Linked Data: Publication, Coreferencing and Usage with the Getty Vocabularies

GREGG GARCIA
GGARCIA@GETTY.EDU
SOFTWARE ARCHITECT, J PAUL GETTY TRUST
What Are the Getty Vocabularies?

AAT      TGN      ULAN      CONA
These presentations are not primers on **controlled vocabularies** nor on **what is LOD**.

They focus on issues surrounding the Getty vocabularies specifically.

For basic information on vocabularies, see *Introduction to Controlled Vocabularies* (Harpring, revised 2013) and its bibliography.

For basic information on what is LOD, see this Europeana video [http://vimeo.com/36752317](http://vimeo.com/36752317). World Wide Web Consortium (W3C) Glossary: [www.w3.org/TR/ld-glossary/](http://www.w3.org/TR/ld-glossary/)
What are the Getty vocabularies?

- *Catherine wheel* or *rose window? Mona Lisa* or *La Gioconda?*

- The AAT, TGN, ULAN and CONA contain multilingual terminology and other related data to describe
  - visual art, architecture, other cultural heritage works, conservation, archaeology, archival materials, visual surrogates, and related bibliographic materials
 Getty vocabularies are valued as authoritative
  - Grow through contributions from experts
  - Contributors and sources are cited

 Compiled, merged, edited, and published by the Getty Vocabulary Program and our talented, tireless technical team

 Data is made available in various ways: via online searching; relational tables, XML format, Web services APIs
  - Now as Linked Open Data: structured and published to make it openly accessible and shareable on the Semantic Web
    - Open Data Commons Attribution License (ODC-By) v1.0
    - Allows sharing, creation, adaptation of data with attribution
- Getty vocabularies are thesauri compliant with national (NISO) and international (ISO) standards for thesaurus construction.

- They are compliant with / map to other standards:
  - CDWA (Categories for the Description of Works of Art)
  - CCO (Cataloging Cultural Objects)
  - VRA Core (Visual Resources Association core categories)
  - LIDO (Lightweight Information Describing Objects)
  - etc.
Among the features that allow the Getty vocabularies to be used in cataloging, retrieval, and linking are the following:

- Each Getty vocabulary records and terms are identified by **unique numeric IDs**
- The Getty vocabularies are linked to each other
- The Getty vocabularies share a core data structure
Scope and Content of the Getty Vocabularies

AAT    TGN    ULAN    CONA
AAT, the Art & Architecture Thesaurus®

Generic terms for describing art and architecture (decorative arts, other material culture, visual surrogates, archival materials, archaeology, and conservation) (e.g., *oil paintings*, *Baroque*, *lithographers*, *sintering*).

Current totals: 36,000 records; 245,000 terms

- AAT is multilingual; large translation projects are underway
- Conceptually organized from terms to describe abstract concepts to generic terms for concrete, physical artifacts
- Facets are the upper levels of the AAT structure
- AAT is not organized by subject matter or discipline
stirrup cups
coaching glasses
hunting glasses

sturzbechers
Sturzbecher
stortebekers

rhyta
rhyton
rhytons
rhea
rheon
rheons
ritón
**Sample record**

**ID:** 300132869

**Terms:**
- bobbin lace (pref, en)
- bone lace (en)
- cushion lace (en)
- 梭心蕾絲 (zh)
- 線軸編織花邊 (zh)
- kloskant (nl)
- dentelle aux fuseaux (fr)
- encaje de bolillos (es)
- encaje de bolillo (es)
- Klöppelspitze (de)
- Klöppelspitzen (de)

**Associative Relationships**
- requires ... lace pillows (<textile fabricating tools and equipment>...Objects Facet) [300132869]

**Hierarchical Relationships (poly)**
- Objects Facet
- ... Visual & Verbal Communication
- ..... Visual Works
- ........ visual works (works)
- .......... <visual works by material>
- ............ needlework (visual works)
- ............... lace (needlework)
- ................ bobbin lace

