



EUROPEANA SOUNDS

Project Number: 620591

D2.1 Crowdsourcing infrastructure and exchange policy

Document Identifier: EuropeanaSounds-D2.1-Crowdsourcing-infrastructure-and-exchange-policy-v1.0.docx

Document link: <http://pro.europeana.eu/web/europeana-sounds/documents>

Date: 03/11/2014

Abstract

This deliverable defines the requirements for enrichment through crowdsourcing in the context of Europeana Sounds. These requirements are based on (1), desk research and literature study; (2), a co-design workshop; and (3), a survey sent out to all the data providers. Enrichment through crowdsourcing will be designed in the form of 'micro-tasks' for two types of audiences; the general public (1) and experts (2). This distinction has been further developed into 'culture snackers' and 'culture vultures'. This document also provides general information on the development plans for the different platforms involved in the crowdsourcing infrastructure. This encompasses the Europeana Core Infrastructure, Europeana Channels, Historypin, Pundit and the Europeana Sounds Widget.

Co-funded by the European Union
Europeana Sounds is coordinated by the British Library



The project is co-funded by the European Union, through the **ICT Policy Support Programme** as part of the **Competitiveness and Innovation Framework Programme (CIP)**.
http://ec.europa.eu/information_society/activities/ict_psp/



| Dissemination level | | |
|---------------------|--|---|
| P | Public | X |
| C | Confidential, only for the members of the Consortium and Commission Services | |
| I | Internal, only for the members of the Consortium | |

I. COPYRIGHT NOTICE

Copyright © Members of the Europeana Sounds Consortium, 2014-2017. This work is licensed under the Creative Commons CC-BY License. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. The work must be attributed by attaching the following reference to the copied elements: “CC-BY Members of the Europeana Sounds Consortium, 2014 <https://creativecommons.org/licenses/by/4.0/>”. Using this document in a way and/or for purposes not foreseen in the license requires the prior written permission of the copyright holders. The information contained in this document represents the views of the copyright holders as of the date such views are published.

II. REVISIONS

| Version | Status | Author | Partner | Date | Changes |
|---------|--------|--|---------|------------|---|
| 0.1 | ToC | Maarten Brinkerink | NISV | 16/6/2014 | |
| 0.2 | Draft | Harry van Biessum | NISV | 27/6/2014 | Draft chapters 1.1., 1.2., 2.1., 2.2. |
| 0.3 | Draft | Sergiu Gordea | AIT | 1/7/2014 | Draft chapter 1.1.1. |
| 0.4 | Draft | Harry van Biessum | NISV | 2/7/2014 | Draft chapters 1.1., 1.2., 1.2.2. |
| 0.5 | Draft | Maarten Brinkerink | NISV | 7/7/2014 | Draft introduction, chapters 1.1.1., 2.2., 3.3., |
| 0.6 | Draft | Maarten Brinkerink | NISV | 25/7/2014 | Changes based on feedback from the project partners |
| 0.7 | Draft | Maarten Brinkerink | NISV | 22/10/2014 | Changes based on feedback from the project partners / reviewers |
| 1.0 | Final | Maarten Brinkerink and Harry van Biessum | NISV | 03/11/2014 | Final version |

III. DELIVERY SLIP

| | Name | Partner/WP | Date |
|------------------------|--|--|------------|
| Document Author | Maarten Brinkerink mbrinkerink@beeldengeluid.nl | NISV / WP2 | 02/11/2014 |
| Reviewed by | Reviewers: Lisette Kalshoven John Ashley Burgoyne | KL / WP2 University of Amsterdam | 27/10/2014 |
| Approved by | Coordinator & PMB | | 03/11/2014 |

IV. DISTRIBUTION

| No. | Date | Comment | Partner / WP |
|-----|------------|---------------------------------------|--------------|
| 1 | 03/11/2014 | Submitted to the Europeana Commission | BL/WP7 |
| 2 | 03/11/2014 | Posted on Europeana Pro | BL/WP7 |
| 3 | 03/11/2014 | Distributed to the Project Consortium | BL/WP7 |

V. 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.

VI. DOCUMENT AMENDMENT PROCEDURE

Amendments, comments and suggestions should be sent to the authors named in the Delivery Slip.

VII. TERMINOLOGY

A complete project glossary is provided at the following page:

<http://pro.europeana.eu/web/guest/glossary>

Further terms are defined below as required:

| TERM | DEFINITION |
|-------|--|
| AB | Advisory Board |
| APEX | Archives Portal Europe network of excellence |
| EC-GA | Grant Agreement (including Annex I, the Description of Work) signed with the European Commission |
| GA | General Assembly |
| PC | Project Coordinator |
| PI | Performance Indicator |
| PM | Project Manager |
| PMB | Project Management Board |
| PSO | Project Support Officer |
| TEL | The European Library |
| TD | Technical Director |
| UAP | User Advisory Panel |
| WP | Work Package |

VIII. 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, Spotify, 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>.

IX. 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.

X. EXECUTIVE SUMMARY: D2.1 CROWDSOURCING INFRASTRUCTURE AND EXCHANGE POLICY

This deliverable defines the requirements for enrichment through crowdsourcing in the context of Europeana Sounds. These requirements are based on the following:

1. desk research and literature study
2. a co-design workshop
3. a survey sent out to all the data providers.

Enrichment through crowdsourcing will be designed in the form of 'micro-tasks' for two types of audiences: the general public (1) and experts (2). This distinction has been further developed into 'culture snackers' and 'culture vultures'.

The goals of the literature study were as follows:

1. to help shape the crowdsourcing classifications for the project
2. to investigate the dominant ideas concerning the motivation of contributors of crowdsourcing projects
3. to find notions about user types in the GLAM (Galleries, Libraries, Archives and Museums) domain
4. to spot possible problems with incorporating different functionalities in advance

In addition, the literature study also pointed out a possible direction concerning the strategy to reach the main target user type of Europeana Sounds, namely through nichesourcing. The desk research was executed simultaneously and complemented the literature study by giving practical examples of crowdsourcing projects and their functionalities and approaches.

The data providers gave good insights into their needs for and ideas about crowdsourcing enrichments by contributing to the survey 'Data Providers Survey on Enrichments through Crowdsourcing'. According to these results, most data providers would like to see (1) more linkage of their rich described items and unsurprisingly (2) more added general information on the poorly described items in their collection(s). One of the results from the survey in terms of current and desirable communities is that many data providers focus on researchers and students - who can be considered experts (or culture vultures) in the audience dichotomy this work package employs.

In addition, this deliverable contains the requirements for designing crowdsourcing functionalities based on the knowledge so far. The user stories are based on the concepts of culture snackers and culture vultures and are displayed according to our initial prioritisation, based on the knowledge so far. Given that Europeana Sounds will establish micro-tasks which will generate user contributions, it is pertinent to clarify how these enrichments and content are exchanged between the (users of the) different Europeana Sounds platforms, Europeana (both the portal and the channels) and original content/data providers. There are several aspects of this to consider. One is the legal aspect, and the second is how the quality of user contributions will be assured. The legal aspect of dealing with the

user contributions will be governed according to the Europeana Terms for User Contributions. With the number of user contributions Europeana Sounds aims to gather, it is not feasible to manually assure the quality of these contributions. Alternatively the crowdsourcing infrastructure needs to rely on other mechanisms for quality assurance, such as constant improvement of metadata quality, through (crowdsourced) feedback loops, recording provenance and letting data providers decide themselves to utilise annotations (or not).

Finally, this document provides general information on the development plans for the different platforms involved in the crowdsourcing infrastructure. This encompasses the Europeana Core Infrastructure, Europeana Channels, Historypin, Pundit and the Europeana Sounds Widget. A more elaborate technical design will be described in D2.2 *Functional design of semantic enrichment*. MS8 *Functional design of semantic enrichment* provides a first draft of the system architecture and describes how all the different platforms are connected to the core infrastructure and how the data flows between them.

TABLE OF CONTENTS

| | | |
|----------|---|-----------|
| 1 | INTRODUCTION | 9 |
| 2 | STATE OF THE ART | 11 |
| 2.1 | Literature Study | 11 |
| 2.1.1 | Classification | 11 |
| 2.1.2 | Motivation..... | 15 |
| 2.1.3 | Engagement | 17 |
| 2.1.4 | User types | 17 |
| 2.1.5 | Functionalities..... | 17 |
| 2.1.6 | Nichesourcing..... | 19 |
| 2.2 | Desk research..... | 20 |
| 2.2.1 | Correction and Transcription Tasks | 20 |
| 2.2.2 | Contextualisation | 21 |
| 2.2.3 | Classification | 21 |
| 2.2.4 | Co-curation..... | 22 |
| 2.2.5 | Motivation..... | 22 |
| 2.3 | Relevant knowledge sources | 23 |
| 3 | DATA PROVIDERS | 26 |
| 3.1 | Information needs..... | 26 |
| 3.2 | Audience engagement..... | 27 |
| 4 | REQUIREMENTS..... | 29 |
| 4.1 | List of user stories | 29 |
| 5 | EXCHANGE POLICIES | 35 |
| 5.1 | Legal conditions | 35 |
| 5.2 | Quality assurance..... | 35 |
| 6 | PLANS FOR TECHNICAL IMPLEMENTATION | 37 |
| 6.1 | Core Infrastructure..... | 37 |
| 6.2 | Europeana Channels..... | 37 |
| 6.3 | Historypin | 38 |
| 6.3.1 | Historypin audiences..... | 38 |
| 6.3.2 | Historypin content | 38 |
| 6.3.3 | Other platforms that we hope to link to..... | 39 |
| 6.4 | Pundit..... | 39 |
| 6.5 | Widget..... | 39 |
| 7 | ANNEX 1: PERSONAS | 41 |
| 7.1 | Daan | 41 |
| 7.2 | Jacob | 42 |
| 7.3 | Jasmine..... | 43 |
| 8 | ANNEX 2: DATA PROVIDERS SURVEY | 44 |
| 9 | REFERENCES | 49 |

1 INTRODUCTION

Work Package 2 (WP2) of Europeana Sounds aims to support the discovery and use of the audio and audio-related metadata in Europeana, by improving metadata through “innovative methods including semantic enrichment and crowdsourcing”. In order to accommodate this, WP2 will “Design and implement mechanisms to improve the quality of existing metadata and contextual information” [REF 1, p. 10].

This Deliverable defines the requirements for enrichment through crowdsourcing in the context of Europeana Sounds. These requirements are based on the following:

1. desk research focused on existing literature within the field and the gathering and evaluation of best-practices in the cultural heritage sector
2. a co-design workshop with work package members and a member of the User Advisory Panel present
3. a survey sent out to all the data providers.

