

SIMILE Demonstrator Storyboard

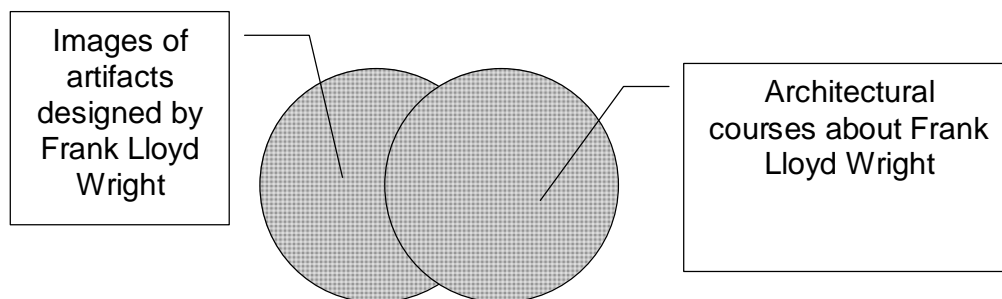
Goals:

1. Demonstrate the practicality and utility of Semantic Web techniques – specifically inferencing search through schema mapping rules – to the problem of information retrieval in the content domain of learning objects and visual images – specifically fine arts images.
2. Frame and illustrate the longer -term SIMILE research agenda.

Introduction to the corpora

In order for this story to be compelling, the user must be interacting with a corpus that is aggregated from different communities where the interest of the user must overlap with the interests of both communities. Therefore the materials in these communities must be related, though that fact may not have been apparent *a priori*. The user is likely to be interested in that subset of the materials from both communities which are related to each other, through her stated interests.

For example, the user's might be a design student working on a thesis about the relationship between architectural styles and design. One particular architect of interest might be Frank Lloyd Wright. Materials of interest could be drawn either photographic images of design work by Frank Lloyd Wright, or from learning objects discussing his work.



So the first requirement for our test corpus is that ***it must contain materials sourced and described by different communities yet which intersect in a common domain of interest.***

We will assume and focus further that the two corpora are photographic images of designed artifacts and course materials. The photograph images are described by high - quality VRA-Core metadata whereas course materials are described by IMS metadata. For actual examples of sample metadata from these vocabularies, see the appendix.

Introduction to the scenario

Who is the user?

Three possible situations:

- The user is a researcher searching for visual images in support of their research
“I want to write a book about Frank Lloyd Wright and his impact on modern design ”
- The user is a learner searching for resources to learn about a particular topic
“Who is Frank Lloyd Wright? Where can I find materials to write a term paper about him?”
- The user is an educator looking to reuse and/or create course materials
“My course is about 20th Century Design. Frank Lloyd Wright is a key figure I need to discuss”

Why are they interacting with the system?

- The user is aware the library is a likely source of images or learning objects that will support their work.

What are they trying to find?

1. Images of works
 - about a particular subject
 - from a particular part of the world
 - created using a particular technique or medium
 - created by a particular individual
 - that demonstrate a particular concept
2. Course materials
 - about a particular subject
 - from a particular part of the world
 - created using a particular technique or medium
 - created by a particular individual
 - that demonstrate a particular concept
 - aimed at certain audience

How important is it that they find it?

- Individuals searching in support of research are probably quite prepared to invest substantial energy in finding materials that can truly be supportive of their research.
- Individuals searching in support of course creation probably have a much lower investment threshold for finding appropriate materials.

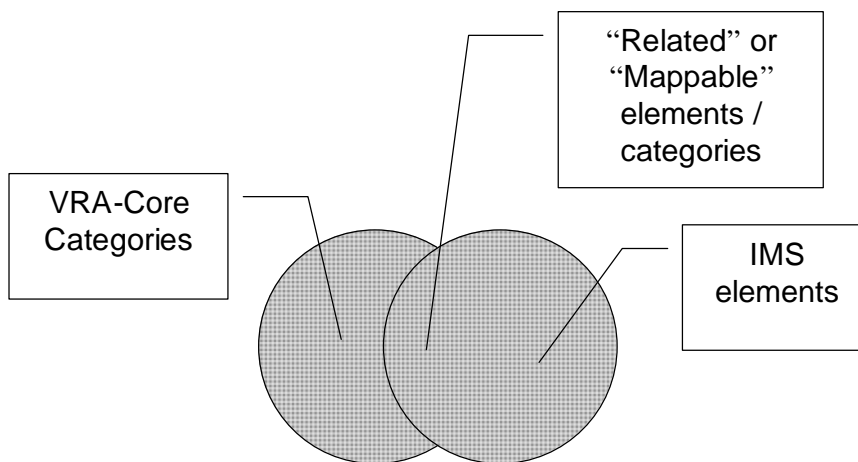
Where do they start?

