Linked.Art & Vocabularies: Linked Open Usable Data

Rob Sanderson, David Newbury
Semantic Architect, Software & Data Architect
J. Paul Getty Trust

rsanderson@getty.edu, dnewbury@getty.edu
@azaroth42, @workergnome
What is RDF?

“RDF is a standard model for data interchange on the Web.”

- A way for computers to work with facts
- A way to express statements about resources
- A W3C (web-based, web-friendly) data standard

Subject -> Predicate -> Object
What is Linked Data?

“The term Linked Data refers to a set of best practices for publishing structured data on the Web.”

- Some rules that help make RDF useful for others
- Every ID is a HTTP URL, and it returns info about itself
- You should use other people’s URLs whenever you can

Subject (URI) -> Predicate (URI) -> Object (URI)
What are Ontologies?

“The formal naming and definition of the types, properties and interrelationships of the entities that exist for a particular domain.”

• The application of the LOD framework to a domain
• Ontologies provide meaning to data
• Ontologies are most useful when used by multiple organizations and datasets
Cultural Heritage Ontologies?

“The CIDOC-CRM provides definitions and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation.”

- An event-based way to look at the world
- Mature ontology in Museum domain
- "Semantic glue" between different information sources
- Highly theoretical
What are Vocabularies?

“A controlled vocabulary is an organized arrangement of words and phrases used to index content and/or retrieve content through browsing or searching.”

• Shared terminology to reduce ambiguity
• In LOD, vocabularies provide reusable identities for shared entities across datasets
• Vocabularies are hubs in the web of data, allowing easy interconnections between systems
Linked Art

https://linked.art/
Linked Art Model

A Linked Open Data model, designed to work across many museums and to enable functional applications.

- **Homogenized** model across many institutions
- Supports varying levels of completeness
- Enables “round-tripping” of data
- Framework: Linked Open Data
- Ontology: CIDOC-CRM (plus a little)
- Vocabularies: AAT, ULAN, TGN
- Focused on Usability
Linked Art Model

Linked Art provides patterns and models that enable cultural heritage institutions to easily publish their data for use by both event-based digital research projects and for non-specialist developers.

90% of the use cases of 90% of the organizations, with 10% of the complexity
Usable vs Complete

Usability vs Completeness

Usable
Complete

Usability

Completeness

GETTY CONSERVATION INSTITUTE + GETTY FOUNDATION + GETTY RESEARCH INSTITUTE + J. PAUL GETTY MUSEUM
Incremental Complexity

Usability

Completeness

IIIF: Interoperability, Linked Art & Vocabularies

@azaroth42
@workergnome
rsanderson; dnewbury
@getty.edu
Successful Data Models...

- Solve actual challenges, documented as use cases
  - Using data that is captured and available
- Allow consistent description of shared use cases
  - Allow for addition of further information
- Can be productively used
  - Via easy-to-implement services
  - With easy-to-implement applications
- Provide interoperability with other data
- Are clearly documented with relevant examples
Successful Data Models...

Are developed ...

• **Iteratively**
  • We will not get it right first time
• **Responsively**
  • We will adapt it in response to feedback
• **Responsibly**
  • We will consider changes/features carefully
• **Collaboratively**
  • We will engage with the community and stakeholders
Successful Data Models...

**Focus** on the primary resources of concern for the domain

Core Classes:
- Object
- Activity
- Person
- Place
- Time
Successful Data Models...

Are Consistent and Unsurprising

The more consistency across the model, the less you need to learn and remember.

Common Patterns designed and used:

- Consistent naming scheme
- Classification of Specifics
- Partitioning
- Statements about Resources
Pattern: Classification of Specifics

We use AAT to provide more specific classifications of resources than are possible using CIDOC-CRM

- Object: Painting vs Sculpture
- Identifier: Accession Number vs Database Id
- Place: Country vs City
- Group: Museum vs Art Dealer
- Activity: Engraving vs Casting
Pattern: Partitioning

Many types of resource can be partitioned, creating identity for further description of more granular parts

- Objects: Parts, Features with different dimensions etc
- Activities: Exhibition vs Venue, Auction vs Lot
- Locations: Country, Region, City, District, Building
- Texts: Set, Volume, Chapter, Page
- Organizations: Institution, Program, Department
Pattern: Statements about Resources

Resources can have **statements with classifications from AAT** for the topic, co-existing with machine readable data.

- Materials
- Provenance
- Attribution
- Biography
- Description
Focus: People and Places

AAT provides a rich list of identities for classification. ULAN and TGN provide identities for actors and locations.

Fill different needs within the model:
• Classification important for shared semantics
• Person/Place important for shared entities

Provides a hub in the web of data for connecting across organizations, without prior agreement.
Vocabularies and CIDOC-CRM?

New Challenge: Vocabularies use different ontologies!

AAT: Minimal impact - labels
ULAN, TGN: Significant impact - entire entity model

How do we ensure consistency and usability of the Vocabularies in the Linked Open Data ecosystem?
APIs & Profiles & Formats & Applications
API: Agreement Preceding Interaction*

The same data can be exposed in different ways to different audiences simultaneously.

- Need to understand **how to access** the data
- APIs can provide different functionality
- Fewer ways with more users is more sustainable
- Web provides HTTP protocol for data transport

(* API is really: Application Programming Interface)
Data Profiles

Need the shape of the data to be **consistent** to be easy to use. Different audiences thus need different shapes.

• Current profile (SKOS-XL, custom GVP) for Researchers
• Add Linked Art profile for Museum audience
• Add schema.org profile for Web/SEO audience

Different representations or descriptions of the same entity, not different entities.
Developer-Friendly Format

```json
{
  "@context": "https://lod.museum/ns/context/1/full.jsonld",
  "id": "https://lod.museum/example/object/1",
  "type": "ManMadeObject",
  "classified_as": "aat:300033618", # by reference
  "label": "Example Painting",
  "made_of": {
    "id": "aat:300015045", # by (minimal) value
    "type": "Material",
    "label": "watercolor"
  }
}
```
Applications

Developers that can rely on consistent format, shape and interactions, can build robust and innovative applications.

• Need many such applications to validate usability
• Notifications of updates to remain synchronized
• Notifications of contributions to dataset
  • ... via reference to external entities?

Translations provide easy internationalization for applications - a huge benefit to the community
IIIF: Interoperability
Linked Art & Vocabularies

Arches

Graph Designer

Node List
Find a node...

Panel:
- Payment
- Amount
- Node
- Label
- Payer
- Currency
- Payee
- value

Panel:
- Node Name
- CRM Class
- Data type

Make card
Data from nodes not collected in other cards will be collected in the root card's form section

Welcome, admin
Search
Help
Notification Infrastructure

- Archives Space
- Provenance
- DOR
- Conserv’n
- Publication

Vocabs

Notificat’n

Harvesting and Indexing Platform

GraphDB

Linked Art & Vocabularies

@azaroth42
@workergnome
rsanderson; dhewbury
@getty.edu
Summary

- Linked Open Data!
- Vocabularies are Identity Providers
- Linked Art – usability focused data model
  - Uses CIDOC-CRM and Vocabularies
- Technical improvements suggested for improving consistency and functionality
Thank You!

Rob Sanderson, David Newbury
rsanderson@getty.edu, dnewbury@getty.edu
@azaroth42, @workergnome
Discuss!