Project Acronym: **Europeana Sounds**
Grant Agreement no: **620591**
Project Title: **Europeana Sounds**

D2.7 Crowdsourcing Evaluation and Impact Assessment

Revision: Final

Date: 03/08/2016

Authors: Maarten Brinkerink (NISV), Harry Van Biessumm (NISV), Alessio Piciolli (NET7), Sergiu Gordea (AIT), Remy Gardien (EF), Vassilis Tzouvaras (NTUA).

Abstract: This document initially reports on the status of the Crowdsourcing Infrastructure, with a focus on the Annotations API that is now in production. It then describes the progress with the development of the end-user facing crowdsourcing applications that have been connected to the crowdsourcing infrastructure to support the enrichment through crowdsourcing which the project aims for. The document then reports on the first experience with, and results gained from, the enrichment through crowdsourcing, by utilising and promoting the crowdsourcing applications in two specific crowdsourcing campaigns. Finally the deliverable reflects on the project's strategic approach towards enrichment through crowdsourcing and its progress towards the key performance indicators.

Dissemination level	
Public	X
Confidential, only for the members of the Consortium and Commission Services	

Revision history

Version	Status	Name, organisation	Date	Changes
0.1	ToC	Maarten Brinkerink, NISV	31/05/2016	Toc
0.2	Draft	Lise Schauer, HP	15/06/2016	Sections 3.1/3.2
0.3	Draft	Alessio Piciolli, NET7	26/06/2016	Section 3.3
0.4	Draft	Harry van Biessum, NISV	04/07/2016	Sections 4.1 / 4.2
0.5	Draft	Sergiu Gordea, AIT	11/07/2016	Section 2.2
0.6	Draft	Remy Gardien, EF	14/07/2016	Section 2.1
0.7	Draft	Maarten Brinkerink, NISV	15/07/2016	Extensive review and Chapter 5
0.8	Draft	Alessio Piccioli, NET7 & Vassilis Tzouvaras, NTUA	18/07/2016	Section 3.3
0.81	Draft	Breandán Knowlton, HP	29/07/2016	Section 3.1 / 3.2 revisions
0.9	Final Draft	Lora Aroyo, VU University Amsterdam, John Ashley Burgoyne, University of Amsterdam , Lizzy Komen, NISV, Joris Pekel, EF, Mia Ridge, BL & PMB members	25-29/07/2016	External review of the final draft
0.91	Final Draft	Maarten Brinkerink, NISV & Laura Miles, BL	01/08/2016	Final contributions, formatting, copy editing
1.0	Final	Richard Ranft, BL	02/08/2016	Layout, minor changes

Review and approval

Action	Name, organisation	Date
Reviewed by	Lora Aroyo, VU University Amsterdam John Ashley Burgoyne, University of Amsterdam Mia Ridge, BL	25-29/07/2016
Approved by	Coordinator and PMB	03/08/2016

Distribution

No.	Date	Comment	Partner / WP
1	03/08/2016	Submitted to the European Commission	BL/WP7
2	03/08/2016	Posted on Europeana Pro website	BL/WP7
3	03/08/2016	Distributed to project consortium	BL/WP7

Application area

This document is a formal output for the European Commission, applicable to all members of the Europeana Sounds project and beneficiaries. This document reflects only the author's views and the European Union is not liable for any use that might be made of information contained therein.

Statement of originality

This document 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.

Project summary

Europeana Sounds is Europeana's 'missing' fifth domain aggregator, joining APEX (Archives), EUscreen (television), the Europeana film Gateway (film) and TEL (libraries). It will increase the opportunities for access to and creative re-use of Europeana's audio and audio-related content and will build a sustainable best practice network of stakeholders in the content value chain to aggregate, enrich and share a critical mass of audio that meets the needs of public audiences, the creative industries (notably publishers) and researchers. The consortium of 24 partners will:

- Double the number of audio items accessible through Europeana to over 1 million and improve geographical and thematic coverage by aggregating items with widespread popular appeal such as contemporary and classical music, traditional and folk music, the natural world, oral memory and languages and dialects.

- Add meaningful contextual knowledge and medium-specific metadata to 2 million items in Europeana's audio and audio-related collections, developing techniques for cross-media and cross-collection linking.

- Develop and validate audience specific sound channels and a distributed crowd-sourcing infrastructure for end-users that will improve Europeana's search facility, navigation and user experience. These can then be used for other communities and other media.

- Engage music publishers and rights holders in efforts to make more material accessible online through Europeana by resolving domain constraints and lack of access to commercially unviable (i.e. out-of-commerce) content.

These outcomes will be achieved through a network of leading sound archives working with specialists in audiovisual technology, rights issues, and software development. The network will expand to include other data-providers and mainstream distribution platforms (Historypin, SoundCloud) to ensure the widest possible availability of their content.

For more information, visit http://pro.europeana.eu/web/europeana-sounds and http://www.europeanasounds.eu

Copyright notice

Contents

Executive summary: D2.7 Crowdsourcing Evaluation and Impact Assessment

The final system architecture for the Europeana Sounds crowdsourcing infrastructure has been designed and is almost fully implemented. At the core of the infrastructure lies the Annotations API. The Annotations API can now support up to 5 types of annotation, which are compatible with the W3C Web Annotation Data Model:

1. Simple Tagging
2. Semantic Tagging
3. Geotagging
4. Object Linking
5. Moderation

Europeana Sounds is continuing to develop a suite of 6 end-user facing crowdsourcing applications that are connected to the crowdsourcing infrastructure to support enrichment through crowdsourcing:

1. Tunepal widget
2. Historypin geotagging interface
3. Pundit (various clients)
4. WITH
5. Europeana Music Collection

Europeana gathered its first experience with the enrichment through crowdsourcing, by utilising and promoting two specific crowdsourcing campaigns in June 2016. One focussed on geolocating radio broadcasts, and utilised the Historypin geotagging interface. The other focussed on enriching music collections with instruments and utilised integration of Pundit and WITH. While the former campaign had disappointing results, the latter was successful (especially because of the combination of online and physical crowdsourcing activities).

Having worked on the challenges of enrichment and participation for two and a half years, and based on the first results with enrichment through crowdsourcing, WP2 predicts a substantial underperformance with KPI 9 "Number of annotations (tags) added by users".

Looking at the current crowdsourcing scenarios, WP2 proposes to focus the remaining project resources on Semantic Tagging and Moderation (more specifically Validation of Semi-automatic Tags). These scenarios do not necessarily result in the largest number of "annotations (tags) added by users" (KPI 9), but they can operate at scale, because they combine curation, crowd intelligence and machine intelligence. They also result in high quality annotations that; (1) unify and link metadata across records and collections, (2) unify and link metadata to the web, and (3) improve multilingual retrieval and representation.

1 Introduction

This document first reports on the status of the crowdsourcing infrastructure, with a focus on the Annotations API (see Chapter 2. Europeana Sounds crowdsourcing infrastructure). It then describes the progress with the development of the end-user facing crowdsourcing applications that have been connected to the crowdsourcing infrastructure to support the enrichment through crowdsourcing that the project aims for (see Chapter 3. Europeana Sounds crowdsourcing applications). The document then reports on the first experience with and results gained with enrichment through crowdsourcing, by utilizing and promoting the crowdsourcing applications in two specific crowdsourcing campaigns (see Chapter 4. Europeana Sounds crowdsourcing campaigns). Finally the document reflects on the project's strategic approach towards enrichment through crowdsourcing and its progress against the KPIs (see Chapter 5. Enrichment strategy and ambitions).

2   Europeana Sounds crowdsourcing infrastructure

This section describes the Europeana Sounds crowdsourcing infrastructure. It will update the information previously reported about the system architecture and then provides an extensive description of the Annotations API.

2.1   Final system architecture

Figure 1 depicts the final system architecture as it has been implemented for the Europeana Sounds crowdsourcing infrastructure. There are two differences between the system architecture as designed in D2.4 (Ref 1), and later reported on in D2.10 (Ref 2):

- The WITH platform, which previously was intended to connect directly with the Annotations API, now acts as a client to the Pundit tools (see Section 3.3 Pundit Integration in the WITH platform), which in return will have a connection to the Annotations API.

- The Historypin platform, which previously was being "crawled" by Europeana for new annotations, now pushes annotations directly to the Annotations API (see Section 3.2 historypin geotagging interface).

The Pundit integration with the Annotations API, the data provider's integration with the Pundit Server, and the Europeana Collections integration with the Annotations API are the three tasks which are still in the process of being implemented at this stage. The previously reported 'round tripping daemon' component in the system architecture (in D2.10, Ref 2) has since been forfeited, for the simple reason that none of the current crowdsourcing applications need it for their data flows. They all can directly 'push' their annotations to the Annotations API.
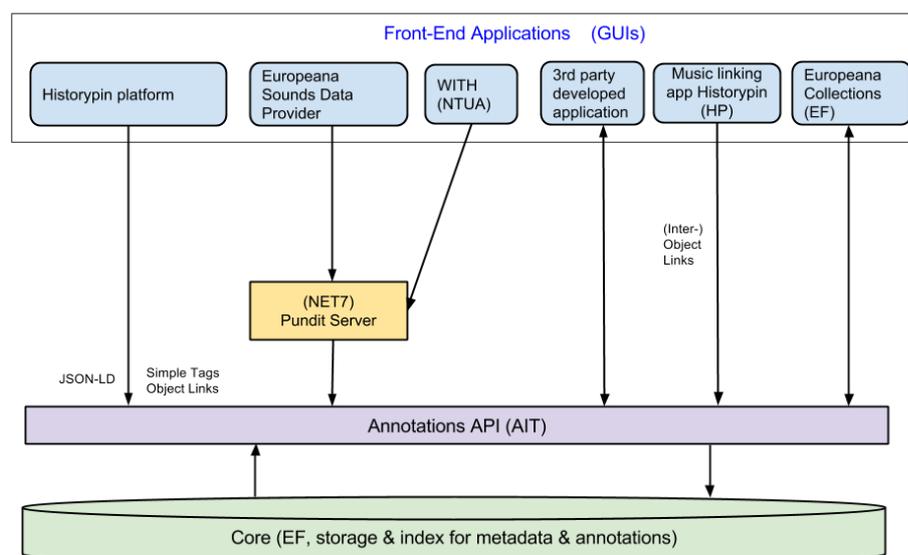

Figure 1: Final system architecture for the Europeana Sounds crowdsourcing infrastructure. All data flows are using JSON-LD as data format.
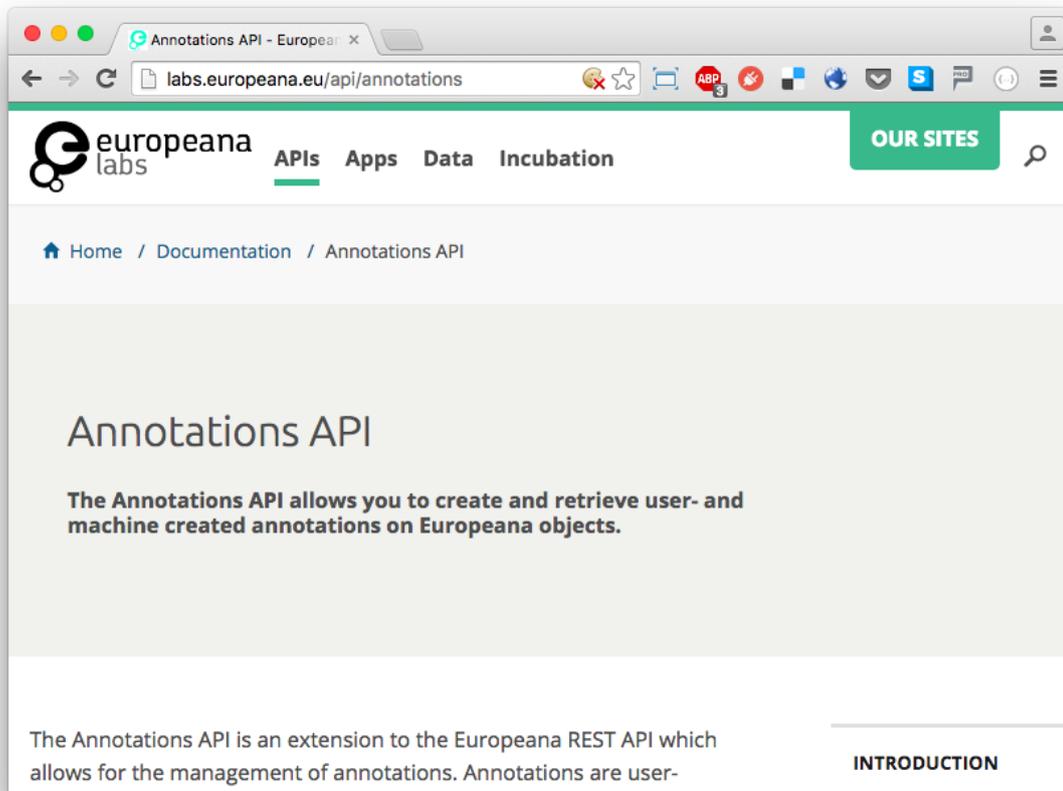
2.2   Annotations API

Figure 2: The Annotations API on Europeana Labs

The role of the Annotations API within the crowdsourcing infrastructure was first presented in D2.4 (Ref 1) and later reported on in D2.10 (Ref 2). The standardisation and interoperability are the main concerns taken into account for the design of the API, given the complexity of the crowdsourcing infrastructure and use cases, and the data flows exchanged between the different components. However, the aim for interoperability extends beyond the borders of the Europeana Sounds ecosystem.

To contribute to international and cross-sector standardisation and interoperability of annotations, the partners involved have actively participated in the development of the W3C Web Annotation Data Model[1] (WA), which is currently in the process of being released as a community recommendation (estimated release date is end of July 2016[2]). The collaboration between the W3C Web Annotation Working Group[3] was strengthened by the face-to-face meeting facilitated by the I Annotate conference[4] [5] (that was attended by various WP2 members[6]).

---

[1] https://www.w3.org/TR/annotation-model/
[2] http://w3c.github.io/web-annotation/model/wd2/#acknowledgments
[3] https://www.w3.org/annotation/
[4] http://iannotate.org/2016/
[5] http://pro.europeana.eu/page/issue-5-annotations#Article4

There were two public releases of the Annotations API in 2016, which were advertised in the Europeana Labs and included the planned development roadmap and changelog[7]. The first alpha release was opened for public testing in March 2016. This was communicated through various mailing lists (such as EuropeanaTech[8] and through the public[9] and closed[10] W3C Web Annotation list). The Annotations API also received its own section on Europeana Labs[11] (see Figure 2). This version included basic functionality for creating and retrieving various annotations types, including but not limited to: Simple Tagging, Semantic Tagging, and Object Linking.

The provided REST interface has already implemented the specifications of the Web Annotation Protocol[12], and was first released in July 2015 by the W3C Web Annotation Working Group[13]. The serialisation of the annotations followed the Web Annotation Data Model, but did not include the last minute changes of the March and July releases[14]. The second public release is fully compliant with the July 2016 version of the web annotation specifications which includes the data model; vocabulary and the protocol[15] (see Section 2.2.2.2 Build and Deployment).

The current functionality of the Annotations API, as available in the second public release, is described in detail in the next section of this document, and is included in the Business Logic and Technical Functionality sections.

2.2.1   Business logic

The development of the Annotations API was driven by the 'User Stories' and 'Epics' defined in D2.2 (Ref 3), which were further analysed in order to provide concrete specifications for individual application scenarios (our 'business logic'). Some of the Europeana Sounds application scenarios - as presented at the I Annotate conference [16] - are not yet fully covered by the Web Annotation Data Model. For these application scenarios several extensions have been proposed to the W3C Working Group by the Europeana Sounds project. It is expected that some of the application scenarios of Europeana Sounds and subsequent extensions will be considered for the second version of the Web Annotation Data Model[17].

---

[6] http://www.europeanasounds.eu/news/iannotateconference

[7] http://labs.europeana.eu/api/annotations-roadmap-changelog

[8] http://pro.europeana.eu/europeana-tech

[9] http://lists.w3.org/Archives/Public/public-annotation/

[10] http://lists.w3.org/Archives/Member/member-annotation/

[11] http://labs.europeana.eu/api/annotations

[12] https://www.w3.org/TR/annotation-model/

[13] https://www.w3.org/TR/2016/WD-annotation-protocol-20160331/

[14] https://www.w3.org/TR/2016/WD-annotation-model-20160331/

[15] http://w3c.github.io/web-annotation/

[16] The presentation given by the Europeana Sounds project representatives:
http://www.slideshare.net/Europeana_Sounds/challenges-on-modeling-annotations-in-the-europeana-sounds-project-62463781

[17] https://github.com/w3c/web-annotation/milestone/1

The API endpoints for creating, retrieving and searching annotations were introduced in D2.10 (Ref 2), and they have the same interfaces in the current version. However, support for more types of specific annotations was introduced, which are presented in the 2.2.1.1 Templates section. Another major functionality that was introduced in this version is the support for user feedback, which is presented in section 2.2.1.6 Moderation.

Apart from the business logic of the Annotations API, there were other technical and infrastructure enhancements including: better authentication and authorisation support, update of the application console, enhanced support for administration of annotations, improvements of the build and deployment infrastructure.

2.2.1.1 Templates

Given the fact that a primer document is not yet provided by the W3C Web Annotation Working Group for consolidating the common understanding and unique representation of different annotation types, the project partners collaborated to consolidate the representation of specific annotation types for the supported application scenarios[18]. Examples for each annotation type supported by the current version have been made available in the technical documentation[19] and in the application console of the Annotations API. These examples are used as templates by testers and client developers, ensuring a correct syntactic representation of annotations.

2.2.1.2 Simple Tagging

The specification of serialisation for Simple Tagging was changed in the latest version of the Web Annotation Data Model. The simple tags are no longer accepted as string literals in the body field, but instead as string literal in the *bodyValue* field, while the extended representation is supported through the *TextualBody* class (see Figure 3). The Annotations API accepts both formats as input, but the retrieval through its endpoint always returns the expanded version.

---

[18] The current working document:
https://docs.google.com/document/d/1Yw1uJdf76v3StXST8x16TReB8FmOLw5LuWOzZz4lSiM/edit#heading=h.l2fg46yn5tej
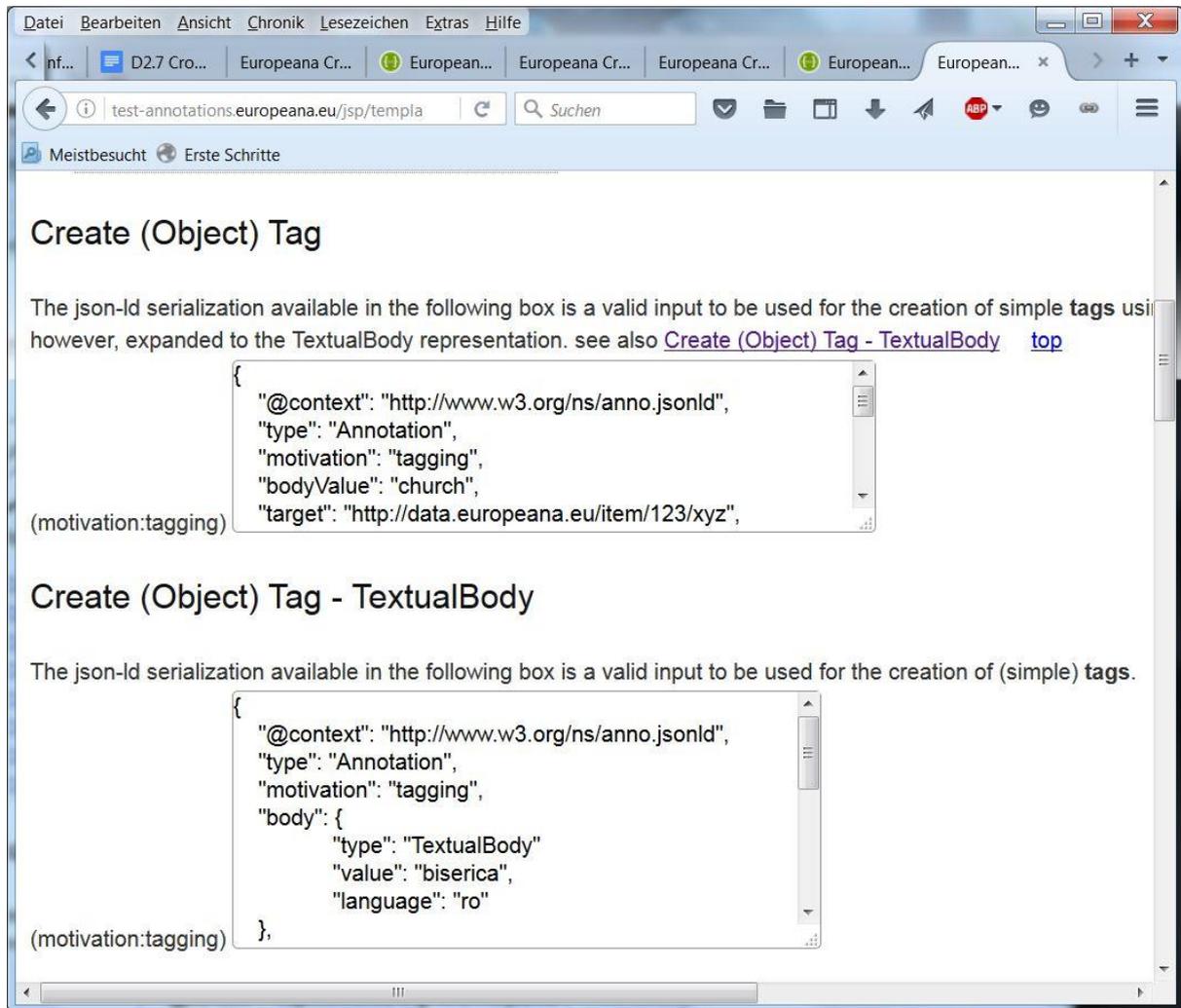[19] http://annotations.europeana.eu/docs/

Figure 3: Templates for Simple Tagging

2.2.1.3   Semantic Tagging

In Europeana Sounds, and other similar projects, Semantic Tagging is used to enrich the objects with semantic resources from Linked Open Data repositories (i.e. DBpedia[20], Wikidata[21], Geonames[22]), or with semantic classification schemes (i.e. genre concepts from the Europeana Genre Vocabulary created by WP1, as documented in D1.3, Ref 4). However, quite often individual named entities (e.g. representing authors, performing artists, painters or locations) are available in several repositories.

Given the huge size of the relevant repositories of semantic resources, and the ambiguity of the search results, it is often not easy to find the most appropriate semantic resources for annotations. In order to address this issue, the project initiated the development of the Entity API[23], which has the goal of providing a unique resource identifier (URI) in the Europeana.eu domain for the entities that are currently referenced in Europeana metadata records. By normalising the usage of named entities in the

---

metadata and annotations, a better support for semantic and multilingual search will be provided. The first, alpha release[24] of the Entity API is expected by mid-August, which will offer support for retrieval and suggestion of entities.

