Title:	**Deep Image Annotation: Making a Difference in Knowledge Organization**
Presented to:	International Society for Knowledge Organization, Fourth Biennial Conference, London July 13th 2015
Author:	**David Clarke**
Affiliation:	Synaptica (www.synaptica.com)
Date:	20150707

ABSTRACT

Visual images are a powerful medium for communicating ideas and information, and they provide a valuable complement to textual content. A vast amount of information resides inside photographs, paintings, diagrams and drawings, which is comprehensible to the human eye but relatively inaccessible to machine queries. Well established techniques exist to support searching and browsing images based on the metadata that has been applied to whole images, but search and browse access to specific features within images is a relatively immature field.

Section two of this paper will briefly examine a number of current tools and standards for annotating images at the sub-image level.

Section three will present a detailed methodology for deep image annotation that has been co-developed by the author and a team of high-definition imagery and Linked Data practitioners. The methodology is based upon three key principles: (i) images must be rendered as multi-resolution pyramids to enable ultra high definition images to be explored via the web; (ii) all image and sub-image metadata must be expressed as Linked Data; and (iii) sub-image visual features, which may be single point coordinates or bound areas, must be directly addressable via HTTP-URIs.

Sections four and five will discuss how Knowledge Organization Systems (KOS) can support search and browse access to the informational content deep within images. Among the methods discussed will be: using hierarchies to decompose and organize the spatial or thematic structure of an image; classifying and indexing visual features to the subjects and named entities they represent; linking from visual features to concepts and from concepts back to visual features; supporting faceted search and graph-based queries.

CONTENTS

1 Introduction

1.1 The Informational Value of Pictorial Content

Visual images are a powerful medium for communicating ideas and information, and they provide a valuable complement to textual content. A vast amount of information resides inside photographs, paintings, diagrams and drawings which is comprehensible to the human eye but relatively inaccessible to machine queries.

> 'Access to image-based resources is fundamental to research, scholarship and the transmission of cultural knowledge. Digital images are a container for much of the information content in the Web-based delivery of images, books, newspapers, manuscripts, maps, scrolls, single sheet collections, and archival materials.'
>
> *International Image Interoperability Framework* (IIIF) Website[1]

1.2 The Inaccessibility of Sub-Image Details

Well established techniques exist to support searching and browsing images based on metadata applied to the whole image, but search and browse access to specific points or regions within images is a relatively immature field.

Image-level metadata may be sufficient for information access to the majority of images, but image-level metadata is insufficient to support access to large or complex images. Examples from art, science and current affairs are provided.

Figure 1: *The Garden of Earthly Delights*, Hieronymus Bosch, c. 1503[2]

Figure 1 portrays *The Garden of Earthly Delights* by Hieronymus Bosch, which is an allegorical work of art with dense figurative detail. The painting's narrative is built up from many separate scenes, each telling a story that is rich in allusion and symbolism. Metadata applied to the whole image does not enable the viewer to locate or interpret the various different scenes and figures in the work.

[1] http://iiif.io/about.html accessed 20150707

[2] https://en.wikipedia.org/?title=The_Garden_of_Earthly_Delights accessed 20150707

Figure 2: *Full Moon*, Gregory H. Revera, 2010[3]

Figure 2 presents an image of the Earth's moon. Metadata applied to the whole image does not enable the viewer to identify topographic features or the location of the Apollo landing sites.

Figure 3: Leaders of the G8, 2013[4]

Figure 3 presents a photograph of ten international leaders at the G8 Summit in Lough Erne Northern Ireland on 18th June 2013. Image level metadata can

[3] https://en.wikipedia.org/wiki/File:FullMoon2010.jpg accessed 20150707

[4] https://en.wikipedia.org/wiki/G7_(forum)#/media/File:Ten_leaders_at_G8_summit,_2013.jpg accessed 20150707

describe the names and titles of the people in the photograph, but not match the specific names to each individual.

1.3 The Value Proposition for Sub-Image Annotation

In all three of the examples, while a certain level of information access can be supported by image-level metadata, a much richer knowledge discovery experience can be provided if specific visual details are individually identified.

An analogy with information access to physical books illustrates the value proposition for sub-image annotation.

In a physical library, card catalogues and bibliographic databases provide a means to identify shelves of books and individual books. Tables of contents and subject indexes then facilitate a deeper level of information access to specific pages, sections and paragraphs within a book.

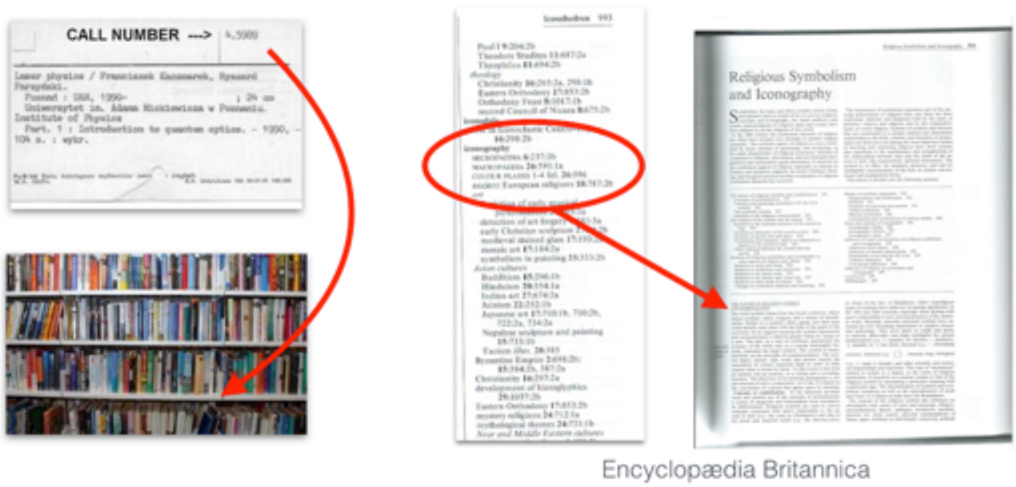


Figure 4: Illustration of Access to Books versus their Interior Content

Image-level metadata may be compared with card catalogues and bibliographic databases. They take the user to a discrete work but cannot take the user inside the work to discover its interior content. Sub-image annotation may be compared with tables of contents and subject indexes. They allow the user to search and to navigate inside images.

