Yan Wang

LIS592: Practicum

Director: Prof. Tim Cole & Prof. Allen Renear

6 Sep 2009

Annotation Data Model and Implementation Research: Analysis & Experimentation with the Annotation Tool-Pliny

I. Introduction

A major aim of the larger research project of which this practicum is a part is to show the potential to adapt existing annotation applications and optimize existing content repositories to exploit a shared annotation data model and interoperability specification. There are several important research issues requiring investigation. The guiding research questions are "How does the web architecture principles and needs of scholarly practice inform and guide the development of the shared data model of annotation?" And, "How should the data model address the usual issues of completeness, expressive power, computational complexity, etc? Particularly relevant to development of an interoperable annotation data model are issues to do with managing generality, representing reference as well as anchoring, and accommodating community-based annotation classification schemes and taxonomies." (The Open Annotation Collaboration, 2009) There are seven draft use cases made public through OAC wiki page in summer 2009 which represent a range of common scholarly practice involving annotation. These initial use cases illustrate how the project is to be grounded in existing practice. A more complete and fully realized set of use case scenarios involving existing annotation applications and suitable for concrete demonstration during Phase 2 of the project is being developed.

My overall practicum was undertaken to support the OAC project and included a literature search to help develop an understanding of how scholars use annotation, both in traditional (i.e., print) context and as emerging in digital context. For my practicum project, I focused on an examination of one particular scholarly annotation tool, Pliny. My intent was to learn how and where Pliny records and stores various kinds of notes (i.e., annotations); how to extract the annotations programmatically rather than via the Pliny's graphical user interface (GUI); and how to write annotations in the same way. Also I defined cross-walk mappings between Pliny and Annotea RDF data model[1]. My overarching research questions, which are inter-related, were:

[1] http://www.w3.org/2002/12/AnnoteaProtocol-20021219

- How scholars using Pliny implicitly define annotation and what facets do they assume annotations have?
- What functionalities is Pliny providing?
- Are these functionalities able to satisfy the major categories of annotation tasks of scholarly practices?
- What information about annotation is considered as important and how is this information stored in annotation application tool- Pliny?
- Based on what I learned analyzing Pliny, how should the larger project approach analyzing other annotation application tools those might be good candidates to support annotation interoperability?

To support my practicum project and my practicum more generally, I participated in conference calls and on-campus team meetings and presentations to do with the OAC grant. I also expanded my readings of the topics, beginning with a systemic review of current scholarly annotation tools and application[2]. I also contributed as a member of the project team involved in synthesizing and using these technologies and developing and refining use cases.

As part of my practicum I performed a literature search, assembled and augmented topical bibliographies, and participated in in-depth experimentation with one annotation tool -- Pliny. Through literature reading of the current scholarly understanding of annotation and examining Pliny and crossing walking Pliny annotation into Annotea RDF, I gained a sense of how to provide detailed descriptions of exsiting data models and implementations with regards to an interoperable data model and provided a basis for a conceptual examination of how such models might be extended.

II. Prior Art

A great deal of research has already been performed on trying to define the scope of annotation and objectives or purposes of annotation. The previous research shows there are still lot of open issues to do with how we understand annotation and how we migh best aproach the development of annotation application tools. Renear (1999) presented an outline of a taxonomy of annoation which illustrates six major categories of annotation regards to the functions in both traditional and digtial content. He pointed out annotations can be categorized as:

- Recording and Schdueling Reading
- Basic Highlighting

[2] https://apps.lis.illinois.edu/wiki/display/openannotation/A+Few+Recent+Papers+To+Do+With+Annotation

- Commentary
- Classifycation
- Copyediting,Editing, Joint-Authoring
- Speech Acts

This outline offers a basic understanding of various annoation activities in both traditional and digital context. As information infrastructure that has shifted from a Print-Post Office nexus to a hybrid system in which electronic computer networks bulk ever larger, annotation in digtial context becomes more common. Meanwhile, the funcationalities of annotation in digital area can be more powerful, as they can be manipulated, shared, filtered and so on.

In contrast to Renear, Jane Hunter (2009) summarized key features for different annoation system. She examined the types of content that each system was designed to annotate and the annotation structures and formats supported, as well as how each system supported annotation authoring.

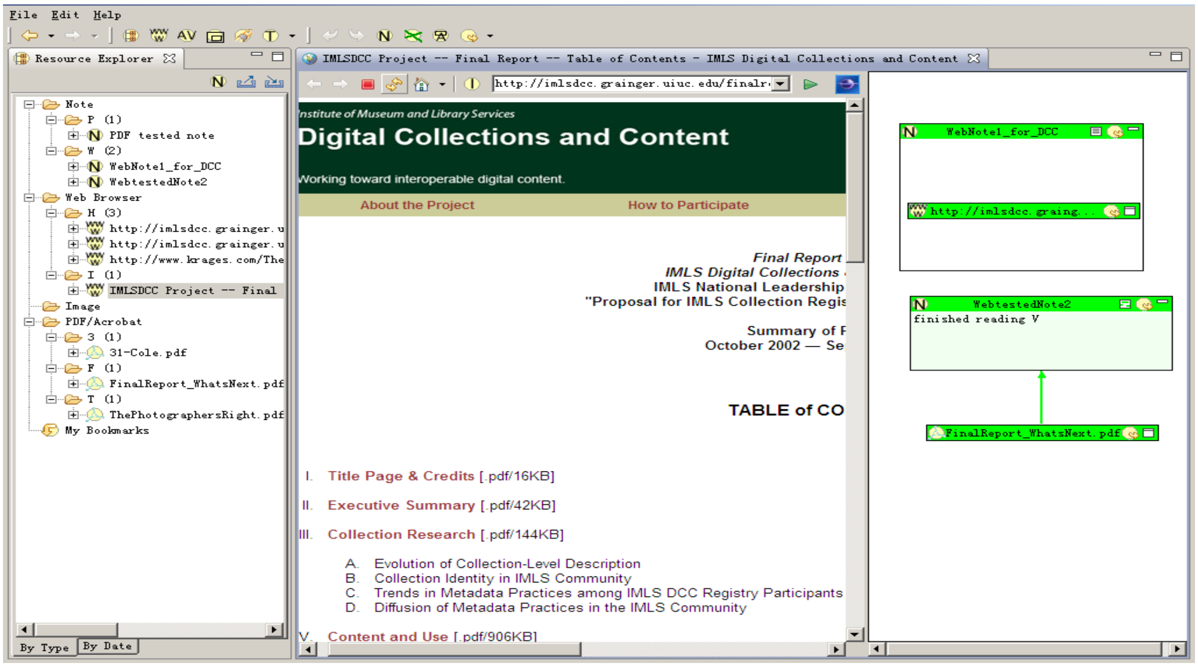
Though the research on anotation model is not new, work on the granularity of understanding the nature of annotation and how users do annoation results in developing annoation systerms only for specific purposes (Agosti, 2007) is only now emerging. For example, text processors like MS word, PDF annotator and so on are designed to support private local annotation. So does Adobe Photoshop for private annotatation of images. However, there are also annotation system such as Annotea and Flickr that are designed to support collaborative annation. Agosti (2007) presented a formal model of annotion in digital content which features a five-layer object model including indentification, cooperation, linking, semantics and materialization. Each layer has serveral elements needing our attention when we think of digital annotation, such as user and user group, author, permission in co-operation layer, unique handle of annotaiton and digital object in identification layer, meanings graphics and meaning of annotation in sementics layer and so on. More and more recent scholarly published papers discuss annotation in digital context. The significance of exchanging and sharing annotation has been highlighted in the growth of digtial scholarship. The W3C Annotea Project established in 2002 is a representive system based on reource decription framework (RDF) infrastructure, and relying on HTTP and XML for manipulating sharable annotations of web contents. (Annotea Protocols, 2002) It allows users to attach data to web pages and store this data in one or more annoation servers. The protocal supports posting, quering, downloading, updating and deleting client-server interactions for annotation creation and collaboration.

Going beyond pior work, OAC is trying to systematically anynalize of current annoation models from multi-perpective such as application design, system architectures, scholarly-focused use cases and so on in order to build a framework for advancing and leveraging annotations across boundaries of annotation clients and servers, and content collections.

III. Analysis Part 1: Examination of Pliny – a Scholarly Annotation Tool

So first, to understand practical issues that will be encountered in trying to share annotations, my approach was to examine a specific existing annotation tool, Pliny. In the next section I then compare this analysis to the general annotation interoperability model proposed by Annotea and see how they relate.