**Note:** With "needle lace," one of two primary types of handmade lace. It is characterized by being made by ...

**Chinese (traditional) ..... 與「針織蕾絲（needle lace）」同為主要的手工蕾絲,其特色是以纏繞於線軸或梭心...

**Dutch ..... Een van de twee belangrijkste soorten met de hand vervaardigde kant; 'naaldkant' is ...

**German ..... Zusammen mit der “Nadelspitze” eine der wichtigsten...

**Spanish ..... Junto a "encaje a aguja", uno de los dos tipos principales de ...

**Contributors:** VP, CHIN, AS, RKD, IfM-SMB-PK, CDPB-DIBAM

**Sources:** Earnshaw, Clabburn, Needleworker's Dictionary (1976); Identification of Lace, 2d ed. (1984); Ginsburg, Illustrated History of Textiles (1991)
Translations are underway

- Getty Conservation Institute (Los Angeles, California)
- Academia Sinica (Nankang, Taiwan)
- Netherlands Institute for Art History (RKD: Bureau AAT) (The Hague, The Netherlands)
- Centro de Documentación de Bienes Patrimoniales (Dirección de Bibliotecas, Archivos y Museos) (Santiago, Chile)
- Staatliche Museen zu Berlin Preussischer Kulturbesitz (Berlin, Germany)
- [Canadian Heritage Information Network (CHIN) (Gatineau, Quebec, Canada), Istituto Centrale per il Catalogo e la Documentazione (Rome, Italy) (ICCD)] and other contributors of partial translations; future translations in French, Italian, Portuguese, other languages
TGN, the Getty Thesaurus of Geographic Names®
Names, other information for current & historical administrative places (cities, nations, empires) and physical features (e.g., Firenze, Roman Empire, Ganges River).
Current totals: 1,241,000; 1,800,000 names

- It includes formerly inhabited places, archaeological sites, and historical places with unknown exact locations (i.e., “lost settlements”)
- It focuses on places important to study of art and related disciplines
- It grows through contributions from large national geographic databases, the expert user community, archaeology projects, scholars, etc.
**Sample record**

**ID:** 7011179

**Names:**
- Siena (it, en, es, fr, nl, de)
- Σιένα (el)
- 锡耶纳 (zh)
- シエーナ (ja)
- Sienese (en-adj)
- Sienna (h-en)
- Sena Julia (h-la)
- Sanna (h)
- Saena (h)

**Hierarchical Rel. (polyhier.)**
- World (facet)
  - Europe (continent)
  - Italy (nation)
  - Tuscany (region)
  - Siena (province)
  - Siena (inhabited place)
- World (facet)
  - Europe (continent)
  - Italian Peninsula (peninsula)
  - Etruria (former confederation)
  - Siena (inhabited place)
  - Historical: Late Etruscan

**Place Types**
- inhabited place
- city
- archiepiscopal see
- commune (administrative)
- cultural center
- university center
- world heritage site

**Lat:** 43 19 07 N (43.3180)
**Long:** 011 19 50 E (11.3300)
**Elevation:** 1056.430 feet (322.0000 meters)

**Note:** Siena was founded as an Etruscan hill town; later was the Roman city of Sena Julia (3rd century BCE). It flourished under the Lombard kings (6th century CE) and was a Medieval self-governing commune...

**Contributors:** BHA, FDA, GRLPSC, VP

Images by be under additional copyright
ULAN, the Union List of Artist Names®
Names, other information for people (artists, patrons, sitters) and corporate bodies (repositories, studios) related to the design, creation, patronage, collection, conservation, etc. of art, architecture, and other cultural materials (e.g., Bartolo di Fredi, National Palace Museum)
Current totals: 223,000 records; 582,000 names

- It includes current and historical associative relationships (e.g., student-teacher, firm-member)
- Contributions to ULAN are from expert communities (museums, art libraries, cataloging projects, etc.)
- ULAN is contributed to the Virtual International Authority File (VIAF) (joint project with US Library of Congress and numerous libraries worldwide to combine name authority)
Sample record

