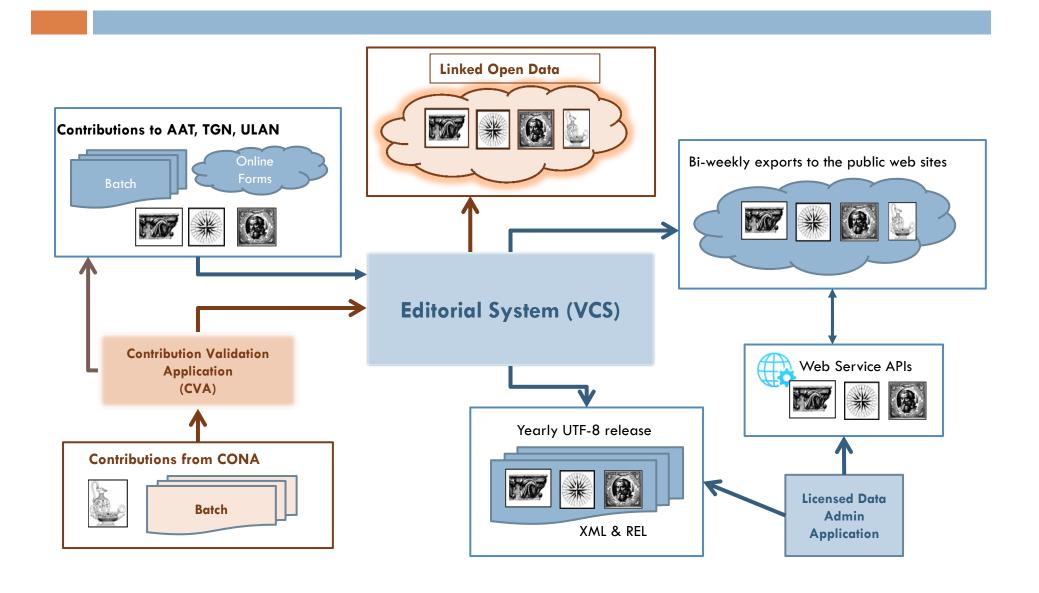
THE GETTY VOCABULARIES TECHNICAL SOLUTIONS OVERVIEW

Joan Cobb IT Specialists Project Manager Getty Information Technology Services



International Working Group Meetings Friday, September 5, 2014

Current Technical Solutions





allow their use in linked data. A project to publish AAT, TGN, ULAN, and CONA to the LOD (Linked Open Data) cloud is underway. The documents on this page contain news and presentations about releasing the Getty vocabularies as LOD. These materials are subject to frequent modification and addition.



- News and Status of the Project
- What Is LOD?

▶ Photo Archive

BHA & RILA

Research Guides & Bibliographies

■ Digital Collections

→ Getty Vocabularies

(ULAN)®

 Contribute Editorial Guidelines

Art & Architecture

Thesaurus (AAT) ®

Authority (CONA) ®

Geographic Names (TGN) ®

▶ Union List of Artist Names

Getty Vocabularies as

■ Frequently Asked Questions Dobtain the Getty Vocabularies

Linked Open Data

■ Training Materials

▶ Getty Thesaurus of

Article & Research Databases

▶ Collecting & Provenance

- Introduction to Getty Vocabularies as LOD (PDF, 3.8 MB, 46pp)
- Linked Open Data Flier (PDF, 1 MB, 1pp) Cultural Objects Name
 - List of External Advisors (PDF, 88KB, 7pp)

The AAT and TGN are now available as LOD. They are published under the ODC-By 1.0

Developers and programmers, technical documentation to help you explore the AAT and TGN data is available at the SPARQL endpoint at vocab.getty.edu

News and Status of the Project

Ontology update: Note that the ontology has changed. Click on the "Developers and

Additional releases: Plans are in place to release the other vocabularies as LOD on the following dates: ULAN in January 2015 and the CONA demo in July 2015.

Other formats: In addition to LOD, we plan to continue providing the data in relational tables and XML releases. If any format is discontinued in the future, users will be given advance notice.

Working with the community: We have established an open community and we welcome collaboration. Previously we were getting feedback at an email address, but now the preferred way of communicating technical questions, comments, and usage stories is to use the SemanticWeb.com Q&A forum. We will be monitoring this forum regularly. The tags to use are: Getty, AAT, TGN, ULAN, CONA. The following links will pre-populate with the appropriate tags:

http://answers.semanticweb.com/questions/ask/?tags=Getty,AAT http://answers.semanticweb.com/questions/ask/?tags=Getty,TGN http://answers.semanticweb.com/questions/ask/?tags=Getty,ULAN http://answers.semanticweb.com/questions/ask/?tags=Getty_CONA

For questions and comments about editorial content or general information regarding the Getty Vocabularies, please contact us at vocab@getty.edu

URIs: A description of the full set of URIs for all four vocabularies is available in "Getty Vocabulary Semantic Representation, section 1.5.3 GVP URLs and Prefixes.

The semantic links to the AAT and TGN subjects are now live. For example: http://vocab.getty.edu/aat/300311452 http://vocab.getty.edu/tgn/7001393

Other vocabularies: Until the other vocabularies are available in LOD format, as requested by several in the community, we have implemented a redirect to the HTML version of the full record. Here are examples in each vocabulary:

http://vocab.gettv.edu/ulan/500125633 http://vocab.getty.edu/cona/700001941

What Is LOD?

A current trend in managing art information is to increasingly make data about art, architecture, and cultural heritage objects available as Linked Open Data (LOD). This applies to the information directly describing the objects, but also to the vocabularies used in the description

When data is linked and open, it means that data is structured and published according to the principles of Linked Data, so that it can be both interlinked and made openly accessible and shareable on the Semantic Web. The goal of linked open data is to allow data from different resources to be interconnected and queried, thus making it more useful. Although the idea of linking data in an open way is not new, the widespread practice of doing so is relatively new, thus the protocols, standards, and licensing options used for linked open data are still evolving.



