

RDF Data and Image Annotations in ResearchSpace

Vladimir Alexiev
Ontotext

47A Tsarigradsko Shosse,
Sofia 1124, Bulgaria
+359 2 974 61 60

vladimir.alexiev@ontotext.com

Stanislav Kostadinov
Sirma ITT

135 Tsarigradsko Shosse, fl 3
1784 Sofia, Bulgaria
+359 82 83 83 95

stanislav.kostadinov@sirma.bg

Jana Parvanova
Ontotext

47A Tsarigradsko Shosse,
Sofia 1124, Bulgaria
+359 2 974 61 60

jana.parvanova@ontotext.com

ABSTRACT

This paper presents the approaches to data and image annotations in ResearchSpace (<http://www.researchspace.org>), an Andrew W. Mellon Foundation funded project led by the British Museum aimed at supporting collaborative internet research, information sharing and web applications for the cultural heritage scholarly community.

Categories and Subject Descriptors

H.3.1 Content Analysis and Indexing

K.4.3 [Organizational Impacts]: Computer-supported collaborative work

General Terms

Annotation. Museum applications. Cultural heritage online.

Keywords

ResearchSpace.

1. INTRODUCTION

The internet, global digitization efforts, Europe's Digital agenda, continuing investments in Europeana, the Digital Public Library of America and many other initiatives, have made millions upon millions of digitized cultural artifacts available on the net. From ancient maps, to bibliographic records, to paintings, to coins and hoards, to paleographic analysis, to prosopography factoids, to ancient manuscripts, to video interviews of Nazi victims... everything is becoming more and more connected. The question is how to make sense of all this information: how to aggregate it, find connections, build narratives, analyze the data, support the scientific discourse, engage users.

There are two key interconnected issues which any application in the cultural heritage domain needs to resolve: what data model to use so that the richness of data is not compromised in the digital world, and how to facilitate different types of users in the process of creation and use of the resources.

Annotation, the process of taking and adding notes to digital resources, is one of the key activities for the study and engagement with digital resources; there are dozens of annotation

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from Permissions@acm.org.

DH-case '13, September 10 2013, Florence, Italy
Copyright 2013 ACM 978-1-4503-2199-0/13/09...\$15.00.
<http://dx.doi.org/10.1145/2517978.2517997>

tools (see [1] for a selection of tools facilitating digital research, and [2] for an overview of tools for semantic annotation). With the change of the environments it also evolves and one of the key issues currently is how to integrate different types of annotation into environments which offer support for various users and activities.

This paper addresses this angle of annotation as part of larger environments and presents from a practical perspective solutions adopted for data and image annotation in ResearchSpace. In Section 2 it introduces briefly ResearchSpace; Section 3 looks into the annotation approach adopted, and finally section 4 presents conclusions and ideas for the future.

2. RESEARCHSPACE

ResearchSpace (RS) is a project of the British Museum (London), funded by the Andrew W. Mellon Foundation (USA). It aims to support collaborative research projects for cultural heritage scholars. RS implements an open source framework and hosted environment for web-based research, knowledge sharing and web publishing. RS intends to provide:

- Data conversion and aggregation;
- Semantic RDF data sources, based on the CIDOC CRM ontology;
- Semantic search based on fundamental relations (FRs);
- Data analysis and management tools;
- Collaboration tools, such as forums, tags, data baskets, sharing, dashboards;
- A range of research tools to support various workflows, e.g. Image Annotation, Image Compare, Timeline and Geographical Mapping;
- Web Publication.

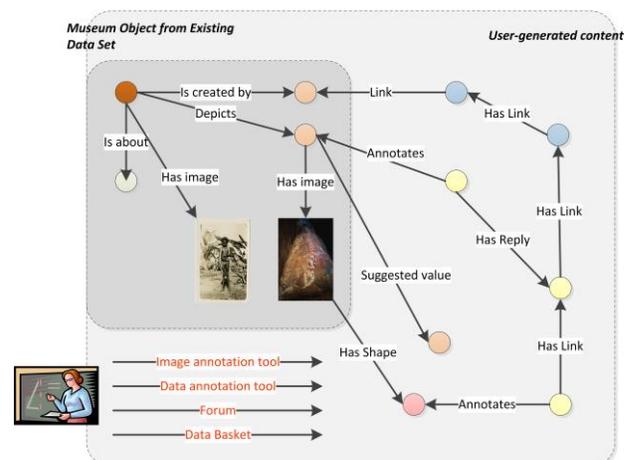


Figure 1. ResearchSpace – using various tools to add user-generated content

From technical perspective, ResearchSpace's goal is to host a variety of existing or newly-created tools that would facilitate collaborative research. Data Annotation and Image Annotation tools are among the basic new tools that have been developed and used in RS.

