Open Annotation Collaboration Phase II Demonstration Experiments Case Study Report

Prepared by Jacob Jett & Carole L. Palmer

6/7/2013

Center for Informatics Research in Science & Scholarship

Graduate School of Library & Information Science

University of Illinois at Urbana-Champaign

Supported by a grant from the Andrew W. Mellon Foundation

Table of Contents

Contents

Table of Contents....2
Executive Summary....3
Open Annotation Project Overview and Goals....4
Methods....6
Demonstration Experiment Profiles....7
Findings....10
Data Model Complexity vs. Simplicity....10
String Literal Body Content (Data Model too complex)....11
Media Fragments (Data Model too complex)....12
Annotation Classification (Data Model too complex)....13
Resource Multiplicity [Data Model too simple]....13
Aggregating Annotations (Data Model too simple)....15
Structured Annotations (Data Model too simple)....15
Community Best Practices....16
Non-information Objects....16
Content Negotiation....16
Annotation beyond the Digital Humanities....17
Recent Developments....19
Summary of Findings & Recommendations....23
References....25

Executive Summary

This report presents results from a case study, conducted as part of Phase II of the Open Annotation Collaboration (OAC), examining nine annotation demonstration experiments and associated use cases. During Phase II, the OAC actively developed and experimented with an RDF-based annotation data model. The primary features of the data model were developed in response to findings in Phase I and evolved during the course of Phase II based on feedback from the demonstration experiments, community discussions, and face-to-face meetings. The case study was based primarily on interviews conducted with project developers and user groups, supplemented with information from final reports submitted by the participating projects.

Developers faced common use cases and challenges during the course of their experimentation with the OAC Beta and OA Draft Specifications. The following three issues emerged collectively from the OAC experiments:

- Parts of the data model are too complex resulting in a heavy burden for compliance.
- Parts of the data model are underdeveloped for certain use cases or common implementation needs.
- Community best practices and worked examples are needed to facilitate implementation of the data model.

Some demonstration experiments had different perspectives on key issues, related to the use cases for their content and user groups. Video annotation was found to need multiple selectors to successfully segment the annotation target, while a tool for annotation of TEI text would find this feature of the data model superfluous. In addition, some demonstration experiments needed methods within the data model to precisely define target regions beyond simple rectangular boxes, while other demonstration experiments could exploit web standards such as Media Fragment URIs to adequately communicate the region being annotated. While most developers had use cases that called for the juxtaposition of resources, only text mining and linguistics use cases absolutely required well-defined multiplicity semantics for choice, union, and sequence.

Implementers experimenting with the OAC Beta and OA Draft data models frequently raised questions about the complexity of the data model in comparison to the simplicity of use cases. They also readily pointed out areas where the specification lacked sufficient guidance for implementers or the data model lacked desired features that would facilitate building implementations for specific use cases. During the course of the case study, the Open Annotation Community Group resolved some of these issues through community discussions on listservs and through face-to-face meetings of community members. A few issues were resolved by fiat; however, the sustainability of these solutions is uncertain and will likely need to be revisited in the future. The OAC demonstration experiments also revealed that there are a number of specific use cases associated with annotation of scientific texts and analysis of texts that need to be further explored.

Further work is needed to better understand the complexity dynamics of data model development. Several of the additions to the data model made by the Community Group to resolve ambiguity and scalability issues should be validated through application and testing, and the long term implications of modeling decisions need to be tracked and assessed. Finally, additional expansions to the OA data model should address better support for aggregation of annotations, evidentiary chaining of annotations, and annotation use in linguistics, text mining, and the sciences.

Open Annotation Project Overview and Goals

Project History & Context

The Open Annotation Collaboration (OAC) is a Mellon-funded project to develop an RDF-based linked data interoperability standard for web annotations. The first phase of OAC (2009-2010) was a partnership led by the University of Illinois at Urbana-Champaign, with collaborators at The University of Queensland, the University of Maryland, Los Alamos National Laboratory, and JSTOR. The primary goal was to develop a standard to guide the development of annotation tools and the sharing of interoperable annotations in the digital environment. As a scholarly primitive, annotation is a fundamental research function common across humanities disciplines (Unsworth, 2000). It is vital to, and often serves to coordinate, numerous activities within the research process, including searching, collecting, reading, and writing (Palmer, Teffeau, & Pirmann, 2009). During Phase 1, the OAC team closely examined use cases presented by the digital humanities community to better understand annotation in practice.

During OAC Phase II (2011-2013), a continuing collaboration among the University of Illinois at Urbana-Champaign, The University of Queensland, the University of Maryland, and Los Alamos National Laboratory, the project focused on active experimentation to further develop the OAC data model and ontology. To expand the range of technical development and use cases informing the OAC data model, a Request For Proposals (RFP) was released in April 2011 to identify five annotation demonstration experiments to participate in OAC II. This report provides an analysis of a case study designed to examine these sub-award experiments and their associated use cases, to assess outcomes of their experimentation with the data model and document lessons learned across both phases of OAC development. The following projects are covered in the report:

Core Phase II Demonstration Experiments:

1. The University of Queensland. *Annotation Supporting Collaborative Development of Scholarly Editions*.
2. Los Alamos National Laboratory (in collaboration with Stanford University). *Annotation of Digitized Medieval Manuscripts*.
3. University of Maryland (in collaboration with Alexander Street Press). *Annotation of Subscription Streaming Video Content*.
4. University of Illinois at Urbana-Champaign (in collaboration with the Herzog August Bibliothek Wolfenbüttel). *Annotation of Digital Emblematica*.

Phase II Sub-Award Demonstration Experiments:

5. Cornell University. *MapHub Phase II*.
6. Brown University. *OAC Compliant Annotation Framework in Fedora*.
7. Meertens Institute, Royal Netherlands Academy of Arts and Sciences (KNAW). *CODA – CATCHPlus Open Document Annotation*.
8. University of Colorado Denver. *Automated Annotation of Biomedical Text*.
9. New York University. *Annotation Middleware for Scholarly Publications and Resources*.

Intersections with Annotation Ontology & Development of the Open Annotation Community

During the course of OAC II the OAC team became aware of a parallel standard being developed by the Annotation Ontology. The Annotation Ontology (AO) is a Boston-based project at MIND Informatics, a research center funded by MassGeneral (Massachusetts General Hospital) Institute for Neurodegenerative Disease. After a series of meetings and teleconferences through Fall 2011, project leads from both initiatives decided it would be beneficial to merge the two data standard development efforts. In January 2012, the Mellon Foundation awarded OAC a supplemental grant to hold two face-to-face meetings of the Open Annotation (OA) Community Group, founded that month through a merger of the Annotation Ontology and Open Annotation

Collaboration communities. The OA Community Group maintains its web pages and email archives through a W3C Community Groups web space.

The continuing development of the data model and specification has since become the responsibility of the OA Community Group. A preliminary specification, the Open Annotation Draft specification, was published in May 2012 following the outcomes of a March 2012 meeting in Boston of core Community Group members. OAC II project members quickly began experimenting with the new Draft data model produced by the community. The interviews conducted for the case study focused on developer experimentation with both the OAC Beta specification and the OA Draft specification as it was published in May 2012.

Methods

The case study was based on interviews conducted with demonstration experiment developers and user groups, supplemented with information from the final reports submitted by each OAC II demonstration experiment as a part of their sub-award deliverables. Each interview adhered to the protocol designed by OAC II project coordinator Jacob Jett and co-principal investigator Dr. Carole Palmer. Developers at each of the nine OAC II demonstration experiments were invited to participate in the case study. While all of the projects agreed to participate, primary interview sessions with Illinois and Los Alamos developers were not completed due to scheduling conflicts; however, since Illinois was the OAC II project lead and Los Alamos led authorship of the OA specification, these institutions had more direct influence on the data model's development. To further explore the kinds of user practices and expectations tool developers needed to accommodate, secondary interview sessions were conducted with groups of users identified at two of the experiment institutions, Illinois and Colorado. All participants agreed to an IRB approved informed consent statement at the beginning of the interview session.