**ID:** 500115493

**Names:**
Dürer, Albrecht (pref, de, en)
Albrecht Dürer
Duerer, Albrecht (en)
Durer, Albrecht (en)
Dürrero, Alberto (it)
Durerus, Albertus (la)
Direr, Albreht
Djurer, Albrecht
Dûrê, Albrecht
Dûrers, Albrehts
Alberto Duro Tedesco
デューラー, アルブレヒト (ja)
阿尔布雷希特·丢勒 (zh)

**Nationalities:**
German (pref)
Bavarian

**Roles:**
artist (preferred)
painter
printmaker
engraver (printmaker)
woodcutter
draftsman
illustrator
designer
mathematician
theorist
portraitist
religious artist

**Gender:** male

**Birth and Death Places:**
Born: Nuremberg (Bavaria, Germany) (inhab place)
Died: Nuremberg (Bavaria, Germany) (inhab place)

**Events:**
active: Nuremberg (Bavaria, Germany) (inhab place)
active: Venice (Veneto, Italy) (inhabited place)

**Associative Relationships:**
apprentice of Wolgemut, Michael; 1486 to 1489
.............. (German painter, printmaker, 1434/1437-1519)
child of Dürer, Albrecht, the elder
.............. (German goldsmith, 1427-1502)
founder of ...Albrecht Dürer Workshop
.............. (German workshop, established 1495)
patron was ...Maximilian I, Holy Roman Emperor
.............. (Roman emperor, 1459-1519)

**Contributors:**
VP, JPGM, Avery, WCI, Prov, GRLPSC, BHA, CCA

**Sources:**
Bartrum, Albrecht Dürer and his Legacy (2002); Bénézit, Dictionnaire des Peintres (1976; Bartsch, Le peintre graveur.

Patricia Harpring © 2014 J. Paul Getty Trust. For educational purposes only. Do not distribute.
Basics of Linked Data

• Universal Resource Identifier (URI) – “an ASCII string used to identify things on the Semantic Web” (http://www.w3.org/wiki/URI)
  http://vocab.getty.edu/aat/300198841
  http://vocab.getty.edu/tgn/1000193

• URIs are linked to each other through triples composed of subject-predicate-object relationships
  <http://vocab.getty.edu/aat/300198841> <http://www.w3.org/2004/02/skos/core#prefLabel> "rhyta"@en .

• The definitions of data elements and links described by ontologies
  http://www.w3.org/2004/02/skos/core#
  http://www.w3.org/ns/prov#
  http://purl.org/dc/elements/1.1/

• Data is delivered to a requesting agent through a triple serialization using HTTP
  RDF/XML, Notation-3 (N3), Turtle, N-Triples, RDFa, and RDF/JSON
Growth of the LOD Cloud since 2007

View of LOD as of October 2007
12 Datasets

LOD as of August 2014
570 Datasets

http://lod-cloud.net/
Where to start the implementation?

- Using the mapping developed by Marcia Zeng, we wanted to see what the Getty AAT looked like as LOD
- D2RQ provided a simple platform for taking our existing RDBMS tables and serving them as URIs
- Attached property definitions to exiting database fields
AAT Subject Record in D2RQ (1st try)

SUBJECT #300198841
Resource URL: http://localhost:2020/resource/AAT/Subject/300198841

<table>
<thead>
<tr>
<th>Property</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>skos:altLabel</td>
<td>rítons</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>ríton</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>ryton</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>ρυτόν</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>角杯</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>角状杯</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>角状杯</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>角沿杯</td>
</tr>
<tr>
<td>skos:altLabel</td>
<td>角杯</td>
</tr>
<tr>
<td>skos:prefLabel</td>
<td>rythka</td>
</tr>
<tr>
<td>gr:recordType</td>
<td>Concept</td>
</tr>
<tr>
<td>gr:facetCode</td>
<td>VTQ</td>
</tr>
<tr>
<td>is:gr:synonym</td>
<td>Greek: ρυτόν, Roman: ríton</td>
</tr>
<tr>
<td>is:gr:subjectId</td>
<td>300198841</td>
</tr>
<tr>
<td>is:skos:broadOf</td>
<td>300265900</td>
</tr>
</tbody>
</table>

The server is configured to display only a limited number of values (limit per property: 50).

