Collaborative Annotation of 3D Crystallographic Models

J. Hunter,* M. Henderson, and I. Khan

The University of Queensland, St. Lucia, Queensland, Australia 4072

Received May 22, 2007

This paper describes the AnnoCryst system—a tool that was designed to enable authenticated collaborators to share online discussions about 3D crystallographic structures through the asynchronous attachment, storage, and retrieval of annotations. Annotations are personal comments, interpretations, questions, assessments, or references that can be attached to files, data, digital objects, or Web pages. The AnnoCryst system enables annotations to be attached to 3D crystallographic models retrieved from either private local repositories (e.g., Fedora) or public online databases (e.g., Protein Data Bank or Inorganic Crystal Structure Database) via a Web browser. The system uses the Jmol plugin for viewing and manipulating the 3D crystal structures but extends Jmol by providing an additional interface through which annotations can be created, attached, stored, searched, browsed, and retrieved. The annotations are stored on a standardized Web annotation server (Annotea), which has been extended to support 3D macromolecular structures. Finally, the system is embedded within a security framework that is capable of authenticating users and restricting access only to trusted colleagues.

INTRODUCTION

Annotations have long been used as a tool to facilitate collaborative scholarly discourse. They enable users to attach additional information (such as comments, notes, queries, assessments, tags, and references) to electronic resources such as documents, images, or data sets and to share this information within their community. When applied to digital resources shared via the Web, they provide a very powerful collaborative tool—enabling the easy capture and wide dissemination of individual and group opinions of particular digital resources.

The main focus of the work described in this paper is the development of an annotation system for 3D crystallographic and structural models for chemical compounds. The aim is to enable geographically distributed teams of crystallographers and chemists to collaboratively discuss, compare, assess, and make comments on macromolecular crystal structures—either before or after they have been published in public online databases. The long-term storage of this commentary on an annotation server provides a method for capturing group or peer knowledge and enabling its sharing and reuse. Such a framework is of significant interest to many communities (including protein crystallography and functional nanomaterials) in which uncertainty and debate exist over the precise 3D structure of a complex protein or compound or where there is significant interest in the ensuing biological implications of such structures.

Critical to the adoption of such a system within a scientific community is the ability to authenticate the source of the annotation and to restrict access to a particular group of trusted colleagues. This is especially important in eScience[1] to protect intellectual property or maintain research competitiveness if the structure or functional description has not yet been published. In many cases, the annotation or interpretation of the 3D structure(s) may be more valuable than the target of the annotation.

Our aim was to avoid reinventing the wheel by identifying existing relevant technologies and integrating, extending, and refining them in order to satisfy the requirements of this particular application. By building a prototype that works in a standard desktop Web browser and employs tools already in common use, we will lower the barrier to entry and ensure maximum usability and add value to the current infrastructure.

Hence, our implementation combines and extends the following existing open-source technologies which are based on open standards:

- Jmol[2] − an open-source 3D crystallography model browser
- Annotea[3] − a Web-based annotation server developed by the W3C as part of the Semantic Web initiative, which we have extended to support the annotation of structural features within crystallographic models
- FOAF[4] − Semantic Web initiative for describing people and their network of relationships
- Shibboleth[5] − an Internet2 middleware initiative that enables identity management and secure access to Web resources shared among a federation of organizations
- XACML (eXtensible Access Control Markup Language)[6] − XML-based language for defining and enforcing access control policies

This paper begins by describing the background of this work and identifying previous related efforts. We then describe an example scenario that illustrates how we envisage the system will be used. Following the example scenario, we describe the system's architecture, implementation details, and user interface. Next, we present the results of the usability testing and system evaluation. Finally, we identify areas where further work is required and present our conclusions.

* Corresponding author e-mail: j.hunter@uq.edu.au.

10.1021/ci700173y CCC: $37.00
Published on Web 09/28/2007

BACKGROUND AND OBJECTIVES

Online 3D Crystal Structure Databases. Three-dimensional crystallographic and chemical structures are an important representation of scientific knowledge in many disciplines. They are highly relevant to both organic chemistry (e.g., biomolecular modeling of proteins, DNA, and RNA for drug design) and inorganic chemistry (e.g., the design of functional nanomaterials). As such, there is a large number of online crystallographic databases. The major public databases of 3D crystal structure data are highly valuable scientific resources in which extensive investment has been made. Together, the databases listed below provide comprehensive access to the majority of published crystal structures, as well as online data validation, deposition, searching, browsing, and visualization tools:

- Cambridge Structural Database (CSD)[7] for organic molecules and metal−organic compounds
- Protein Data Bank (PDB)[8] for polypeptides and polysaccharides having more than 24 units
- Nucleic Acids Data Bank (NDB)[9] for oligonucleotides
- Inorganic Crystal Structure Database (ICSD)[10] for inorganics and minerals
- The Metals Database (CRYSTMET)[11] for metals and alloys
- Incommensurate Structures Database (ICSDB)[12]

Within these databases, a variety of formats are used to represent 3D chemical structural information. The most commonly used formats are as follows:

- CIF − crystallographic information file[13]
- mmCIF − macromolecular crystallographic information file[14]
- CML − chemical markup language[15]
- PDB − original Protein Data Bank format for macromolecular structures[16]
- MDL MOL − file format created by MDL for molecular structures[17]
- CBF − crystallographic binary file[18]
- NeXus − a common data format for neutron and X-ray data[19]

The different online databases provide tools to support the preparation, validation, and deposition of different sets of file formats.

3D Crystal Structure Viewers. There are a large number of viewers available for rendering, displaying, and manipulating 3D crystal structures. The software varies from expensive, computationally intensive packages capable of calculating atomic forces to free Web browser "plugins". Despite the wide range of viewers, as far as we are aware, none of them supports the following functionalities:

- the attachment of annotations to molecules or other structural components and the storage of the annotations on a secure Web server
- the viewing of more than one structure simultaneously and the annotation of links between structures (although DeepView/Swiss-PdbViewer[20] does enable the comparison of several proteins simultaneously, it does not support the annotation of links)
- real-time collaborative viewing, discussion, and annotation of 3D structures by geographically disparate groups of users

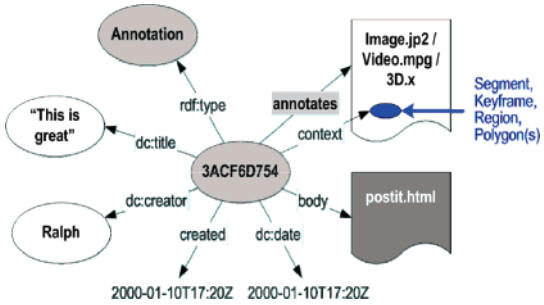

Figure 1. Extending the Annotea data model for images, videos, and 3D objects.