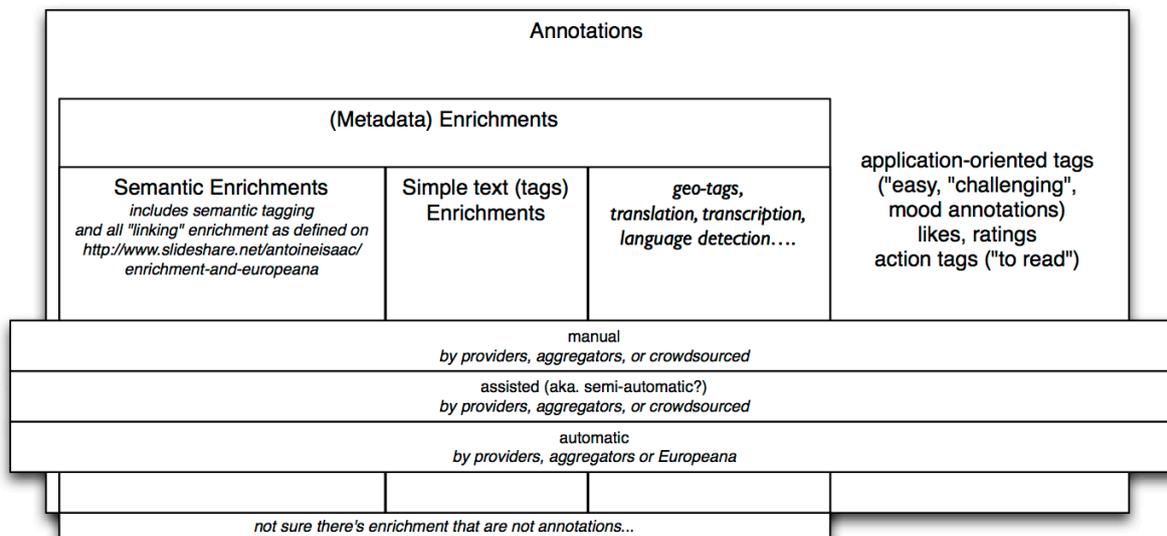
Further requirements relating to the enrichment of specific collections can best be developed after a programme of consultation with the end-users of such enriched data; these explorations will be carried out in the context of the technical product development supporting enrichment, under the general categories and types of enrichment defined in this document.

Enrichment through crowdsourcing, as supported by WP2, will be designed in the form of “micro-tasks” [REF 1, p. 10] for two types of audiences: the general public (1) and experts (2). As described in *MS7 End-user contribution defined*, this distinction has been further developed, according to Chenchen Shen’s paper “Design for User Engagement on Europeana Channels” into ‘culture snackers’ and ‘culture vultures’ [REF 4].

To summarise, this means that culture snackers are perceived as users with a casual interest in Europeana or the type of content it serves. Culture vultures are perceived as users who work professionally with the Europeana portal, or are culture enthusiasts for the type of content it serves. In the context of Europeana Sounds special attention will go out to the users - of both kinds - with a specific interest in audio and/or music. It’s important to realise that culture snackers and culture vultures are not static or mutually exclusive roles. Depending on context a single user is sometimes a snacker and sometimes a vulture.

This document focuses on the requirements for enrichment through crowdsourcing. The figure below depicts different types of annotations within the technical infrastructure. The scope of D2.1 includes application-oriented tags and (metadata) enrichments of a manual or assisted nature. The latter - assisted (metadata) enrichments - is also dealt with in the specific case of the improvement of existing enrichments, which are the result of the automatic enrichment processes, and are also part of the figure below.

Figure 1: Different types of annotations within the technical infrastructure



NB: this doesn't take into account the target of annotation, in Open Annotation parlance: cultural object itself, image, image area, A/V selection, text, metadata values, other annotations (e.g. enrichments to be improved by refinement or correction. This would need another table!

2 STATE OF THE ART

This chapter contains the findings from the literature study and the desk research. The goals of the literature study were as follows:

1. to help shape the crowdsourcing classifications for the project
2. to investigate the dominant ideas concerning the motivation of contributors of crowdsourcing projects
3. to find notions about user types in the GLAM (Galleries, Libraries, Archives and Museums) domain
4. to spot possible problems with incorporating different functionalities in advance

In addition, the literature study also pointed out a possible direction concerning the strategy to reach the main target user type of Europeana Sounds, namely through nichesourcing. The desk research was executed simultaneously and complemented the literature study by giving practical examples of crowdsourcing projects and their functionalities and approaches.

2.1 Literature Study

2.1.1 Classification

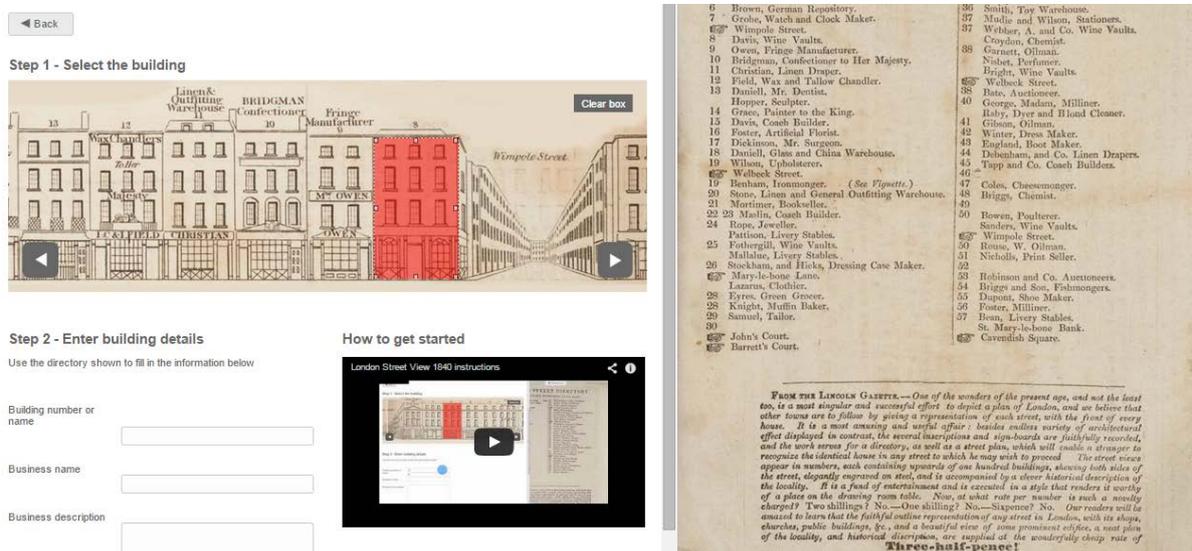
Oomen and Aroyo created six types of crowdsourcing classifications which encompass working practices that are relevant for GLAM domains [REF 2]. Below, these classifications will be explained in relation to types of contributions defined in Milestone MS7 (user-generated metadata, subjective contributions and user profile related interactions) and annotation types (manual, assisted and application oriented enrichments) [REF 2].

Correction and Transcription Tasks “[invite] users to correct and/or transcribe outputs of digitisation processes” [REF 2, p. 140]. In many cases, the Correction and Transcription Tasks can be classified as both assisted enrichments (semi-automatic), and as manual enrichments at the same time, and provide user generated metadata contributions that are not necessarily subjective. The assisted part can be found in crowdsourcing projects that provide the specific selections of the digitised objects next to the relevant place where the metadata enrichments have to take place.

For instance, within the project [London Street Views 1840¹](http://crowd.museumoflondon.org.uk/lsv1840/), users are invited to transcribe scanned pamphlets about London’s street history (including a business directory). Users are able to mark a certain house in the scanned pamphlet and manually enrich it with relevant information from the scanned business directory which is automatically provided on the right side of the screen.

¹ <http://crowd.museumoflondon.org.uk/lsv1840/>

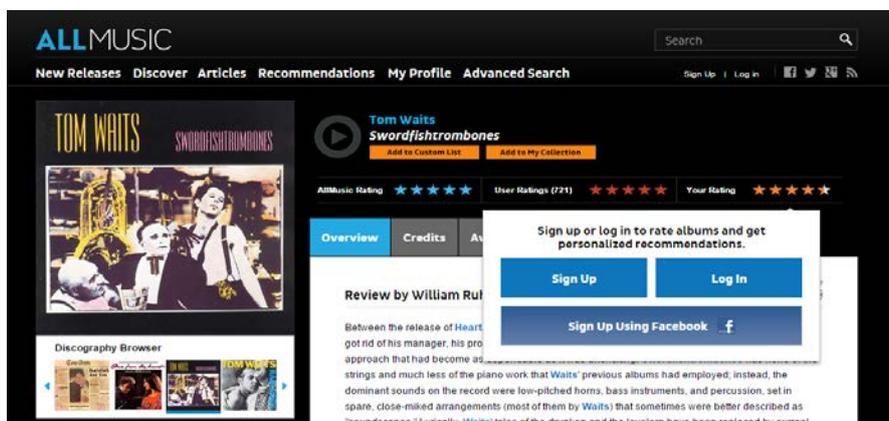
Figure 2: Example of manual enrichment possibilities that incorporated an assisted component as well (London Street Views 1840)



The screenshot displays the 'London Street View 1840' interface. It features a street view on the left with a building highlighted in red. A 'Clear box' button is visible in the top right of the view. To the right is a directory list with numbered entries and names. Below the street view is a 'Step 2 - Enter building details' form with fields for 'Building number or name', 'Business name', and 'Business description'. A 'How to get started' video player is also present. At the bottom right, there is a historical text snippet from 'The Lincoln Gazette'.

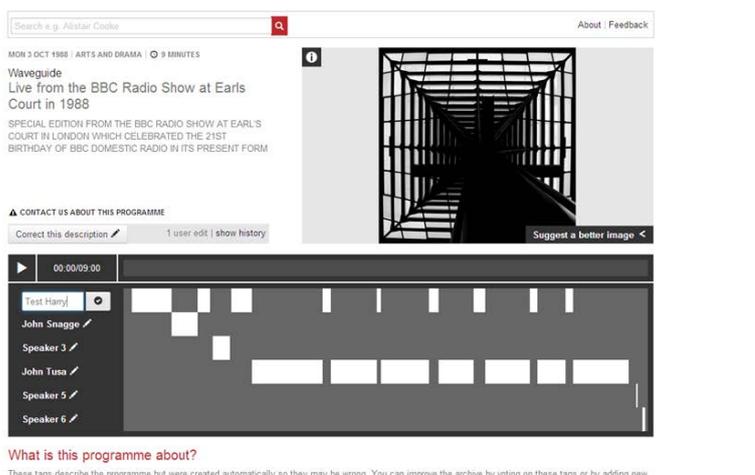
The Contextualisation of an object will be performed by end-users ‘adding contextual knowledge to objects, e.g. by telling stories or writing articles/wiki pages with contextual data’ [REF 2 p. 140]. Contextualisation emerges in various forms, as an objective or a subjective enrichment for instance. The objective contextualisation can be executed by the linking of a media object (image, video, sound) to a related lemma on a platform such as Wikipedia. This reusing of media objects fits the broader concept of objective contextualisation, which puts information gained from an existing (trusted) source into another relevant context. The subjective contextualisation of an object concerns all kinds of manual enrichments by end-users, whether it is a comment or a personal story about an object. The subjective contextualisation can also appear in the form of application oriented enrichments, such as rating a song, whereby user profile related interactions are most common.

Figure 3: Example of subjective contextualisation encompassing user profile related interaction (Allmusic.com)



The task of 'gathering descriptive metadata related to objects in a collection' is the fourth category of the authors and is labelled as Classification [REF 2 p. 140]. This type of crowdsourcing overlaps, to a certain degree, with the objective variant of Contextualisation, because many crowdsourcing projects deploy the use of controlled vocabulary. For example, the World Service Radio Archive project uses terms from Wikipedia in a controlled field (based on auto-completion), when giving end-users the possibility to add tags to their radio broadcasts. When using a free text field for this tagging functionality, the Contextualisation becomes more subjective, as the tag has no explicit link to another trusted source of information; instead it depends entirely on the faculties of the end-user.