Metadata


© 2016 J. Paul Getty Trust, author: Gregg Garcia. For educational purposes only. Do not distribute.
Issues Discovered Through Prototype Development

• How to represent lookup lists, associative relationships, N-ary relations, revision history

• Tool limitations – no functionality for getting language attributes from a database column

• Use RDBMS or triple store?

• Environment limitations – RDF uses much more space on filesystems

• Need for a semantic/ontology expert to help connect data to established standards

• What to do with lookup lists
Opportunities for Connecting Data - Coreferencing

- Language list made into its own hierarchy in AAT
- Languages coordinated with IANA Subtag Registry
- TGN place types connected to AAT subjects
- ULAN roles connected to AAT subjects
- ULAN nationalities connected to AAT subjects
- ULAN events connected to AAT subjects
- ULAN places connected to TGN subjects
Query Example

- Get all nations from TGN for a lookup list
Many links cannot be made automatically.

Matching ULAN Nationality to AAT

- Nationality/Culture/Race/Ethnicity in ULAN was conceptual, now needed to be linked to AAT.
- Certain matches must be resolved by hand.

<table>
<thead>
<tr>
<th>code</th>
<th>Nat/Culture ULAN</th>
<th>AAT subject_id</th>
<th>AAT term_id</th>
<th>AAT term</th>
</tr>
</thead>
<tbody>
<tr>
<td>901614</td>
<td>Languedocian</td>
<td>300020847</td>
<td>1000020847</td>
<td>Burgundian (Migration</td>
</tr>
<tr>
<td>901615</td>
<td>Burgundian (Migration period)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>901620</td>
<td>Parisian</td>
<td>300379855</td>
<td>1000540772</td>
<td>Gallic</td>
</tr>
<tr>
<td>901630</td>
<td>Gallic</td>
<td>300111192</td>
<td>100011192</td>
<td>German</td>
</tr>
<tr>
<td>901710</td>
<td>Swabian</td>
<td>300014076</td>
<td>100014077</td>
<td>hessian</td>
</tr>
<tr>
<td>901820</td>
<td>West German</td>
<td>300105675</td>
<td>1000105675</td>
<td>Thuringian</td>
</tr>
<tr>
<td>901830</td>
<td>Prussian</td>
<td>300379868</td>
<td>1000540441</td>
<td>Saxon</td>
</tr>
<tr>
<td>901840</td>
<td>Bavarian</td>
<td>300263656</td>
<td>1000263656</td>
<td>Wendish</td>
</tr>
<tr>
<td>901850</td>
<td>Hamburger</td>
<td>30019478</td>
<td>100019478</td>
<td>Frankish</td>
</tr>
<tr>
<td>901860</td>
<td>Hessian</td>
<td>30015912</td>
<td>100015912</td>
<td>Merovingian (culture)</td>
</tr>
<tr>
<td>901920</td>
<td>Merovingian (scripts)</td>
<td>300210134</td>
<td>1000210134</td>
<td>Merovingian (scripts)</td>
</tr>
<tr>
<td>901950</td>
<td>Carinthian</td>
<td>300385997</td>
<td>1000568044</td>
<td>Carinthian</td>
</tr>
</tbody>
</table>

This hessian is bad match, type of burlap.
Our vocabularies express as LOD can be used to connect other resources.
Triple Store vs. RDBMS

Triple Stores (GraphDB, 4store, Virtuoso)
• Built for storage and retrieval of triples
• Queried using SPARQL
• Import/export using RDF and SPARQL Update

Relational Database Management Systems (Oracle, SQL Server, MySQL)
• Data is organized into flat, two-dimensional tables containing fields
• Queried using SQL
• Import/export using DML statements along with whatever formats are supported by RDBMS such as comma-delimited files and XML
Tools for Transforming RDBMS Data into RDF Triples