2 Image Annotation Tools and Standards

2.1 Tools

There are numerous proprietary and open source tools that allow images to be annotated with single-point and/or bound-area markers plus text labels and descriptions. Examples include:

Annotorious: http://annotorious.github.io

IIPImage: http://iipimage.sourceforge.net

MarQueed: https://www.marqueed.com

SZoter: http://www.szoter.com

Zoomify: http://www.zoomify.com

Some image sharing applications, such as Flickr (https://www.flickr.com) have also incorporated image annotation features.

2.2 Standards

An international cross-domain standard for image annotation does not exist, but there are several initiatives that aim to standardize image metadata and annotations within particular domains.

Medical and Clinical:

AIM: https://wiki.nci.nih.gov/display/AIM/Annotation+and+Image+Markup+-+AIM

DICOM: http://dicom.nema.org

Academic and Research:

ImageClef: http://www.imageclef.org

National and Research Libraries:

IIIF: http://iiif.io

Additionally, the W3C has created a general data model in RDF for adding annotations to content in any media including image-based content:

Open Annotation Data Model: http://www.openannotation.org/spec/core/

3 OASIS Deep Image Annotation Methodology

This paper describes a novel approach to image annotation that builds upon existing annotation tools, standards and methods and combines them with tools and methods from the disciplines of controlled vocabularies (KOS), semantic indexing and Linked Data.

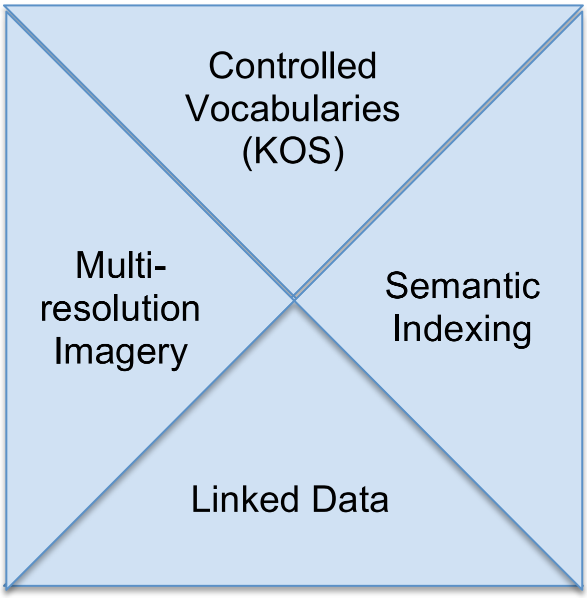


Figure 5: The convergence of four disciplines

The convergence of these disciplines was pursued in an R&D project under the name OASIS: Open Annotation Semantic Indexing System.

The OASIS R&D team determined several high-level design guidelines:

1. The system should be able to process images of any size using multi-resolution imagery
2. All image and annotation metadata should be stored as Linked Data in an RDF graph database
3. Each individual annotation should be assigned a unique HTTP-URI
4. Images and annotations should be semantically indexed using controlled vocabularies (KOS)

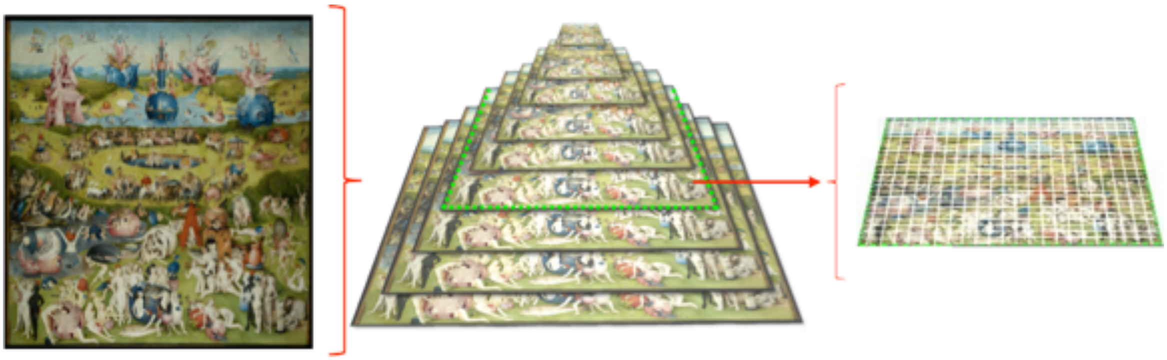
The core standard chosen for the high-level data model is the W3C Open Annotation Data Model[5]. The Open Annotation specification was selected to support Linked Data and also because the model provides extensibility options for the future annotation of content in other media types.

[5] http://www.openannotation.org/spec/core/

3.1 Multi-resolution Imagery

Very large and complex images can be Gigabytes in size, making it impractical to deliver them over the web as discrete image files. Multi-resolution imaging technology solves this problem by converting a large discrete source image into a pyramidal stack of images rendered at different resolutions.

Each resolution is then split into a mosaic of small image tiles that can be stitched together based on their relative coordinates.



Source image

Pyramidal stack of multi-resolution images

Image tiles at a particular resolution

Figure 6: Illustration of multi-resolution image processing

Multi-resolution imaging technology supports pan-and-zoom browsing of images of any size over the web, including on mobile devices.

3.2 Image and Annotation Metadata as Linked Data

Image and annotation metadata and associated indexing terminology is stored as Linked Data in an RDF graph database, providing several strategic advantages:

1. **Data portability** is achieved by using an open standards data model
2. **Data standardisation** is promoted by the ability to adopt ontologies and predicates from external authorities such as Dublin Core, CIDOC, FOAF, GeoNames, OWL, PROV and SKOS
3. **Data reuse** is promoted by the ability to adopt and integrate externally curated controlled vocabularies and data sets such as LC, Getty, GeoNames and DBpedia
4. **Data analytics** is facilitated by the ability to construct pattern-based queries of the graph database
5. **Data sharing** is enabled by the option to publish a SPARQL endpoint

Linked Data enables globally distributed data sets to be connected to work together as a coherent database. Federated SPARQL queries gather and merge results from distributed data that includes the native graph database of image and annotation data, and may also include live queries to external SPARQL endpoints and to internally cached copies of reference data sets.