Depending on the level of additional information that is provided by the end users, there are three ways of representing semantic tags (i.e. as minified representations): (1) the *minimal* representation providing only the URI of the semantic resource, (2) tagging with *simple resource*, and (3) tagging with *specific resource* (see Figure 4).

Figure 4: Templates for Semantic Tagging

2.2.1.4   Geotagging

The geotagging type of annotation (see Figure 5) was implemented to support the Geolocating Radio Broadcasts Crowdsourcing Campaign (see Section 4.1 Geolocating Radio Broadcasts). With Geotagging, users can associate geo-coordinates to Europeana metadata records adding to the existing spatial metadata of the record, or suggest a more precise location for the locations that are already present in the Europeana metadata records. For example, most of the existing locations do not identify a specific location, but only indicate a town, a village or a state. This information will enable alternative explorations of Europeana metadata records, for example, through map representations or location-based search.
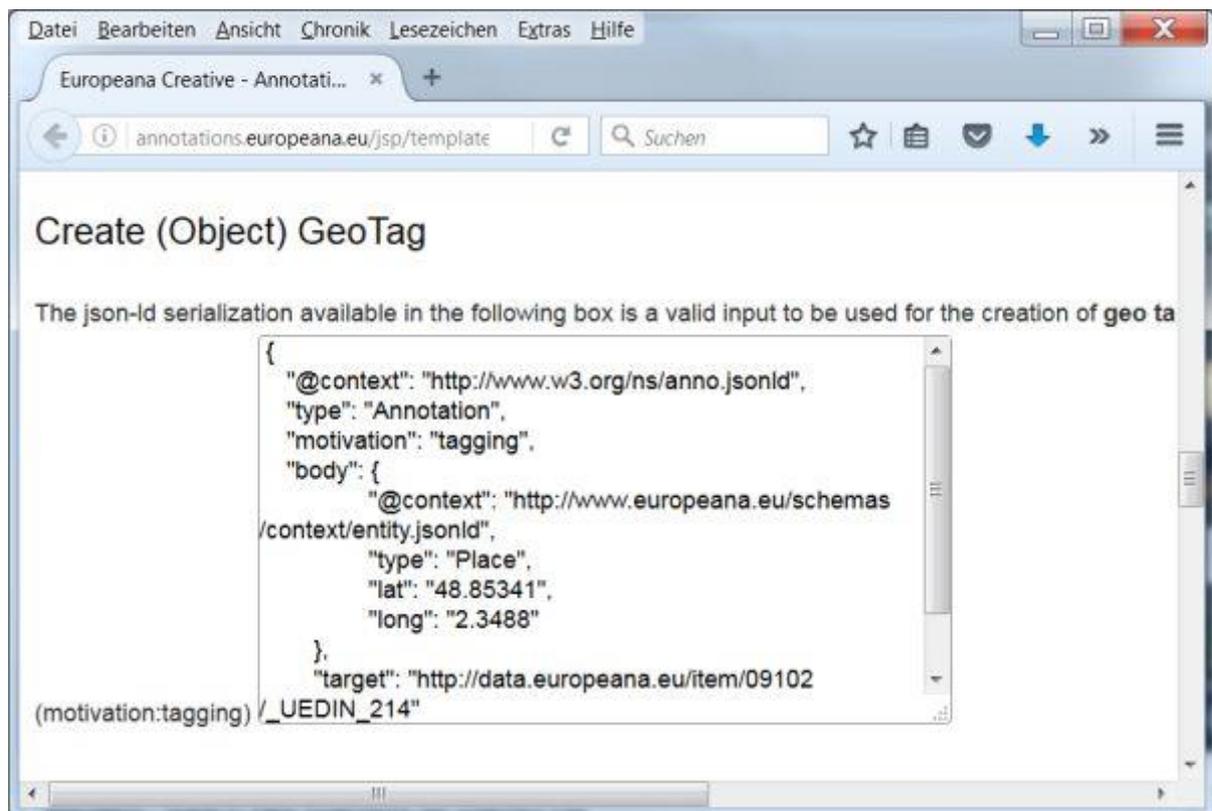

Figure 5: Template for Geotagging

2.2.1.5   Object linking

There are two types of annotations for linking Europeana metadata records (i.e. Object Linking). The first - and the one currently supported by the Web Annotation Data Model - is used to relate two or more Europeana metadata records together, without specifying the nature of the relationship. However, many types of relationships may exist between metadata records (e.g. same object, same work or original-copy). Some of these relationships are symmetric, in the sense that both resources have the same relationship to each other. Others are asymmetric, meaning that there are different relationships between two metadata records, depending on the direction (i.e. parent-child).

To be able to support an explicit type of relationship between two metadata records within an annotation, a second type of annotation for the linking of Europeana objects was defined (see Figure 6)

and implemented. These use the RDF *graph specification*[25] which makes it possible to express the specific type and direction of the relationship. The model makes no restriction on the type of relation for the link, however, it is encouraged that these correspond to the EDM properties that can be used to relate objects together in the metadata (i.e. all extensions of dc:relation available in the Europeana Data Model[26]). The object linking with a specific type of relationship is used in the Tunepal[27] application scenario, in which Europeana resources are linked with resources from TheSession.org[28] (see Section 3.1.4 Integration with TheSession.org).
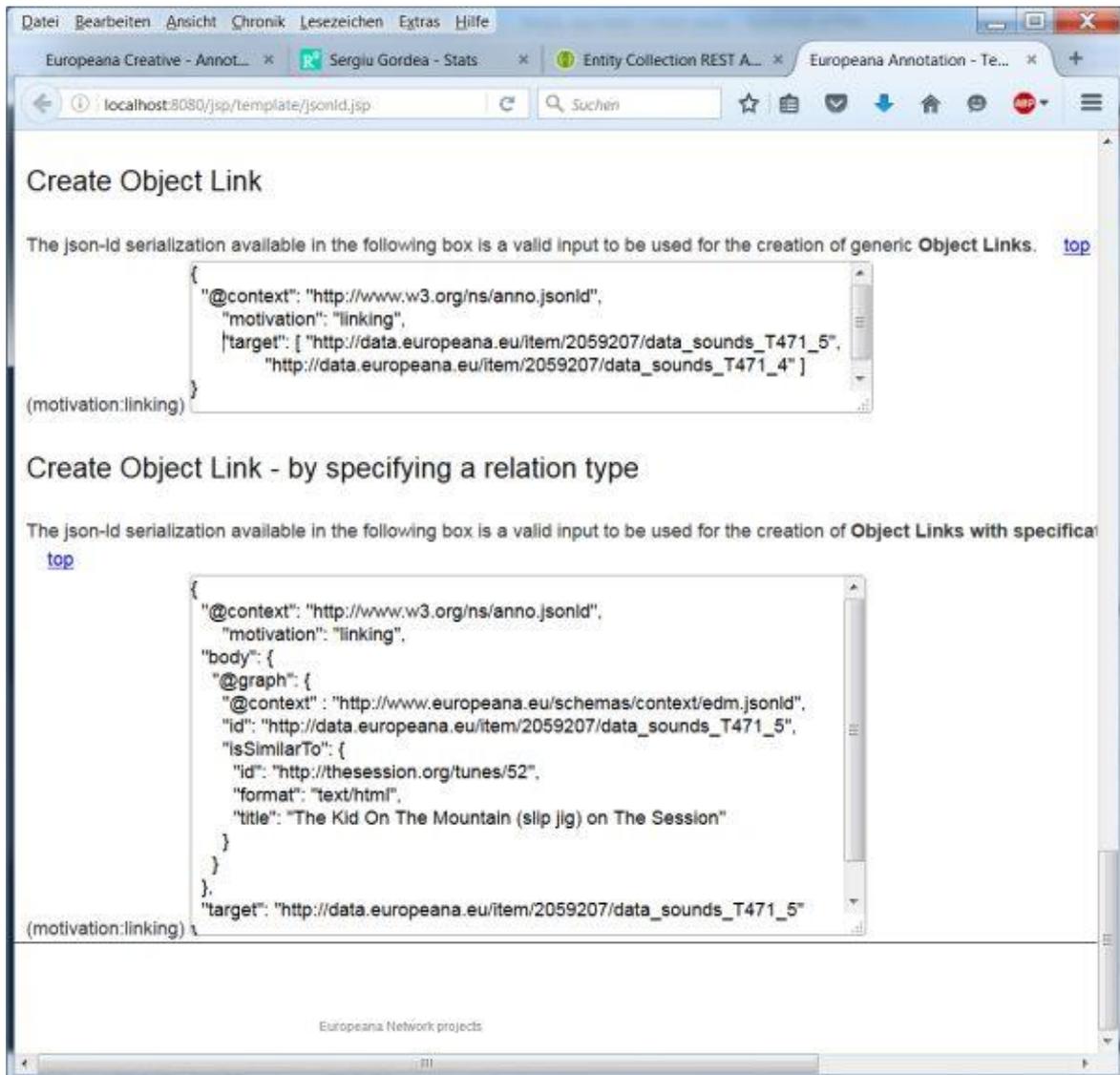

Figure 6: Templates for Object Linking

---

[25] https://www.w3.org/TR/trig/
[26]
http://pro.europeana.eu/files/Europeana_Professional/Share_your_data/Technical_requirements/EDM_Documentation//EDM_Definition_v5.2.7_042016.pdf
[27] http://tunepal.org/tunepal/index.php
[28] http://www.thesession.org

2.2.1.6   Moderation

D2.10 *Development of the Crowdsourcing Infrastructure* introduced the Moderation Principles (Ref 2) which can be summarised as the following:
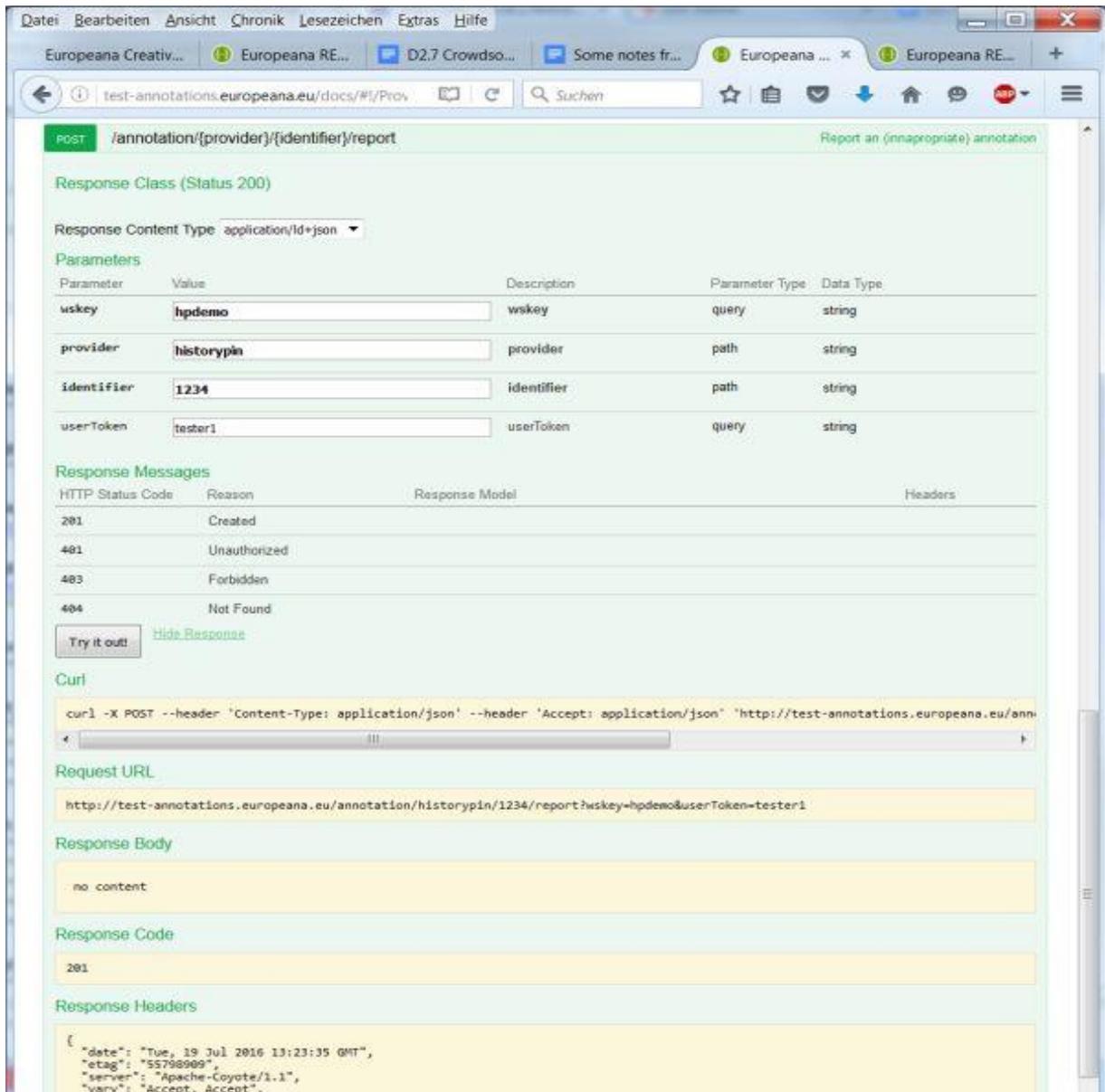
User Generated Annotations are always stored separately from the original metadata records.

1. End-users of the various crowdsourcing platforms connected to the crowdsourcing infrastructure can only create new annotations, or comment on annotations from other end-users.
2. Instead of supporting end-users with the editing or deleting of annotations by other end-users, we support the evaluation (e.g. 'flagging' and 'liking') of annotations from other end-users.
3. The target of semantic enrichments is restricted to resources from trusted repositories, in order to counter the spamming of links.
4. Utilisation of the annotations that can be retrieved from the Annotations API is left up to the policy of the respective data re-users.
5. Only administrators of the Annotations API can edit or delete annotations.

Some of the annotations are created by users, which are biased by the user level of domain knowledge or interest, and other annotations are created by software, which do not always inspire a high level of confidence. In such cases, the end users may be interested in providing their feedback on existing annotations.

The main goal of the moderation functionality (see Figure 7) - in line with the principles that have been established - is to assess the quality of user annotations, providing a mechanism for the users to indicate which are the most useful and precise annotations (for an example of an implementation, see Section 3.3.12 New Pundit Features, Step 3). In addition, the moderation functionality allows the leveraging of crowdsourcing as a confirmation mechanism for (semi-)automatically generated enrichments.

In time the amount of annotations for the Europeana metadata records will grow, which will create difficulties for their display, when embedding them in existing portals for end-users (e.g. Europeana Collections). By computing a popularity score, we provide support for the ranking of annotations based on the user feedback (through moderation). Additionally, this mechanism can be used to identify annotations which are incorrect or inappropriate, so that their owners or moderators can be notified (see Figure 8).
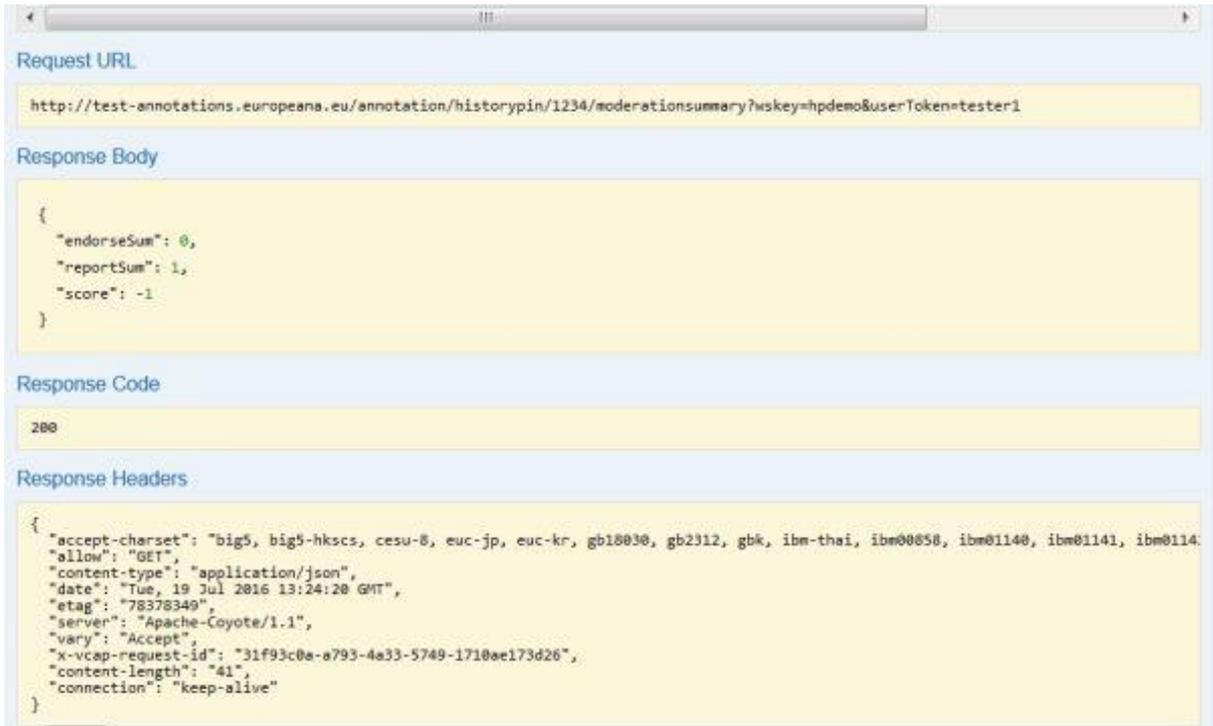
Figure 7: Providing feedback on Annotations in the application console

Figure 8: Visualization of the Moderation summary in the application console

2.2.2   Technical functionality

2.2.2.1   Administration Console

The Administration Console is integrated into the application console, but the access to the administration functionality is restricted to the users belonging to the administrators group. Additionally, the Administration Console is accessible only through the intranet in the production environment.

An update of the Swagger[29] version was required to increase the configurability of the application console, which also allowed for the separation of the Swagger GUI from the implementation. In the latest release the Annotations API uses the same code base for the GUI section of the application console as the other Europeana APIs. For the development of the next releases of the Annotations API, more effort is allocated to increase the reusability and create more artefacts that are reusable within all Europeana APIs.

An important module to be shared between various APIs provides the implementation of a common authentication and authorisation functionality. In order to achieve this goal, we chose to implement the OAuth2[30] protocol. In the current version, we adapted the implementation of the authorisation mechanism to comply with the spring definition of OAuth2 interfaces[31].

---

[29] http://swagger.io/
[30] http://oauth.net/2/
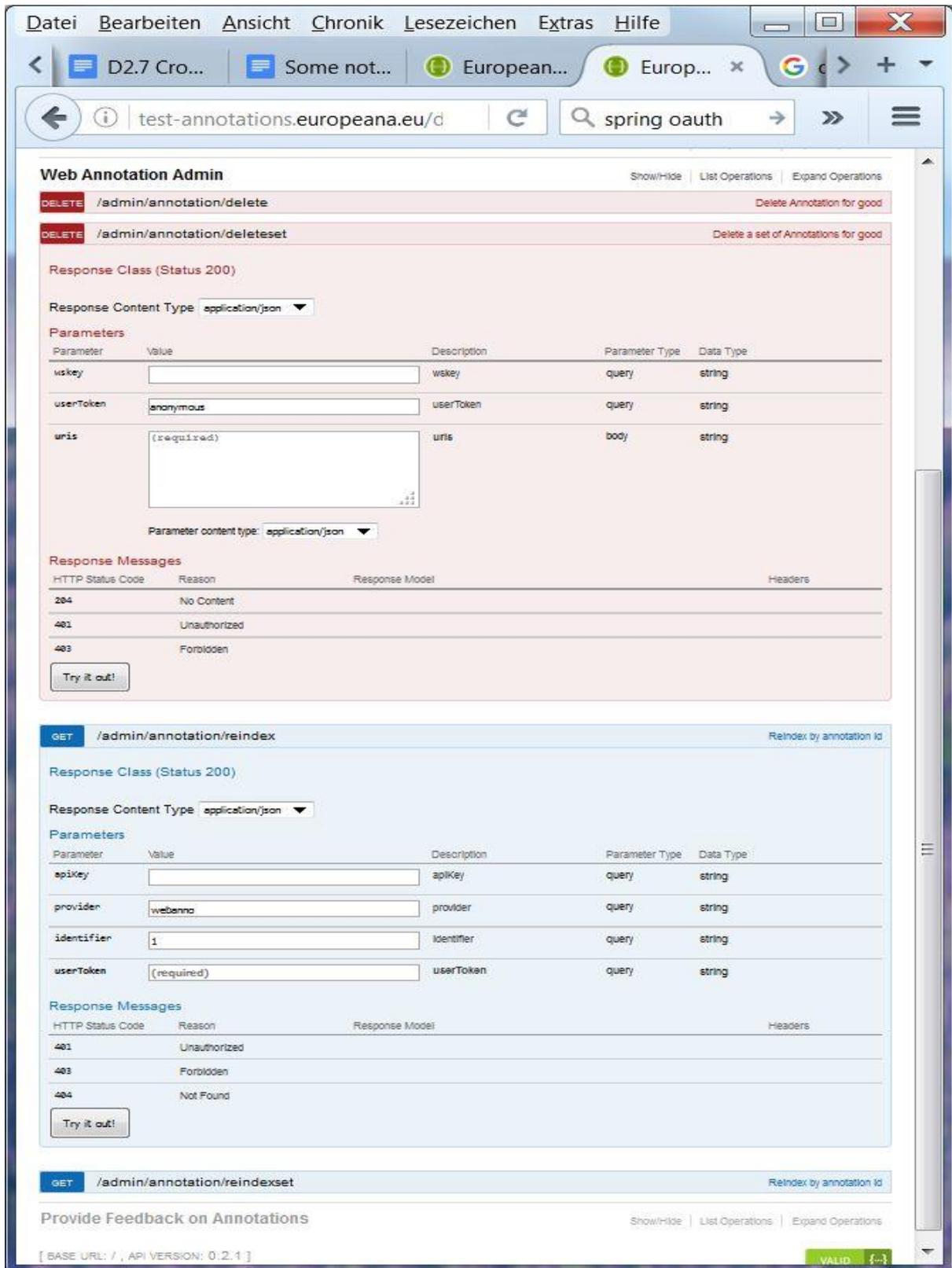[31] http://projects.spring.io/spring-security-oauth/

The whitelist management functionality is used to specify the list of external resources that are permitted to be used in the annotations. This was already introduced within D2.10 (Ref 2) and did not suffer significant changes. The newly added functionality is related to the deletion and reindexing of annotations. The regular users are provided with the delete functionality, which disables the annotations in the Annotations API, and is the equivalent to the standard "*move to recycle bin*" functionality provided by operating systems, or popular document management systems. In addition to this, the administrators are provided with the authorisation to completely remove the annotations from the database. There are two operations that support this functionality: one is used for removing individual annotations basing on their identifiers. The second is used to remove a set of annotations, identified by their URIs as demonstrated in Figure 9.