Pliny was evaluated as an example of a scholarly annotation application. The exact definitions used by Pliny are fluid and malleable. It is not a Web application in the sense that it is not a Web service that responds to requests from other computers -- but it is a Web-aware application in that it is designed to interact with Web resources. In some sense, It’s more like a personal software application, in which users can organize and collect resources and notes of immediate interest to their work. Since Pliny allows annotation of websites, PDFs and images and provides an environment for scholars to deal with their annotations- developing interpretations of the resources they are researching; organize their notes; make connections of difference resources. It aims to “bring the benefits of computing to existing scholarly practice rather than requiring the user to adopt new strategies to benefit." (See Wang and Cole’s PowerPoint introduction to Pliny in the Appendix). Pliny is a three-frame interface providing a separate frame for each of a hierarchical resource explorer (map), a resource viewing area, and a notes area. (See Fig.1)


File Edit Help
Resource Explorer
IMLSDCC Project -- Final Report -- Table of Contents - IMLS Digital Collections and Content
http://imlsdcc.grainger.uiuc.edu/finalr
Note
P (1)
PDF tested note
W (2)
WebNote1_for_DCC
WebtestedNote2
Web Browser
H (3)
http://imlsdcc.grainger.u
http://imlsdcc.grainger.u
http://www.krages.com/The
I (1)
IMLSDCC Project -- Final
Image
PDF/Acrobat
3 (1)
31-Cole.pdf
F (1)
FinalReport_WhatsNext.pdf
T (1)
ThePhotographersRight.pdf
My Bookmarks
By Type
By Date
WebNote1_for_DCC
http://imlsdcc.graing...
WebtestedNote2
FinalReport_WhatsNext.pdf

finished reading V

Institute of Museum and Library Services

Digital Collections and Content

Working toward interoperable digital content.

About the Project How to Participate

Final Report
IMLS Digital Collections
IMLS National Leadership
"Proposal for IMLS Collection Regis

Summary of F
October 2002 — Se

TABLE of CO

I. Title Page & Credits [.pdf/16KB]

II. Executive Summary [.pdf/42KB]

III. Collection Research [.pdf/144KB]

A. Evolution of Collection-Level Description
B. Collection Identity in IMLS Community
C. Trends in Metadata Practices among IMLS DCC Registry Participants
D. Diffusion of Metadata Practices in the IMLS Community

V. Content and Use [.pdf/906KB]

Fig.1. User’s interface of Pliny on note-taking for a web page

i) Pliny Supports existing scholarly research practice (ESP. Humanities)

Conventionally, we can think humanities scholarly workflow as three phases: First as Public -- discovering & reading resources, Second as Private -- taking & organizing notes, finding relationships, writing up ideas and third as Publishing -- new articles, books, etc. See as below,

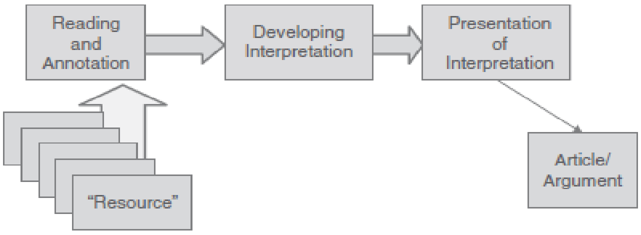


Fig.2. Humanities scholarship modle in Bradley's article

Pliny is designed to support especially the second phase, personal note-taking and management. (Bradley, 2008) However, Pliny's definition of the second phase begs the question how private should the private phase be? Increasingly in Digital Humanities there are many practical opportunities and reasons for collaboration which suggests that the 2nd phase could need a more interoperable approach. Though not formally published, notes, preliminary ideas & sketches, and the like can be useful in their own right; scholars should be able to preserve and in some cases share these notes. (The Open Annotation Collaboration Phase I , 2009)

ii) How & Where Pliny store the annotation --- also what Pliny thinks an annotation is (From Larry and Yan's email)

- Pliny used Derby version of the apache software foundation - Apache Derby to store the data in the database directory as in the example C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny\pliny which is the root directory for most of the user-generated information stored by Pliny.

- Under the root directory, C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny\pliny\log\log1.dat contains multiple versions of the text of notes, suggesting it may provide a roll-back capability. The text portion is in ASCII, but the file uses primarily binary.

- Under the root directory, C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny\pliny\seq0 contains many binary files which contain a mix of java code and SQL commands.

- Parallel to the root directory, the directory C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny\pliny\webIconCache contains a thumbnail JPEG of the target webpage.

- Outside of the root directory, C:\Documents and Settings\Sony\pliny\.metadata\.plugins\org.eclipse.ui.workbench\workbench.xml contain the designation of the URL targets of each annotation, separately. That is, for multiple annotations targeting one webpage, there will be here multiple annotation descriptors, each targeting the same page. There's some form of serial number differentiating the annotations.

- The directories which are also outside of the root directory are where Pliny keeps cached copies of PDF and image files when image and PDF are annotated as local resources. Images and PDFs are saved respectively under the directory C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny.imageRes \imageCache and C:\Documents and Settings\Sony\pliny\.metadata\.plugins\uk.ac.kcl.cch.jb.pliny.pdfAnnot\pdfCache

iii) **Extract the annotations not via the Pliny's graphical user interface.**

In this section Pliny database table schema will be described and what data each table contains will be explained. (This summary draws on work done collaboratively with Larry Jackson and Tim Cole)

The below captured from the commend line are major Pliny tables in Derby database:

```
TABLE_SCHEM | TABLE_NAME     |REMARKS
------------------------------------
1)  PLINY      |LINK          |
2)  PLINY      |LINKABLEOBJECT |
3)  PLINY      |LOTYPE        |
4)  PLINY      |NOTE          |
5)  PLINY      |OBJECTTYPE    |
6)  PLINY      |RESOURCE      |
```

The LINK (1) table stores relations between pairs of annotations which indicate one annotation can point to the other to show the relation. The LINKABLEOBJECT (2) table implements the adjuncts to Resource's which are shared by both annotations and links. Table (3) states the type, and defines the associated display color, for annotations and/or links. The NOTE table (4) stores the content of annotation bodies. The OBJECTTYPE table (5) lists the viewer to be used for objects, and thereby the type classification of the object. The RESOURCE table (6) is Pliny's principal table, and the starting point for the definition of all objects.

The graphy synthesized by Larry as below shows the complex relationship between six major tables in Pliny:

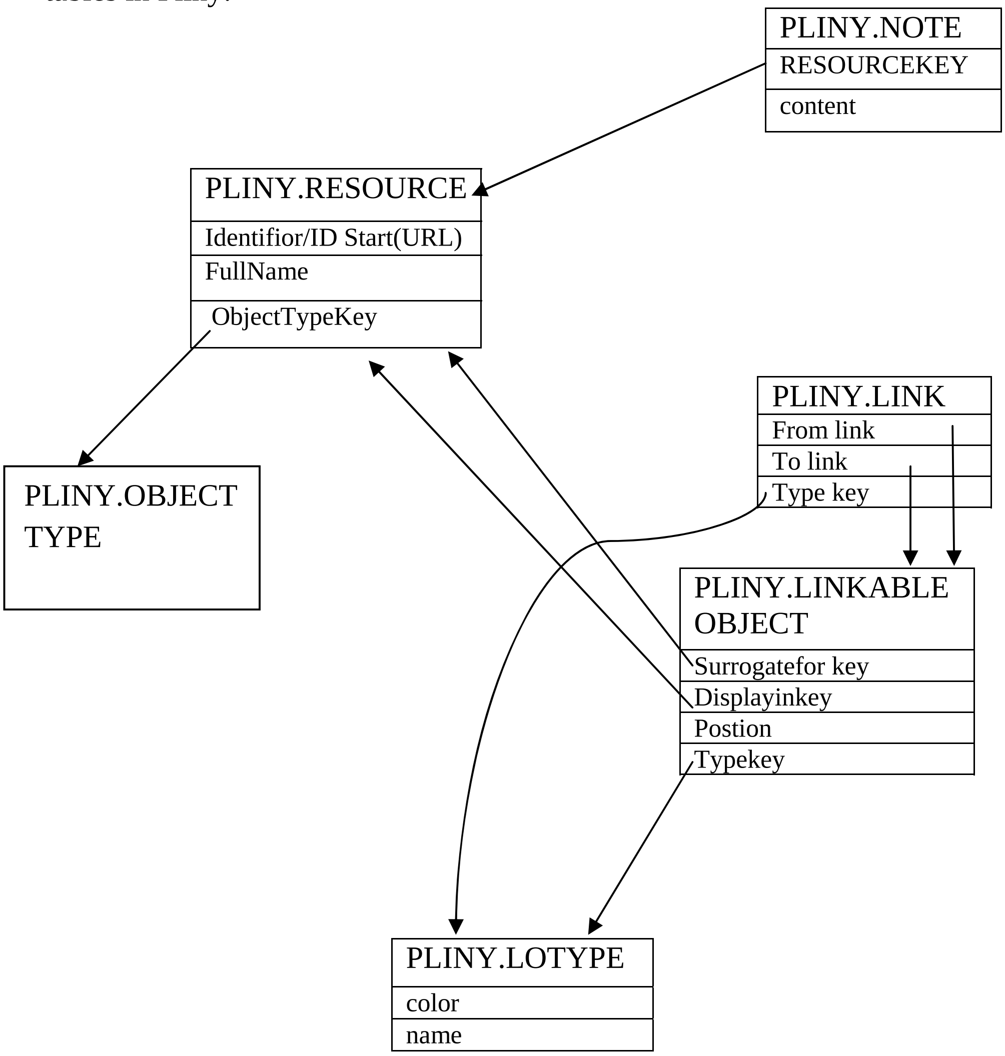


Fig.3. Relationships of Pliny major tables

Fig.3. illustrates how Pliny stores notes, and how it keeps track of references to Web resources. We can also see how Pliny links to external objects referenced by the notes -- e.g., Web pages, PDFs, and Images. The above example shows that Pliny associates the web page and the note a resource key which is foreign key for the note table and also linkable object table. However, the keys assigned and used in the Pliny data store are database generated identity columns and so non-persistent and not useful outside local user context. Fig.3. also suggests what kind of metadata Pliny keeps about each note or container entity as each object is created and where this information is maintained, for example from PLINY.NOTE table, contains the timestamp information for each note indicating when it was last modified or created. (see details in the following sample)