- Individuals searching in support of their research:
 - Browsing: they start with collections, collection overviews, try to match their needs to collections, then look for the right instances within those collections.
 - Using free text search: they may be able to express their needs to other humans, but unable to express using schema/vocab constraints because they (a) don't know which vocabs exist and may be appropriate to use or (b) they don't understand the vocab well enough to map their needs to the vocab.

- Using templated search: they may start with a rather specific definition of their needs, for example expressed in a VRA -core template.
- Individuals searching in support of course creation:
 - Probably start with materials already used in courses. Again, they locate these via browsing, using free text search or using templated search as appropriate.
 - If they cannot find finished courses, they may start to assemble their own course material. Here they will use browsing, free text search and templated search also.
- What problems do they encounter?
 - If they cast their search widely it is hard to get sufficient expressivity in the search to achieve acceptable precision.
 - If they target their search across a particular corpus, they may not become aware of, or miss potentially useful and relevant materials.

Key Issues

If materials from these two corpuses intersect in a domain of interest, how can that intersection be expressed in terms of relationships between the elements / categories of these two metadata sets? The aim of demonstrator is to improve search results through introduction of schema synonyms and/or subproperty relationships.



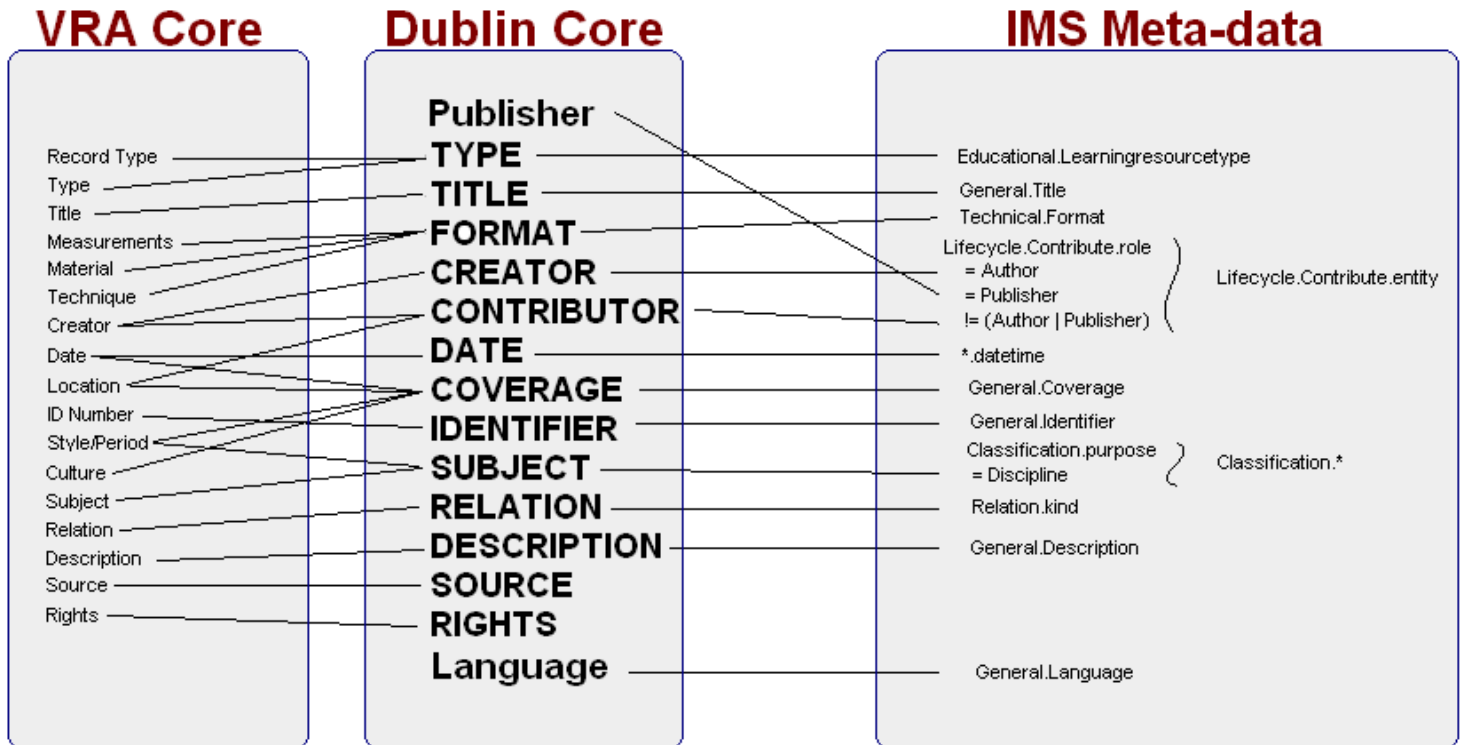
Example Script for Demonstrator

1. Start with two islands of metadata e.g. VRA and IMS.
2. Search across one set (e.g. VRA), show hits purely from the collection in this format.
3. Introduce maps between the metadata vocabularies e.g. creator/author, subject
4. Search across both sets, show more useful hits, including some from island2

Message is "Maps are metadata: all you've done is add more metadata, and now you have more functionality"

Mapping between vocabularies

Both VRA Core and IMS provide mappings to Dublin Core which are summarised in the following diagram.



However, as the example metadata in the appendix shows, the mapping is actually a bit more problematic. For example consider the following situations:

1. "I want to find out about Frank Lloyd Wright"

There are two ways of doing this query - either do a free text search on FLW (Frank Lloyd Wright) or do a templated search where FLW is the creator. In order to illustrate the issues involved, here are the relevant fields in each of the examples in the appendix:

Example 1:

```
Creator.Personal Name=Wright, Frank L. (1867 -1959)
Location.CurrentRepository =Scottsdale (AZ, US), Frank Lloyd Wright
Foundation, Taliesin West
Rights=Frank Lloyd Wright Foundation, Taliesin West, Scottsdale, AZ US
```

Example 2:

```
general.title = Frank Lloyd Wright
general.description = A collection of images of buildings by Frank
