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

D1.7 Aggregation Report 2

Revision: Final

Date: 02/03/2016

Authors: Tom Miles (BL)

Abstract: This document reports on the second year of metadata aggregation in the Europeana Sounds project. It should be seen as a continuation of D1.5 Aggregation Report 1. It outlines how Data Providers have prepared their data for publication on Europeana, the progress made in year two of the three-year project, the challenges encountered and how these were resolved.

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

Coordinated by the British Library, the Europeana Sounds 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/

Revision history

Version	Status	Name, organisation	Date	Changes
0.1	ToC	Tom Miles, BL	03/12/2015	
0.2	1st draft	Tom Miles, BL	30/01/2016	
0.3	2nd draft		16/02/2016	Input from reviewers
0.4	Final draft	Tom Miles, BL	29/02/2016	Input from PMB
1.3	Final	Richard Ranft, BL	02/03/2016	Layout, minor changes

Review and approval

Action	Name, organisation	Date
Reviewed by	Anastasia Gasia, The European Library/European Foundation Stephan Bartholmei, German Digital Library	15/02/2016 15/02/2016
Approved by	Coordinator and PMB	22/02/2016

Distribution

No.	Date	Comment	Partner / WP
1	02/03/2016	Submitted to the European Commission	BL/WP7
2	02/03/2016	Posted on Europeana Pro website	BL/WP7
3	02/03/2016	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This document contains original unpublished work except where clearly indicated otherwise.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, 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

Copyright notice

Contents

Executive summary: D1.7 Aggregation Report 2

This deliverable reports on WP1 Aggregation activities of all 18 Data Providers in the Europeana Sounds project. D1.7 provides details on the amount of audio and audio-related content required by the project to be published on Europeana by the end of year two: 250,000 audio items and 90,000 audio-related items, of which 40,000 are required to be freely reusable.

The deliverable outlines how the Data Providers worked with the National Technical University of Athens (NTUA) and the Europeana Foundation (EF), coordinated by the Work Package Lead (British Library) in order to meet these objectives.

At the end of Year 2, all five Key Performance Indicators for WP1 were met or exceeded.

D1.7 continues with an account of the remaining work required in Year 3 in order to meet the project requirements. The document discusses what lessons have been learned from this work; and looks ahead to how Europeana Sounds might continue to work with associate partners in the last year of the project

1 Introduction

1.1 WP1 tasks and documents providing context

As outlined in the Europeana Sounds' Description of Work (DOW), the WP1 tasks are as follows:

- T1.1: Content Selection Policy – completed (D1.1 *Content Selection Policy*, MS1)
- T1.2: Ontologies – completed (D1.3 *Ontologies*, MS3)
- T1.3: EDM profile – completed (D1.4 *EDM Profile for Sound*)
- T1.4: Aggregation management – ongoing. The first tranche of content and metadata was ready for ingestion in month 13 (D1.5 *Aggregation Report 1*, MS5).
- T1.5: Training and support in content selection policy and aggregation workflow (lead: NTUA) – ongoing (D1.6 *Training Report 1*).

This document, D1.7 *Aggregation Report 2*, deals with the continuing aggregation management from month 12 to month 24.

Upcoming documents are: MS6 *Final tranche of content and metadata ready for ingestion* (month 30); D1.8 *Final Aggregation Report* (month 36); D1.9 *Training report 2* (month 36).

1.2 Europeana Sounds Data Providers

Table 1 show a list of the Data Providers, the consortium partners involved in the Europeana Sounds project which supply approved metadata records, with their project partner number and corresponding country.

Table 1: list of Data Providers

Code and short name	Full name	Country
01 BL	The British Library	UK
02 NISV	Beeld en Geluid	Netherlands
06 BNF	Bibliothèque nationale de France	France
10 CNRS	Centre national de la recherche scientifique	France
11 DIZI	UAB DIZI	Lithuania
12 DNB	Deutsche Nationalbibliothek	Germany
13 FMS	Syllogos Oi Filoi Ths Mousikhs Σύλλογος Φίλων της Μουσικής	Greece
14 ICCU	Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche	Italy
15 ITMA	Irish Traditional Music Archive / Taisce Cheol Dúchais Éireann	Ireland
16 TLA	Max Planck Gesellschaft	Netherlands
17 NLL	Latvijas Nacionālā Bibliotēka	Latvia
18 OEM	Österreichische Mediathek	Austria
19 RBB	Rundfunk Berlin-Brandenburg	Austria
20 TAD	Sabhal Mòr Ostaig	UK
21 SB	Statsbiblioteket	Denmark
22 ONB	Österreichische Nationalbibliothek	Austria
23 FCSH	Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa	Portugal
24 CCE	Comhaltas Ceoltóirí Eireann	Ireland

1.3 Description of content

The audio content that was aggregated is a rich mixture of sound recordings, from wax cylinder recordings of the end of the 19th century to more contemporary recordings. The subjects recorded vary; including oral histories, field recordings, commercial recordings, environmental and natural sounds, radio broadcasts and other performance-based recordings. In addition to the recordings, there are music manuscript collections – mainly from the Austrian National Library but also from the British Library – photographic images, videos, and woodblocks depicting musical scenes. Therefore, the entire combined collection spans worldwide content from the sixteenth century to the present day.

1.4 Technical platforms

1.4.1 Data Providers platforms

Data Providers host their own webpages containing metadata and the digital object, the “content”. The content may be streaming only from the Data Providers’ webpage, or directly accessible by a download link, or both. All content is freely accessible worldwide.

1.4.2 Metadata Ingestion Tool (MINT)

MINT is hosted and run by the National Technical University of Athens, who are also work package lead for the Europeana Sounds’ technical infrastructure (WP5). Data Providers upload their metadata to

MINT and use this tool to transform their metadata to the EDM Sounds profile in XML format. MINT is a dark aggregator – in other words, the metadata on it cannot be accessed by the public or external visitors.

