Project Acronym: Europeana v2
Grant Agreement number: 270902
Project Title: Europeana Version 2

D7.3: REPORT ON INNOVATIVE APPLICATIONS

Revision	FINAL
Date of submission	30 May 2014
Author(s)	Maarten Brinkerink, Nederlands Instituut voor Beeld en Geluid
Dissemination Level	Public

REVISION HISTORY AND STATEMENT OF ORIGINALITY

Revision History

Revision No.	Date	Author	Organisation	Description
0.5	26/5/2014	Maarten Brinkerink	Beeld en Geluid	First draft, to gather feedback from WP7
0.6	27/5/2014	Maarten Brinkerink	Beeld en Geluid	Second draft, based on WP7 feedback and suggestions
0.7	28/5/2014	Maarten Brinkerink	Beeld en Geluid	Pre-final

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise. Acknowledgement of previously published material and of the work of others has been made through appropriate citation, quotation or both.

Table of Contents

1. Introduction

This deliverable reports on the results of the further development of the innovative applications, as part of Task 7.2. The plans for this further development have been described in MS40 'Report on innovative applications, second iteration'[1]. This was a plan that built upon the work reported on in 'MS39: Report on innovative applications, first iteration', as part of Subtask 7.1.1., to select a number of key applications that were beneficial to Europeana and the wider European cultural heritage domain. A first cycle of further development of these innovative applications has already happened, following the specifications that were part of 'D7.5: Functional specifications for social semantic functions (incl. prototype description)'[2]. Each of the applications that have been further developed, are described in this report; *Waisda for Europeana*, *PyBossa for Europeana* and *Annotorious for Europeana*. Focus of the report is on the new features that resulted from the further development.

---

[1] http://pro.europeana.eu/group/europeana-v2.0/documents?p_p_auth=NuB6B9cv&p_p_id=20&p_p_lifecycle=1&p_p_state=exclusive&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_20_struts_action=%2Fdocument_library%2Fget_file&_20_groupId=866067&_20_folderId=983534&_20_name=65842
[2] http://pro.europeana.eu/group/europeana-v2.0/documents?p_p_auth=c66gpuOe&p_p_id=20&p_p_lifecycle=1&p_p_state=exclusive&p_p_mode=view&p_p_col_id=column-2&p_p_col_count=1&_20_struts_action=%2Fdocument_library%2Fget_file&_20_groupId=866067&_20_folderId=1660973&_20_name=50007

2. Waisda? for Europeana

In MS40 the further development of *Waisda? for Europeana* within the context of Task 7.2 was described. It consists of four different types of further development:

1. Connecting Waisda? to OpenSKOS instances, allowing for a generic method to utilize controlled-vocabularies that are available in a SKOS format for the matching mechanisms of the game.

2. Extending Waisda? with a harvestable export mechanism, allowing Europeana – and potentially also other third parties – to reuse the user generated metadata that results from Waisda?

3. Creating an updated Waisda? prototype for a multitude of video collections that are available through Europeana, to allow the Europeana Network to investigate the possibilities that Waisda? offers them.

4. Updating the open source code repository of Waisda?, allowing third parties to also benefit from the above mentioned further development.

2.1 OpenSKOS Connector

OpenSKOS[3] is a repository and management system for controlled vocabularies that have been modelled according tot the SKOS[4] format. It provides users with the ability to manage several controlled vocabularies, edit terms and provides web services on top of its respository. OpenSKOS is a CATCHPlus[5] project and is currently being utilized in the production environment of both the Europeana Foundation (EF) and Nederlands Instituut voor Beeld en Geluid (Beeld en Geluid).

The Waisda? code base has been expanded with an OpenSKOS Connector, allowing for a generic method to utilize controlled vocabularies that are available in a SKOS format for the matching mechanisms of the game. The OpenSKOS Connector utilizes the API that OpenSKOS provides on top of the SKOS framework, to connect different instances of OpenSKOS to the tag matching mechanisms of *Waisda? for Europeana*.