Rather than reinvent the wheel and build our own viewer, our aim was to identify and use one of the widely used free open-source plugins. Some of the more commonly used open-source tools include RasMol,[21] PyMOL,[22] ProteinExplorer,[23] and Jmol.[1] For the AnnoCryst system, we decided to use Jmol—a free open-source Java molecule viewer for 3D chemical structures. Jmol supports a wide range of molecular file formats, including Protein Data Bank (pdb), crystallographic information file (cif), mmCIF, MDL Molfile (mol), and Chemical Markup Language (CML). It enables users to render, pan, zoom, rotate, and spin the 3D structures It is written in Java and runs on Windows, Mac OS X, Linux, and Unix systems. There is both a stand-alone application and a development tool kit that enables easy integration with other Java applications. The most notable feature is an applet that can be seamlessly integrated into Web pages to display molecular structures in a variety of ways. For example, structures can be displayed as "ball and stick models", "space filling models", "wireframes", and either color-coded or black-and-white models. We selected Jmol as our viewer, because it is popular, is available as a Java applet that runs on multiple environments, and supports a wide range of file formats.

Annotation Systems and The Semantic Web. Annotea[2] is a Web-based annotation system that uses the resource description framework (RDF)[24] to model annotations as a set of statements or assertions made by the author about a resource. Annotations are stored remotely from the target resources on a HTTP server. This enables clients such as Annozilla[25] or Amaya[26] to query, update, post, delete, and reply to annotations. Currently, there are three publicly available implementations of annotation servers which use Annotea: W3C Annotea Server, Zannot, and PyNotea.[27] Figure 1 illustrates the RDF annotation schema used to describe various properties of an annotation including a unique id (3ACF60754), creator, title, type, date of creation, body, target, and context. A key strength of the Annotea protocol is that it uses open W3C standards such as RDF, XPointer, XLink, and HTTP. The use of RDF makes it possible to easily adapt or extend the existing scheme to incorporate additional information and support more structured annotations. For example, we may want to include metadata that describes the language of the annotation, its media type, or the type of resource it annotates. Using machine-processable RDF graphs for representing annotations offers additional advantages associated with Semantic Web technologies [e.g., RDF, Web Ontology Language (OWL),[28] and Semantic Web Rule Language (SWRL)[29]]. These include the following:

- easy linking to external but related Semantic Web resources such as Connotea,[30] FOAF,[3] or discipline-specific ontologies such as the Protein Ontology[31]
- more sophisticated, complex ontology-based searching and retrieval across annotations and related resources
- dynamic inferencing of new relationships and knowledge via a reasoning engine such as Pellet[32]

Previous work has shown that Annotea can easily be extended to enable the annotation of media types other than text. For example, image annotation has been implemented through the additional use of scalable vector graphics;[33] video annotation has been demonstrated within Vannotea.[34] Our objective was to extend Annotea again—but this time, to enable the annotation of 3D crystal structures.

Security Framework. The main focus of this work is to provide annotation tools for collaborative scientific research teams. A critical requirement for such applications is the ability to restrict access to the annotations to particular trusted groups or individuals. This is especially important within eScience, where the annotation or interpretation of the raw document or data may be more valuable than the target of the annotation. Also, by providing researchers with a robust, reliable security infrastructure, they may be more willing to engage in the exchange of views and ideas—a key to successful interorganizational collaboration. There are two levels of security associated with annotations:

- protecting the annotation server on which the annotations are stored, through some form of identity management and authentication
- authenticating the source of the individual annotations and restricting access to them through the specification of access policies that define permissable types of access (e.g., list, create, read, edit, and delete) by individual users or groups

A previous related project[35] implemented the security framework for a generic annotation system. This project used Shibboleth[5] for the authentication and authorization. Shibboleth is an Internet2 middleware initiative that enables identity management and authenticated access to Web resources shared among multiple organizations, via a single sign-on. Shibboleth is being widely adopted in higher education institutions in the U.S., U.K., Europe, and Australia for federating access to distributed repositories. In addition, XACML[6] was used for defining access control policies to restrict access to the annotations.

Additional Objectives. The primary aim of this project was to develop an open-source Web-based annotation tool for 3D crystallographic structures that allows the attachment and sharing of annotations between distributed research teams. A user requirements survey was undertaken with the Protein Structure Group at the University of Queensland's Institute of Molecular Biology (IMB), at the project onset. This survey identified the following additional user requirements:

- the ability to respond/reply to existing annotations; this provides functionality similar to a Wiki, but focused around a particular 3D structure
- The ability to browse, search, and retrieve annotations stored on distributed servers by attributes that include creator, laboratory, date, type, keywords, and free text
- The ability to link the structure to multiple associated documents (e.g., related publication(s) or structures)
- the ability to carry out a side-by-side comparison of multiple structures
- the ability to attach annotations both asynchronously and synchronously with other project team members through a shared annotation application and simultaneous audio/video conferencing.

An additional outcome of the user requirements functional specification was an example usage scenario describing how the users and developers envisaged the system would be employed.

USAGE SCENARIO

The protein crystallography group at the IMB is collaborating with the structural genomics team at the School of Molecular Sciences on a joint project that aims to solve the 3D structure of specific proteins in order to better understand their function and interaction with other proteins. A student, jointly supervised by the research leaders of both groups, has been attempting to solve the crystal structure of oxidized disulfide bond-formation protein (DSBA; 1A2J) and to compare it with the reduced DSBA crystal structure (1A2L) which has already been solved, published, and uploaded to the Protein Data Bank. The student stores the draft mmCIF file for 1A2J in a local Fedora[36] repository. She then logs onto AnnoCryst, opens the mmCIF file, renders it using the Jmol plugin, and attaches annotations (queries and comments) to those features of interest or in question and saves the annotations to the local annotation server. The default access policy specifies that her group can list the annotations and they are read-only. A Really Simple Syndication (RSS)[37] feed is automatically sent to the team members notifying them of the upload of the new mmCIF file and associated annotations. Her supervisor receives the notification, logs on to AnnoCryst, and views the draft 1A2J structure (via Jmol) and associated annotations which are listed in an annotation sidebar embedded in the browser. As he clicks on each of the students' annotations, the structure rotates and zooms to the optimum view and highlights the annotated segment and displays the annotation. The supervisor can then create new annotations that respond to the student's queries. As the supervisor responds to each annotation, his responses are stored on the server and displayed in the sidebar juxtaposed below the original annotation. Default access controls are attached to each annotation—however, these may be edited to restrict access to individual members of the team (e.g., just the student and co-supervisor). Using the browser, the supervisor is also able to search the Protein Data Bank to locate, retrieve, and view the related 1A2L structure—alongside 1A2J to enable comparison and add a link to this as one of his annotations. Each time a new annotation is attached, an RSS feed is sent to the student (as well as other team members who have subscribed and have access) notifying her that responses to her queries have been posted on the annotation server.

