GVP LOD: ONTOLOGIES AND SEMANTIC REPRESENTATION

Vladimir Alexiev, Data and Ontology Group, Ontotext Corp

CIDOC Congress, Dresden, Germany
2014-09-05: International Terminology Working Group
2014-09-09: Getty special session

full version (HTML), short version (PDF, slideshare).

Press O for overview, H for help.
Proudly made in plain text with reveal.js, org-reveal, org-mode and emacs.
ONTOTEXT SCOPE OF WORK

http://vocab.getty.edu

• Ontology development: http://vocab.getty.edu/ontology
• Contribution to ISO 25964 ontology (latest thesauri standard)
• Complete mapping specification
• Help with R2RML conversion scripts, contrib to RDB2RDF (Perl), rrx:languageColumn extension
• GraphDB (OWLIM) repository. Enterprise Edition (clustered)
• Sem app dev (customized Forest UI), tech consulting
• SPARQL 1.1 endpoint: http://vocab.getty.edu/sparql
• Documentation (100 pages): http://vocab.getty.edu/doc
• Lots of sample queries, incl charts, geographic, etc
• Per-entity export files, explicit/total data dumps
• Help desk / support
• Presentations, scientific papers
**SEMANTIC RESOLUTION & CONTENT NEGOTIATION**

All GVP, AAT and TGN URLs resolve, returning human or machine readable content through content negotiation (303 redirect). Eg about the ontology:

<table>
<thead>
<tr>
<th>URL</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://vocab.getty.edu/ontology">http://vocab.getty.edu/ontology</a></td>
<td>semantic URI, content-negotiated</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/ontology.html">http://vocab.getty.edu/ontology.html</a></td>
<td>page (application/xhtml+xml)</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/ontology.rdf">http://vocab.getty.edu/ontology.rdf</a></td>
<td>application/rdf+xml</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/ontology.ttl">http://vocab.getty.edu/ontology.ttl</a></td>
<td>text/turtle</td>
</tr>
</tbody>
</table>

Eg about an AAT subject:

<table>
<thead>
<tr>
<th>URL</th>
<th>Format</th>
</tr>
</thead>
<tbody>
<tr>
<td><a href="http://vocab.getty.edu/aat/300011154">http://vocab.getty.edu/aat/300011154</a></td>
<td>semantic URI, con-neg</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/aat/300011154.html">http://vocab.getty.edu/aat/300011154.html</a></td>
<td>page (application/xhtml+xml)</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/aat/300011154.rdf">http://vocab.getty.edu/aat/300011154.rdf</a></td>
<td>application/rdf+xml</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/aat/300011154.ttl">http://vocab.getty.edu/aat/300011154.ttl</a></td>
<td>text/turtle</td>
</tr>
<tr>
<td><a href="http://vocab.getty.edu/aat/300011154.nt">http://vocab.getty.edu/aat/300011154.nt</a></td>
<td>NTriples</td>
</tr>
</tbody>
</table>
GVP VOCABULARY DATA

Scope includes:

- Subjects: Concepts but also non-concepts
- Obsolete subjects (and dct:isReplacedBy)
- Terms: plain (SKOS) & rich (SKOS-XL). Term characteristics
- Languages (some custom lang tags)
- Hierarchical rels: custom & standard, distinguish BTG,BTP,BTI
- Associative rels (170 subprop of skos:related)
- Historic info on rels (rdf:Statement) and terms
- Alignment (exactMatch to LCSH)
- Sources (bibo:Document, bibo:DocumentPart with locator)
- Contributors (foaf:Agent)
- Revision history (prov:Activity)
- Thesaurus-specific (TGN place types, coordinates)