- An sample example to see how an annotation to a web page store in PLINY.RESOURCE table

RESOUR-CE KEY	FULL NAME	INIT CHAR	OBJECT TYPE KEY	IDENTIFIER	ID START	ATTRIBUTES	CREATION DATE	CREATION TIME
1	IMLSDCC Project – Final Report – Table of Contents – IMLS Digital Collections and Content	I	2	url:http://imlsdcc.grainger.uiuc.edu/finalreport2007	url:http://imlsdcc.grainger.uiuc.edu /finalreport2007	sash=50	2009-08-30	15:09:20
2	WebNote1_for_DCC	W	1				2009-08-30	15:10:08

NB:The resourcekey is incremented value assign by derby database. Fullname attribute is named by the annotator given to resource or note at the front-end Pliny inferface as shown in Fig.1.

- The WebNote1_for _DCC Note in PLINY.NOTE table

RESOURCEKEY	CONTENT	TSTAMP
2	To READ I, II and III part	2009-08-30 15:10:08.667

NB:The resourcekey value in Note table here is corresponding to the reourcekey value in Resource table.

- PLINY.LINK

LINKKEY	FROMLINK	TOLINK	TYPEKEY	ATTRIBUTES
generated	Index into LINKABLEOBJECT.	Index into LINKABLEOBJECT	Index into LOTYPE, to obtain the text label and colors for the type of this link.	N/A

NB: Link table is useful which shows the relation between two related annotations.

- PLINY.LINKABLEOBJECT

LINKABLE OBJECTKEY	TYPE KEY	POSITION	DISPL PAGENO	SURR PAGENO	DISPLAYED INKEY	SURROGATE FORKEY	IS OPEN	SHOWING MAP
gernerated	Index into LOT YPE, providing the color and type of this annotation.	Display coordinates , in an encoded string.			DisplayedInKey is the index into RESOURCE for the superordinate to this annotation. Typically that's the webpage with which the annotation is associated.	Index into RESOURCE of this annotation.	Whether diplay the content of note or not	Whether having contained annotation or not

NB: Linkableobject table joined the Resource table and Note table together as shown in attribute displyinkey and surogateforkey .

- PLINY.LOTYPE

LOTYPE KEY	NAME	TITLEFORECOLOUR	TITLEBACKCOLOUR	BODYFORECOLOUR	BODYBACKCOLOUR	SOURCEROLE KEY	TARGETROLE KEY
gernerated							

NB: the major function of LinkableObjectType table is used for changing background color of the note or note title, because Pliny provided a self-defined color code to category notes. (Fig. 8)

- PLINY.OBJECTTYPE

OBJECTTYPEKEY	NAME	PLUG INKEY	IDSTRING	EDITORID	ICONID
1	Note	1		uk.ac.kcl.cch.jb.pliny.noteEditor	NULL
2	Web Browser	1		uk.ac.kcl.cch.jb.pliny.browserEditor	
3	Image	2		uk.ac.kcl.cch.jb.pliny.imageRes.editor	editor:uk.ac.kcl.cch.jb.pliny.imageRes.editor
4	PDF/ Acrobat	3		uk.ac.kcl.cch.jb.pliny.pdfAnnot.PDFEditor	editor:uk.ac.kcl.cch.jb.pliny.pdfAnnot.PDFEditor

Our intent is to understand Pliny well enough that we can define a Web service with which Pliny could interact. There are also implicit metadata such as authorship about each note in Pliny we will need to make explicit when we share a note created in Pliny with others. Longer term the objective is identify how Pliny would need to be modified and extended so that it can expose notes and references a user creates or captures using Pliny to such a external Web service.

IV. Analysis Part 2: Mapping Pliny into Annotea RDF

i). what Annotea is and model of annotation

The Semantic Web initiative of the W3C is an effort to make the Web more amenable to automated analysis. The potential to apply Semantic Web approaches to annotation was recognized early on. Recent papers, such as OAC research proposal and Haslhofer’s paper, show that annotations play increaslying important role in contributing metadata for search and retrieve digital resources. Applying Semantic Web approaches to make annotation sharable in digtial context is very necessary to benefit scholarship. In 2002, the Annotea protocal was established to support the creation and publication of shareable annotations. Kahan et al. (2002) defined Annotea as “a Web-based shared annotation system based on a general-purpose open resource description framework (RDF) infrastructure, where annotations are modeled as a class of metadata.” The Annotea Metadata schema defined seven annotation properties a set of annotation type classes.These annotation metadata values accomodated a context attribute based XPointer for locating the annotation targets within the annotated document. In Annotea a web annoation can be viewed as metadata and a remark or note about a document identified by a URI and stored in specialized servers. Figure 2 illustrates the commucation betwwen annotation servers and clients and how annotations can be shared through Annotea Protocol. (Kahan, 2002)

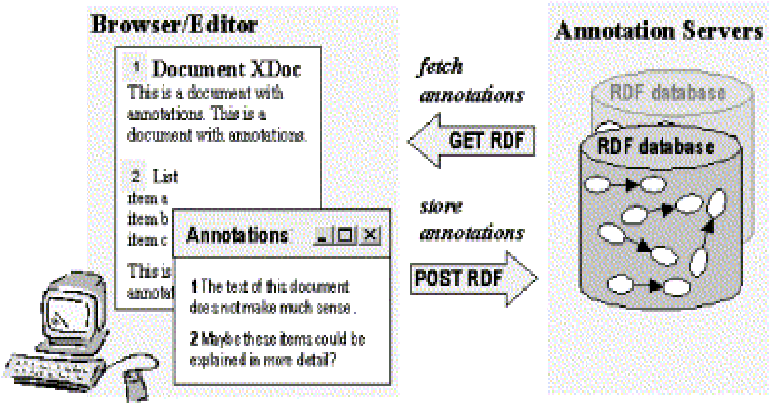

Browser/Editor
1 Document XDoc
This is a document with annotations. This is a document with annotations.
2 List
item a
item b
item c
This is
annotat
Annotations
1 The text of this document does not make much sense.
2 Maybe these items could be explained in more detail?
fetch annotations
GET RDF
store annotations
POST RDF
Annotation Servers
RDF database
RDF database

Fig. 4. General architecture of Annotea

As mentioned, the Annotea model allows for six attributes to be attached to an annotation as depicted in figure 5: (Kahan, 2002)

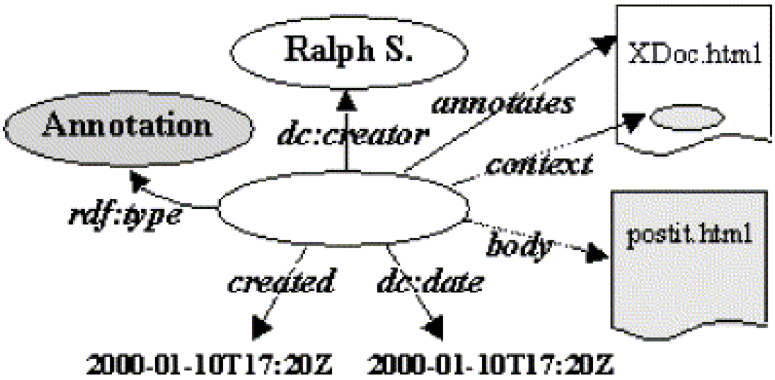
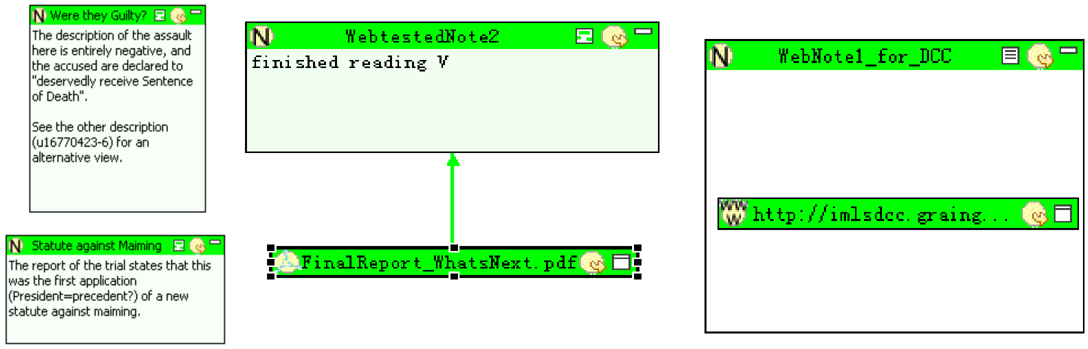


Fig. 5 The RDF model of an annotation.