In order for data to be understood and processed automatically by computers, data in records or about resources must be expressed in a standard format. Each thing (for example, a museum object, a place, or a person) must be represented by a persistent identifier (known as a Uniform Resource Identifier (URI)). A Resource Description Framework (RDF) is a language or format for describing things as well as the relationships between things in terms of simple properties and values (known as 'triples'), while things are represented using URIs. Among the most often-used formats for publishing art vocabularies are the Simple Knowledge Organization System (SKOS) and Web Ontology Language (OWL).

If data is to be open to the community for linking and discovery, traditional licensing and copyright practices for art information, images, and associated vocabularies must be adjusted. Data is considered open if the community is free to use, reuse, and redistribute the data, subject either to no restriction or to only the requirements of attribution or share -alike, Among the licenses most often applied to art information are Creative Commons and Open Data Commons licenses, each of which offers a range of levels of openness.

The Art & Architecture Thesaurus® (AAT), the Union List of Artist Names® (ULAN), the Getty Thesaurus of Geographic Names® (TGN), and the Cultural Objects Name Authority® (CONA) are copyrighted by the J. Paul Getty Trust.

Revised 19 August 2014

Image caption: The Getty Vocabularies as LOD will allow linking and discovery of artworks, architecture, people, places, styles, techniques, materials, and other information related to cultural heritage. Clockwise from top left, images are from the following collections: 2012.R.17, 2003.R.22, 2679-146, 2003.R.22, 96.R.19



Getty Vocabularies: LOD

SPARQL

Any -

Search

Brief -

The Getty Vocabularies

Welcome to the SPARQL endpoint vocab.getty.edu

The Getty Vocabularies: The AAT, TGN, ULAN, and CONA contain structured terminology for art and other material culture, archival materials, visual surrogates, and bibliographic materials. Compliant with international standards, they provide authoritative information for catalogers and researchers, and can be used to enhance access to databases and Web sites. The Getty Vocabularies are produced by the Getty Vocabulary Program (GVP) and grow through contributions.

NOW Available:



The Art & Architecture The

Catherine wheel or rose wind and architecture.



The Getty Thesaurus of Ge London or Londinium? TGN

Ontology Update Note: There are cha

Documentation and Downloa

- GVP Semantic Representation: Sections GVP URLs and Prefixes a Comprehensive Sample Queries a
- GVP ontology: HTML, RDF/XML
- Datasets: (NTriples Zip, see doc s
 - AAT: full.zip (all statements), e
 - TGN: full.zip (all statements), e
- VOID: void.ttl (see doc section Des
- AAT Relationship Types: Crossv
- · General information about the L

Explore the Dataset:

- Text search: Use 'All/AAT/TGN' to search only terms. Use 'Full' to sea
- SPARQL Query: Click on the 'SPA provides a lot more detail!

Questions and Comments:

For technical questions and comments, please post them on the SemanticWeb.com Q&A forum. We will be monitoring that forum regularly. The tags to use are: Getty, AAT, TGN, ULAN, CONA. The following links will pre-populate with the appropriate tags:

- · http://answers.semanticweb.com/questions/ask/?tags=Getty,AAT
- · http://answers.semanticweb.com/questions/ask/?tags=Getty,TGN
- · http://answers.semanticweb.com/questions/ask/?tags=Getty,ULAN
- http://answers.semanticweb.com/questions/ask/?tags=Getty,CONA

For questions and comments about editorial content or general information regarding the Getty Vocabularies, please contact us at vocab@getty.edu.

Usage Stories:

It is our hope that we can all learn from each other, so please don't be shy about letting the community know how the LOD versions of the Getty Vocabularies are useful in your work. The best way to do this is to post to the SemanticWeb.com Q&A forum as described in the 'Questions and Comments' section above.

Openness

The data is released under Open Data Commons Attribution License. Adherence to ODC Attribution instructions for the correct assertion of attribution is encouraged. The preferred form of attribution for AAT is:

"This [title or report or article or dataset] contains information from Art & Architecture Thesaurus (AAT)®, respectively Thesaurus of Geographic Names (TGN)® which is made available under the ODC Attribution License."

Special Case: In circumstances where providing the full attribution statement above is not technically feasible, the use of canonical AAT/TGN URIs is adequate to satisfy Section 4.3 of the ODC Attribution License.

Future Publications



The Union List of Artist Names (ULAN) ®

Titian or Tiziano Vecellio? ULAN is a structured vocabulary, including names, biographies, and other information about artists and architects

January 2015

July 2015



The Cultural Objects Name Authority (CONA) ®

Mona Lisa or La Gioconda? CONA, a new vocabulary now accepting contributions, includes titles, attributions, and other information for art and architecture.

Additional Links:

- · The Getty Vocabularies
- . The Getty Research Institute
- · The Getty (Main)