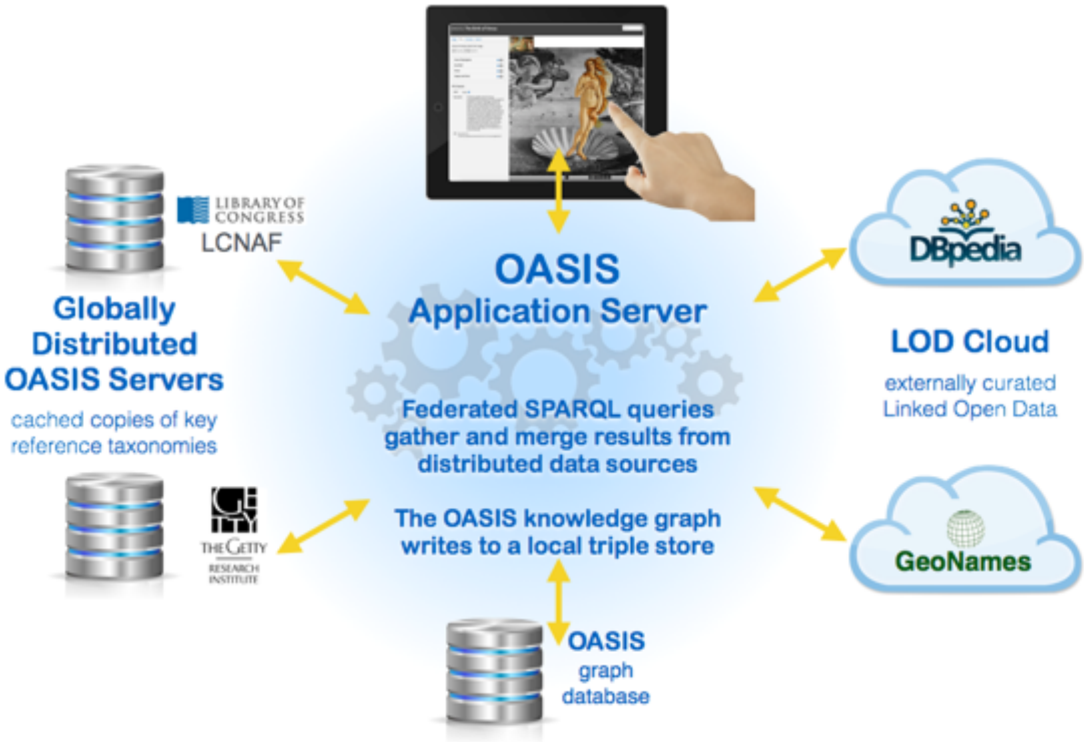


Figure 7: Linked Data unifies globally distributed data sets

3.3 Annotations are Assigned Unique HTTP-URIs

The assignment of unique identifiers to individual annotations is an essential enabler for sub-image indexing and information access. In the illustration below a region of an image of the Mona Lisa is defined as *The smile of the Mona Lisa* and assigned its own unique HTTP-URI.

Figure 8: A region is annotated as *The smile of the Mona Lisa* and assigned an HTTP-URI

An extensible set of metadata can then be created for each unique annotation including the coordinates of the point or bound area, the level zoom at which the feature should be viewed, a label and descriptive text.

3.4 Images and Annotations may be Semantically Indexed

Both whole images and sub-image annotations may then be *semantically indexed* using *controlled vocabulary terminology*.

Semantic indexing specifically means that semantically expressive relationships, such as *topic*, *location*, *creator*, may be used to index items of content (i.e. images or sub-image annotations) to descriptive terminology. By using Linked Data the indexing relationships become predicates in a subject-predicate-object triple, and predicates can be adopted from authoritative sources (such as Dublin Core Terms, FOAF, SKOS, and OWL), further promoting data portability and reuse.

Controlled vocabulary terminology specifically means that indexing is performed using conceptual entities rather than plain text strings. Conceptual entities also have unique HTTP-URIs and associated descriptive metadata. Concepts are members of a concept scheme that may organize the concepts into a Knowledge Organization System (KOS) using hierarchical, associative and equivalency relationships. By using Linked Data and authoritative controlled vocabularies (such as LCNAF, GeoNames, Getty AAT and IconClass), the concepts used to describe images and sub-image annotations can themselves become gateways to connect to additional external data sets that are also related to the same concepts.

