Harvesting Community Annotations on 3D Models of Museum Artefacts To Enhance Knowledge, Discovery and Re-Use

Jane Hunter, Anna Gerber
University of Queensland
St Lucia, Queensland, Australia
(617) 33654311
{jane, agerber}@itee.uq.edu.au

Abstract

Many cultural heritage organizations responsible for providing access to large online collections recognize the potential value that social tagging systems can add to their collections. Projects such as Steve.Museum aim to give online users a voice in describing the content of publicly-held collections of digital heritage, through online social tagging and annotation tools. However there are a number of unresolved challenges associated with re-using community tags, aggregating them within the museum's authoritative metadata stores and incorporating them within museum "metasearch" services. Although social tagging sites provide simple, user-relevant tags, there are issues associated with the quality of the metadata, the scalability compared with conventional indexing systems and a lack of interoperability across social tagging and annotation systems.

In this paper we propose an integrated system that overcomes many of the limitations of social tagging systems and maximizes their potential value-add within the context of museum collections. The Harvana system firstly enables communities to attach tags/annotations to digitized 3D museum artefacts through Web-based annotation services. The annotations/tags are represented using a standardized but extensible Resource Description Framework (RDF) model and an ontology-directed folksonomy. This approach facilitates interoperability between tags/annotations. Secondly, the system uses the Open Archives Initiative Protocol for Metadata Harvesting (OAI-PMH) API to automatically harvest the annotations/tags from distributed community servers. The harvested annotations are aggregated with the authoritative museum metadata in a centralized metadata store. The HarvANA system provides a streamlined, interoperable, scalable approach that enables cultural organizations to leverage community enthusiasm for tagging and annotation, augment their institutional metadata with community tags and enhance their discovery and browse services over 3D models.

Keywords
Social Tagging, Annotation, 3D digital objects, Harvesting, Metadata, Digital Collections, Ontology, Folksonomy

1 Introduction

Over the past few years, collaborative tagging and annotation systems that involve communities of users creating and sharing their own metadata, have exploded on the Internet. Sites such as Flickr, Del.icio.us, Connotea and LibraryThing are considered exemplary of the Web 2.0 phenomena because they use the Internet to harness collective intelligence. Such systems provide a community-driven, "organic" approach to classifying information and resources on the Web, so that they can be browsed, discovered and re-used.

In response to the explosion of social tagging tools, many cultural organizations are exploring how they might harness these tools to deliver improved online services. Organizations such as museums, libraries and archives recognize the potential value that community taggers can add to their digital heritage collections. Projects such as Steve.Museum [1] aim to give online users a voice in describing the content of publicly-held collections, through online social tagging and annotation tools. However they are also struggling with a number of technical and social challenges associated with managing both community-generated metadata and curatorial metadata.

Proponents of social tagging systems claim that because the terms used to describe the resources are community-defined, they are more topical, adaptive and relevant to users than traditional museum cataloguing systems that use complex, relatively fixed, hierarchical thesauri and authority files. Controlled vocabularies (such as the Art and Architecture Thesausrus (AAT)) do not evolve with popular language and many of the terms are irrelevant or anachronistic. Searches by non-experts often fail to yield results of relevance or that the users expect or understand. Authoritative metadata is also very expensive as it requires the time and effort of expert cataloguers. Social tagging and community annotation systems on the other hand, offer a mechanism by which the time consuming and expensive task of metadata generation can be distributed across communities. It is also argued that they provide a better measure of usefulness than software-based systems (e.g., Google) that rank resources based on the number of external links that point to a resource.

However, recent analyses of tagging data [2] have shown that the indexing terms input by untrained users are often inconsistent and inaccurate – causing resources to go undiscovered or discovered in the wrong category. There is also the problem of scalability. It is difficult to predict how Flickr, del.icio.us, and other folksonomy-dependent sites will scale as content volume escalates. Unlike hierarchical thesauri and ontologies, folksonomies don't: organically evolve into synonymous clusters; identify preferred terms; accrue into broader and narrower terms. Finally, a major limitation of social tagging systems is their lack of interoperability. Many of the popular social tagging systems are centralized, non-interoperable with other systems, don't support multiple levels of sharing and generally don't employ standards. Initiatives such as TagCommons [3] are investigating mechanisms and open standards to improve the interoperability of these popular community tools so tags can be shared across communities and resource types.

Our hypothesis (and that of the organizations involved in projects such as Steve.Museum) is that significant value can be added to a collection by augmenting authoritative metadata with community-generated metadata (tags and annotations) that has undergone some form of quality control. However there are a number of unresolved challenges associated with implementing such a hybrid approach, that deter many organizations from incorporating community-generated metadata in their metadata stores and search services. These include:

- How to carry out quality control of the community metadata without adversely impacting on the spontaneity and simplicity of the tags, and with minimal cost and effort?
- How to provide a measure of the authority of the source or author of the community-generated annotation? And how best to use this to rank retrieved results?
- How to access annotations and tags distributed over multiple different sites, systems and platforms – with different APIs?
- How to streamline the retrieval and aggregation of the community metadata with the authoritative metadata? Given that there are no standards for representing or defining tags and annotations;

- How best to exploit, distinguish and display the community metadata through enhanced search, discovery and presentation services?
- How to identify terms from folksonomies that are candidates for inclusion in existing controlled vocabularies, to improve the relevance of authoritative metadata and the success rate of discovery services?