Lloyd Wright arranged in chronological order prepared for a course at
Boston College.
```

```
educational.description = The learning object has several collections
of photographs of buildings by Frank Lloyd Wright arranged in
chronological order and grouped together by common subject. It is
intended to support lectures about his work, giving students the
opportunity to view a variety of examples of his work.
classification.taxonpath.taxon.entry = 20th Century \Architect\Frank
Lloyd Wright
```

Example 3:

```
general.title = Frank Lloyd Wright
general.description = An introduction to Frank Lloyd Wright in the
Library of Congress America's Story Exhibition
educational.description = The learning object explains who Frank Lloyd
Wright was, in the context of recent American History. It has a
portrait photograph and links to further resources.
classification.taxonpath.taxon.entry = Amazing Americans \Writers &
Artists\Frank Lloyd Wright
```

The free text search does not require the use of schema based inference to map between different vocabularies. It will successfully return all three records, assuming it knows that “Wright, Frank L.” is equivalent to “Frank Lloyd Wright”. However the templated search is more problematic as assuming it uses a mapping based on the previous diagram, it will infer that `vra:creator` maps onto `ims:lifecycle.contribute`. However in the IMS examples, FLW is not listed in `ims:lifecycle.contribute` but is listed in the `ims:general.title`, `ims:general.description` and `ims.taxonpath` classification of both learning objects. So the demonstration needs to consider how to perform this mapping.

2. *“I want to find out more information about 20th century designers?”*

This query is even hard just for free text search. Consider the relevant fields in each of the examples:

Example 1:

```
Creator.Personal Name = Wright, Frank L. (1867 -1959)
Creator.Role = designer
```

Example 2:

```
general.keyword = architecture, chicago school, modernist, prarie style,
20th Century
general.coverage = 1886 -1997;US
classification.taxonpath.taxon.entry = 20th Century \Architect\Frank
Lloyd Wright
classification.keyword = architecture, chicago school, modernist,
prarie style, 20th Century
```

Example 3:

```
general.keyword = architecture, chicago school, modernist, prairie style,
20th Century, interior design, furniture design
general.coverage = June 8, 1867 - April 9, 1959;US
classification.keyword = architecture, chicago school, modernist,
prairie style, 20th century, interior design, furniture design
```

So it is quite possible such a search would return no records. If the word “designers” is stemmed to “design”, then it will return Example 3. If the system knows can infer from the lifetime information in Creator.Personal Name in Example 1 that Frank Lloyd Wright worked in the 20th century, then it could return Example 1 although this is non-trivial.

Demonstrator stage 2

The second demonstrator will demonstrate the network effects of multiple pairwise maps. Here we add island 3, with maps to island1 (for example we add CIDOC with existing map to VRA) and search using island2 terms, get hits from island 1 and island 3.