• R2RML standard developed by the W3C for mapping RDBMS tables to RDF triples

• Use perl RDF-RDB2RDF module for implementing R2RML scripts

• Custom scripts for creating staged download files in N3, NT, RDF and JSON serializations

• Use OWLIM triple store and customized Forest application for HTTP server to handle user requests
Triple Extract Process

Vocabulary Coordination System (Oracle RDBMS) → R2RML + Perl → NT Files

OWLIM Triple Store → Sesame Console → vocab.getty.edu/sparql → Internet
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 Thesaurus (AAT)** ©
  Catherine wheel or rose window? AAT is a structured vocabulary, including terms, descriptions, and other information for generic concepts related to art and architecture.

- **The Getty Thesaurus of Geographic Names (TGN)** ©
  London or Londinium? TGN is a structured vocabulary, including names, descriptions, and other information for places important to art and architecture.

- **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.

Ontology Update Note: There are changes to the ontology and mapping, see Version 2.0 Changes and Version 3.0 Changes.

Documentation and Downloads:

- **GVP Semantic Representation**: HTML (for linking), PDF (for printing).
  Sections GVP URLs and Prefixes and Semantic Resolution describe the used URLs and provide examples. All prefixes that we use are in prefixes ttl. Comprehensive Sample Queries are included.

- **GVP ontology**: HTML, RDF/XML, Turtle

- **Datasets**: (NTriples Zip, see doc section Export Files before using!):
  - AAT: full.zip (all statements), explicit.zip (only explicit statements)
  - TGN: full.zip (all statements), explicit.zip (only explicit statements)
  - ULAN: full.zip (all statements), explicit.zip (only explicit statements)

- **VOID**: void ttl (see doc section Descriptive Information)

- **Associative Relationship Types**: Full Matrix by Codes PDF|Excel

- **General information about the Linked Open Data projects**

Explore the Dataset:

- **Text search**: Use 'AAT/TGN/ULAN' to select the vocabulary. Enter either text or concept ID in the query box. Multiple words in the text are 'AND'ed. Use 'Brief' to search only
Documentation and Downloads:

- **GVP Semantic Representation**: HTML (for linking), PDF (for printing). Sections GVP URLs and Prefixes and Semantic Resolution describe the used URIs and provide examples. All prefixes that we use are in prefixes.ttl. Comprehensive Sample Queries are included.
- **GVP ontology**: HTML, RDF/XML, Turtle
- **Datasets**: (NTriples Zip, see doc section Export Files before using!):
  - AAT: full.zip (all statements), explicit.zip (only explicit statements)
  - TGN: full.zip (all statements), explicit.zip (only explicit statements)
  - ULAN: full.zip (all statements), explicit.zip (only explicit statements)
- **VOID**: void.ttl (see doc section Descriptive Information)
- **Associative Relationship Types**: Full Matrix by Codes PDF|Excel
- **General information about the Linked Open Data projects**
Existing Standards (Ontologies) Used

- **Thesaurus information:** SKOS, SKOSXL, ISO 25964
- **Common properties:** Dublin Core (DC), Dublin Core Terms (DCT)
- **Sources and contributors:** Bibliographic Ontology (BIBO), Friends of a Friend (FOAF)
- **Geographic information:** W3C Geo Ontology (WGS)
- **Revision History:** Provenance (PROV)
- **System properties:** Resource Description Framework (RDF), RDF Schema (RDFS), Web Ontology Language (OWL), and XML Schema Definition (XSD)
- **Implementation of the conversion:** RDB to RDF Mapping Language (R2RML)

Link to the HTML version of the full documentation: [http://vocab.getty.edu/doc/](http://vocab.getty.edu/doc/)
Discussion Groups

Community Support Forum

We hope the community will use this getty-vocabulary forum to ask questions, discuss issues, and find solutions related to the technical aspects of this publication. Usage examples are especially welcome.