1.4.3 Europeana

Once metadata is published on MINT by Data Providers by regular deadlines throughout the project, the metadata is harvested and ingested by Europeana, as directed by the WP1 lead. The process of harvesting and the timetable of ingestions have recently been revised by the Europeana Foundation. As a result, metadata published on MINT can be accessible on Europeana within a matter of days. Previously, metadata was added to Europeana in monthly instalments.

2 Aggregation processes

2.1 Main stages of ingestion

2.1.1 Content selection

The content selected by Data Providers was from material already available online; most Data Providers had already selected the content and were working on its ingestion.

2.1.2 Metadata preparation

As Data Providers began work on the project, it became clear that there were different ways of preparing metadata, depending on methodology and the way in which Data Providers stored their data. The key task was to prepare metadata so that it was compliant with European Data Model (EDM) Sounds standards. Some Data Providers were able to import their metadata from XML files to MINT and map data to the European Data Model (EDM) Sounds profile. However, for other providers it was more effective to upload spreadsheets in .csv format as this meant that metadata could be edited before upload. In the case of the National Library of France (BnF) it was necessary for NTUA to provide conditional formatting in MINT before the original records from BnF could be imported to MINT and mapped to EDM.

2.1.3 Metadata ingestion

Throughout the year, staff from the Data Providers developed their skills in using MINT, working with NTUA, Europeana and the WP1 lead in order to solve problems. The project's Basecamp tool proved a valuable resource and learning space – this reduced the need for the same questions to be asked again and again. Data Providers have gradually started to answer each other's questions within this forum, before NTUA has a need to investigate.

A main benefit of using MINT to prepare metadata for ingestion has been that prevents incorrectly processed data proceding to publication: any mistakes made in mapping are rejected by MINT and this means that, when it came to the ingestion of metadata for Europeana, only a tiny percentage of records

were invalid. Data Providers were encouraged to diagnose their own errors and this has been made possible by previewing records in EDM Sounds XML format, or in the Europeana preview pane. These two panels on MINT provide a quick way of assessing whether mandatory fields are missing, whether links to webpages or media files are working, whether images are correctly displaying, etc. The EDM panel displays the newly formatted XML: this was the data that would be harvested by Europeana. A green tick confirms the metadata is valid; a red cross denotes a problem, and the area of concern is highlighted.

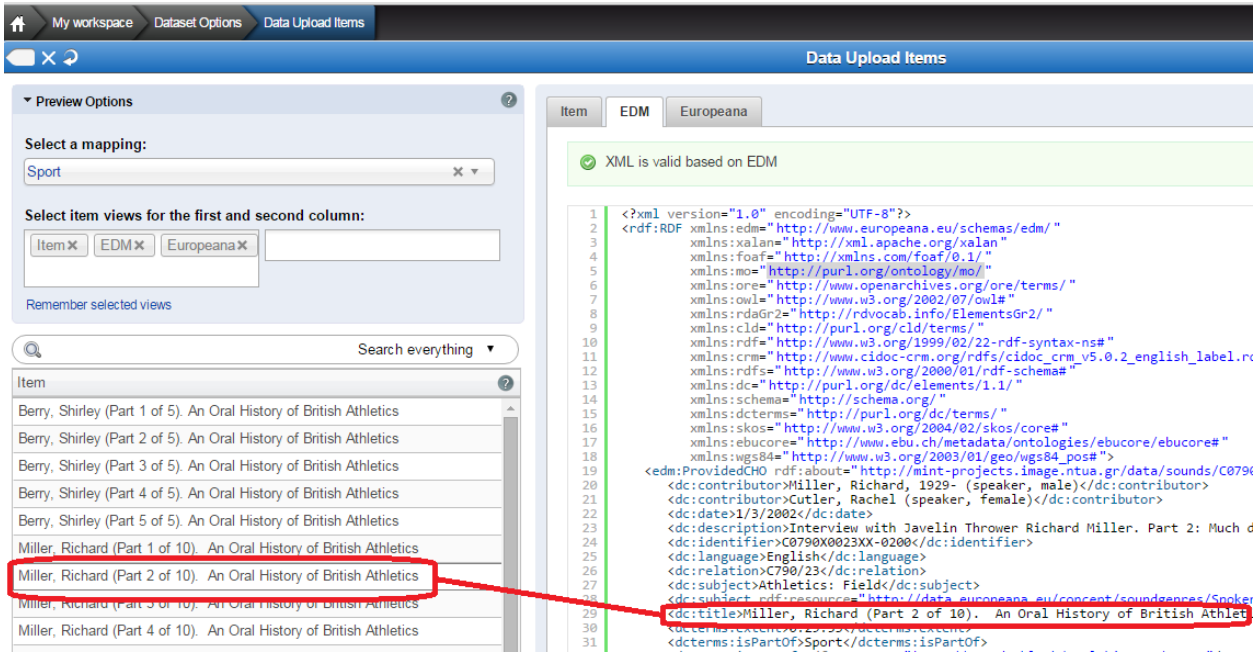


Figure 1: a preview of a record on MINT and its corresponding XML in EDM profile which has been validated