SYSTEM IMPLEMENTATION

Figure 2 illustrates the overall system architecture and its five main components:

1. The Jmol plugin for rendering and manipulating 3D objects, selecting features to be annotated, and highlighting already annotated features

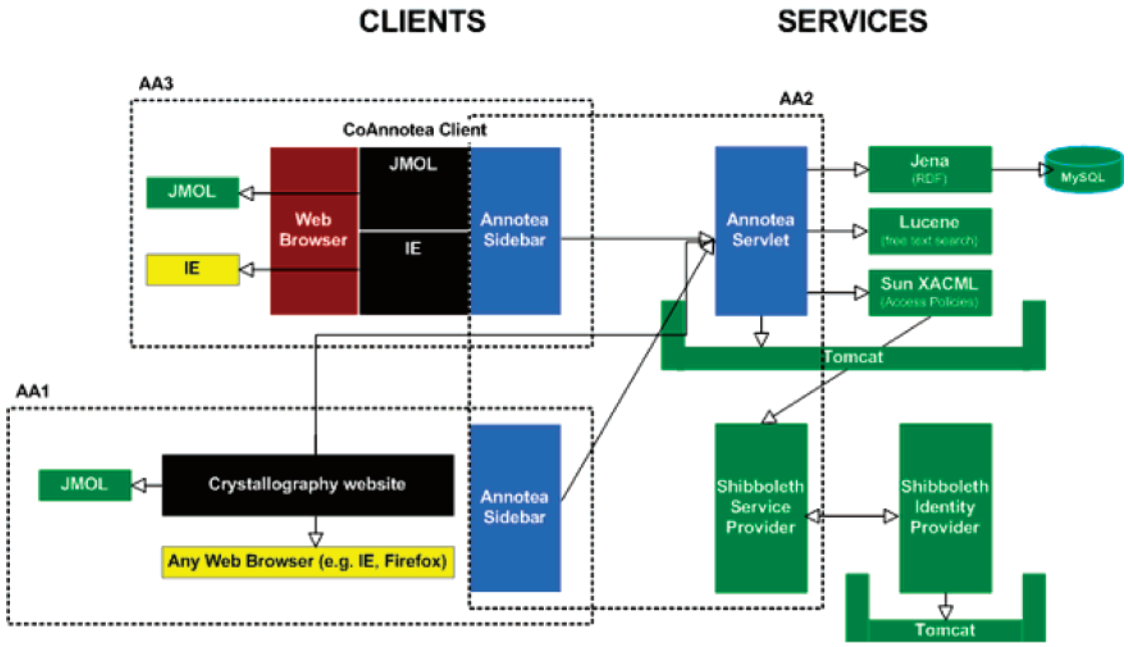

Figure 2. Architectural overview of the AnnoCryst system.

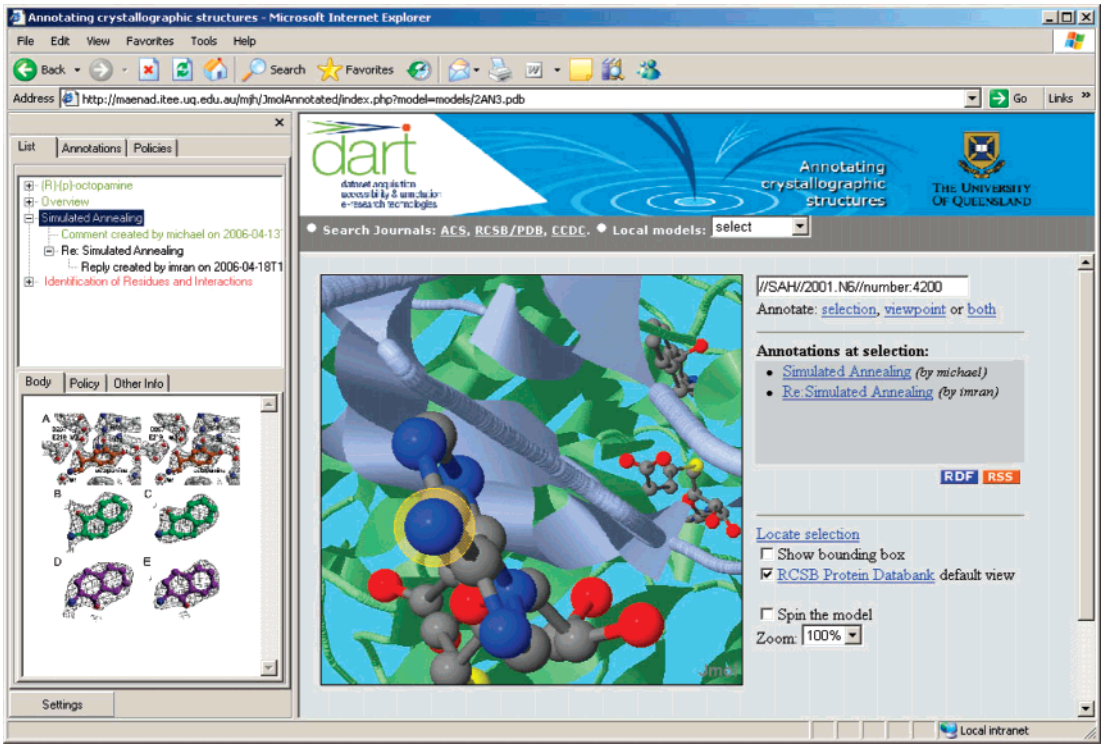

Figure 3. AnnoCryst user interface showing the Annotea sidebar (left) and Jmol plugin (center).