Within the task of Classification there are various ways end-users can execute this task: manually by adding tags (in free or controlled fields) (1), or assisted whereby an automatic process had occurred beforehand (2). An example is the task provided by the above-mentioned World Service Radio Archive. The assisted part comes from the automatic speech recognition (ASR), which is performed before presenting the yet unknown speakers in a radio broadcast to the potential contributors [REF 12]. In this way, end-users can add the name of a speaker just once, without the need to enter all the time-codes in the audio-file as and when the speaker appears again.

Figure 4: Example of assisted enrichments (World Service Radio Archive)


Another relevant crowdsourcing category that is relevant to investigate further, is the task of Co-curation, which is defined as “using [the] inspiration/expertise of non-professional curators to create (Web) exhibits.” [REF 2 p. 140] The user research undertaken by Chenchen Shen provides many indicators that this functionality is quite popular among current Europeana users [REF 4]. The task of Co-curation involves user profile related interactions, in most cases, for example, the creation of exhibits by users on the website of the International Movie Database. In this example, the functionality to create a curated Web exhibition is combined with the possibility to add subjective contextualisation to the different objects within the exhibition. The user that created the exhibition displayed in Figure 5 enriched this self-made collection with texts about the collection in general (“people whose films I shall always watch if I want”) and the objects within the exhibition in particular (this user for instance watches Wes Anderson films “to get inspiration”). Objective contextualisation happens in this example without the explicit help of the end-users; the subsequent Web exhibitions are automatically linked with the specific digital objects in the users created collections.

Figure 5: Example of co-curation with the implementation of subjective contextualisation (imdb.com)



The classification of Complementing Collections is described by Oomen and Aroyo as the “active pursuit of additional objects to be included in a (web) exhibit or collection.” [REF 2 p. 140] Complementing collections in this sense is out of scope of this Work Package, and for the Europeana Sounds project as described in the Description of Work, since it involves the pursuit of user-generated content. Furthermore, the concept of Crowdfunding, defined as the “collective cooperation of people who pool their money and other resources together to support efforts initiated by others”, is also not considered as a target for this Work Package, since it has no relation to enrichment of metadata [REF 2 p. 140].

2.1.2 Motivation

An extensive amount of literature has been written which focuses on the reasons why people might choose to contribute to crowdsourcing projects. Below, the main motivations for end-users to contribute to crowdsourcing are summarised. Oomen and Aroyo claim that users who engage in crowdsourcing activities will do so for either extrinsic or intrinsic motivations. An example of extrinsic motivation is payment. Oomen and Aroyo give the example of Amazon’s Mechanical Turk, a platform in which users perform simple tasks such as creating subtitles, categorising websites and so on [REF 2].

The GLAM crowdsourcing initiatives Oomen and Aroyo discuss in their paper all focus on intrinsic motivations, as do the websites discussed in the previous section. This is because within these initiatives both GLAMs and their users benefit from mutual recognition [REF 2] and will foster a more profound type of engagement, as Oomen and Aroyo paraphrase Leadbeater [REF 5]. Oomen and

Aroyo cite Shirky in defining the merits of intrinsic motivation: “amateurs are sometimes separated from professionals by skill, but always by motivation; the term itself derives from the Latin amare - to love. The essence of amateurism is intrinsic motivation: to be an amateur is to do something for the love of it” [REF 6]. Shirky is cited again when he discusses the work of Benkler and Nisembaum about social motivations: “they divide social motivations into two broad clusters - one around connectedness or membership and the other around sharing and generosity.”[REF 6]

When analysing the crowdsourcing projects in their paper, Oomen and Aroyo organise these projects according to these two clusters of motivations. Other important incentives for users to participate are cited from a study by Oomen (et al.): “altruism, fun and competition.” [REF 7] Oomen and Aroyo claim that “the creation of a strong sense of belonging to an altruistic community, and making explicit the mutual benefits of the contributed tags, attracts users with diligence and ethical behaviour.” [REF 2] Furthermore, the authors claim that a combination of technological and interaction aids, psychology principles and community building rules can help to establish behavioural norms, build an image of the desired quality of content and filter or correct erroneous information.

Oomen and Aroyo provide the example of [Waisda?](#)² to explain that the community itself can also act as a filter (in the sense of validating user-generated metadata). Waisda? checks if there is mutual agreement between players on a chosen term during the game, before it is considered for inclusion in the archive [REF 2]. In section 5.2 of this document, other ways in which the community can act as a filter will be explored as well.

Mia Ridge elaborated further on what the aforementioned intrinsic motivation encompass [REF 8]. Ridge discovered that the intrinsic motivations for participating in museum crowdsourcing projects include: “fun, pleasure in doing hobbies, enjoyment in learning, mastering new skills, practicing existing skills, recognition, community, passion for the subject.” [REF 8 p. 441] In addition, Ridge notes that motivations can be grouped into “altruistic, intrinsic and extrinsic.” [REF 8 p. 441] Furthermore, Ridge draws a distinction between the motivation of contributors involving commercial on one hand, and non-commercial (cultural heritage) crowdsourcing projects on the other: “unlike commercial crowdsourcing, participation in cultural heritage is driven by pleasure, not profit. Rather than monetary recompense, GLAM projects provide an opportunity for altruistic acts, activated by intrinsic motivations, applied to inherently engaging tasks, encouraged by a personal interest in the subject or task.” [REF 8 p. 438]

In these studies, the intrinsic and altruistic aspect of contributors’ motivation were marked as the main reasons for users to contribute to crowdsourcing projects and were specifically mentioned in projects helping cultural heritage organisations with enriching their metadata. Therefore, in designing the crowdsourcing tasks for Europeana Sounds, WP2 will focus primarily on these two aspects of users’ motivation.

² <http://woordentikkertje.manbijthond.nl/>

2.1.3 Engagement

In addition to these findings regarding contributors' motivations, the literature study also provided a relevant model for user engagement. According to Ridge, this model, described below, applies equally well to both physical and online participations/visitors. Although the model focuses on crowdsourcing games, it is relevant to a more general user engagement perspective: "deep engagement can be its own reward. 'Flow' is experienced as a state of deep, enjoyable focus or engagement." [REF 8] Supporting flow through content and interaction design helps keep players engaged with an activity, and therefore helps crowdsourcing projects be more productive. [REF 8 p. 443]

This *flow* depends on three requirements:

1. clear goal
2. immediate feedback on the success of your attempts to reach that goal
3. good match between the skills of the individual and the challenges faced

2.1.4 User types

Ridge defines different user types in her study, and although they are not specifically described as either culture vultures or culture snackers, Ridge's user types still provide a useful insight into the target groups of the Europeana Sounds project. The culture vulture type, for instance, shares traits with what Ridge labelled as digital volunteers: "those who intentionally participate in crowdsourced tasks for the intrinsic rewards. Include people who are passionate about the relevant subject, people who like doing the task offered in the project or the source material used, and people who are unable to volunteer in venue opening hours or locations." [REF 8 p. 438] Furthermore the term 'proams' (professional amateurs) is cited by Ridge as referring to "self-motivated, enthusiastic, and dedicated amateurs who work to professional standards." [REF 8] This fits the concept of the culture vulture type as well, and thereby exemplifies that culture vultures are not necessarily restricted to the domain of professionals.

A further examples example of a user type that is in line with the culture vultures' concept is the group of people labelled as 'experts' by Ridge and are in this specific case "trained historians who work in other fields while enjoying hobbies that let them keep up their historical research skills on the job; and self-taught researchers with decades of experience as practicing historians." [REF 8 pp. 438-439] Opposing these descriptions, Ridge classified a group of participants that can fit both target groups of Europeana Sounds: "those whose contributions are a side effect of their participation in other core activities." [REF 8 p. 438]

2.1.5 Functionalities

The literature review also presented multiple sets of functionalities as well as corresponding insights, and possible problems, related to these functionalities. In order to anticipate possible problems that

the different functionalities could bring into this project, it is useful to consider a list of possible tagging system problems, as pointed out by Haslhofer, Robitza, Guimbretiere and Lagoze [REF 9]:

- Polysemy (a tag can be ambiguous and have many related meanings)
- Synonymy (multiple tags can have the same meaning)
- Semantics of a tag might range from very specific to very general because people describe resources along a continuum of specificity

For their study, Haslhofer (et al.) created Maphub, an open source web application. Maphub is an early adopter of the Open Annotation model [REF 10]. The study uses four main tag creation methods as the main varying condition:

- Label-based tagging (user enters comma separated label-based tags)
- Suggestive tagging (the system randomly selects and suggests label-based and semantic tags that were entered by the same or other users before)
- Semantic tagging (tags are generated and suggested based on the text and geographical area of an annotation)
- Semantic tagging with Context Display (tags are generated and suggested as in the semantic tagging condition, but the user is shown additional contextual information (Wikipedia page abstract) when hovering over it)

This study shows that their semantic tagging implementation does not affect tag production, the types of and categories of obtained tags, or user task load, while providing tagging relationships to well-defined concept definitions. Compared to label-based tagging, their technique also gathers positive and negative tagging relationships, which can be useful when improving tag recommendation and resource retrieval [REF 9]. Furthermore, Marlow (et al.) created a classification in the different ways a project can organise their tagging options, distinguished in three categories, cited by Haslhofer (et al.):

- Blind tagging (user cannot view tags assigned to the same resource by other users)
- Viewable tagging (where users sees tags associated with a resource)
- Suggestive tagging (where the system suggests possible tags from the same or other users or from a resource's context to the user)

Haslhofer (et al.) follow Sen (et al.) for classifying the different types of tags as either factual or personal tags. Sigurbjörnsson (et al.) provide Haslhofer (et al.) with the classification of semantic categories: locations, artefacts or objects, people or groups, actions or events and time. According to Haslhofer (et al.), "contextual information [...] can only be determined after reconciling label-based tags with data entries in other data sources." [REF 10] The authors of the article also define a possible solution for this problem as being: "mapping label-based tags to concepts defined in knowledge contexts." [REF 9] One of the projects studied in the desk research (World Service Radio Archive) uses this solution in the form of utilising Wikipedia as the knowledge source for their

semantic tagging. Section 1.2.2 of this document provides further examples of knowledge sources within the scope of the Europeana Sounds project.

2.1.6 Nichesourcing

The concept of nichesourcing provides an example of how Europeana Sounds can reach its primary target user type – namely the culture vulture. De Boer (et al.) cite a study by Schroer and Hertel on voluntary engagement within the Wikipedia community:

“Nichesourcing is a specific type of crowdsourcing where complex tasks are distributed amongst a small crowd of amateur experts [...] rather than the ‘faceless’ crowd. A niche is gathered from either distributed experts on a specific topic or from an existing network centered around the same culture, location or topic. In both cases the members have domain knowledge and an intrinsic motivation to contribute and provide high quality results.”