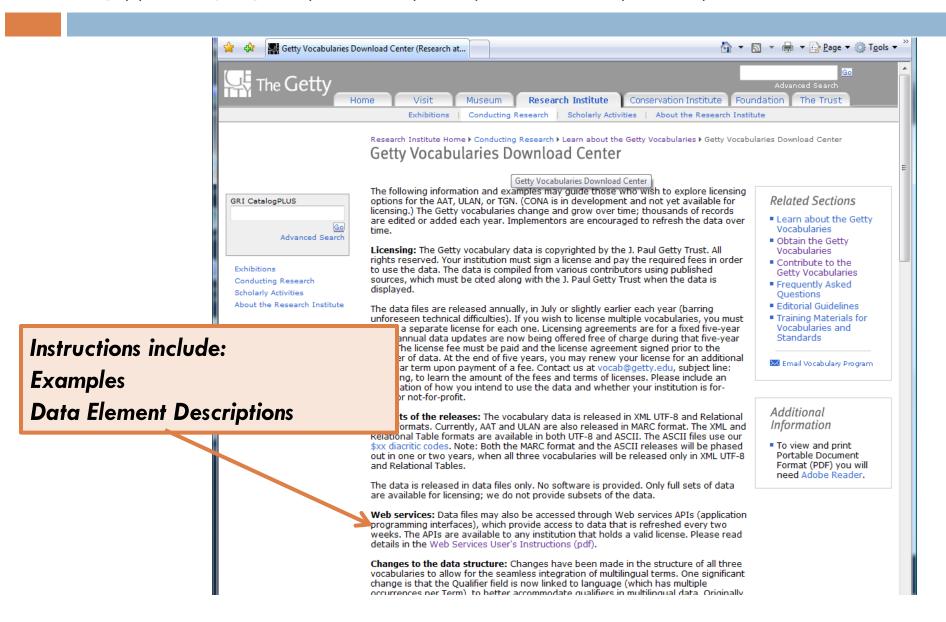
Vocabularies on the Web

Automatically updated bi-weekly



Web Services - User's Instructions

http://www.getty.edu/research/tools/vocabularies/obtain/download.html



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

Yearly Exports

Sample Data available in UTF-8

XML

Relational Tables Download sample records from the Art & Architecture Thesaurus (AAT). The full AAT is a hierarchical vocabulary of around 34,000 records, including 134,000 terms, descriptions, bibliographic citations, and other information relating to fine art, architecture, decorative arts, archival materials, archaeology, and other material culture. The full XML data file for the AAT is 20,425 kilobytes in size. The sample available here is a small subset of the AAT. The next data release will take place in July 2012. For further information, see About the AAT.

AAT Sample Data
Format options:

Data Dictionary for the AAT Data Release

XML UTF-8 (2.38 MB)
Relational Table UTF-8 (40.4 MB)

XML format PDF

Relational Table format PDF

Download sample records from the Union List of Artist Names (ULAN). The full LAN is a vocabulary of around 162,000 records, including 453,000 names and biographical and bibliographic information for artists, architects, firms, shops, and art repositories, including a wealth of variant names, pseudonyms, and language variants. The full XML data file for the ULAN is 97,383 kilobytes in size. The sample available here is a small subset of the ULAN. The next data release will take place in July 2012. For further information, see About the ULAN.

ULAN Sample Data

Data Dictionary for the ULAN Data Release

Format options:

XML UTF-8 (14 MB)

XML format PDF

Relational Table UTF-8 (92 MB)

Relational Table format PDF

Download sample records from the Getty Thesaurus of Geographic Names (TGN). The full TGN is a hierarchical vocabulary of around 895,000 records, including 1.1 million names, place types, coordinates, and descriptive notes, focusing on places important for the study of art and architecture. TGN is not a GIS; coordinates are included for many records, but they are for finding purposes only (i.e., to find the place on a map). The full XML data file for the TGN is 133,312 kilobytes in size. The sample available here is a small subset of the TGN. The next data release will take place in July 2012. For further information, see About the TGN.

TGN Sample Data

Data Dictionary for the TGN Data Release

Format options:

XML format PDF

XML UTF-8 (1.10 MB)

Relational Table format PDF

Relational Table UTF-8 (7.86 MB)