- Welcome to our community discussion... announcement
  By Getty Vocabularies LOD - 1 post - 75 views
  6/11/15

- Why ISO-8859-1 can't be used with content-negotiation encoding
  By Christopher Johnson - 7 posts - 18 views
  Apr 10

- large results download limits performance
  By Matthew Lincoln - 2 posts - 22 views
  Mar 8

- Getty vocabulary in Spanish
  By Sandor Kopacci - 13 posts - 69 views
  Feb 19

- Getting more data out of ULAN
  By Vladimir Alexiev - 3 posts - 50 views
  Jan 31

- ULAN Import Error Completed
  By Nabeel Ahmed - 2 posts - 18 views
  Jan 13

- How to get all Subjects of a Given Parent in SPARQL
  By Ykje - 3 posts - 23 views
  12/22/15

- Help with a query
  By Karan - 2 posts - 25 views
  12/22/15

- Wrongly merged biographies
  By Gabriel Kerneis - 2 posts - 13 views
  12/15/15
rhyta (drinking vessels, <vessels for serving and consuming food>, ... Furnishings and Equipment (Hierarchy Name))

Note: Refers to vessels from Ancient Greece, eastern Europe, or the Middle East that typically have a closed form with two openings, one at the top for filling and one at the base so that liquid could stream out. They are often in the shape of a horn or an animal’s head, and were typically used as a drinking cup or for pouring wine into another vessel. Drinking was done by holding the rhyton above the drinker’s head and catching the stream of liquid in the mouth.

Terms:
- rhyta (preferred, C,U,LC,English-P,D,U,PN)
  - Greek (transliterated)-P,D,U,PN
  - Spanish,UF,U,PN
- rhyton (C,U,English,AD,U,SN)
  - Greek (transliterated),AD,U,SN
  - Spanish,AD,U,SN
- Rhyton (C,U,English,AD,U,SN)
- rhytons (C,U,English,UF,U,N)
  - French-P,D,U,PN
  - Spanish-P,D,U,PN
- rhea (vessels) (C,U,English,UF,U,N)
- rheons (C,U,English,UF,U,N)
- rheon (C,U,English,UF,U,N)
- ἔρυθον (C,U,Ancient Greek,UF,U,U)
- rhūtōn (C,U,Ancient Greek (transliterated),UF,U,U)
### Full Text Search – Brief Results

Results for "yute" (6 of 6)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Term</th>
<th>Parents</th>
<th>Scope Note</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>aat:300375512</td>
<td>Hibiscus cannabinus (species)</td>
<td>Hibiscus (genus), Malvaceae (family), ... Agents Facet</td>
<td>Species of fast growing herbaceous annual plant, with stalks growing to 5.5 m (18 feet) in height. It is probably native to southern Asia, though its exact natural origin is unknown. It is widely d...</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300014244</td>
<td>jute board</td>
<td>fiberboard, fiber products, ... Materials Facet</td>
<td>A strong plyboard containing no jute fiber but made typically from sulfate and wastepaper pulp, used especially for shipping containers.</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300014045</td>
<td>jute (fiber)</td>
<td>plant fiber, natural fiber, ... Materials Facet</td>
<td>Bast fiber derived from either of two East Indian plants (Corchorus olitorius and C. capsularis). The pale brown fibers are soft, lustrous, and coarse, ranging in length from 4 to 10 feet. The brit...</td>
<td>Concept</td>
</tr>
<tr>
<td>tgn:2786554</td>
<td>Yutes Run</td>
<td>Allegheny, Pennsylvania, ... World</td>
<td></td>
<td>PhysPlaceConcept</td>
</tr>
<tr>
<td>tgn:7475035</td>
<td>Yutengping</td>
<td>Miaoli Xian, Chung-hua Min-kuo, ... World</td>
<td></td>
<td>AdminPlaceConcept</td>
</tr>
<tr>
<td>tgn:7415324</td>
<td>Yuteriaca, Rio</td>
<td>Oaxaca, México, ... World</td>
<td></td>
<td>PhysPlaceConcept</td>
</tr>
</tbody>
</table>
## Full Text Search – Full Results