Most schemas can of course have some relationships with DC as well.
variants: controlled vocabularies, subject folding, index folding

Comments on scale: a compelling demo would include 3 collections of ~40K records each.

Example VRA Metadata: Example 1

(Example metadata taken from VRACore 3.0 Specification,
<http://www.vraweb.org/vracore3.htm>)

Record Type = work
Type = architectural furniture
Type = seating furniture
Type = dining chair
Type = tall back chair
Type = spindle-back chair
Title = Frederick C. Robie House dining chair
Measurements.Dimensions = 52.5 x 18 x 19.25 cm
Material.Medium = oak
Material.Medium = leather
Technique = cabinet making
Technique = upholstering
Creator.Personal Name = Wright, Frank L. (1867 -1959)
Creator.Role = designer
Date.Design = 1906
Date.Completion = 1909
Location.Current Repository = Chicago (IL,USA),University of Chicago,David & Alfred Smart Museum of Art
Location.Former Site = Frederick C. Robie House, Chicago, IL, US
ID Number.Current Repository = 1965.2.14fu rn
Style/Period = Arts and Crafts
Culture = American
Relation.Part of = Frederic C. Robie House
Description = The dining chair is part of a set of six designed specifically for the dining room of the Frederick C. Robie House.
Rights = David & Alfred Smart Museum of Art, University of Chicago, IL, US

Record Type = work
Type = photograph
Type = gelatin silver print
Title = interior view of Frederic C. Robie House dinning room with furnishings
Measurements.Dimensions = 8x10"
Material.Medium = gelatin
Material.Medium = silver
Material.Support = photo paper
Technique = photography
Technique = gelatin silver process
Creator.Personal name = Fuermann, Henry
Creator.Role = photographer
Date.Creation = 1910
Location.Current Repository = Scottsdale (AZ, US), Frank Lloyd Wright Foundation, Taliesin West
ID Number.Current Repository = 0908.018
Culture = American
Subject = Frank C. Robie House
Subject = dining room
Subject = dining table
Subject = dining chair

Subject = stained glass window
Rights = Frank Lloyd Wright Foundation, Taliesin West, Scottsdale, AZ,
US

Record Type = image
Type = slide
Title = interior view of Frederick C. Robie House dining room with
furnishings
Measurements.Dimensions = 2x2"
Measurements.Format = 35 mm
Measurements.Format = horizontal
Technique = photography
Creator = Mole, Christopher
Creator.Role = copy photographer
Date.Creation = 1985
Location.Current Repository = Albuquerque (NM, US), University of New
Mexico, Bainbridge Bunting Slide Library
ID Number.Current Repository = UNM d000614
ID Number.Current Repository = FURN/AMER/Wright/Robie.383787
Source = gift of Christopher Mole
Rights = Frank Lloyd Wright Foundation, Taliesin West, Scottsdale, AZ,
US

Example IMS Metadata: Example 2

```
general.identifier =
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/FLW.html
general.title = Frank Lloyd Wright
general.catalogentry.catalog = URI
general.catalogentry.entry =
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/FLW.html
general.language = en-US
general.description = A collection of images of buildings by Frank
Lloyd Wright arranged in chronological order prepared for a course at
Boston College.
general.keyword = architecture, chicago school, modernist, prarie style,
20th Century
general.coverage = 1886 -1997;US
general.structure = collection
general.aggregationlevel = 3
lifecycle.version = 1.0
lifecycle.status = Final
lifecycle.contribute.role = Contributor
lifecycle.contribute.entity.vcard
lifecycle.contribute.entity.vcard.fn = Jeffrey Howe
lifecycle.contribute.entity.vcard.org = Boston College
lifecycle.contribute.entity.vcard,email = howejbc@bc.edu
lifecycle.contribute.date = 1998
metametadata.identifer =
http://www.simile.org/examples#franklloydwright001
metametadata.contribute.role = Creator
metametadata.contribute.entity = Mark Butler
metametadata.contribute.role = Validator
metametadata.contribute.entity = John Gilbert
metametadata.contribute.date = 15/09/2003
metametadata.metadatascheme = LOMv1.0
metametadata.language = en-US
technical.format = txt/html
technical.size = 8940 bytes
technical.location =
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/FLW.html
technical.requirement.type = browser
technical.requirement.name = Internet Explorer
technical.requirement.type = browser
technical.requirement.name = Netscape
technical.installationremarks
technical.otherplatformrequirements = Internet connection required.
educational.interactivitytype = Expositive
educational.learningresourcetype = collection
educational.interactivitylevel = low
educational.semanticdensity = medium
educational.intendedenduserrole = Le arner, Teacher, Author
educational.context = Higher Education
educational.typicalagerange = 18 -99
educational.difficulty = low
educational.typicallearningtime = 00:02:00
educational.description = The learning object has several collections
of photographs of buildings by Frank Lloyd Wright arranged in
chronological order and grouped together by common subject. It is
```