In the remainder of this paper, we investigate solutions to the questions above, through the development of an integrated community annotation system, called HarvANA (Harvesting and Aggregating Networked Annotations) and its application to a collection of 3D models of museum artefacts.

2 Objectives

The objectives of the HarvANA system described in this paper are to identify, implement and evaluate approaches to the issues listed above and hence enable cultural organizations to exploit the full potential of community annotation/tagging systems. More specifically, our aim is to develop an efficient streamlined system (based on open standards and comprising a set of open source services) that improves discovery and reasoning services across cultural heritage collections by meeting the following objectives:

- To identify a common model for representing tags and annotations on 3D digital objects and to enhance the interoperability of tags/annotations from distributed sources e.g., different communities using different systems;
- To develop a set of easily deployable tools and services for attaching annotations to 3D objects, harvesting annotations, aggregating distributed annotations with metadata, searching and browsing aggregated metadata and presenting search results and associated annotations;
- To evaluate the system by enabling the attachment, harvesting and aggregation of annotations to a testbed of 3D museum models.

The remainder of this paper describes the testbed system that we have developed and its evaluation. In the next section (Section 3), we discuss previous related work and the technical issues that influenced our system design and implementation. In Section 4, we describe the common, extensible model that we employ for representing the annotations/tags. Section 5 describes the overall system architecture. Section 6 describes the implementation details and the user interface. Section 7 provides an evaluation of the system and finally we provide future work plans and conclusions in Sections 8 and 9.

3 Background and Related Work

Museums are increasingly employing 3D models to represent their physical artefacts digitally and enhance audience accessibility and interactivity through Web interfaces. Projects such as the EU-funded SCULPTEUR project[1] have been developing new techniques for describing, indexing and searching 3D models over the Web. In particular they have been using ontology-based metadata tools based on the CIDOC-CRM, for describing and searching collections via concepts defined in the ontology. Based on these manually-annotated collections, they have subsequently developed:

- search services that combine content-based queries (e.g., colour, pattern, shape) with text-based and concept-based queries.
- classification agents that automatically cluster art objects based on a combination of content semantics (colour, pattern, shape) and application semantics (who, what, where, when etc.) e.g., roman coins.

The SCULPTEUR system was designed for museum professionals who understand the CIDOC/CRM concepts and who require advanced search and browse tools. Although SCULPTEUR uses a combination of machine-

[1] http://www.sculpteurweb.org/

learning and manual input to generate both textual and ontology-based metadata, it does not take advantage of community-generated annotations and tags. Similarly, the MultimediaN E-Culture project [4] has been applying semantic web techniques to the annotation of 2D images of art works – and has demonstrated the improvements in keyword search performance that semantic web annotation can provide.

Our hypothesis is that cultural heritage systems such as the two described above, would be further enhanced, were they able to leverage the current enthusiasm for social tagging. Hence, the aim of the HarvANA system is to demonstrate how the collective intelligence of communities (external to the museum) can be harvested to enhance the metadata and "meta-services" for 3D models of museum artefacts. In the remainder of this section, we discuss the specific technological issues and decisions that required consideration during the design and development of HarvANA.

3.1 Web-based Annotation Services for 3D Objects

Most existing 3D annotation tools have been developed for disciplines other than the cultural heritage domain. For example, SpacePen allows a team to collaboratively work on a building design by annotating Java3D models using a Web browser and a pen-based interface for drawing suggested modifications on the model [5]. AnnoCryst [6] enables users to annotate 3D crystallography models (CIF, MOL files) through a JMOL viewer and store the annotations on a shared Web server. 3DSEAM is designed to enable 3D scenes represented using the Extensible 3D (X3D) standard to be annotated using MPEG-7 [7]. All of these systems enable users to attach annotations to 3D models and to browse annotations added by others, asynchronously. Annotations may also be attached to specific parts of a 3D model – this requires both the object view and an attachment point or 3D locator to be saved with the annotation.

However, the survey of existing systems failed to reveal any collaborative Web-based annotation systems for 3D models of museum artefacts, that enable either descriptive text or semantic tags to be attached (either to the whole object or a point or region on the object) – and then saved to enable later, asynchronous searching, browsing and response, by other users.

3.2 Architectural Models

Most Web annotation systems comprise: an annotation creation/ authoring and attachment interface; an annotation browse, search and retrieval interface; and an annotation storage and indexing component. Figure 1 provides a high-level view of a collaborative Web annotation system.