The basic artifacts in ResearchSpace are:

- *RDF representation of museum objects.* The objects are represented in deep CIDOC-CRM data structures. A typical object's graph consists of 100-400 nodes and is 6-10 levels deep
- *Collection of images related to each object.* Images vary from small pictures suitable for web publication to large gigapixel photos intended for scientific usages.
- *User-generated content – annotations, tags, links – on both Museum Objects and Images*

Data models that have been applied include:

- CIDOC-CRM for cultural heritage data
- OAC for annotations, tags, links
- SKOS for thesauri

A consortium led by Ontotext did software development for RS Stage 3 (Working Prototype) between November 2011 and April 2013. Development of Stage 4 is expected to start in 2013, with more museums and galleries coming on board.

3. THE ANNOTATION APPROACH IN RESEARCH SPACE

ResearchSpace offers two types of annotation – data and image annotation. Data and Image annotation tools are very important part of ResearchSpace and are its most enhanced mechanism for creating user-generated content.

The annotations in ResearchSpace have the following features:

- They can target a variety of objects: museum objects, images, specific statements about objects or arbitrary user-defined regions inside images
- They form threaded discussions.
- Annotation's content is in rich-text format.
- Links can be embedded in annotations – both to external and internal entities (incl. saved searches, other annotations etc.). RS provides reports about linking-by-annotation (e.g. for any selected object we can see all related objects whose annotations make a mention of it).

Both tools feature rich and dynamic user interface, implemented in JavaScript. The technical approach allows for easy reuse in other applications or even as standalone tools.

Annotations are stored using a version of OAC available prior to final specification. Certain extensions are made to incorporate additional information (replies, shapes in an image etc.).

An example of objects graph for a data annotation on a statement about a Rembrandt's painting is presented on Fig. 2.

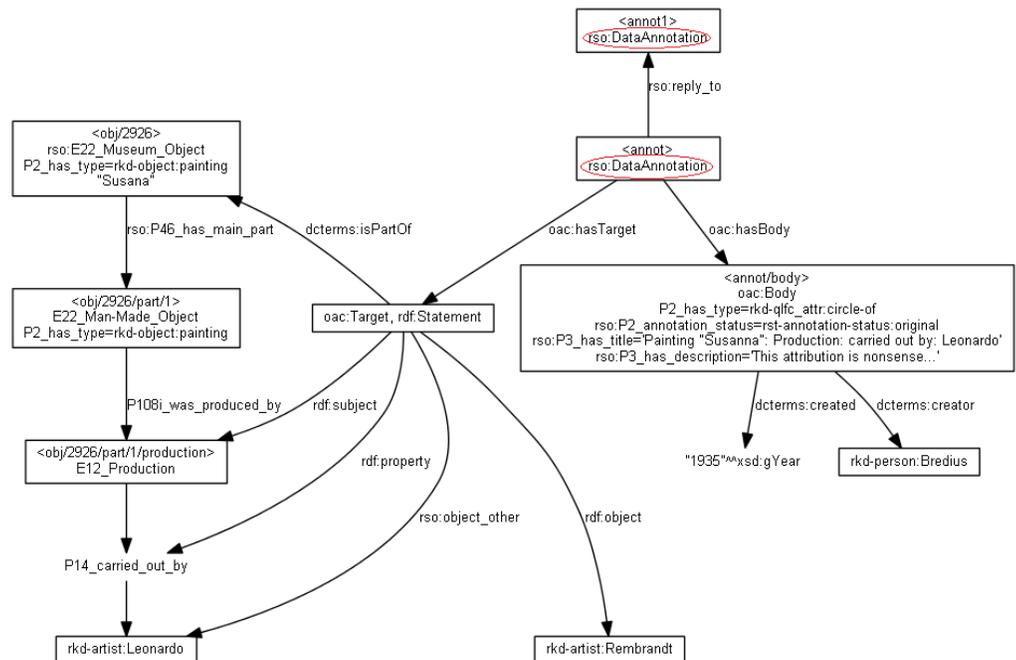


Figure 2. An example of objects graph for a data annotation.