The seven primary interviews were carried out between March and June 2012 via Skype® and audio recorded. Each developer was interviewed individually except for a pair of developers at New York University Libraries who were interviewed together. The secondary interviews with the user groups consisted of in-person sessions with 3 participants at Illinois and with 2 participants at Colorado, conducted via Skype®. User group participants agreed to an IRB informed consent statement at the beginning of the interview, with each session adhering to the interview protocol.

Case study analysis applied the experiential knowing methodology outlined by Stake (1995). Analysis within cases was based on naturalistic interpretations of the interviews, supplementary case materials, and observations, followed by categorical analysis of patterns across the cases.

The body of the report is organized in four sections. The first two sections, Demonstration Experiment Profiles and Findings, are drawn directly from the interviews and final reports. The profiles provide context for each experiment and frame the overall scope of the group of experiments conducted within OAC. The findings are presented in general categories that represent key issues highlighted during the interviews. The third section covers OAC project developments that emerged during the course of the case study. The forth section provides a summary of the key findings followed by selected recommendations based on the case study analysis.

Demonstration Experiment Profiles

Annotation Supporting Collaborative Development of Scholarly Editions

Synopsis: The University of Queensland built a suite of annotation tools that let humanities scholars collaborate on the production of scholarly editions of digitized manuscripts. As part of the OAC II project, tools were based on the OAC Beta Specification (Aug. 2011). The annotation tool suite, LORE (Literature Object Re-use and Exchange),[1] allows users to interact with multiple versions of a text at the same time and create annotations that span them. Annotations made using LORE are stored within an annotation repository, LOREStore,[2] available for reference and retrieval throughout the lifespan of LORE.

Annotation Use Cases: In addition to the supporting scholarly note-taking activities in an online environment, the Queensland experiment accommodated the need of scholars to use their annotations as a medium for discourse during collaboration. One of the more complex use cases that Queensland highlighted during OAC I was the need not just to segment the targets of the annotation, i.e., to select specific portions of each text, but also to link the targets together to preserve the context of the annotation, as needed, for example, when comparing two targets. In addition, with the annotations created within LORE stored in an annotation repository, the use case also indicated the need for differentiating between different kinds of annotations to support groupings of annotations for retrieval by a query.

Annotation of Subscription Streaming Video Content

Synopsis: The Maryland Institute for Technology in the Humanities (MITH) collaborated with researchers at Alexander Street Press (ASP) to develop a series of browser-based plug-ins supporting annotation of video content. The plug-ins employ the OA Draft Specification (May 2012) to create annotations that have a standard, and therefore interoperable, data architecture independent of the video client software they are used with.

Annotation Use Cases: The MITH / ASP experiment was designed to address the use case where the annotations require simultaneous segmentation across multiple dimensions. In the case of video, segmentation of the annotation target is needed by play time and surface region. The two segmentation conditions required to properly identify a portion of video for annotation is a different kind of multiplicity than seen in the Queensland use case of multiple linked targets. An important implication is that resource multiplicity issues within the OAC and OA data models may be generalizable beyond the multiple targets case, i.e., there is also a need for a multiple selector resource.

Annotation of Digital Emblematica

Synopsis: The OAC researchers at Illinois carried out an in depth study of the annotation needs of scholars of digital emblems. Emblems, a bimodal genre, are an interesting area of humanities scholarship due to the brief period of time in which they were created (ca. 1530-1750) and the manner in which they combined text and images to express aspects of the culture during that time. The Illinois team engaged a group of emblem scholars to use existing annotation tools, Pliny and Dannotate, and document their experiences with them. The goal was to map the scholars' actual annotation usage to annotations conforming to the OAC, and later the OA, data models to better understand how the model met or failed to meet end user needs.

Annotation Use Cases: The emblem scholar case demonstrated a variety of annotation practices that aligned with other digital manuscript use cases. For the most part emblem scholars used annotation for note-taking but

[1] http://itee.uq.edu.au/~eresearch/projects/aus-e-lit/lore.php
[2] http://openannotation.metadata.net/lorestore/

also as part of their scholarly discourse process. In particular, emblems, as a form of composite media, illustrated the need to segment very specific portions of a resource for annotation. Like the Queensland example, emblem scholars also frequently needed to compare resources, requiring that the data model can accommodate multiple targets.

During the course of the user study it became apparent that the scholars had developed hierarchical, and in one case quite complex, vocabularies to index their annotations as they were produced, annotating the annotations as explanations, comparisons, references, etc. These labels were later used to retrieve specific groups of annotations. With regards to the OA specification, these types of indexing and retrieval activities could be supported by assigning sub-classes of the main annotation class, oa:Annotation, and provided additional evidence for the Queensland observation on the role of classing for retrieval from an annotation repository; however, the Queensland project also noted that sub-classing annotations, especially across multiple domains, did not seem like an entirely scalable solution. Illinois developers also noted that the rdf:type property was not intended to be used to preserve these kinds of finely grained taxonomic distinctions between resources.

MapHub Phase II

Synopsis: Cornell University applied the OA Draft (May 2012) data model to the latest version of their existing map annotation software application, MapHub. Using a corpus of 6,000 historic maps from the Library of Congress, Map Division, prototype web and smartphone-applications were developed to allow users to annotate digitized maps in three ways: by adding textual notes, by adding geotags, and by overlaying other maps. As an integral part of the experiment, the Cornell team assessed the utility of semantic tags verses normal text string tags, recording the frequency of use of semantic tags suggested by the Maphub applications and comparing semantic tagging activities to more traditional, label tagging activities.

Annotation Use Cases: In addition to further examples of note-taking and target segmentation uses, the Cornell project contributed some unique examples of georeferencing points on the maps, overlaying maps, and tagging parts of the maps. For georeferencing, users make annotations that link points on the map to other web resources, such as a DBPedia page on a specific location. The tagging activity involved minting annotations with multiple bodies. Some annotations were machine-suggested semantic tags that link points on the map to non-geographic concepts.

One of the more novel use cases emerging from the MapHub work was the users' desire to annotate portions of one map with those from another map. This demonstrated that there was a need to segment annotation bodies in the same way as annotation targets and suggested that bodies and targets shared properties. Further explorations by the larger community, through the resource multiplicity and selector discussions, eventually confirmed that bodies and targets are symmetrical with respect to the range of properties that they possess and that segments of an annotation body can be selected in exactly the same manner as target segments.

An OAC Compliant Annotation Framework in Fedora

Synopsis: Brown University created a prototype web service that allows developers to build OA Draft (May 2012) compliant annotation tools that store data in a Fedora repository framework. As part of their proof of concept testing, they built a prototype annotation client that allows users to annotate TEI texts and used the web service to store and retrieve annotations in the Fedora repository.

Annotation Use Cases: Like the experiments at Queensland, Illinois, and Cornell, the Brown annotation tools support both note-taking and discourse activities, however the group was somewhat unique in their years of prior work with annotation client software to support digital humanities scholarship. Their application of annotation to TEI texts spanned multiple resources—the Women's Writers Project, Virtual Humanities Lab, US

Epigraphy, and Inscriptions of Israel & Palastine—and suggested that annotations had vastly increased utility when treated as web resources, leading to the next steps achieved with the OAC framework in Fedora.

CODA – CATCHPlus Open Document Annotation

Synopsis: The Meertens Institute CODA experiment, part of the Dutch CATCHPLus program implementing tools to support digital humanities research, implemented an OAC Beta (Aug. 2011) compliant annotation toolset and annotation repository. Two of the CATCH corpuses, the Queen's Cabinet collection at the Dutch National Archives and the Sailing Letters collection at the National Library of the Netherlands, were used to evaluate the CODA toolset. In addition, Stanford's Shared Canvas annotation tool client was tested to compare approaches for annotating digitized images of textual manuscripts.

Annotation Use Cases: Toolset development responded to two primary use cases. The first was the need to apply Optical Character Recognition (OCR) results to the digitized text as annotations for further processed using Named Entity Recognition tools. The second was a technological requirement to detect the bounding boxes used to help the OCR software transcribe selected portions of text, with testing indicating a preference for the Shared Canvas approach. In addition, in developing an annotation repository and related search and retrieval services to aid with assessments of the two annotation toolsets, analysis of querying requirements found a need to type and sub-type annotations to improve retrieval.