http://www.getty.edu/research/tools/vocabularies/obtain/download.html

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

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

Loading the data

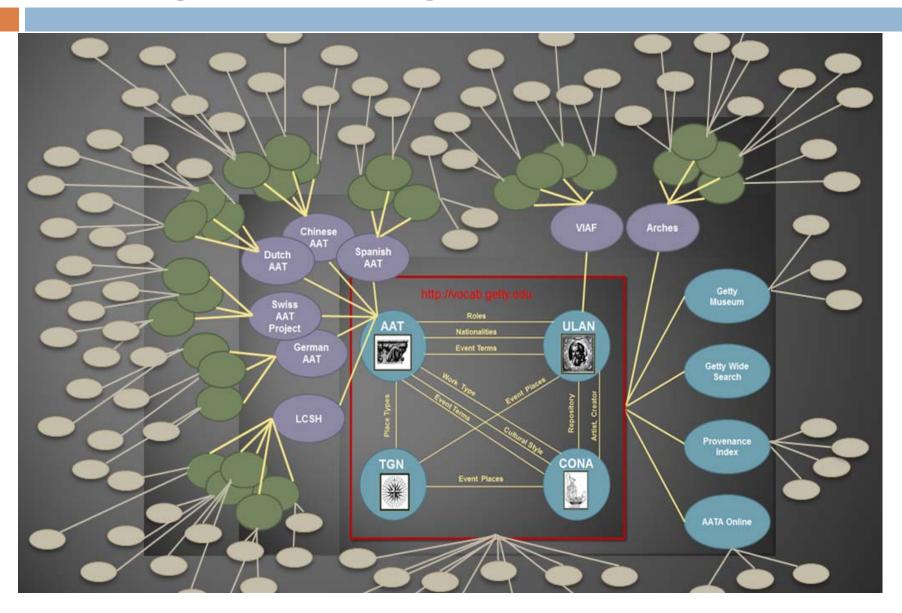
Multi-lingual data

- 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

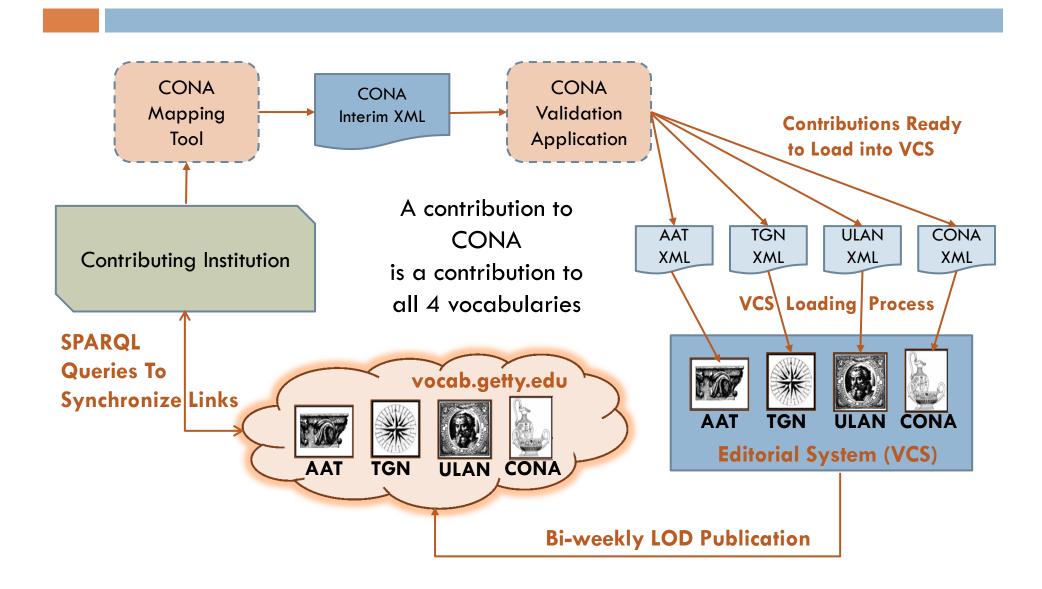
Contribution XML schemas

- 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

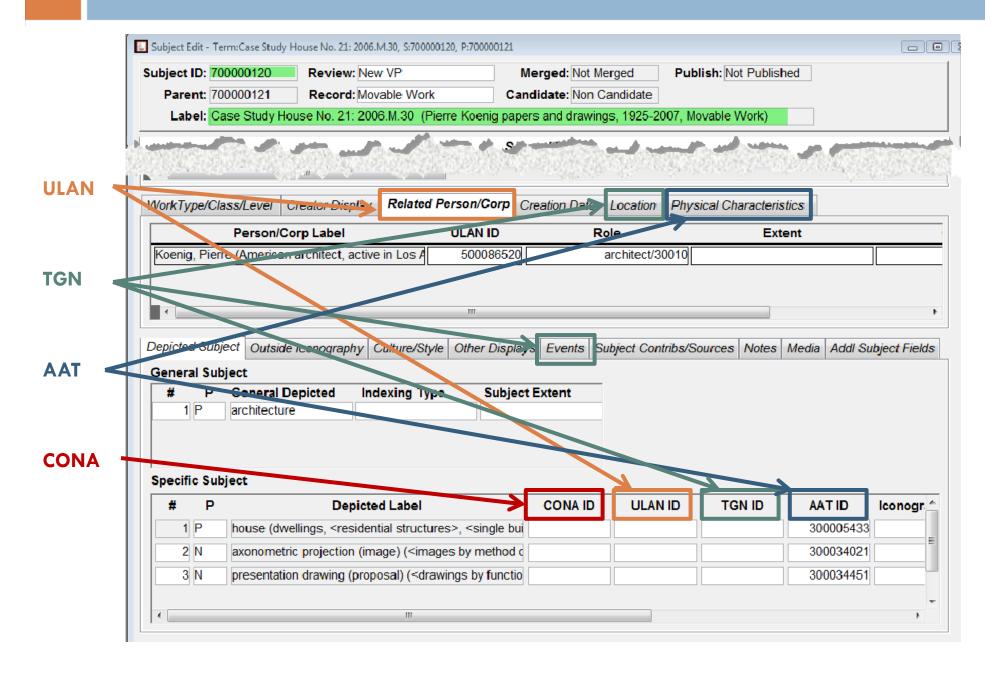
The Getty Vocabularies as Conduits to Other Resources