From the above RDF model, we can see the major attributes of an annotation. The RDF type is the top-level class "annotation." Subclasses of annotation defined by Annotea, defined by an RDF schema, are: SeeAlso, Question, Explanation, Example, Comment, Change, and Advice. Use of subclassing is optional. Annotea also defines seven properties of Annotation Schema which are: Annotates, Author, Context, Created, Body, Modified, and Related. Creator is the author who made the annotation. Body is the 'content' of an annotation which could contain textual description and also resources. Created is the date and time when annotation is created. XPointer provides standard way to locate the annotation within the web content in the context property of Annotea schema. Extensions are possible for the model through adding attributes from other namespaces such as language, format, and access policy which restrict the annotation available to specific users group. (Hunter, 2009)

ii). Implementation & Intellectual

In this section, the initial intellectual mapping decisions will be given from PLINY to Annotea RDF. Though the annotation itself can be in any language and any formats, such as textual document, audio, video, image, in Pliny annotation can only be represented as a note box with its content inside as shown below, therefore in our Annotea RDF mapping process, under the top-level class annotation, we think using comments as subclass is appropriate.



The description of the assault here is entirely negative, and the accused are declared to "deservedly receive Sentence of Death".

See the other description (u16770423-6) for an alternative view.

finished reading V

The report of the trial states that this was the first application (President=precedent?) of a new statute against maiming.

Fig. 6 Annotation examples in PLINY.

iii). Mapping PLINY notes to Annotea Annotations:

In this section, Table 1 shows the initial decision for mapping the information from Pliny tables into Annotea. In other words, which attribute in which table from Pliny can be mapped into annotates or author for example. There is some granularity between Pliny data model and Annotea, since Pliny is more personal standalone annotation application while Annotea is towards to support publishing sharable annotations. Therefore, not all information from Pliny table can be mapped into Annotea.

Properties in Annotea Defined by [http://www.w3.org/2000/10/annotation-ns#]	**Corresponding data in Pliny**

Annotates Relates an Annotation to the resource to which the Annotation applies. The inverse relation is 'hasAnnotation' Should be a URL.	For Notes annotating Web resources, URL is in PLINY.RESOURCE.INDENTIFIFER column; these notes can only refer to the entire Web resource. For Notes annotating specific parts of PDF or image resources, the original URL for the resource being annotated is contained within the PLINY.RESOURCE.ATTRIBUTES column, but the column value must parsed to retrieve this information. NB: • PDFs and images are always cached • If URL is not type file, there will always be another resource which is the PDF as a Web resource. • PDFs and images may also be annotated as Web resources.
Author The name of the person or organization most responsible for creating the Annotation.	Pliny does not have analog for this annotea property. For each note, since in a stand-alone Pliny instance, one can assume that all notes were created by the same individual.
Body Relates the resource representing the 'content' of an Annotation to the Annotation resource.	Include values from both PLINY.RESOURCE.FULLNAME (title of annotation) and PLINY.NOTES.CONTENT. NB: Notes may have embedded resources – i.e., contained other notes, or contained images, PDFs, or other Web resources. These are linked through the PLINY.LINKABLEOBECT.SURROGATEFORKEY
Context The context within the resource named in 'annotates' to which the Annotation most directly applies.	We assume closet analog in Pliny is value of PLINY.LINKABLEOBJECT.POSITION. This is, for example, the pixel rectangle within the annotated object (PDF or Image) which is given by the DISPLAYINKEY column of the same row. (Key for Note annotating that rectangle appears as SURROGATEFORKEY in same row). Notes on Web resources are assumed to have context of the whole.
Created The date and time on which the Annotation was created. yyyy-mm-ddThh:mm:ssZ format recommended.	Analog in Pliny is PLINY.NOTE.TSTAMP NB: Time Format should be converted to W3CDTF

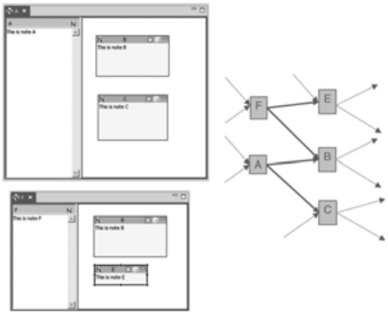
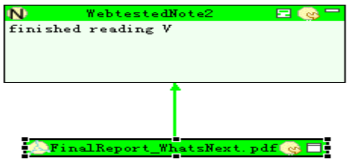
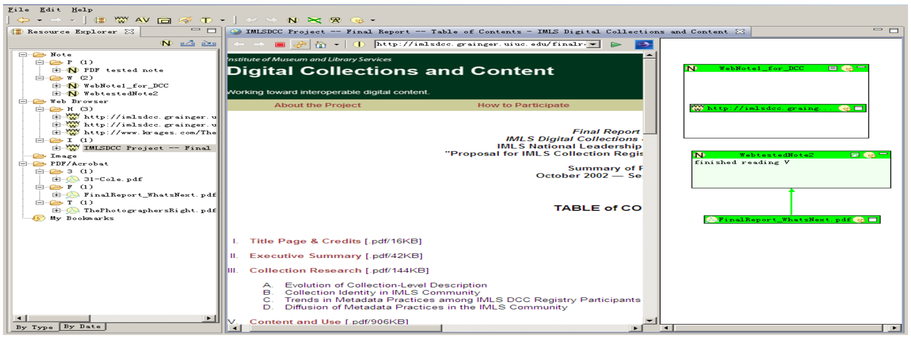
Modified The date and time on which the Annotation was modified. yyyy-mm-ddThh:mm:ssZ format recommended.	Not available in Pliny
Related A relationship between an annotation and additional resources that is less specific than 'body'. The 'related' property is expected to be subclassed by more specific relationships.	Pliny offers various types of relationships between annotations, but most must be inferred. For example:1) Contained annotations  This is note A This is note B This is note C This is note F This is note B This is note E F E A B C (taken from John's article) 2) refer to annotations:  WebtestedNote2 finished reading V FinalReport_WhatsNext.pdf

Table 1. Mapping Pliny into Annotea

Two Examples of our Mapping from PLINY to Annotea for different relations of Notes:


File Edit Help
Resource Explorer
Note
P (1)
PDF tested note
W (2)
WebNote1_for_DCC
WebtestedNote2
Web Browser
H (3)
http://imlsdcc.grainger.u
http://imlsdcc.grainger.u
http://www.krages.com/The
I (1)
IMLSDCC Project -- Final
Image
PDF/Acrobat
3 (1)
31-Cole.pdf
F (1)
FinalReport_WhatsNext.pdf
T (1)
ThePhotographersRight.pdf
My Bookmarks
By Type
By Date
IMLSDCC Project -- Final Report -- Table of Contents - IMLS Digital Collections and Content
http://imlsdcc.grainger.uiuc.edu/finalr
WebNote1_for_DCC
http://imlsdcc.graing
WebtestedNote2
finished reading V
FinalReport_WhatsNext.pdf

nstitute of Museum and Library Services

Digital Collections and Content

Working toward interoperable digital content.