[REF 11 p. 17]

This strategy of nichesourcing is suitable for culture vultures, since this type of user is to be considered an expert in a specific field. De Boer (et al.) make comparisons between main aspects of crowdsourcing and nichesourcing. One aspect that is relevant for this WP is the task and its complexity:

“Crowdsourcing deals with large complex tasks, by dividing it up into smaller atomic tasks, the latter which do not require specific knowledge or skills from the crowd members. [...] Atomic tasks, that require a specific level of knowledge or effort, can be outsourced to niches in which the members possess particular knowledge or motivation is present.”

[REF 11 p. 17]

Another interesting aspect lies in the resource pool of users needed to realise a task, an aspect on which Historypin has expertise to bring into the Europeana Sounds project:

“Building and maintaining a dedicated crowd is essential to crowdsourcing applications that use altruism, reputation and (to a lesser extend) enjoyment as motivation. [...] In nichesourcing, composite and complex tasks are distributed within existing communities. [...] Although communities, in contrast to crowds, provide smaller pools to draw resources from, their specific richness in skill is suited for the complex tasks with high-quality product expectations found in nichesourcing. Moreover, the peer resources receptive to complex tasks may exploit their own social trust relations to transitively trigger other communities that may offer reinforcing resource pools.”

[REF 11 p. 18]

These findings on nichesourcing are in line with the strategy of designing micro-tasks as described in MS7 [REF 3] and the DoW [REF 1 p. 10], in that they explicitly mention the breaking down of large

complex tasks into smaller atomic tasks. In some cases, these smaller tasks require a specific level of domain knowledge which is embodied in the culture vulture target group as described in MS7. [REF 3, pp. 7-8]

2.2 Desk research

In this phase of the project we compared different crowdsourcing projects that enrich cultural heritage objects organised by organisations inside and outside Europe. Since the material that the Europeana Sounds project will aggregate contains more than just audio (images, sheet music, etc.), the focus is not just on projects that aim to enrich audio, but the enrichment of images, texts and videos was taken into account as well.

Due to this broad scope many different types of crowdsourced enrichments were considered during the desk research. In addition to the different types according to the classification of crowdsourcing initiatives by Oomen and Aroyo [REF 2] we have:

- Correction and Transcription Tasks: [Transcribe Bentham](#)³, [London Street Views 1840](#),
- Contextualisation: [Historypin](#)⁴, [Amateurfilm Platform](#)⁵, [Citizen Archivist Dashboard](#)⁶
- Classification: [World Service Radio Archive](#)⁷, [Waisda?](#), [Vele Handen](#)⁸
- Co-curation: [International Movie Database](#)⁹, [Spotify](#)¹⁰

As mentioned earlier, Co-curation, Complementing Collection and Crowdfunding are not considered, since this work package is essentially about enrichments and these categories are therefore out of scope.

2.2.1 Correction and Transcription Tasks

In the words of Oomen and Aroyo, several of the websites we have considered so far are concerned with “inviting users to correct and/or transcribe outputs of digitisation processes.” [REF 2] However, the websites are each inviting the users in slightly different ways. For example, on the website London Street Views 1840, the user can add information about London buildings in the past, for each individual building (house number). The tool provides the user with the scanned texts where all the necessary information is already available, except for the business type (for this a controlled drop down list is provided). In this way, the website forces the contributor to not only transcribe the scanned texts, but also to classify them according to the website creators’ classification system.

³ <http://blogs.ucl.ac.uk/transcribe-bentham/>

⁴ <http://www.historypin.com/>

⁵ <http://www.amateurfilmplatform.nl/>

⁶ <http://www.archives.gov/citizen-archivist/>

⁷ <http://worldservice.prototyping.bbc.co.uk/>

⁸ <http://velehanden.nl/>

⁹ <http://www.imdb.com/>

¹⁰ <https://www.spotify.com/nl/>

2.2.2 Contextualisation

Some crowdsourcing projects allow the crowd to add contextual knowledge to objects, by letting them tell their stories or write articles and / or Wikipedia pages with contextual data [REF 2]. On [Historypin](#), a user is given the opportunity to provide comments and suggestions to the various geo located photographs. Like most of these crowdsourcing initiatives with 'Contextualisation' functionality, the comments are displayed in a thread format. The [Amateurfilm Platform](#) is another example of a website that displays their comments this way. However, on Historypin you need to be registered as a contributor, which is not the case for contributing comments to the Amateurfilm Platform. An advantage of having registered contributors is the additional context the other visitors see from the displayed name of the person or organisation who contributed the comments. In the case of Historypin, an extra advantage is that the names of the contributors are hyperlinked to their profile, which allows other users to keep track of their activities (by becoming a fan in Historypin's terminology).

When no registration is needed however, this lowers the threshold for users to add contextual knowledge and this can create more activity in the threads. The [Amplifon Sounds of Street View](#) is an open source platform making use of the Web Audio API and Google Map's Street View feature to allow users to embed sounds from that certain area (instead of using markers to tag Google Street View with secondary images), giving visitors to the page a sonic and visual interaction with a specific location at street level, as another example of Contextualisation.

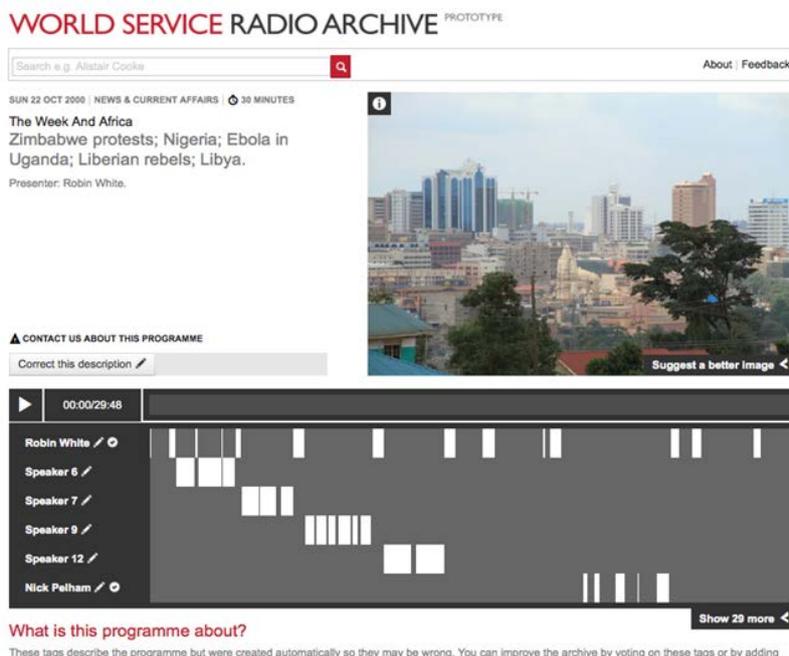
2.2.3 Classification

The gathering of descriptive metadata related to objects in a collection through crowdsourcing is labelled by Oomen and Aroyo as 'Classification' [REF 2]. The [World Service Radio Archive](#) is a good example of a way to gather descriptive metadata for radio programmes. It is possible for the user to contribute new topics, add the names of speakers, correct descriptions and add images to the radio programmes. The new topics the user can add to the radio programme are controlled ([Wikipedia](#)¹¹/[DBPedia](#)¹²). The names the user can provide to the unknown speakers in the radio programmes are not controlled, but they do interlink with the names in the other radio programmes on the website. Plus their system has already divided the audio into segments in which the same speaker appears again. In this way a contributor only needs to add one name per speaker in the total audio item and does not have to find out on what timecodes a speaker appears again, since this is automatically generated for them.

¹¹ http://en.wikipedia.org/wiki/Main_Page

¹² <http://dbpedia.org/About>

Figure 6: The different speakers divided into segments by the BBC's World Radio Service Archive



It proved to be a very successful crowdsourcing initiative. Within ten months they received “more than 40,000 positive and negative votes against automatically generated topics, covering around 6,000 distinct programmes. Around 10,000 new topics were added [...] to around 3,000 distinct programmes’ from 2,000 registered users.” [REF 12]

2.2.4 Co-curation

Another crowdsourcing category is the task of Co-curation, which is described as “using inspiration/expertise of non-professional curators to create (Web) exhibits.” [REF 2 p. 140] An example of this would be the creation of exhibits by users on the [International Movie Database](#). The creation of playlists of favourite songs by end-users on [Spotify](#) is an example of co-curation in the scope of the Europeana Sounds project, whereas the idea to enrich such a playlist with videos or images of birds or sheet music could also be an idea to look into, in order to make use of the full extent of resources available in Europeana.

2.2.5 Motivation

Having considered a number of inspiring types of crowdsourced enrichments, it is clear there are several successful methods for approaching potential contributors. For example, the transparency of a website such as the [World Service Radio Archive](#) is a good example of how one can accommodate the participants’ motivations encompassing sharing and generosity, a motivational factor named by Benkler and Nisembaum, and cited by Oomen and Aroyo [REF 2]. Displaying the tagline ‘Help tag this collection of BBC radio programmes from the past 45 years’ on their homepage makes the desired altruistic trait of the visitor clear, for they explicitly request help from the visitor.

[Waisda?](#) is an example of a project that gives the contributor a feeling of connectedness and membership, another motivational factor described by Benkler and Nisembaum [REF 2]. On the Waisda? Homepage, the user will immediately see how many other contributors (termed ‘players’ by Waisda?) are online, the actual community at that very moment. An image of the colourful building of the archive (Netherlands Institute for Sound and Vision) is displayed so that the contributors actually see and know which organisation they are helping with their contributions. When a user contributes frequently, he or she will see their own profile name on a leader board, with the other best contributors (top scorers) of the past seven days. Furthermore, Waisda? provides free tickets to the museum of the Netherlands Institute for Sound and Vision to the best contributor of each week and month, and furthers the connectedness with the archive on a physical level (by providing a very small extrinsic motivation). The project [Vele Handen](#) rewards contributors by sending them photocopies of objects from the archive which the contributors can choose themselves. This creates a feeling of connectedness with the archive and the content as well, since it gives the organisation a sense of their most popular items from the perspective of their contributors.

Beyond a general sense of altruism or loyalty to a particular public institution, users can be motivated by the development of social capital (expressed as the net present value of relationships mutually entered into) within and between affinity groups perceived as valuable to the user. For example, in the context of the Europeana Sounds project, communities of practice exist which continue the performance and preservation of culturally significant intangible heritage, for reasons of identity, patriotism, ethnic self-assertion and other motivations relating to membership within these groups. The project can recognise these intrinsic and socially-grounded purposes and support the development of these values by presenting heritage-related crowdsourcing and enrichment activities at the events and participatory gatherings of these communities. Data providers can help identify these existing communities.

2.3 Relevant knowledge sources

In order to provide proper support for the objective contextualisation task that future contributors will carry out, a list of relevant knowledge sources is compiled. The relevancy is foremostly based on the knowledge sources that are audio related, but general knowledge sources (for example [RAMEAU](#)¹³) were not excluded, considering that users not necessarily enrich metadata of audio related objects with audio related concepts only. The compiled list is extracted from [T1.2](#) (Europeana Sounds WP1).