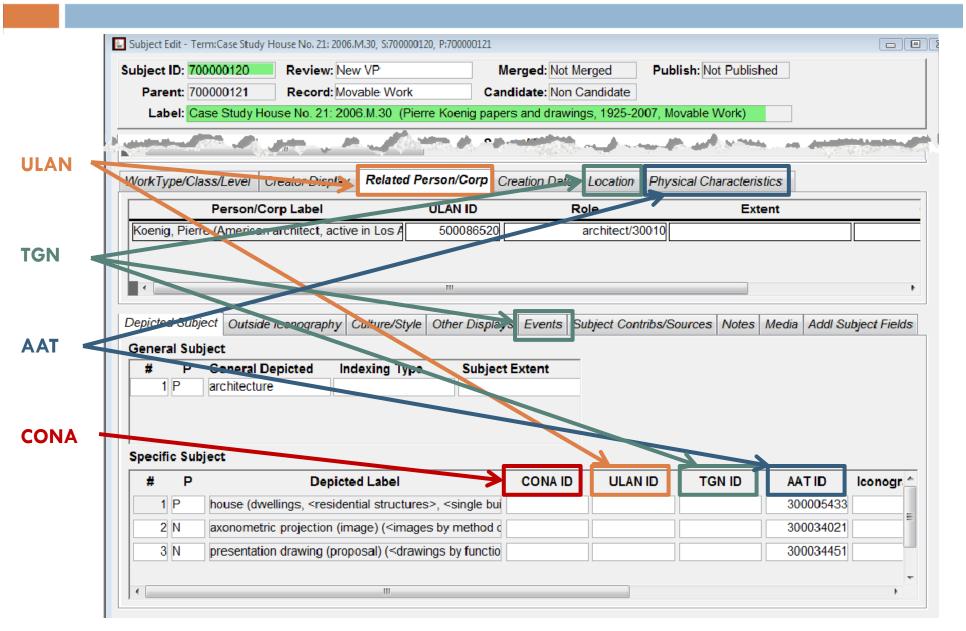
CONA Can Help Connect Resources



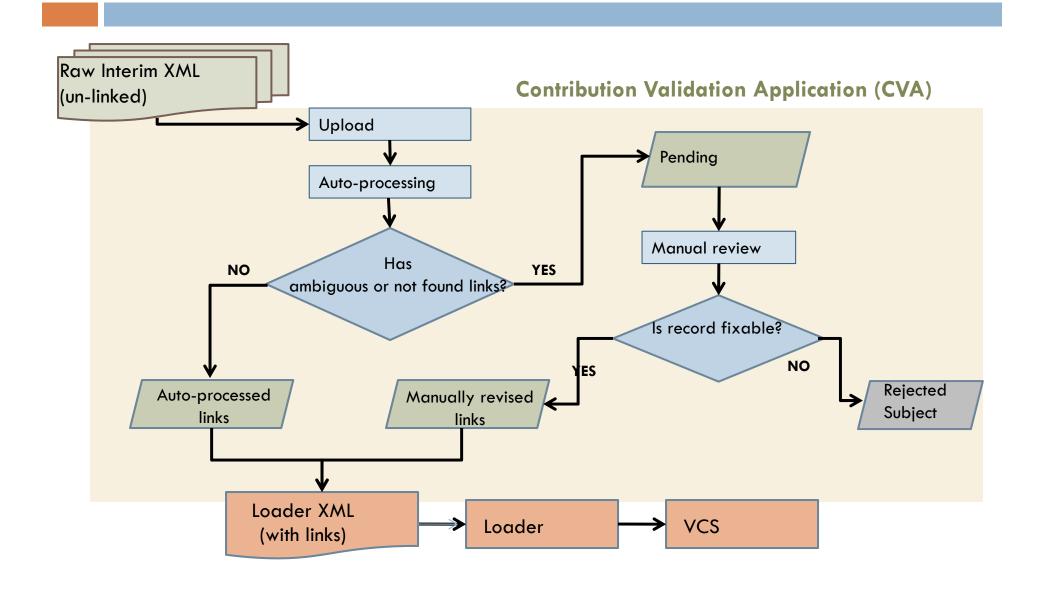
CONA-VCS: Live links to AAT, TGN & ULAN



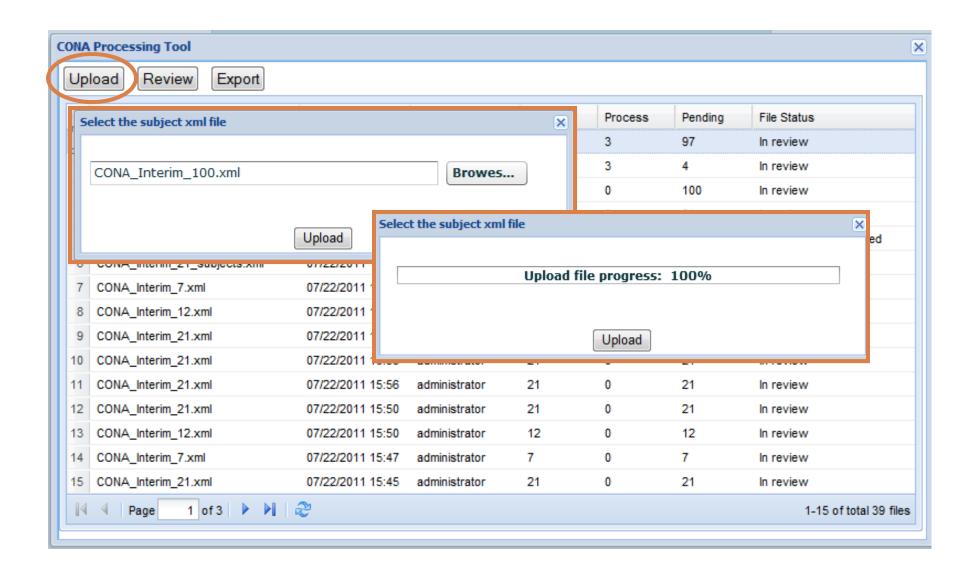
A CONA contribution may also be a contribution to AAT, TGN & ULAN