Automated Annotation of Biomedical Text

Synopsis: University of Colorado Denver developed tools to add linguistic and part of speech annotations to biomedical articles. Built using the Colorado Richly Annotated Text Corpus (CRAFT), machine learning techniques were applied to build an automated workflow linking concepts in the biomedical texts to multiple gene ontologies. The end goal is to prepare the biomedical dataset with annotations rich enough to facilitate reasoning algorithms to aid in document retrieval. The Colorado project implemented the OAC Beta (Aug. 2011) data model.

Annotation Use Cases: Two use cases were involved in preparing text data for further machine processing. Part of speech annotations are layered into the texts to support linguistic analysis to identify claims, hypotheses, theories, and research outcomes. Text is also prepared by adding a layer of concept annotations that link segments of text to concepts in various gene ontologies. These annotations are then used to support data mining applications and the production of data visualizations. For the part of speech process, it was found that the provenance of annotations needed to be recorded for assertions based directly on other annotations.

Annotation Middleware for Scholarly Publications and Resources

Synopsis: New York University (NYU) developed a proof of concept demonstrator that supports annotation of several samples of digitized text. The prototype stores annotations in an annotation repository, and the demonstrator as a whole has been built to be compatible with Drupal content management systems. The NYU group used the OAC Beta (Aug. 2011) data model in their experiment.

Annotation Use Cases: The use cases confronting NYU mostly conformed to the scholarly note-taking and discourse use cases reported by other experiments. However, unlike other projects, the contexts for the discourse and note-taking activities were educational classroom use of digital objects, and workflow related—NYU cataloging staff had a need to annotate specific segments of digitized documents with metadata as an alternative to crafting a separate metadata records. This reaffirmed the already known general day-to-day use of annotations on the web for commenting, tagging, etc.

Findings

The findings across the cases are organized into three broad themes:

1. Data model complexity vs. simplicity.
2. Accommodating best practices in the technical community.
3. Emerging scholarly and scientific practices.

The first two categories address concerns reported by the participants in relation to both their specific OAC sub-award demonstration experiments and to the larger Open Annotation community. The third finding relates to the one science-based experiment at the University of Colorado Denver, which is unique among the OAC projects.

Data Model Complexity vs. Simplicity

The range of experiments presented a dichotomy of data model requirements, demonstrating a need for a complex model to support annotation tools that empower users and enrich their work, as well as need for a simple, accessible data standard that can be readily implemented.

In general, a robust and thorough data model allows developers to build tools that satisfy a large variety of use cases. Having multiple ways to express data allows modeling of many kinds of use cases, including exceptions to the norm. The same level of expressivity can also be used to build multiple alternative models for the simplest and most common of use cases. In turn, these multiple simple solutions exert pressure on developers to build tools that either have relaxed validity conditions or which are not truly interoperable with other tools conforming to the same specification. Data models are frequently complex (Batra, 2007), in part because people find the features of objects or systems with many simple parts difficult to understand (Casti, 2001; Flood & Carson, 1988; Niekerk & Buhl (2004)).

While the essential data model promulgated by the OAC and the subsequent OA is quite simple (see Figure 1), the realities of Web architecture, RDF and Linked Data necessitated additional features, properties, and relationships, which when taken as a whole make the OA data model quite complex.

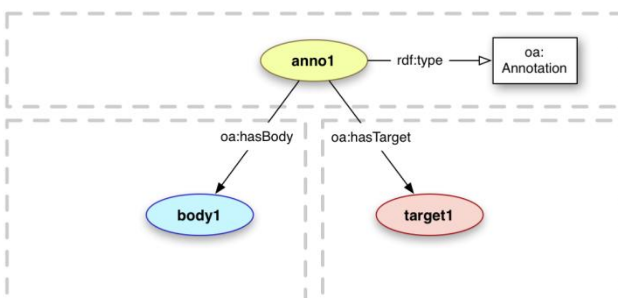


Figure 1: Basic Annotation Model[3]

During the course of the case study interviews it became apparent that while the complexity of the OAC and OA data models facilitated implementing tools in response to use cases, about half of the developers found it could deter interoperability. In particular, developers frequently expressed concern about the ability to nest data structures to an arbitrary depth, which results in the need for much more complex query structures for annotation retrieval from repositories. Three additional complexity problems are discussed in more detail below: string literal body content, media fragments and annotation classification. In other cases, developers required a more complex structure to deal with certain technically sophisticated use cases. Three kinds of limitations in the vocabulary and properties, and in some cases best practice guidance, were identified in cases of resource multiplicity, aggregating annotations, and structured annotations. These limitations are discussed further below.

[3] http://www.openannotation.org/spec/core/core.html#BodyTarget

String Literal Body Content (Data Model too complex)

Developers expressed concern with how frequently the specifications called for use of intermediary nodes with newly minted UUIDs (Universally Unique ID).[4] A specific example is a common case of annotation of arbitrary web resource with a simple string of text.

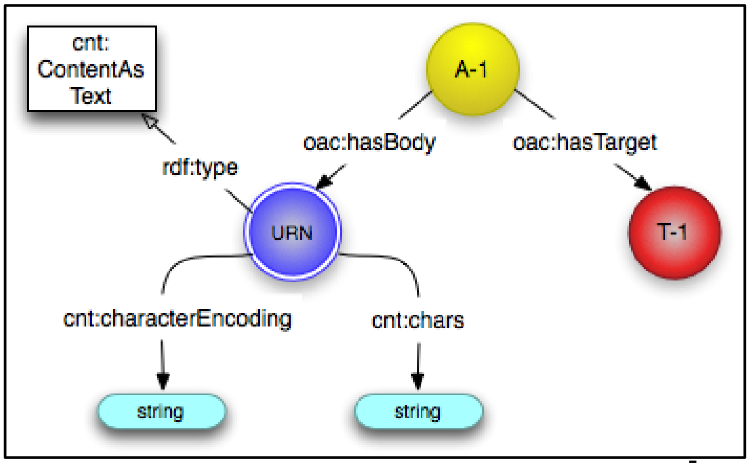


Figure 2: OAC Beta Data Model – Inline Body[5]

The OAC and OA data models (Fig. 2 & 3) call for a Body node with UUID to be created and the string to be expressed using the Content in RDF standard[6] (e.g., by typing the Body node as rdf:type cnt:ContentAsText and providing a cnt:characterEncoding string, e.g., "utf-8"). This approach allows the annotation to preserve contextual information about the string literal, for example, during ingest client software can detect what encoding standard was used and then render the string literal to the end user correctly. Some developers argued that this was unnecessary for the majority of use cases (Fig. 4) which are simply linking text content to a web resource.

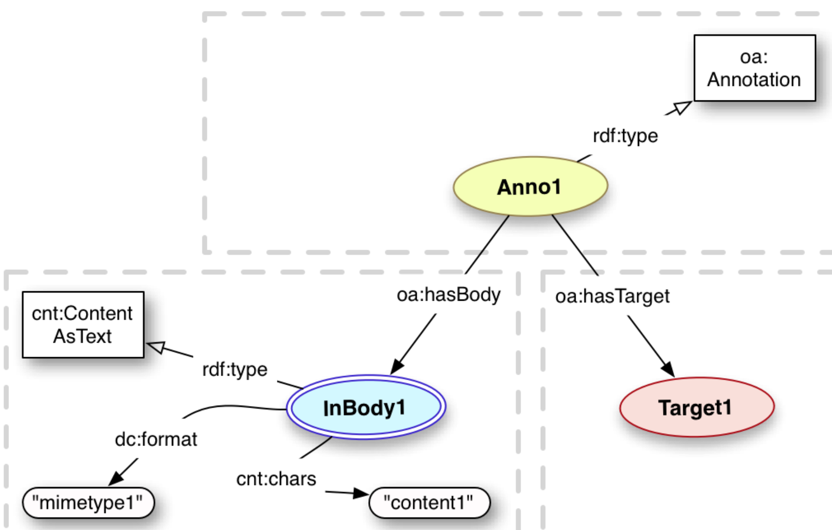


Figure 3: OA Core Data Model Draft – Inline Body[7]