3.1 Data Annotation

Data annotations target specific statements about the objects, e.g. date of production, frame's material, subject or creator (see Fig. 3). A statement is addressed via the object's URI and a triple in the object's graph.

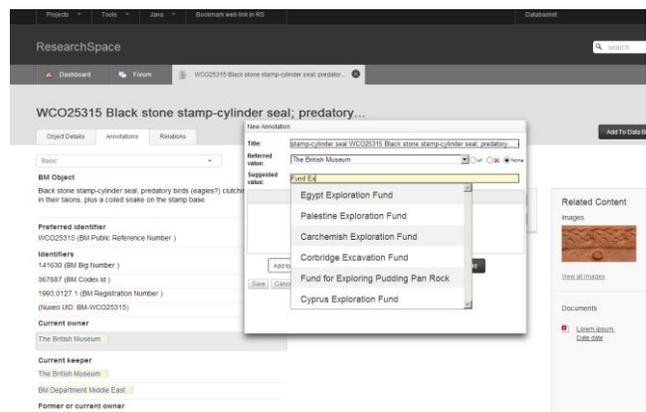


Figure 3. A screen shot on data annotation.

In addition to adding a comment or reply on the statement, users can agree, disagree or suggest an alternative statement. For example, users may disagree on the given list of creators and may suggest another artist to be added to it.

These suggested values are a subject to approval/rejection workflow performed by authorized users. Fig. 4 presents the typical lifecycle of facts/statements in a project.

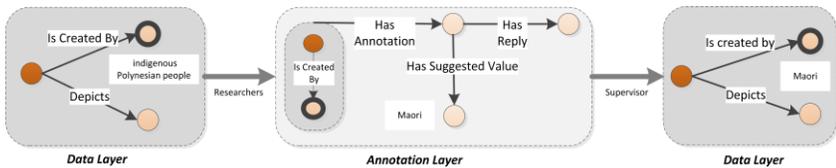


Figure 4. The lifecycle of annotated data.

1. Existing objects are imported into system. Each object is a graph of statements. Import can be done as a part of migration procedure for a data set or on demand.
2. Researchers can access objects via search, links, and data basket items. They can add comments on the facts about object, reply to previous comments, express agreement/disagreement, suggest alternative facts. All this forms a discussion about the object and is a convenient way for supporting team work.
3. Users with special roles (supervisors) can review discussions, and can either approve or reject suggested statements. This becomes part of the project's output. New facts and related argumentation can be published.

3.2 Image Annotation

Image annotations tool allows users to create arbitrary shapes on images and add their comments or notes on them. Again, other users can reply to these notes, forming a discussion and engaging with the digital resources.

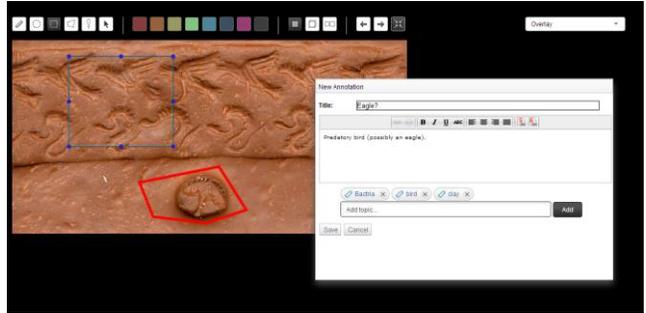


Figure 5. A screen shot on selecting an image area and annotation.