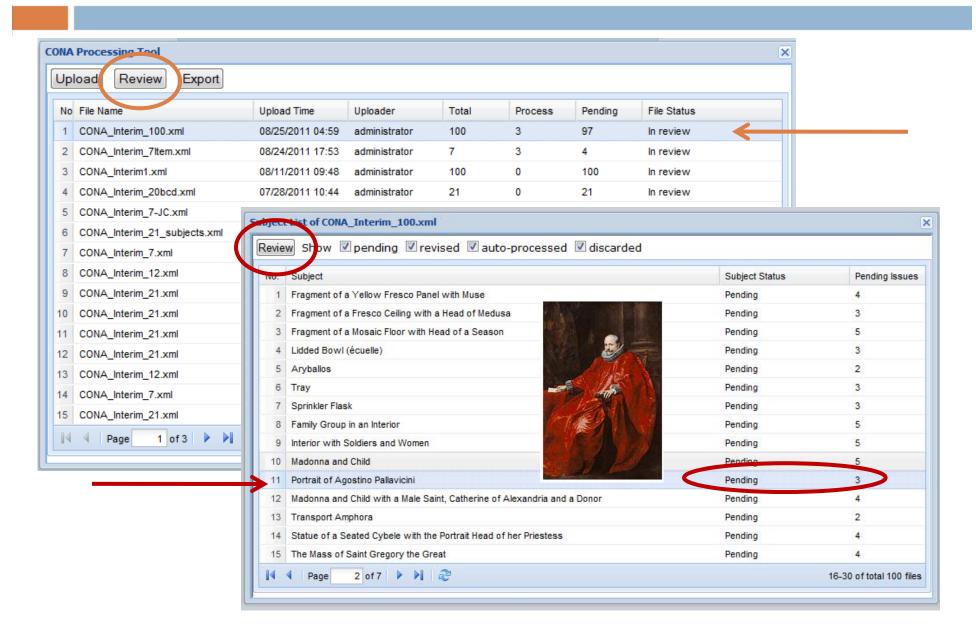
Loading CONA data



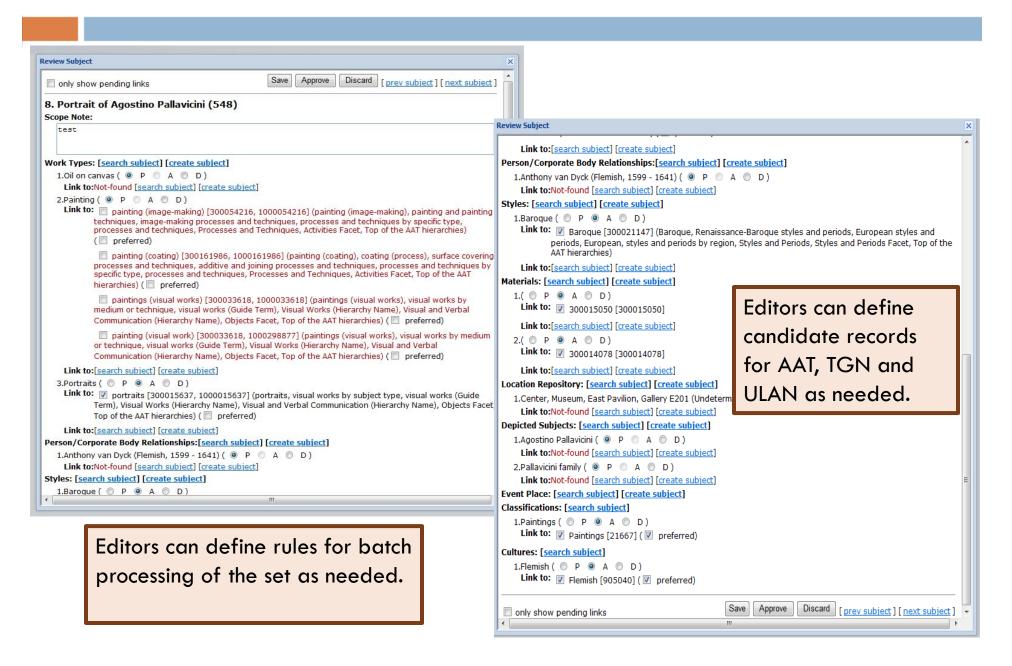
Upload Raw Interim XML file



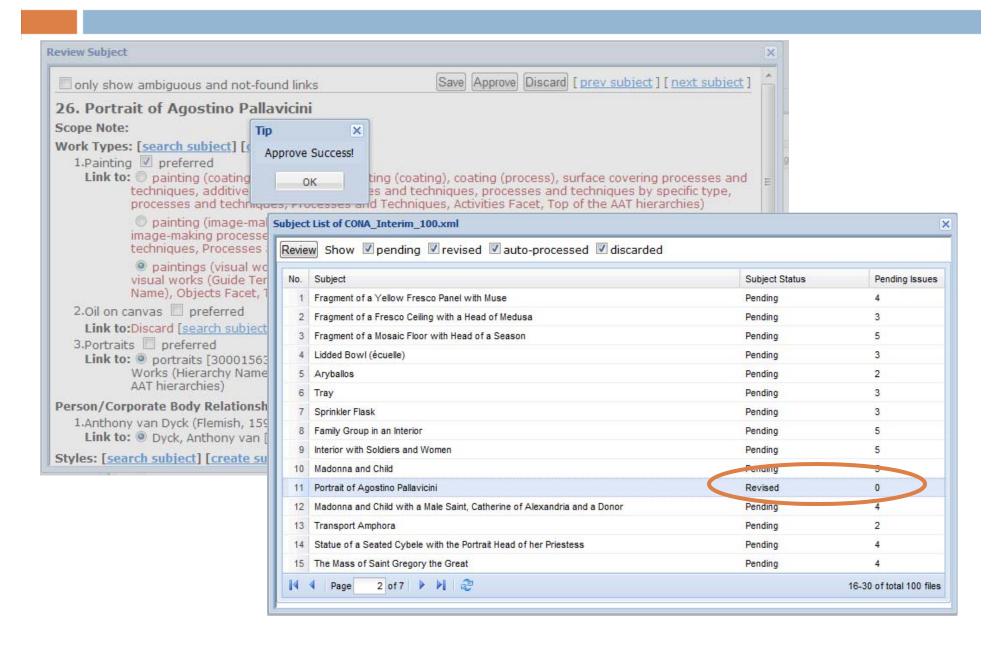
Manually review the list



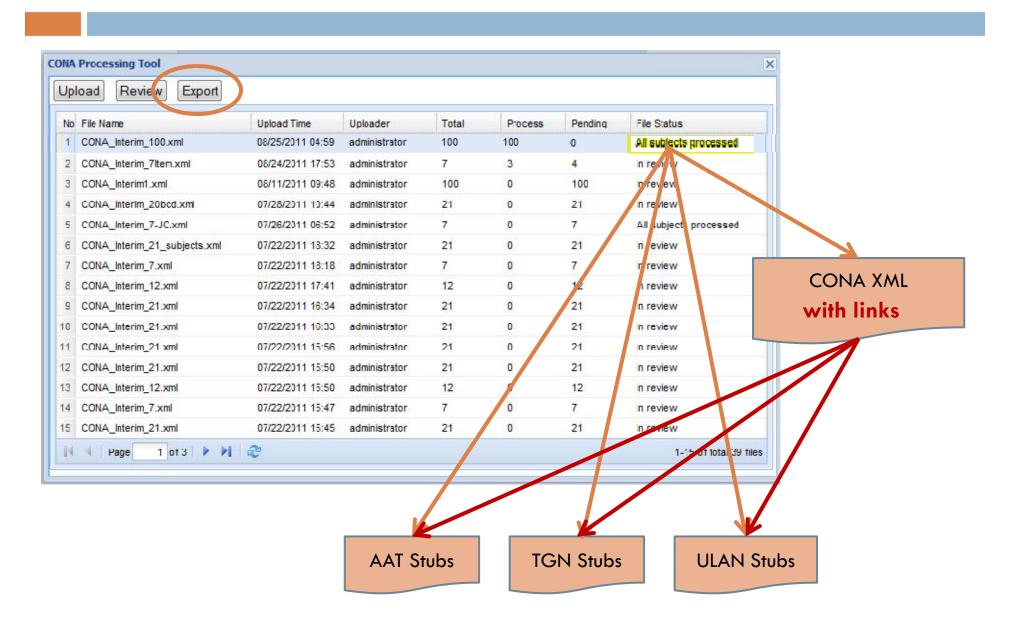
Review the record



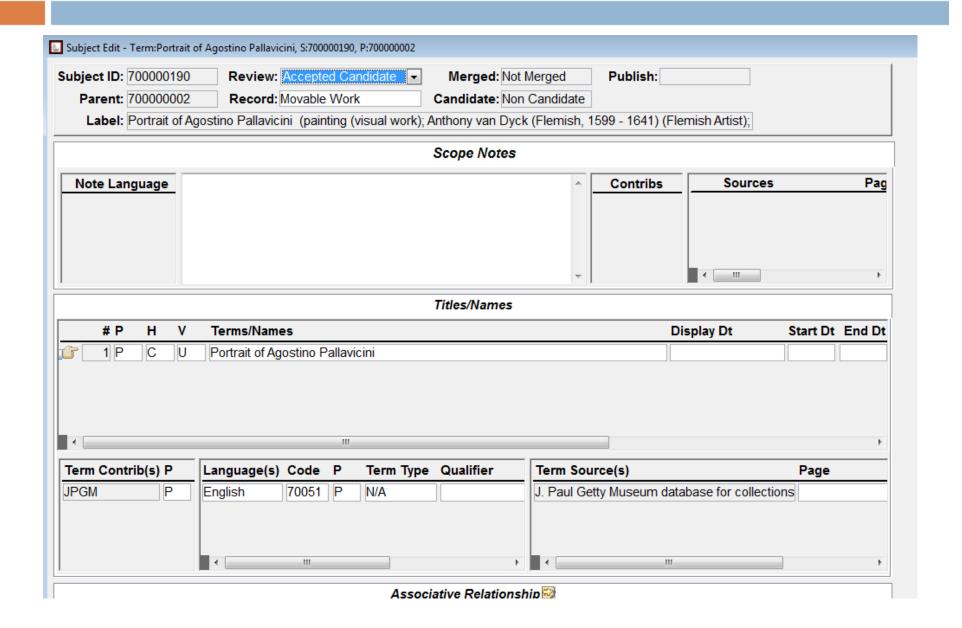