2. The Annotea sidebar, for listing, searching, and browsing annotations

3. The Annotation server (an extended version of W3C's Annotea)

4. The SPARQL[38] interface, for querying annotations

5. The Security framework, for authenticating users and restricting access

A screenshot of AnnoCryst's user interface (UI) can be seen in Figure 3. The UI consists of the following embedded components:

- A Web browser (Internet Explorer)
- The Jmol 3D browser (in the center)
- The Annotea sidebar (left-hand side)

The Jmol Plugin. We extended the Jmol plugin to support the highlighting and attachment of annotations. The Jmol API made this a relatively quick and easy procedure. Jmol offers a rich set of functionality for rendering and manipulating 3D crystal structures, retrieved from either local (in our case, Fedora) repositories or public online databases (e.g., PDB or ICSD). Clicking on the annotation tab on the browser menu bar opens the annotation sidebar (to the left of the Jmol window) and displays a list of annotations for this

particular structure. Clicking on an individual annotation in the list retrieves the associated context (the Jmol view, feature identifier, and coordinates of the targeted feature). This information is passed to Jmol, which rotates and zooms to the specific view and highlights the targeted feature. At the same time, the annotation content is displayed in the lower left-hand window of the sidebar. Users can create new annotations, by first highlighting a particular feature within the current Jmol view and then selecting the menu item for creating and saving a new annotation. One of the limitations of Jmol is that it currently only supports annotation at the molecular level (this is because only molecules are assigned unique IDs). Ideally, we would like to be able to select a group of molecules (e.g., those molecules that comprise a ligand or strand), assign a unique identifier, and attach an annotation to this group.

The AnnoCryst prototype also allows multiple 3D models or Jmol windows to be opened simultaneously. This enables users to compare structures side-by-side and for annotations to be added that reference a combination of these views. The system is currently limited in that only features within one view can be highlighted at a time and saved within the context. When a user selects an annotation associated with multiple structures, the system opens all of the associated views and highlights selected features.

The Annotea Sidebar. The Annotea sidebar is a sidebar plugin for Internet Explorer, originally designed to annotate standard Web pages. We extended it to support the attachment of annotations to 3D crystal structures.

Annotation Retrieval and Display. The Annotea sidebar displays a list (or staggered tree for discussion threads) of all the annotations that refer to the current active 3D file. This list appears in the top left-hand frame of the sidebar. It uses the W3C's Annotea Protocol over HTTP, to retrieve the annotations from the list of servers specified during system configuration. The sidebar only shows those annotations that the user is permitted to access. Annotations are color-coded depending on the author—this makes it easy to identify distinct author contributions while browsing. Clicking on an annotation rotates the current structure to the annotated view, highlights the annotated feature, and displays the annotation body in the bottom left-hand frame of the sidebar.

Annotation Creation. Once a particular structural feature and view has been selected, the user can create and attach an annotation to it. The "new annotation" button opens a popup window allowing the annotation metadata and body to be entered. Figure 4 illustrates the user interface for creating and posting a new annotation. We have extended Annotea to support structured annotations that contain a number of fields (title, type, rank, language, and body). The actual body of the annotation can be a hyperlink (URL), file, free text, or tag from a controlled vocabulary or ontology (e.g., the Protein Ontology). Tags are selected from a drop-down menu and populated from the associated ontology or controlled vocabulary. The system is flexible in that different annotation schemas and associated vocabularies can be accommodated—they are explicitly specified during system configuration.

After a user has finished entering the data associated with a new annotation, they can also edit the associated policy. This is described in detail in the *Security Framework* section.

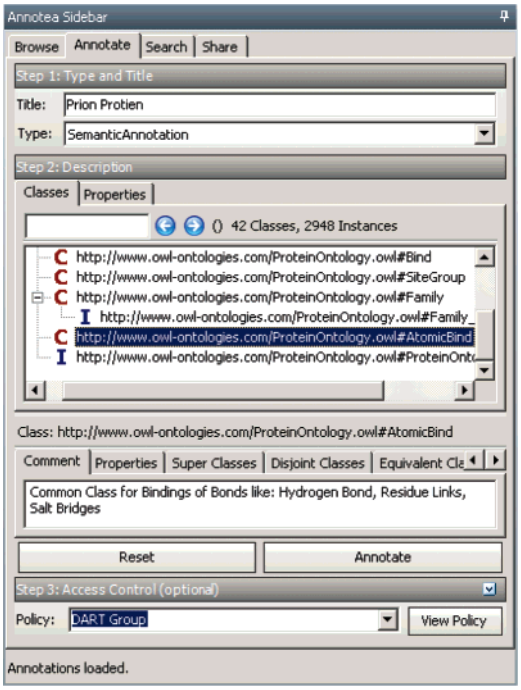

Figure 4. User interface for creating an annotation.

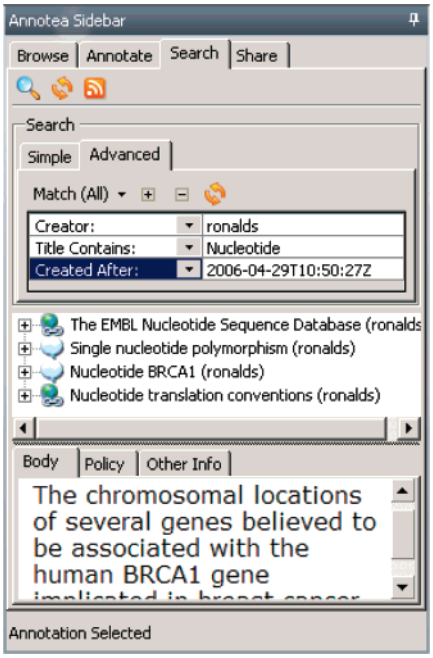

Figure 5. Search interface for annotations.

They are also able to attach a Creative Commons license[39] to the annotation. This specifies how the annotation can be reused by others.

The Annotation Server. The annotations are stored on a secure annotation server (an extended version of W3C's Annotea) as RDF statements. The system supports searching across either the metadata (*creator*, *date*, *type*, *language*, etc.) or the body of annotations, for example, "*show me all annotations by a particular author, between given dates and containing reference to a particular nucleotide*". This functionality is provided through the search interface (Figure 5), which allows users to quickly and intuitively define SPARQL queries. Although we have chosen to implement

the search interface within the browser sidebar, we are also considering the development of a Portal-style interface which does not require the installation of client software specific to Internet Explorer. Such a portal could also provide alternative methods for displaying and browsing annotations, for example, a graphical view of the distribution of annotations across a chosen collection—highlighting those resources with the greatest number of annotations or the most recent annotations.