- [Dismarc mdGenre](#)¹⁴
- [FAST \(Faceted Application of Subject Terminology\)](#)¹⁵
- [DBpedia](#)

¹³ <http://rameau.bnf.fr/>

¹⁴ <http://www.dismarc.org/info/>

¹⁵ <http://www.oclc.org/research/activities/fast.html?urlm=159754>

Library of Congress Genre/Form Terms for Library and Archival Materials

- [Freebase](#)¹⁶

Knowledge sources that encompasses subjects: RAMEAU

- [RAMEAU](#) (French and English)

Knowledge sources that encompasses mood

- Affect and mood model have been proposed, more references [here](#)¹⁷
- [The Circumplex Model of Affects](#)¹⁸ is one of those proposed in 1980

Knowledge sources that encompasses musical instruments

- [Freebase instruments](#)¹⁹
- [MIMO](#)²⁰ (multilingual)
- [Library of Congress Medium of Performance Thesaurus](#)²¹ (LCMPT)

Knowledge sources that encompasses genres

- [EBU genres vocabulary](#)²²

Knowledge sources that encompasses tales

- [Aarnt and Thompson classification](#)²³

Knowledge sources that encompasses roles

- [MARC Code List for Relators](#)²⁴
- [DISMARC Roles](#)²⁵

Knowledge sources that encompass traditional music entities

Traditional performative culture resources tend to include lists of entities that will be relevant to the contextualisation and linking of the associated archives. For example, [Traditional Tune Archive](#)²⁶,

¹⁶ https://www.freebase.com/user/gsergiu/views/music_genres

¹⁷ http://www.academia.edu/4403225/Automatic_Music_Mood_Classification_of_Hindi_Song

¹⁸ <https://www2.bc.edu/~russeljm/publications/Russell1980.pdf>

¹⁹

<https://www.freebase.com/query?autorun=1&q=%5B%7B%22id%22:null,%22name%22:null,%22type%22:%22/music/instrument%22,%22/music/instrument/family%22:%5B%5D%7D%5D>

²⁰ <http://www.mimo-international.com/MIMO/instrument-families.aspx>

²¹ <http://www.loc.gov/catdir/cpsol/medprf.html>

²² http://www.ebu.ch/metadata/ontologies/skos/ebu_ContentGenreCS.rdf

²³ http://en.wikipedia.org/wiki/Aarne%E2%80%93Thompson_classification_system

²⁴ <http://www.loc.gov/marc/relators/relacode.html>

²⁵

<http://www.dismarc.org/index.php?form=admin.thesauri&task=showThesaurus&thesaurus=dmsvAccuralMethod>

www.thesession.org²⁷, TunePal²⁸ and others will list performers, tune classifications, relevant commercial recordings, tune settings and related commentary that could make this information very relevant for crowdsourcing and semantic enrichment activities.

²⁶ <http://www.tunearch.org/wiki/TTA>

²⁷ <http://www.thesession.org/>

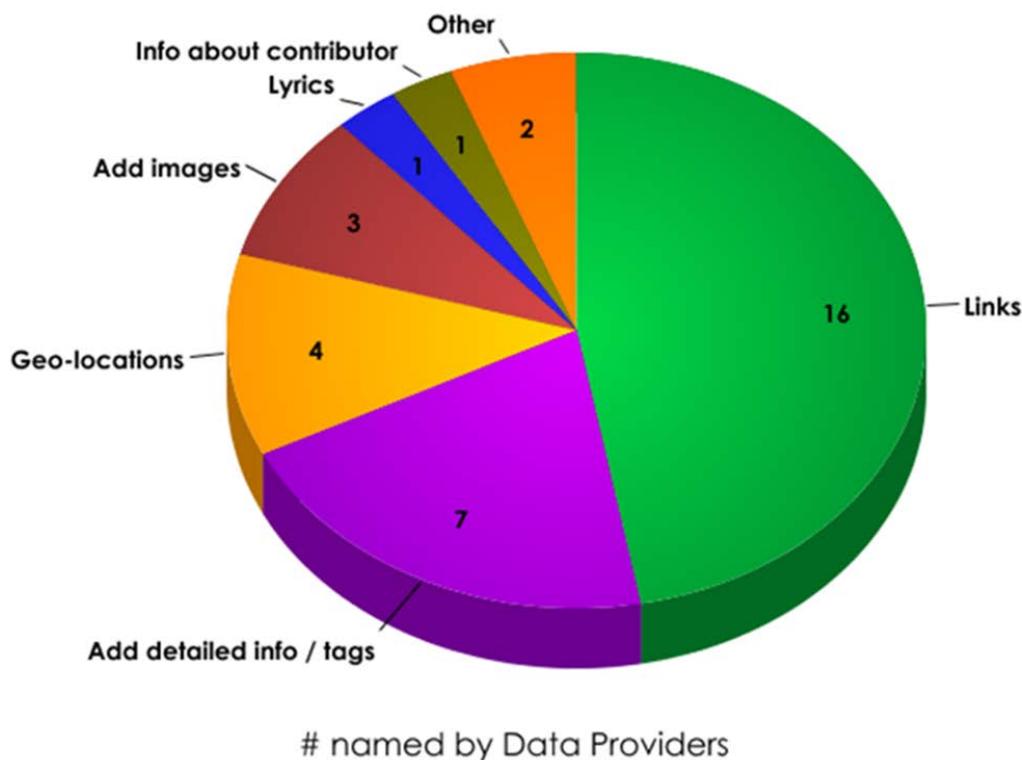
²⁸ <http://tunepal.org/tunepal/index.php>

3 DATA PROVIDERS

3.1 Information needs

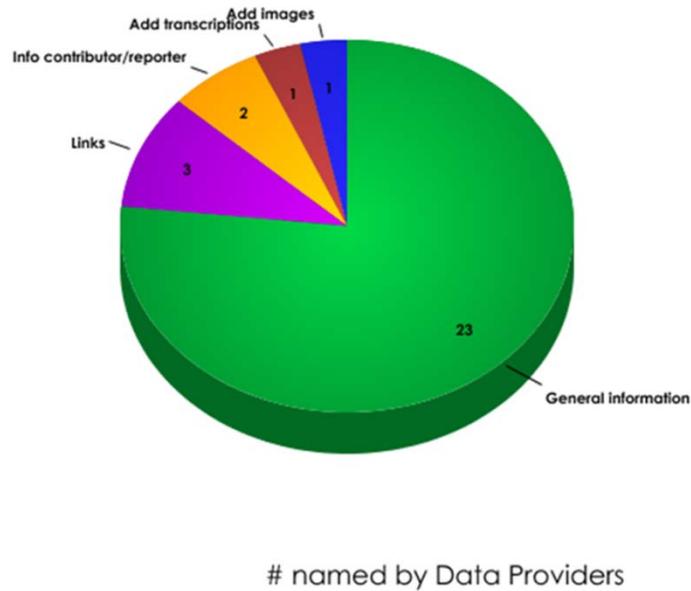
The data providers gave good insights into their needs for and ideas about crowdsourcing enrichments by contributing to the survey '[Data Providers Survey on Enrichments through Crowdsourcing](#)' issued by WP2 in May 2014 [REF 13]. One way of providing this insight was achieved by asking them to name two items from their own collection(s) in the survey, one particularly well described item (the rich item) and one item that is lacking all kinds of information (the poor item). As a follow up the data providers were asked what could make these items more interesting as a result of enrichment through crowdsourcing. Underneath the answers of the data providers are displayed:

Figure 7: Enriching an item with high metadata quality



According to these results, most data providers would like to see (a) more linkage of their rich described items and unsurprisingly (b) more added general information on the poorly described items in their collection(s). For the low metadata quality item, some examples of the missing general information named by data providers were: location, genre, name of performer, date of recording, etc. Almost half of the improvements the data providers mentioned for their rich described item consisted of linkage.

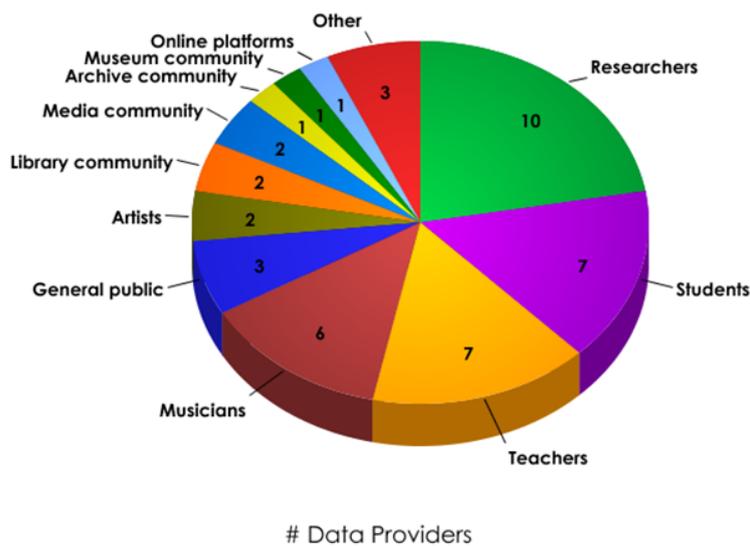
Figure 8: Enriching an item with low metadata quality



3.2 Audience engagement

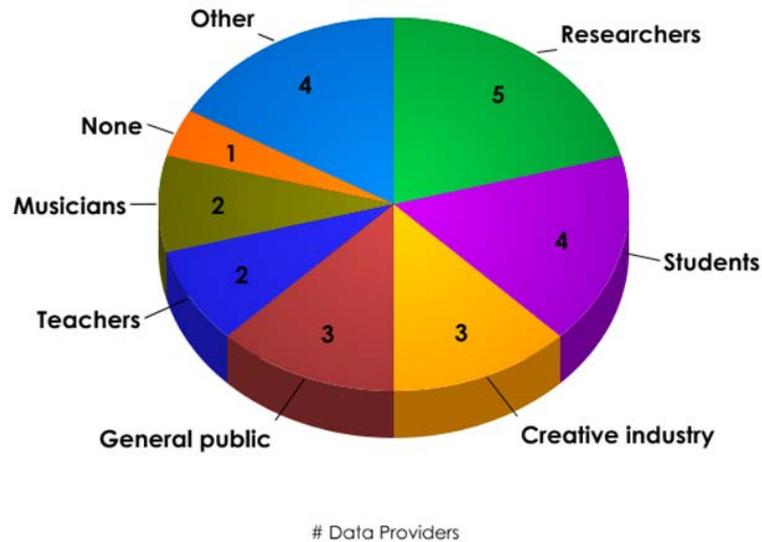
The data providers also answered questions about their engagement with communities and their experience with crowdsourcing projects in the past [REF 13]. Following on from that, they mentioned which communities they desired to engage with more.

Figure 9: Currently engaged communities



One of the results from the survey in terms of (a) current and (b) desirable communities was that many data providers focus on researchers and students - who can be considered experts in the audience dichotomy this work package employs - and that the general public (culture snackers) was named less often than expected.

Figure 10: Desirable communities



4 REQUIREMENTS

This chapter contains the requirements for designing crowdsourcing functionalities based on what is known so far. These requirements will probably be slightly altered in the future since it needs further prioritising once the technical aspects, and thus the possible limitations of these aspects, emerge later on in the project. The requirements are presented using the concept of user stories, which are based foremost on the culture vulture user type since this is the primary target group for Europeana Sounds. This is not to say that the culture snackers are not present in these user stories.