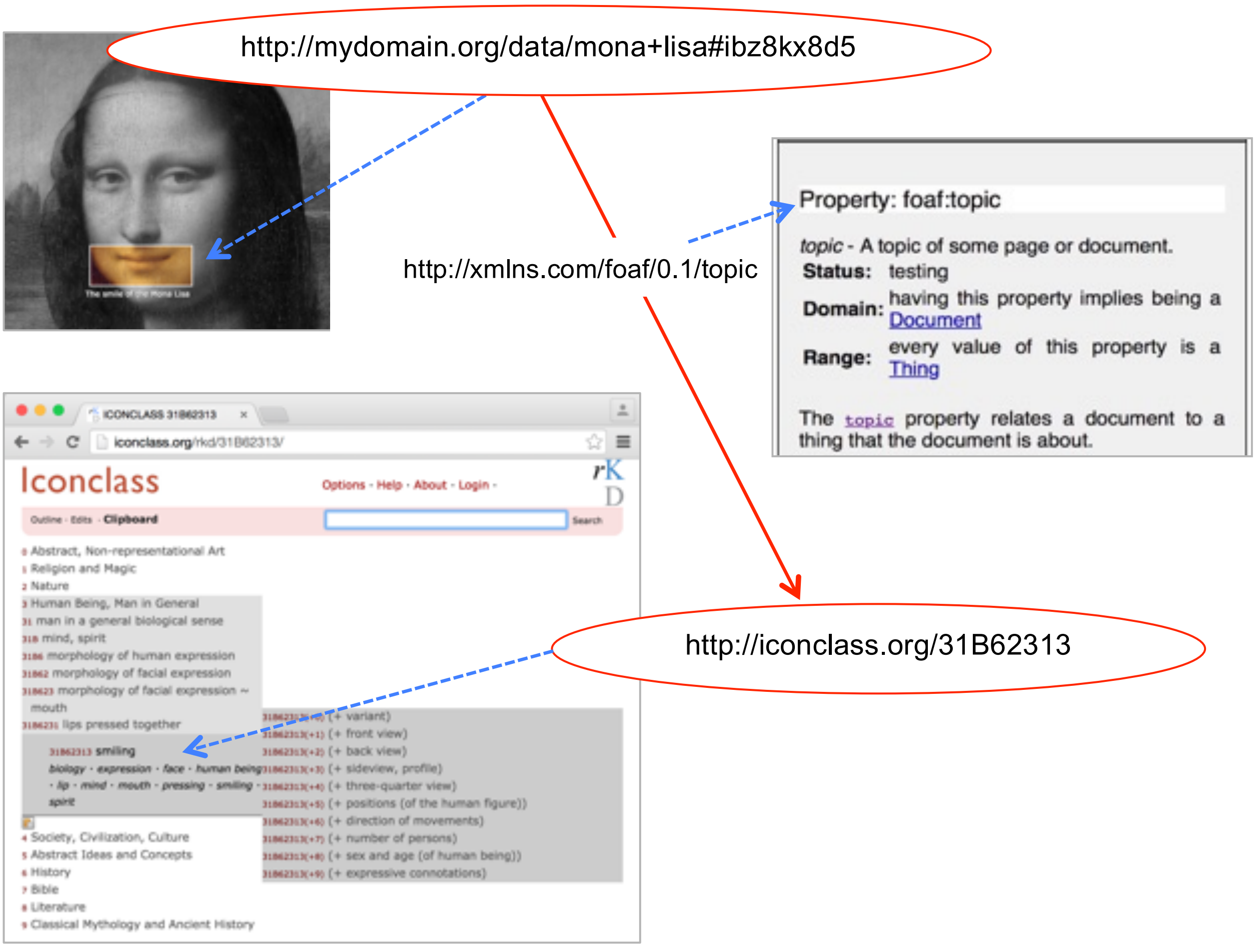


Figure 9: *The smile of the Mona Lisa* is indexed via foaf:topic to the IconClass concept *Smiling*

4 Knowledge Organization Methods

While numerous image annotation tools and standards currently exist, none have attempted to fully integrate the process and methods of image annotation with the process and methods of Knowledge Organization (KO). The integration of these processes and methods is a key objective of the OASIS R&D project.

OASIS identified six primary methods for organizng image and sub-image content.

4.1 Image-Level Metadata

Image-level metadata describes the whole image. At a minimum, descriptive properties usually include an image title, long description, date and creator. They could also record metadata automatically captured by cameras according to the Exif standard[6]. Different knowledge domains and disciplines will also each have the need to capture additional information. In the cultural heritage community it may be desirable to distinguish the date and creator of the photograph from the date and creator of the object depicted in the photograph, and to add provenance information about the depicted object. The image-level metadata may be further extended and linked to collection management system records. In the healthcare and life-sciences community it may be desirable to include metadata captured according to the AIM or DICOM standards (see section on Standards).

Figure 10: Image-level metadata

A generalized prescriptive approach to image-level metadata would fail to satisfy most real-life user needs. The OASIS design solution was to treat image-level metadata as a completely extensible ontology. A base set of properties and attributes can be defined for any knowledge domain and then be quickly and easily edited and extended by end-users using GUI-driven tools.

Image-level metadata provides a catalogue of the images in a collection. The metadata elements can support faceted search and navigation of the collection.

[6] Exif http://www.exif.org

4.2 Annotation-Level Metadata

Annotation-level metadata describes particular points of interest or regions of interest in an image. At a minimum, descriptive properties usually include an annotation label and long description. Additionally metadata is required to define the spatial co-ordinates of the point or area and the level of zoom defined for the point or region. Spatial co-ordinates may be expressed in pixels offset from the top-left corner or in percentages of the width and height. In OASIS an HTTP-URI property is also generated for every annotation.

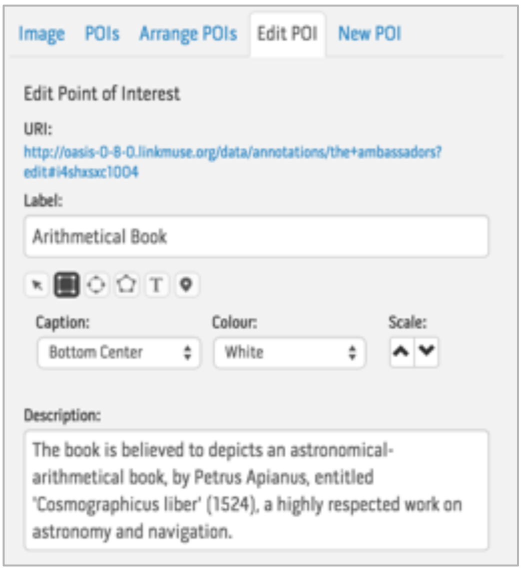


Figure 11: Annotation-level metadata

As with Image-level metadata extensibility is essential in order to meet the real-world needs of users in a diversity of knowledge domains and disciplines.

Annotation-level metadata provides an inventory of the visual features within an image. The metadata can support a *search inside* function and table-of-contents style browsable navigation.

4.3 Listing Features as a Table of Contents

Textual works are frequently divided into sections supporting access via a table of contents. Pictorial works may also be decomposed into sections based on regions of interest and points of interest.

Alphabetical lists of annotated features provide a table of contents for an image. These features can also be arranged into hierarchies, as described in the following section.

4.4 Arranging Features into Hierarchies

The visual features of an image may be arranged into hierarchies based on thematic or associative sets and sub-sets, or to represent the spatial decomposition of an image.

For example, *The Garden of Earthly Delights* triptych by Hieronymus Bosch comprises three discrete panels each of which contains a number of separate scenes and within each scene there are a number of vignettes and/or individual figures.