Semantic Search Functionality. SPARQL[38] is an extremely powerful and expressive search language. However, SPARQL queries are quite complex, and direct entry requires the user to understand the underlying RDF and OWL data structures. To reduce the learning curve, we built a GUI that enables users to browse the ontology and construct a query interactively from drop-down lists.

The use of RDF enables more sophisticated SPARQL queries across the annotation server, that go beyond the statement stored in each annotation. Queries can also span statements made in related documents. An example of this is the Friend-of-a-Friend (FOAF) RDF vocabulary, which links additional information about the user's social network or colleagues. Queries can be made that incorporate this information, for example, "give me all annotations made by people who work with me". This demonstrates a major advantage of the Semantic Web: the workplace is not stored in the original data set nor in the schema, but nevertheless it can be indirectly incorporated within the queries via the FOAF link. Examples of useful queries include requests such as

- all annotations released under Creative Commons[39] license X (or some combination of license types: BY, ND, SA, or NC)
- all annotations tagged with the label "X"
- all annotations by user X within a given date range
- all annotations created by colleagues of Michael Henderson (as identified through his FOAF file)
- all annotations relevant to paper Y, derived from the Connotea tag collection

The 3D nature of the data also enables complex spatial searching. For example: show me all annotations within a particular spatial context between (x1,y1,z1) and (x2,y2,-z2).

Figure 6 shows an example SPARQL query. Each PREFIX defines a shorthand notation for referring to the ontology OWL files. Each statement within the WHERE clause indicates the specified triple pattern that must be satisfied. Values within a question mark represent variables. Results are those that satisfy the WHERE statements and the FILTER expression. The search results can also be fed into stylesheets and converted into alternative views, for example, RSS feeds.

The Security Framework. When users log on to the system, they are authenticated through the Shibboleth identity management and authentication system. A specific menu tab in the Annotea sidebar allows users to define access policies for the new annotation using XACML.[6]

Figure 7 shows the interface which was developed to define access policies. It consists of two main parts: the definition of access control rights for particular user groups and the definition of user groups (based on particular eduPerson attributes[40]). A default access policy is automatically assigned to each new annotation, but the creator can edit this. Access permissions include List, Read Only, Edit, and Delete.

```
PREFIX xml: < http://www.w3.org/2001/XMLSchema#>
PREFIX dc: <http://purl.org/dc/elements/1.0/>
PREFIX Jmol: <http://metadata.net/Jmol/Jmol.rdfs#>
SELECT *
WHERE {
?annotation dc:date ?date .
?annotation Jmol:Model <http://rcsb.org/pdb/2an3.pdb> .
?annotation Jmol:Coordinates ?coordinates .
?coordinates Jmol:X ?x .
?coordinates Jmol:Y ?y .
?coordinates Jmol:Z ?z .
FILTER
(xml:dateTime(?date)
> xml:dateTime("2006-08-01T00:00:00Z")
&&
xml:dateTime(?date) <
xml:dateTime("2006-09-01T00:00:00Z") &&
?x > 50 && ?y > 200 && ?z > 100)}
```

Figure 6. Example of a SPARQL query.

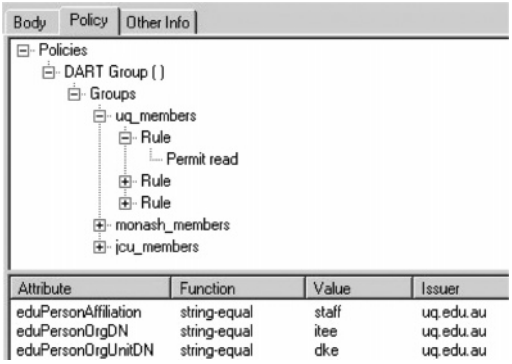

Figure 7. User interface for defining access policies.

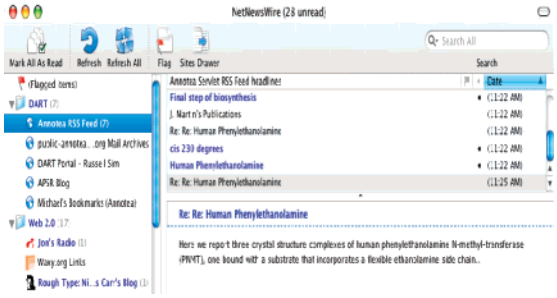

Figure 8. RSS feed (in NetNewsWire reader).

RSS Syndication and Notification. As a result of user feedback, an additional RSS[37] functionality was added to the system. RSS enables users to subscribe to a feed by entering a link of the feed into their local reader program. The reader then checks the user's subscribed feeds for new information, retrieves new content, and presents it to the user. Using RSS, users can be notified whenever new annotations, to which they have access, are added to a server. Figure 8 illustrates an example of a RSS feed from the annotation server. Users are sent the title, date, and a link to the annotation content (if they have read access).

RSS also provides a mechanism by which the authors of a structure (e.g., protein structures in PDB) can be notified, if an annotator identifies errors in their structure. Although, it may be advisable for an accredited moderator to confirm