### Results for "yute" (11 of 11)

<table>
<thead>
<tr>
<th>Subject</th>
<th>Term</th>
<th>Parents</th>
<th>Scope Note</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>aat:300375512</td>
<td>Hibiscus cannabinus  (species)</td>
<td>Hibiscus (genus), Malvaceae (family), ... Agents Facet</td>
<td>Species of fast growing herbaceous annual plant, with stalks growing to 5.5 m (18 feet) in height. It is probably native to southern Asia, though its exact natural origin is unknown. It is widely d...</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300014076</td>
<td>burlap</td>
<td>canvas, &lt;textile materials by process or technique&gt;, ... Materials Facet</td>
<td>Coarse canvas made of jute, used mainly for sacks and wrapping.</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300014244</td>
<td>jute board</td>
<td>fiberboard, fiber products, ... Materials Facet</td>
<td>A strong plywood containing no jute fiber but made typically from sulfate and wastepaper pulp, used especially for shipping containers.</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300253474</td>
<td>buckram</td>
<td>&lt;textile materials by finishing process&gt;, &lt;textile materials by process or technique&gt;, ... Materials Facet</td>
<td>Heavy weave cotton, jute, or linen textile stiffened with glue, size, or starch and used for interlinings in garments, box making, bookbinding, etc.</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300015203</td>
<td>oakum</td>
<td>caulking compound, sealing compound, ... Materials Facet</td>
<td>Loosely twisted hemp or jute fiber impregnated with tar or a tar derivative and used in caulking seams, as of wooden ships, and packing joints, as of pipes.</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300014078</td>
<td>canvas</td>
<td>&lt;textile materials by process or technique&gt;, textile materials, ... Materials Facet</td>
<td>Closely woven textile made in various weights, usually of flax, hemp, jute, or cotton, used especially for sails, tarpaulins, awnings, upholstery, and as a support for oil painting. Also used for a...</td>
<td>Concept</td>
</tr>
<tr>
<td>aat:300131081</td>
<td>oilcloth</td>
<td>&lt;textile materials by finishing process&gt;, &lt;textile materials by process or technique&gt;, ... Materials Facet</td>
<td>Textile of woven cotton, jute, or hemp, treated with oil and pigment and used as a waterproof covering.</td>
<td>Concept</td>
</tr>
</tbody>
</table>
### 2.6 Full Text Search Query

This is the query used for the Full Text Search.

```sparql
PREFIX gvp: <http://vocab.getty.edu/gvp/>
PREFIX foaf: <http://xmlns.com/foaf/0.1/>
PREFIX schema: <http://schema.org/>
PREFIX dct: <http://purl.org/dc/terms/>
PREFIX gvp_lang: <http://vocab.getty.edu/aat/gvp_lang/>
PREFIX skos: <http://www.w3.org/2004/02/skos/core#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#

  ?subject luc:term "fishing" AND vessel" a ?type.
  FILTER (?type != gvp:subject)
  OPTIONAL {?subject gvp:placeTypePreferred [gvp:prefLabelGVP [xml:literalForm ?type1]]}
  OPTIONAL {?subject gvp:agentTypePreferred [gvp:prefLabelGVP [xml:literalForm ?type2]]}
  OPTIONAL {?subject gvp:prefLabelGVP [xml:literalForm ?term]}
  OPTIONAL {?subject gvp:parentStringAbbrev ?parents}
  OPTIONAL {?subject foaf:focus gvp:biographyPreferred/schema:description ?descr}
  OPTIONAL {?subject skos:scopeNote [dct:language gvp_lang:en; rdfs:value ?scopeNote]}
}
```

- If the user selected Brief, we use predicate luc:term (just before the red text), for Full, we use predicate luc:term
- If the user selected only one of the vocabularies (e.g. AAT), we add a clause like

```sparql
?subject skos:inScheme "aat".
```

The following result columns are included:
- Subject
- GVP preferred Term
- Abbreviated Parent string

Two field pairs that the FTS concatenates (if present):
- Description (ULAN one-line biography) and
What the artists of these works of art have in common?