By organizing these visual features into a hierarchical structure the complexity of the image can be decomposed thus making the image more accessible and navigable.

Deep Image Annotation: Making a Difference in Knowledge Organization

David Clarke – ISKO Fourth Biennial Conference, London, July 2015

Figure 12: The complete Bosch triptych can be accessed by a hierarchical arrangement of panels, scenes, figures and features

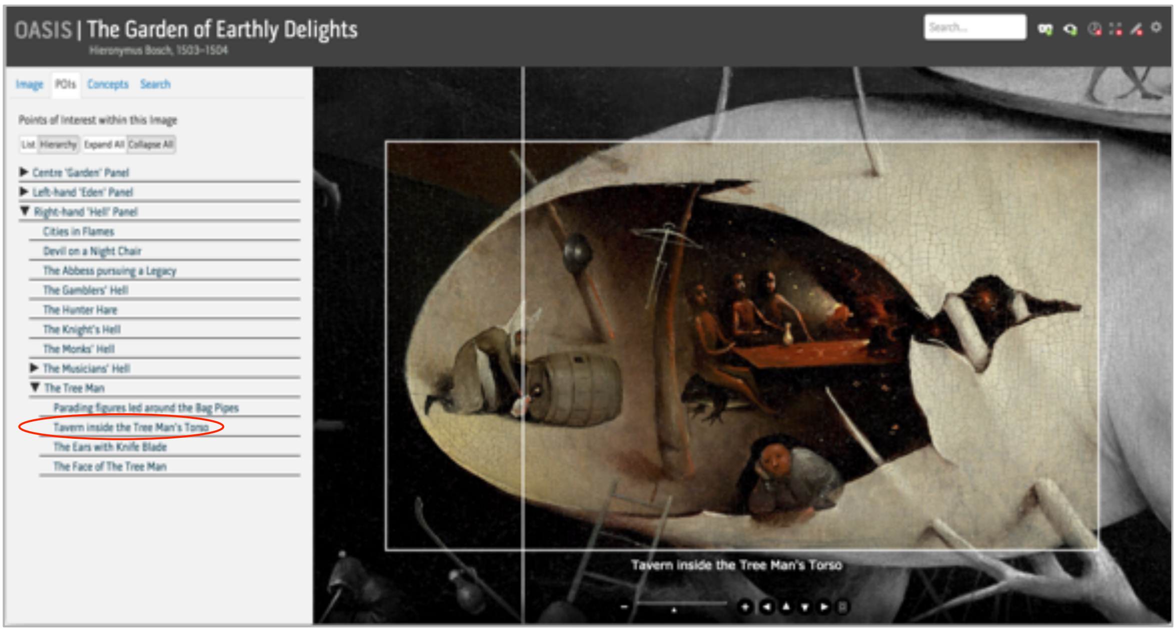


Figure 13: When an individual feature is selected it triggers pan-and-zoom to that feature

4.5 Building and Adopting KOS

Knowledge Organization Systems (KOS) are key enablers for accessing an internal content set by its concepts and also for connecting content to other internal or external content sets.

The OASIS design required the ability to create and build KOS directly within the image annotation system and also to link to and/or ingest cached copies of external KOS schema.

Scheme management is modular, allowing KOS to be defined using building blocks of resources from other internal or external ontologies. It is also extensible, providing end-user GUI-driven tools to add, modify and adopt KOS schema.

The adoption of external authoritative KOS schema from the Linked Open Data cloud is encouraged as this maximizes the opportunity to connect to additional data sets. In a cultural heritage prototype internal KOS schema were created and large controlled vocabularies and name authority files published as Linked Open Data were ingested including Library of Congress Name Authority Files (LCNAF)[7], The Getty Research Institute Art and Architecture Thesaurus (AAT)[8], and IconClass[9].

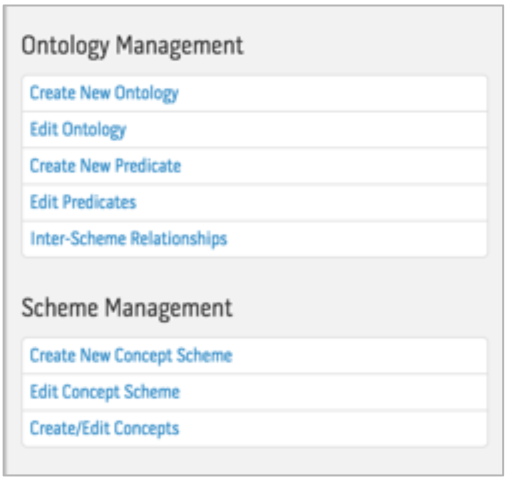


Figure 14: OASIS Ontology and KOS configuration tools

4.6 Semantic Indexing

As previously noted in section 3.4, ***Semantic indexing*** specifically means that a semantically expressive relationship is used to index an item of content (i.e. image or sub-image annotation) to descriptive terminology (derived from a KOS controlled vocabulary where possible). Consistent with the general design guidelines adopted by the OASIS team, end-user configuration and extensibility of semantic indexing relationships is supported.

The OASIS method for managing semantic indexing relationships comprises three steps:

1. Create or adopt predicates to serve as semantically expressive indexing relationships using external authoritative ontologies where possible.
2. Define inter-KOS rules to govern which predicates may be used as relationships between entities from any pair of KOS schemes.

[7] LCNAF: http://id.loc.gov/authorities/names.html
[8] AAT: http://www.getty.edu/research/tools/vocabularies/lod/
[9] IconClass: http://www.iconclass.org/help/lod

3. Support rules-based indexing where the system assists indexers with the choice of appropriate indexing relationship types.

Each of these three methods are illustrated below by screens from the applicable OASIS interface where the methodology is supported.