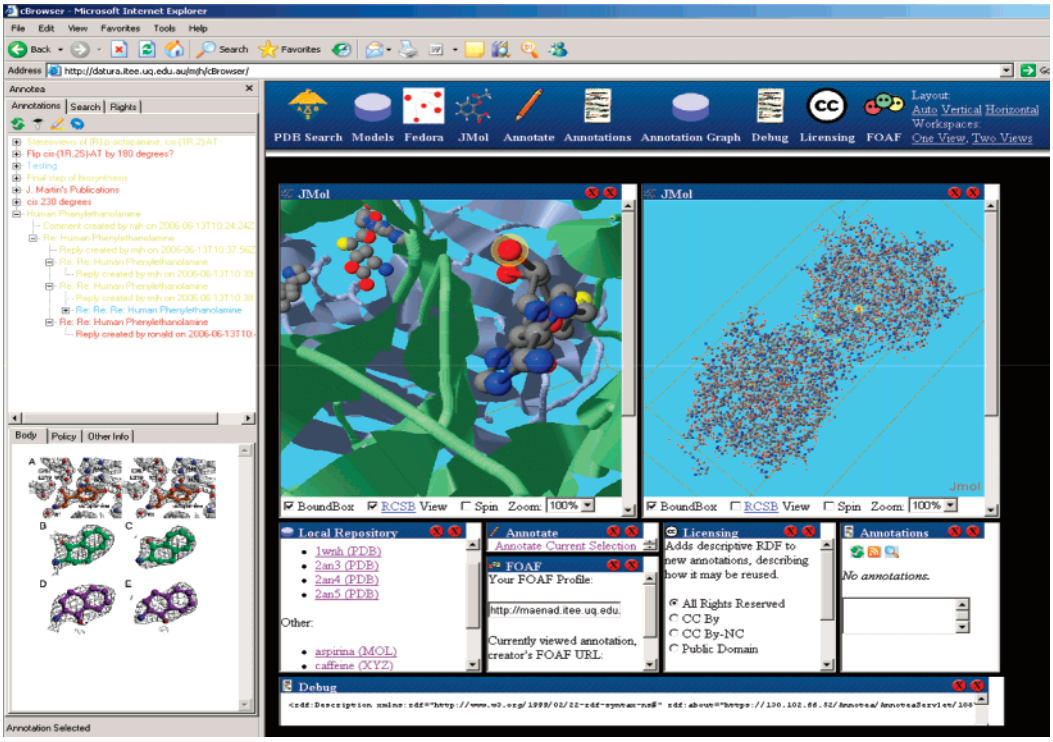

Figure 9. Using AnnoCryst to compare and annotate two CIF files simultaneously.

the validity of such annotations prior to forwarding them to the original author.

Annotating Multiple Structures. The user groups we were working with specifically requested the ability to view multiple 3D structures side-by-side for comparison and to either annotate them individually or annotate relationships between the entire structures or structural components. Hence, we extended the original AnnoCryst system by embedding a tabbed content viewer that allows multiple Jmol plugin windows to be opened simultaneously (Figure 9). We also had to extend the underlying annotation data model by creating a new class of annotations—an association. Associations annotate the link between multiple structures or between segments of those structures. Retrieving a stored association retrieves all of the structures and views linked to that annotation. This functionality is expected to be very useful for identifying and discussing protein docking mechanisms.

Synchronous Annotation. The application described here operates in an asynchronous manner. While users are browsing and contributing to a shared annotation server, they are unaware of others using the system simultaneously. Through a previous project, we developed Vannotea,[34] an application that allows real-time collaborative indexing, annotation, and discussion of audiovisual content.

Figure 10 illustrates how we have extended Vannotea to allow this kind of interaction to take place. In Figure 10, the distributed users are discussing and annotating a 3D model together with the associated diffractometry images and publication. As each participant views the 3D model, any changes, such as rotation and zoom, are mirrored instantly on each user's workstation. Jabber message passing is used to ensure precise synchronization across the shared applications. Because Vannotea uses the Annotea sidebar, annotations are created and edited in precisely the same manner as the asynchronous version. Since they are being saved in the same database, using the same markup and protocols, the annotations can be viewed in either AnnoCryst or Vannotea.

EVALUATION AND DISCUSSION

A usability study of AnnoCryst was undertaken with the structural genomics group at the University of Queensland, who agreed to trial the system and provide feedback. User feedback was very positive—users could foresee immediate benefits. They anticipate that the collaborative discussions and comparisons enabled by the system will expedite the resolution of structures and facilitate more rapid understanding of protein−protein interactions and functional behavior. The use of Jmol provides users with a familiar interface for visualizing 3D structures but, in addition, provides the capabilities necessary to rapidly develop and disseminate a structured layer of comprehensive group knowledge around these models.

The usability study and user feedback identified the following functionalities that are not currently supported but that users would like to see implemented:

- integration of AnnoCryst software with other viewers (e.g., PyMol[22] or ProteinExplorer[23])
- the ability to annotate specific structural features, for example, strands, ligands, and helixes

As the developers of the system, we were also aware of a number of additional limitations, the most significant being that the context describing the annotated region (identifier, coordinates, and view) is currently a Jmol-specific string. Community effort is required to standardize the specification of 3D regions and represent it in RDF to ensure interoperability across a wide range of 3D crystallography applications.

Users also identified the following features as significant advantages or benefits of the system:

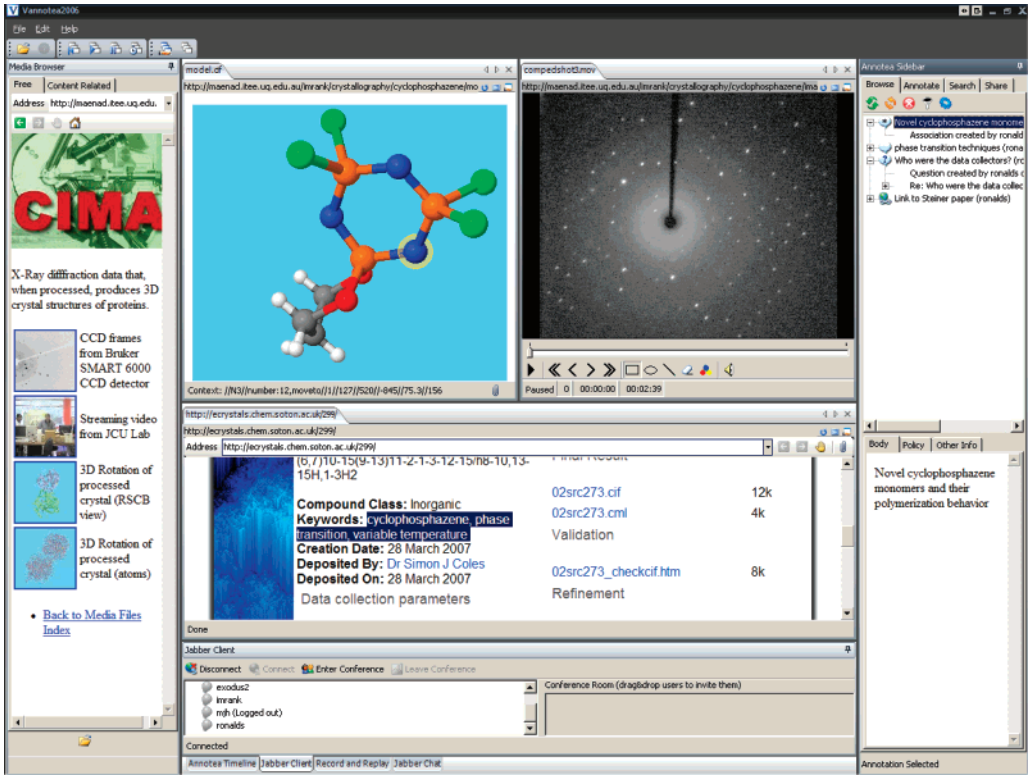

Figure 10. Synchronous discussion by multiple distributed users using Vannotea.