One of the richest thesauri I've seen
<table>
<thead>
<tr>
<th>Prefix</th>
<th>Ontology</th>
<th>Used for</th>
</tr>
</thead>
<tbody>
<tr>
<td>bibo:</td>
<td>Bibliography Ontology</td>
<td>Sources</td>
</tr>
<tr>
<td>dc:</td>
<td>Dublin Core Elements</td>
<td>common</td>
</tr>
<tr>
<td>dct:</td>
<td>Dublin Core Terms</td>
<td>common</td>
</tr>
<tr>
<td>foaf:</td>
<td>Friend of a Friend ontology</td>
<td>Contributors</td>
</tr>
<tr>
<td>iso:</td>
<td>ISO 25946 (latest on thesauri)</td>
<td>iso:ThesaurusArray, BTG/BTP/BTI</td>
</tr>
<tr>
<td>owl:</td>
<td>Web Ontology Language</td>
<td>Basic RDF representation</td>
</tr>
<tr>
<td>prov:</td>
<td>Provenance Ontology</td>
<td>Revision history</td>
</tr>
<tr>
<td>rdf:</td>
<td>Resource Description Framework</td>
<td>Basic RDF representation</td>
</tr>
<tr>
<td>rdfs:</td>
<td>RDF Schema</td>
<td>Basic RDF representation</td>
</tr>
<tr>
<td>schema:</td>
<td>Schema.org</td>
<td>common, geo (TGN)</td>
</tr>
<tr>
<td>skos:</td>
<td>Simple Knowledge Org System</td>
<td>Basic vocabulary representation</td>
</tr>
<tr>
<td>skosxl:</td>
<td>SKOS Extension for Labels</td>
<td>Rich labels</td>
</tr>
<tr>
<td>wgs:</td>
<td>W3C World Geodetic Survey geo</td>
<td>Geo (TGN)</td>
</tr>
<tr>
<td>xsd:</td>
<td>XML Schema Datatypes</td>
<td>Basic RDF representation</td>
</tr>
</tbody>
</table>
HIERARCHICAL RELATIONS

Use iso:ThesaurusArray to allow Guide Terms below Concepts. Infer cross-threading SKOS/ISO broader relations.
### Key Values (Flags) Are Important

Excel-driven Ontology Generation™ (getty-codes.xls to getty-codes.ttl)

Key values can be mapped to Custom sub-class, Custom (sub-)prop, Ontology Value (eg <term/kind/Abbreviation>)

<table>
<thead>
<tr>
<th>vocabulary</th>
<th>field</th>
<th>val</th>
<th>ObjectProperty</th>
<th>Class</th>
<th>label</th>
<th>domain</th>
<th>range</th>
<th>subProperty</th>
<th>subClassOf</th>
<th>ConceptSchema</th>
</tr>
</thead>
<tbody>
<tr>
<td>subject</td>
<td>record_type</td>
<td>F</td>
<td></td>
<td>gvp:Facet</td>
<td>Facet</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT subject</td>
<td>record_type</td>
<td>H</td>
<td></td>
<td>gvp:Hierarchy</td>
<td>Hierarchy Name</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT subject</td>
<td>record_type</td>
<td>C</td>
<td></td>
<td>gvp:Concept</td>
<td>Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subject</td>
<td>record_type</td>
<td>-</td>
<td></td>
<td>gvp:ObsoleteSubject</td>
<td>Obsolete Subject</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGN subject</td>
<td>record_type</td>
<td>P</td>
<td></td>
<td>gvp:PhysPlaceConcept</td>
<td>Physical Place Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGN subject</td>
<td>record_type</td>
<td>A</td>
<td></td>
<td>gvp:AdminPlaceConcept</td>
<td>Administrative Place Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TGN subject</td>
<td>record_type</td>
<td>B</td>
<td></td>
<td>gvp:PhysAdminPlaceConcept</td>
<td>Physical and Administrative Place Concept</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>subject_rols</td>
<td>preferred</td>
<td>P</td>
<td></td>
<td>gvp:broderPreferred</td>
<td>Preferred Parent</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:broder</td>
<td>gvp:broder</td>
<td></td>
</tr>
<tr>
<td>subject_rols</td>
<td>preferred</td>
<td>N</td>
<td></td>
<td>gvp:broderNonPreferred</td>
<td>Non-PREFERRED Parent</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:broder</td>
<td>gvp:broder</td>
<td></td>
</tr>
<tr>
<td>subject_rols</td>
<td>preferred</td>
<td>G</td>
<td></td>
<td>gvp:broderGeneric</td>
<td>Generic</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:broder</td>
<td>gvp:broder</td>
<td></td>
</tr>
<tr>
<td>subject_rols</td>
<td>preferred</td>
<td>P</td>
<td></td>
<td>gvp:broderPart</td>
<td>Parent</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:broder</td>
<td>gvp:broder</td>
<td></td>
</tr>
<tr>
<td>subject_rols</td>
<td>preferred</td>
<td>I</td>
<td></td>
<td>gvp:broderInheritance</td>
<td>Parent</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:broder</td>
<td>gvp:broder</td>
<td></td>
</tr>
<tr>
<td>term</td>
<td>preferred</td>
<td>P</td>
<td></td>
<td>gvp:PrefLabelGVP</td>
<td>Preferred Label</td>
<td>gvp:Subject</td>
<td>gvp:Subject</td>
<td>gvp:PrefLabel</td>
<td></td>
<td></td>
</tr>
<tr>
<td>term</td>
<td>name</td>
<td>V</td>
<td>gvp:termFlag</td>
<td>Term Flag</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termFlag/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>term</td>
<td>other_flags</td>
<td>C</td>
<td>gvp:termKind</td>
<td>Term Kind</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termKind/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT term</td>
<td>other_flags</td>
<td>C</td>
<td>gvp:termKind</td>
<td>Term Kind</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termKind/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT term</td>
<td>other_flags</td>
<td>F</td>
<td>gvp:termKind</td>
<td>Term Kind</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termKind/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT term</td>
<td>other_flags</td>
<td>N</td>
<td>gvp:termKind</td>
<td>Term Kind</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termKind/</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AAT term</td>
<td>other_flags</td>
<td>S</td>
<td>gvp:termKind</td>
<td>Term Kind</td>
<td>skos:Label</td>
<td>skos:Concept</td>
<td>termKind/</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| gvp:Facet a owl:Class ;
| rdfs:isDefinedBy <http://vocab.getty.edu/ontology> ;
| rdfs:subClassOf gvp:Subject, iso:ThesaurusArray ;
| rdfs:label "Facet" ;
| rdfs:comment "One of the major divisions of a vocabulary" ;
| skos:example "Objects Facet (AAT), World (TGN)" ;
| dct:description "One of the major divisions of a vocabulary.\nExample: Objects Facet (AAT), World (TGN)." ;
ASSOCIATIVE RELATIONS ARE VALUABLE