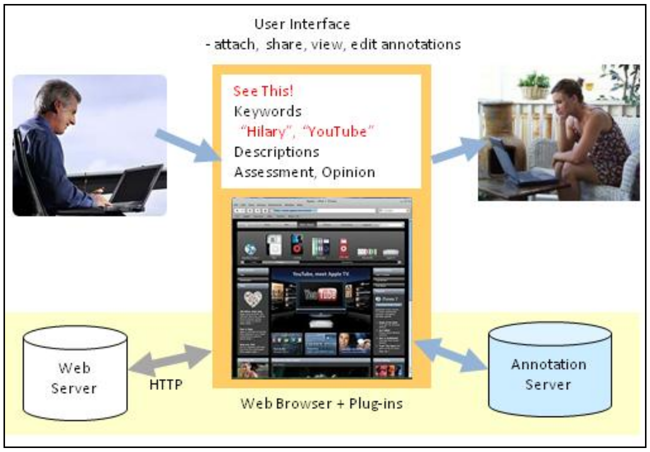

Figure 1: High level View of a Web Annotation System

The main dichotomy between system architectures, is whether the annotations are stored separately from the documents they are annotating – or whether they are stored together on the same centralized site. Flickr is an example of a centralized application that stores the resources and the tags on the same system and that requires both resource providers and users to submit their contributions via the one site. The separate storage approach calls for Web servers and data storage dedicated to annotations which include hyperlinks to the original target resources (via URIs). Although separate storage carries overheads associated with maintaining HTTP access to indexed storage repositories and web servers, and maintaining links between annotations and documents, it has a number of advantages. The decoupling of the annotations from the content allows more control and flexibility over how the annotations are accessed, processed, presented and re-used. If the annotations are on a separate server, then access, authorization and posting of responses can more easily be controlled and restricted to a particular community of users. Separation allows a single resource to be annotated in many different ways by users on the same annotation server or on different annotation servers, using different community-specific terminologies or ontologies. Separating the annotations from the resources also avoids the copyright issues that arise when having to store a copy of the digital resources on the social tagging site.

For the purposes of the HarvANA project, we assume an architecture in which the annotations are stored on a server that is separate to the server hosting the 3D models (which would typically be the museum's server). Our assumption is that there may be multiple communities external to the museum who may be annotating the same set of objects.

3.3 Interoperability and Aggregation of Annotations

The importance of being able to aggregate metadata from a range of sources has been recognized by a number of projects. Annotea [8], TagCommons [3], Steve.museum [1] have all recognized the need for standardized ways of defining annotations and tags so they can be shared between communities. However the problem of annotation aggregation is still largely unresolved. The challenge is that a central server has to download remote RDF annotations to the local server as quickly and efficiently as possible, whilst maintaining synchronization with the most recently added remote annotations and performing data merging, inferencing and querying over the integrated data sets – perhaps with the inclusion of some ranking or weighting mechanism to give more or less weight to those annotations provided by most trusted colleagues or organizations. Previous approaches to querying and aggregating distributed (RDF) annotation servers can be divided into two categories:

- RSS Feeds – using this approach, the central agency subscribes to RSS feeds from the distributed annotation servers. Because this approach involves the continual transmission of small amounts of RDF data associated with any updates, it is not controllable by the agency receiving the feeds and is not scalable when databases of considerable dimensions are involved.
- Peer-to-peer approaches - each annotation server is considered as a peer with its own local RDF database. RDF queries are broadcast to each peer and the results are aggregated on retrieval. However such systems suffer from scalability issues because of the computational burden when many users are simultaneously querying each server.

Within the HarvANA project, we propose a novel, third approach - retrieving annotation data stored on multiple distributed servers via an OAI-PMH [31] interface overlaid on top of the annotation servers. This approach involves mapping the annotations stored on an Annotea server to the museum's metadata schema and periodically harvesting the annotations/tags by having the central agency send OAI-PMH (HTTP) requests to the server. The advantages of this approach include:

- It is controlled by the harvesting organization;
- It can be configured to run periodically at set times (e.g., hourly or nightly when network traffic is low);
- Date stamps (*from* and *until*) can be used to only harvest updates since the last harvest;
- It provides a platform-independent, standardized approach that is independent of the actual tagging system or the type of resources being tagged;
- OAI-PMH has been widely adopted and has been shown to be scalable for large distributed collections.

Table 1 illustrates the relationships and mapping between:

- a museum record for a 3D model (of a cooking pot), available through the Web interface to the online collections for the Qld Museum and
- a community annotation attached to the same object, by Anna Gerber, which is stored on a remote annotation server.

This example illustrates the richness of the description that has been enabled through Anna Gerber's annotation.

	Metadata	**Annotation Record**
Identifier	**http://www.qldmuseum.com/3Dcollection/A3168**	http://maenad:8080/Annotea/OAI1/anno_14_121207
Title	Cooking pot from Roman occupation 18th century	
Creator		Anna Gerber
Date	[ca. 1750 – 1800]	12 December 2007
Description		Cooking pot or olla, from London Wall site, part of collection of Romano-British antiquities selected to illustrate the life of the people who lived on the present site of London from the Roman occupation to the 18th century ...from excavation in city of London
Subject/Keywords	Cooking pot	http://metadata.net/AAT#earthenware http:// metadata.net/AAT#crockpot
Rights	Reproduction rights owned by the Queensland Museum	Creative Commons license
Source	Queensland Museum	http://maenad:8080/Annotea/OAI1/
Type	3D model	annotation
Format	X3D	text
Relation		**http://www.qldmuseum.com/3Dcollection/A3168**

Table 1: The Relationship between the Museum metadata for a 3D model and a corresponding annotation

3.4 Quality Control

Despite the enthusiasm and hype around social tagging systems, recent analyses [3] have revealed that there is a significant degree of idiosyncrasy, inconsistency, contradiction and inaccuracy within folksonomic tags.
Consequently, significant research effort is currently focused on mechanisms by which tags can be improved and more useful formal semantics can be derived from simple community tagging systems. The aim is to optimize the trade-off between the simplicity and freedom of community tagging and the benefits to search engines of hierarchical structured vocabularies. For the HarvANA system we have adopted the "ontology-directed-folksonomy" approach. When entering tags, users are provided with suggested and popular tags from an ontology (specified at system configuration) but they still have the option to define their own unique tags. This approach is stable, flexible, ensures maximum semantic richness of the metadata and facilitates easy migration to a different community simply by changing the backend ontology. A pull-down menu provides the interface to the ontology when inputting and searching on tags. When a user searches on a parent tag, all items with the parent tag, synonym tags or children tags are retrieved. The class hierarchies are also incorporated within the tag cloud to embed multi-level structuring.