About the Project How to Participate

Final Report
IMLS Digital Collections
IMLS National Leadership
"Proposal for IMLS Collection Regis

Summary of F
October 2002 — Se

TABLE of CO

I. Title Page & Credits [.pdf/16KB]
II. Executive Summary [.pdf/42KB]
III. Collection Research [.pdf/144KB]
A. Evolution of Collection-Level Description
B. Collection Identity in IMLS Community
C. Trends in Metadata Practices among IMLS DCC Registry Participants
D. Diffusion of Metadata Practices in the IMLS Community
V. Content and Use [.pdf/906KB]

Fig. 7 Different relations of Notes in Pliny

Example a) below shows how to mapping contained view of notes in Pliny to Annotea,

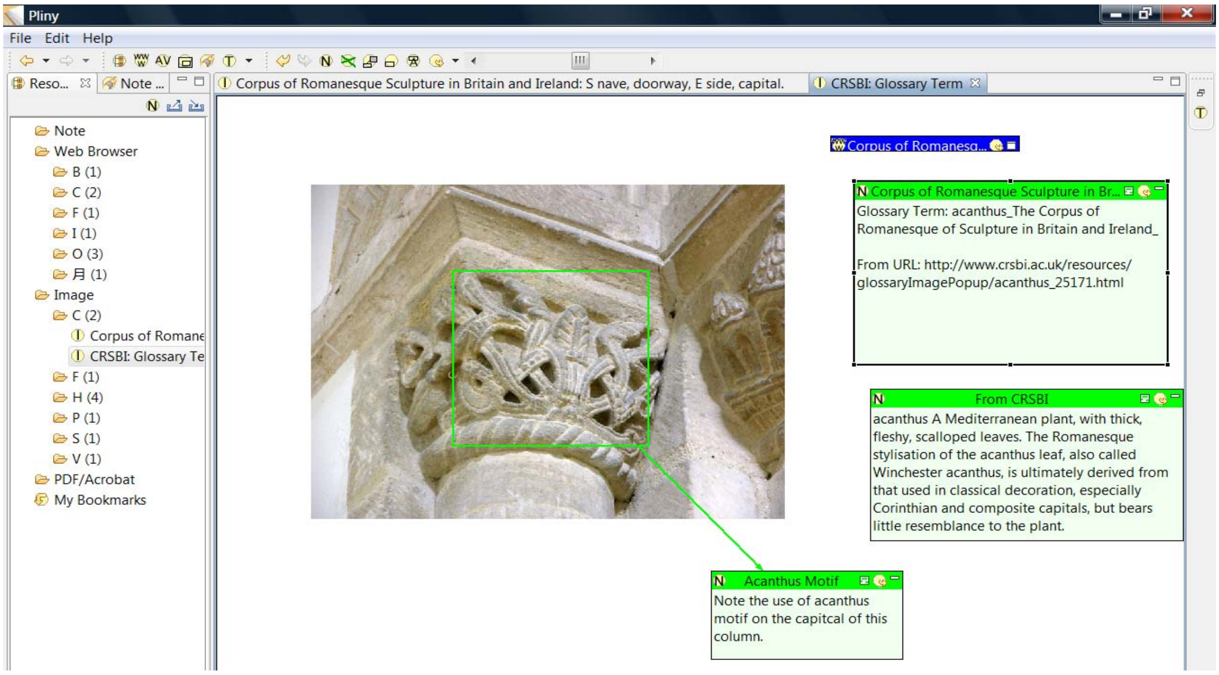
```
<?xml version="1.0" encoding="UTF-8"?>
<r:RDF xmlns:r="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
          xmlns:h="http://www.w3.org/1999/xx/http#"
          xmlns:a="http://www.w3.org/2000/10/annotation-ns#"
          xmlns:d="http://purl.org/dc/elements/1.1/">
  <r:Description r:nodeID='containerTop'>
    <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Annotation"/>
    <r:type r:resource="http://www.w3.org/2000/10/annotationType#Comment"/>
    <a:annotates r:resource="http://imlsdcc.grainger.uiuc.edu/finalreport2007/"/>
    <a:related r:nodeID='contained1'/>
    <d:title>WebNote1_for_DCC</d:title>
    <a:author>Yan Wang</a:author>
    <a:created>2009-08-30T21:10:08.667Z</a:created>
    <a:body>
      <r:Description>
        <h:ContentType>text/html</h:ContentType>
        <h:Body r:parseType="Literal">
          <html xmlns="http://www.w3.org/1999/xhtml">
            <head>
              <title>WebNote1_for_DCC</title>
            </head>
            <body>
              <h2>WebNote1_for_DCC</h2>
              <p>TO READ
                <br/>
                I, II and III PartTO READ
                <br/>
                I, II and III Part
              </p>
            </body>
          </html>
        </h:Body>
      </r:Description>
    </a:body>
  </r:Description>
  <r:Description r:nodeID='contained1'>
    <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Annotation"/>
    <r:type r:resource="http://www.w3.org/2000/10/annotationType#Comment"/>
    <a:annotates r:resource="http://imlsdcc.grainger.uiuc.edu/finalreport2007/"/>
    <a:related r:nodeID='containerTop'/>
    <d:title>http://imlsdcc.grainger.uiuc.edu/docs/31-Cole.pdf</d:title>
    <a:author>Yan Wang</a:author>
    <a:created>2009-08-30T21:10:08.667Z</a:created>
    <a:body r:resource='http://imlsdcc.grainger.uiuc.edu/docs/31-Cole.pdf'/>
  </r:Description>
</r:RDF>
```

Example b) below shows how to mapping refering view of notes in PLINY to Annotea,

```
<?xml version="1.0" encoding="UTF-8"?>
<r:RDF xmlns:r="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
         xmlns:h="http://www.w3.org/1999/xx/http#"
         xmlns:a="http://www.w3.org/2000/10/annotation-ns#"
         xmlns:d="http://purl.org/dc/elements/1.1/">
   <r:Description r:nodeID='Target1'>
      <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Annotation"/>
      <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Comment"/>
      <a:annotates r:resource="http://imlsdcc.grainger.uiuc.edu/finalreport2007/"/>
      <d:title>WebNote1_for_DCC</d:title>
      <a:author>Yan Wang</a:author>
      <a:created>2009-08-30T21:10:08.667Z</a:created>
      <a:body r:resource='http://sample.org/sample.html'/>
   </r:Description>
   <r:Description r:nodeID='Source1'>
      <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Annotation"/>
      <r:type r:resource="http://www.w3.org/2000/10/annotation-ns#Comment"/>
      <a:annotates r:resource="http://imlsdcc.grainger.uiuc.edu/finalreport2007/"/>
      <a:related r:nodeID='Target1'/>
      <d:title>http://imlsdcc.grainger.uiuc.edu/docs/31-Cole.pdf</d:title>
      <a:author>Yan Wang</a:author>
      <a:created>2009-08-30T21:10:08.667Z</a:created>
      <a:body r:resource='http://imlsdcc.grainger.uiuc.edu/docs/31-Cole.pdf'/>
   </r:Description>
 </r:RDF>
```

The above two examples successfully validate through RDF validation service[3]. However, the a:context for the mapping result of web pages in Pliny is the whole webpage which doesn’t make any sense to XPointer whose mechanism is figuring out which part of the recourses that the Annotation most directly applies to. It causes problem when we saved the mapping results into Annotea server to queried from the results, as Annotea server require the information of a:context.

[3]



N Corpus of Romanesque Sculpture in Br...

Glossary Term: acanthus_The Corpus of Romanesque of Sculpture in Britain and Ireland_

From URL: http://www.crsbi.ac.uk/resources/glossaryImagePopup/acanthus_25171.html

N From CRSBI

acanthus A Mediterranean plant, with thick, fleshy, scalloped leaves. The Romanesque stylisation of the acanthus leaf, also called Winchester acanthus, is ultimately derived from that used in classical decoration, especially Corinthian and composite capitals, but bears little resemblance to the plant.

N Acanthus Motif

Note the use of acanthus motif on the capitcal of this column.

Fig. 8 An example of image annotation in Pliny

For Images and PDFs, the a:context is applicable in Pliny which can be viewed from the above picture that the rectangle on the images which was stored in PLINY.LINKABLEOBJECT.POSITION gives the pixel information of the area which has been annotated. Base on Haslhofer's (2009) article, we are thinking of a possible approach for how to manipulate exporting the pixel information from Pliny's back-end derby database for images or PDFs into Annotea. We may be able to create a transparent image (and PDF) with same size of the original one in SVG format with the rectangle and make sure the original image (and PDF) have superposition of the new SVG file. This possible approach may help us access the fragment of image or PDF through Annotea server.

V. Discussion

There are several issues that might need our attention when reconciling PLINY data model to generalized, interoperable models of annotation discussed in the literature. From the above examination and analysis, we know Pliny offers various types of relationships between annotations, though these are arbitrary and sometimes ambiguous. But is it enough to simply assert that some untyped relationship exists between pairs of annotations? Possibly we could infer some annotation relationships exposed through Pliny and extend the Annotea related property with Qualified DC relationships. For example: We could assume that Pliny note X contained within note Y has a qdc:isPartOf relationship to note Y, or Note A that points to Note B has a qdc:references relationship to Note B. However, because it is not really clear in Pliny exactly what these relationships are meant to convey, this may not be wise. Further work with Pliny users would be needed to reach a conclusion on this matter. Refinements beyond general assertion of relationship between Pliny notes may have to be left to communities of users. Other issues that arose in the process of mapping Pliny "notes" to Annotea RDF annotations:

1) Indefinite "targets" for PLINY notes

    a. Notes can "target" i.e., annotate a complete web page, PDF resource, or image resource. Ambiguity is whether target is entire resource or some sub-part. See also c, e, & f – can't always distinguish these from other cases. a: annotates is suspect since context is unknown which leaves a:annotates imprecise.

    b. Notes can also "target" sub-parts of PDF and images. Interpretation here is more straightforward – such notes are clearly annotations of one kind or another of a specific sub-part of a resource, a:annotates and a:context are more easily derived.

    c. Arrows can be drawn from one note to another, but not clear what is meant by arrows (note they can be 'typed' using color codes); possible meanings: note2 annotates note1; note2 provides context for note 1 (see e below); note2 is a sibling of note1 – i.e., both note1 and note2 annotate same resource, same context and are talking about similar facet of the target (same FRBR level). a:annotates, a:related, a:context may all be in doubt.