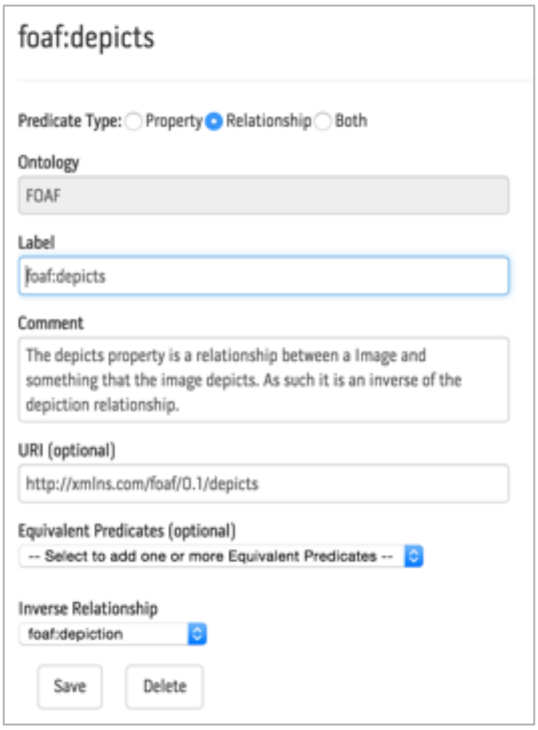


Figure 15: Predicates from external authorities may be adopted

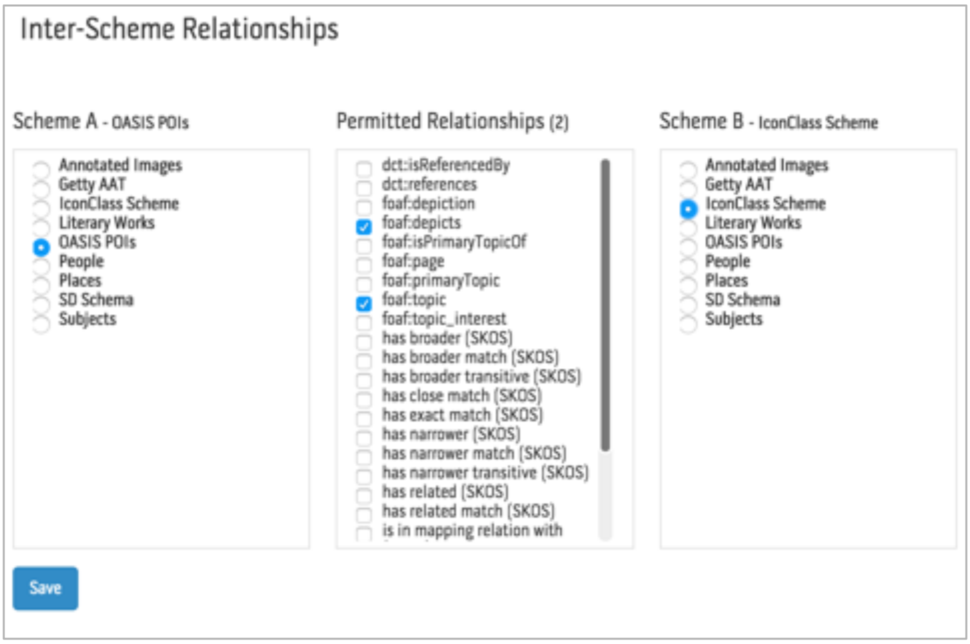


Figure 16: Rules constrain which predicates can be used between which schemes

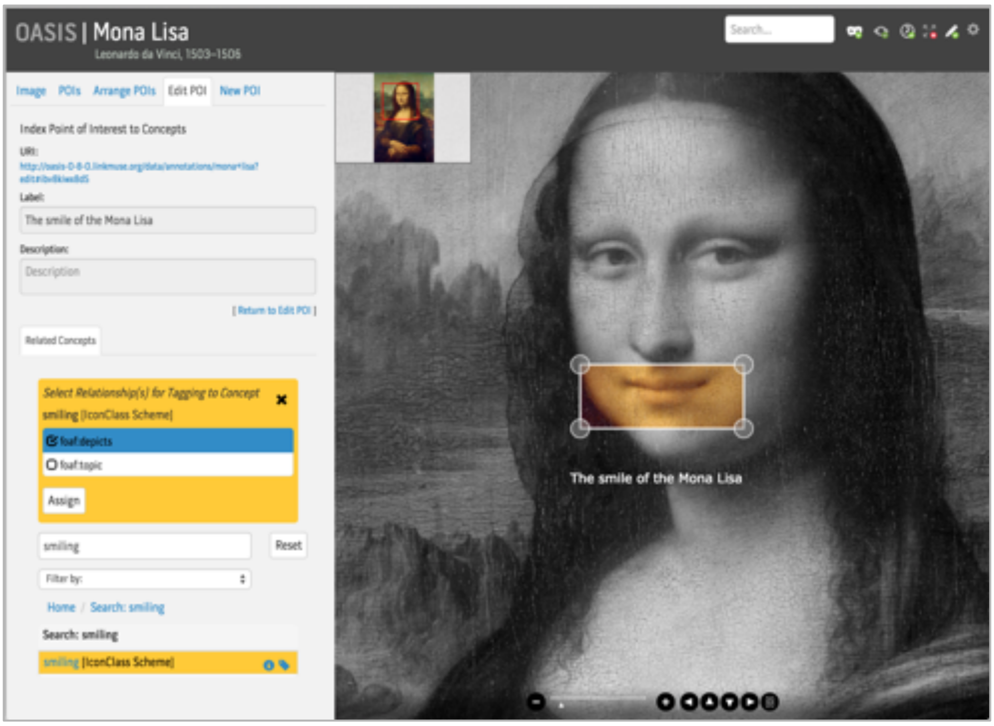


Figure 17: Rules-based filters guide and constrain semantic indexing

5 Search, Refine, Browse and Discover

Implementing knowledge organization processes and methods should deliver improvement in any or all of the following aspects of information access:

1. **Search**, which starts with a user's expression of their question then follows iterative refinement
    a. disambiguation: did you mean X or Y
    b. expansion: user types X, query uses X plus synonyms or language equivalents
    c. redirection: user types X, query redirects to preferred term Y
    d. filtering: user directs criteria to specific facets and/or creates a compound query
2. **Refinement**, which is an iterative process for refining search based on retrieved metadata
    a. metadata in the current results list or displayed items of content is extracted and ranked then grouped by facet
    b. facets of ranked metadata present additional filters
    c. selections from the ranked metadata filters refine the results list, which triggers another refresh of the facets of ranked metadata filters
3. **Browse**, which starts with a presentation of organised lists or graphs of related things then follows a user's selected pathway
    a. alphabetical: alphabetically sorted lists
    b. hierarchical: hierarchically sorted classifications
    c. ranked: relevance-ranked or enumerated lists

    d. visualisations: visualisations of classified sets of results or connected nodes
    e. matrices: multi-parameter tables with sort by column support
    f. galleries: scrolling carousels or matrix arrays of results including images
4. **Discovery**, which interrupts the flow of a user's search or browse upon the surfacing of associated concepts or content of interest
    a. associations: content or concepts related to the currently displayed content (directly or indirectly)
    b. intersections: content or concepts that represent the union of criteria derived from multiple results sets or facet filters

The OASIS integration of knowledge organization processes and methods with image annotation processes and methods has delivered several new ways to search, refine, browse and discover the information inside pictorial content. The following sub-sections provide examples to illustrate some of the ways that visual content can be explored and interrogated.

5.1 Search Across Image Collections

Searching across an image collection can be supported by any image management system assuming the images have textual metadata. The use of KOS vocabularies to index images should improve search precision and recall and may also offer faceted search and pattern-based query options.

5.2 *Search Inside* Images

Searching inside images, i.e. searching for discrete visual features within images, can be supported by most image annotation tools. The use of KOS vocabularies to index annotations greatly improves the retrieval of image content by allowing users to search on the ideas and concepts behind the pictorial details. This is particularly powerful if the visual material employs allusion, metaphor and symbolism.

Figure 18 illustrates how a search inside Holbein's *The Ambassadors* using the query 'mortality' identifies numerous visual details which are related to the idea of mortality but which do not use the word *mortality* in the annotation text.

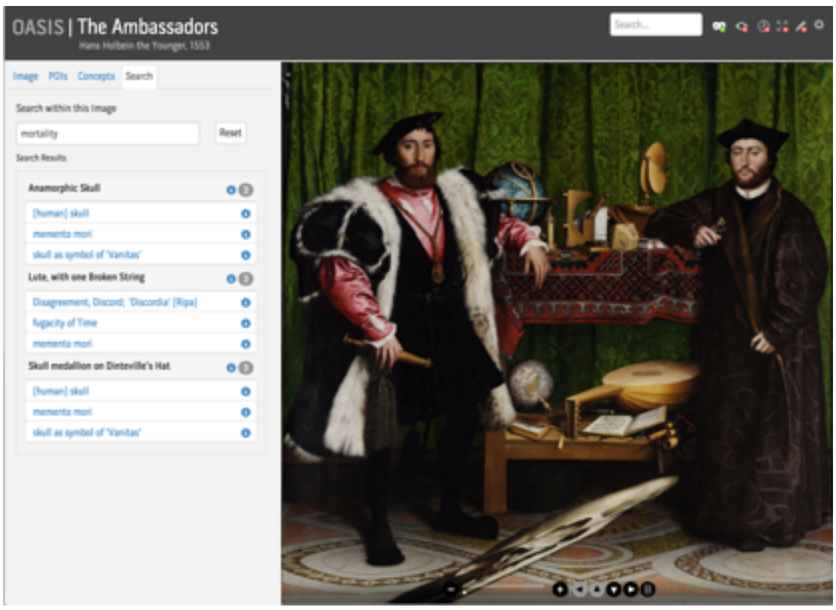


Figure 18: Conceptual search inside image reveals visual details related to the idea

5.3 Idea to Image and Image to Idea

Users interact with knowledge discovery applications in different ways and may often want to switch rapidly between opposite modalities, such as browsing feature lists and concepts to find related content, and browsing content to discover the associated features and concepts. OASIS supports switching between these two modalities from the same screen. The following two images illustrate this. The first uses a hierarchical list of visual features to navigate from a feature to the part of the image it describes. The second presents a concept cloud that dynamically updates as the user pinches and swipes to move around the image.

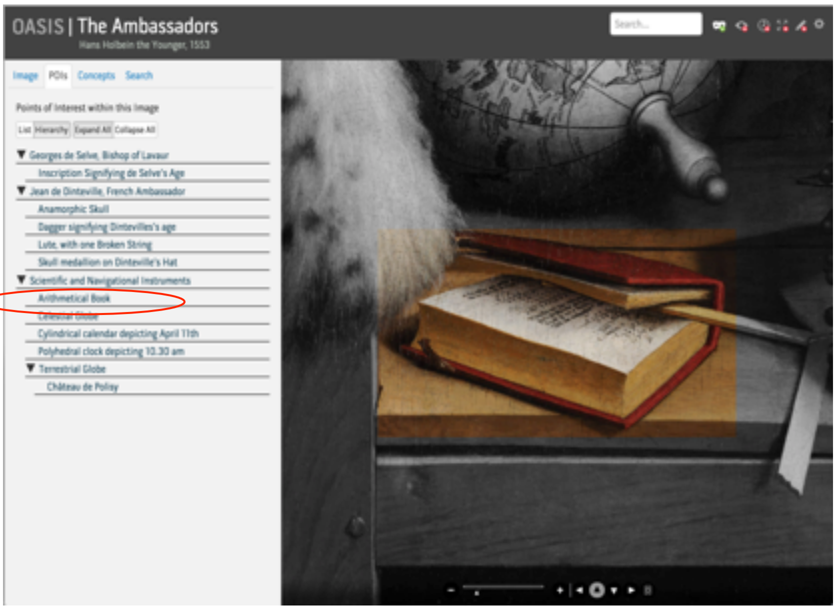


Figure 19: Browsing a feature list to pan-and-zoom to the image detail it describes