The creation of annotations is considered a success when the objects are stored in the database of the Annotations API. For search purposes the annotations are also processed and written in a Solr[32] based, text index. The indexing of annotations is typically invoked during creation. However, if the indexing fails, the annotations are still accessible through the generated URI. Additionally, the API clients can disable the indexing during the creation of annotations. This functionality is typically used when transferring a large number of annotations from external application servers (e.g. Pundit, HistoryPin, WITH, etcetera), or when generating annotations automatically using enrichment tools (e.g. automatic enrichments using the Musical Instrument Museums Online (MIMO) Vocabulary and Thesaurus and Cultuurlink[33]). In such cases, it is important to ensure high throughput for the storage of annotations and it is advised to disable indexing on creation. The administration console allows administrators to explicitly invoke the (re-)indexing of annotations. Administrators have the option to reindex individual items or a larger dataset composed of a list of annotations identified by their URIs.

---

[32] http://lucene.apache.org/solr/
[33] http://cultuurlink.beeldengeluid.nl/

Figure 9: Administration Console

2.2.2.2 Build and deployment environment
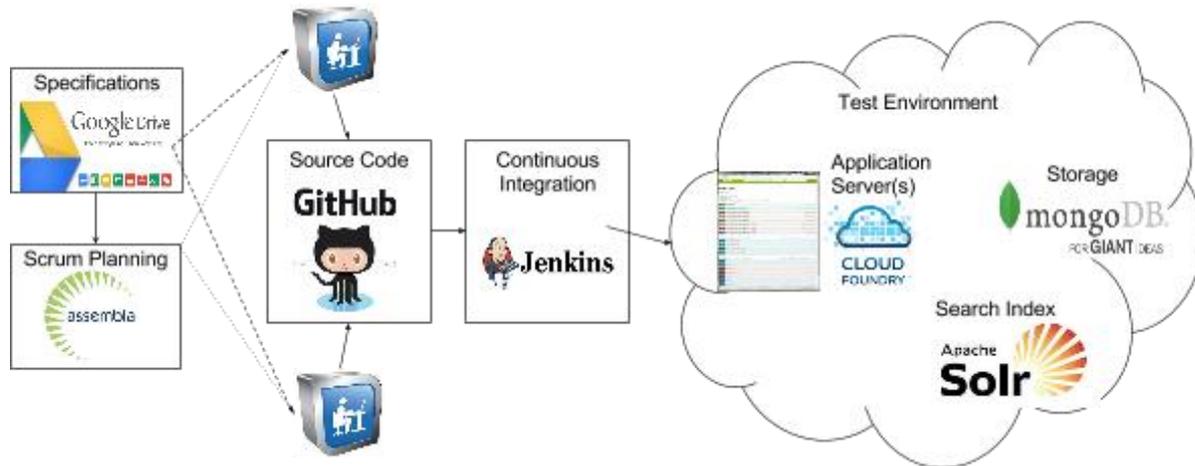
Figure 10: Development tools and test environment

The building and deploying of the Annotations API has used the standard tools and infrastructure of Europeana (see Figure 10), following the Europeana agile development practices[34]. The build configuration and deployment is managed using the Jenkins[35] build tool. Starting with the first public release of the Annotations API, the source code management in GitHub[36] uses a master branch where the stable source code for releases is stored and a development branch used for continuous integration and testing purposes (see Figure 11). Consequently there are two different build job configured which are automatically deploying releases into the production environment and into the test environment, respectively.

The compilation of the source code is performed by using the maven plugins using the appropriate java version for the target environment. The packaging subtask is configured to include the application configurations files (i.e. including database and solr connectivity configs, logging configurations, OAuth configs, swagger configs, etc) and the Swagger UI[37].

---

[34] https://app.assembla.com/spaces/europeana-npc/wiki/Our_Scrum_principles
[35] https://jenkins.io/
[36] https://github.com/europeana/annotation/branches
[37] http://swagger.io/swagger-ui/
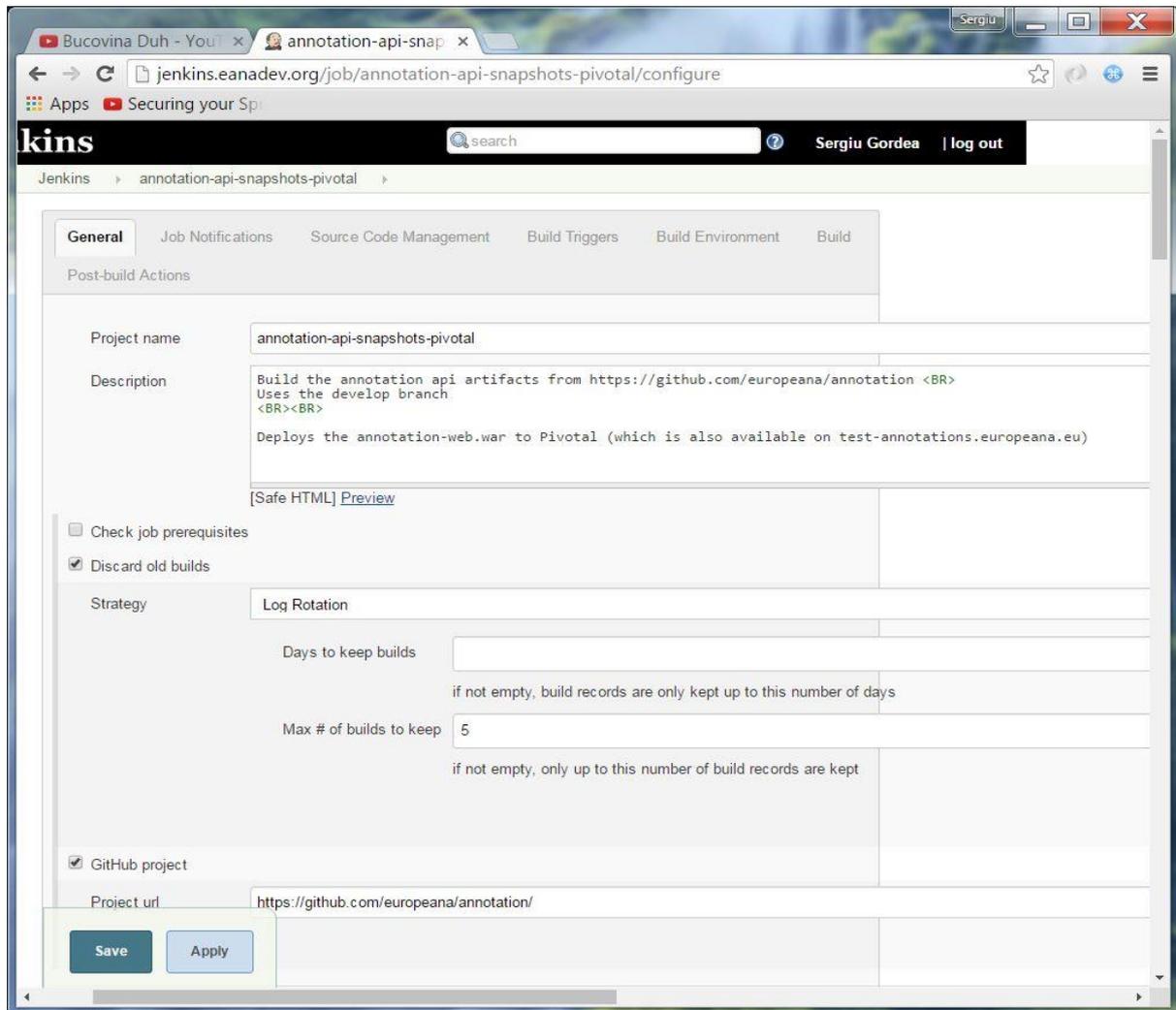
Figure 11: Build job configuration in Jenkins

In the final step, the application is remotely deployed to the cloud environment where it is hosted, by using the Cloud Foundry client tool[38]. The required application server (i.e. Tomcat[39]) is installed in a fresh virtual machine on which the packaged application is uploaded and physically deployed. If the application is successfully started, the routing of web requests using the Europeana subdomains is installed and the service is ready to be used. The latest release of Annotations API is available in the production environment (see Figure 12) which is accessible through the URL http://annotations.europeana.eu/docs/, while the test version is accessible through the URL http://test-annotations.europeana.eu/docs/.

---

[38] https://docs.cloudfoundry.org/cf-cli/
[39] http://tomcat.apache.org/

Figure 12: The overview of Annotations API functionality in the Application Console

3   Europeana Sounds crowdsourcing applications

This section describes the progress with the development of the end-user facing crowdsourcing applications that have been connected to the crowdsourcing infrastructure to support enrichment through crowdsourcing. The table below (Table 1) summarises some of the main characteristics of the Europeana Sounds crowdsourcing applications:

Table 1: Main characteristics of the Europeana Sounds crowdsourcing applications

Application	Technical partner	Type of annotation	Data requirements	Target audience	Status
Tunepal	Historypin	Object Linking	Limited to traditional music content Requires a direct link to the audio file	Culture vultures	In development
Historypin	Historypin	Geotagging	Limited to field recordings or audio recordings that have a geographical link Content needs to be hosted on SoundCloud or audioBoom	Culture Vultures	In production
Pundit	NET7	Simple Tagging, Semantic Tagging, Moderation	Metadata record needs to include the Europeana Resource ID	Culture Snackers and Culture Vultures	In production
WITH	NTUA	Simple Tagging (currently) To be implemented: Semantic Tagging and Moderation	Metadata needs to be harvested and published by Europeana	Culture Snackers	In production
Music Collection	EF	None (currently) To be implemented: Simple Tagging, Semantic Tagging and Object Linking	Limited to music content Requires a direct link to the audio file	Culture Vultures	In development

3.1 Tunepal widget

This section describes the advancement of the *Crowdsourcing through specialised platforms scenario*, as introduced in D2.2 (Ref 3) D2.4 (Ref 1) and D2.10 (Ref 2), which focusses on the challenge of music identification. Historypin has proposed a traditional music pilot that brings archival sounds holdings to knowledgeable users that might not be aware of these collections by themselves.

3.1.1 Identifying tunes

As part of the work for D2.4 (Ref 1), Historypin undertook user consultation in the form of a survey to answer questions about how traditional musicians hear new tunes, how they identify and find them, and how they subsequently learn these new tunes.

The result of this survey was an insight in the workflow that traditional musicians follow when they are identifying and learning new tunes, which are detailed below. To identify and learn a tune, the musician would:

1. Record the tune at a session or a feis (traditional music festival)
2. Upload it into Tunepal to find a name
3. Use the name to search for recordings of that tune on Google, Youtube or TheSession.org
4. Pick a recording they like best and learn that by ear

The traditional music pilot that Historypin has developed combines steps 2 and 3 in the above workflow and eliminates the need for musicians to first find the name of a tune and then use different sources to find high-quality archival recordings of it.

3.1.2 Approach

As part of our work on Europeana Sounds, and to provide real connections between the traditional music scene and music archives, Historypin is working together with the developer of Tunepal[40] to improve the Tunepal app in such a way that it becomes a website-based app, into which you can play a traditional Irish or Scottish tune and which will then surface archival recordings of that tune, pulled from the Europeana database. These tunes are supplied to Europeana Sounds by data providers Tobar an Dualchais (TAD) and Comhaltas Ceoltóirí Éireann (CCÉ). The platform is extensible to any Europeana-provided datasets, though at the moment only these two providers offer data at a suitable level of access and description, so the Europeana search query is tailored to just these providers.

Between March and August 2015, Historypin worked closely together with Professor Bryan Duggan, the creator of Tunepal, to link the archival recordings to Tunepal through the Europeana API. These archival recordings were linked by using the existing Tunepal transcription functionality to first transcribe the music and then to identify candidate track names against a corpus of user-contributed tune titles at TheSession.org. Using these track names as a pivot, the Europeana Search API was used to query the provided datasets. In addition, Tunepal has been updated to work in a web browser rather than in a

---

[40] http://www.tunepal.org

mobile app, and a bespoke interface to display the archival recordings next to the tune names has been created.

3.1.3 Description of the applications

This Tunepal app has taken the form of a HTML/CSS/JavaScript widget that:
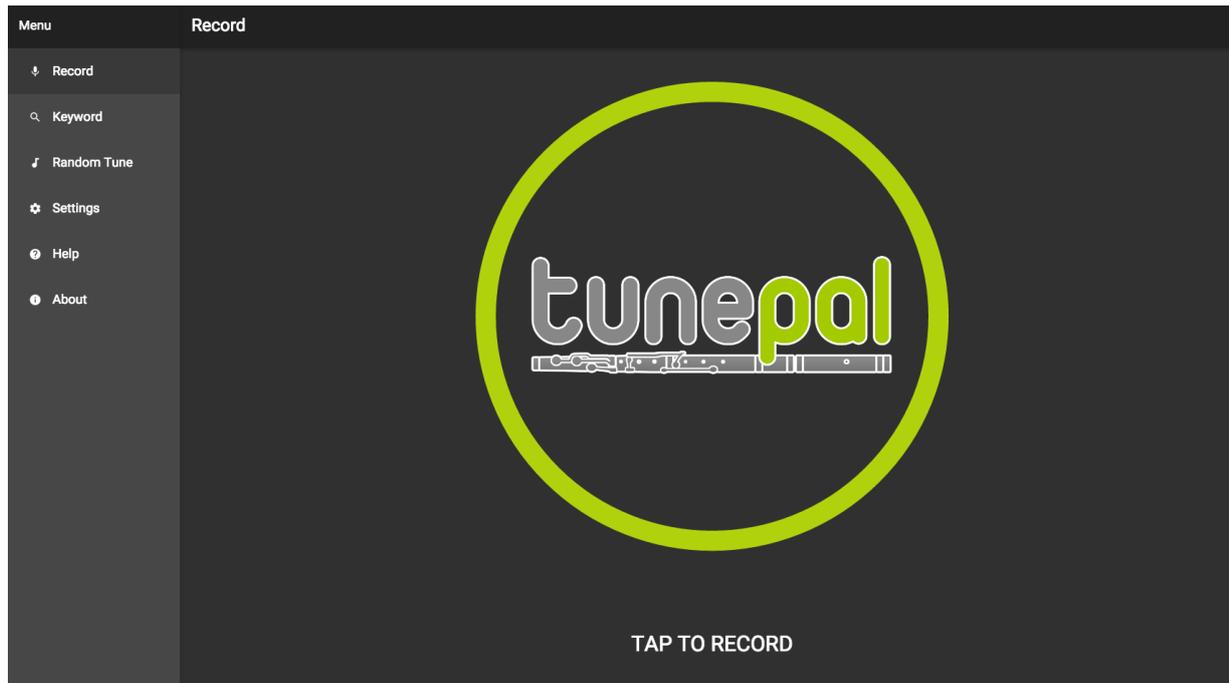
Figure 13: Tunepal homescreen

1. Allows the user to record 12 second pieces of music:

2. Connects to Tunepal query-by-playing search engine to get some suggested high-probability tune names:
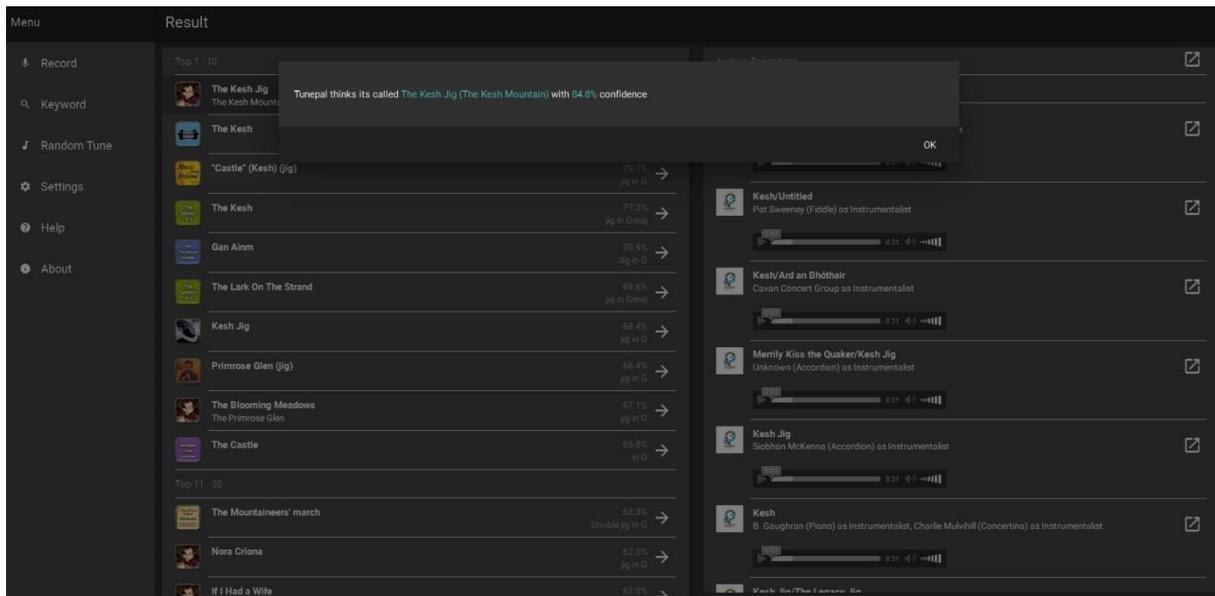
Figure 14: Tunepal query-by-playing result

3. Queries the Europeana Search API to get back archive recordings that are likely matches:
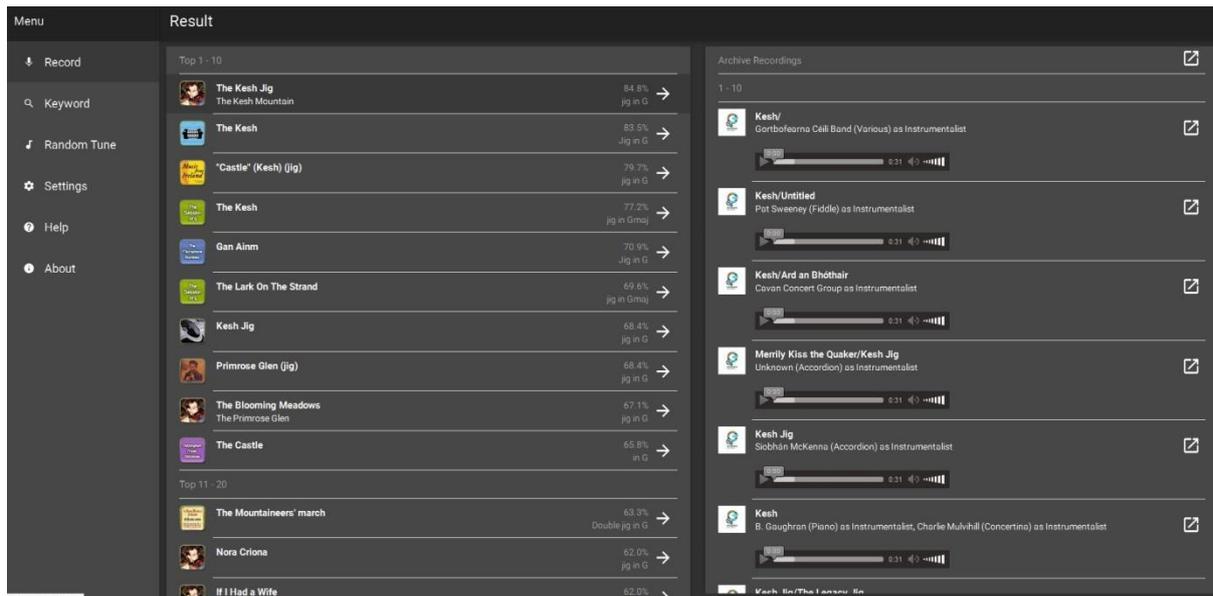
Figure 15: Tunepal query-by-playing results

4. Lets the user listen to some of those audio candidates from Europeana:
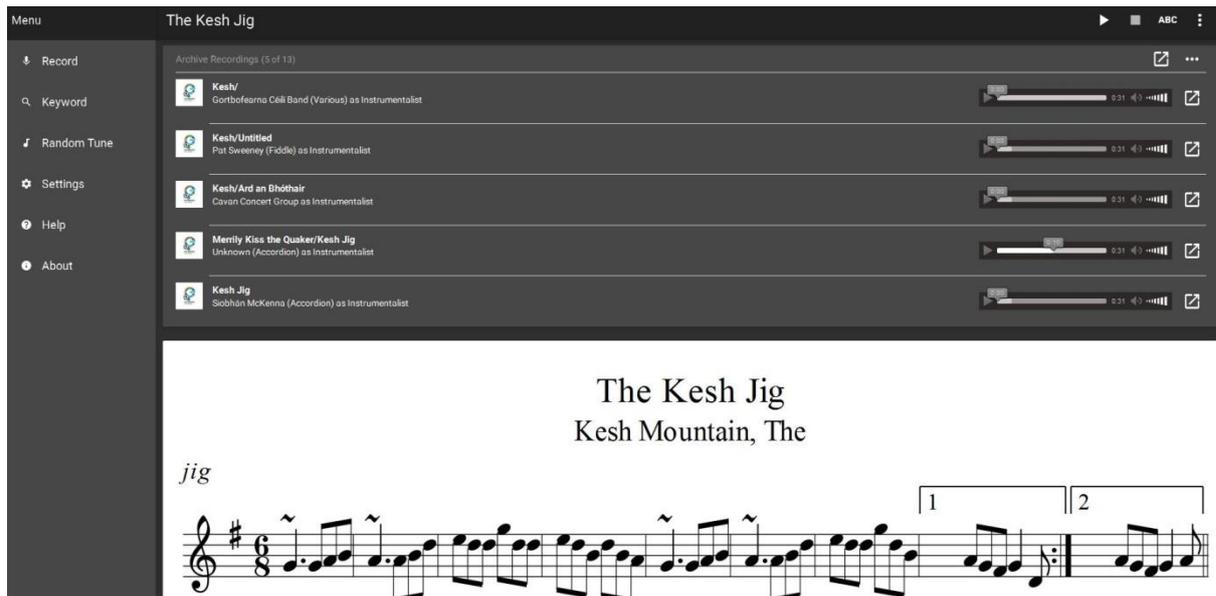
Figure 16: Tunepal individual tune result with archive recordings and staff notation
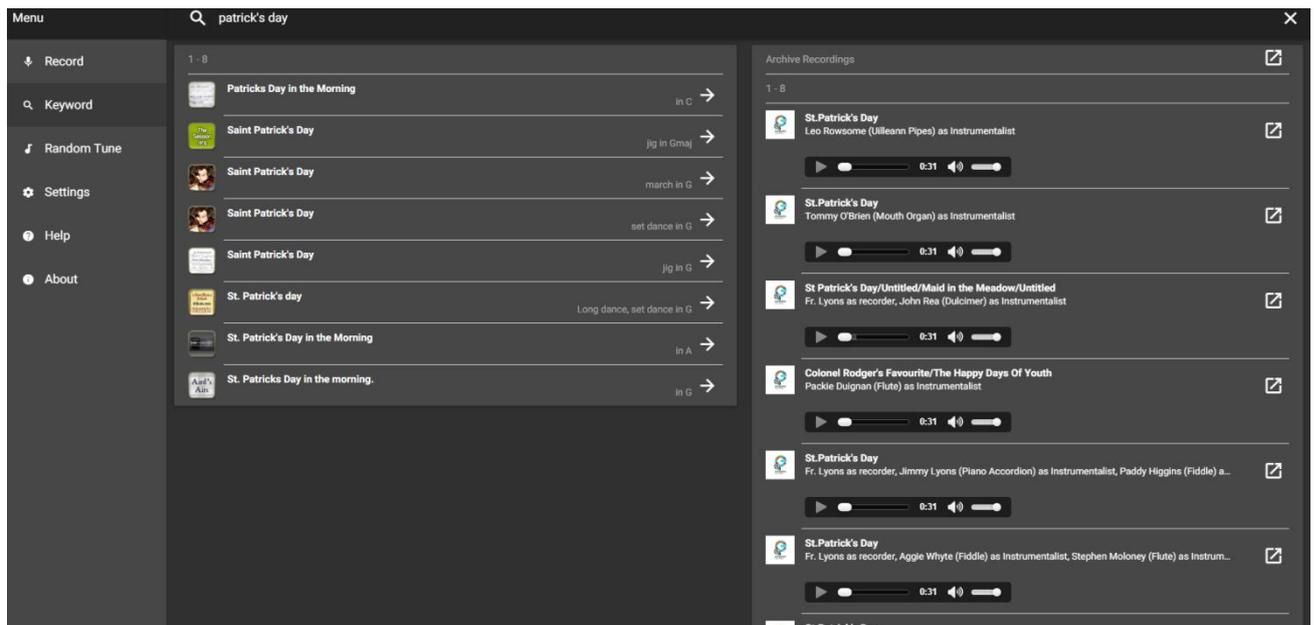
Figure 17: Tunepal search result

5. Records (as an Object Link, see Section 2.2.1.5 Object Linking) the link between input and matched recording, if the user indicates that there is a match.