Added benefit of this OpenSKOS Connector is the fact that this also opens opportunities for multilingual vocabularies. The previous instance of *Waisda? for Europeana* made use of the Dutch GTAA[6] thesaurus utilized by Beeld en Geluid. This restricted users who were tagging the (also) Dutch newsreel footage to using Dutch terms to describe the content, if they aimed at matching the controlled vocabulary terms. The OpenSKOS Connector allows the game to be configured in such a way that it uses different controlled vocabularies for different sets of content. For *Europeana 1914-1918* Europeana has deployed its own OpenSKOS instance[7], containing terms in various categories such as military aspects, operations, campaigns, and social aspects. All the concepts are translated in seven languages and are linked to the Library of Congress Subject Headings[8]. The Waisda? prototype will be the first project to utilize this vocabulary from the OpenSKOS repository hosted by EF. This will allow users to utilize seven different languages when tagging the video content with relevant concepts.

---

[3] http://openskos.org/
[4] http://www.w3.org/2004/02/skos/
[5] http://www.catchplus.nl/
[6] http://gtaa.beeldengeluid.nl/
[7] http://skos.europeana.eu/api
[8] http://id.loc.gov/authorities/subjects.html

2.2 User Generated Metadata Exporter

To enable reuse of the social tags by Europeana, the integration of the application with Europeana mentioned in MS39, the *Waisda? for Europeana* prototype has been extended with a harvestable export mechanism, allowing third parties to extract the user generated metadata from the database of the game. This data has been made harvestable via the Open Archives Initiative Protocol for Metadata Harvesting[9], which already forms the basis of the Europeana aggregation workflow.

User generated metadata that was gathered within the game has been modelled following the Open Annotation data model[10]. The implementation has been the result of an in-depth exchange between Beeld en Geluid and the Europeana Foundation R&D team. This first version of the export mechanism for the user generated metadata focusses on exporting the terms that have matched controlled-vocabularies (including time codes relative to the video items and URIs of the corresponding vocabulary terms).

The data feed for the User Generated Metadata Exporter can be found here:
http://waisda.tuxic.nl:8080/oai?verb=ListRecords&metadataPrefix=rdf

This is an example of the output format:

```
<oa:Annotation rdf:about="oai:waisda.beeldengeluid.nl:annotation:1311">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#">
  <oa:hasTarget>
   <rdf:type rdf:resource="http://www.w3.org/ns/oa#SpecificResource">
   <rdf:type rdf:resource="http://purl.org/dc/dcmitype/MovingImage">
   <oa:hasSource rdf:resource="http://www.openbeelden.nl/media/2583/"/>
   <oa:hasSelector>
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#FragmentSelector">
    <rdf:value>t=npt:14,15</rdf:value>
    <dcterms:conformsTo rdf:resource="http://www.w3.org/TR/media-frags/"/>
   </oa:hasSelector>
  </oa:hasTarget>
  <oa:hasBody>
   <oa:Tag>
      <rdf:type rdf:resource="http://w3.org/2011/content#ContextAsText">
      <cnt:chars>veluwe</cnt:chars>
   </oa:Tag>
  </oa:hasBody>
 </oa:Annotation>
 <oa:Annotation rdf:about="oai:waisda.beeldengeluid.nl:annotation:1312">
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#">
  <oa:hasTarget>
   <rdf:type rdf:resource="http://www.w3.org/ns/oa#SpecificResource">
   <rdf:type rdf:resource="http://purl.org/dc/dcmitype/MovingImage">
   <oa:hasSource rdf:resource="http://www.openbeelden.nl/media/2583/"/>
   <oa:hasSelector>
    <rdf:type rdf:resource="http://www.w3.org/ns/oa#FragmentSelector">
    <rdf:value>t=npt:18,19</rdf:value>
    <dcterms:conformsTo rdf:resource="http://www.w3.org/TR/media-frags/"/>
   </oa:hasSelector>
  </oa:hasTarget>
  <oa:hasBody>
   <oa:Tag>
      <rdf:type rdf:resource="http://w3.org/2011/content#ContextAsText">
```