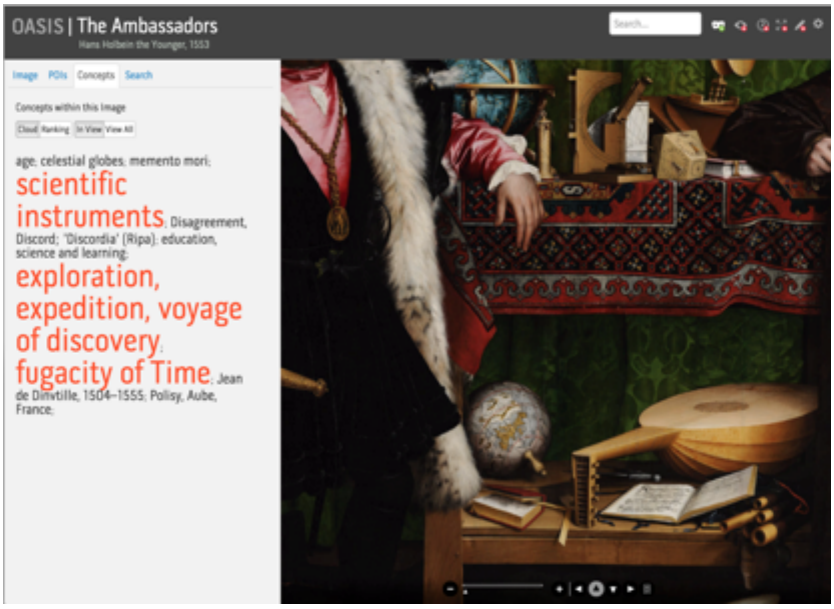


Figure 20: Pan-and-zoom around the image updates the concept cloud to reveal what is in view

5.4 Idea to Idea

When an information system uses KOS users can navigate from idea to idea as well as between ideas and content. This allows users to broaden, narrow or redirect the focus of their search.

In the example below the user has selected the concept 'scientific instruments' from the Getty AAT thesaurus. This presents a concept explorer window providing options to broaden or narrow the focus of the query.

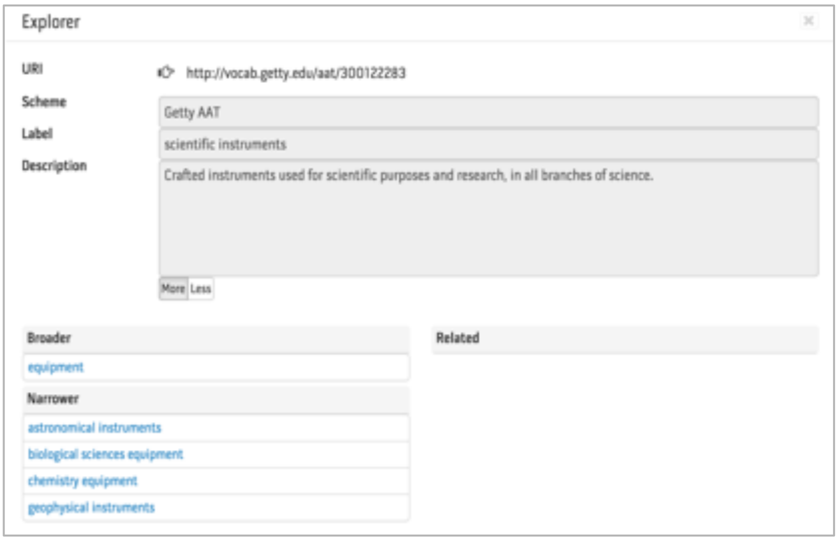


Figure 21: Navigating between concepts to narrow or broaden a search

5.5 Image to Image Discovery

One of the most powerful knowledge discovery capabilities that can be facilitated by a KOS is to draw to the users attention to unexpected connections between related ideas and content.

In the example below the OASIS system was able to suggest related images and details of images while viewing a detail on a specific image. The visual artifacts did not have an obvious connection: a knight, a lute, a vase of flowers and several skulls, but the connection was made via the semantic indexing, in this case using the Getty AAT knowledge organization system.

<table>
<tr>
<td>
The Ambassadors[10]
Hans Holbein the Younger,
1553</td>
<td rowspan="2">Concept Details
Label:
memento mori [Getty AAT] [11]
Description:
Symbols, often objects such as skulls or hourglasses, but sometimes written inscriptions, intended to remind the viewer of death.
Related Points of Interest on this Image:
• Fading flowers
Related Images / Points of Interest:
• Portrait of a Knight
• St. Francis in the Desert
  o Skull
• The Ambassadors
  o Anamorphic Skull
  o Skull medallion on Dinteville's Hat
  o Lute, with one Broken String</td>
<td>
St. Francis in the Desert[12]
Giovanni Bellini
1480</td>
</tr>
<tr>
<td>
Portrait of a Knight[13]
Vittore Carpaccio
1510</td>
<td>
The Merchant Georg Gisze[14]
Hans Holbein the Younger,
1532</td>
</tr>
</table>

[10] Source: http://en.wikipedia.org/wiki/File:Hans_Holbein_the_Younger_-_The_Ambassadors_-_Google_Art_Project.jpg

[11] Source: http://vocab.getty.edu/aat/300257053

[12] Source: http://en.wikipedia.org/wiki/File:Giovanni_Bellini_-_Saint_Francis_in_the_Desert_-_Google_Art_Project.jpg

[13] Source: http://en.wikipedia.org/wiki/File:Vittore_Carpaccio_-_Young_Knight_in_a_Landscape_-_Google_Art_Project.jpg

[14] Source: http://en.wikipedia.org/wiki/File:Hans_Holbein_der_Jüngere_-_Der_Kaufmann_Georg_Gisze_-_Google_Art_Project.jpg

6 Conclusions

The OASIS project has demonstrated that knowledge organization systems and methods can greatly enhance the ability to annotate and provide access to image content.

The use of controlled vocabularies and semantic indexing plays a mission-critical role in making connections between related content that is inherently non-textual and thus non-searchable.

The W3C Open Annotation Data Model, released in February 2013, has provided a successful high-level model for the annotation of image content.

7 Further Research and Collaboration

Synaptica is currently running trials of the OASIS system within the cultural heritage community, and invite enquiries from any organizations wishing to explore deep image annotation.

OASIS is an ongoing R&D program and further enhancements and extensions are under development.

Enquires about OASIS may be directed to: research@synaptica.com.