In addition, HarvANA restricts access to the annotation server via Shibboleth identity management. Our approach is to provide annotation services for closed communities with specific knowledge or expertise - rather than the general public. This reduces the proportion of incorrect, inappropriate or misleading tags and obviates the need for a moderator to check annotations.

4 A Common Model

A common extensible model for representing annotations and tags is essential to ensure compatibility and interoperability of annotation and tagging systems and the sharing and re-use of tags generated within different communities and tagging sites. Although a number of initiatives have proposed alternative models, the W3C's Annotea RDF model [8] is emerging as a defacto standard for modelling annotations and tags – having been adopted by a large array of both clients and servers. Representing the annotations/tags semantically using RDF or OWL is especially valuable because it enables the annotated resources to become accessible to the larger Semantic Web, including inferencing and reasoning engines

The Annotea model specifies the following attributes associated with an annotation:

- *keywords (*the tags or labels used to classify the artifact);
- *body* (a free-text description);
- *type* (the top-level class is *annotation*, but possible sub-classes include: *Comment, Question, SeeAlso, Reply, Ranking*);
- *creator* (the author of the annotation or tag);
- *date_created* (the date the tag/annotation was attached and published).

Extensions are possible through the addition of further optional attributes that might include fields such as: *language, media_type, format* (e.g., the annotation may be a URL, audio, image, video). Figure 2 shows how we have extended the Annotea model to represent annotations on 3D models. The *context* field enables annotations to be attached to a part of a resource. For 3D objects, the *context* needs to record both the object view (rotation and position) and the locator for the annotation (if it is attached to a point, region or component ID).

In addition, security in the form of access control can be implemented by the addition of an "access_policy" attribute, which points to a machine-interpretable XACML policy associated with the annotation itself. Such a policy specifies who can read, edit or delete the annotation. The enforcement of such a policy relies on the authentication and identification of users of the system through an Identity Management system such as Shibboleth.

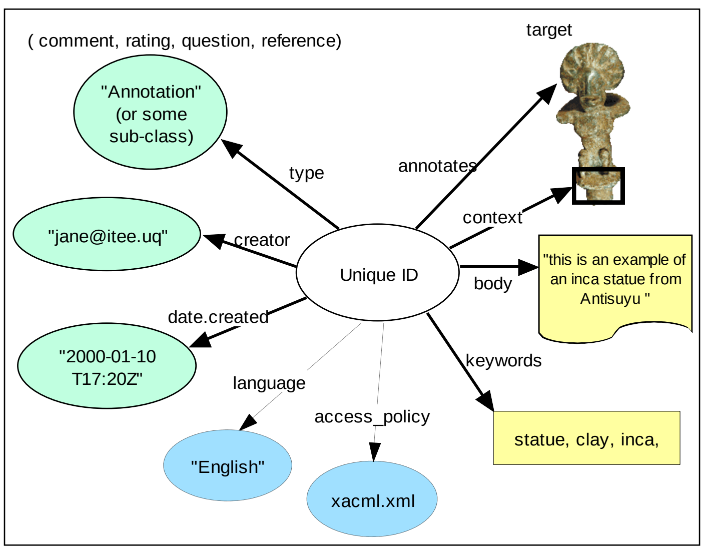

Figure 2: Annotea RDF Model showing Web Annotation of a 3D Model

5 System Architecture

Within the HarvANA system, community annotations are stored on (one or more) Annotea-compliant annotation servers that are separate from the collections that they are annotating. An OAI-PMH interface has been built on top of the Annotation server(s). This enables the periodic harvesting of new annotations (since the last harvest) by sending OAI-PMH (HTTP) requests to the server(s). The harvested annotations are then aggregated with the institutional metadata, to enrich the metadata store with community knowledge. Figure 3 provides a high-level view of the HarvANA system architecture.

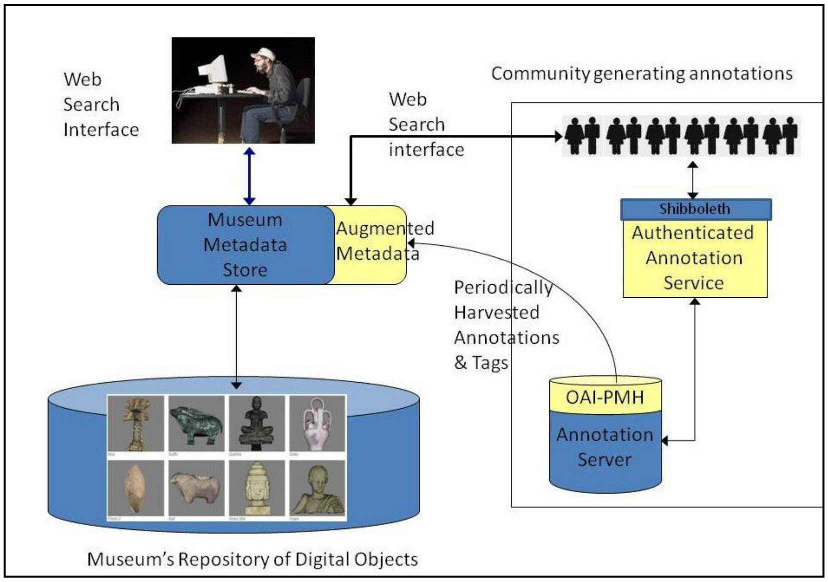

Figure 3: High-level View of the HarvANA Architecture

HarvANA uses W3C's Annotea annotation protocol [8] and an RDF Jena data store for storing and querying annotations. An Annotea client plug-in for Internet Explorer and Firefox has been developed that enables users to create and attach annotations to resources (3D models) retrieved via a Web Search Interface. The system supports the annotation of 3D objects, as well as providing a user interface for browsing and searching annotations. Users can search across annotation attributes that include: creator, date, keywords or free-text searching over the description. Quality control of the annotations is provided by validating annotations/tags against the schema and restricting tags to the specified ontology – accessible via pull-down menus within the annotation creation interface.