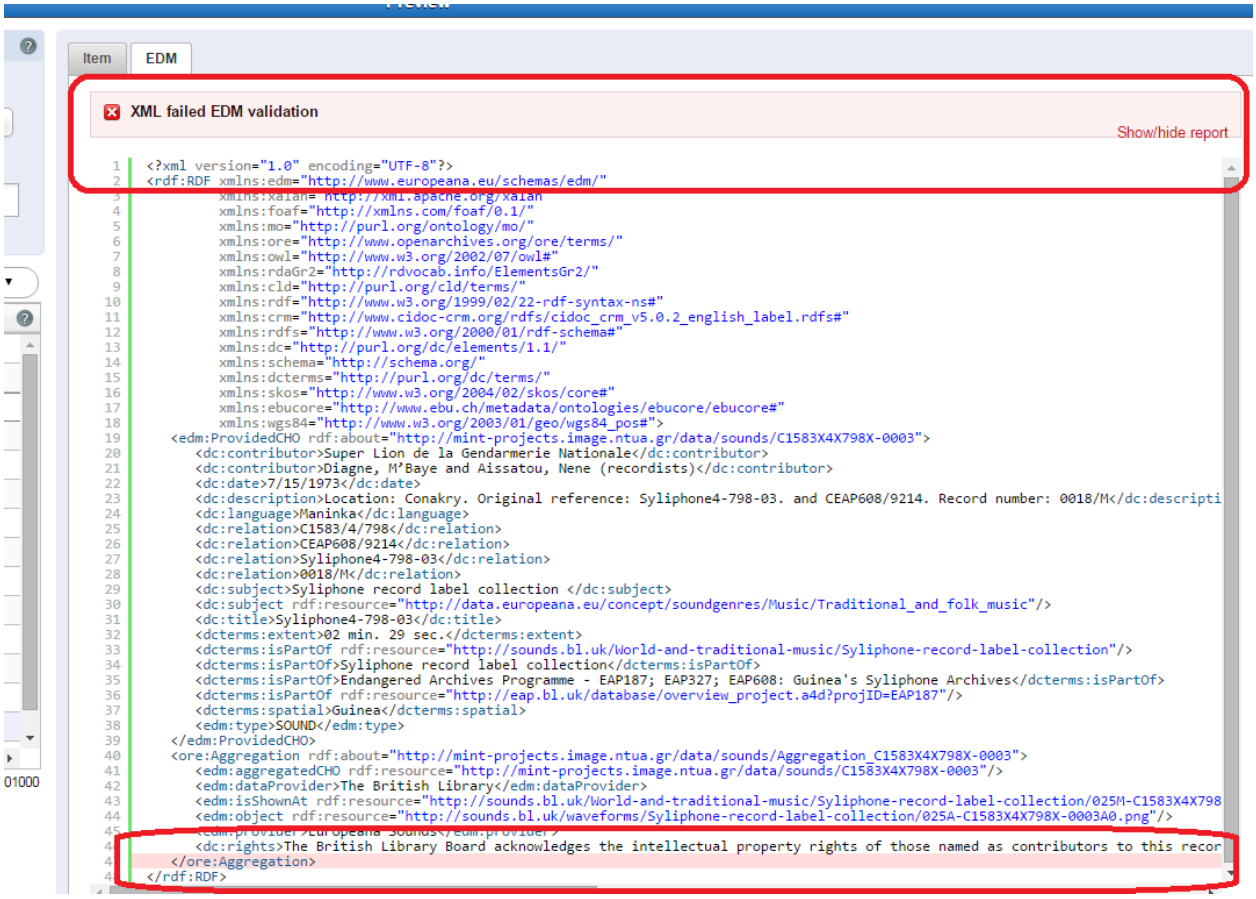


Figure 2: an example of a record on MINT which has failed EDM validation, highlighting the section of the record requiring attention

2.1.4 Metadata curation

Previewing work

Data Providers needed to consider how the metadata they provided to MINT would display on Europeana. The preview portal on MINT provided an opportunity to do this. However, at the time of writing, the Europeana preview panel shows an earlier version rather than the new version of Europeana so, for this to remain a useful function of MINT, a new preview would need to be built.

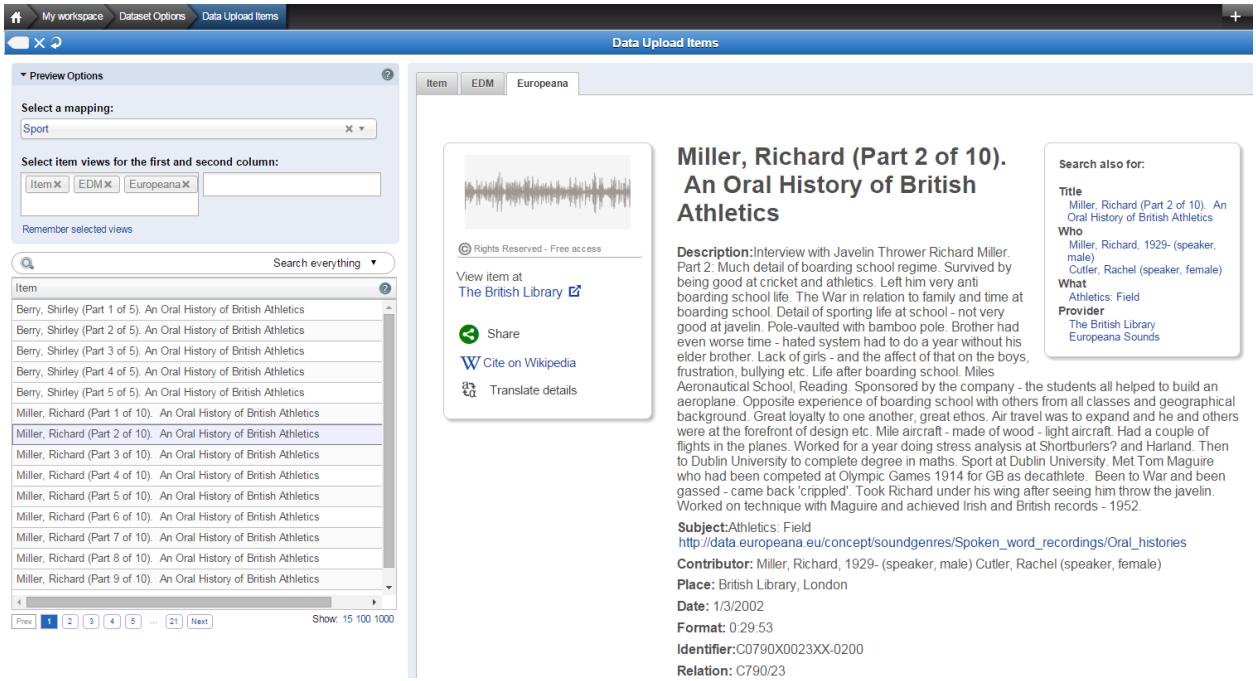


Figure 3: an example of the Europeana preview pane in MINT, displaying a record as it would be seen on the earlier view of Europeana