4.1 List of user stories

The user stories presented below are based on the concepts of Culture Snackers and Culture Vultures and are displayed according to our initial prioritisation based on the knowledge we possess now. Hence they are likely to change over time. The definitions of these end-user concepts are included in *MS7 End-user contributions defined*, chapters 2.1 and 2.2 respectively [REF 3]. Multiple sources were used for extracting these user stories. A great source of information was found in the user research done by Chenchen Shen [REF 4], the brainstorm session and discussions within WP2 and initial user research done by Historypin on www.thesession.org. Also some relevant user stories were extracted from the user research done by Mak and van Lieshout [REF 14]. The contexts of these different sources are fleshed out underneath the user stories.

Table 1: User Stories

| As a / an | I want to | Because | Derived from: |
|-----------------|--|--|-------------------------------------|
| culture vulture | add the date of the creation of a sound | as a culture vulture I want to order my search results chronologically | Mak and van Lieshout [REF 14 p. 22] |
| culture vulture | add a period | I want to share my knowledge | Discussion within WP2 |
| culture vulture | add locations to sounds and in return be able to find recordings on a geographical map | in order to find the sound of a specific animal I want to study for my profession and to hear the actual animal on a holiday later | WP2 brainstorm session |
| culture vulture | add a location to a media object | I want to share my knowledge | Discussion within WP2 |
| culture vulture | add a location that is mentioned in an object | I want to share my knowledge | Discussion within WP2 |
| culture vulture | add the name of a person that is in a media object | I want to share my knowledge | Discussion within WP2 |
| culture vulture | edit already made annotations by others | I want to improve incorrect annotations | Shen [REF 4 Appendix, p. 73] |

| As a / an | I want to | Because | Derived from: |
|---------------------------|--|--|--|
| culture vulture | remove already made annotations by others | I want to improve incorrect annotations | Shen [REF 4 Appendix, p. 73] |
| culture vulture | comment on an item | I can share my knowledge/thoughts/questions on an item with other users, the institution and Europeana | Mak and van Lieshout [REF 14], Backlog |
| culture vulture | create multiple tags for a single work of art | in that way I can make the connection between different objects and can afterwards show these linked objects to my students | Shen [REF 4 Appendix, p. 73] |
| culture vulture | add a tag where I specify the subject(s) that the object encompasses | I want to share my knowledge | Discussion within WP2 |
| culture vulture | add the title of a media object | I want to share my knowledge | Discussion within WP2 |
| culture vulture | edit or delete my annotation | sometimes I will annotate something wrongly and want to be able to undo my fault | Shen [REF 4 Appendix, p. 71] |
| Europeana website visitor | comment about the essence of a media object | on the one hand it helps the project, but on the other hand it shows me that I care and know about the subjects I mention in my comments | Shen [REF 4 p. 26] |
| Europeana website visitor | write comments about media objects | I think knowledge and opinion exchange is important, I like to learn from other people's questions and gather information via their comments and links | Shen [REF 4 p. 24] |
| culture vulture | mark an item and write what I know about this item | I like to share my knowledge | Shen [REF 4 Appendix, p. 73] |
| culture vulture | relate an item to another item / exhibition | the context of an item becomes more complete / clear | Mak and van Lieshout [REF 14], Backlog |
| culture vulture | create a link between objects | I want to share my knowledge | Discussion within WP2 |
| Europeana website visitor | recommend related materials on content | in this way I connect future visitors to trusted sources about the subject they're interested in | Shen [REF 4 p. 27] |
| culture vulture | add links to connected literature or other resources that are somehow connected to an item | in this way I connect future visitors to trusted sources about the subject they're interested in | Shen [REF 4 Appendix, p. 73] |

| As a / an | I want to | Because | Derived from: |
|---------------------------|---|---|--|
| culture culture | define the type of relation between objects that I linked | I want to share my knowledge | Discussion within WP2 |
| culture culture | make an annotation in which I can point out the name of the book that I have written about this topic | the visitors must know that I have written this book about that subject | Shen [REF 4 Appendix, p. 73] |
| Europeana website visitor | create or participate in a discussion on a topic that I am interested in | I am interested what other people think about topics that I am interested in and I enjoy sharing my knowledge with other people | Shen [REF 4 p. 27] |
| Europeana website visitor | write a review about a media object | in that way I create a memory and it reflects my own consumption of media | Shen [REF 4 p. 26] |
| culture culture | have the option to explain why I like an object instead of just like or not like the object | I want to understand why other people like a particular item, instead of just knowing that they like a specific item | Shen [REF 4 Appendix, p. 72] |
| culture culture | place item in a genre / channel | the context of an item becomes more complete / clear | Mak and van Lieshout [REF 14], Backlog |
| culture culture | have the option of making a visual mark when annotating | that makes it easy to point other users to the particular feature of the object | Shen [REF 4 Appendix, p. 70] |
| culture culture | have a scientific debate about a work of art with other users (for instance in the unfolded annotation box) | this will improve my understanding of a certain topic | Shen [REF 4 Appendix, p. 73] |
| culture culture | give songs a place in the corpus where they belong | others can find out more easily what the name of a particular tune that they listen to is | Initial research on thesession.org |
| culture culture | point out which tunes work well as sets (a group of three tunes usually in the same key or the same rhythm) | I want to share such a list with likeminded musicians | Initial research on thesession.org |
| culture culture | view the image annotations of other users | in that way I will get a better understanding of the annotated image | Discussion within WP2 |

| As a / an | I want to | Because | Derived from: |
|--|--|---|--|
| culture vulture | clearly distinguish user created info from professionally created info | I can better judge the trustworthiness of an item. | Mak and van Lieshout [REF 14], Backlog |
| culture vulture | create a collection | it is inspiring | Shen [REF 4 Appendix, p. 71] |
| culture snacker | curate a collection of songs | in order to share my favourite music with my friends and to invite more friends to my profile | WP2 brainstorm session |
| culture vulture | have a shortcut icon to collect an item in my collection | in that way I do not always have to go through the item page first | Shen [REF 4 Appendix, p. 70] |
| culture vulture | recommend items for a collection | I want to help people/organisations out in creating a good exhibition | Shen [REF 4 Appendix, p. 71] |
| culture snacker/vulture | add my own curated collection | in order to share my own knowledge of music | WP2 brainstorm session |
| Europeana website visitor | share my own collections | I like to show my own curated collections to others | Shen [REF 1 p. 27] |
| participant in the expert evaluation of the pre-concepts and ideas for the Channels design | be invited for collaborating on a collection | then I would feel a sense of acknowledgement from the one who invites me | Shen [REF 4 Appendix, p. 54] |
| participant in the expert evaluation of the pre-concepts and ideas for the Channels design | collaborate on making a collection with friends, classmates and colleagues | it is fun and this would be another way of socializing with each other and know more about others' interest | Shen [REF 4 Appendix, p. 54] |
| culture vulture | write the name of my collection in multiple languages | in that way I can show it to my Hungarian colleagues as well | Shen [REF 4 Appendix, p. 72] |
| culture vulture | be able to share my created collection publicly | people in this collection all have the same interest as me. So if I join in this topic, we can discuss the thing we like together | Shen [REF 4 Appendix, p. 72] |
| culture vulture | have the option to choose if my collection is visible to others or not | sometimes I don't want everybody to look at the collections I made | Shen [REF 4 Appendix, p. 71] |
| culture vulture | create a collection that I can share privately | I might want to add these objects into a new media object that I am currently creating with a group of people | Shen [REF 4 Appendix, p. 72] |

| As a / an | I want to | Because | Derived from: |
|-------------------------|--|--|------------------------------|
| culture vulture | complete a quest | I will feel the sense of achievement afterwards, partly by scrolling through the objects I gathered within the quest (watching my own results) | Shen [REF 4 Appendix, p. 54] |
| culture vulture | participate in a quiz | it will enhance my eagerness to learn | Shen [REF 4 Appendix, p. 54] |
| culture vulture | share or collect items from Europeana and put them on my own website that is about my family history | I want to get the best overview about my own family's history as possible | Shen [REF 4 Appendix, p. 71] |
| culture vulture | store the information that I like all the collections that I like | in that way people can see how many collections I like | Shen [REF 4 Appendix, p. 71] |
| culture vulture | see the user's picture in the annotation tooltips | in that way you can feel more valuable of yourself and will increase the visibility | Shen [REF 4 Appendix, p. 72] |
| culture vulture | automatically list the most popular annotations at the top of the annotations | that makes it more easy to see the most interesting annotations first | Shen [REF 4 Appendix, p. 72] |
| culture snacker/vulture | add upcoming events | in order to physically meet the community | WP2 brainstorm session |
| culture snacker | rate the music I listen | in order to share my favourite music with my friends and to invite more friends to my profile | WP2 brainstorm session |
| culture snacker | link this music to my social media accounts (Facebook) | in order to share this with my friends | WP2 brainstorm session |

Shen conducted extensive user research to provide a recommendation for the way the design development of Channels should be followed from the user's perspective. This research was done within the scope of WP4. More design recommendations for WP4 followed shortly thanks to the user research done by User Intelligence [REF 14]. This research also provided some relevant user stories.

Furthermore, a brainstorm session during the Co-design Workshop on the 2 June 2014 in Hilversum gave valuable input from the members of WP2 as well. The brainstorm sessions' starting point were the personas that Brigitte Jansen from NISV created: Jasmine (culture snacker), Jacob (culture vulture) and Daan (shares traits from both the culture vultures and the snackers). These personas were based on the personas Chen created for the Art & Art History Channel for WP4. See Annex 1 of

this document for the actual description of the three personas as used during the co-design workshop.

In addition to the culture vultures and the culture snackers, there are also the traditional Irish and Scottish musicians, who could be classified both as music vultures and experts in their particular field. They generally engage with traditional music in their spare time. The social aspect of the sessions where they meet other musicians to play, learn, teach and have fun is very important. Next to these physical meetings the traditional Irish and Scottish musicians gather online on platform www.thesession.org, a forum-type website for traditional musicians. Historypin undertook some initial user research on this online platform, providing user stories from the point of view of these music vultures.

5 EXCHANGE POLICIES

Given that Europeana Sounds will establish micro-tasks which will generate user contributions, it is pertinent to clarify how these enrichments and content are exchanged between the (users of the) different Europeana Sounds platforms, Europeana (both the portal and the channels) and original content/data providers. There are several aspects of this to consider. One is the legal aspect, and the second is how the quality of user contributions will be assured.

5.1 Legal conditions

The legal aspect of dealing with the user contributions will be governed according to the [Europeana Terms for User Contributions](#)²⁹, which are part of the general [Europeana Terms and Policies](#)³⁰.

The majority of the user contributions that will result from the crowdsourcing micro-tasks that Europeana Sounds WP2 will develop, will be annotations in the form of (metadata) enrichment, whether that is classification, contextualisation, etc. As stated in the Europeana Terms for User Contributions, “all Metadata contributed by Users to Europeana will be made available under the terms of the [Creative Commons CC0 1.0 Universal Public Domain Dedication](#)³¹. This means that the User irrevocably grants Third Parties the right to freely use such Metadata without any restrictions.” This is aligned with the [Europeana Data Exchange Agreement](#)³² which states that the same conditions apply as to the metadata that is contributed to Europeana by the data providers. This ensures that it is legally possible to reuse the user contributed annotations in the same way as the original metadata, and that it is also possible to reconcile them, if required.