---

[4] http://www.ietf.org/rfc/rfc4122.txt
[5] http://www.openannotation.org/spec/beta/
[6] http://www.w3.org/TR/Content-in-RDF10/
[7] http://www.openannotation.org/spec/core/20120509/index.html

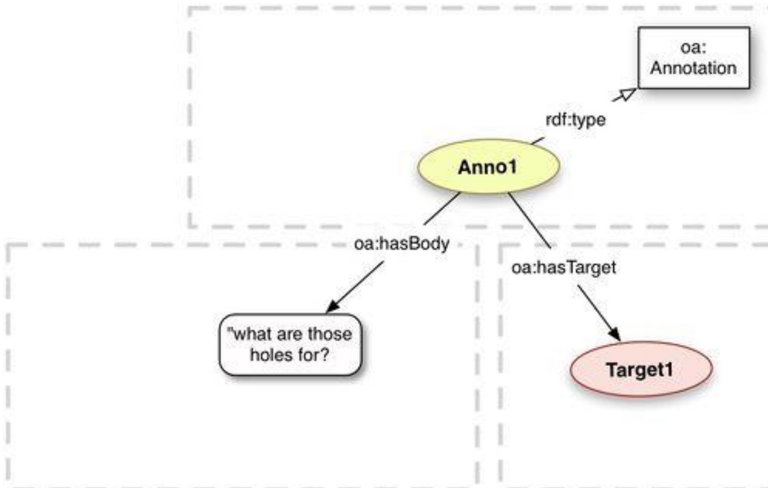


Figure 4: Common string literal annotation use case

The Brown University project team also found that the need to mint UUIDs, whenever other types of URIs were unavailable, was a barrier to building OAC (and OA) compliant annotations within their Fedora repository framework. This was due to Fedora's internal architecture, which simply does not use UUIDs to name resources. To overcome this architectural problem, developers at Brown developed a mapping layer from Fedora's PID (Persistent ID)[8] resource naming scheme, which is a local Fedora-specific resource identity scheme, to the UUIDs, a "universal" web identity scheme, used in the OAC and OA models.

Media Fragments (Data Model too complex)

In the case of annotating media, NYU developers preferred existing web architecture standards, e.g., media fragments, anchor tags, etc., over the more cumbersome OAC approach of referring separately to a specific target, source, and selector (Fig. 5, left side). They questioned why a target URI containing a media fragment could not be used as a single node replacement (Fig. 5, right side) for the specific target and fragment selector graph. The NYU team was not alone in this concern. Several of the other projects noted that using a Fragment URI presented a low barrier for implementing the standard for communities whose use case are expected to consist only of targets that can be segmented using hash URIs.

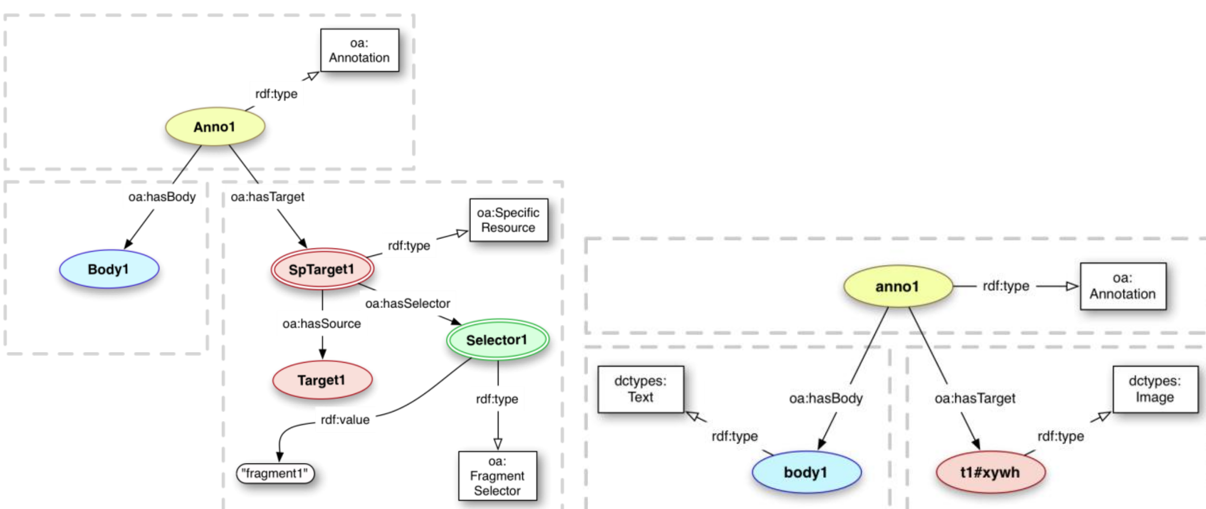


Figure 5: Fragment Selectors[9] (left) vs. proposed Fragment URIs[10] (right)

[8] https://wiki.duraspace.org/display/FEDORA35/Fedora+Identifiers
[9] http://www.openannotation.org/spec/core/20120509/index.html#SelectorFragment

Examining the two models juxtaposed in Figure 5, it is clear that the data model on the right hand side is much simpler to implement. While the Community Group eventually added Fragment URI support, the specification authors noted the following reasons why separate and distinct selectors are preferred.

- Most fragments are defined with respect to individual media types, and not every media type has a fragment specification.
- Even if a media type does have a fragment definition, it is often not possible to describe the segment of interest sufficiently precisely. For example, fragments for HTML cannot be used to describe an arbitrary range of text.
- It is not possible to determine with certainty what is being identified without knowing the media type, as the same fragment string might be possible in different specifications. For example, the same fragment string could identify either a rectangular area in an image, or a strangely named section of an HTML document.
- Fragment URIs are not compatible with other methods of describing the segment more specifically, described in the Specific Resources module of this specification. It is recognized that this additional level of description is not required in all scenarios, however.
- As URIs are considered to be opaque strings, annotation systems may not discover annotations with fragment URIs when searching by means of the URI without the fragment. For example, an Annotation with the Target http://example.com/image.jpg#xywh=1,1,1,1 would not be discovered in a simple search for http://example.com/image.jpg, even though it is part of it.

Annotation Classification (Data Model too complex)

As mentioned in the demonstration experiment profile section above, retrieval of groups of annotations is a core use case guiding development of the OA data model. Taxonomic classification schemes are frequently used to index and divide large bodies of data into more manageable pools from which query result sets can be derived. The interviews demonstrate that this practice is mirrored by at least a small portion of scholars who tag and organize their own notes for later finding and reuse.

The Queensland demonstration experiment specifically was engaged in using the taxonomic classification of annotations to facilitate retrieval of groups of annotations from an annotation repository by querying for annotation type. During the course of their experiment, developers at Queensland found that an extensible classification vocabulary can quickly become unwieldy in size. Their classification vocabulary became overly elaborate, posing complications to building efficient annotation queries. While the OAC, and subsequent OA, data models respond to the need for grouping annotations by providing a limited set of base annotation sub-classes,[11,12,13,14] scope creep remains a significant risk. Since classification vocabularies are highly customized for particular domains (and sometimes organizations or individuals) they can expand very rapidly and present a stumbling block for efficient query writing.

Resource Multiplicity [Data Model too simple]

An early use case that emerged during OAC I was the need by humanities scholars to annotate juxtaposed targets. Early versions of the OAC data model accomplished this through multiple targets (Fig. 6).