- the seamless interface to files stored in public online databases or in local Fedora or SRB databases
- the ability to share the annotations asynchronously and instantly; replies to existing annotations generate conversation threads or a Wiki focused around a particular structure
- users' ability to compare multiple structures by opening multiple models simultaneously and attaching annotations that compare multiple models, selections, and viewpoints
- the use of Shibboleth and XACML, providing user authentication through single a sign-on and access control over the annotations
- the use of RDF and SPARQL, enabling more sophisticated and intelligent semantic searching that can harness RDF information stored beyond the immediate annotation server, such as a user's FOAF network or external Web APIs, such as Connotea and del.icio.us
- the ability of RSS feeds to be specified in terms of a SPARQL query, allowing complex monitoring of annotated data
- the ability to attach CreativeCommons licenses to annotations that specify the rights and permissible future uses of published annotations

Comparison with Related Work. The only similar system that we are aware of is the C-ME 2D/3D annotation tool,[41] recently developed by the Collaborative Molecular Modelling Environment at the Scripps Research Institute. This is a proprietary system that uses the Microsoft Office Server System 2007 (MOSS 2007) for storing annotations, Vista's Windows Presentation Foundation for the graphics display, and Visual Studio 2005/C# as the development platform. It only supports PDB files, and authentication uses Kerberos.[42] Kerberos is limited compared to Shibboleth, which is capable of authentication across organizations or institutions.

The advantage of AnnoCryst is that it employs open-source, platform-independent components that are based on current open standards. This ensures maximum system portability, flexibility, and interoperability. It ensures that the system supports a wide range of 3D file formats (including CIF, PDB, MOL, and XYZ) and can easily be integrated within a range of popular 3D crystal structure viewers (e.g., PyMol). Regardless of their preferred viewer, users will still be able to share annotations created by colleagues who may be using different viewers. AnnoCryst stores the annotations on an extended version of the W3C's Annotea server. The annotations are separate from the actual 3D structural file and in a format that is application-independent. The annotation server is accessible through standard Web formats (XML) and protocols (HTTP), which are supported by almost every platform and programming language. The use of RDF for representing annotations maximizes their potential accessibility, reuse, and integration with external but related RDF information sources (such as Connotea, FOAF, and the Protein Ontology). This approach also optimizes the potential for future knowledge extraction. The choice of technologies also offers the benefit that XSL stylesheets can be used to quickly and easily transform annotations into different views: a visual graph, showing annotations and their replies; a tag cloud, showing the frequency of tags added to annotations; and RSS feeds, for notifying subscribers of changes to annotations.

FUTURE WORK

In the immediate future, we are planning to investigate the following issues and extensions:

- We would like to test the system within a community from a different discipline (e.g., a materials science group

at the Centre for Functional Nanomaterials who are investigating inorganic compounds).

- We would like to incorporate the same functionality within other 3D crystal structure viewers such as PyMol[22] to allow the annotation of specific features such as electron density.
- Adding annotation capabilities to protein−ligand docking programs such as Dock, AutoDock, or Gold[44] would allow the discussion and annotation of simulations of small module docking, for example, to identify and annotate likely binding sites between ligands.
- The current prototype is limited to the annotation of features uniquely identified by Jmol, such as atoms and views. In the future, we plan to enable the selection and annotation of complex composite features, such as strands, ligands, helixes, and loops.
- The current system has, to-date, only been implemented on a single annotation server. We would like to investigate the issues associated with managing multiple distributed servers hosted by different institutions but accessible to users from cross-institutional communities or research groups.

CONCLUSION

This paper describes the AnnoCryst system that we have developed to enable the collaborative annotation of 3D crystallographic models by distributed groups of users. AnnoCryst enables communities to quickly and easily create secure, structured, searchable Wikis around individual 3D structures. By harnessing the collective knowledge of a community or group, it is anticipated that the system will expedite the structural resolution process and facilitate more rapid understanding of complex 3D structures (such as proteins). It offers significant potentia as both a research and teaching tool. As related online databases continue to expand, the system will be able to seamlessly tap into and integrate these external related resources through its underlying foundation of Semantic Web technologies.

ACKNOWLEDGMENT

The DART project is a collaboration between the University of Queensland, James Cook University, and Monash University, funded by the Australian Commonwealth Department of Education, Science and Training, under the Research Information Infrastructure Framework for Australian Higher Education, from 2005−2006. The work described here represents the key outcomes of work package AA1 of the DART project. We would also like to thank the Structural Genomics group at the University of Queensland for their valuable feedback.