The latest version of Tunepal - which incorporates steps 1-4 - is available here: www.tunepal.org

Every user that uses Tunepal to search for a tune either through playing or by keyword will be shown Europeana recordings for this tune if they are available. Tunepal receives around 73,000 unique visitors a year and roughly 40% of these (or 29,200 visitors) will have seen and interacted with the Europeana recordings, as around 40% of the tunes available in Tunepal have recordings associated with them.

3.1.4   Integration with TheSession.org

TheSession.org is a forum and online community for traditional musicians and the ideal place to deploy Tunepal and reach a large audience of users interested in tune identification.
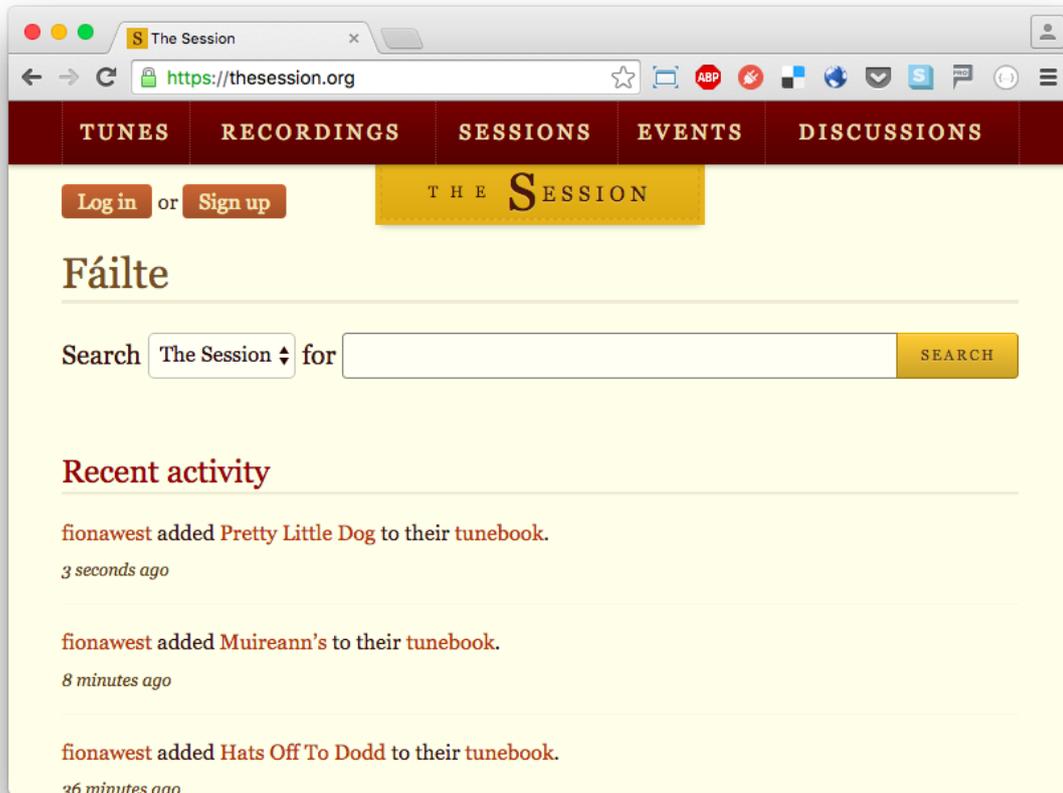
Figure 18: Homepage of TheSession.org traditional music forum

To facilitate the integration of Tunepal on the TheSession.org website, Historypin has set up meetings with the developers of Tunepal and TheSession.org to design and plan the required development. At this stage, the developer from TheSession.org is working on embedding Tunepal into the website, so people can use it and pull up the archival recordings within the TheSession.org website. It will also be possible to create annotations that can be pushed back to Europeana based on the usage of Tunepal in TheSession.org. These annotations will take the shape of geotags (see Section 2.2.1.4 Geotagging) and will showcase where an archive recording has been listened to, based on the location information embedded in the user's browser.

Unfortunately, development has been delayed by the lack of full-length archival traditional tunes for playback. Many of the tunes provided by the traditional music data providers in the project have been restricted to a 30-second preview. This is especially problematic in the case of traditional music, where tunes are played in sets of three. This means 30-second previews in most cases likely will not represent all three works represented in the archival recording (only the first one). The tune that Tunepal suggests as a candidate might however be the last tune in the set and therefore not available to the user (so the link cannot be validated through crowdsourcing).

For the integration both the Tunepal and TheSession.org developer are required to make high-impact changes to their technical environments. They hence need to be convinced of the utility of their effort. Since successful algorithmic tune identification relies on full length tracks, this issue needs to be resolved, before the respective developers can start changing their technical environments. This situation has made it difficult to adhere to the roadmap that has been laid out in previous deliverables and to proceed with the evaluation of the Tunepal / TheSession.org integration.

Recently data provider Comhaltas has started providing access to full length tracks on their own website[41]. However, these full length tracks are not accessible through Europeana (Sounds) yet, since they require an update of the aggregated metadata records. WP2 will collaborate with WP1 to ensure this dataset will be updated swiftly and will then get back to the developers of Tunepal and TheSession.org, through Historypin.

3.1.5   Next steps

Based on the user research - as reported in D2.10 (Ref 2) - which confirmed that traditional musicians will use Tunepal as a way of identifying tunes as well as discovering archival material, Historypin can list next steps, which connect to the roadmap outlined above.

1. Improving Tunepal based on user feedback, especially the accuracy of the music recognition and the usability of the interface.
2. Historypin will do more testing of the Tunepal app, this time with adult musicians, focusing mostly on the usability of the interface and to try and get clearer answers to what value the Europeana search functionality adds to the app, as well as how the app adds value to Europeana Sounds.
3. Historypin will deploy the next iteration of Tunepal on www.thesession.org, a forum for traditional musicians. It can be integrated as a tune recognition tool there. TheSession.org users will be asked to complete a short online questionnaire, after they use the tool, to allow us to gather more data about accuracy, availability of archival records and how the tool is perceived by our target audience.

3.2   Historypin geotagging interface

As a parallel track of work alongside the Tunepal widget development, Historypin has created a geotagging interface to enrich Europeana Sounds collections on the Historypin.org[42] website.

This new interface allows logged-in Historypin users to suggest more accurate or more precise locations for pins. If this enrichment is applied to objects with a Europeana ID, these annotations are pulled into the Annotations API as a Geotag (see Section 2.2.1.4 Geotagging). This synchronisation with the Annotations API happens courtesy of the connection that has been established between Historypin and

---

[41] https://archive.comhaltas.ie/
[42] http://www.historypin.org

the Europeana APIs, as part of the technical demonstrator app (developed as part of the Europeana Food and Drink project[43]).
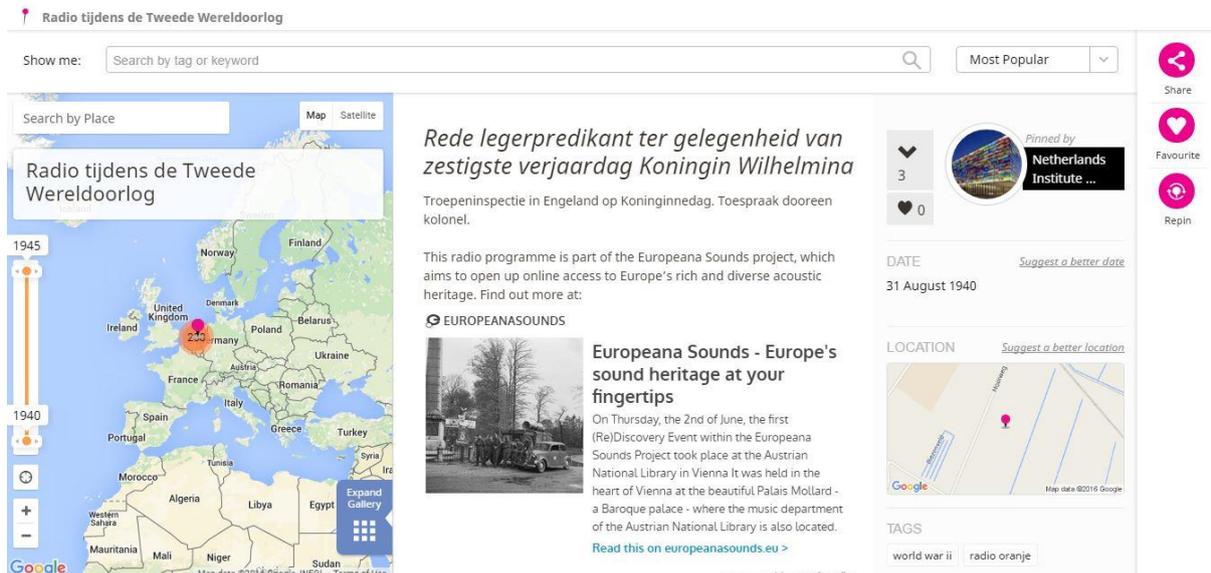
Figure 19: A screenshot that shows the location of the "Suggest a better location" link above the small map on the right

---

[43] http://foodanddrinkeurope.eu/
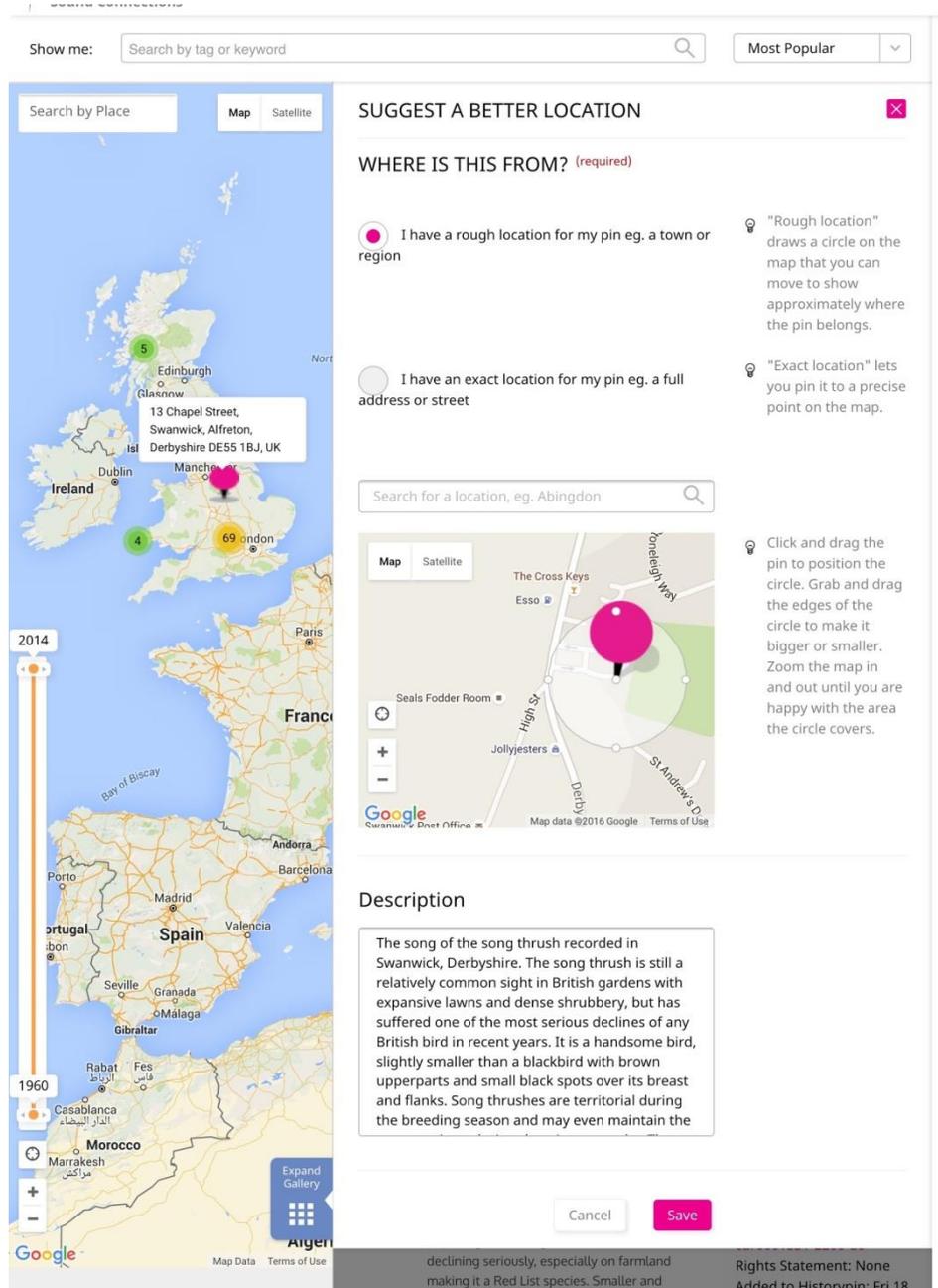
Figure 20: The "Suggest a better location' panel that opens when the user clicks the link on a pin

Europeana will ingest the annotations provided by Historypin as part of a specialisation of the Semantic Tagging scenario that has already been implemented for Geotagging (see Section 2.2.1.4 Geotagging). The specialisation required for this application scenario provides the possibility of submitting a 'body' with the geographic data, even if only the information about latitude and longitude coordinates has been provided, without a place name (label or controlled vocabulary term).

The first planned test of this Historypin geotagging interface took place in April 2016 in conjunction with the preparations for the Geolocating Radio Broadcasts Crowdsourcing Campaign (see Section 4.1 Geolocating Radio Broadcasts).

3.3   Pundit integration in the WITH platform

3.3.1   Introduction

After analysing the results obtained from the first crowdsourcing pilots - as reported in D2.4 (Ref 1) - WP2 decided to develop the integration between Pundit[44] (NET7) and the WITH platform (NTUA) in order to (1) allow the simplification of the public dissemination of resources suitable for enrichment through crowdsourcing, by using a centralised collection platform (WITH); and (2) improve the annotation process, by providing an end-to-end workflow.

NET7 and NTUA managed the integration of Pundit and WITH in two successive steps, to first get a set simple and working solution (first step), and then work on a second solution (second step), which was more complicated to produce, but more convenient for the end-user:

- First step: integration of Pundit - as a widget - in the WITH platform. Implemented in April 2016

- Second step: direct integration of Pundit inside the WITH platform. Implemented in May 2016

In the first case a button (widget) was added to the item pages on WITH, which opened a new browser window with an instance of Pundit. This step was used as a proof-of-concept, to verify that all the enrichments were successfully saved the Pundit Server. In the second step the Pundit code was inserted directly into the items pages on WITH, simplifying the user experience.

The integration of Pundit and WITH has allowed the initial crowdsourcing campaign to run (June 2016), in which users had the ability to manually create Semantic Tags (see Section 2.2.1.3 Semantic Tagging). Through a revised and simplified interface, which will be described in detail below, the user could associate a metadata record aggregated by Europeana with a concept from a controlled vocabulary (in this case concepts from the MIMO Vocabulary and Thesaurus[45]).

Users were also given the opportunity to interact with existing annotations created by other users, or semi-automatically (see Section 2.2.1.6 Moderation): users can indicate reactions, such as"Like," "Dislike", "Report" or "Endorse" or they can add comments.

The integrated platform used to carry out the crowdsourcing campaign is described later in this document and the results obtained demonstrate that it is the favoured approach (see Section 4.2 Enriching Music Collections with Instruments, and more specifically Section 4.2.2 Results).

3.3.2   New Pundit features

In order to obtain the Pundit and WITH integration the following activities were carried out:

- The Pundit client was modified in order to work in a single page website context like the WITH platform is.

---

[44] http://thepund.it
[45] http://www.mimo-international.com/vocabulary.html

- A custom CSS file was developed for the Pundit client, in order to solve some conflicts that arose between Pundit and WITH.

- The user interface for Pundit was revised.

- A JavaScript callback function was created, synced with the event of saving annotations in the Pundit client. This is used by WITH in order to connect the annotation to the WITH user (this is - among other things - used to assign a score to the user based on the number of the annotation performed by the user).

- The MIMO Connector in the Pundit resource panel - which retrieves the concepts from the MIMO vocabulary and thesaurus via auto-completion - was modified in order to sort the results when the user puts the search string in the form. Now the instruments that are closer to the root of the MIMO tree appear at the top (concepts are sorted from broad terms to narrower terms).

- We have developed the Pundit Annotator Pro for the Europeana Chrome Browser Extension[46] which allows users to instantiate the Pundit client from the item page of a metadata record on the Europeana Portal or Collections (including, but not limited to the Music Collection).

3.3.3   Pundit client interface improvements

Some modifications were made on the Pundit client user interface in order to improve the user experience. NET7 created simple tutorial videos divided into 6 steps to explain the workflow of the annotation process to the end-user:

1. Find your item on the WITH platform[47]
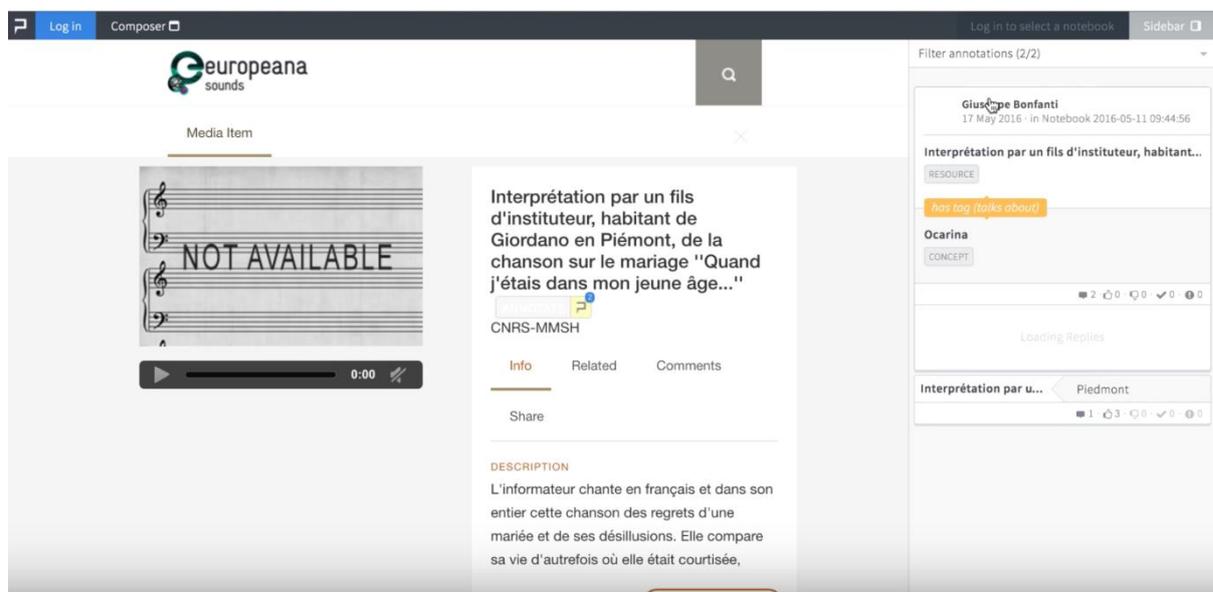
Figure 21: Item page on WITH

---

[46] http://thepund.it/annotatorEuropeana
[47] https://youtu.be/5gX8P6QlNjc

The user types the URL in the navigation bar to access the WITH platform. Then clicks on the item that they want to annotate. Next to the title of the item, the user can then click on the "Launch Pundit" button (see Figure 21). Previously created annotations will appear in the ''Sidebar'' on the right side of the screen.