---

[10] http://www.openannotation.org/spec/core/core.html#FragmentURIs
[11] http://www.openannotation.org/spec/beta/#DM_Types
[12] http://www.openannotation.org/spec/beta/subclasses.html
[13] http://www.openannotation.org/spec/core/20120509/index.html#Motivation
[14] http://www.openannotation.org/spec/extension/index.html#Classes

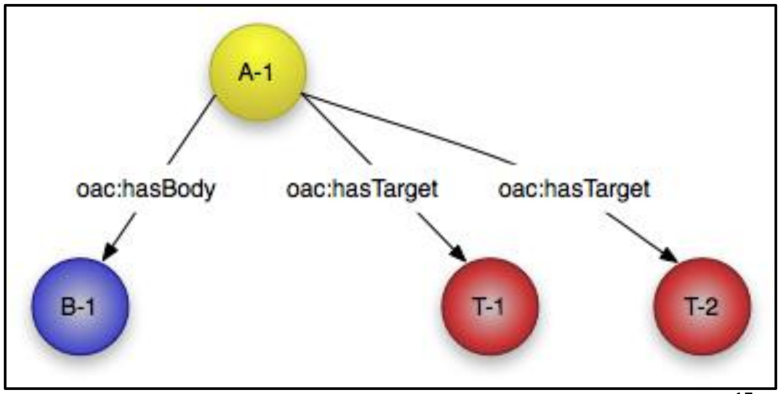
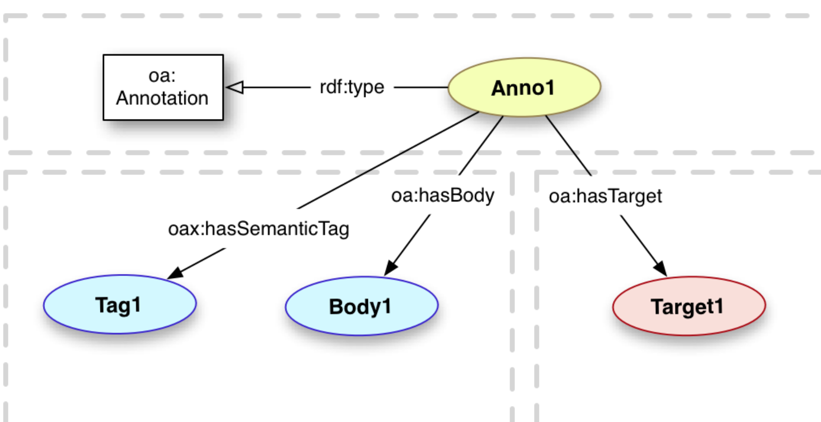


Figure 6: OAC Beta Data Model – Multiple Targets[15]

Researchers at Colorado found that the ambiguity within the data model's vocabulary, demonstrated in Figure 5, presented a significant barrier to their annotation workflows which require machine interpretable data structures. Since the Colorado team was working with a number of machine processes they needed a consistent manner in which to interpret the model's data structures, and since the meaning of multiple targets is completely ambiguous, it is impossible to develop reliably decidable machine algorithms to cope with the data structure. The ambiguity of the data model is also difficult for human developers building implementations of the data model to cope with; they have to choose ahead of time, as they build their implementation, how to interpret the multiple target situation.

Since it is unclear if the body of the annotation is targeting both targets individually, sequentially, alternatively, or as a kind of composite target, an annotation consumer has a variety of choices for how to interpret the annotation, some of which lead to further questions. For example, if the targets are interpreted as two individuals then the annotation may be more clearly modeled through two separate and distinct annotations. (e.g., Annotation A-1 where Body B-1 is about Target T-1 and Annotation A-2 where Body B-1 is about Target T-2.)

Interpreting instances of multiplicity can introduce considerable overhead during the implementation process, and if implementers within the community cannot agree on a consistent interpretation of the multiplicity semantics, then the ambiguity will significantly impact annotation interoperability. For instance the annotator's intent could be lost if the annotation consumer decided that the two targets represent a choice, e.g., Body B-1 is about Target T-1 OR Target T-2, but the annotator really means to juxtapose Target T-1 with Target T-2, e.g., Body B-1 is about the union of the two targets (T-1 ∪ T-2).



[15] http://www.openannotation.org/spec/beta/#DM_Multiple

Figure 7: OAX Extension Document – Semantic Tags[16]

Early versions of the subsequent OA Draft specification did not address the ambiguity of multiple targets but rather inadvertently extended it to annotation bodies through the introduction of the oax:hasSemanticTag property (Fig. 7). Semantic tags are a specialized form of annotation body introduced to satisfy the use cases in the Annotation Ontology science and data mining communities. The introduction of such a specialized type of annotation Body, whose cardinality varied radically from that of an ordinary annotation Body, proved to be controversial.

The savings in numbers of annotations presented by applying multiple tags to a single target resource was a key use case to the Annotation Ontology community. Further, investigators at Queensland noted that they also had use cases that called for annotations with multiple bodies, e.g., an annotation A-1 where Body B-1 is a remark in English about the Target T-1. Body B-2 is the equivalent remark in French. They needed a method to model the choice the annotation system has when displaying the annotation to the end user, i.e., the end user only needs to see the remark in English or in French but not in both French and English.

During the initial meeting of the OA Community Group in March 2012 it was established that the range of allowed resources for annotation bodies was the same as those allowed for annotation targets. Both during the course of the case study interviews and at OAC and OA Community meetings several implementers began raising questions about whether or not it might also be appropriate to allow multiple Bodies and eliminate Semantic Tags as a property distinct from an annotation Body. During the course of the OAC project review meeting, specification authors were finally convinced by the community members to allow multiple annotation Bodies. This decision compounded the ambiguity problem.

Aggregating Annotations (Data Model too simple)

The ability to group and aggregate annotations on the "canvas" was an important feature of the Shared Canvas annotation tool. In comparing the Shared Canvas implementation with the OAC and OA, developers at Meertens recommended expanding the specification with additional inline selectors to help support accommodation of aggregated annotation data structures.

Structured Annotations (Data Model too simple)

The annotation of parts of speech and gene concepts in the Colorado case suggested that the OAC and subsequent OA data models did not adequately support annotations of structured resources such as RDF graphs (Fig. 8), a key use case for data mining and linguistics annotation. Better support for annotating data structures directly would facilitate adoption of the data model by communities that regularly annotate structured data.

[16] http://www.openannotation.org/spec/extension/index.html#SemanticTag

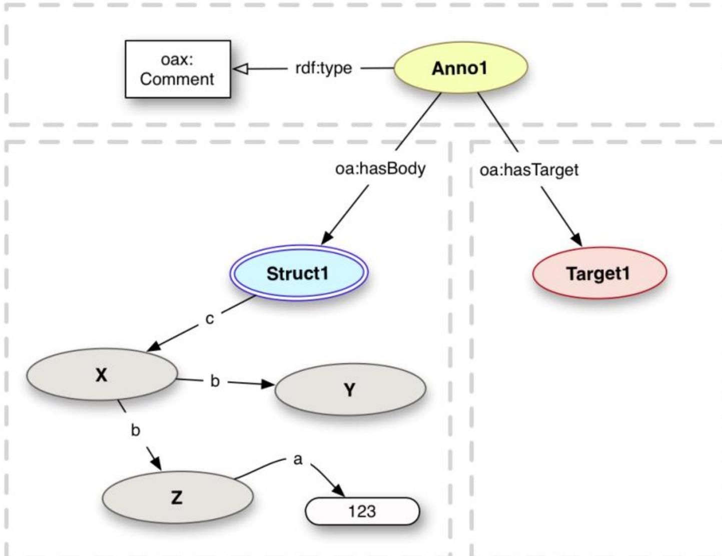


Figure 8: OAX Extension Document – Structured Resources[17]

Community Best Practices

Most participants interviewed during the course of the case study indicated that a lack of developed best practices and illustrative examples posed a barrier to implementing the OAC and OA data models. Specifically, developers desired more guidance from the lead authors of the specification and the broader community on non-information resources and content negotiation. In addition to these two specific examples, discussed further below, well developed best practices would also be highly valuable, and were sometimes sought, for resolving all of the complexity issues described above. For example, with regards to the problem of proliferating annotation classes, the Queensland team settled on a set of ad hoc best practices for when to classify an annotation and when not to. Best practice solutions for both Media Fragment URIs and String-literal Bodies were also widely discussed.