More Excel-driven Ontology Generation™ (assoc-rels.xls to assoc-rels.ttl)

- Relations come in owl:inverseOf pairs (or owl:SymmetricProperty self-inverse)

<table>
<thead>
<tr>
<th>code</th>
<th>icode</th>
<th>domain (C1)</th>
<th>LOD rel</th>
<th>range (C2)</th>
<th>LOD rel</th>
<th>Editor rel - From C1 to C2</th>
<th>Editor rel - From C2 to C1</th>
<th>example</th>
<th>example</th>
</tr>
</thead>
<tbody>
<tr>
<td>2000</td>
<td></td>
<td>any</td>
<td>related to</td>
<td>any</td>
<td>any - related to - any</td>
<td>same</td>
<td>gulf red is related to light red (pigment)</td>
<td>light red (pigment) is related to gulf red</td>
<td></td>
</tr>
<tr>
<td>2001</td>
<td></td>
<td>any</td>
<td>formerly referred to</td>
<td>any</td>
<td>any - formerly referred - any</td>
<td>same</td>
<td>gigues formerly referred to tiddles</td>
<td>tiddles formerly referred to gigues</td>
<td></td>
</tr>
<tr>
<td>2100</td>
<td></td>
<td>any</td>
<td>distinguished from</td>
<td>any</td>
<td>any - distinguished from - any</td>
<td>same</td>
<td>historic farms are distinguished from abandoned farms; naive art is distinguished from outsider art</td>
<td>abandoned farms are distinguished from historic farms; outsider art is distinguished from naive art</td>
<td></td>
</tr>
</tbody>
</table>

```sparql
gvp:aat2000_related_to a owl:ObjectProperty;
  rdfs:subPropertyOf skos:related;
  rdfs:domain skos:Concept; rdfs:range skos:Concept;
  # domain "any"; range "any";
  dc:identifier "2000";
  skos:prefLabel "aat2000_related_to";
  dc:title "related to - any";
  skos:example "gulf red is related to light red (pigment)";
  skos:scopeNote "generic relationship, not explained";
  dct:description """"any - related to - any; generic relationship, not explained. Example: gulf red is related to light red (pigment)"" .
gvp:aat2000_related_to a owl:SymmetricProperty.
```
 OBSOLETE SUBJECTS