2. Login to Pundit with Google or Facebook[48]
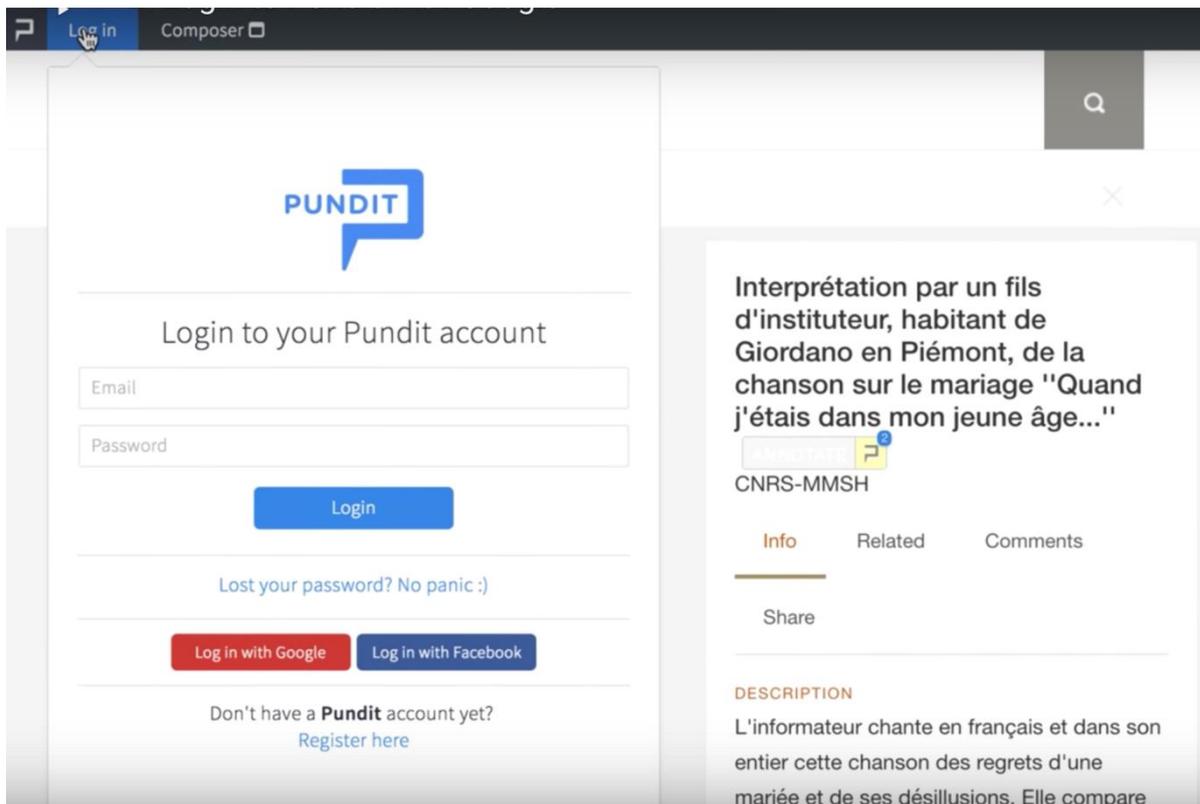
Figure 22: Login screen of Pundit

To create an annotation the user needs to log into Pundit by clicking on the "Login" button (see Figure 22). The login window will appear and then the user can use a social media account (Google or Facebook) or login with an email address to create a Pundit account.

---

[48] https://youtu.be/Qw1gN5m9Dmo

3. Use social![49]
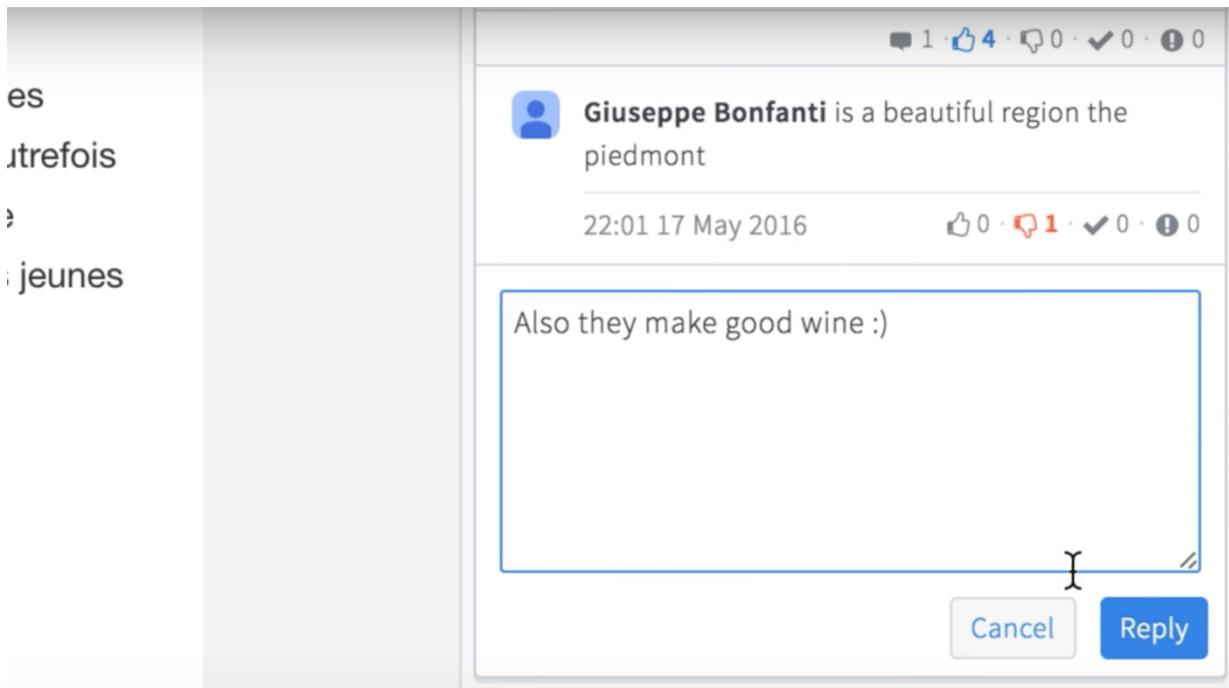
Figure 23: Example of comments and (dis)likes in the Pundit interface

Users can like, dislike or comment on annotations via the Pundit client interface (see Figure 23). This is an implementation of the Moderation functionality of the Annotations API (see Section 2.2.1.6 Moderation).

4. Create your annotation: connect the item to a musical instrument[50]

Figure 24: Drop-down list with all available options for enrichment

---

Users can create an annotation using the musical instrument concepts available in the MIMO Vocabulary and Thesaurus. The user clicks "Annotate", then selects ''Instrument Tag'' (see Figure 24), and types in the name of the relevant instrument (see Figure 25) in order to create the Semantic Tag (see Section 2.2.1.3 Semantic Tagging).
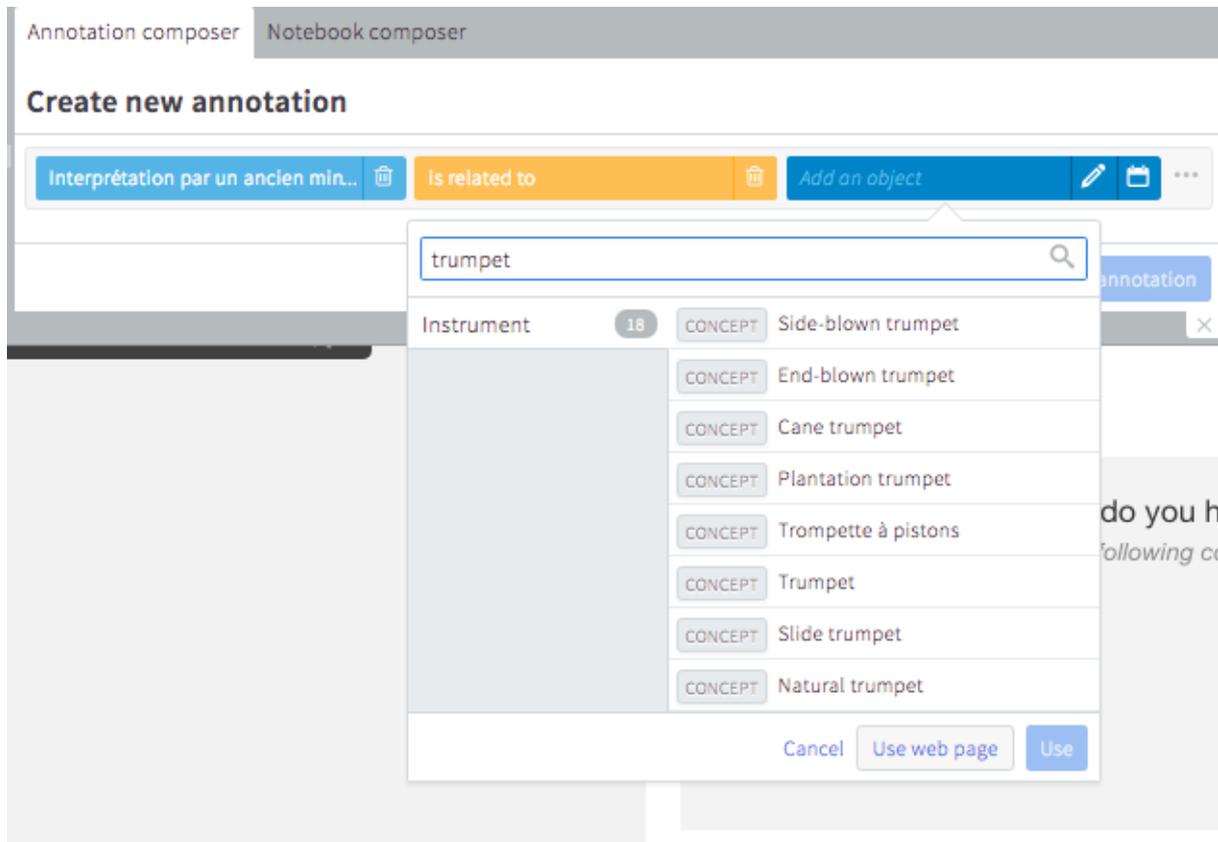
Figure 25: Drop-down list with instrument terminology fetched from the MIMO Vocabulary and Thesaurus

5. Install the Pundit Annotator Pro for Europeana Chrome browser extension[51]
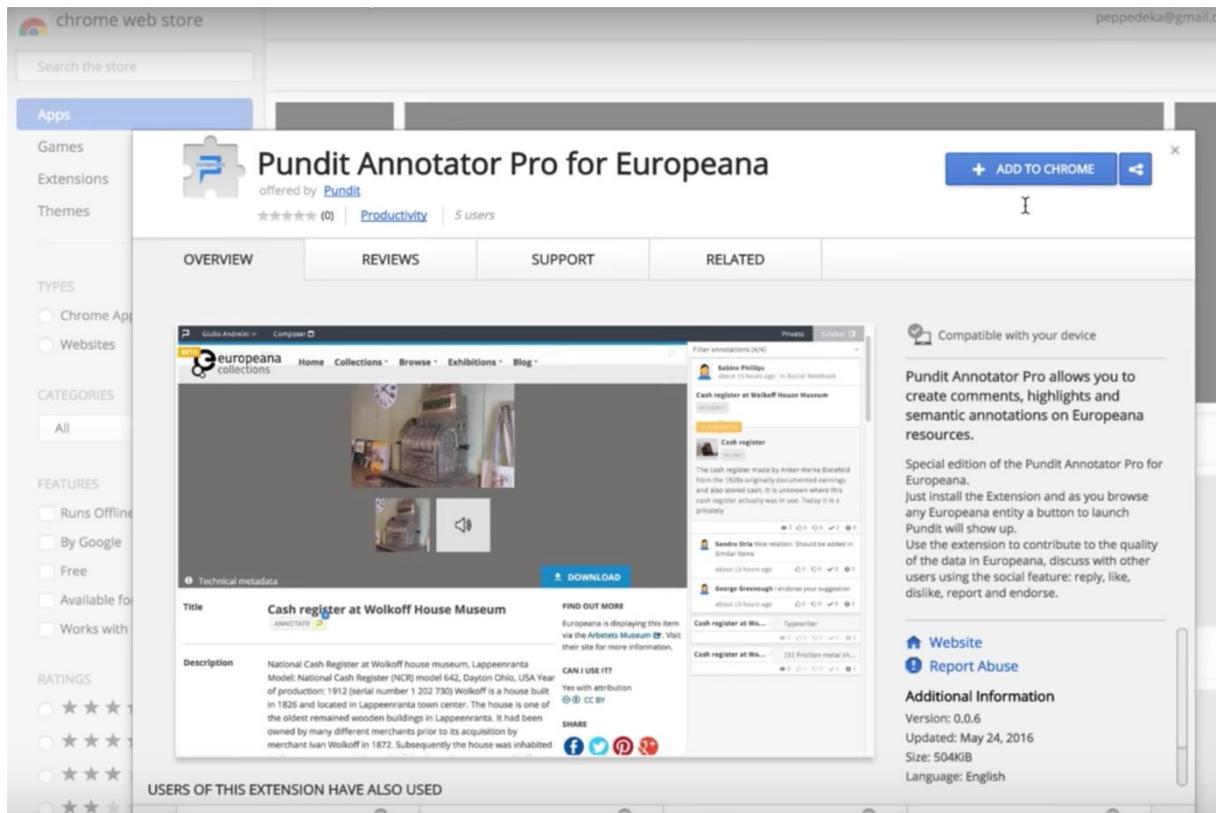
Figure 26: ''Pundit Annotator Pro for Europeana'' Chrome Browser Extension

The user is able to install the "Pundit Annotator Pro for Europeana" Chrome Browser Extension by navigating to: http://thepund.it/annotatorEuropeana (see Figure 26, only for Google Chrome users)

6. All platforms are linked to Europeana[52]
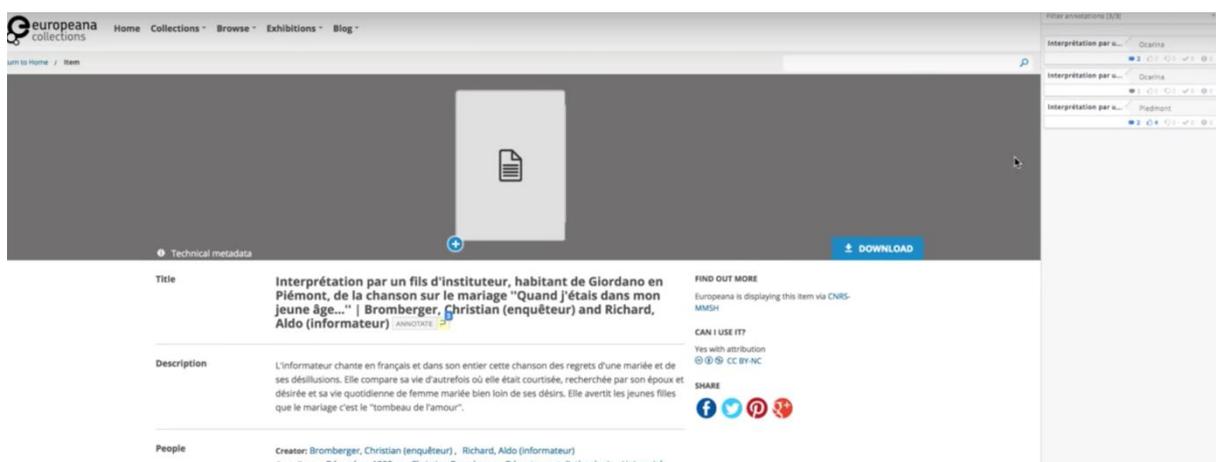
Figure 27: Item page on the Europeana Collection

---

[51] https://youtu.be/hrvycYUj6SM
[52] https://youtu.be/fN4U0bjdFVM

All platforms are linked to Europeana, through the same Resource ID. This allows the Pundit Annotator Pro for Europeana Chrome Browser Extension to show all annotations connected to the Europeana metadata records, regardless of the context or platform where the Pundit is represented.

3.3.4   The WITH platform

The WITH platform[53] - as introduced in D2.4 (Ref 1) - is available for cultural institutions and organisations, professional users and third party developers. It allows them to easily search for the cultural resources that meet their retrieval criteria so as to collect, use and reuse them to promote innovation and demonstrate the social and economic value of cultural content. This is achieved through the delivery of APIs that facilitate the development of applications based on cultural content.

WITH is a platform for storing, accessing and processing content and metadata. It is designed and developed in alignment with complementary services used and produced in the Europeana ecosystem.

These are the high level functionalities of WITH:

- Aggregate multiple sources of cultural heritage content

- Create and curate collections of digital resources

- Add your own metadata and content to the search base

- Maintain interoperability with data models and standards using the services of the MINT tool

- Store metadata in several formats and serialisations; support for widely used domain models

- Serve collections as specific back ends for specialized front-end applications

When reaching out to potential user groups and stakeholders, the methods of engagement can be broken down to:

- Discover (leveraging APIs to cultural content from around the world)

- Aggregate (upload metadata, map, transform, publish to external repositories)

- Create (collections, exhibitions and stories)

- Annotate (link and enrich using SKOS thesauri and linked data repositories)

- Participate (share, tag and rate collections, follow users & spaces, join user groups)

- Build (use the API to access data and services)

The workflow for setting up the crowdsourcing campaign and for end-users to create annotations in the crowdsourcing platform is the following:

---

[53] http://with.image.ntua.gr

Step 1 (for data providers): upload the collections that have been selected by the data providers and are suitable for enrichment through crowdsourcing in order to import them to WITH using the Europeana import functionality. Data providers input the Dataset ID from Europeana to import the dataset to WITH. This process is illustrated in Figure 28:
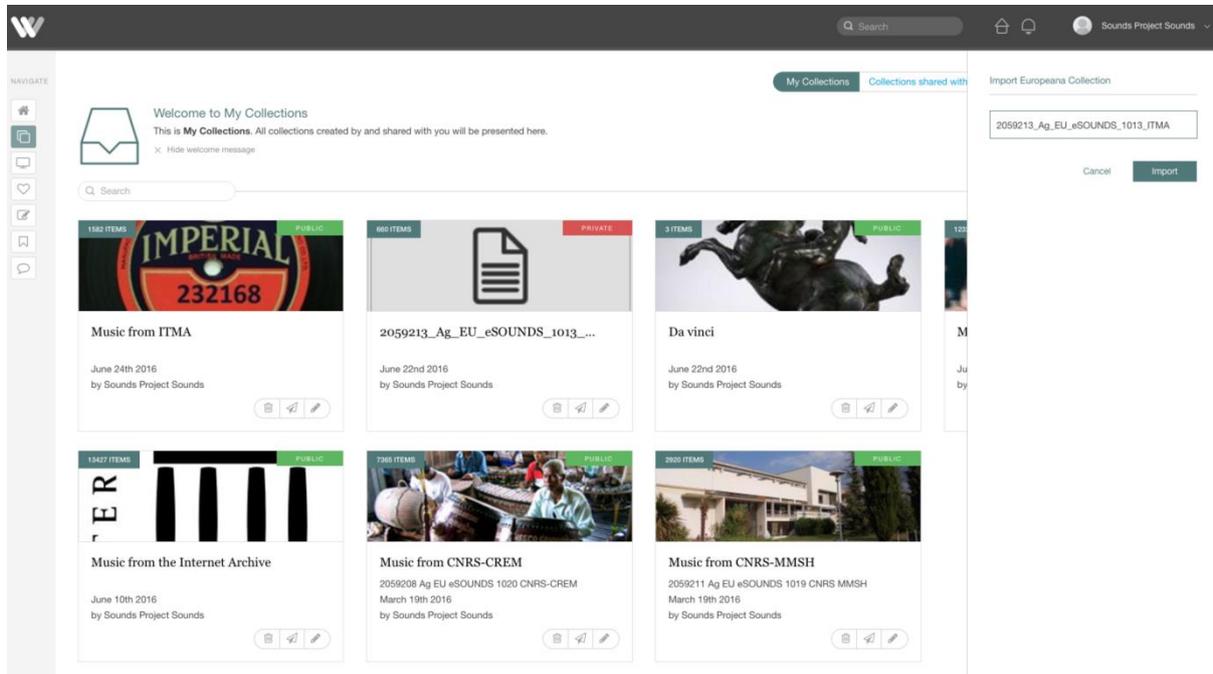
Figure 28: Import collections from Europeana using the Europeana Collection ID

Step 2 (for data providers): data providers can share their imported collections with the Crowdsourcing Space in WITH. In the ''my collection'' page of their user profile, all collections have a ''Share'' button. Collections can be shared with other users and/or user groups, as is illustrated in Figure 29:
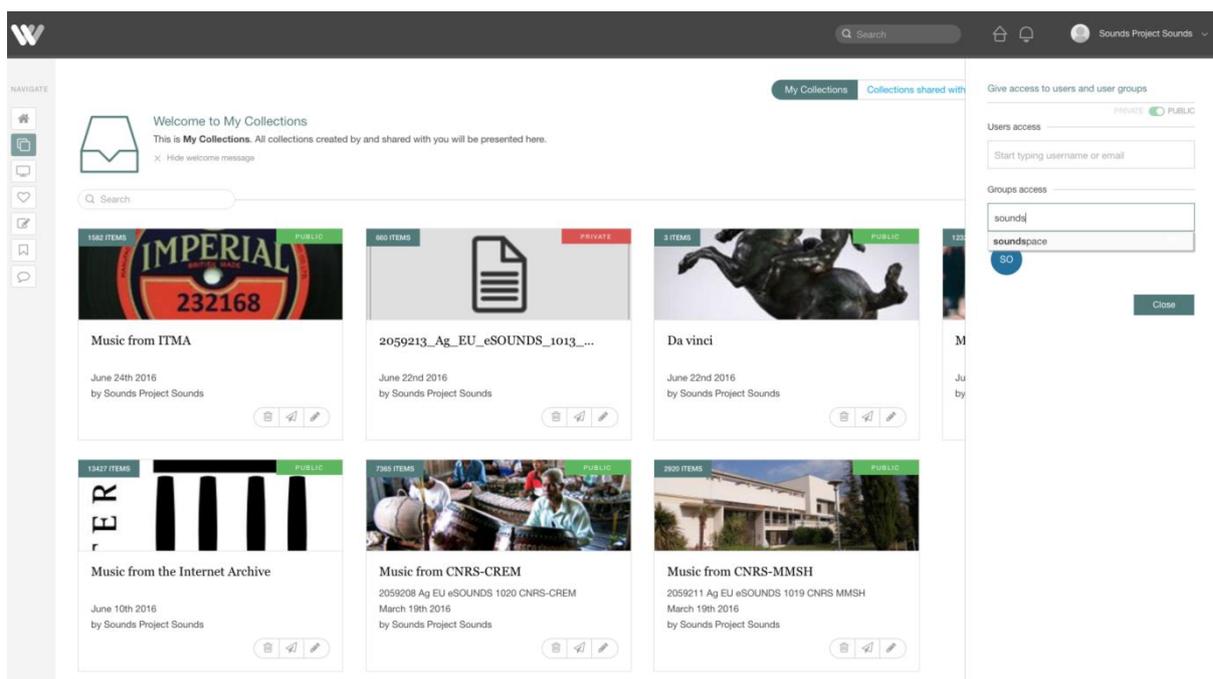
Figure 29: Share collection with the Crowdsourcing Space

Step 3 (for campaign managers): Initiate the crowdsourcing campaign by setting the campaign parameters. These parameters are (example numbers, based on the crowdsourcing campaign ran as described in Section 4.2 Enriching Music Collections with Instruments): The duration (twelve days), the target of tags (2,500 annotations) and the badges (bronze for 10 annotations, silver for 20 annotations and gold for 50 annotations).