Non-information Objects

During the course of their work with text annotation, developers at Queensland found that the true nature of the target resource of many of their annotations were not necessarily the specific digital documents that the annotations were attached to but the general work that the digital documents represent, in the FRBR (Functional Requirements for Bibliographic Records)[18] sense. In the FRBR document model a document is typically represented by multiple levels of abstraction. The most abstract level is called a "Work," a work is then realized by an "Expression", e.g, a novel or play, which in turn is made tangible through a "Manifestation" e.g., an edition of that novel or play. Documents on the web are frequently modeled as the intermediate levels of a work, e.g., an expression or manifestation, which could represent a specific version of a document or a document in a specific language. The lowest level of abstraction is the "Item" level, which is frequently used to refer to specific physical documents (e.g., a specific annotated copy of Moby Dick). The question posed by the Queensland group is whether or not FRBR works and other abstract resources are valid as annotation targets and bodies. As the data model specification does not address the range of resources that are valid for annotation bodies and targets, there is a need for the OA community to examine the issue and suggest best practices.

Content Negotiation

The video annotation experiment at MITH revealed that recent innovations in web architecture can invalidate annotation target selectors by arbitrarily scaling images to the most suitable size for the end user's display. In this situation, even though the target and source resources have not changed, the segmentation conditions

---

[17] http://www.openannotation.org/spec/extension/index.html#StructuredBody
[18] http://archive.ifla.org/VII/s13/frbr/frbr1.htm

communicated by the annotation selectors will no longer apply due to the shift in scale. Developing a set of best practices for how to respond to arbitrary rescaling of resources by web clients such as web browsers would represent a method to rectify the problem without adding additional complexity to the specification.

Annotation beyond the Digital Humanities

Some of the annotation use cases, such as those with video and map content, have implications for a broad range of disciplines, however the OAC initiative was largely focused on annotation applications in the humanities. Annotation is clearly an important scholarly primitive beyond the humanities, common within all areas of scholarship and beyond. Through merged goals and work on the data model specification, the Annotation Ontology (AO) was able to bring this perspective to the Open Annotation Community Group.

A core use case AO was exploring in parallel to OAC's annotation work was the enrichment of corpuses of biomedical literature by annotating them with links to bioinformatics ontologies (Ciccarese, et al, 2010). This approach was designed to facilitate the text mining techniques increasingly employed to help researchers identify collections of documents relevant to their research needs. The semantic tag use case in the larger OA effort was developed in response to this AO use case. In the context of OAC, this issue was highlighted by the Colorado and Cornell (MapHub) sub-award demonstration experiments.

In addition to the need to support text mining applications through interlinked ontologies, biomedical scholars and scientists in general, like humanists, want to annotate their respective literatures with curatorial, discourse and other types of annotation (Ciccarese, et al, 2010; Ciccarese, et al, 2011). The recognition of annotation practices common to scientists and humanists underpinned the desire of the AO and OAC communities to merge their efforts to build an interoperable annotation data model.

Additionally, the Colorado experiment, the one science focused rather than humanities focused OAC II demonstration experiment, was already engaged with the AO team prior to their OAC sub-award. Their work demonstrated that annotation can play a more direct role in scientific workflows used as part of the data preparation process. In their OAC use case, biomedical documents are first prepared with a layer of syntactic annotations to support linguistic machine learning tools and techniques employed for:

- Sentence boundary detection,
- Tokenization,
- Part-of-speech tagging,
- Syntactic parsing, and
- Named entity recognition, specifically of gene names

The linguistic tags, tokens, and named entities are then further exploited to annotate claims within the documents (Verspoor, et al, 2011).

In their OAC experiment, the Colorado team found that this type of analysis can be greatly facilitated by introducing a new property that relates annotations to one another through evidentiary chains, e.g, Annotation 2 based upon Annotation 1 (see Fig. 9). In comparing the OA data model with the Linguistics Annotation Framework (LAF), developed as an interoperability standard by the linguistic community (Ide, Romary & de la Clergerie, 2003), they identified a number of areas of overlap and suggested a way to reconcile the two models (Verspoor & Livingston, 2012).

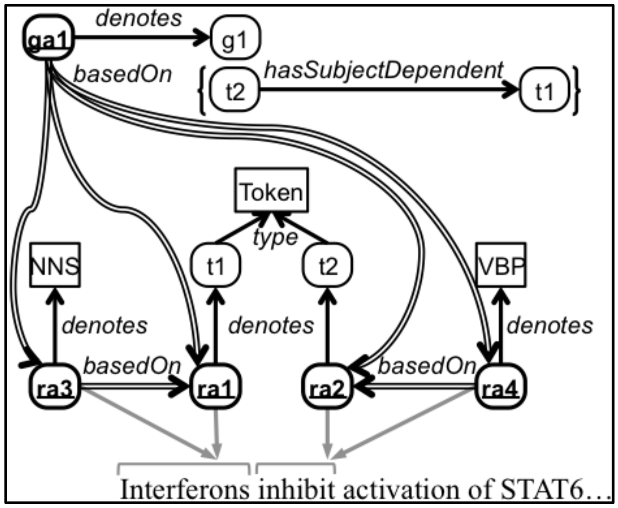


Figure 9: Annotations based on other annotations (Verspoor & Livingston, 2013)

While there have been a number of preliminary interactions, the OA Community has not yet engaged directly with the linguistics annotation community, due in part to the ready availability of satisfactory annotation tools within the linguistics community's domain. However, there are clearly important areas of overlap between the OA and LAF models. Nonetheless, the OA data model should be well equipped for the text mining applications of 21st Century humanities scholars (Underwood, 2012), since it has integrated vocabulary and properties to support science text mining annotation use cases. Future work needs to explore annotation properties that highlight, define, and record evidentiary relationships between annotations. Moreover, best practices to facilitate the use of annotations in text mining applications, either in data preparation or as data in and of themselves need to be established.

Recent Developments

During the time span the OAC II case study interviews were conducted in, the OA Community Group was constantly engaged in discussions and work on the core data model. Many of the issues and concerns reported above were discussed by the larger community of annotators and modeling decisions and best practice recommendations were put into place.

As of 13 February 2013 the 1.0 production level version of the Open Annotation Data Model specification was published by the OA Community Group. The final version of the specification addressed many of the concerns raised by project developers during the interviews and in their sub-award project final reports. Specifically, the Community Group's production level specification was able to address the following issues:

1) **String Literal Body Content:** String literal bodies were discussed at length within the OA community. The majority consensus was that the current method of embedding textual bodies should be maintained for the following list of reasons:[19]
    a. It would be inconsistent with the rest of the model which allows any resource as a Body or Target, and thus would be a special case just for text in the Body.
    b. It is not possible to represent a single property in OWL-DL as having a range of either a literal or a resource, and this is considered important for both reasoning and integration with other systems.
    c. It makes both the JSON-LD serialization and implementation more complex, as the type of the Body would always have to be determined.
    d. While literals can have their language and datatype associated with them in RDF, there are other aspects of text that are important for interpretation that cannot be associated with a literal. Examples include the media type (text/plain vs text/html), the directionality (right-to-left versus left-to-right), the encoding (utf-8 vs ascii), plus of course metadata such as authorship, date of creation and so forth.
    e. Given the previous point, the possibility for a resource to encode an embedded textual Body is essential. As the cost of using a blank node is minimal, the consistency of a single method for embedding content is deemed more important than the option of sometimes using a literal.
    f. As described in the next section, it is important to distinguish textual tags from general comments. This would not be possible with just a literal.

2) **Annotation Classification:** Sub-classing of the oa:Annotation type and construction of annotation class hierarchies was ultimately found to not scale very well and, due to variances in vocabulary practices, unnecessarily detract from the interoperability of annotation tools built to the community's specification. By promoting this type of taxonomic classification of annotations, the OAC and early versions of the OA data models ran substantial risks of collisions between terms through the creation of redundant terms and proliferation of terms within the overall vocabulary. For instance, if one community uses annotation type oax:Link to denote an annotation whose body consists of a URL linking to another web resource, while another community uses annotation type oax:Reference to represent that exact same annotation, then it becomes unclear to consuming software which annotation type is meant.

    Constraining the classification vocabulary is one obvious solution that could be contentious in a community as diverse as OA. For example, if constrained in this way, there is a risk of the vocabulary not meeting the needs of all of the community stakeholders. An alternate modeling solution, Motivation, suggested by earlier versions of the Annotation Ontology's nascent data model, was adopted (Fig. 10).