Approve the record



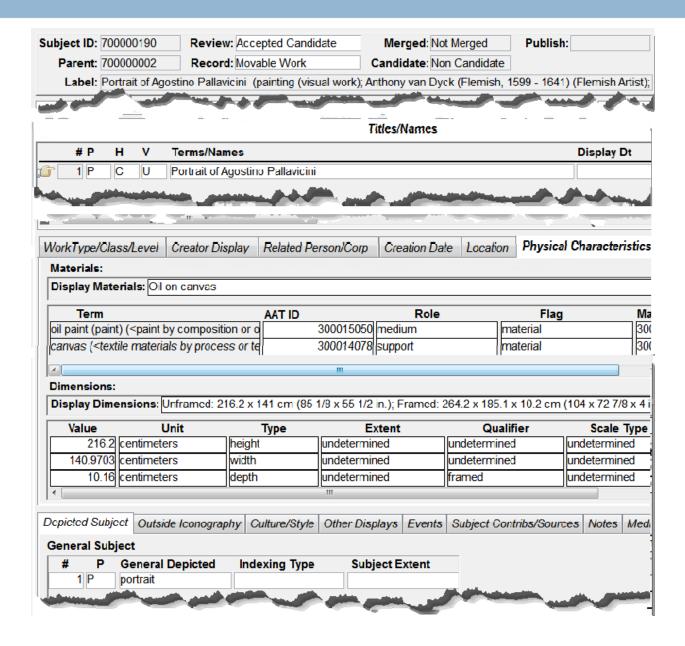
Ready to load into our editorial system



Edit the record in the editorial system if necessary

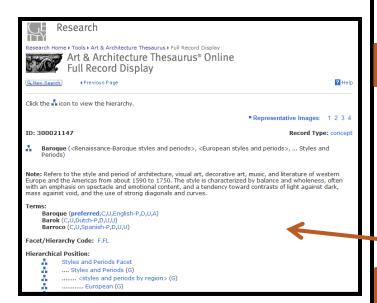


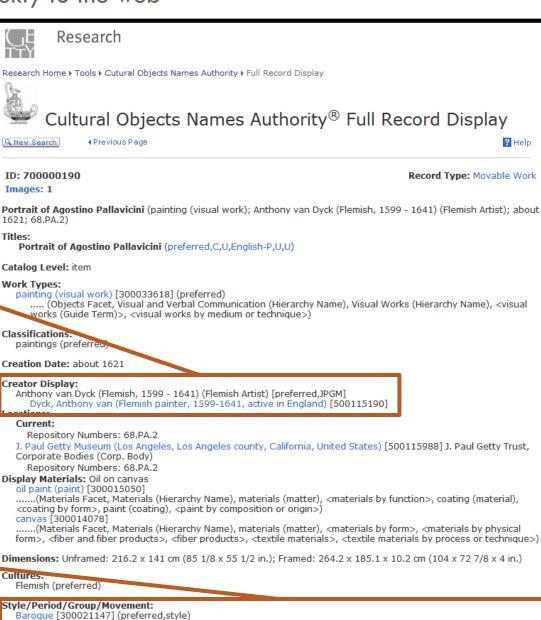
The links established in CVA are already there