intended to support lectures about his work, giving students the opportunity to view a variety of examples of his work.
educational.language = en-US
rights.cost
rights.copyrightandotherrestrictions = See
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/copyrite.html
rights.description = Copyright Boston College, Mass, USA
relation.kind = PartOf
relation.resource.identifier =
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/default.html
relation.resource.description = A digital archive of American Architecture
relation.resource.catalogentry.catalog = URI
relation.resource.catalogentry.entry =
http://www.bc.edu/bc_org/avp/cas/fnart/fa267/default.html
annotation.person
annotation.date
annotation.description
classification.purpose = Discipline
classification.taxonpath.source = A Digital Archive of American Architecture
classification.taxonpath.taxon.id
classification.taxonpath.taxon.entry = 20th Century \Architect\Frank Lloyd Wright
classification.description = A digital archive of American Architecture
classification.keyword = architecture, chicago school, modernist, prarie style, 20th Century

Example IMS Metadata: Example 3

```
general.identifier = http://www.americaslibrary.gov/cgi -
bin/page.cgi/aa/writers/wright
general.title = Frank Lloyd Wright
general.catalogentry.catalog = URI
general.catalogentry.entry = http://www.americaslibrary.gov/cgi -
bin/page.cgi/aa/writers/wright
geeral.catalogentry.catalog = Librar y Of Congress
general.catalogentry.entry = America's Story
general.language = en-US
general.description = An introduction to Frank Lloyd Wright in the
Library of Congress America's Story Exhibition
general.keyword = architecture, chicago school, modernist, prarie style,
20th Century, interior design, furniture design
general.coverage = June 8, 1867 - April 9, 1959;US
general.structure = linear
general.aggregationlevel = 2
lifecycle.version = 1.0
lifecycle.status = Final
lifecycle.contribute.role = Publisher
lifecycle.contribute.entity = Library of Congress
lifecycle.contribute.role = Content Provider
lifecycle.contribute.entity = Al Ravenna
lifecycle.contribute.date = 1954
metametadata.identifer =
http://www.simile.org/examples#franklloydwright002
metametadata.contribute.role = Creator
metametadata.contribute.entity = Mark Butler
metametadata.contribute.role = Validator
metametadata.contribute.entity = John Gilbert
metametadata.contribute.date = 15/09/2003
metametadata.metadatascheme = LOMv1.0
metametadata.language = en-US
technical.format = txt/html
technical.size
technical.location = http://www.americaslibrary.gov/cgi -
bin/page.cgi/aa/writers/wright
technical.requirement.type = browser
technical.requirement.name = Internet Explorer
technical.requirement.type = browser
technical.requirement.name = Netscape
technical.installationremarks =
technical.otherplatformrequirements = Internet connection required.
educational.interactivitytype = Expositive
educational.learningresourcetype = Narrative text
educational.interactivitylevel = low
educational.semanticdensity = low
educational.intendedenduserrole = Learner, Teacher, Author
educational.context = Primary Education, Secondary Education
educational.typicalagerange = Suitable for children over 9
educational.difficulty = low
educational.typicallearningtime = 00:00:10
educational.description = The learning object explains who Frank Lloyd
Wright was, in the context of recent American History. It has a
portrait photograph and links to further resources.
```

educational.language = en-US
rights.cost
rights.copyrightandotherrestrictions =
<http://www.americaslibrary.gov/about/copyright.html>
rights.description = Copyright Library of Congress, USA
relation.kind = IsPartOf
relation.resource.identifier = <http://www.americaslibrary.gov/cgi-bin/page.cgi>
relation.resource.description = America's Story
relation.resource.catalogentry.catalog = URI
relation.resource.catalogentry.entry =
<http://www.americaslibrary.gov/cgi-bin/page.cgi>
annotation.person
annotation.date
annotation.description
classification.purpose = Idea
classification.taxonpath.source = America's Story, Library of Congress
classification.taxonpath.taxon.id
classification.taxonpath.taxon.entry = Amazing Americans \Writers &
Artists\Frank Lloyd Wright
classification.description = America's Story
classification.keyword = architecture, chicago school, modernist,
prarie style, 20th Century, interior design, furniture design