Thumbnails

A key issue in metadata curation for sounds is the inclusion of thumbnails, which are mandatory for images, in the Europeana portal. For images, a low resolution version of the image itself would almost always suffice. For video, a still would suffice but, for audio, there is not necessarily a thumbnail that can be used, resulting in audio records looking rather dull.

The British Library and CNRS have supplied audio waveforms as unique thumbnails – these were generated as part of their own online web portals. In the case of ICCU, BNF, NLL and ITMA, photographs of the contributors (see Figure 4), or disc labels (see Figure 5), of each recording are used.

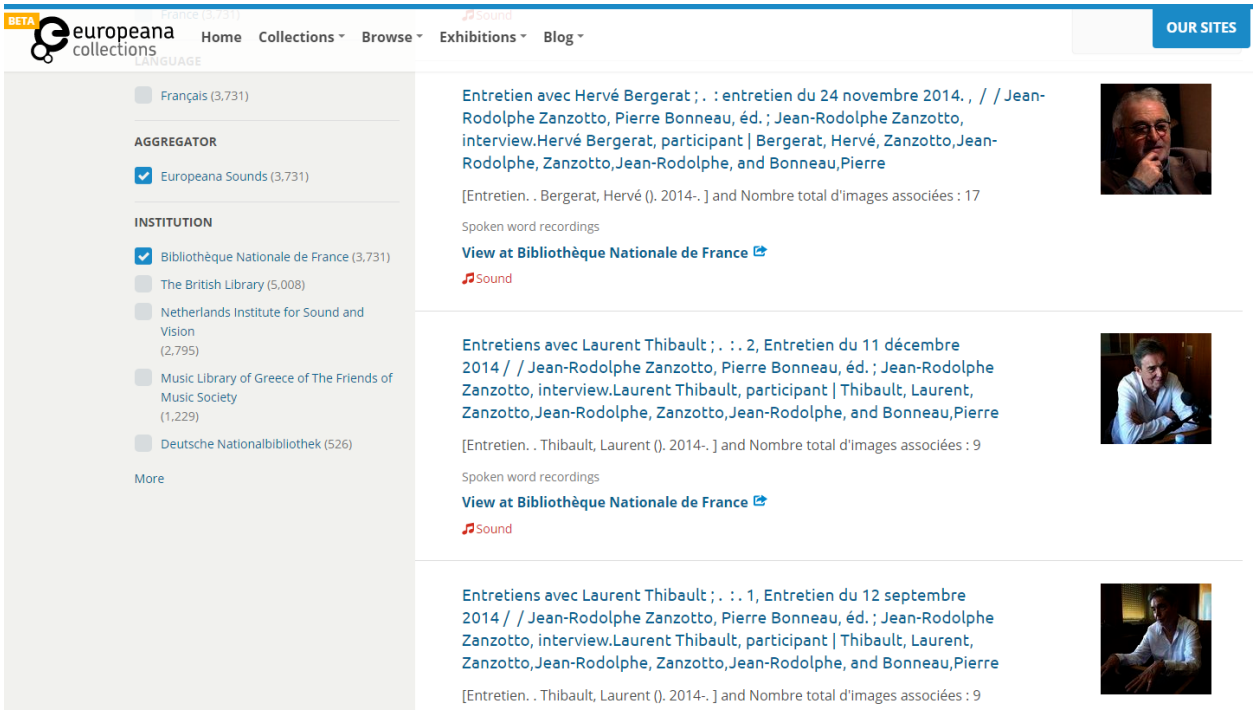


Figure 4: list of recordings from National Library of France (BNF) showing photographs of contributors used as thumbnails in edm:object fields