The Annotea server is implemented using a Tomcat Java Servlet. The RDF annotations are stored using the Jena API over a MySQL database. The OAI-PMH interface on the Annotea server was developed by mapping the RDF annotations to unqualified Dublin Core and incorporating OCLC's OAICat Java servlet within the Tomcat Java Servlet container. This enables HTTP requests to be periodically sent to the Annotea Server to retrieve any new or updated annotations as XML records. These are then incorporated within the original institutional metadata store and indexed using Lucene.

In addition, a security interface based on Shibboleth user authentication and XACML access controls has been implemented which restricts access to the annotation server (and individual annotations) to the members of specific online communities [10].

6 Testbed and Case Study

To evaluate HarvANA within the context of a museum collection, we used a set of 3D objects acquired from the Image-based 3D model Archive of Telecom Paris[2]. These models were developed using an image-based 3D modeling technique based on silhouette and stereo fusion that generates 3D virtual objects with high-resolution colour texture. The original models were available in VRML97. This is a legacy format that has since been superseded by the Extensible 3D (X3D) standard[3]. Hence we converted the models from VRML to X3D, using the VrmlMerge tool[4]. We then built a simple Web search and browse interface to a collection of 3D museum models (Figure 4). The "museum metadata" is stored in a MySQL database on a Web server at the University of Queensland.

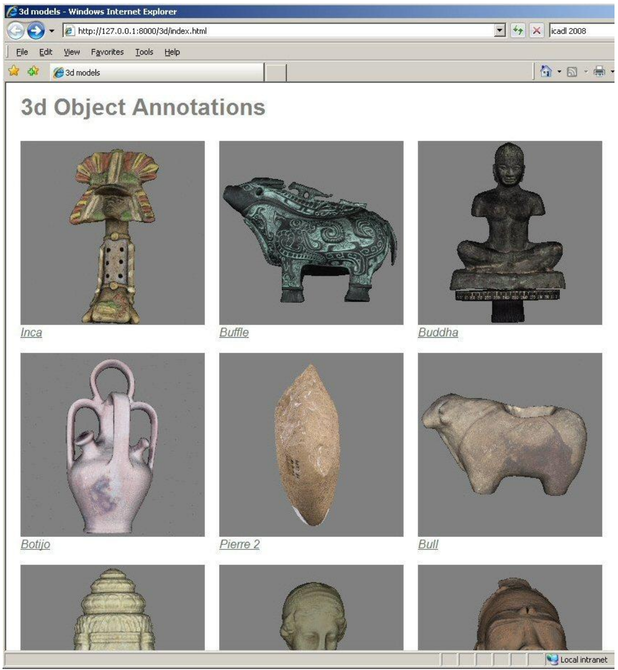

Figure 4: Browse Interface for a Collection of 3D Objects

For the creation and attachment of annotations/tags, we extended the secure Annotation sidebar developed within the DART project [10]. This is a Web plug-in that allows users to create new annotations and search and browse existing annotations through a sidebar embedded in the browser to the left of the main panel. The main panel of the browser displays the target object being annotated.

To support the attachment and viewing of annotations for the X3D models, we extended the Sidebar using the X3D Scene Access Interface (SAI) and javascript to retrieve and set the current annotation context (position, rotation, zoom values for the 3D model). To enable the display of and interaction with X3D models, we used the Vivaty player - an X3D SAI-compliant browser plugin (See Figure 5).

