



SIMILE Demo

Objectives, Status and Future Plans

Ryan Lee <ryanlee@w3.org>
Stefano Mazzocchi <stefanom@mit.edu>

WWW2004 Conference
May 22, New York, USA



Objectives

SIMILE Goals

- Make semantic interoperability of metadata a reality for digital libraries by:
 - providing reusable software for browsing, searching and mapping heterogenous metadata
 - using semantic web technologies
 - identifying issues, gaps and best practices

SIMILE Vision

- tools should help humans focus on their abilities, amplifying, not replacing them!
- metadata quality is a function of heterogeneity
- serendipitous discovery is a value that should not get lost
- empower recombinant metadata

SIMILE Participants

- MIT Libraries (MacKenzie Smith)
- MIT CSAIL (David Karger)
- HP Labs (Mick Bass)
- W₃C (Eric Miller)

Status

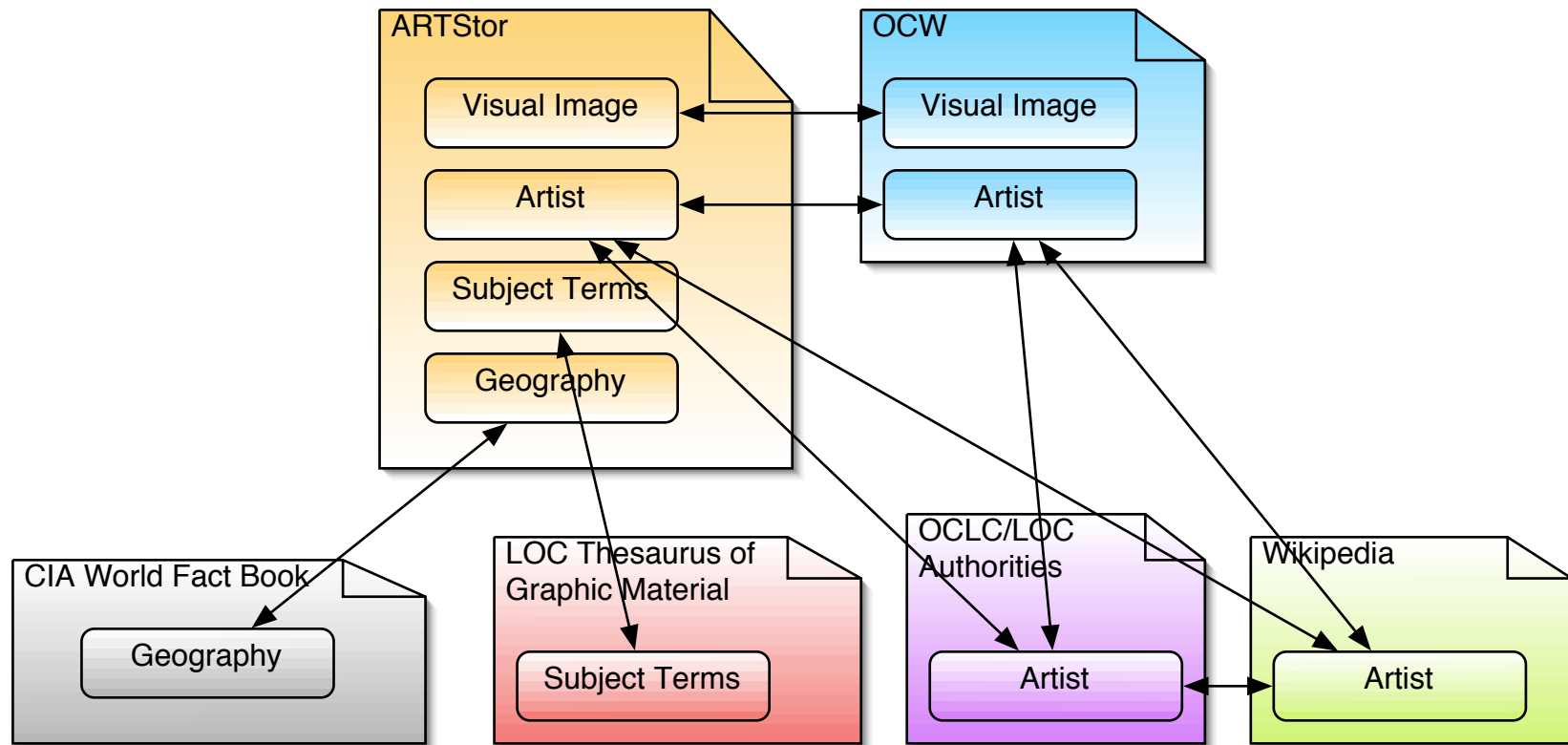
- faceted metadata browser
 - aimed at end users
 - goal is to show max functionality with min complexity (maximize usability)

- RDF browser
 - aimed at semantic web specialists
 - goal is to enable simple browsing of complex models

Datasets

- ARTStor: works of art
- MIT OpenCourseWare: MIT courses that discuss visual images
- Wikipedia: bibliographic background on artists
- CIA World Fact Book (in progress): geographic information