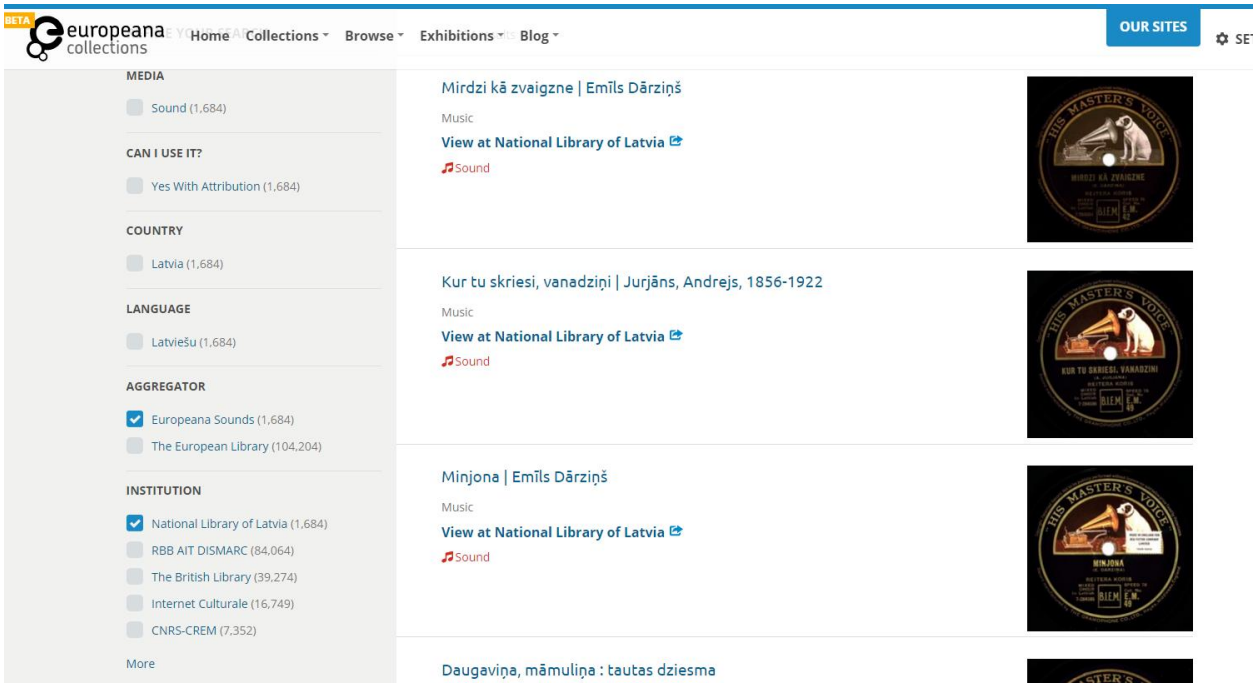


Figure 5: a list of recordings from National Library of Latvia (NLL), using photographs of disc labels as thumbnails in edm:object fields

Accompanying images

Data Providers enhanced their collections by providing additional media where relevant. This could include transcripts of interviews (Figure 6), transcripts of musical recordings, photographs of contributors (Figure 7), etc.