[19] http://www.openannotation.org/spec/core/core.html#BodyEmbed

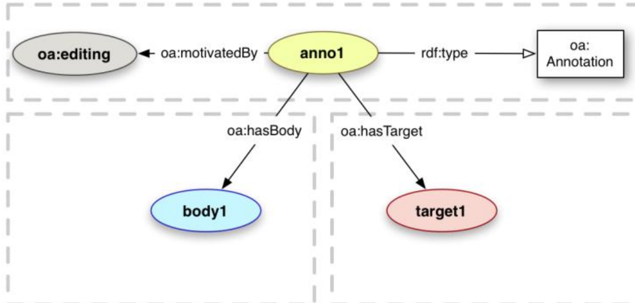


Figure 10: OA Core Data Model Final - Motivations[20]

3) **Media Fragments:** The membership of the OA Community Group did elect to provide a simplified method of targeting segmented resources that conformed to media fragments and similar fragment URI specifications (Fig. 11). This allowed easier implementation of the data model for the simplest use cases.

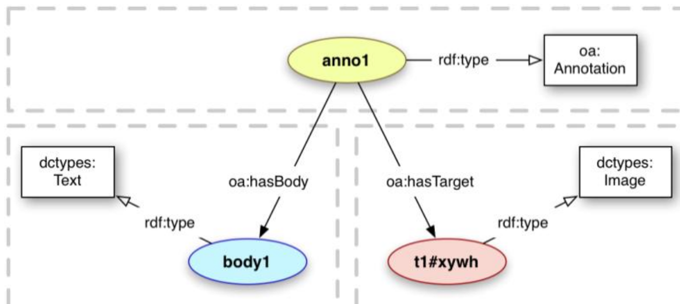


Figure 11: OA Core Data Model Final – Fragment URIs[21]

4) **Resource Multiplicity:** One of the more divisive issues confronted by the OA Community Group was the multiplicity issue. While some in the community did not believe that the ambiguity introduced through multiplicity needed to be addressed, a small group, championed by researchers at Colorado pushed on the issue. Combined with the discovery that annotation bodies and targets are completely symmetrical properties and the need to accommodate multiple selectors, especially for video annotations which must be segmented both spatially and chronologically, the OA Community Group dedicated the majority of one face-to-face meeting to resolving the issue.

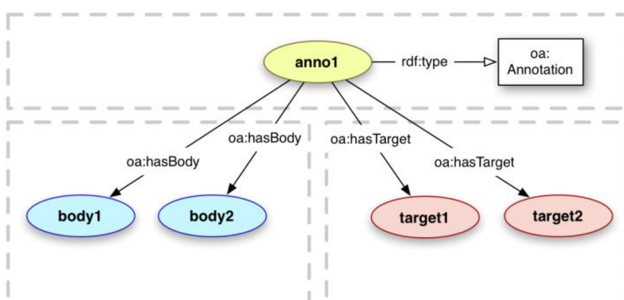


Figure 12: OA Core Data Model Final – Multiple Bodies / Targets[22]

Since the ambiguities surrounding multiplicity in resources are mostly logical ones a set of solutions conforming to logical operations have been developed (Fig. 12-15). An initial understanding was forged that multiple bodies and multiple targets referred to several annotation bodies referring individually to several

[20] http://www.openannotation.org/spec/core/core.html#Motivations
[21] http://www.openannotation.org/spec/core/core.html#FragmentURIs
[22] http://www.openannotation.org/spec/core/core.html#MultipleBodyTarget

annotation targets (Fig. 12). This relationship corresponds to the logical OR condition where a consuming agent can render any or all of the bodies and targets arbitrarily as the truth of any subset of relationships is independently true of the other subsets.

For cases where only one subset of relations can be true at a time, the oa:Choice resource classification (Fig. 13) was developed. This model conforms to XOR (eXclusive OR) relationships. This model was developed with content negotiation and cases with bodies that use different human languages to express the same intellectual content in mind.

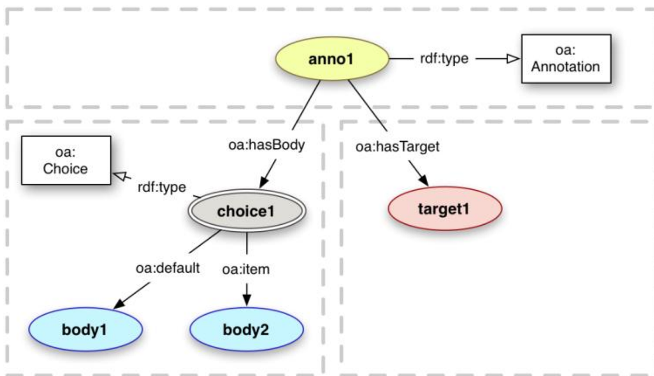


Figure 13: OA Core Data Model Final – Choice[23]

Developed for cases where the annotation relationship holds for unions of multiple resources, the oa:Composite resource classification (Fig. 14) was developed. This model conforms to logical AND (or Union) relationships and was developed to help express such resources as a composite selector, e.g., a video selector that must select both spatial and chronological coordinates, or to allow targets consisting of juxtaposed resources.

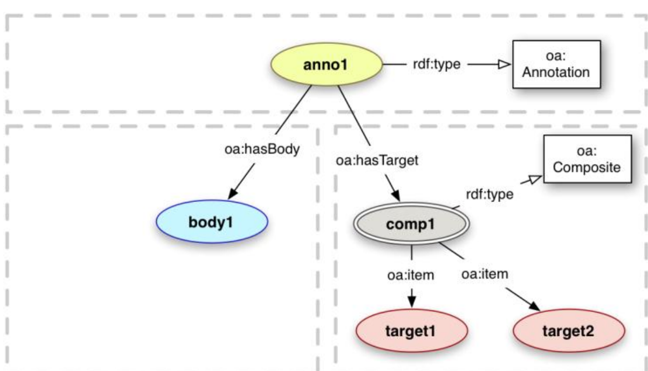


Figure 14: OA Core Data Model Final - Composite[24]

Finally, a model was developed for ordered sets of resources (Fig. 15).

[23] http://www.openannotation.org/spec/core/multiplicity.html#Choice
[24] http://www.openannotation.org/spec/core/multiplicity.html#Composite

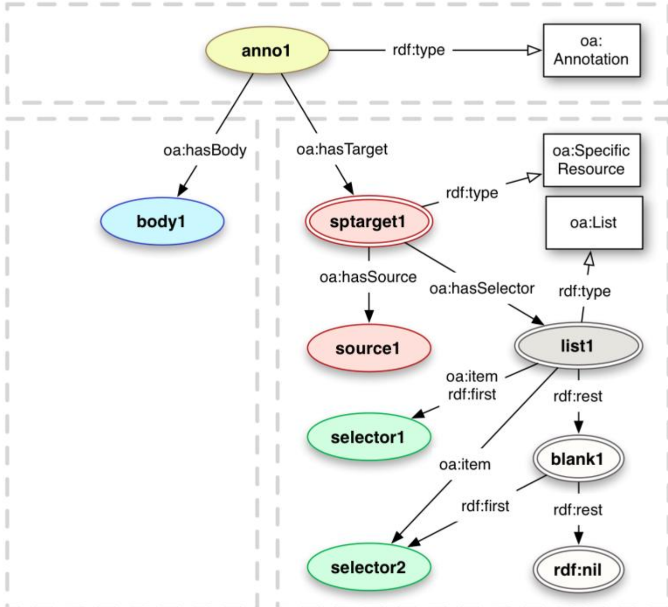


Figure 15: OA Core Data Model Final - List[25]

5) **Structured Annotations**: The OA Community Group refined and included some guidelines for annotating structured data (Fig. 16) but few of the communities members have made use of this feature of the core data model yet.