[2] http://www.tsi.enst.fr/3dmodels/

[3] http://www.web3d.org/x3d/

[4] http://www.deem7.com/vrmlmerge

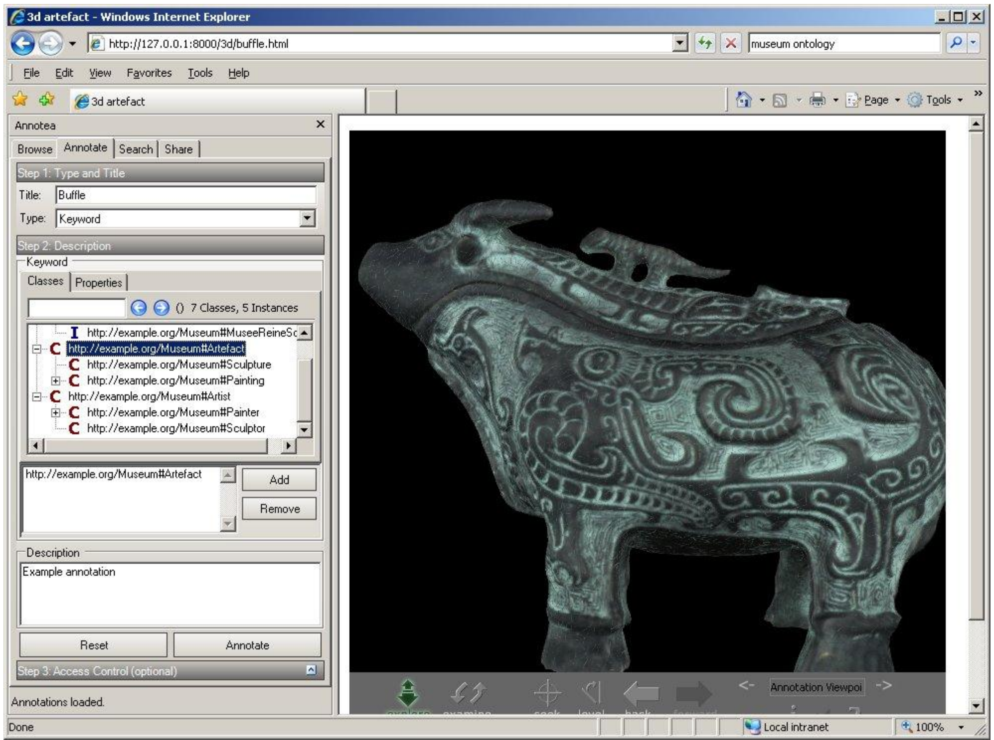

Figure 5: User Interface Showing the Annotation Sidebar and Vivaty Player within a Web Browser

We then developed a simple (OWL) ontology based on a sub-set of the Art and Architecture Thesaurus (AAT) that limits the keywords or tags to a set of controlled, machine-processable terms. The ontology (OWL file) is specified during system configuration – so each community can use their own set of preferred terms. The generic annotation creation interface is dynamically generated from the underlying annotation schema. Currently the system supports the creation of 5 types of annotation (*Comment*, *Question*, *SeeAlso*, *Reply* and *Keywords*). A *SeeAlso* annotation enables the user to specify a URL to another resource. A *Reply* is a response to a currently selected annotation – it is displayed staggered to the right of the original annotation in the list. When attaching *Keywords*, a pulldown menu is generated from the specified ontology.

6.1 User Interface

6.1.1 Creating and Editing Annotations

Figure 5 illustrates the Annotea sidebar (a client plug-in for IE and Firefox) being used to attach Keywords to a 3D Model of a Buffalo. The sidebar provides a browser-based interface for creating and attaching annotations to 3D objects retrieved via the Web search interface. When creating an annotation, users specify a *Title*, *Type* (*Comment*, *Question*, *SeeAlso*, *Reply* or *Keywords)* and the actual content (free text description, URL and/or keywords).

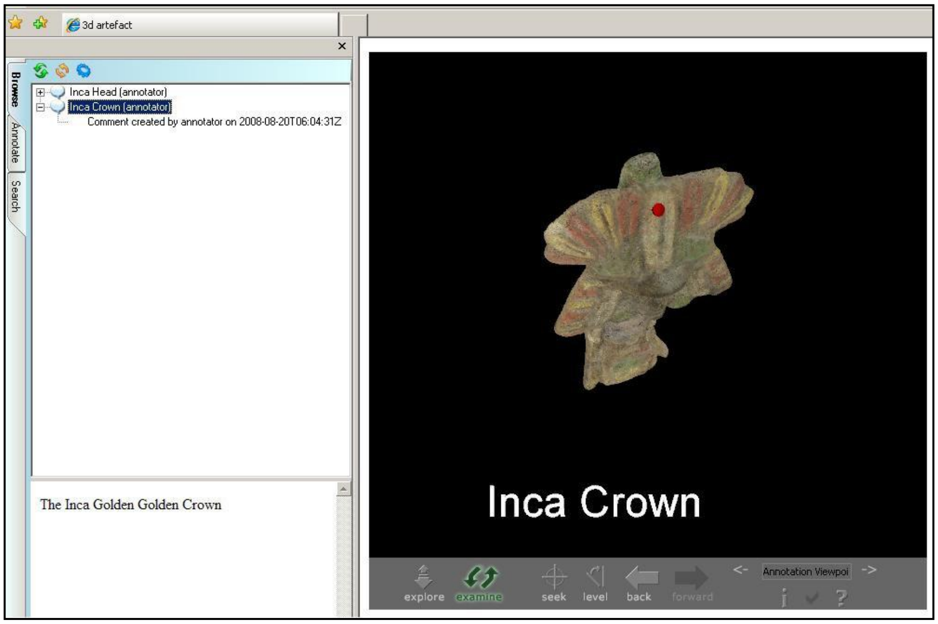

Figure 6: Attaching an Annotation to a Point/Region

Keywords or tags can either be selected/auto-completed from the pre-defined ontology or newly defined. When creating a new annotation, the user may also define an (XACML) access policy or attach a Creative Commons license. The newly created annotation is validated against the RDF model and assigned a unique persistent identifier, before being saved.

Existing annotations are displayed and can be browsed via the sidebar. Clicking on an existing annotation, causes the Vivaty player to rotate and zoom onto the stored object view and the location and title of the currently selected annotation are displayed (Figure 6). A red pushpin identifies the location of an annotation attached to a point on the 3D model.

6.1.2 The Enhanced Search & Presentation Interface

To evaluate the system, we first set up an annotation server for storing the annotations (see Section 5.1) and set about creating tags and annotations on the 3D models and storing them on the server.