---

[9] http://www.openarchives.org/pmh/
[10] http://www.openannotation.org/spec/core/

```
            <cnt:chars>brand</cnt:chars>
        </oa:Tag>
      </oa:hasBody>
    </oa:Annotation>
    <oa:Annotation rdf:about="oai:waisda.beeldengeluid.nl:annotation:1313">
        <rdf:type rdf:resource="http://www.w3.org/ns/oa#">
        <oa:hasTarget>
         <rdf:type rdf:resource="http://www.w3.org/ns/oa#SpecificResource">
         <rdf:type rdf:resource="http://purl.org/dc/dcmitype/MovingImage">
         <oa:hasSource rdf:resource="http://www.openbeelden.nl/media/2583/"/>
         <oa:hasSelector>
           <rdf:type rdf:resource="http://www.w3.org/ns/oa#FragmentSelector">
           <rdf:value>t=npt:23,24</rdf:value>
           <dcterms:conformsTo rdf:resource="http://www.w3.org/TR/media-frags/"/>
         </oa:hasSelector>
        </oa:hasTarget>
        <oa:hasBody>
         <oa:Tag>
           <rdf:type rdf:resource="http://w3.org/2011/content#ContextAsText">
           <cnt:chars>brandweer</cnt:chars>
         </oa:Tag>
        </oa:hasBody>
    </oa:Annotation>
```

2.3 Updated Content Offering

An updated prototype for *Waisda? for Europeana* can be found at
http://waisda.tuxic.nl:8080/. Apart from the incorporation of the above-mentioned
functionality, new content has been added to the prototype. It shows how collections included
in Europeana can also be added into the prototype. In collaboration with EF a series of video
interviews[11] that have been produced in the context of Europeana 1914-1918[12] has been
selected for inclusion. It fits the criteria for inclusion in *Waisda? for Europeana*, which include
the fact that it is harvested by Europeana, that it is a video collection and the fact that direct
links to the digital objects are available (in a video format supported by JW Player[13]). It is also
openly licensed for reuse under a Creative Commons license (CC-BY-SA)[14]. The collection is
featured next to the currently already implemented newsreel collection from Sound and
Vision.

The above-mentioned inclusion of another video collection within the game also provides a
perfect opportunity to showcase the potential of the OpenSKOS Connector. To demonstrate
this the new content is configured for matching against the above-mentioned instance of
OpenSKOS hosted by EF. Because both the content and the vocabulary are results from of
Europeana 1914-1918, this provides a suitable test scenario.

2.4 Code Repository Update

All of the above-mentioned changes to the Waisda? code base have been made available as
'branches' on Github, including documentation and released as open source software (under
a GNU General Public License version 3[15]):

---

[11]
http://europeana.eu/portal/search.html?query=provider_aggregation_edm_isShownBy:*mp4&rows=24&qf=PROVI
DER:%22Europeana+1914-1918%22
[12] http://www.europeana1914-1918.eu/
[13] http://www.jwplayer.com/
[14] https://creativecommons.org/licenses/by-sa/3.0/
[15] http://gplv3.fsf.org/

Code for the OpenSKOS Connector can be found here:
https://github.com/beeldengeluid/waisda/tree/OAC_SKOS_MERGE

A section called 'Using RESTService instead of dictionary' has been added to the documentation and described how to utilize the OpenSKOS Connector. It can be found here:
https://github.com/beeldengeluid/waisda/blob/OAC_SKOS_MERGE/docs/customizing.md#using-restservice-instead-of-dictionary

Code for the User Generated Metadata Exporter can be found here (including documentation):
https://github.com/beeldengeluid/waisda/tree/WAISDA_OAIPMH_EXPORTER

3. PyBossa for Europeana

There is one important deviation from the plans reported in MS40 and D7.5 to mention here: Because of circumstances beyond the control of WP7, further development of *PyBossa for Europeana* has regrettably been cancelled. The main reason has been the difficulty to influence the roadmap and activities of the open source community that forms the core of PyBossa. PyBossa is a project initiated by the Open Knowledge Foundation (OKFn).