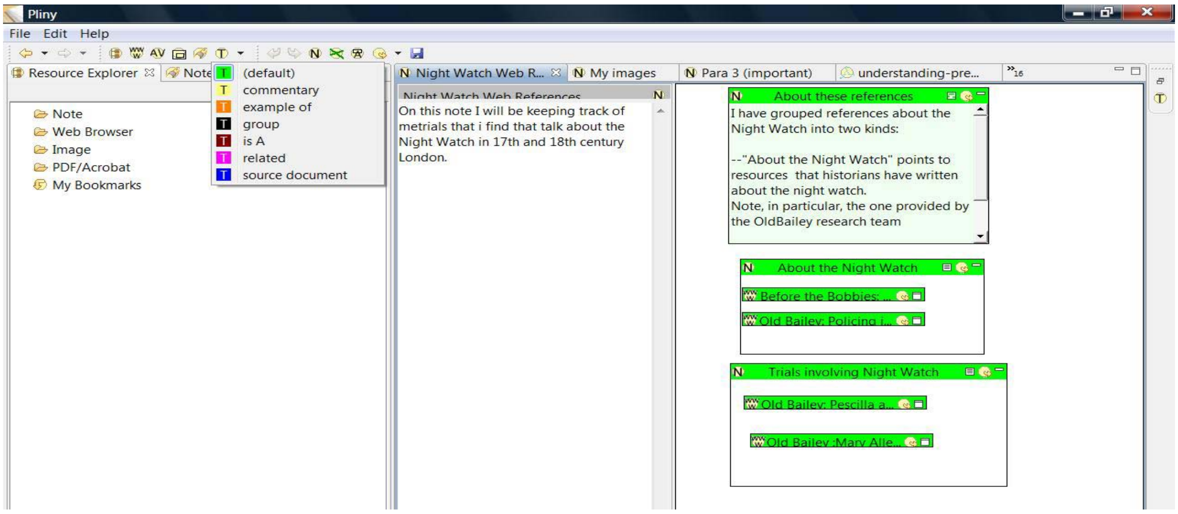
    d. Notes can be contained one in the other. Pliny will display as if contained note annotates container note; however, container as mechanism to group notes that all share target or other facet in common is not accounted by this approach. a:annotates, a:related may be in doubt.

    e. Notes can be created by cut and paste from resource. This creates a note that may serve as a:context for another note or may be self-referencing as context for a note. a:annotates and a:context in doubt.

    f. Contents of a note can be another resource. Is this creating a resource 1 annotates resource2 annotation or is it creating a multi-target annotation or a multi-sourced annotation?

2) Locally defined classes of annotation and classes of relationships between annotations.

    a. Pliny allows typing of both notes and relationships between notes, but classing of types is left entirely to the user. Pliny allows users to color code their annotations according to their types or relation. (see Fig.9 as below) We could perhaps seed Pliny implementations with Annotea annotation types, but not clear they would be used consistent with intent. This may have to be left to communities of users. Pliny offers self –defined notes type mechanism in its type manger which color coded different types of notes, therefore , we may be able to map PLINY self –defined types into Annotea subclasses of types.



On this note I will be keeping track of metrials that i find that talk about the Night Watch in 17th and 18th century London.

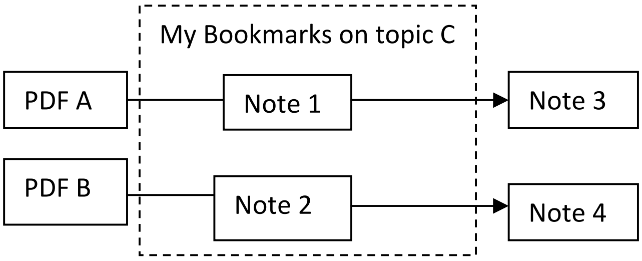
I have grouped references about the Night Watch into two kinds:

--"About the Night Watch" points to resources that historians have written about the night watch.
Note, in particular, the one provided by the OldBailey research team

Fig. 9. The example of user self-defined note type in Pliny

b. In PLINY one note can "annotate" 2 resources easily just dragging the note with its entire attribute including content, title, and type to beside the resource. However, a note annotating different resources may have different types when considered as an annotation of each resource singly. For example, note A in resource 1 may be color coded as red which means important, but note A for resource 2 may be color coded differently, i.e, may not be that important as in resource 1.

c. In PLINY, note 1 annotating PDF A and note 2 annotating PDF B can be brought together in the same container through creating a new note and organizing the new note into "My Bookmarks" in PLINY resource explore pane. This then brings note1 and note2 together. However, it's hard to decide how many relationships should be preserved when exporting such joined annotations. For example as shown below, note 1 annotates PDF A, but related to note3, note2 annotates PDF B and related to note4, how can make decision to bring note1 and note2 together and exclude note3 and note4?



VI. Conclusion

This report provides an initial approach to examining an existing annotation application tool-Pliny. I was able to explore in a cursory way the why and how Pliny data might be mapped into Annotea RDF annotation. Hopefully these provide some initial insights into how to begin building extended services for humanities scholarship that can support the sharing and interoperability of annotations. Clearly the issues of what annotation properties are most important and being able to express context for an annotation target that makes clear exactly what intellectually is being annotated for Web resources are crucial considerations when trying to construct a shared, interoperable annotation data model.

We are still working on how to solve the uncertainties of annotation types which in Pliny is defined by community or user self. More often the relationships between annotation is more than one to one relationship and showing contrast, comparison and similarity of common relationships of annotation in humanities scholarships are also challenging for make annotation interoperable. In addition, Pliny papers are based on traditional scholarship which was viewed to be built upon personal musing of individual scholars, however, traditional scholarship is gradually changing in the digital era and great potential of sharing annotation in digital context will benefit the new merging scholarship in which collaborative work is an important part. The informal bloggish annotations or messages are so popular for which we can see the potential of a web and resource-centric interoperable annotation environment for future scholarship. Therefore, we have to pay more attention to scholar needs in ways that prior work in this domain sometimes has not done and to get larger range of common scholarly practices involving annotation to develop the model of collaborative annotation system.

Acknowledgments The work reported in this report was under the supervision of Prof. Tim Cole and Prof. Allen Renear. I also would like to thank Larry Jackson (GSLIS senior scientist) in offering help to examine Pliny and build annotea server in UIUC to test mapping results.

Bibliography

The Open Annotation Collaboration. (2009). *Phase I Proposal Narrative.* Retrieved June 15, 2009, from OAC Open Annotation Collaboration: http://www.openannotation.org/phaseI.html

Bradley, J. (2008).Thinking about Interpretation: Pliny and Scholarship in the Humanities. *Literary and Linguistic Computing*, 23(3), 263-279.

Renear, Allen H.; DeRose, Steve J.; Mylonas, Elli; van Dam, Andries. (1999). *An Outline for a Functional Taxonomy of Annotation*. White Paper and part of a presentation given to Microsoft Research, April 1999. Available IDEALS: http://hdl.handle.net/2142/9098

Hunter, Jane. (2009). Collaborative, Semantic Tagging and Annotation Systems. Chapter in *Annual Review of Information Science and Technology* (43). Medford, N.J.: Learned Information, Inc.

Agosti M. & Ferro N. (2007). A formal model of annotations of digital content. *ACM Transactions on Information Systems,* 26(1), Article 3.

Kahan, J.; Koivunen, M.; Prud'Hommeaux, E.; Swick, R. (2002). Annotea: an open RDF infrastructure for shared Web annotations. *Computer Networks,* 39(5), 589-608.

Haslhofer, B.; Jochum, W.; King, R.; Sadilek, C.; Schellner, K. (2009). The LEMO annotation framework: weaving multimedia. *International Journal of Digital Library,* 10, 15-32.