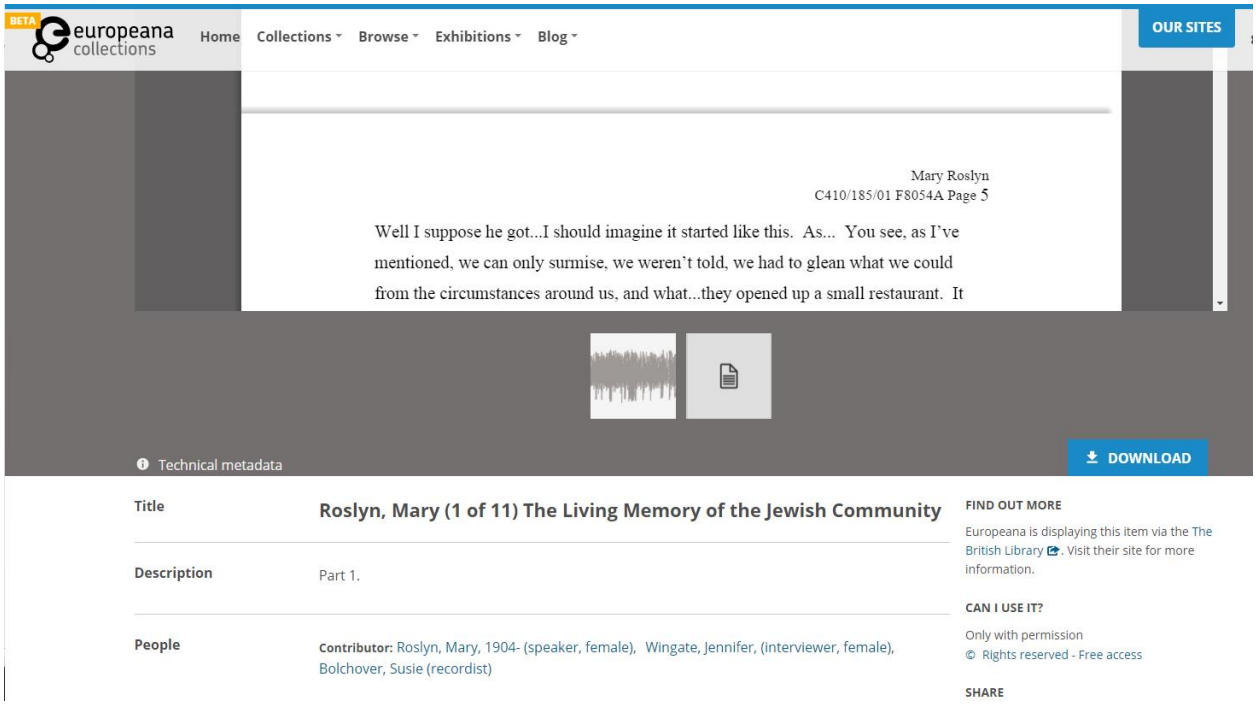


Figure 6: Europeana record showing accompanying transcript in hasView field

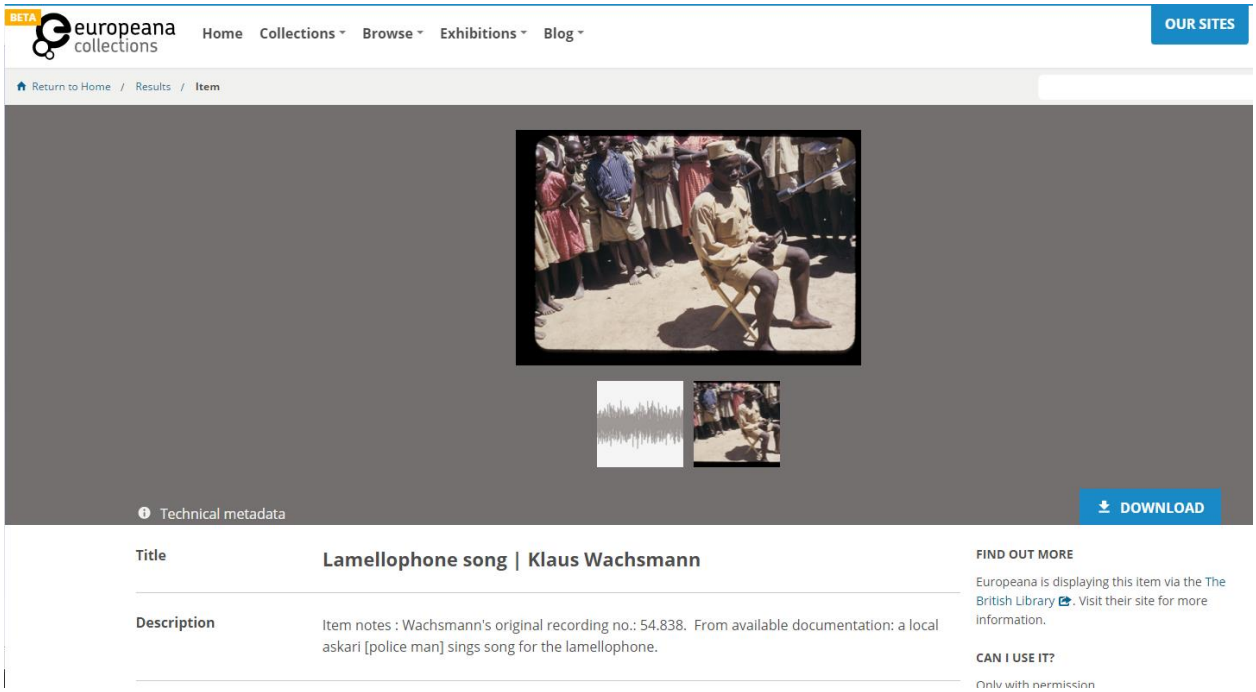


Figure 7: Europeana record showing accompanying image in hasView field

Music Collections

Europeana's Music Collections (Figure 8) were developed during Year 2 of the Europeana Sounds project as part of WP4 Channels Development. This was an opportunity to showcase music related content - from Europeana Sounds and elsewhere on Europeana - thereby presenting a richer selection to the user. Other collections are also being prepared (for example, Europeana Art History and Europeana Fashion).

The Music Collections require directlylinked media and therefore Data Providers have been encouraged to provide direct links from Europeana to their digital objects, wherever legally and technically possible (this is discussed in more detail in 4.1.1).

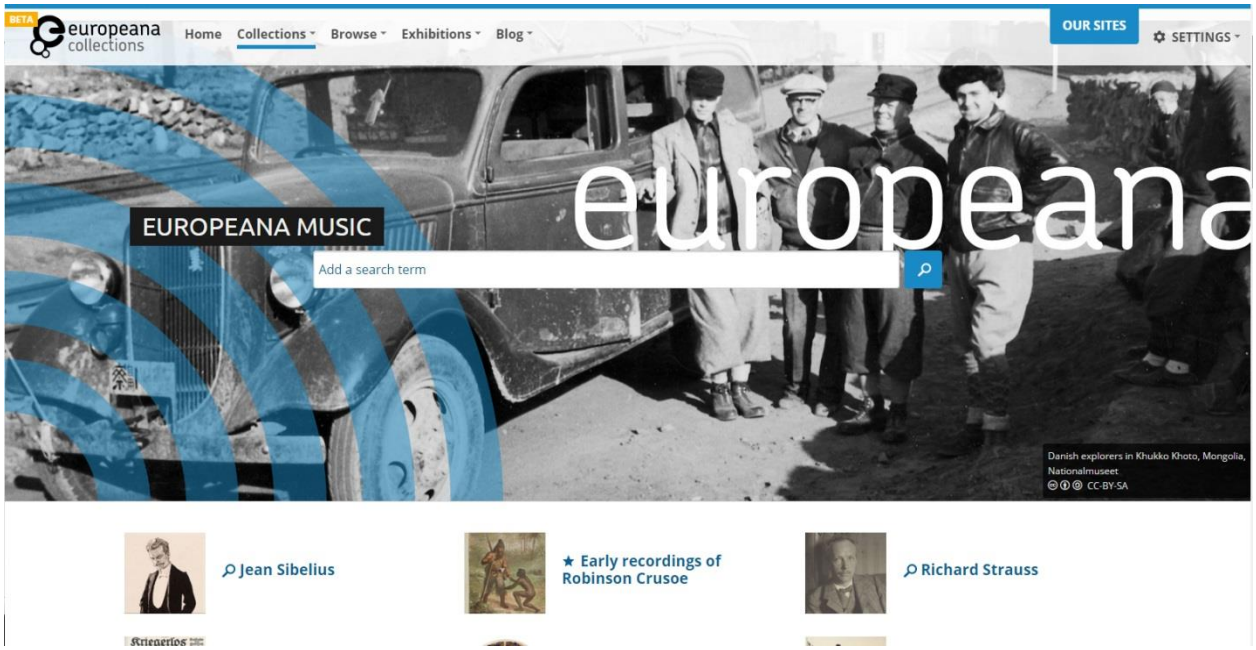


Figure 8: homepage of the Europeana Music Collections with hero image and browse entry points below

2.2 Year 1 performance indicators and targets

As a reminder, table 2 below outlines the amount of aggregation required by the project by the end of each year. Table 3 outlines the actual progress made during Year 1.