Figure 30: The Crowdsourcing Space

Step 4 (for end-users): Users have two options for creating an annotation: 1. Click on the "Start tagging" button and the system will randomly display 10 items (one after the other). 2. Select items manually from the available collections.
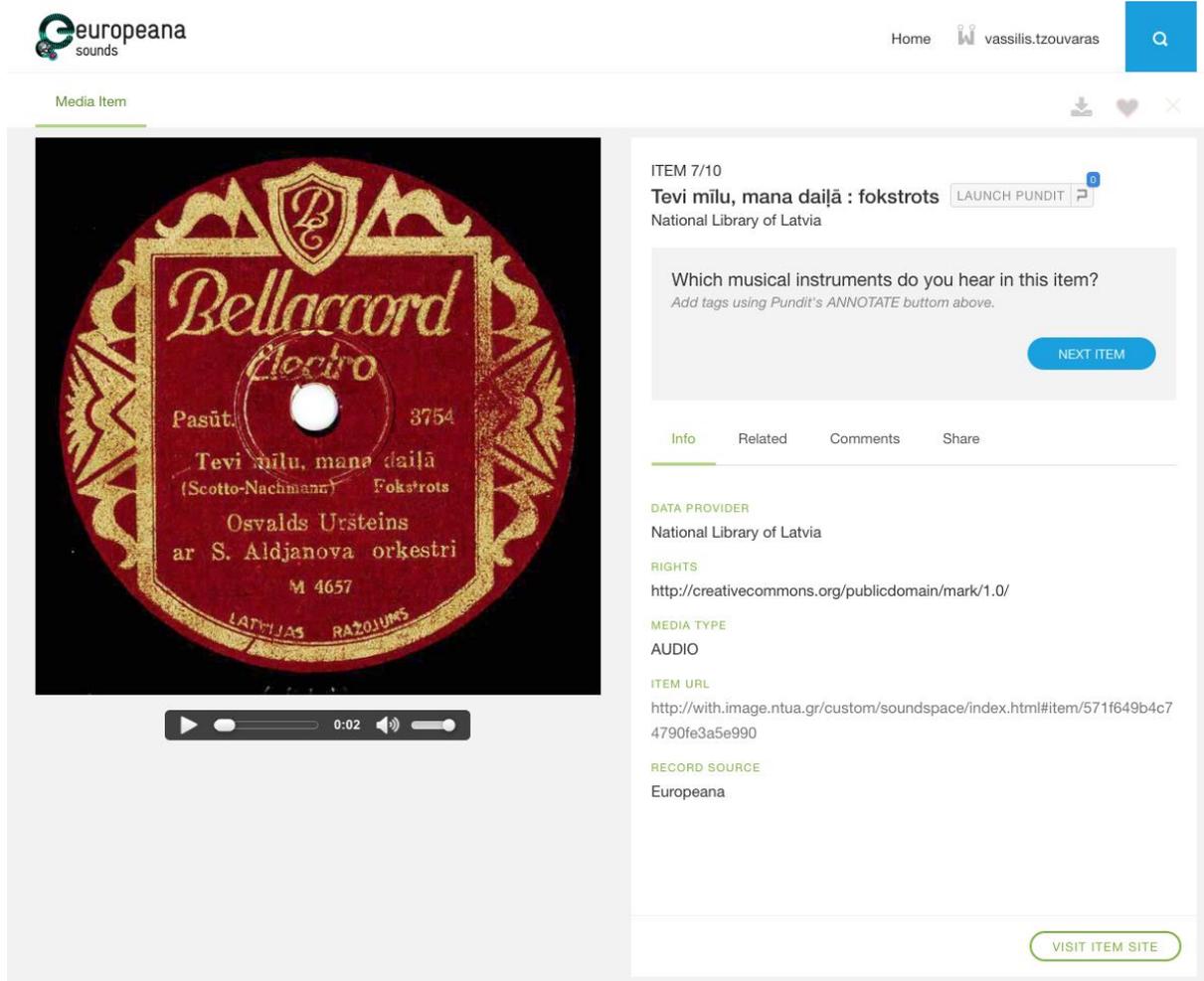
Figure 31: Item preview where users can add annotations

Step 5 (for end-users): The user clicks on the "Annotate" button and selects the "Instrument Tag" option (see more info in the PUNDIT section 3.3.2).
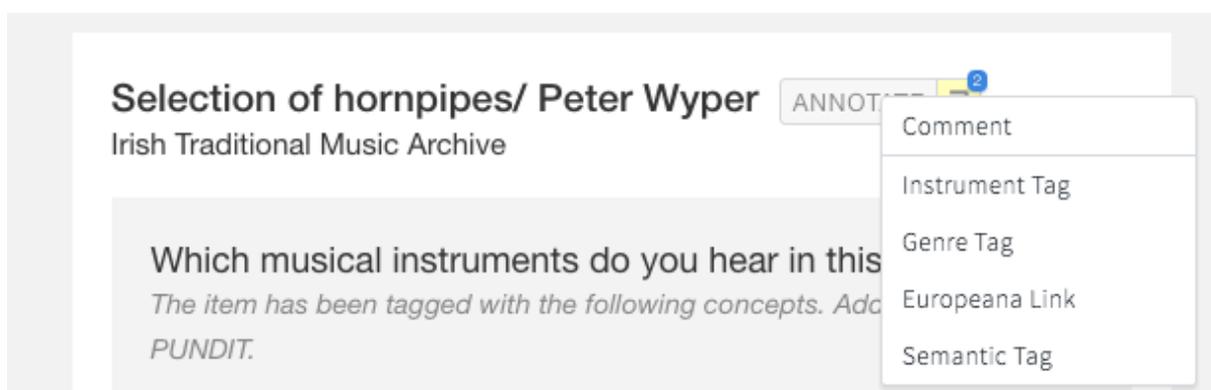
Figure 32: Initiate Pundit for inserting the tags

Step 6 (for end-users): Select the instrument from the MIMO vocabulary and thesaurus lookup service in Pundit. The lookup service uses auto-completion to quickly find the instrument.
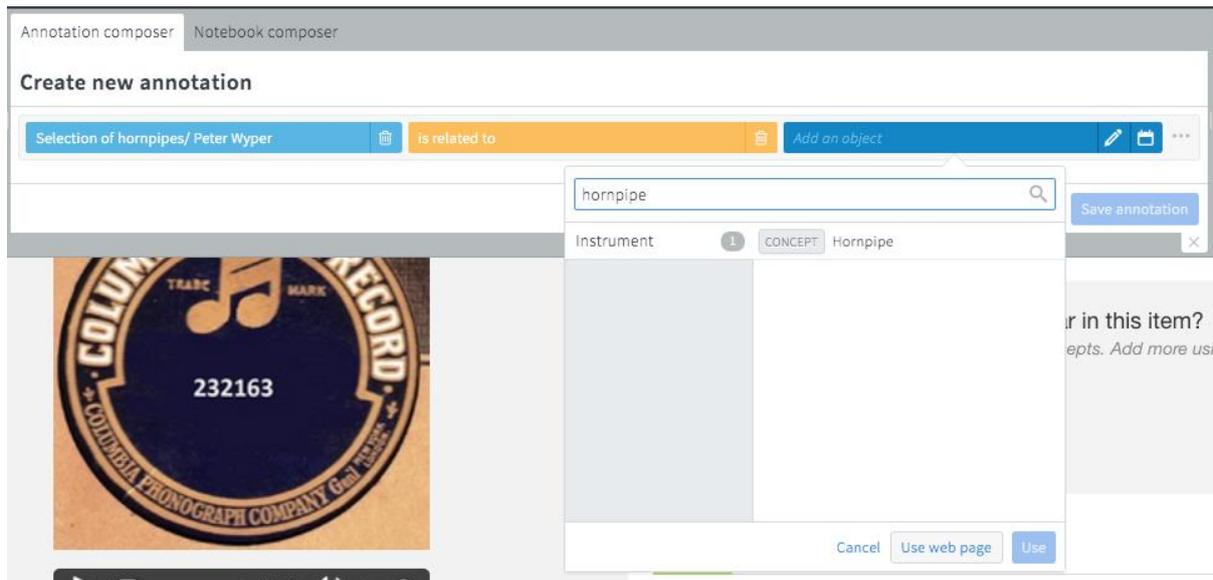
Figure 33: MIMO lookup service
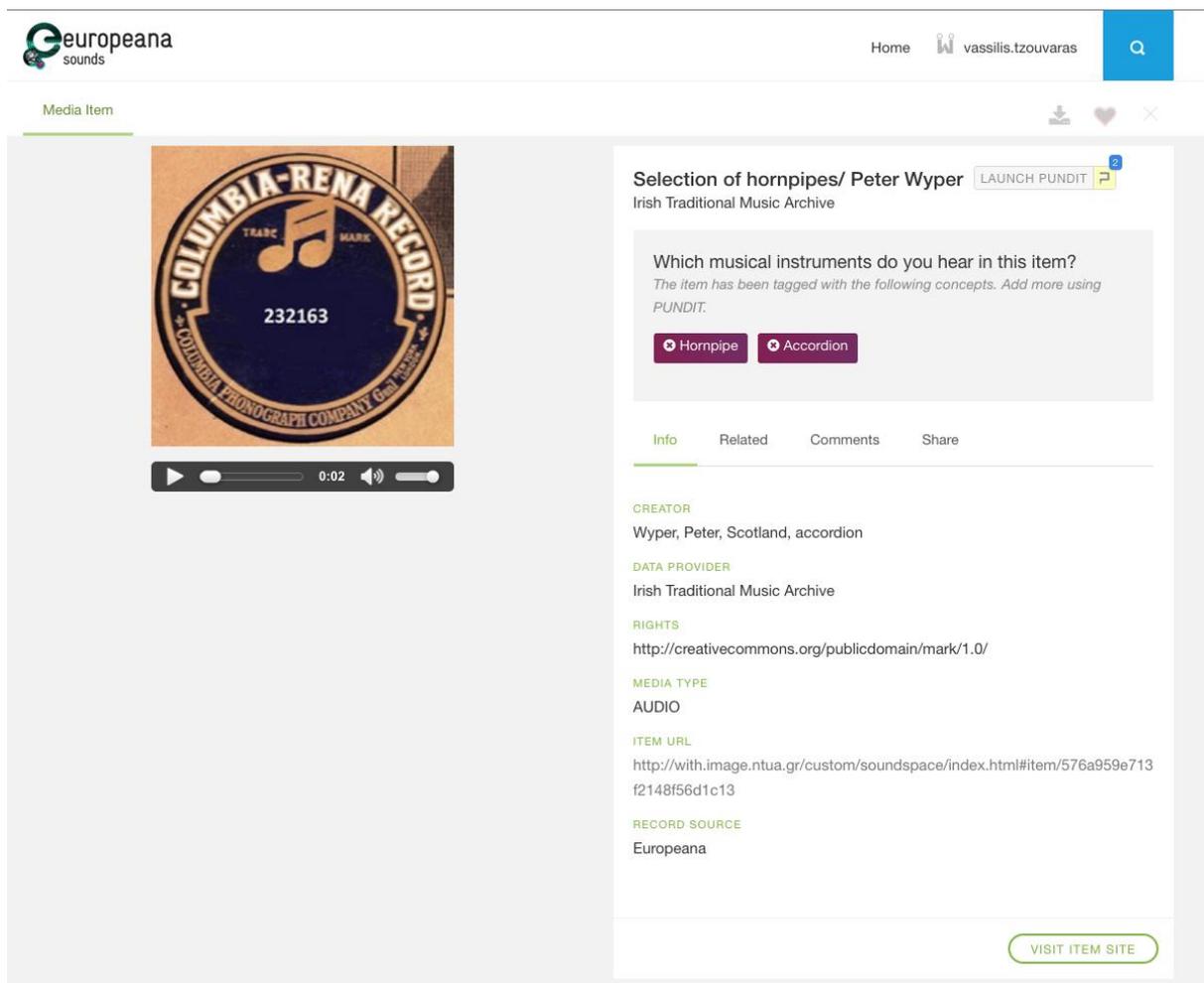
Figure 34: Selected tags appearing on the WITH interface

Step 7 (for end-users): Review the completed tags in the "My Annotations" page. Users can review the tagged items and delete or add more tags.
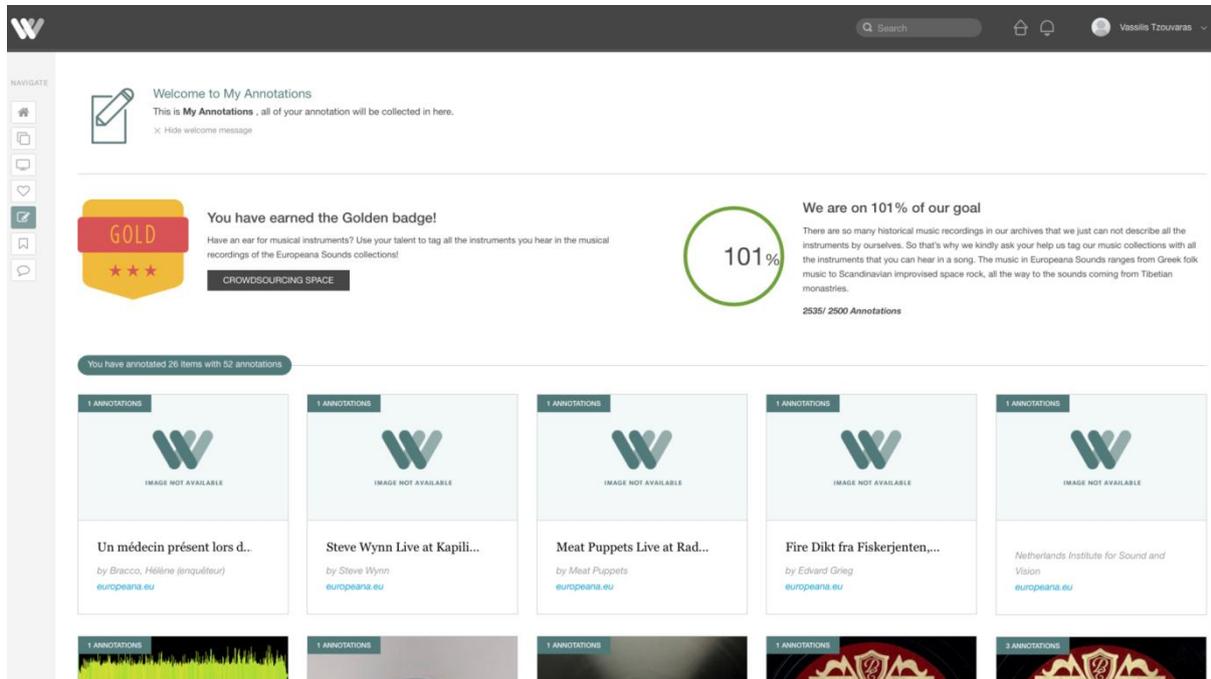
Figure 35: My Annotations page

3.3.5   Next steps

For further development to Pundit, WITH and the integrated solution for Semantic Tagging (which was developed to support the crowdsourcing campaign as described in Section 4.2 Enriching Music Collections with Instruments), the focus will be on:

- Further improving the user experience of the support for manual Semantic Tagging with the integrated solution for (future) online crowdsourcing campaigns

- Deployment of the Pundit widget on Europeana Sounds data providers websites for both manual Semantic Tagging and validation of (semi-)automatic enrichments (more on this in Section 5.1.5 Validation of Semi-automatic Tags)

- Promotion of the Pundit Chrome Browser Extension for manual Semantic Tagging (and possibly also validation of (semi-)automatic enrichments)

- Native functionality with the WITH platform to support centralised crowdsourcing campaigns with a focus on the validation of (semi-)automatic enrichments (this will also support automatic enrichment with various thesauri when data providers import their collections)

3.4   Europeana Music Collection

At the time of writing this report, the Europeana Music Collections have not implemented a connection with the Annotations API yet. This will happen in the next six months and will be outlined as follows:

- Displaying existing annotations in Europeana Collections: to display annotations that have been sent to the annotations API in Europeana Collections.
  Timeframe: August 2016
- Allowing users to annotate musical items with music genres (see Figure 36).
  - On all item pages where items are either from the provider "Europeana Sounds" or are known to be retrieved via the Music Collections, we will show the ability for a user to specify or enhance the music genre.
  - The user will be shown a dropdown or a search input box to select or search a music genre from the Europeana Sounds genre vocabulary, and add this as an annotation to the item.
  - When a user wants to do so, they can login with a MyEuropeana[54] account first in order to link the annotations to their user account.
  Timeframe: October 2016

- Allowing users to annotate any Europeana item with tags and links.
  Timeframe: Gradually implemented until the end of the project (January 2017).

Figure 36: Wireframe with call-to-action for genre tagging in Europeana Music Collection

---

4 Europeana Sounds crowdsourcing campaigns

This section reports on the first experience with, and results gained from the enrichment through crowdsourcing, by utilising and promoting the crowdsourcing applications in two specific crowdsourcing campaigns. It describes the concept, results and findings from the evaluation of the two crowdsourcing campaigns that were held from the 1st of June till the 30th of June 2016. In these two campaigns, WP2 utilised two of the Europeana Sounds crowdsourcing applications: Historypin (see Section 3.2 Historypin geotagging interface) and the Pundit integration in WITH (see Section 3.3 Pundit Integration in the WITH platform).

The general concept for the crowdsourcing campaigns originated in the Europeana Sounds plenary meeting in Lisbon (28-29 January 2016). During this meeting WP2, together with the data providers and WP6 discussed the deployment of crowdsourcing applications that the project is developing. WP2 proposed to the project partners that the focus of the first crowdsourcing activities should be on dedicated crowdsourcing tasks, organised centrally (meaning involving as many data providers with the same information needs and disseminating the crowdsourcing activities to the public through the main Europeana Sounds communication channels as a unified project). This happened in collaboration with WP6, who agreed to give priority to the online promotion required for a whole month, in order to turn these activities into a promotion campaign in the month of June 2016.
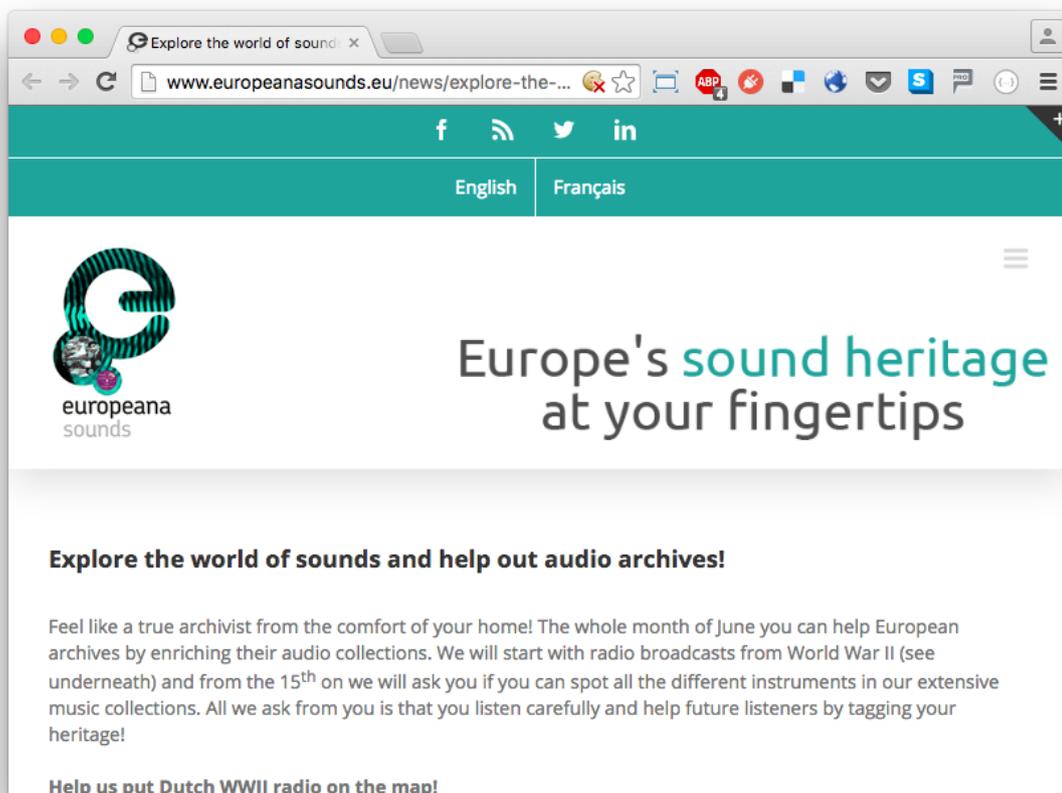
Figure 37: Blogpost on the Europeana Sounds website announcing the crowdsourcing campaigns in June 2016

To involve the data providers in the campaigns, WP2 organised a workshop at the metadata training workshop in Dublin (28-29 April 2016)[55]. The data providers were asked to provide datasets, and the workshop went through the process of integrating these datasets into the crowdsourcing applications, to enable enrichment through crowdsourcing. Through a hands-on session data providers could also experiment with the enrichment process themselves.

4.1 Geolocating radio broadcasts

4.1.1 Concept
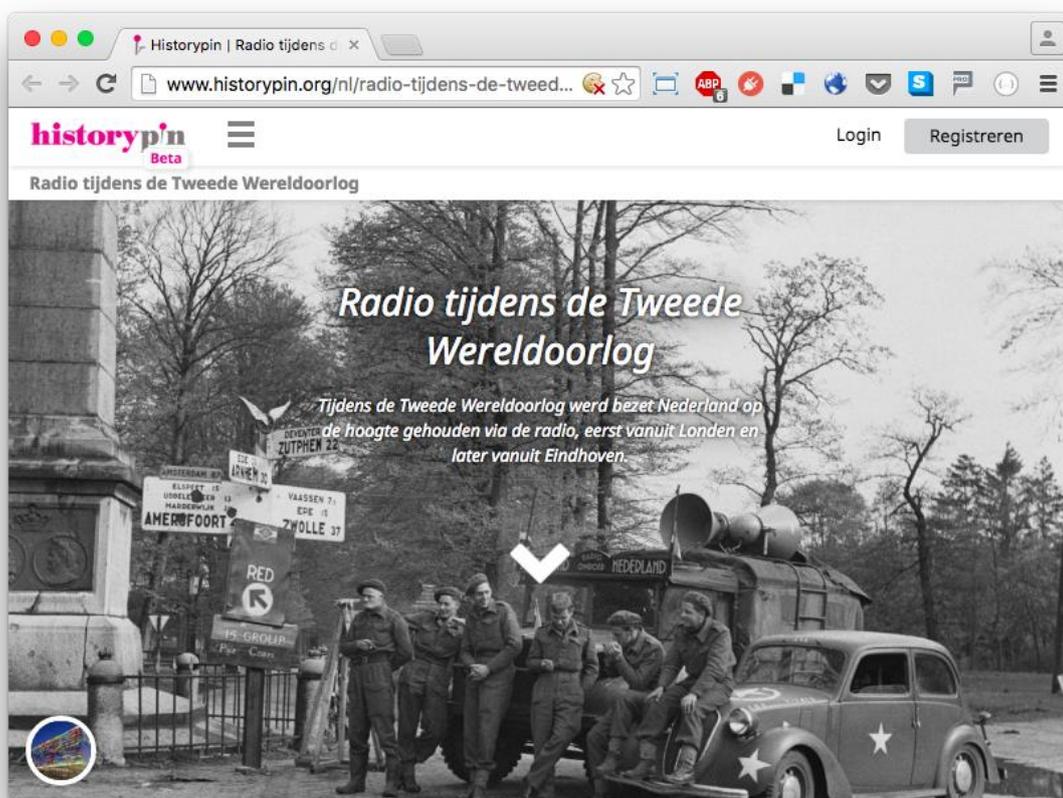
Figure 38: Landing page on Historypin for the Geolocating Radio Broadcasts Crowdsourcing Campaign

For enriching and improving the geographic location of records from the Europeana Sounds project, through crowdsourcing, a Geolocating Radio Broadcasts Crowdsourcing Campaign ran from Wednesday the 1st of June until Wednesday the 15th of June. The Historypin geotagging interface facilitates users to easily pin an item to a specific geolocation (identification of the real-world geographic location, as described in Section 3.2 Historypin Geotagging Interface).