- AAT obsolete subjects are 4.4% of valid subjects, which shows a good rate of editorial actions
- Obsolete subjects may have been used in client data. In order not to leave such data hanging, we publish minimal information:

```reasonml
aat:300123456 a gvp:ObsoleteSubject; # Was made non-publishable
  skos:prefLabel "Made up subject";
  skos:inScheme aat: ;
  schema:endDate "2012-12-31T12:34:56"^^xsd:dateTime.

aat:300386746 a gvp:ObsoleteSubject; # Was merged to a dominant Subject
  skos:prefLabel "Buncheong";
  skos:inScheme aat: ;
  dct:isReplacedBy aat:300018699; # Punch'ong
  schema:endDate "2012-12-31T12:34:56"^^xsd:dateTime.
```
USE OF ISO:THESAURUSARRAY IN GVP

Use rdf:List for ordered children.
Novelty: if parent is Concept, use anonymous ThesaurusArray
CONTRIBUTION TO ISO 25946

- Contributed to ISO 25946 ontology (LOV entry)
- First industrial use of ISO 25946
- Defined appropriate combinations of BTG, BTP, BTI relations (first formally defined in ISO).


<table>
<thead>
<tr>
<th></th>
<th>BTGx</th>
<th>BTPx</th>
<th>BTIx</th>
</tr>
</thead>
<tbody>
<tr>
<td>BTGx</td>
<td></td>
<td>BTGE</td>
<td>BTPE</td>
</tr>
<tr>
<td>BTPx</td>
<td></td>
<td>BTPE</td>
<td>BTPE</td>
</tr>
<tr>
<td>BTIx</td>
<td></td>
<td>BTIE</td>
<td>no</td>
</tr>
</tbody>
</table>

- Eg: anvil components BTP <anvils and anvil accessories> BTG <forging and metal-shaping tools> => BTPE
- Mount Athos BTI Orthodox religious center BTG religious center => BTIE
TERMS

Support multilingual labels: both SKOS (plain)...

```plaintext
aat:300198841 a skos:Concept, gvp:Subject, gvp:Concept;
skos:prefLabel "rhyta"@el-latn, "rhyta"@en, "rhytons"@es, "rhytons"@fr;
skos:altLabel "rhyta"@es, "rhyton"@es, "rhyton"@en, "rhyton"@el-latn ...

skosxl:prefLabel aat_term:1000198841-en, aat_term:1000198841-el-Latn ...

skosxl:altLabel aat_term:1000198841-es, aat_term:1000297235-en ...
```

... and rich info in SKOS-XL:

```plaintext
aat_term:1000198841-en a skosxl:Label;
dc:identifier "1000198841";
dct:language aat:300388277, gvp_lang:en; # owl:sameAs
dct:contributor aat_contrib:10000000, aat_contrib:10000131;

skosxl:literalForm "rhyta"@en; # with Qualifier if applicable
gvp:term "rhyta"@en; # no qualifier

gvp:displayOrder "1"^^xsd:positiveInteger;
gvp:termType <term/type/Descriptor>; # Descr/AltDescr/UseFor
gvp:termPOS <term/POS/PluralNoun>; # Part of Speech
gvp:contributorPreferred aat_contrib:10000000, aat_contrib:10000088;
gvp:contributorNonPreferred aat_contrib:10000131;
gvp:sourcePreferred aat_source:2000051089-term-1000198841 ...;
dct:source aat_source:2000024811, aat_source:2000052946 ...;
gvp:sourceNonPreferred aat_source:2000052946;
```
SOME CUSTOM LANGUAGE TAGS

Despite the richness of IANA tags (9000), we had to define new tags, using several extension mechanisms:

- Private language, e.g.
  - x-byzantin-Latn: Byzantine Greek (transliterated)
  - x-frisian (IANA/ISO has codes for predecessor Old Frisian and dialects West, Saterland and North Frisian)
- Private language used in specific region, e.g.
  - qqq-002: African language (not specified which)
  - qqq-ET: Ethiopian (not specified: Boro/Borna, Karo…)
- Private modifier, e.g.
  - grc-Latn- x-liturgic: Liturgical Greek
  - ber-Latn- x-dialect: Berber Dialects (transliterated)
  - fa-Latn- x-middle: Persian, Middle (transliterated)
  - zh-Latn-pinyin- x-notone: transliterated Pinyin without tones
SOURCES

bibo:Document or bibo:DocumentPart

```
aat_source:2000051089 a bibo:Document;
  dc:identifier "2000051089"
  bibo:shortTitle "AATA database (2002-)";
  dct:title "Getty Conservation Institute (GCI). database of AATA Online...".

aat_source:2000051089-term-1000198841 a bibo:DocumentPart;
  dct:isPartOf aat_source:2000051089;
  bibo:locator "128257 checked 26 January 2012".
```