Records are published bi-weekly to the web



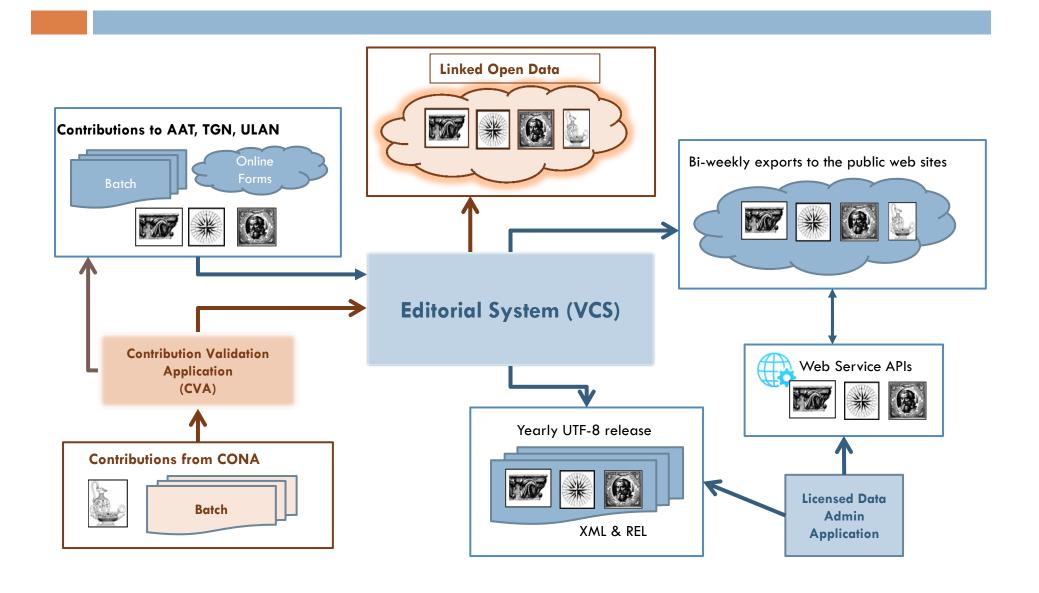




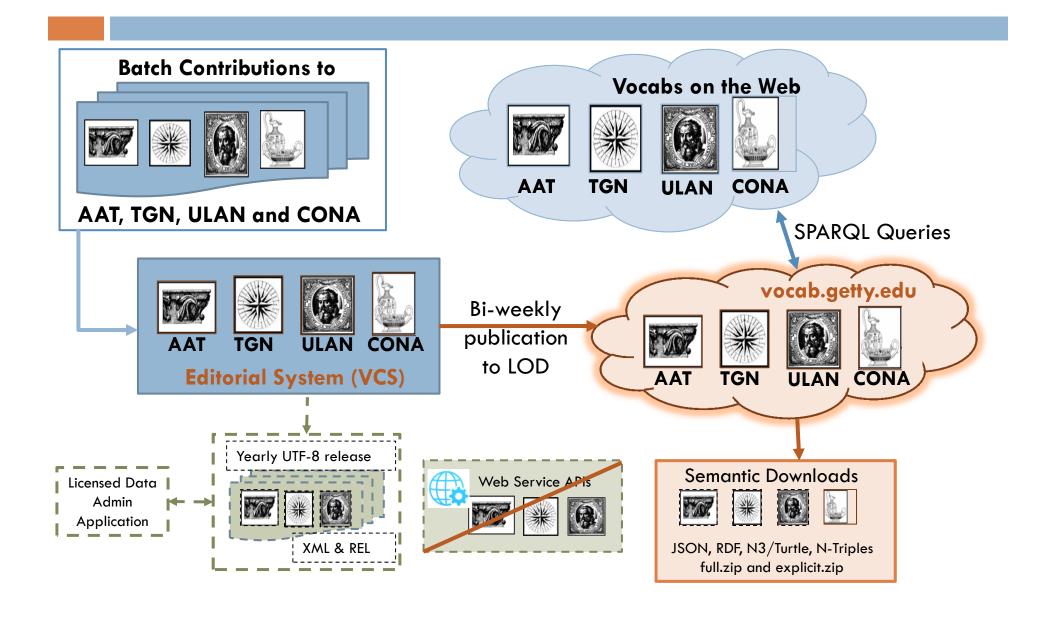
.....(Styles and Periods Facet, Styles and Periods, <styles and periods by region>, European, <European styles and

neriods> <Renaissance-Baroque styles and periods>)

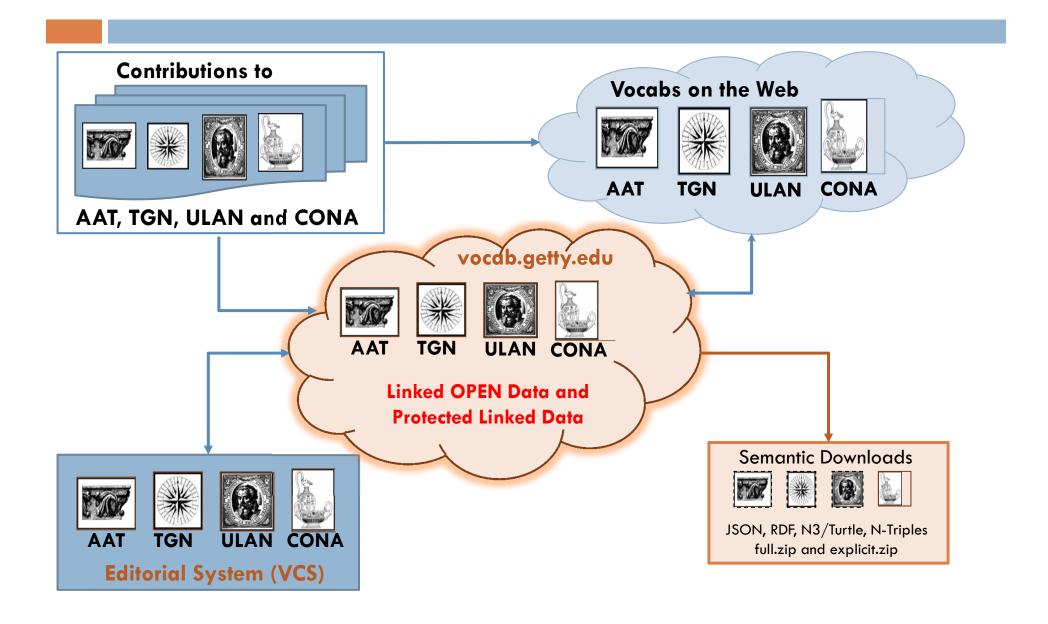
Current Technical Solutions



Possible Mid Term Technical Solutions



Possible LongTerm Technical Solutions



Joan Cobb

IT Specialist Project Manager Information Technology Services

> 1200 Getty Center Drive Los Angeles, CA 90049

> > jcobb@getty.edu



The J. Paul Getty Trust