Table 2: list of Key Performance Indicators (KPIs) for WP1 to be reached at the end of each year of the project

Indicator no	Relating to objective / result	Indicator name	Expected progress		
			Year 1	Year 2	Year 3
1	Aggregation (WP1)	Number of audio items aggregated	50,000	250,000	500,000
2	Aggregation (WP1)	Number of other items aggregated	30,000	90,000	225,000
3	Aggregation (WP1)	Number of items freely available for re-use	10,000	40,000	90,000
4	Aggregation (WP1)	Number of Data Providers using new EDM profile	50%	100%	100%
5	Aggregation (WP1)	Number of consortium partners to have made use of training sessions	33%	66%	100%

Table 3: WP1 KPIs and actual progress in Year 1 (compare table 4)

Indicator no	Relating to objective /result	Indicator name	Expected progress Year 1	Actual progress Year 1
1	Aggregation (WP1)	Number of audio items aggregated	50,000	52,439
2	Aggregation (WP1)	Number of other items aggregated	30,000	31,682
3	Aggregation (WP1)	Number of items freely available for re-use	10,000	39,571
4	Aggregation (WP1)	Number of Data Providers using new EDM profile [ref 7]	50%	89%
5	Aggregation (WP1, WP5)	Number of consortium partners to have made use of training sessions	33%	100%

2.3 Metadata ingestion plan

2.3.1 Explanation of statistics

According to D1.5 *Aggregation Report 1* [Ref.1]:

We define “items aggregated” as fulfilling these requirements:

- EDM Sounds mandatory fields are present in the metadata, or can be added in MINT when appropriate (e.g. genre concepts from the new vocabulary)
- a rights statement from the *Europeana Available Rights Statements*[Ref 2] has been determined
- links to digital objects are present and working, and the objects are accessible worldwide, without a password

In addition to the above requirements being met:

- The metadata is now viewable on the Europeana or, has been harvested from MINT by Europeana and will be published at the earliest opportunity

In deliverable D1.5: *Aggregation Report 1*, it was explained that KPIs 1-3 were met in the following ways:

- the aggregation of metadata published on MINT;
- already published on Europeana;
- or harvested by Europeana but not yet ingested.

One recommendation from Year 1's technical review was that, as the main aim should be to make metadata publicly available on Europeana the aggregation KPIs as defined above did not measure publication on Europeana but only metadata aggregated. We therefore report here on both measures for aggregation on MINT and publicly accessible on Europeana.

At the end of Year 2, the metadata published on MINT was harvested by Europeana on 3 February 2016 and published on Europeana on 5 February 2016. The statistics gathered for this report have been compiled from records accessible from Europeana, as well as metadata aggregated on MINT.

In reviewing statistics the following should be considered:

- Some data records on Europeana provide access to multiple digital objects – for example, a manuscript of a symphony can link back to the institutional website where one hundred digital images can be found (Figure 9 and Figure 10)
- Some data records on Europeana which, for example, refer to one sound recording, contain additional, related digital objects. These can be musical transcriptions of the sound, transcriptions of an interview, photographs of the contributors or, in the case of commercial discs, photographs of the disc labels which provide more information. These objects have been entered in the edm:hasView and edm:object fields (photographs of disc labels make useful, unique thumbnails).

Because of this, the figures seen on the Europeana portal may not represent the total number of the audio and audio-related media actually accessible via Europeana.

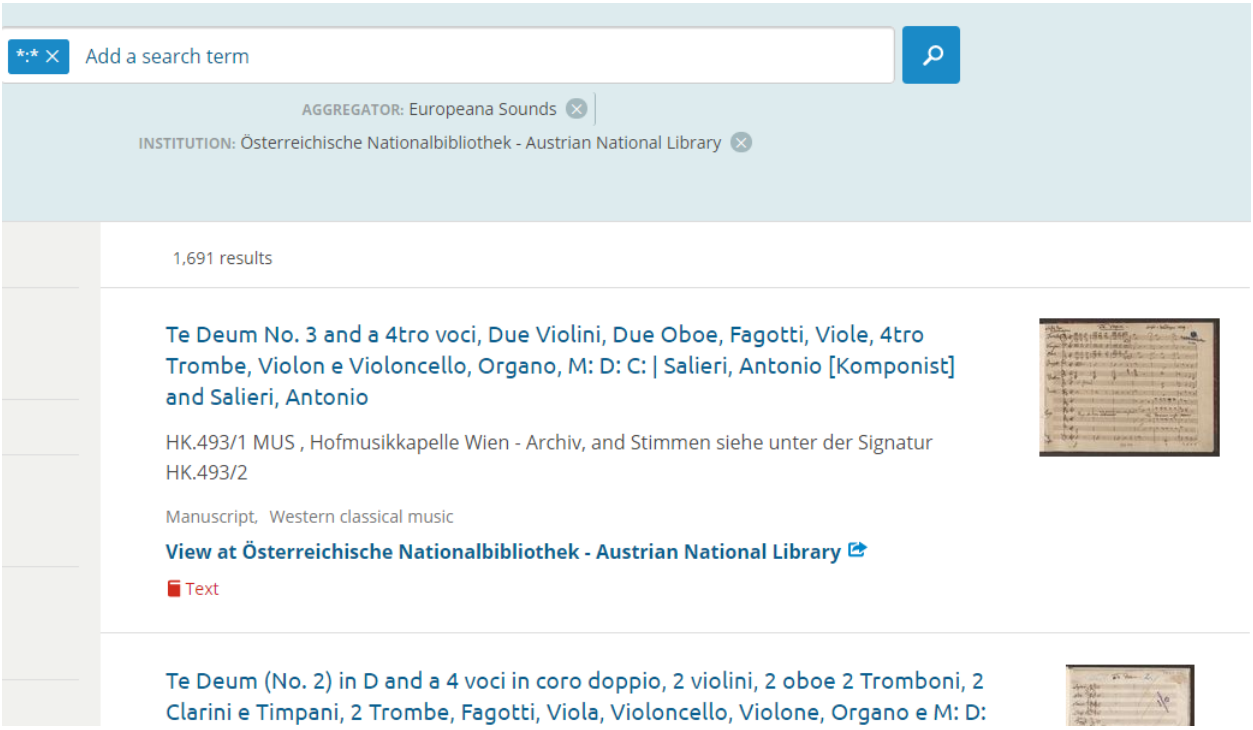


Figure 9: Search results on Europeana from the Österriechische Nationalbibliothek - Austrian National Library (ONB)