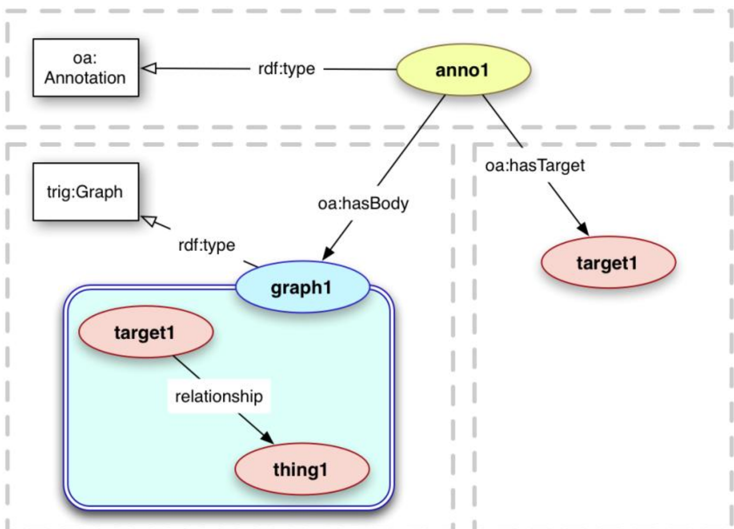


Figure 16: OA Core Data Model Final – Named Graph Serialization[26]

6) **Content Negotiation**: Rather than develop best practices to accommodate content negotiation, members of the OA Community Group elected to expand the specification's vocabulary in two ways to resolve the issue. First, support for choices (see 4 above) was added as part of the resource multiplicity discussion. The second step taken was to add selector support for machine states (Fig. 17). Only two kinds of states are currently supported; time and HTTP Request headers.

---

[25] http://www.openannotation.org/spec/core/multiplicity.html#List

[26] http://www.openannotation.org/spec/core/publishing.html#Graphs

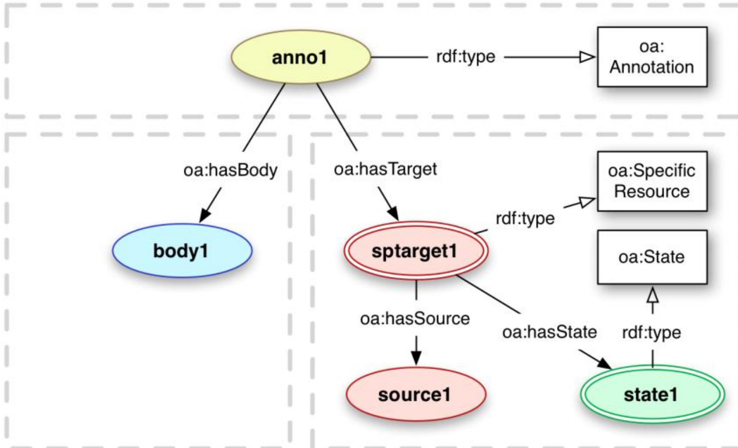


Figure 17: OA Core Data Model Final - State[27]

Other concerns that were raised by interviewees and in sub-award demonstration experiment final reports remain open issues. These include: developing support for aggregated annotations, developing community best practices (including advice on how to accommodate non-information objects such as FRBR works), and developing modeling vocabulary to support evidentiary relationships between annotations.

Summary of Findings & Recommendations

The case study interviews identified a number of issues and concerns revealed through the demonstration experiments with the annotation data model. Fortunately, the advent of the Open Annotation Community Group and additional face-to-face meetings allowed developers to resolve many of the issues through discussion with this well-rounded and active community forged from the founding OAC and Annotation Ontology communities. The developers engaged with the community group on the pros and cons of vocabulary, data structures, and use case support. Each meeting led to major agreements on critical modeling decisions, including the use of Fragment URIs, multiple annotation Bodies, how to model provenance, and methods for modeling complex multiplicity structures, such as choice, composite, and list.

While considerable progress was made by the OA Community Group, several issues raised in the interviews and the final reports from the demonstration experiment remain unresolved. They are summarized in the following recommendations and points of information, for consideration by the Open Annotation Community Group.

- Continue evolving the data model to better support non-humanities annotation communities through:
  - Development of methods and best practices for aggregating annotations, such as through ORE Aggregations;
  - Development of vocabulary and methods to support layered annotation use cases such as evidentiary relationships, e.g, by adopting some form of oa:basedOn property.

- Develop domain specific best practice guidelines.

  As use cases are frequently community and domain driven, the OAC project risked excluding significant stakeholders if the resulting data model was not complex or expressive enough, or too cumbersome to meet the needs represented across the nine projects. The overall OA Core specification is flexible enough to meet the needs of multiple domain communities, but not all of the data model's advanced features are useful for all domain use cases. Several guideline documents are needed to increase accessibility and facilitate uptake throughout the greater community of annotators:

[27] http://www.openannotation.org/spec/core/specific.html#States

    - Guidelines illustrating what parts of the data model are necessary to be minimally compliant;
    - Domain specific guidelines illustrating how to implement the data model in domain specific situations, e.g., for video annotations.

- Examine non-information resources more fully, especially FRBR "works," which have an important role in humanities initiatives.
    - Develop best practice guidelines on how to model or not model these resources, similar to the explanation developed on the community's decision not to allow string literal bodies as part of the Core Data Model.

References

- Batra, D. (2007). Cognitive complexity in data modeling: causes and recommendations. *Requirements Engineering*, 12(4), 231-244.
- Casti, J. L. (2001). Complexity. In *Encyclopaedia Britannica*. Retrieved June 6, 2013, from http://www.britannica.com/EBchecked/topic/130050/complexity
- Ciccarese P., Ocana M., Das S., Clark T. (2010). *AO: An Open Annotation Ontology for Science on the Web*. Paper presented at Bio-ontologies 2010, Boston, USA.
- Ciccarese P., Ocana M., Castro L.J.G., Das S., Clark, T. (2011). An Open Annotation Ontology for Science on Web 3.0. *Journal of Biomedical Semantics*, 2(Suppl 2). doi:10.1186/2041-1480-2-S2-S4
- Flood R. L., Carson E. R. (1988). Dealing with complexity: an introduction to the theory and application of systems science. Plenum, New York
- Ide, N., Romary, L., de la Clergerie, E. (2003). International Standard for a Linguistic Annotation Framework. In *Proceedings of the HLT-NAACL 2003 Workshop on Analysis of Geographic References*. Edmonton, Alberta.
- Niekerk K.vK., Buhl H. (2004). Introduction: comprehending complexity. In Niekerk K.vK., Buhl H. (eds.) *The significance of complexity: approaching a complex world through science, theology, and the humanities*. Ashgate, Aldershot
- Palmer, C. L., Teffeau, L., Pirmann, C. (2009). *Scholarly Information Practices in the Online Environment: Themes from the Literature and Implications for Library Science Development*. Dublin, OH: OCLC Research. Retrieved June 6, 2013, from www.oclc.org/programs/publications/reports/2009-02.pdf
- Underwood, T. (2012, August 14). Where to Start with Text Mining. Message posted to The Stone and the Shell. Retrieved June 6, 2012, from http://tedunderwood.com/2012/08/14/where-to-start-with-text-mining
- Unsworth, J. (2000). *Scholarly Primitives: What Methods Do Humanities Researchers Have in Common and How Might Our Tools Reflect This?* Presentation at Humanities Computing, Formal Methods, Experimental Practice Symposium. Kings College, London. Retrieved June 6, 2013, from http://jefferson.village.virginia.edu/~jmu2m/Kings.5-00/primitives.html
- Verspoor, K., Cohen, C. B., Lanfranchi, A., Warner, C., Johnson, H. L., Roeder, C., Choi, J. D., Funk, C., Malenkiy, Y., Eckert, M., Xue, N-W, Baumgartner, W. A. Jr., Bada, M., Palmer, M. & Hunter, L. E. (2011). A corpus of full-text journal articles is a robust evaluation tool for revealing differences in performance of biomedical natural language processing tools. *BMC Bioinformatics*.
- Verspoor, K., Livingston, K. (2012). Towards Adaptation of Linguistic Annotations to Scholarly Annotation Formalisms on the Semantic Web. In *Proceedings of the Sixth Linguistic Annotation Workshop (LAWVI) at The 50th Annual Meeting of the Association for Computational Linguistics (ACL)*. Jeju, Korea.