Datasets Map



Schemata

- Dublin Core
- VRA
- LOM
- SKOS
- SKOS extension
- SIMILE's own glue ones

ARTStor Example

```
@prefix vra:      <http://web.mit.edu/simile/www/2003/10/vraCore3#> .
@prefix art:      <http://web.mit.edu/simile/www/2003/10/artstor#> .
@prefix vc:       <http://www.w3.org/2001/vcard-rdf/3.0#> .
@prefix person:   <http://web.mit.edu/simile/www/2003/10/person#> .
@prefix dc:       <http://purl.org/dc/elements/1.1/> .
@prefix rdfs:     <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:      <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix :         <#> .
```

[...]

```
art:ControlledTerm
  a rdfs:Class ;
  rdfs:label "Controlled Term"@en .
```

```
art:Collection
  rdfs:subClassOf art:ControlledTerm ;
  rdfs:label "An element from the controlled list of
collections"@en .
```



MIT OpenCourseWare

```
@prefix rdfs:    <http://www.w3.org/2000/01/rdf-schema#> .
@prefix rdf:    <http://www.w3.org/1999/02/22-rdf-syntax-ns#> .
@prefix lomEdu: <http://www.imsproject.org/rdf/
imsmd_educationalv1p2#> .
@prefix ocw:    <http://web.mit.edu/simile/www/2004/01/ocw#> .
@prefix dc:    <http://purl.org/dc/elements/1.1/> .
@prefix dcq:   <http://dublincore.org/2000/03/13/dcq#> .
@prefix :     <#> .
```

[...]

ocw:Lecture

```
rdfs:subClassOf lomEdu:LearningResourceType ;
rdfs:label "Lecture"@en .
```

ocw:Bibliography

```
rdfs:subClassOf lomEdu:LearningResourceType ;
rdfs:label "Bibliography"@en .
```



OWL Mappings

Using String Distance Metrics (Levenshtein)

```
<http://web.mit.edu/simile/www/metadata/ocw/  
Contributor#picasso,_pablo,1881-1973>  
  rdf:type disp:PreferredTerm ;  
  owl:sameAs <http://web.mit.edu/simile/www/metadata/artstor/  
Subject#picasso,_pablo,1881-1973> ;
```

.

```
<http://web.mit.edu/simile/www/metadata/artstor/  
site#1new_york,_metropolitan_museum_of_art>  
  rdf:type disp:PreferredTerm ;  
  owl:sameAs <http://web.mit.edu/simile/www/metadata/artstor/  
site#new_york,_metropolitan_museum_of_art> ;
```

Achieved Results

- Usable implementation of both Longwell and Knowle as web-based general purpose RDF browsers
- Passed the 0.5 Megatriples wall (100k original, 400k inferred)
- Successful use of XSLT2 as XML->RDF bridge
- Use of the Levenshtein distance on literals to evaluate potential mappings between datasets

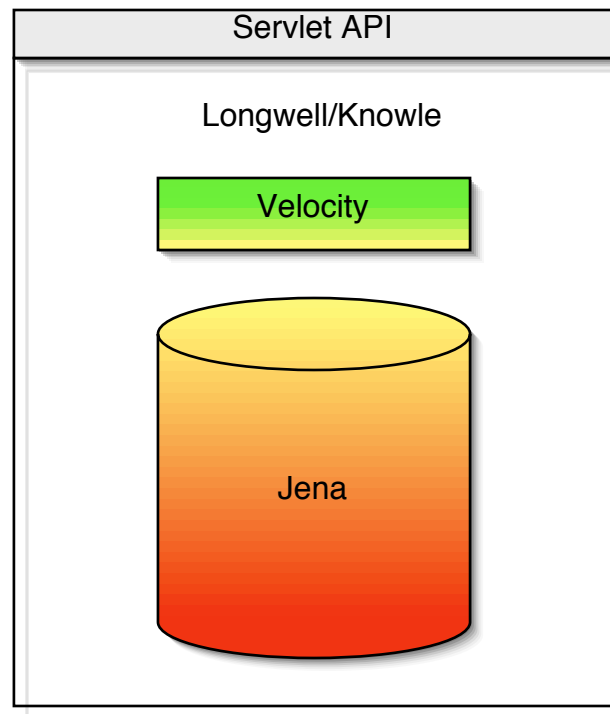
Simile
سیمیله

Demo



Here we go!

Current Architecture





Open Source!
Available Today!

<http://simile.mit.edu/longwell/>



Open Questions

Scalability

- How more complex can the model grow before saturating our computational capacity?
- How can we design a distributed architecture and still be fast enough to be useable?

Connectivity

- How can we increase the connectivity when merging models with reasonable costs and without compromising perceived metadata quality?

Provenance

- How should provenance influence the reasoning on aggregated models?

Evolution

- How can we deal with the evolution of models and their impact on previous inferenced interpretations?
- Can time be another provenance or we need a different dimension?

Disagreement

- How well can the semantic web model cope with disagreement?
- How do we distinguish disagreement from mistakes?

Simile
سیمیله

Thanks!



Simile
سِمْيَلَة

Q&A