Robert Macpherson
The Campagna near Rome

Joseph Heintz, the elder
The Toilette of Venus

Jean-Honoré Fragonard
Ruins of an Imperial Palace, Rome
All three were active in Italy but were not Italian.

**SPARQL Query**

```
select ?x ?name ?bio ?birth {
  {select distinct ?x
   {?x foaf:focus/bio:event/(schema:location | (schema:location/gvp:broaderExtended)) tgn:1000080-place}
   ?x gvp:prefLabelGVP/xl:literalForm ?name;
   foaf:focus/gvp:biographyPreferred [schema:description ?bio;
   gvp:estStart ?birth].
  filter ("1550"^^xsd:gYear <= ?birth && ?birth <= "1900"^^xsd:gYear)
  filter exists {?x gvp:broaderExtended ?facet.
    filter (?facet in (ulan:500000002))
  filter not exists {
      ?x foaf:focus/(schema:nationality | (schema:nationality/gvp:broaderExtended)) aat:300111198}
}
```

- Just one of the ways to get interesting results using all three vocabularies.
- The full URIs are:
  - Italy: [http://vocab.getty.edu/tgn/1000080-place](http://vocab.getty.edu/tgn/1000080-place)
  - Person, Artists facet: [http://vocab.getty.edu/ulan/500000002](http://vocab.getty.edu/ulan/500000002)
  - Italian (culture or style): [http://vocab.getty.edu/aat/300111198](http://vocab.getty.edu/aat/300111198)

© 2016 J. Paul Getty Trust, author: Joan Cobb. For educational purposes only. Do not distribute.
Query Examples/Exercises

- Keyword phrase search with stemming
- Combination term and scope note search
- Expanded search using term and scope note
- Exact match on term
- Get count of subjects having a Chinese term
- Get count of subjects having a Chinese term in AAT
Registered with Databh
http://datahub.io/organization/the-getty-trust

Open Data Commons Attribution License (ODC-By) v1.0
http://opendatacommons.org/licenses/by/1-0/

© 2016 J. Paul Getty Trust, author: Joan Cobb. For educational purposes only. Do not distribute.
Available on the web (whatever format) but with an open licence, to be Open Data

Available as machine-readable structured data (e.g. excel instead of image scan of a table)

Available as (2) plus non-proprietary format (e.g. CSV instead of excel)

All the above plus, Use open standards from W3C (RDF and SPARQL) to identify things, so that people can point at your stuff

All the above, plus: Link your data to other people’s data to provide context
Some Useful Resources and Links

- **Tim Berners-Lee TED Talk**
  - [http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html](http://www.ted.com/talks/tim_berners_lee_the_year_open_data_went_worldwide.html)

- **Europeana Video**
  - Sometimes a picture is worth a thousand words – in this case, it’s a video
  - [http://vimeo.com/36752317](http://vimeo.com/36752317)

- **Eero Hyvönen’s book**
  - *Publishing and Using Cultural Heritage Linked Data on the Semantic Web*

- **Linked Open Data in Libraries Archives and Museums (LODLAM)**
  - [http://lodlam.net/](http://lodlam.net/)

- **Open Memory Project**
  - [http://summit2015.lodlam.net/2015/04/21/challenge-entry-open-memory-project](http://summit2015.lodlam.net/2015/04/21/challenge-entry-open-memory-project)
Links to the live demo

Developer SPARQL Endpoint
http://vocab.getty.edu

General Information about our LOD project
http://www.getty.edu/research/tools/vocabularies/lo/index.html

About the Getty Vocabularies in general
http://www.getty.edu/research/tools/vocabularies/index.html

Public Forum
https://groups.google.com/forum/#!forum/gettyvocablod

Examples of ID 300198841
Human readable: http://vocab.getty.edu/page/aat/300198841
Machine readable: http://vocab.getty.edu/aat/300198841