Although there is active development on the source code of the PyBossa project[16], most activity is project related. Apart from that, the community is limited to Python developers (the code language the project is based on). Voluntary developer resources on the side of the Open Knowledge Foundation have proved to be very scarce, since OKFn related developers mostly voluntarily contribute to open source project like PyBossa, next to their professional assignments (which understandably receive higher priority).

The developers that contributed to the first instalment have been unavailable for further development as a result of the above-mentioned reality of this open source project. To mitigate this setback, development on a third prototype – based on Annotorious[17] – has been instigated. More on this in the next section.

Previous work on *PyBossa for Europeana* in the context of Task 7.2 remains available at the previously reported locations, and has been archived at Beeld en Geluid (for backup purposes):

- Prototype: http://culttag8.herokuapp.com/
- Europeana client API: https://github.com/mk270/europeana-search
- PyBossa 'upstream' features: https://github.com/PyBossa/pybossa
- Bespoke PyBossa tools: https://github.com/mk270/pybossa

---

[16] http://crowdcrafting.org
[17] https://annotorious.github.io/

4. Annotorious for Europeana

Annotorious is being developed under the leadership of the Austrian Institute of Technology (AIT). The software has its roots in the YUMA Universal Media Annotator prototype[18], developed as part of the EuropeanaConnect research project. Annotorious is licensed under the terms of the MIT License[19]. Because of the established nature of AIT and its familiarity with Europeana and involvement in several Europeana-related project, Annotorious was selected as an alternative 'innovative application' to mitigate the lack of further development of *PyBossa for Europeana*. It aims to show that the mechanisms that have been designed and – in the case of Waisda? – have been developed in the context of Task 7.2, can indeed also be implemented for the enrichment of still images.

4.1 Demo Page with Europeana API Connection

A demo page (see *Image 1*) that allows search against the Europeana API and which will retrieve full-resolution images for annotation (using an appropriate mechanism of the API, such as the *edm:isShownBy* fields in the Europeana API JSON response) has been created for *Annotorious for Europeana*. It is possible to specify a particular collection to which the search can be limited.
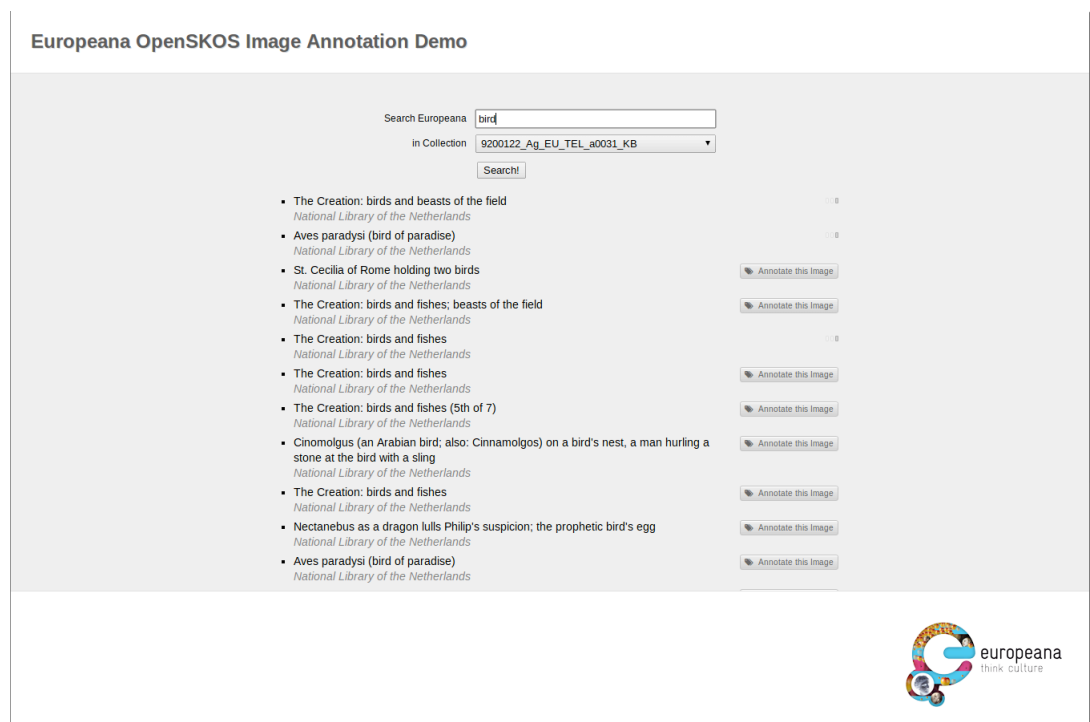
The demo page can be found here: http://ait-ngcms.github.io/annotorious-openskos-demo/