The OAI-PMH annotation harvester was configured (using the Quartz scheduling library) to harvest updates to the annotation server every hour. The harvested annotation records are retrieved and incorporated within the museum metadata store (MySQL database), but are saved as "annotation records", distinguishable from the original institutional metadata records. The link to the 3D model is via the "relation" field in the annotation metadata record. Table 1 shows both an authoritative museum metadata record and an annotation record for a model of a cooking pot in the Qld Museum.

The existing web-based search interface (Figure 4) was then extended to enable users to search across only institutional metadata, only community annotations or both. Figure 7 illustrates the enhanced search interface. If the option to search community annotations is selected, users may choose to search on *tags, description, subject, title, date* or *creator* fields. For example, a user may search for all objects annotated by 'Anna Gerber'. Selecting the "more information" button beside each thumbnail displays the authoritative metadata above the annotations (in chronological order) using colours to distinguish between them.

A tag cloud showing the most popular tags is also generated and displayed at the top of the search page (Figure 7). Clicking on a tag in the cloud triggers a search for items with that tag, synonyms or children tags. When the results are displayed, a related tag cloud (showing tags that are related by subClass or equivalentClass relationships) is displayed at the bottom of the page. The search can be refined by clicking on a related tag name, or expanded to search on all related tags.

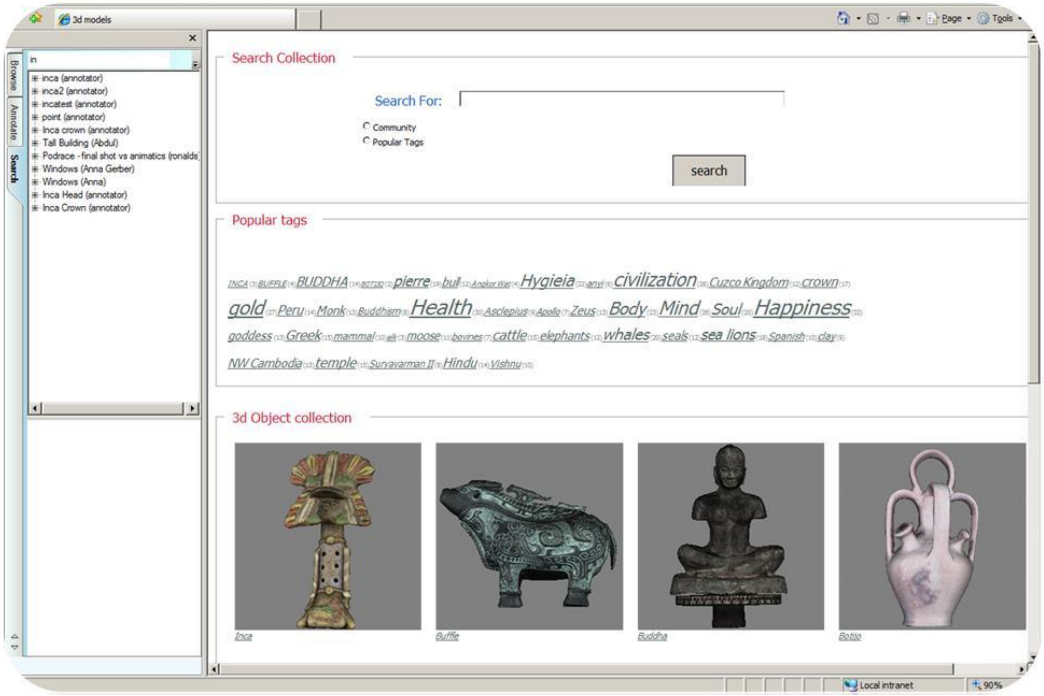

Figure 7: Enhanced Search and Browse Interface

7 Evaluation

The aim of this work is to enhance the discovery services for a museum's online collection of 3D models, by harvesting tags and annotations attached by a specific community, and aggregating them with the existing authoritative or institutional metadata to enrich it and improve findability. In order to evaluate the effectiveness of the approach, we assessed three key criteria:

- Usability of the system from an end-user point of view;
- Improvements to the search and discovery services;
- System design, efficiency and deployment from an administrative point of view.

End-users of the system comprise two types: the authenticated community members who attach the annotations and the general public who search the collection. Users attaching the annotations found the Annotea browser plug-in, easy to download, install and configure but requested auto-completion of the ontology-based tags as well as via a pull-down multi-level menu. They also requested the ability to see existing metadata and tags for the current resource they are annotating. Currently this is only visible whilst searching. They also requested the ability to annotate regions on the surface of 3D objects. Feedback from users searching the collection, indicated that addition of the tag cloud to the search interface was a positive improvement. They felt it added a sense of social interaction and 'fun' to the interface. Although, many users said that they would continue to use the free text search field. Feedback on relevance of search results indicated that the addition of ontology-based tags by domain experts, improved the discovery of relevant images, for both novice and expert users, as well as enriching the collection's inherent educational value. Feedback from a collection management perspective was provided by our museum collaborators. The museum collections managers who were surveyed, approved of the use of OAI-PMH with which they are already familiar. The use of OpenID instead of Shibboleth for the user authentication was raised. They believe there needs to be more detailed evaluation with respect to the value-add of the community annotations. The separation of the annotation servers from the centralized metadata store was seen as an advantage as it removed them of responsibility for hosting and maintaining the annotation servers. The ability to easily reconfigure the system and to add new annotation servers to be harvested was seen as a great asset.

8 Future Work

To date, we have implemented and evaluated a version of HarvANA in the context of a museum collection. This process has revealed a number of key areas that warrant further investigation and development.