An NISV collection of World War II Dutch radio programs broadcasted from the (occupied) Netherlands was made available on the landing page on Historypin for this campaign (see Figure 38). Most of the

---

[55] http://www.europeanasounds.eu/news/europeana-sounds-team-in-dublin

radio programs in this collection have a specific location as subject, for example, the radio coverage of the liberation of Tilburg. For the purpose of this crowdsourcing campaign, the Historypin interface was translated into Dutch, as the collection that was used for this crowdsourcing campaign was only in that language.

In order to perform the bulk upload from Europeana to Historypin, a spreadsheet was created by Historypin to gather all the metadata records and links to SoundCloud from NISV. For this geolocation crowdsourcing campaign, only one Europeana Sounds dataset was re-published on Historypin. A limiting factor for inclusion in Historypin was that only datasets that are hosted on either SoundCloud[56] or audioBoom[57] (previously called Audioboo) could be re-published. Originally this crowdsourcing campaign was meant to include datasets from multiple data providers but the requirement to include audio hosted on only SoundCloud or audioBoom turned out to be more of a limiting factor than anticipated, especially as SoundCloud is also used for re-distribution of selected Europeana Sound material for promotional purposes.

Figure 39: Tweet from NISV to promote the crowdsourcing campaign

(Translation: "Help us pin radio from WW II on the map!")

As the 258 available radio broadcasts are all in Dutch the target audience was restricted to users with sufficient knowledge and understanding of the Dutch language. The main calls-to-action for the crowdsourcing campaign were written in Dutch, as was the Historypin landing page with the NISV

---

[56] https://soundcloud.com/
[57] https://audioboom.com/

collection.[58] As well as the blog posts on the Europeana Sounds website [59] and NISV[60] there were four Facebook posts published by NISV and Europeana Sounds, and nine tweets were sent out from Europeana Sounds[61], NISV[62] (see Figure 39), NISV R&D[63] and Historypin[64] Twitter accounts with a combined reach of 29,821 Twitter followers. The internal NISV newsletter and the external platform WO2 actueel[65] published a-call-to-action as well. This was all coordinated in collaboration with WP6.

4.1.2   Results

These combined communication efforts from the 1st of June till the 15th of June 2016 resulted in the landing page on Historypin obtaining 176 page views during the crowdsourcing campaign (an average of 11.73 visits a day), in comparison there were 37 visits in the 32 days after the campaign was finished (average of 1.16 visits a day). The blog post that announced the campaign on the NISV website attracted 374 page views (343 unique visitors) from the 1st till the 15th of June, with an average time on page of 05:19 min. In the 15 days of the Geolocating Radio Broadcasts Crowdsourcing Campaign there were only 3 enrichments made by 2 contributors (see for example the enrichment depicted in Figure 40).
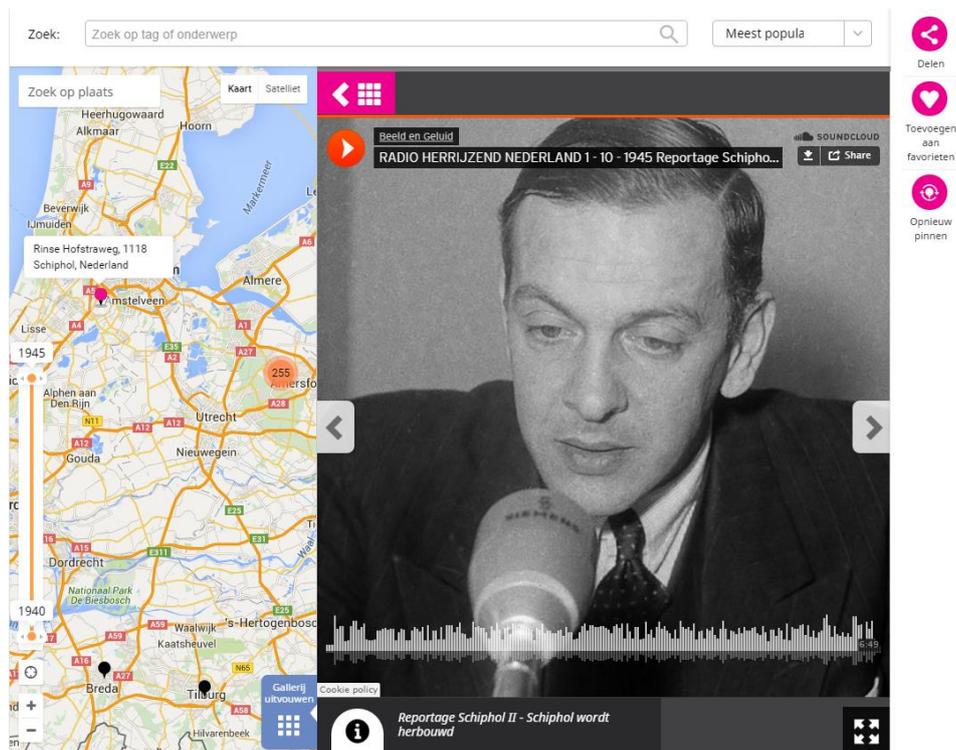
Figure 40: One of the radio items that was enriched with a geolocation during the crowdsourcing campaign, a radio broadcast that informed people about the rebuilding of Schiphol airport.

---

[58] http://www.historypin.org/nl/radio-tijdens-de-tweede-wereldoorlog/
[59] http://www.europeanasounds.eu/news/explore-the-world-of-sounds-and-help-out-audio-archives
[60] http://www.beeldengeluid.nl/blogs/research-and-development/201606/help-ons-radio-uit-woii-op-de-kaart-te-zetten
[61] https://twitter.com/eu_sounds
[62] https://twitter.com/BeeldenGeluid
[63] https://twitter.com/benglabs
[64] https://twitter.com/historypin
[65] http://www.wo2actueel.nl/artikel/4901/Help-radio-uit-de-Tweede-Wereldoorlog-op-de-kaart-te-zetten.html

4.1.3   Evaluation

To evaluate the unexpectedly low result of the Geolocating Radio Broadcasts Crowdsourcing Campaign, the circumstances that have affected this will be described here:

- The crowdsourcing task is rather time consuming, since the location of the radio broadcast is not always mentioned at the start of an audio item. This means the audience needs to be invested in the crowdsourcing task, in order to complete it. One could conclude that online promotion towards a (mostly) general audience of 'culture snackers' is not sufficient for this specialised type of annotation (see Section 2.2.1.4 Geotagging). In relation to this point, the crowdsourcing campaign would have benefitted from an extension with a physical event (with more specialised communities of interest), similar to the ones that will be described for the other crowdsourcing campaign that were held (see Section 4.2 Enriching Music Collections with Instruments).

- It is also worth noting that the collection is not a perfect match for the Historypin geotagging interface. Some radio items in the collection deal with multiple locations. This is largely due to that fact that audio is a time-based medium. However, the Historypin geotagging interface is treating audio as a static item, meaning users can only pin the item to one geolocation.

- Finally the limited amount of visits that the landing page on Historypin attracted during the crowdsourcing campaign (11,73 visits a day) can be attributed to the fact that the crowdsourcing task could only be performed by a Dutch-speaking audience. In order to attract a larger and more international audience for this crowdsourcing task, it would be beneficial to include more diverse datasets that require geotagging.

4.2   Enriching Music Collections with instruments

4.2.1   Concept

The second crowdsourcing campaign was aimed at enriching the musical recordings aggregated by Europeana Sounds with musical instruments. In order to gather uniform terms, the users were given the option to create Semantic Tags (see Section 2.2.1.3 Semantic Tagging) using the Musical Instrument Museums Online (MIMO) Vocabulary and Thesaurus.

Users were guided to a dedicated landing page (see Figure 30) on the WITH platform (see Section 3.3.4 The WITH platform), specifically created for the Enriching Music Collections with Instruments Crowdsourcing Campaign, that provided a random entry point to the music recordings from Europeana Sounds that were selected by the data providers as potentially benefitting from enrichment through crowdsourcing. Eight data providers re-published (part of) their collection on WITH, a total of 31,888 musical recordings (see Figure 43).

When consulting a music recording via the WITH platform, users were able to create a Semantic Tag via the Pundit crowdsourcing application that provided them with the ability to annotate the item with musical instruments from the controlled vocabulary that MIMO provides.

The Enriching Music Collections with Instruments Crowdsourcing Campaign started at Monday the 20th of June and lasted until Thursday the 30th of June. The campaign was mainly promoted through Twitter, Facebook, the Europeana Sounds website[66] and newsletter, the Europeana newsletter and Music Collection, and several websites from data providers[67].

Two physical events with sessions dedicated to the Enriching Music Collections with Instruments Crowdsourcing Campaign were also organised. The first event took place on Monday the 20th of June by CNRS-CREM in Paris, to mark the start of the campaign (see Figure 41). A total of 11 people attended the event, ethnomusicology tutors and students, people working in information management, and one computer scientist participated. The event was focused on enriching the French collections (CNRS-CREM, CNRS-MMSH, BnF), because it was easier for the participants to consult the original metadata records in their native language. The day started with presentations about Europeana Sounds, and the MIMO Vocabulary and Thesaurus. Training on how to use the Pundit and WITH crowdsourcing applications was also provided.
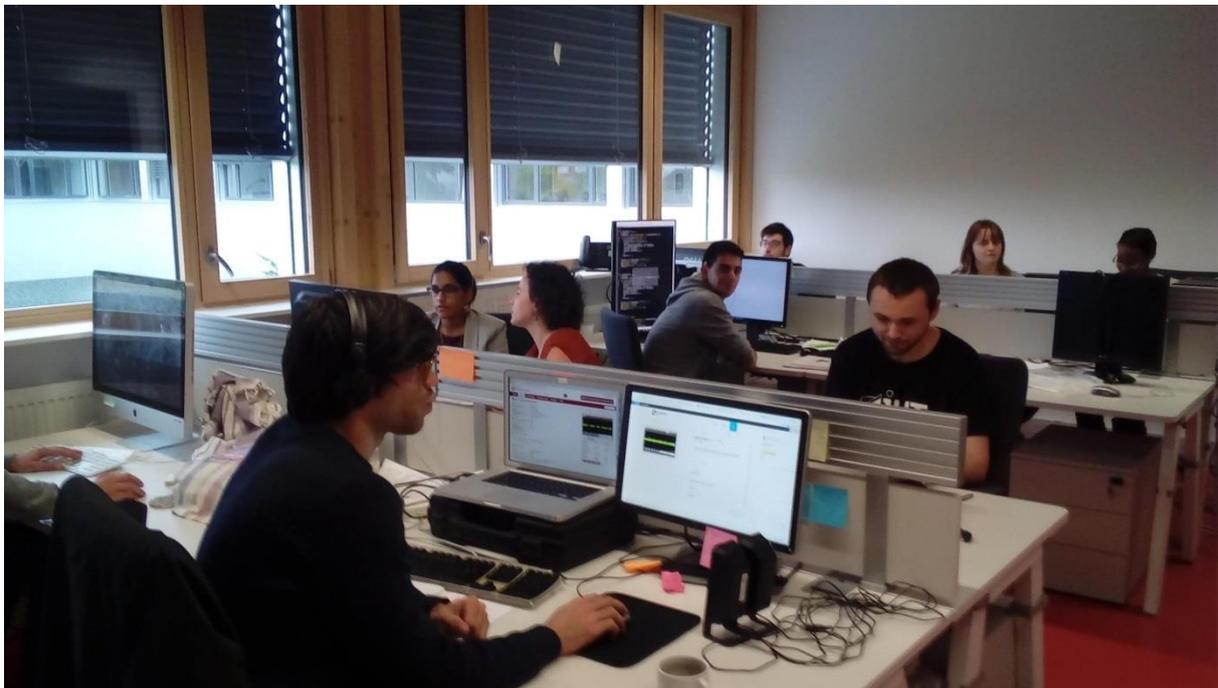
Figure 41: Participants of the physical crowdsourcing event organised by CNRS-CREM on the 20th of June.

In their work process, the participants opened the original metadata records of an item in one web browser tab and the corresponding historical recording in WITH in another tab. The participants mostly used the unstructured instrument names that were in the original metadata as a starting point, and then listened to the audio item on the WITH platform to decide whether the corresponding term from the (structured) MIMO Vocabulary and Thesaurus could be added to the object as an enrichment, or not.

From post-event interviews with a selection of top contributors to the physical events and a survey distributed among all participants, we have learnt that the participants and organisers enjoyed the

---

[66] http://www.europeanasounds.eu/news/identify-musical-instruments-in-our-recordings
[67] https://www.beeldengeluid.nl/blogs/research-and-development/201606/tag-instrumenten-onze-muziekcollecties

competitive element of the event. The tutors also liked the fact that students could experience the use of technical environments outside of the regular environments in use at CNRS-CREM and that everyone had the chance to discover more about Europeana and its database. After the event the organisers explained that it would have been good to see if an item is already annotated in the result page of WITH (this has since been resolved by NTUA).

The other event was organised by FMS and NTUA on Thursday the 30th of June in Athens as the closing event for this crowdsourcing campaign (see Figure 42). Ten people participated in this event, employees of both FMS and NTUA participated, but also regular members of the Music Library of Greece were present.

The primary aim of the closing event was to enrich the FMS collection. However, during the event participants discovered that not all Greek traditional instruments were represented in the MIMO Vocabulary and Thesaurus and thus the goal was extended to enrich all Europeana Sounds collections. The event lasted two hours and started with a short training for the Pundit and WITH crowdsourcing applications, after which the actual enrichment through crowdsourcing began.

Figure 42: Crowdsourcing event organised by FMS and NTUA on the 30th of June

Both participants and organisers were positive about the event; the fact that there were musicologists as well as people with knowledge of the tool was considered an important benefit of this physical meeting. A minor critical point that came from the organisers, as well as the participants, was the double registration (for WITH and Pundit) before starting the annotating; they would rather only login

once. Another lesson learned from the organisers was to start promoting the event earlier in order to increase the number of participants.

4.2.2   Results

In 12 days there were 41 tweets about this particular Europeana Sounds crowdsourcing campaign (using the hashtag #CrowdsourcingMonth, see Figure 39). 16 of those tweets came from Europeana Sounds (2,712 followers at the beginning of the campaign).

From the Europeana Sounds tweets, users clicked 99 times on the link to the Europeana Sounds blog posts and 122 times on the link to the landing page on the WITH platform directly. The tweets coming from Europeana Sounds were retweeted 160 times and received 110 likes in total. The other 25 tweets came from 7 individuals and 7 organisations, with a combined reach of 22,947 followers. Furthermore 10 Facebook posts about the Enriching Music Collections with Instruments Crowdsourcing Campaign were published via the Europeana Sounds Facebook account from Monday the 20th of June till Friday the 1st of July. 6 posts were shared 17 times by 6 individuals and 6 organisations. The posts received 65 likes and were loved twice.

The public dissemination efforts in collaboration with WP6 led to 352 page views (350 unique visitors) of the main Europeana Sounds blog post about this campaign, and users spent an average of 3 minutes and 56 seconds on this blog post that linked to the WITH landing page and provided the reader with instructions on how to create an annotation.

From the 20th of June till the 30th of June there was 51 contributors that created a total of 2,581 annotations (see Figure 43).
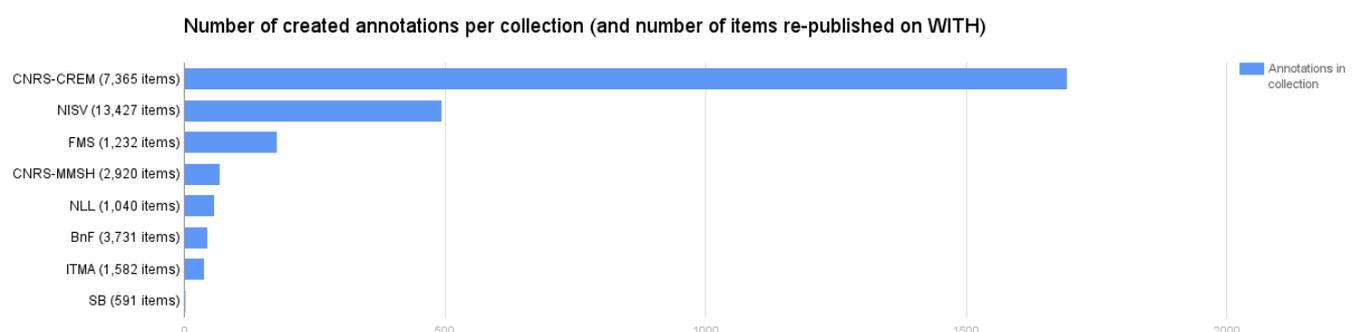
Figure 43: The number of items re-published on WITH and the number of annotations per collection

The top four contributors to the Enriching Music Collections with Instruments Crowdsourcing Campaign created 69.7% of all the annotations that were created. The 21 participants who attended one of the 2 physical crowdsourcing events in Paris and Athens created 90% of all the annotations that were created. The remaining 30 contributors can be divided into a group of 16 people that are affiliated with the Europeana Sounds project (4.2% of all annotations) and 14 external contributors who created 5.8% of all the annotations that were created during this crowdsourcing campaign.

Percentage of created annotations per group

- CNRS-CREM event (11 contributors)
- FMS NTUA event (10 contributors)
- Europeana Sounds (16 contributors)
- Other (14 contributors)
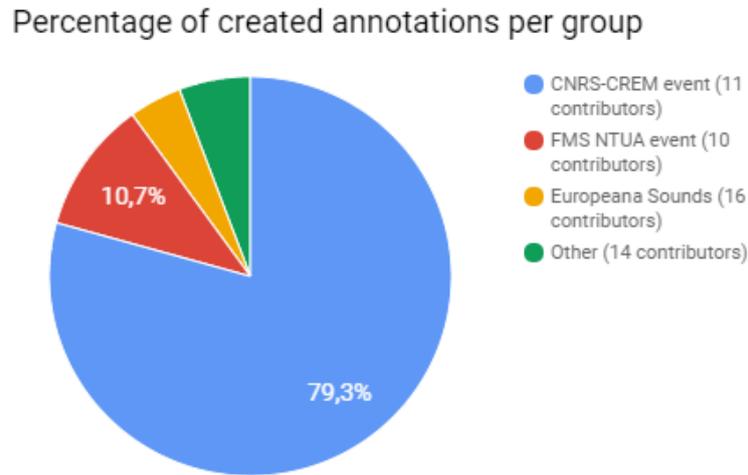
Figure 44: Percentage of annotations of the contributors to the crowdsourcing campaign categorised

4.2.3   Evaluation

The results show that the small group of contributors is responsible for a significant amount of annotations (69.7% of all annotations were created by four persons) and that the participants of the two crowdsourcing events generated 90% of all the annotations gathered in the Enriching Music Collections with Instruments Crowdsourcing Campaign. The over-representation of a small amount of "super taggers" to the bulk of the contributions is a common characteristic of crowdsourcing initiatives (Oomen & Aroyo, Ref 5).

The top four contributors were all attendees of a physical crowdsourcing event. This shows that the physical crowdsourcing events were greatly beneficial to this campaign, when - for instance - compared to the online promotion and the engagement which resulted from that (just 10%) and compared to the Geolocating Radio Broadcasts Crowdsourcing Campaign, which was promoted through online communication channels and struggled to entice substantial engagement (see 4.1 Geolocating Radio Broadcasts).

To gain more insight in the motivation and enrichment process, and to get feedback on the annotation tool from engaged users, several questions, through a post-event email survey, were asked to the most active contributors:

- The contributors explained that they were motivated because they could see their progress on the WITH platform: ''the fact that the annotations were saved and displayed was very motivating.'' The competition aspect was mentioned as another motivating factor, the chance to discover sounds from the different institutions that participated was also appreciated by the participants.

Contributors distribution
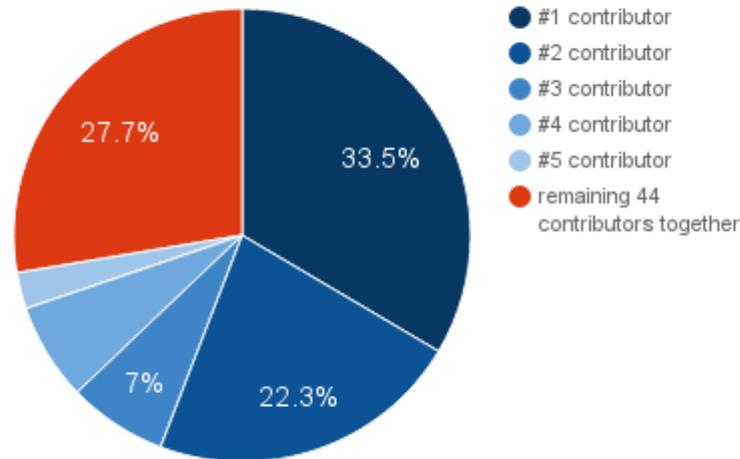
Figure 45: Distribution of contributors in percentages of total annotations

- When asked what would encourage the use of the crowdsourcing application again in the future, the contributors mentioned that another physical event would be encouraging and that participating together in a team of people would be motivating. One suggested to apply the same concept of crowdsourcing to music video content as well (concert registrations, etcetera): "because you can get a lot of information when you see a musician play."

Figure 46: The "My Annotations" page in WITH, seeing progress was mentioned as motivating by the top contributors

- The contributors also gave feedback on their experience with using the crowdsourcing application. The experience in general was good, and was rated on average at 3.75 (with 5 being a great

experience and 1 a poor experience). One user praised the swift workflow that the tool facilitates: "you can find the instruments quickly and get a lot of work done quickly."

Critical points were given as well:

- Two contributors reported that it would have been better if a single sign-on system had been in place, instead of the need to register for WITH as well as for Pundit to start creating annotations.

- The other two contributors needed some time to become familiar with the functionality: "the first time I created an annotation, I had to get some help from someone who already knew the crowdsourcing application, otherwise it would have taken much longer for me to understand how to create an annotation and to get used to the tool."

- The contributors also had some trouble with the terminology and language of the instruments in the MIMO Vocabulary and Thesaurus: "sometimes you need to click first on the term in another language first, in order to find the proper term in French" "it seems to me that there are several different names which actually refer to one single instrument."

- Furthermore, it was suggested to make available the option of annotating vocal parts of the music as well, using an additional controlled vocabulary.