Image 1: Search results on the demo page, loading API results suitable for annotation

4.2 Semantic Tagging Widget with OpenSKOS Connection

A tagging widget, based on a text-entry box with auto-completion, has been implemented as an Annotorious plugin. Auto-completion is driven directly off the OpenSKOS prefix-search API method (see *Image 2*).

---

[18] http://pro.europeana.eu/thoughtlab/user-generated-content#EConnect
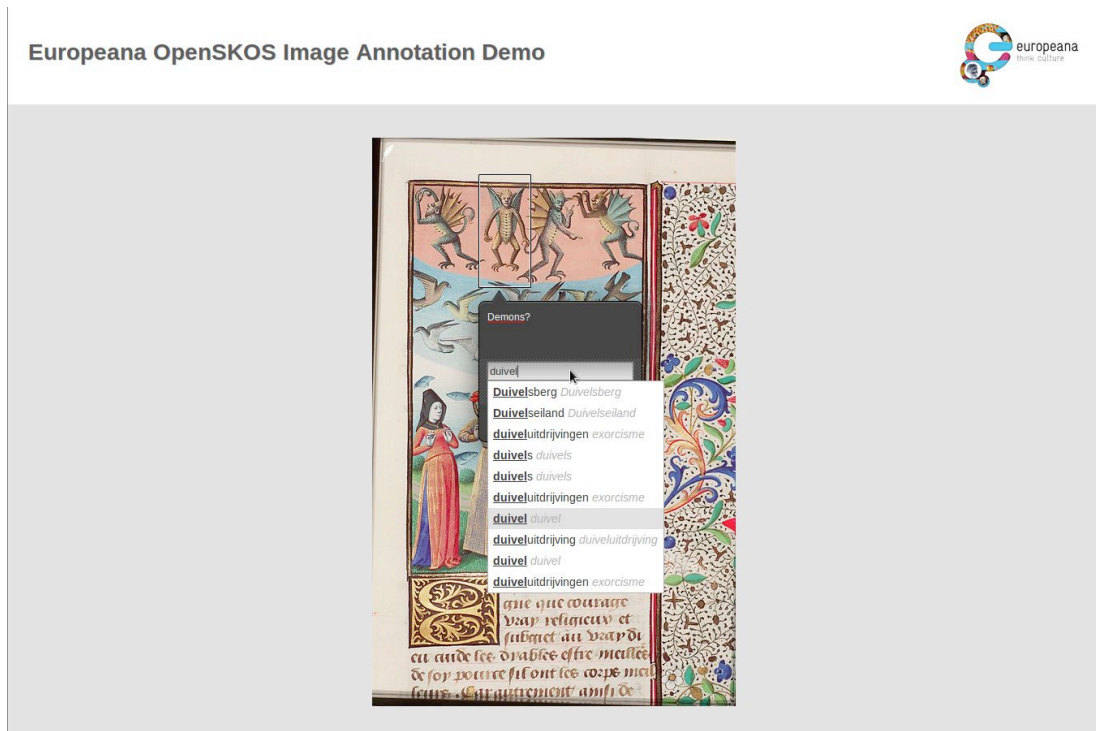[19] http://opensource.org/licenses/MIT