References Consulted

Marshall, C.C. (1997). Annotation: from Paper Books to the Digital Library. *Proceedings of Digital Libraries '97. NewYork: ACM Press 1997.*

Initial "First Draft" OAC Use Cases. (2009). Retrieved June 15, 2009, from The Open Annotation Collaboration Wiki: https://apps.lis.illinois.edu/wiki/pages/viewpage.action?pageId=11568056

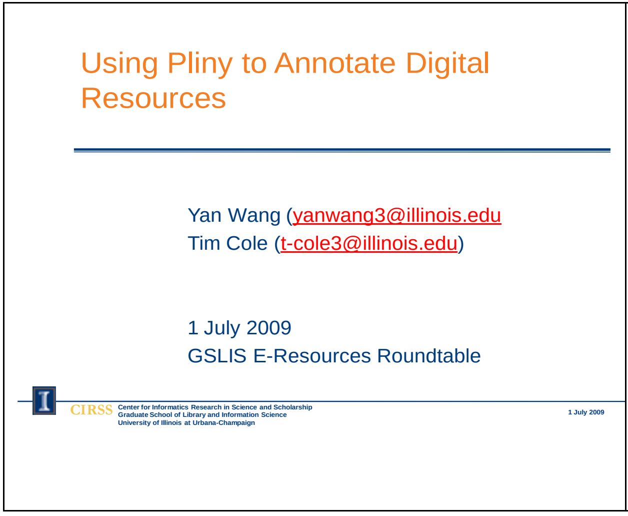
Agosti, Maristella;Bonfiglio-Dosio, Giorgetta; Ferro, Nicola. (2007). A historical and contemporary study on annotations to derive key features for systems design. *International Journal on Digital Libraries*, 8(1), 1-19.

Annotea Protocols. (2002, December 20). Retrieved August 25, 2009, from Annotea Protocols: http://www.w3.org/2002/12/AnnoteaProtocol-20021219

Annotea Annotation Schema. (2004, March 23). Retrieved August 25, 2009, from Annotea Annotation Schema: http://www.w3.org/2000/10/annotation-ns

Appendix

1) PPT for Metadata round table on July 1st ,2009 http://cirss.lis.illinois.edu/Rtable/errt.html



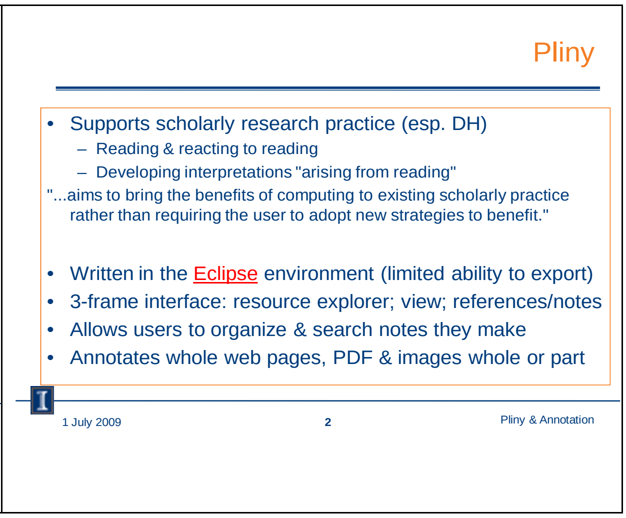
Using Pliny to Annotate Digital Resources

Yan Wang (yanwang3@illinois.edu
Tim Cole (t-cole3@illinois.edu)

1 July 2009
GSLIS E-Resources Roundtable

CIRSS Center for Informatics Research in Science and Scholarship
Graduate School of Library and Information Science
University of Illinois at Urbana-Champaign

1 July 2009



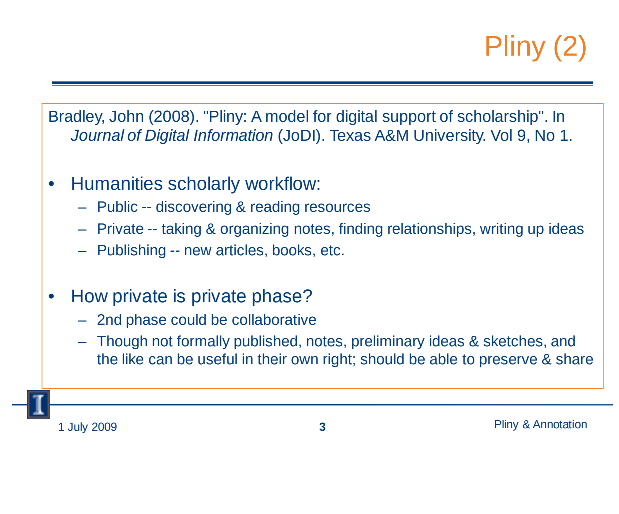
Pliny

- Supports scholarly research practice (esp. DH)
  - Reading & reacting to reading
  - Developing interpretations "arising from reading"

"...aims to bring the benefits of computing to existing scholarly practice rather than requiring the user to adopt new strategies to benefit."

- Written in the Eclipse environment (limited ability to export)
- 3-frame interface: resource explorer; view; references/notes
- Allows users to organize & search notes they make
- Annotates whole web pages, PDF & images whole or part

1 July 2009 2 Pliny & Annotation

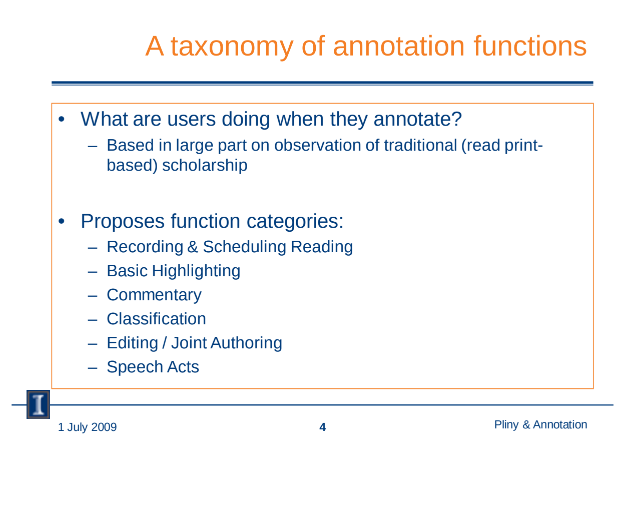


Pliny (2)

Bradley, John (2008). "Pliny: A model for digital support of scholarship". In *Journal of Digital Information* (JoDI). Texas A&M University. Vol 9, No 1.

- Humanities scholarly workflow:
  - Public -- discovering & reading resources
  - Private -- taking & organizing notes, finding relationships, writing up ideas
  - Publishing -- new articles, books, etc.
- How private is private phase?
  - 2nd phase could be collaborative
  - Though not formally published, notes, preliminary ideas & sketches, and the like can be useful in their own right; should be able to preserve & share

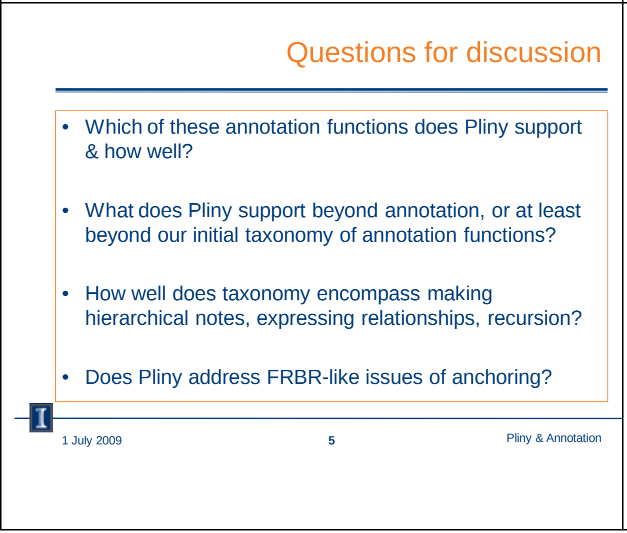
1 July 2009 3 Pliny & Annotation



A taxonomy of annotation functions

- What are users doing when they annotate?
  - Based in large part on observation of traditional (read print-based) scholarship
- Proposes function categories:
  - Recording & Scheduling Reading
  - Basic Highlighting
  - Commentary
  - Classification
  - Editing / Joint Authoring
  - Speech Acts

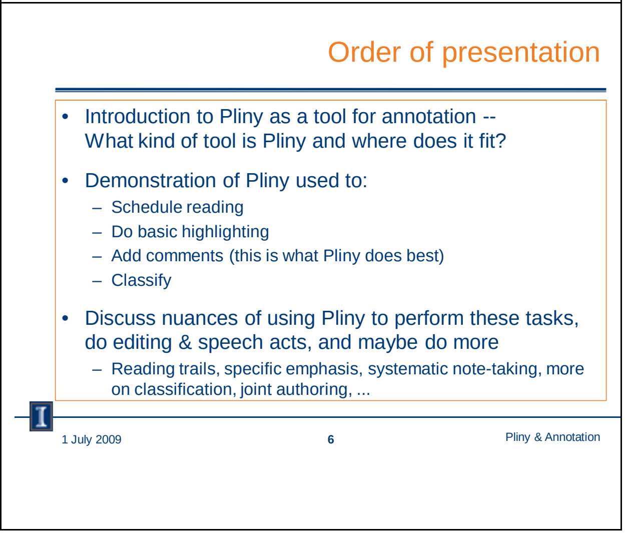
1 July 2009 4 Pliny & Annotation



Questions for discussion

- Which of these annotation functions does Pliny support & how well?
- What does Pliny support beyond annotation, or at least beyond our initial taxonomy of annotation functions?
- How well does taxonomy encompass making hierarchical notes, expressing relationships, recursion?
- Does Pliny address FRBR-like issues of anchoring?