- To date, we have been able to attach annotations to whole objects or points on 3D objects, but further work is required to enable the attachment of annotations to surface regions and labeled components of 3D models.
- We would like to use social network information to rank search results according to the trust profiles of annotators. For example, FOAF (Friend-Of-A-Friend)[5] trust profiles enable users to rate how much they trust their web of acquaintances on particular subjects. Ideally, the search results that match annotations of the most trusted colleagues, should be ranked higher.
- Identifying new tags that should be incorporated in the ontology. Collections managers need tools to assist them to identify which tags to incorporate and where they should be added;
- Investigating the use of SPARQL to perform semantic inferencing and querying and to derive new knowledge;
- The harvested annotations may be of significant interest to other remote communities. We are interested in using RSS feeds to stream the harvested annotations back out to these and other sites;
- Machine-learning - in recent years a number of systems have combined machine-learning techniques with manual ontolog-based annotation to index web pages, textual documents and multimedia. However they fail to incorporate or leverage community-driven social tagging systems. We are keen to investigate tripartite classification systems that combine social tagging, machine-learning and traditional museum classification approaches, integrated through common structured ontologies.

9 Conclusions

In this paper we have described a standardized approach by which social tags and annotations can be attached to 3D artefacts and readily accessed, harvested and re-used by meta-search services. A comparative analysis of our approach reveals the following advantages:

- The common RDF model for representing annotations/tags together with OAI-PMH:
  - provides a standardized mechanism by which cultural heritage web sites can expose their tag data, so that it can be harvested and re-used by metasearch services.
  - enables machine-understanding of the tags and annotations and more sophisticated semantic inferencing by meta-search services;
  - enables aggregation of tags from multiple distributed annotation servers, regardless of the underlying tagging platform or system.
- The decision to use the already widely adopted OAI-PMH and Annotea protocols provides a high level of openness, enhances interoperability and ensures a low barrier to entry;
- The adoption of Shibboleth and XACML provide a means of authenticating the source of the annotations and restricting access to, editing and re-use of the annotations based on author-specified policies. These security mechanisms also help protect against malicious taggers and tag spam;
- The “ontology-directed-folksonomy” approach (using a community-defined light-weight ontology to suggest keywords) provides quality control at time of capture, and reduces the need for post-processing without restricting the freedom of taggers;
- The flexible system design (based on a separation of the annotation servers from the museum metadata and object repository) enables easy re-configuration and customization. Administrators can easily modify the OWL ontology that defines community-specific tags or the frequency of harvesting of the annotations. They can also easily add new annotation servers to be harvested.

[5] http://www.foaf-project.org/

By combining OAI-PMH with Annotea, HarvANA delivers a scalable, interoperable method by which custodians of digital collections can effectively and efficiently leverage community enthusiasm for collaborative tagging systems. The adoption of a common model for representing annotations and a light-weight adaptable ontology for tags, provides the optimum combination of formal and informal metadata and helps to address the current weaknesses in tagging systems. By combining the best aspects of the Social and Semantic Webs, HarvANA represents an early exemplar of a Web 3.0 paradigm – a collective knowledge system.

10 Acknowledgements

We would like to acknowledge the use of 3D computer models of museum objects that were developed by Carlos Hernández Esteban and Francis Schmitt from Télécom Paris and that have been acquired from the Image-based 3D Models Archive, Télécom Paris.

References

[1] Chun, S, Cherry R, Hiwiller D, Trant J. and Wyman, B 2006. "Steve.museum: An Ongoing Experiment in Social Tagging, Folksonomy, and Museums," Museums and the Web 2006.

[2] Golder, S. and Huberman B.A. 2006. "Usage Patterns of Collaborative Tagging Systems." *Journal of Information Science*, 32(2). 198-208

[3] TagCommons 2007. Ontologies Vs. Formats Vs. Schema Vs. API. March 2, 2007. http://tagcommons.org/

[4] J. van Ossenbruggen, A.Amin, L.Hardman, et al "Searching and Annotating Virtual Heritage Collections with Semantic Web Techniques", Museums and the Web, 2007

[5] T. Jung, M. D. Gross, and E. Y.-L. Do, "Annotating and sketching on 3D web models," in 7th International Conference on Intelligent User Interfaces, San Francisco, California, USA, 2002, pp. 95-102.

[6] J. Hunter, M. Henderson, I. Khan, "Collaborative Annotation of 3D Crystallographic Models", *ACS Journal of Chemical Information and Modeling (JCIM)*. 47(5). September 2007.

[7] Ioan Marius Bilasco Bilasco,, Jerome Gensel, Marlène Villanova-Oliver, Herve Martin, "3DSEAM: a model for annotating 3D scenes using MPEG-7," *ism*, pp. 310-319, Seventh IEEE International Symposium on Multimedia (ISM'05), 2005

[8] Koivunen, M.-R. and Kahan, J., Annotea: an open RDF infrastructure for shared Web annotations. In Proceedings of the 10th international conference on World Wide Web. Hong Kong. ACM Press (2001)

[9] Lagoze, C, Van de Sompel, H., Nelson, M., and Warner, S. "The Open Archives Initiative Protocol for Metadata Harvesting, Version 2.0". June 2002. http://www.openarchives.org/OAI/2.0/openarchivesprotocol.htm

[10] Khan, I., Schroeter R. and Hunter, J. 2006. "Implementing a Secure Annotation Service", *International Provenance and Annotation Workshop*, Chicago, USA. 3 - 5 May 2006.