Image 2: Annotorious plugin with auto-completion running off of OpenSKOS

If the text box is left empty, users can see the tags that already exist for that specific item, with the option to remove these existing tags (see *Image 3*).
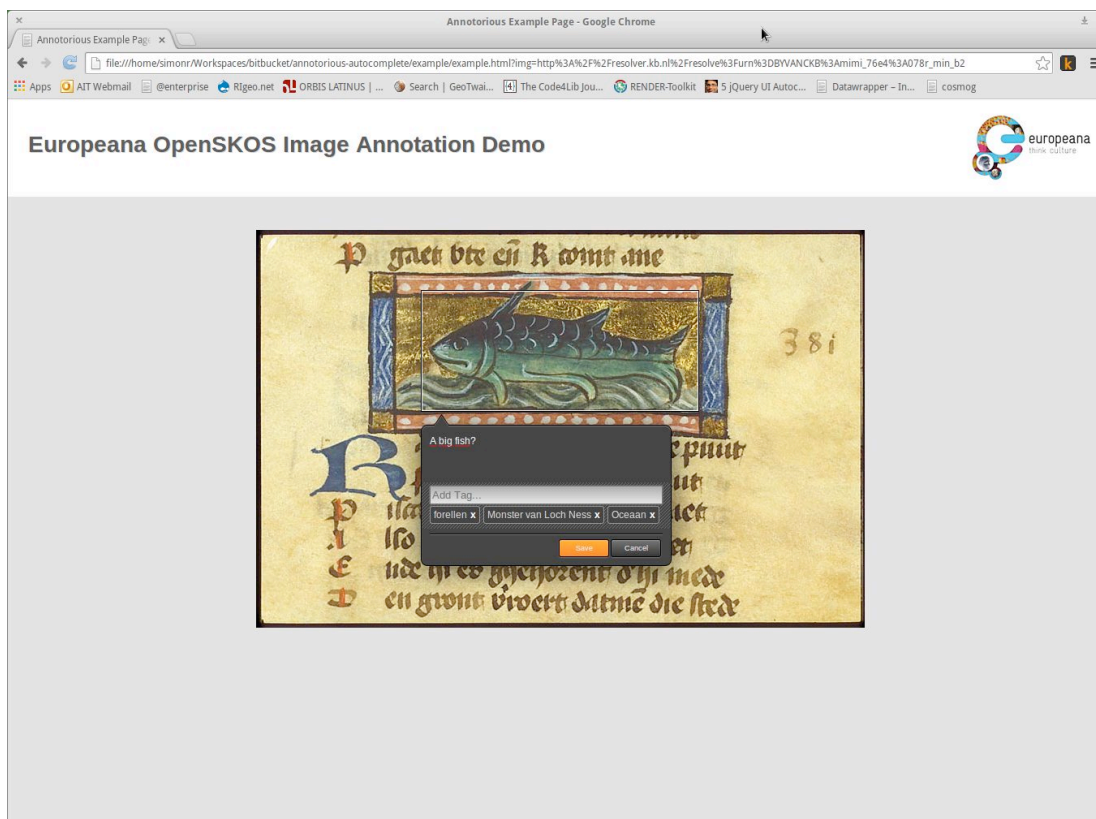