REFERENCES AND NOTES

(1) Hey, T.; Trefethen, A. Cyberinfrastructure for e-Science. *Science* **2005**, *308*, 817−821.
(2) Jmol: an open-source Java viewer for chemical structures in 3D. http://jmol.sourceforge.net/ (accessed July 20, 2007).
(3) Kahan, J.; Koivunen, M-R.; Prud'Hommeaux, E.; Swick, R. Annotea: An Open RDF Infrastructure for Shared Web Annotations. In *Proceedings of the 10th International World Wide Web Conference*, Hong Kong, May 1−5, 2001; ACM: New York, 2001; pp 623−632.
(4) Brickley, D.; Miller, L. Friend of a Friend (FOAF) Vocabulary Specification 0.9.2007. http://xmlns.com/foaf/0.1/ (accessed July 20, 2007).
(5) Internet2. Shibboleth. http://shibboleth.internet2.edu/ (accessed July 20, 2007).
(6) OASIS, eXtensible Access Control Markup Language (XACML), version 2.0. http://docs.oasis-open.org/xacml/2.0/ (accessed July 20, 2007).
(7) Allen, F. H. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.
(8) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(9) Berman, H. M.; Olson, W. K.; Beveridge, D. L.; Westbrook, J.; Gelbin, A.; Demeny, T.; Hsieh, S.-H.; Srinivasan, A. R.; Schneider, B. The Nucleic Acid Database: A Comprehensive Relational Database of Three-Dimensional Structures of Nucleic Acids. *Biophys. J.* **1992**, *63*, 751−759.
(10) Belsky, A.; Hellenbrandt, M.; Karen, V. L.; Luksch, P. New developments in the Inorganic Crystal Structure Database (ICSD): accessibility in support of materials research and design. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 364−369.
(11) White, P. S.; Rodgers J. R.; Le Page Y. CRYSTMET: a database of the structures and powder patterns of metals and intermetallics. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 343−348.
(12) Kroumova, E.; Luna, J. A.; Madariaga, G.; Perez Mato, J. M. Incommensurate Structures Database (ICSDB). http://www.cryst.ehu.es/icsdb/ (accessed July 20, 2007).
(13) Brown, I. D.; McMahon, B. CIF: the computer language of crystallography. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 317−324.
(14) Bourne, P. E.; Berman, H. M.; McMahon, B.; Watenpaugh, K. D.; Westbrook, J.; Fitzgerald, P. M. D. The Macromolecular Crystallographic Information File (mmCIF). *Methods Enzymol.* **1997**, *277*, 571−590.
(15) Murray-Rust, P.; Rzepa H. S.; Wright, M. Development of Chemical Markup Language (CML) as a System for Handling Complex Chemical Content. *New J. Chem.* **2001**, 618−634.
(16) Protein Data Bank, Protein Data Bank Contents Guide: Atomic Coordinate Entry Format Description, Version 3.1, July 19, 2007. http://www.wwpdb.org/documentation/format3.1−20070719.pdf (accessed July 20, 2007).
(17) Dalby, A.; Nourse, J. G.; Hounshell, W. D.; Gushurst, A. K. I.; Grier, D. L.; Leland, B. A.; Laufer, J. Description of Several Chemical Structure File Formats Used by Computer Programs Developed at Molecular Design Limited. *J. Chem. Inf. Comput. Sci.* **1992**, *32*, 244−255.
(18) Bernstein, H. J.; Hammersley A. P. Specification of the Crystallographic Binary File (CBF/imgCBF). *Int. Tables Crystallogr.* **2006**, *G-2.3*, 37−43.
(19) Klosowski, P.; Koennecke, M.; Tischler J. Z.; Osborn, R. NeXus: A common format for the exchange of neutron and synchrotron data. *Physica* **1998**, *241−243*, 151−153.
(20) Guex, N.; Peitsch, M. C. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **1997**, *18*, 2714−2723.
(21) Sayle R. A.; Milner-White, E. J. RasMol: Biomolecular graphics for all. *Trends Biochem. Sci.* **1995**, *20*, 374−376.
(22) DeLano, W. L. PyMOL Reference Guide, 2004. http://pymol.sourceforge.net/newman/refman.pdf (accessed July 20, 2007).
(23) Martz, E. Protein Explorer: Easy Yet Powerful Macromolecular Visualization. *Trends Biochem. Sci.* **2002**, *27*, 107−109. http://proteinexplorer.org (accessed July 20, 2007).
(24) Brickley, D.; Guha, R. V. RDF Vocabulary Description Language 1.0: RDF Schema, W3C Recommendation, 2004. http://www.w3.org/TR/rdf-schema/ (accessed July 20, 2007).
(25) Annozilla (Annotea on Mozilla). http://annozilla.mozdev.org/ (accessed July 20, 2007).
(26) Amaya. http://www.w3.org/Amaya/ (accessed July 20, 2007).
(27) Annotea-based Annotation Servers. http://www.w3.org/2001/Annotea/Projects#servers (accessed July 20, 2007).
(28) McGuinness D.; van Harmelen F. OWL Web Ontology Language Overview, W3C Recommendation, February 2004. http://www.w3.org/TR/owl-features/ (accessed July 20, 2007).
(29) Horrocks, I.; Patel-Schneider, P. F.; Boley, H.; Tabet, S.; Grosof, B.; Dean M. SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission, 2004. http://www.w3.org/Submission/SWRL/ (accessed July 20, 2007).
(30) Lund, B.; Hammond, T.; Flack, M.; Hannay, T. Social Bookmarking Tools (II): A Case Study − Connotea. *D-Lib Magazine* **2005**, 11. http://www.dlib.org/dlib/april05/lund/04lund.html (accessed July 20, 2007).
(31) Sidhua, A. S.; Dillon, T. S.; Chang, E. Protein Ontology Project: 2008 Updates. In *9th Data Mining and Information Engineering*; Zanasi, A., Brebbia, C., Ebecken, N. F. F., Eds.; WIT Press: Southampton, U.K., 2008.
(32) Sirin, E.; Parsia, B.; Cuenca Grau, B.; Kalyanpur A.; Katz, Y. Pellet: A practical OWL-DL reasoner. *J. Web Semantics* **2007**, *5*, 51−53.

(33) Scalable Vector Graphics (SVG) Full 1.2 Specification W3C Working Draft, 2005. http://www.w3.org/TR/SVG12/ (accessed July 20, 2007).
(34) Schroeter, R.; Hunter, J.; Guerin, J.; Khan I.; Henderson, M. A Synchronous Multimedia Annotation System for Secure Collaboratories *2nd IEEE International Conference on E-Science and Grid Computing*, Amsterdam, December 4−6, 2006; IEEE Computer Society Press: Washington, DC, 2006; p 41.
(35) Khan, I.; Schroeter R.; Hunter, J. Implementing a Secure Annotation Service. *Proceedings of International Provenance and Annotation Workshop*, Chicago, 2006; ACM: New York, pp 212−221
(36) Staples, T.; Wayland R.; Payette, S. The Fedora Project: An Open-source Digital Object Repository System. *D-LIb Magazine* **2003**, *9* (4).
(37) Winer, D. RSS 2.0 Specification, 2003. http://cyber.law.harvard.edu/rss/rss.html (accessed July 20, 2007).
(38) SPARQL Query Language for RDF, W3C Candidate Recommendation, June 2007. http://www.w3.org/TR/rdf-sparql-query/ (accessed July 20, 2007).
(39) CreativeCommons. http://creativecommons.org/ (accessed July 20, 2007).
(40) Internet2 Middleware Architecture Committee for Education, Directory Working Group (MACE-Dir), EduPerson Object Class Specification (200604a), May 2007. http://www.nmi-edit.org/eduPerson/internet2-mace-dir-eduperson-200604.html (accessed July 20, 2007).
(41) Kolatkar A. C-ME: a smart client to enable collaboration using a two- or three-dimensional contextual basis for annotations. Presented at Microsoft eScience Workshop, October 2006.
(42) Clifford Neuman, B.; Ts'o, T. Kerberos: An Authentication Service for Computer Networks. *IEEE Commun.* **1994**, *32* (9), 33−38.
(43) Taylor, R. D.; Jewsbury, P. J.; Essex, J. W. A review of protein-small molecule docking methods. *J. Comput.-Aided Mol. Des.* **2002**, *16*, 151−166.

CI700173Y