Based on the answers about the motivation of the most active contributors it seems beneficial to organise more physical crowdsourcing events in the future, to further optimise user feedback to reflect their progress (see Figure 46) and continue to build a diverse offering of music collections. The user feedback received during the crowdsourcing campaigns will be taken into serious consideration for the further development of Pundit and WITH (as described in Section 3.3.5 Next Steps). The issues related to the terms in the MIMO Vocabulary and Thesaurus are unfortunately related to limitations of their API, and therefore cannot be improved by WP2. We will however relay these issues to the MIMO project.

5 Enrichment strategy and ambitions

This section reflects on the project's strategic approach towards enrichment through crowdsourcing, and its progress towards the key performance indicators.

5.1 Enrichment through crowdsourcing scenarios

Although the Europeana Sounds Description of Work (Ref 6) and its included KPIs mostly foresee enrichment through crowdsourcing taking place in the form of "annotations" (KPI 9) and "new connections" (KPI 10), the current experience of the project members (especially WP2 members) has led to a more extensive set of crowdsourcing scenarios, within the wider context of data enrichment and user engagement.

As identified by the reviewers at the Europeana Sounds Technical Review for Year 2, these findings are not only relevant for the crowdsourcing tasks and KPIs that the project aspires to achieve, but also as insights in relation to user engagement for the network that Europeana Sounds has established as a consortium representing the sound domain in the Europeana ecosystem (and beyond):

> *Crowdsourcing activities in WP2 [...] provide new learnings in terms of engagement, usability, technologies. Crowdsourcing activities have also been the occasions to inform the consortium about how end-users conduct specific tasks and content-related activities as part of their day-to-day work.*

Ref 7: Extract from the Europeana Sounds Technical Review Report (Year 2) focussing on WP2

The table below (Table 2) provides a summary of possible ways to strategically utilise crowdsourcing within an enrichment and/or user engagement strategy, which are then described in more detail in the following sections below:

Table 2: Summary of WP2 crowdsourcing strategies

Crowdsourcing scenario	Crowdsourcing application support	Impact at scale	Benefit	Target audience
Simple Tagging	Pundit (NET7) Music Collection (EF, to be implemented)	Low	Bridges the semantic gap	Culture Snackers (broad)
Semantic Tagging	Pundit (NET7) Music Collection (EF, to be implemented)	High	Unifies and links metadata across records and collections Unifies and links metadata to the web Improves multilingual retrieval and representation	Culture Vultures (niche)
Geotagging	Historypin (Historypin)	Medium	Enables geographic data manipulation and visualisation ('mapping')	Culture Vultures (niche)
Object Linking	Tunepal (Historypin, to be implemented) Music Collection (EF, to be implemented)	Medium	Link metadata across records and collections	Culture Vultures (niche)
Validation of Semi-automatic Tags	Pundit (NET7) WITH (NTUA, to be implemented)	High	Combines curation, crowd intelligence and machine intelligence	Culture Snackers (broad)

5.1.1 Simple Tagging

Simple Tagging is supported as a type of annotation in the Annotations API (see Section 2.2.1.2 Simple Tagging), and implemented in Pundit (see Section 3.3 Pundit Integration in the WITH platform), Historypin (see Section 3.2 Historypin geotagging interface) and (soon) also in the Europeana Music Collection (see Section 3.4 Europeana Music Collection). Given its unstructured 'free' tagging nature, and the fact that Simple Tags are language specific, the impact at scale - Europeana now includes over 50 million metadata records - is low. This does not mean the resulting enrichments are not of value. The unstructured nature of these tags, also comes with the benefit of bridging the so-called "semantic gap" through the use of non-expert terms to enrich the metadata records (Ref 5). Next to that it is also an accessible activity, for a broad set of users (Culture Snackers, MS7 Ref 8) and can be disseminated to a broad public.

5.1.2   Semantic Tagging

Semantic Tagging is supported as a type of annotation in the Annotations API (see Section 2.2.1.3 Semantic Tagging), and implemented in Pundit (see Section 3.3 Pundit Integration in the WITH platform) and (soon) also in the Europeana Music Collection (see Section 3.4 Europeana Music Collection). Given the usage of controlled vocabularies and thesauri, these enrichments benefited from being structured, hierarchical and multilingual. This means that at scale Semantic Tagging has high impact, since one enrichment often results in multiple (multilingual[68]) labels being associated with the metadata records (and greatly adds to the data quality of that metadata record). Next to that, the structured nature of the enrichment links the metadata records to other records that use the same URI, across collections and even across domains, extending the links to other sources on the web. However, this type of enrichment through crowdsourcing requires some domain knowledge and is therefore more suitable for a niche audience of Culture Vultures (Ref 8) and requires a targeted public dissemination strategy.

5.1.3   Geotagging

Geotagging is supported as a type of annotation in the Annotations API (see Section 2.2.1.4 Geotagging), and implemented in Historypin (see Section 3.2 Historypin Geotagging Interface). Considering the specificity of this type of annotation (not all metadata records can be associated with a geographical location), the impact at scale is medium. It does however provide the benefit of re-using the metadata in interesting ways (like map visualisations for example), seen from the end-user perspective. Enriching metadata records with geotags through crowdsourcing required targeting local (historic) communities with a geographical awareness. It is therefore for a niche audience of Culture Vultures, and - like Semantic Tagging - requires a targeted public dissemination strategy.

5.1.4   Object Linking

Object Linking is supported as a type of annotation in the Annotations API (see Section 2.2.1.5 Object Linking), but currently remains unsupported by any of the Europeana Sounds crowdsourcing applications. It will be implemented in Tunepal in the near future (see Section 3.1 Tunepal widget) and (soon) in the Europeana Music Collection (see Section 3.4 Europeana Music Collection). At scale, its impact is medium. Not all metadata records have a meaningful relation with other sources or entities in the (Europeana) ecosystem. Object Links do provide the benefit of being able to compute relevant related content, based on the links, especially if the type of relationship has been made explicit. Like Semantic Tagging, Object Linking requires some domain knowledge and is therefore more suitable for a niche audience of Culture Vultures and requires a targeted public dissemination strategy.

5.1.5   Validation of semi-automatic tags

Validation of Semi-automatic Tags is supported as a type of annotation in the Annotations API, through the Moderation functionality (see Section 2.2.1.6 Moderation). This crowdsourcing scenario is currently supported by Pundit (see Section 3.3 Pundit Integration in the WITH platform) and will (soon) be supported by WITH. Crowdsourcing the validation of semi-automatic tags offers high impact at scale,

---

[68] For example, the MIMO URI provides labels in 8 languages. For illustration, see: http://www.mimo-db.eu/InstrumentsKeywords/3573

since it combines the scalability of automated processing of large datasets, with relatively simple human cognitive task that can be systematically be leveraged and moderated through crowd intelligence. Therefore these tasks can be disseminated to a broad public of Culture Snackers.

Its potential in volume does however rely on how much metadata can be pre-processed, for instance through alignment. Within enrichment experiments conducted by Europeana R&D[69] and AIT[70] WP2 has identified the great enrichment potential of effectively utilising collection expertise and (semi-) automatic alignment tools (like CultuurLink) to generate suggestions for enrichment of the related metadata records. Together both experiments focussed on a corpus of four datasets from Europeana Sounds data providers (CREM, MMSH, NISV & ONB) and resulted in 84,221 Object Links (KPI 10).

Alignment as a (semi-) automatic enrichment mechanism works at scale, because it does not apply enrichment at the metadata record level, but aggregates the unique entities found in a dataset and tries to align these terms to controlled vocabularies. If relevant candidates for alignments are found, collection experts and/or data managers can assess the results. These alignments are then applied to all instances of the original (unstructured) unique entity in the metadata records that belong to that dataset. Because alignment is applied at the dataset level, validity of a specific enrichment at the metadata record can vary. This is where validation through crowdsourcing is of great added value to further improve upon the (semi)-automatic enrichments.

5.2 Predicted results and reflection on the enrichment KPIs

This section responds to the Europeana Sounds Technical Review Report for Year 2 and the thoughtful comments about the expectations of crowdsourcing within the project. It starts by providing a current status of the WP2 KPIs and the recent results from the crowdsourcing campaigns specifically. Based on the recent findings and the crowdsourcing strategies and applications available to the project, it then provides the predicted results towards the end of the project. It ends with a reflection on the KPIs that were set at the beginning of the project.

5.2.1 Current status of the WP2 KPIs

The table below (see Table 3) provides an updated overview of the current progress that has been made with the WP2 related KPIs. In relation to enrichment through crowdsourcing, KPI 9 and KPI 10 are most relevant. KPI 8 is related to the automatic enrichment that occurs at the point of data ingestion, and is not influenced by crowdsourcing (but simply by the amount of sound metadata records that Europeana aggregates as a whole). KPI 11 is not measured in amount of enrichments, but in the number of people participating in GLAMwiki edit-a-thons. And finally KPI 12 is related to a closed task that greatly outperformed the expected amount of enrichments through the application of MIR methods.

---

[69] Internal working document describing the results of alignment of Europeana Sounds metadata (musical instruments) to the MIMO Vocabulary and Thesaurus (using the CultuurLink application): https://docs.google.com/document/d/1aW21npspoGW2oDnFVVlfHotOS7uW2rsTJ2MPg8_S4b4/edit#heading=h.bz78ahr21sjd

[70] Internal working document describing the results of linking Europeana Sounds metadata (composers) to Linked Open Data resources: https://docs.google.com/document/d/1v6GZTCWL1vxKs4bpAadEQwjYBh7X5d-zUBuTH9z4G7s/edit#heading=h.9d2ybzea9efd

Table 3: Status of WP2 KPIs in Q10

No	Relating to objective / result	Indicator name	Progress (cumulative)			
			YR1	YR2	Q10	Target YR3
8	Enrichment (WP2)	Number of metadata records enriched through semantic enrichment	34,813	667,690	1,345,051	2,000,000
9	Enrichment (WP2)	Number of annotations (tags) added by users	NA	0	2,539	1,500,000
10	Enrichment (WP2)	Number of new connections among records established by users & automatically	NA	55,000	84,221	10,000
11	Enrichment (WP2)	Number of participants in the GLAMwiki edit-a-thons	12	45	101	200
12	Music information retrieval (WP2)	Number of items accessible through the music retrieval service	NA	312,096	312,096	25,000

Looking more closely at KPI 9, it is evident that the project is greatly under-achieving (at 0.2 %) at this stage. As the crowdsourcing campaign described in Section 4.1 Geolocating Radio Broadcasts was the first to utilise the crowdsourcing applications for enrichment through crowdsourcing (beyond the pilot and testing phase), we can only report 3 Geotags and 2,536 Semantic Tags. KPI 10 is greatly over-achieving (at 842 %). It is however important to note that the results with linking (more specifically alignment) up until have been the result of expert-assisted semi-automatic enrichment (see section 5.1.5 Validation of Semi-automatic Tags), not crowdsourcing in the purist sense.

The under-achievement of the enrichment through crowdsourcing had already been announced by WP2 lead Maarten Brinkerink (NISV) at the Europeana Sounds Technical Review for Year 2, and the circumstances at play have been reinforced by the reviewers in their report:

> *The crowdsourcing aspect may offer more resistance tha[n] anticipated, but this is a general phenomenon in a time where the possibilities of self-expression via a plethora of social media are overwhelming.*

Ref 7: Extract from the Europeana Sounds Technical Review Report (Year 2) focussing on WP2

5.2.2   Predicted results towards the end of the project

At the Europeana Sounds Technical Review for Year 2, the reviewers requested WP2 to provide them with a realistic prediction of how the project would perform on the crowdsourcing-related KPIs.

The table below demonstrates this prediction and is a conservative estimate (see Table 4), based on the practical experience with the first two crowdsourcing campaigns and the analysis of the potential of different crowdsourcing scenarios, as described in Section 5.1 Enrichment through Crowdsourcing Scenarios.

Table 4: Prediction of the results for KPI9 and KPI 10 in Q11 and Q12

KPI	Activity	Crowdsourcing application	Enrichment potential	Coverage	Expected result	Target audience
9	At least 6 physical events aimed at Semantic Tagging of Musical Instruments	WITH combined with Pundit	30,000 metadata records at minimum (as currently available in WITH	15 participants per event, create 30 Semantic Tags per hour, for at least 3 hours[71]	8,100	Culture Vultures
9	Continuous online promotion of the Enriching Music Collections with Instruments Crowdsourcing Campaign	WITH combined with Pundit	30,000 metadata records at minimum (as currently available in WITH	1,000 Semantic Tags per month through online promotion	6,000	Culture Snackers
9	Semantic Tagging of Music Genres	Europeana Music Collection	200,000 metadata records without a Music Genre	5% of the page views on the Europeana Music Collection[72] in Q12 result in a Music Genre enrichment	3,750	Culture Vultures
9	Validation of MIMO Alignments	WITH (with or without Pundit, depending on when Semantic Tagging is implemented in WITH natively) Pundit widget on data providers' websites (depending on the amount of deployments) Pundit Annotator Pro for Europeana Chrome Browser Extension	280,000 suggestion for Semantic Tags based on the MIMO Vocabulary and Thesaurus have been generated, as part of R&D experiments by EF and AIT	5% coverage through crowdsourcing, in the form of a simple validation task	14,000	Culture Snackers
				Total	31,850	

The conservative estimate provided above (see Table 4) predicts that KPI 9 can still realistically reach 34,350 annotations (tags) added by users within Q11 and Q12 of the project (cumulative result). This is still a great underachievement at 2.3 % of the original goal.

A less conservative estimate could also take into account the possibilities of a few options that have dependencies that are partly out of reach for the core WP2 members and more specifically the technical partners:

---

[71] This estimate is based on the experience with Enriching Music Collections with Instruments Crowdsourcing Campaign physical events.

[72] Based on a monthly traffic of 25,000 page views.

More physical events focused on (various forms of) Semantic Tagging would be extremely beneficial to a higher KPI 9 result (as evidenced with the Enriching Music Collections with Instruments). But the likelihood of this depends on the assistance from the data providers, which in return depends on them having collections that require specific types of Semantic Tagging (music instruments and genres to start with). Currently there are still at least more than ten data providers with more than one person month available to utilise for WP2 activities during the remaining project months. We will suggest the organisation of a physical event in support of a crowdsourcing campaigns as a valid activity to utilise these remaining resources.

More coverage of the current potential for enrichment through crowdsourcing (as listed in Table 3) could be gained, if reach was extended beyond the (online promotion of) the crowdsourcing campaigns and Europeana Music Collections. Deployments of the Pundit widget on data provider websites for example could be beneficial to reach this goal. Again assistance from the data providers is required here. NET7 is currently exploring this option with two data providers. Again, we will emphasise to the data providers that this is a valid way to utilise their resources within WP2.

Extending the generation of suggestions for Semantic Tags that can then be validated through simple crowdsourcing tasks would also be beneficial to further expanding the realistically achievable results for KPI 9 (and further increasing the results for KPI 10 at the same time). This could consist of a continuation of the alignment experiments by EF R&D and AIT, where datasets are - with the help of collection data manager - are aligned with the MIMO Vocabulary and Thesaurus, extending this work to new datasets/collections. But it could also be an expansion of this type of (semi-)automatic enrichment beyond musical instruments.

NISV is currently - together with Europeana - exploring the possibility of utilising the hackathon event (T6.5.3, Ref 6) to challenge developers to apply an algorithmic approach for determining music genres to all 240,000 items in the Europeana Music Collection. The results of the winning approach that surfaces from the hackathon could then also be the starting point for validation through a simple crowdsourcing task and result in additional Semantic Tags. To gather the resources needed for the valorisation of the hackathon outcome(s), for the benefit of enrichment through crowdsourcing, we are collaborating with the Europeana DSI-2 project[73].

Finally NISV is currently - together with AIT - exploring the possibility of reusing some of the data analysis that was applied for the selection of items from the Internet Archive etree.org Live Music Archive[74] as a possibility to create automatically generated semantic enrichments for the location and artist data. As part of the work in WP1, NISV used various open APIs to process the Internet Archive collection to filter out European artists and European live venues relevant for aggregation by Europeana Sounds. These processes can be reused to generate suggestions for Semantic Tags that can be validated through crowdsourcing.

---

[73] http://pro.europeana.eu/project/europeana-dsi-2
[74] https://archive.org/details/etree

5.2.3   Towards a holistic view of the enrichment KPIs

Given our current experience and lessons-learned with enrichment (through crowdsourcing) within the Europeana Sounds project (and beyond), we feel that some shifting of priorities and project goals is called for.

Our recent experience with the crowdsourcing campaigns and our predicted results for the remainder of the project show that the original goal set for KPI 9 is not realistically achievable by any means.

Our reflection of the enrichment strategy and the different crowdsourcing scenarios available to the project suggest that most value for the project at scale lie in Semantic Tagging and Validation of Semi-automatic Tags (through Moderation, see Section 2.2.1.6 Moderation). These scenarios do not necessarily result in the largest number of "annotations (tags) added by users" (KPI 9), but they can operate at scale, because they combine curation, crowd intelligence and machine intelligence). They also result in high quality annotations that; (1) unify and link metadata across records and collections, (2) unify and link metadata to the web, and (3) improve multilingual retrieval and representation.

To reflect these insights of the project in our KPIs, we propose two alternative options to adjust the KPIs as they exist currently:

Option 1 would be to set KPI 9 to a realistic goal, based on the possible results that have been extrapolated (conservatively) in Section 5.2.2 Predicted Results towards the End of the Project: 40,000

Option 2 would be to combine the goals of KPI 9 and KPI 10 in a single KPI for the "Number annotations established by users and automatically" (corresponding to the types of annotations described in Section 2.2.1 Business Logic), at a level we consider realistic, given the aforementioned prediction and including the current (over-achieving) result in KPI 10: 125,000

While we feel both options would reflect a realistic level of ambition for the crowdsourcing tasks in the project, the second option also reflects the interdependencies between different sources of enrichments within a workflow that operates at scale and aims for high quality metadata.

We regret not seeing a realistic opportunity to meet the ambition of the original KPI 9, as was determined at the start of the project. However, we are confident that our current approach to enrichment is both valuable and realistic. Valuable, because it looks at the scalability of enrichment through different processes (automatic and human-assisted) and the specific added value of crowdsourcing as part of a workflow. Realistic because it takes the practical experience with crowdsourcing of the project as a baseline for the future results.

In light of the reviewers at the Europeana Sounds Technical Review for Year 2 (Ref 7) valid observation that the project learnings are not only relevant for the crowdsourcing tasks, but also as insights in relation to user engagement for the Europeana Network (and beyond), we feel a responsibility to disseminate our experiences at all possible venues.

We will present our experience with enrichment and participation at the Europeana Sounds conference 2016 in Vilnius[75]. WP2 has also submitted a paper proposal on "Linking subject labels in Cultural

---

[75] http://www.europeanasounds.eu/europeana-sounds-conference-2016

Heritage Metadata to the MIMO vocabulary using CultuurLink" to 16th European Networked Knowledge Organization Systems (NKOS) Workshop at the International Conference on Dublin Core and Metadata Applications 2016 (DC-2016)[76]. Furthermore we will ensure that our experience and efforts related to the crowdsourcing infrastructure, applications and scenarios find their way into the Europeana DSI-2 Core project.

The upcoming public deliverable *D2.9 Evaluation report on implementation of semantic enrichment* will - among other things - provide recommendations on how the Europeana Sounds crowdsourcing infrastructure, applications and scenarios can continue to be used and improved by data providers and other Europeana Network members, in order to successfully enrich our shared cultural heritage (within and beyond the sound archiving domain).

---

[76] https://at-web1.comp.glam.ac.uk/pages/research/hypermedia/nkos/nkos2016-dc/call-for-papers.html

6   Conclusions

The final System Architecture for the Europeana Sounds crowdsourcing infrastructure has been designed and is almost fully implemented. At the core of the infrastructure lies the Annotation API. The Annotations API can now support up to five types of annotation, which are compatible with the W3C Web Annotation Data Model:

1. Simple Tagging
2. Semantic Tagging
3. Geotagging
4. Object Linking
5. Moderation

Europeana Sounds is continuing to develop a suite of six end-user facing crowdsourcing applications that are connected to the crowdsourcing infrastructure to support enrichment through crowdsourcing:

1. Tunepal widget
2. Historypin geotagging interface
3. Pundit (various clients)
4. WITH
5. Europeana Music Collection

Europeana gathered its first experience with the enrichment through crowdsourcing, by utilising and promoting two specific crowdsourcing campaigns in June 2016. One focussed on geolocating radio broadcasts, and utilised the Historypin geotagging Interface. The other focussed on enriching music collections with instruments and utilised integration of Pundit and WITH. While the former campaign had disappointing results, the latter was successful, especially because of the combination of online and physical crowdsourcing activities.

After working on the challenges of enrichment and participation for two and a half years, and based on the first results with enrichment through crowdsourcing, WP2 predicts a substantial underperformance with KPI 9 "Number of annotations (tags) added by users".

Looking at the current crowdsourcing scenarios, WP2 proposes to focus the remaining project resources on Semantic Tagging and Moderation (more specifically Validation of Semi-automatic Tags) These scenarios do not necessarily result in the largest number of "annotations (tags) added by users" (KPI 9), but they can operate at scale, because they combine curation, crowd intelligence and machine intelligence). They also result in high quality annotations that; (1) unify and link metadata across records and collections, (2) unify and link metadata to the web, and (3) improve multilingual retrieval and representation.

7 References

Ref 1	*D2.4 Crowdsourcing infrastructure V1 Assessment and Recommendations* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D2.4-Crowdsourcing-infrastructure-V1-assessment-and-reccommendations-v1.0.pdf
Ref 2	*D2.10 Development of the Crowdsourcing Infrastructure* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D2.10-Development-of-crowdsourcing-infrastructure-v1-0.pdf
Ref 3	*D2.2 Functional design of semantic enrichment* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D2.2-Functional-design-of-semantic-enrichment-v1.0.pdf
Ref 4	*D1.3 Ontologies for Sound* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D1.3-Ontologies-for-sound-v1.2.pdf
Ref 5	Oomen, J, & Aroyo, L. (2011). Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges. In C&T *'11, Proceedings of the 5th International Conference on Communities and Technologies*. ACM New York. doi:10.1145/2103354.2103373
Ref 6	*Europeana Sounds Annex I - "Description of Work"* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Other%20documents%20related%20to%20the%20project/Dow%20Europeana%20Sounds%20620591%202015-10-19.pdf
Ref 7	*Europeana Sounds Technical Review Report (year 2)* https://basecamp.com/1936492/projects/2141070/messages/56888907?enlarge=227224009#attachment_227224009 (available to the project consortium only)
Ref 8	*MS7 End-user contributions defined* http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Milestones/EuropeanaSounds-MS7-End-user-contributions-defined.pdf

Appendix A: Terminology

A project glossary is provided at: http://pro.europeana.eu/web/guest/glossary. Additional terms are defined below:

Term	Definition
API	Application Programming Interface
APEX	Archives Portal Europe network of excellence
EC-GA	Grant Agreement (including Annex I, the Description of Work) signed with the European Commission
JSON-LD	JavaScript Object Notation for Linked Data
MIMO	Musical Instruments Museums Online
PI	Performance Indicator
PMB	Project Management Board
WP	Work Package