Some actual content may also be generated as part of Europeana Sounds, for example as part of a contextualisation task (an article or elaborate comment, describing the content). Again, under the Europeana Terms for User Contributions, this content will be licensed under a [Creative Commons Attribution-ShareAlike license](#)³³.

It is important to explicitly communicate to users the fact that these Europeana Terms for User Contributions apply to their activity, for all platforms involved. In the case where user accounts are involved, this should be an explicit action in the registration procedure.

5.2 Quality assurance

With the number of user contributions Europeana Sounds aims to gather, it is not feasible to manually assure the quality of these contributions. Alternatively the crowdsourcing infrastructure needs to rely on other mechanisms for quality assurance. As part of Task 2.2, WP2 will develop mechanisms for improving four types of existing enrichments:

²⁹ <http://europeana.eu/portal/rights/terms-for-user-contributions.html>

³⁰ <http://europeana.eu/portal/rights/terms-and-policies.html>

³¹ <http://creativecommons.org/publicdomain/zero/1.0/>

³² <http://pro.europeana.eu/web/guest/data-exchange-agreement>

³³ <http://creativecommons.org/licenses/by-sa/3.0/>

1. Automatic enrichments that are applied during the ingestion of metadata
2. Semi-automatic enrichments, resulting from the application of the ontologies by data providers, as specified in D1.3 *Ontologies for Sound* [REF 15]
3. Assisted structured enrichments by users, as a result of crowdsourcing micro-tasks
4. Manual enrichments by users, as a result of crowdsourcing micro-tasks

The latter two types of existing enrichments show that the improvement of existing enrichments not only leverages a crowd of culture vultures in order to correct and/or refine the result of (semi-) automatic enrichments, but also to correct and/or improve the results of its own collective contributions to the enrichments (community sourcing). This enables a constant process of improvement of the metadata quality, through crowdsourcing.

Another mechanism, which the crowdsourcing infrastructure will implement to ensure the integrity of the original metadata, is based on recording the provenance of user contributions by formatting them as annotations. This will separate them from the source metadata that is provided by the institutions.

Lastly, it is entirely up to the data providers themselves whether or not they will reuse the results of crowdsourcing and user contributions within their own systems. The crowdsourcing infrastructure will provide methods for making this possible, but the final decision remains with the data providers themselves.

6 PLANS FOR TECHNICAL IMPLEMENTATION

This section provides general information on the development plans for the different platforms involved in the crowdsourcing infrastructure. A more thorough technical design will be described in D2.2 *Functional design of semantic enrichment* in M10 (November 2014). MS8 *Functional design of semantic enrichment* provides a first draft of the system architecture and describes how all the different platforms are connected to the core infrastructure and how the data flows between them [REF 16].

6.1 Core Infrastructure

Within the Europeana Creative project a prototype [Europeana Annotations Service](#)³⁴ and [API](#)³⁵ has been developed and will be further developed. The goal is to make it a core service in the Europeana Platform and use it for Channel development within WP4, but it should be noted that the service will be made available to non-Europeana Foundation users as well. This allows services/communities such as [Waisda?](#) or [Pybossa](#)³⁶/[Crowdcrafting](#) to integrate with the Europeana platform, to fetch object metadata from it and write user annotations into it in real-time.

As of October 2014, the prototype currently supports tagging and image annotations (adapted for use with Annotorious in the front-end). The next type of annotations that will be supported will be user created sets/collections.

Europeana (EF) plans are to further develop the Annotations Service within Europeana Sounds, extending its functionalities to support the specific requirements of the users targeted within the Europeana Sounds project. Based on our user research, allowing users to link objects together or to select external resources (like Wikipedia articles) are the top annotation type candidates to extend support for.

6.2 Europeana Channels

Within the Europeana Sounds project, the Europeana (EF) plans to exploit the Europeana Annotations Service and API within the planned Music Channel as well as in other channels with different themes.

The outcomes of the first phase of User Experience research and design for the Music Channel were reported by the end of September 2014 as part of WP4. Currently our user research indicates that only a small proportion of culture vultures say they are interested in contributing annotations. This indicates that while Channels will feature annotation capabilities, it will not be central to their design and user experience. Instead the strategy from the Europeana Foundation will be to partner with dedicated services for user annotations such as Historypin. Therefore Europeana will develop the means to exchange metadata, including user annotations, with Historypin.

³⁴ https://www.assembla.com/spaces/europeana-creative/wiki/Annotation_Service_and_Architecture

³⁵ <https://github.com/europeana/annotation>

³⁶ <http://crowdcrafting.org/about>

Irrespective of the type of content (or Channels theme), allowing users to create sets of existing objects and linking objects together or to external resources is rated as important. Other types of annotations are rated differently depending on the theme - art interested users for example rate image annotation in high-resolution zoomable images highly.

6.3 Historypin

Historypin will be used as a platform to support local community activities related to the identification, enrichment, geotagging or annotation of sound-related material. This will be enabled by creating special 'project' sections on the Historypin platform, in which sound-related material from partners (and possibly end-users) relating to a single theme or content type will be presented in a coherent way and contributions solicited.

Historypin will enable mechanisms to support enrichment or addition of related material; this might include contextualisation through association of related media, through media-based or textual responses, through selection or generation of tagging or through manual or semi-automatic provision of structured enrichments such as geo-location or entity identification.

Historypin will develop the specific requirements for this crowdsourcing platform based on a close consultation with the projected end-users of the platform; experience has suggested that these user needs can dramatically simplify the technical complexity needed.

6.3.1 Historypin audiences

Historypin intends to approach traditional musicians as a target audience. As explained above, they are experts (or 'music vultures') in their particular area and have clearly defined 'jobs' that they are seeking to get done.

Further research needs to be done into this audience segment and into their needs, wants and expectations, but initial research as shown that traditional musicians are using www.thesession.org as a platform to answer their questions about tune identity.

Allowing these traditional musicians to use Historypin to refer more easily to archival materials from ITMA, Comhaltas and Tobar an Dualchais will increase their user experience and encourage reuse and appreciation of archival data.

6.3.2 Historypin content

Historypin intends to ingest materials from ITMA, Comhaltas and Tobar an Dualchais as these three archives have similar content around Irish and Scottish traditional music. In addition, there is also the possibility to ingest materials from www.thesession.org, where many tunes are transcribed and linked to various audio and video files.

This body of content can then be used by visitors to www.thesession.org to point towards when answering questions other musicians might have about tune names and sets.

6.3.3 Other platforms that we hope to link to

Historypin will link to www.thesession.org and is hoping to also link to www.tunepal.com, which is a platform for the automated transcribing of tunes.

6.4 Pundit

[Pundit](#)³⁷ is software that allows you to semantically annotate web pages. It is composed of four components: client, feed, ask, annotation server that together allow you to make large parts of web content 'annotatable'.

The main idea behind Pundit is to enable users not only to comment, bookmark or tag web pages, but also to create semantically structured data while annotating, thus enriching the so-called Web of Data.

The ability to express semantically typed relations among resources, relying on ontologies and specific vocabularies, not only enables users to express unambiguous and precise semantics, but also, more interestingly, fosters the reuse of such collaboratively created knowledge within other web applications. For example: provide a powerful semantic search, build innovative ad-hoc data visualisations or ultimately improve the way users explore the web.

Parts of Pundit will be developed and/or improved as part of Europeana Sounds. Firstly, some customisations will be undertaken through the use of appropriate configuration files that allow for the adaptation of the annotator to the ontology of the project. Then, one or more templates of annotations will be realised - an annotation template can greatly simplify the operation of annotation performed by the end user. Finally, some changes to the graphics of the web page which allows you to make annotations (feed) will be performed. The details of the customisations will be described in the forthcoming D2.2 *Functional Design of Semantic Enrichment*.

6.5 Widget

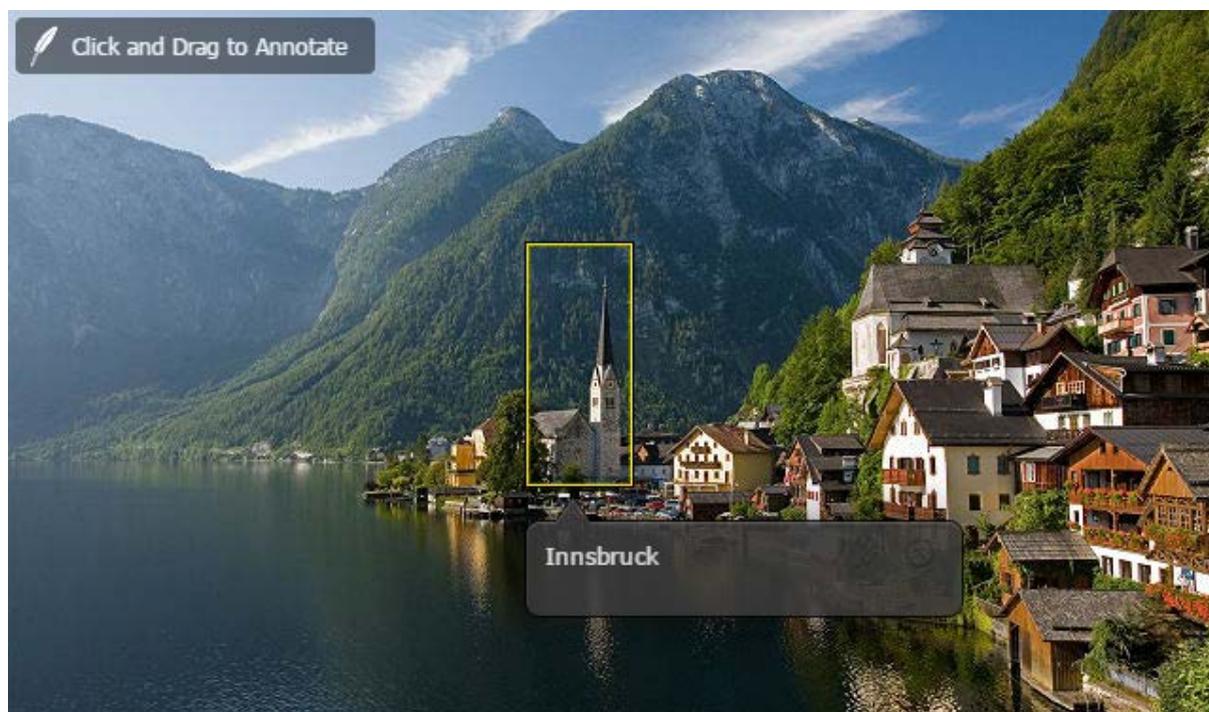
There is still a common interest to store and provide access to these enrichments through a central UGC repository and to provide lightweight solutions for integrating them into domain specific applications through html widgets. The plan is therefore to develop an Annotorious Plugin for the UGC service, in order to address these requirements.

Annotorious [REF 17] is a light-weight, Java-script based tool for annotating images on the Web. This is an extendable framework based on plug-ins and plug-outs, which provided a simple mechanism for developing new plugins and integrating them into web applications.

New plug-ins can easily developed for supporting additional annotation [REF 18] and media types, or storage systems that provide access through standard Web interfaces [REF 19].