1 July 2009 5 Pliny & Annotation



Order of presentation

- Introduction to Pliny as a tool for annotation -- What kind of tool is Pliny and where does it fit?
- Demonstration of Pliny used to:
  - Schedule reading
  - Do basic highlighting
  - Add comments (this is what Pliny does best)
  - Classify
- Discuss nuances of using Pliny to perform these tasks, do editing & speech acts, and maybe do more
  - Reading trails, specific emphasis, systematic note-taking, more on classification, joint authoring, ...

1 July 2009 6 Pliny & Annotation

2) One Example of Java programs to insert annotation to PDF file in PLINY not via UGI

```
import java.sql.Connection;
import java.sql.DriverManager;
import java.sql.SQLException;
import java.sql.SQLWarning;
import java.sql.*;
import java.sql.DatabaseMetaData;
import java.net.*;
import java.io.*;

public class testPlinyPDF2{
private static final String driver = "org.apache.derby.jdbc.ClientDriver";
private static final String url = "jdbc:derby://localhost:1527/C:\\Documents and
Settings\\sony\\pliny\\.metadata\\.plugins\\uk.ac.kcl.cch.jb.pliny\\pliny;create=false";
```

```
final static int size=1024;
public static void main(String[] args) {
  Connection con = null;
  DatabaseMetaData dbmd = null;
  Statement stmt = null;
  ResultSet rs = null;
  int var = 0;
  String fAddress="http://www.krages.com/ThePhotographersRight.pdf";

  try{ // RETRIEVE AND INSERT THE PDF INTO THE DISK FOR RETENTION.
    URL Url;
    byte[] buf;
    int ByteRead, ByteWritten=0;
    InputStream is=null;


    OutputStream outputstream=null;
    URLConnection Ucon=null;
    Url = new URL(fAddress);
    var = highestResourceKey();
    outputstream = new BufferedOutputStream(new FileOutputStream("C:\\Documents and 
Settings\\sony\\pliny\\.metadata\\.plugins\\uk.ac.kcl.cch.jb.pliny.pdfAnnot\\pdfCache\\c"+(var+1)+".pd
f" ));

    Ucon = Url.openConnection();
    is=Ucon.getInputStream();
    buf = new byte[size];
    while ((ByteRead = is.read(buf)) != -1 )  {
      outputstream.write(buf, 0, ByteRead);
      ByteWritten += ByteRead;
    }
    outputstream.close();
    System.out.println("ByteWritten: " + ByteWritten);
    System.out.println("Downloaded Successfully.");
    System.out.println("File name");
  } catch(Exception se) {
   // printSQLException(se) ;
       se.printStackTrace();
  }


  try { // TRY BLOCK NUMBER 2: UPDATE THE PLINY FIELDS FOR THE PDF INSERTION.
    Class.forName(driver) ;
    con = DriverManager.getConnection(url);
    dbmd = con.getMetaData() ;
    stmt=con.createStatement();
    //When in Pliny drag online pdf resource into pdf folder there are two resources rows will be
assigned in the PLINY.RESOURCE table
        //that's why string query0a was added to excute
        var = highestResourceKey();
                String query0="INSERT INTO PLINY.RESOURCE (FULLNAME, INITCHAR, OBJECTTYPEKEY,
IDENTIFIER, IDSTART, ATTRIBUTES, CREATIONDATE, CREATIONTIME)
VALUES('ThePhotographersRight.pdf','T',4,'','','#attr #Mon Aug 10 17:04:00 CDT 2009 cache="+(var+1)+"
url=file\\:/C:\\Documents and
Settings\\sony\\pliny\\.metadata\\.plugins\\uk.ac.kcl.cch.jb.pliny.pdfAnnot\\pdfCache/c"+(var+1)+".pdf
','2009-08-04','10:15:00')";
                String query0a="INSERT INTO PLINY.RESOURCE (FULLNAME, INITCHAR, OBJECTTYPEKEY,
IDENTIFIER, IDSTART, ATTRIBUTES, CREATIONDATE, CREATIONTIME)
VALUES('ThePhotographersRight.pdf','T',2,'" +(String fAddress)+"','" +(String fAddress)+"','#attr #Mon
Aug 10 17:04:00 CDT 2009 url=" +(String fAddress)+" ','2009-08-04','10:15:00')";
                String query1 ="INSERT INTO PLINY.RESOURCE (FULLNAME, INITCHAR, OBJECTTYPEKEY,
IDENTIFIER, IDSTART, ATTRIBUTES, CREATIONDATE, CREATIONTIME) VALUES('Derby-inserted
annotation','D',1,'','','','2009-08-04','10:17:00')";

                stmt.execute(query0);
                stmt.execute(query0a);
                stmt.execute(query1);

    String query2 = "INSERT INTO PLINY.NOTE (RESOURCEKEY, CONTENT, TSTAMP) VALUES ("+var+",'Yan
test Java','2009-08-04 10:17:00')";
    String query3 = "INSERT INTO PLINY.LINKABLEOBJECT (TYPEKEY, POSITION, DISPLPAGENO,
SURRPAGENO, DISPLAYEDINKEY, SURROGATEFORKEY, ISOPEN, SHOWINGMAP)
VALUES(1,'rect:12,25,190,99',1,0,"+(var-1)+","+var+",'Y','N')";
    String query4 = "INSERT INTO PLINY.LINKABLEOBJECT (TYPEKEY, POSITION, DISPLPAGENO,
SURRPAGENO, DISPLAYEDINKEY, SURROGATEFORKEY, ISOPEN, SHOWINGMAP)
VALUES(1,'rect:32,231,252,46',1,0,"+(var-1)+",0,'Y','N')";


        stmt.execute(query2);
        stmt.execute(query3);
        stmt.execute(query4);

        String query1B = "SELECT IDENTITY_VAL_LOCAL() FROM PLINY.LINKABLEOBJECT";
        rs=stmt.executeQuery(query1B);
        int val2=0;
        while(rs.next()) {
          val2 = rs.getInt(1);
        }

        String query5 ="INSERT INTO PLINY.LINK (ATTRIBUTES, FROMLINK, TOLINK, TYPEKEY) VALUES ('',
" +val2+","+(val2-1)+", 1)";
        stmt.execute(query5);
    rs.close();
    stmt.close();
System.out.println("\n--------------------------------------------------") ;
System.out.println("Database Name    = " + dbmd.getDatabaseProductName()) ;
System.out.println("Database Version = " + dbmd.getDatabaseProductVersion()) ;
System.out.println("Driver Name      = " + dbmd.getDriverName()) ;
System.out.println("Driver Version   = " + dbmd.getDriverVersion()) ;
System.out.println("Database URL     = " + dbmd.getURL()) ;
System.out.println("--------------------------------------------------") ;
   // Use the database connection somehow.

  } catch (SQLException se) {
    // printSQLException(se) ;
       se.printStackTrace();
  } catch (ClassNotFoundException e) {
    System.out.println("JDBC Driver " + driver + " not found in CLASSPATH") ;
  } finally {
    if(con != null){
      try {
        con.close() ;
      } catch (SQLException se) {
        printSQLException(se) ;
      }
    }
  }
}

static void printSQLWarning(SQLWarning sw) {
  while(sw != null) {

    System.out.print("SQLWarning: State=" + sw.getSQLState()) ;
    System.out.println(", Severity = " + sw.getErrorCode()) ;
    System.out.println(sw.getMessage());

    sw = sw.getNextWarning();
  }
}

static void printSQLException(SQLException se) {
  while(se != null) {

    System.out.print("SQLException: State:   " + se.getSQLState());
    System.out.println("Severity: " + se.getErrorCode());
    System.out.println(se.getMessage());

    se = se.getNextException();
    }
}

static int highestResourceKey() {
    int var = 0;
    Connection con = null;
    DatabaseMetaData dbmd = null;
    Statement stmt = null;
    ResultSet rs = null;
    String query1A = "SELECT IDENTITY_VAL_LOCAL() FROM PLINY.RESOURCE";
    try {
        Class.forName(driver) ;
        con = DriverManager.getConnection(url);
        dbmd = con.getMetaData() ;
        stmt=con.createStatement();
        rs=stmt.executeQuery(query1A);
        while(rs.next()) {
            var = rs.getInt(1);
        }
        rs.close();
        stmt.close();
        con.close();
    } catch (SQLException se) {
      // printSQLException(se) ;
            se.printStackTrace();
    } catch (ClassNotFoundException e) {
        System.out.println("JDBC Driver " + driver + " not found in CLASSPATH") ;
    } finally {
        if(con != null){
            try {
                con.close() ;
            } catch (SQLException se) {
                printSQLException(se) ;
            }
        }
    }
    return var;
}

} // public class testPlinyPDF2
```