Applied to subject, term, scopeNote:

```
aat:300198841 # subject (rhyta)
  dct:source aat_source:2000030301-subject-300198841;
  dct:source aat_source:2000052378.

aat_term:1000198841-en # term "rhyta"@en
  gvp:sourceNonPreferred aat_source:2000049728;

aat_scopeNote:34904 # scopeNote
  dct:source aat_source:2000046502.
```
CONTRIBUTORS

foaf:Agent

aat_contrib:10000131 a foaf:Agent;
  dc:identifier "10000131";
  foaf:nick "CDBP-DIBAM";
  foaf:name "Centro de Documentación de Bienes Patrimoniales...".

Applied to subject, term, scopeNote:

aat:300198841 # subject "rhyta"
  dct:contributor aat_contrib:10000131;
aat_term:1000198841-en # term "rhyta"@en
  gvp:contributorNonPreferred aat_contrib:10000131;
  gvp:contributorPreferred aat_contrib:10000000.
aat_scopeNote:34904 # scopeNote
HISTORIC INFO

Includes dates of applicability, historicFlag, comment. Applied to terms; hier & assoc rels, place types (using rdf:Statement)

aat_term:1000002693-en a skosxl:Label;
skosxl:literalForm "lambruscatura"@en ;
gvp:historicFlag <http://vocab.getty.edu/historic/historic> ;
schema:startDate "0900"^^xsd:gYear ;
schema:endDate "1700"^^xsd:gYear ;
rdfs:comment "Medieval term for wainscoting".

aat_rel:300020271-aat2812_followed-300020269 a rdf:Statement;
rdf:subject aat:300020271;        # Second Dynasty (Egyptian)
rdf:predicate gvp:aat2812_followed;
rdf:object aat:300020269;        # First Dynasty (Egyptian)
rdfs:comment "Second Dynasty began ca. 2775 BCE";
schema:startDate "-2785"^^xsd:gYear;
schema:endDate "-2765"^^xsd:gYear.

tgn:7011179-placeType-300008347 a rdf:Statement;
rdf:subject tgn:7011179;          # Siena
rdf:predicate gvp:placeTypePreferred;
rdf:object aat:300008347;        # inhabited place
rdfs:comment "settled by Etruscans (flourished 6th century BCE)";
schema:startDate "-0800"^^xsd:gYear;
gvp:displayOrder "1"^^xsd:positiveInteger.
PROV is too complex, so we simplify:

aat:300018699
  prov:wasGeneratedBy aat_rev:12345;
  dct:created  "2014-01-02T01:02:03"^^xsd:dateTime;
  dct:modified  "2014-01-03T01:02:03"^^xsd:dateTime;
  dct:issued   "2014-01-04T01:02:03"^^xsd:dateTime.

aat_rev:12345 a prov:Activity, prov:Create;
  dc:type "created";
  prov:startedAtTime "2014-01-02T01:02:03"^^xsd:dateTime.

aat_rev:12346 a prov:Activity, prov:Modify;
  prov:used aat:300018699;
  dc:type "term added";
  dc:description "leggings, puttee (1000248060)";
  prov:startedAtTime "2014-01-03T01:02:03"^^xsd:dateTime.

aat_rev:12347 a prov:Activity, prov:Publish;
  prov:used aat:300018699;
  dc:type "issued";
  prov:startedAtTime "2014-01-04T01:02:03"^^xsd:dateTime.
TGN SPECIFICS: CONCEPT-PLACE DUALITY

Duality between Concept and its denotation (ala VIAF, UK BL, FR BnF, SE KB…)

![Image of diagram showing the relationship between concept and place entities]
TGN SEMANTIC REPRESENTATION

Place types (TGN->AAT), Concept-Place duality, coordinates
CONSTRUCT QUERY: GET & CACHE ALL DATA FOR SUBJECT

- All data for these subsidiary objects is served by the resource URL
- Cached, thus served quickly
- Served in RDF/XML, N3/Turtle, NTriples, JSON, soon JSON-LD
SAMPLE QUERY: BAR CHART WITH SPARQL

Number of UN members per year. See doc or jsfiddle with it.
THANKS FOR YOUR TIME!

mailto:vladimir.alexiev@ontotext.com