³⁷ <http://www.thepund.it/>

Figure 11: Annotorius Semantic Tagging Plugin [REF 20]



7 ANNEX 1: PERSONAS

Personas used during the Co-design Workshop created by Brigitte Jansen, based on the personas used by Chenchen Shen for the Art & Art History Channel.

7.1 Daan

Keywords: Bass player, in for a good laugh

Personal:

Daan is a 27 year old bass player. He is certainly not a night person (though he really enjoys the handpicked gigs he visits) and lives with his girlfriend and cat in a small apartment. Daan loves to write and perform his music and is having a blast at it with his band mates.

Interests:

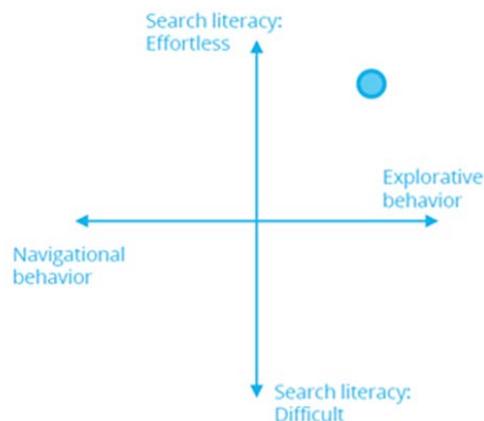
Daan likes a good barbeque on a hot summer day, but when it rains you can find him trying out things on his bass guitar. Daan enjoys listening vinyl records and can stay hours in a record store.

Media use:

Daan doesn't really like the whole social media hype. He is a bit sceptical about it and also still has his old Nokia. Daan strolls the internet for music he likes. He also really likes to research the faces behind the music.

Search strategy:

Daan doesn't just want to hear new music, but also wants to know more about the band (not the rumours, just the former bands and collaborations, etc.) He isn't really interested in opinions of others about music.



7.2 Jacob

Keywords: Researcher, wants to know it all and is a verifier

Personal:

Jacob is a 47 year old researcher at the British Trust for Ornithology. Here he researches bird trends. Jacob is very enthusiastic about nature and likes to take his wife and his son on hikes through the forest.

Interests:

Jacob loves nature, but also enjoys reading a good book with a glass of wine. Once a year he goes on a bike hike through France with some of his friends.

Media use:

Jacob is used to working on the computer every day. He possesses a smartphone, but mostly uses it for calling and Whatsapp. He rarely plays a game on it and never searches anything on his phone. He finds his screen too small to read comfortably.

Search strategy:

Jacob doesn't just type something in at Google. He is really annoyed about the amount of irrelevant outcomes. Therefore he only searches on sites he knows that provide him good and relevant information.



7.3 Jasmine

Keywords: Social, loves to dance and enjoys a good conversation.

Personal:

Jasmine is a 23 years old Communication student. She enjoys her student life and knows here she can buy the best beers in town. Jasmines best makes her homework with some nice country music on the background.

Interests:

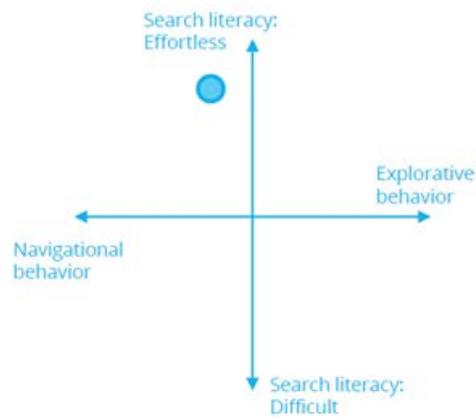
She loves to go out with her friends and likes to dance all evening.

Media use:

Jasmine is always online via her laptop and smartphone. She loves sharing picture through Facebook and Instagram.

Search strategy:

Jasmine loves to know more about the things she sees online and always browses online to find new information. She always has dozens of tabs opened in her browser and bookmarks a lot of interesting sites.



8 ANNEX 2: DATA PROVIDERS SURVEY

Below is the survey on enrichments through crowdsourcing, as distributed to all data providers.

Link to the completed survey -

<https://basecamp.com/1936492/projects/5105397/uploads/17288804>

Data Providers Survey on Enrichments through Crowdsourcing

To get a clear understanding on where we want to go with the enrichments through crowdsourcing that Work Package 2 (Enrichment) will support, we like you to fill in this survey. Our efforts will be informed by the results of this survey, so please express your requirements and wishes in relation to gathering enrichments from the crowds!

The questions in this survey deal with the information needs of you as a Data Provider in order to better describe our collections on the one hand, and the communities and target audiences you think can be enabled to provide this information through crowdsourcing on the other hand.

We are interested in all the information that you can provide in response to our questions, so please don't feel uncomfortable to write lengthy responses to our questions. The **deadline** for this survey is **Friday the 23th of May**.

General

1) Please state your name:

2) State the acronym of your institution as written in the [DoW](#):

Your collection(s) in general

3) What are the highlights of your collection and what makes them special?

4) What kind of information about your collection is still missing / could be richer?

5) Why is it important to supplement this information?

6) How and where would you record this information, once you've got it?

7) Would you like to cross link your music content with other repositories providing complementary information (e.g. music lyrics, music scores, artists)?

8) Which repositories would be relevant for your collection (e.g. other content providers, VIAF, online music lyrics, Dbpedia, etc.)?

9) What part of your collection(s) would you like to make available for crowdsourced enrichments?

Communities

10) Describe the communities that you are currently engaging with (if any):

11) In what ways and on which platforms is your community engaging with the collection:

12) Describe a community that you would like to engage with:

13) How would you reach this community?

14) Have you ever been involved in a crowdsourcing project? If yes, please give a description about his project.

How to make a rich item better?

15) Please name a very well described item in your collection (fill in the title and the link to the description of this item on your website):

Title:
Link:

16) What could make this item even more interesting? (think as broadly as you can, so not only about enriching the existing metadata fields, but for example the relation of this item with other objects and events (inside and outside your institution) and perhaps user's subjective input on the item)

- a)
- b)
- c)
- d)
- e)

17) Where do you think the people are located that can provide this valuable information? (please make sure answer 17a correspond with 16a, 17b with 16b, etc.). And do you already know these people?

- a)
- b)
- c)
- d)
- e)

How to make a poor item richer?

18) Please name an example of an item from your collection that is lacking essential information (fill in the title and the link to the description of this item on your website):

Title:
Link:

19) What could be valuable information for this item? (think as broadly as you can, so not only about enriching the existing metadata fields, but for example the relation of this item with other objects and events (inside and outside your institution) and perhaps user's subjective input on the item)

- a)
- b)
- c)
- d)
- e)

20) Where do you think the people are located that can provide this valuable information? (please make sure answer 20a correspond with 19a, 20b with 19b, etc.). And do you already know these people?

- a)
- b)
- c)
- d)
- e)

21) What would in your opinion be the best way to get this community to enrich items from your collection? (so think about motivating functionalities on the website, but also about certain events or other offline stimuli for engaging people with your collection)

General

22) Please comment here if you have anything to add to the survey in terms of information we might need:

23) If you have any feedback on the survey, it would be very welcome:

9 REFERENCES

| | |
|--------|---|
| Ref 1 | EC-GA including Annexe I ("Description of Work") http://pro.europeana.eu/documents/2011409/2034825/Description+of+Work |
| Ref 2 | Crowdsourcing in the Cultural Heritage Domain: Opportunities and Challenges (J. Oomen and L. Aroyo, 2011). http://www.iisi.de/fileadmin/IISI/upload/2011/p138_oomen.pdf |
| Ref 3 | MS7 End-user contributions defined http://pro.europeana.eu/documents/2011409/ed3b1dbb-ca33-4754-8d6f-abb921e38b46 |
| Ref 4 | Design for User Engagement on Europeana Channels (C. Shen, 2014). https://basecamp.com/1936492/projects/4984678/attachments/100527076/download |
| Ref 5 | <i>We-Think: The Power of Mass Creativity</i> (Leadbeater, C. London: Profile, 2007) |
| Ref 6 | <i>Cognitive surplus: Creativity and generosity in a connected age</i> (Shirky, C. New York: Penguin Press, 2010) |
| Ref 7 | 'Emerging Practices in the Cultural Heritage Domain - Social Tagging of Audiovisual Heritage' (Oomen, J., Baltussen, L., Limonard, S., van Ees, A., Brinkerink, M., Aroyo, L., Vervaart, J., Asaf, K., Gligorov, R. in <i>Proceedings of the WebSci10: Extending the Frontiers of Society On-Line</i> , Raleigh, NC, US: April, 2010) |
| Ref 8 | 'From tagging to theorizing: Deepening engagement with cultural heritage through crowdsourcing' (Ridge, M. <i>Curator: The Museum Journal</i> , 56 (4), 2013, p.435 - 450). |
| Ref 9 | 'Semantic Tagging on Historical Maps' (Haslhofer, B., Werner, R., Lagoze, C., Guimbretiere, F. ACM WebScience 2013. Paris: May, 2013). http://arxiv.org/abs/1304.1636 |
| Ref 10 | Open Annotation Data Model (H. Van de Sompel et al, 2013). http://www.w3.org/ns/oa#d4e731 |
| Ref 11 | 'Nichesourcing: Harnessing the Power of Crowds of Experts' (Boer, V. de, Hildebrand, M., Aroyo, L., Leenheer, P. De., Dijkshoorn, C., Tesfa, B., Schreiber, G. in <i>Proceedings of the 18th International Conference on Knowledge Engineering and Knowledge Management, EKAW 2012</i> , p. 16-20.). |
| Ref 12 | 'Using the past to explain the present: interlinking current affairs with archives via the Semantic Web' (Raimond, Yves, Smethurst, M., McParland, A., Lowis, C. (BBC R&D) August, 2013) http://downloads.bbc.co.uk/rd/pubs/whp/whp-pdf-files/WHP260.pdf |
| Ref 13 | Completed Data Providers Survey on Enrichments through Crowdsourcing https://basecamp.com/1936492/projects/5105397/uploads/17288804?enlarge=114233678#attachment_114233678 |

| | |
|---------------|--|
| Ref 14 | Europeana Channels: Report User Research version 1.0 (Mak, M., Lieshout, B. van., 2014) https://basecamp.com/1936492/projects/6415078/uploads/15995497?enlarge=105018311#attachment_105018311 |
| Ref 15 | D1.3 Ontologies for Sound http://pro.europeana.eu/documents/2011409/ed7f27af-65ff-4ea3-9d00-8345c31489cf |
| Ref 16 | MS8 Functional design of semantic enrichment http://pro.europeana.eu/documents/2011409/d1fa3069-8568-44dc-9c0c-1870017b4057 |
| Ref 17 | Annotorious http://annotorious.github.io/ |
| Ref 18 | Annotorious: Plugins for semantic tagging http://annotorious.github.io/plugins.html#semantic-tagging |
| Ref 19 | Annotorious: Plugins (parse) http://annotorious.github.io/plugins.html#parse |
| Ref 20 | Sample image source: Hallstat, Austria by Nick Csakany (Public Domain) http://commons.wikimedia.org/wiki/File:Hallstatt_300.jpg |