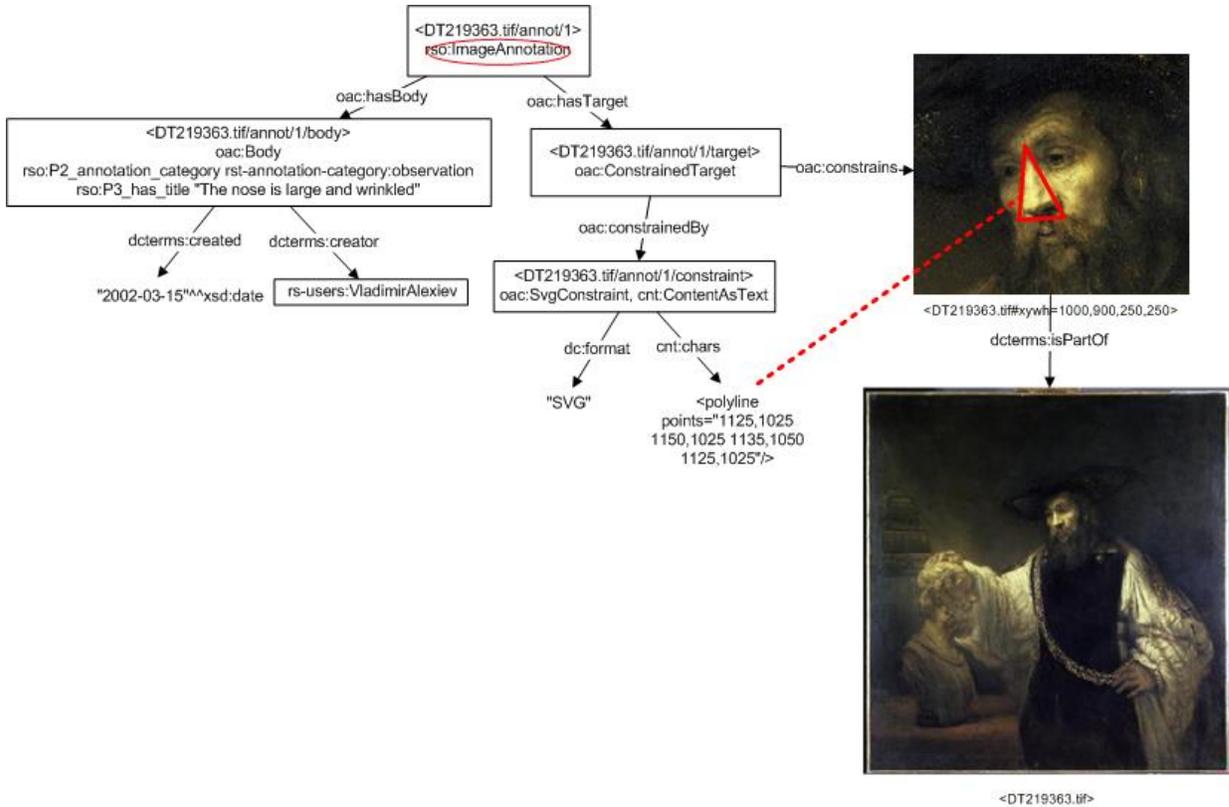


Figure 6. An example of describing regions.

(this feature will be implemented in Stage 4).

Each annotation can contain links to data records, images, saved searches, other annotations as well as external links. Annotations can be filtered and sorted.

A list of annotations that mention a given entity is also supported. References through both supplied and user-generated content can be traced so that user-generated content implicitly creates links between entities.

Predefined and custom tags can be applied to annotations.

The RDF data is visualized using a modification of the rforms open-source project [3].

Image annotation tool is combined with deep-zoom capabilities and supports high-resolution images. IIPImage server (<http://iipimage.sourceforge.net>) is used to provide tiles in the proper resolution. Annotations can be created at any zoom level and when annotation is visualized, original setup is restored.

Image annotation tool is based on svg-edit open-source project (for the editor see [4]). It is used to construct vector images based on XML successor language called SVG (Scalable Vector Graphics). SVG rendering is supported by all major contemporary browsers.

The tool also supports annotations on images based on just a URL, without zooming capabilities. The image viewer part consists of one or more raster layers and a vector layer where shapes are created and manipulated. More than one raster layer is used when user overlays or blends images (experimental features). Users can control opacity of raster layers.

This tool provides filtration by zoom level and users can switch to mode where only shapes from the current zoom are visible. Another useful functionality is the full screen mode where the image is maximized to the whole browser window.

Regions are described using SVG and OAC model is extended to handle this additional information (see Fig. 6).