Image 3: Annotorious plugin showing previously added tags, with the option to remove/correct them

On mouse-over, Annotorious will display a popup showing the annotation text for that particular image (see *Image 4*).
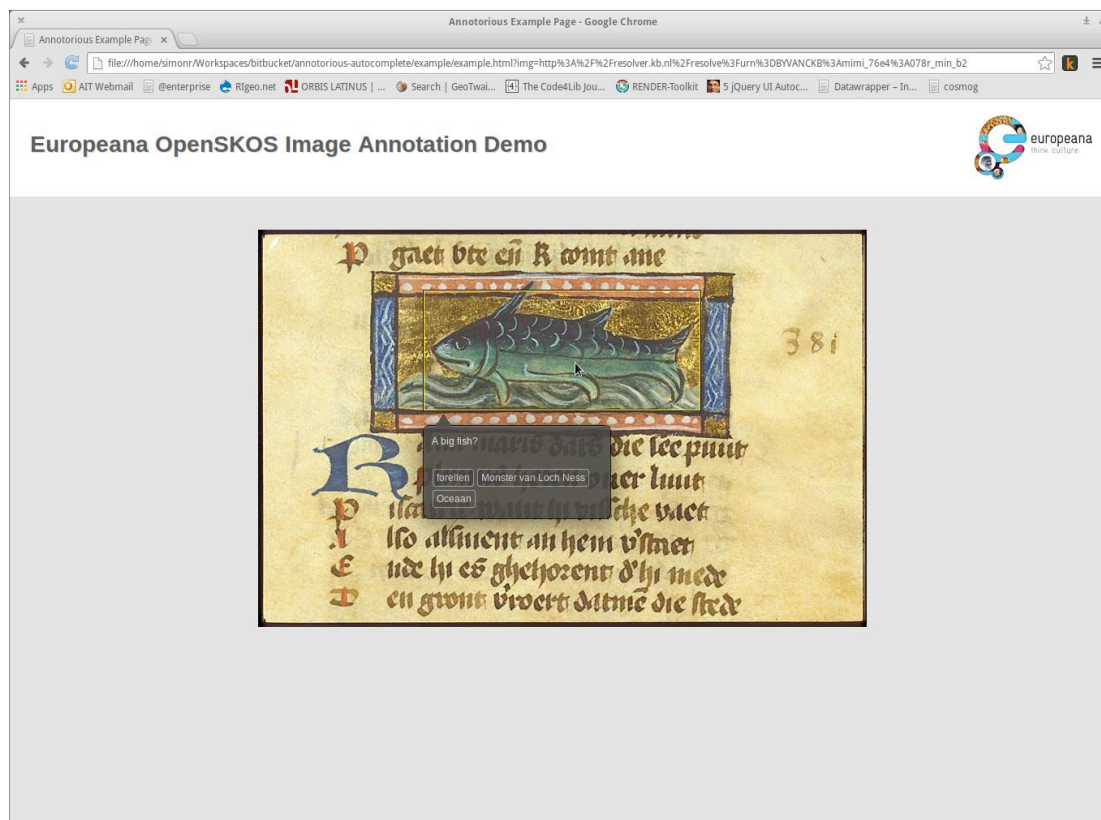

Image 4: Viewing existing annotations for an image, on mouse-over

For a secure JavaScript connection to this OpenSKOS API method, JSONP[20] support has been added to OpenSKOS. This addition to the OpenSKOS code base in the slipstream of this Task, is available as part of the open source repository for OpenSKOS at: https://github.com/CatchPlus/OpenSKOS

4.3 Code Release

All of the above-mentioned functional further development to the Annotorious code base, has resulted in the new code being released as open source software and documented on https://github.com/ait-ngcms/annotorious-openskos-demo/, under the terms of the above-mentioned MIT License.

---

[20] https://en.wikipedia.org/wiki/JSONP