The tool is implemented in JavaScript and loads data dynamically via AJAX (see Fig. 7).

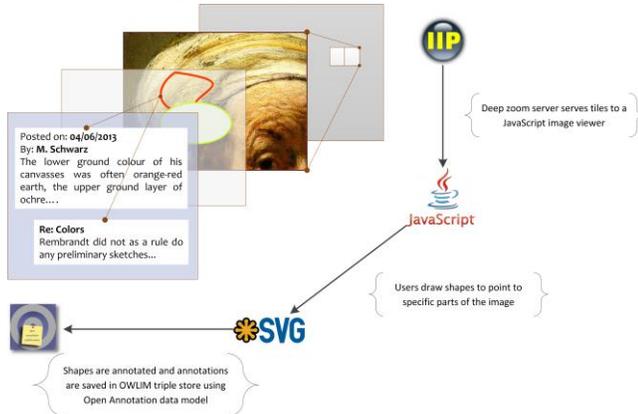


Figure 7. Implementation components for image annotation in ResearchSpace.

4. Conclusion, Future

This paper presents ongoing work on data and image annotations which are conceived as tools to answer the needs of an integrated environment for the presentation, discovery and use of cultural heritage objects. The tools allows integrating the annotation data naturally within the generic data structures used, and in the particular case means that the annotations are benefitting from the overall linked open data approach adopted in ResearchSpace.

After the working prototype stage, a number of presentations and discussions were organized by the project's owner - the British Museum. Based on the received feedback and the general goals and priorities of the ResearchSpace project, the following enhancements are currently under discussion:

- Further improvements in the presentation of RDF data. In next stages we will focus on improving presentation so that it becomes more user-friendly for non-technical users. We consider different approaches to this problem that would allow users to enter data or content in more unstructured way (even free text) and RS would be able to relate it to well-structured data.
- We will work on implementing more image manipulation techniques that would bring more value to RS's abilities for scientific research. One of the

techniques that we plan to look into is RTI (Reflectance Transformation Imaging).

- The process of publishing of research results is still to be well defined and implemented. Two of the major topics to consider there are: results approval workflow and publishing reproducible research materials as part of the results

From a more generic point of view, there are a number of issues around annotation of cultural resources and semantic annotations in particular; some key issues were discussed in [5]:

- Production versus post-production annotation – in the case of the presented tools they can be used at both stages, and by different types of users.
- Generic vs task-specific annotation – the tools again offer flexibility of use and can be applied in both cases; in the further months user studies will help to summarize the feedback of specialized and general uses.
- Manual versus automatic annotation and the "Semantic Gap" – our tools are supporting manual annotation, with a particular strength in facilitating the collaboration of groups of users.

A further issue is the reliability of user annotations, which was discussed in [6]; in the case of ResearchSpace the current usage is oriented towards collaboration of curators and researchers, and also with the introduction of users in different roles and having the special role of a supervisor who can review discussions, and can either approve or reject suggested statements.

5. ACKNOWLEDGMENTS

We would like to express our gratitude to Dominic Oldman, the principal investigator of ResearchSpace.

6. REFERENCES

- [1] BAMBOO DiRT (Digital Research Tools). <http://dirt.projectbamboo.org/categories/annotation>
- [2] Andrews, P., Zaihrayeu, I., and Pane, Juan (2010) A Classification of Semantic Annotation Systems. Technical report. <http://eprints.biblio.unitn.it/1934/>
- [3] rforms. Javascript library providing configurable forms for editing and presenting RDF <https://code.google.com/p/rforms/>
- [4] svg-edit, open source editor. <http://svg-edit.googlecode.com/svn-history/r1771/trunk/editor/svg-editor.html>
- [5] van Ossenbruggen, J., Troncy, R., Stamou, G., and Pan, J. (eds) 2007. Image Annotation on the Semantic Web. W3C. http://www.w3.org/2001/sw/BestPractices/MM/image_annotation.html
- [6] Nowak, S., and Rüger, S. 2010. How reliable are annotations via crowdsourcing: a study about inter-annotator agreement for multi-label image annotation. In *Proceedings of the international conference on Multimedia information retrieval (MIR '10)*. ACM, New York, NY, USA, 557-566. DOI= <http://doi.acm.org/10.1145/1743384.1743478>