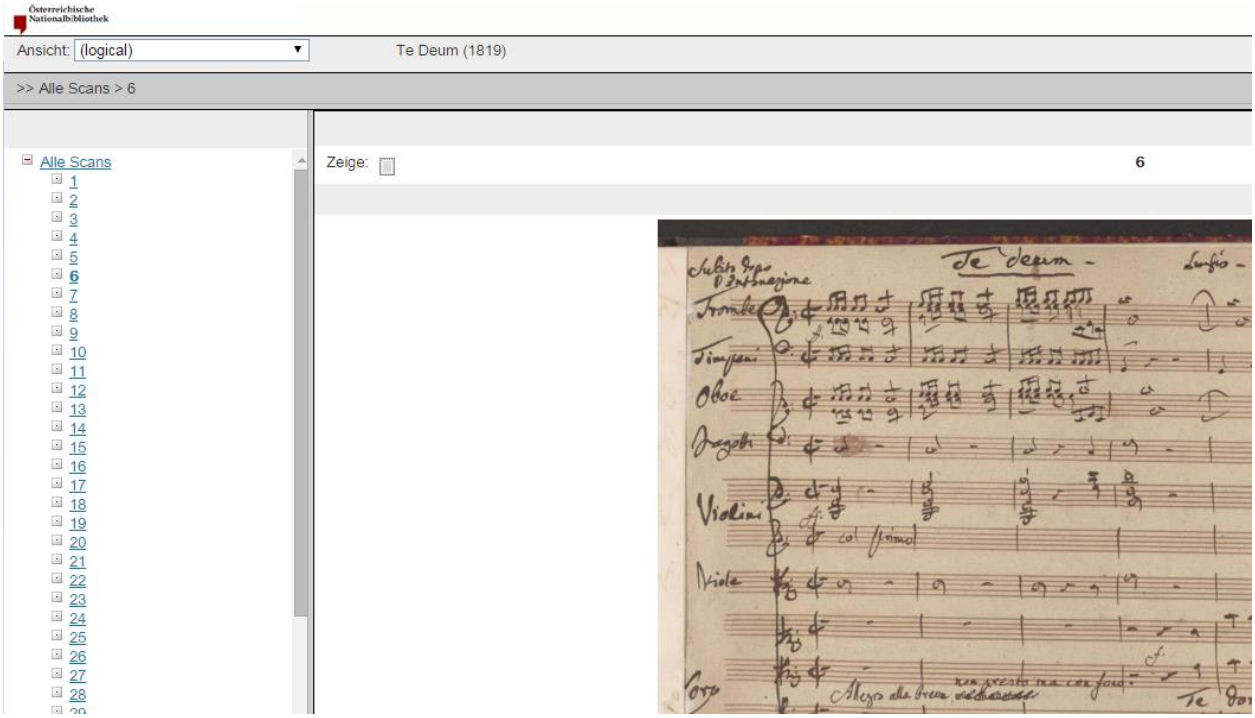


Figure 10: ONB's corresponding web page from the table above, showing multiple pages of a music manuscript

2.3.2 Statistics and ingestion planning

At the start of Year 2, the aggregated metadata had been published on MINT according to plan but was not visible on Europeana. It was difficult to check progress on the Europeana portal and the WP1 lead referred to reports on MINT to analyse how partners were progressing. This progress of publication on MINT was plotted against a backdrop of the rate required to meet the Year 2 targets (Figure 11).

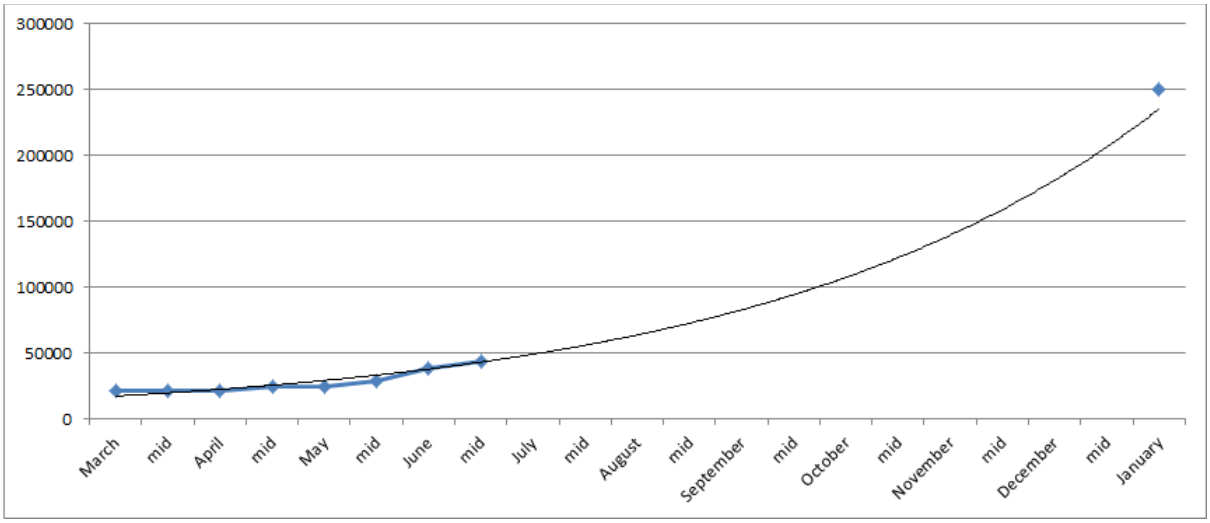
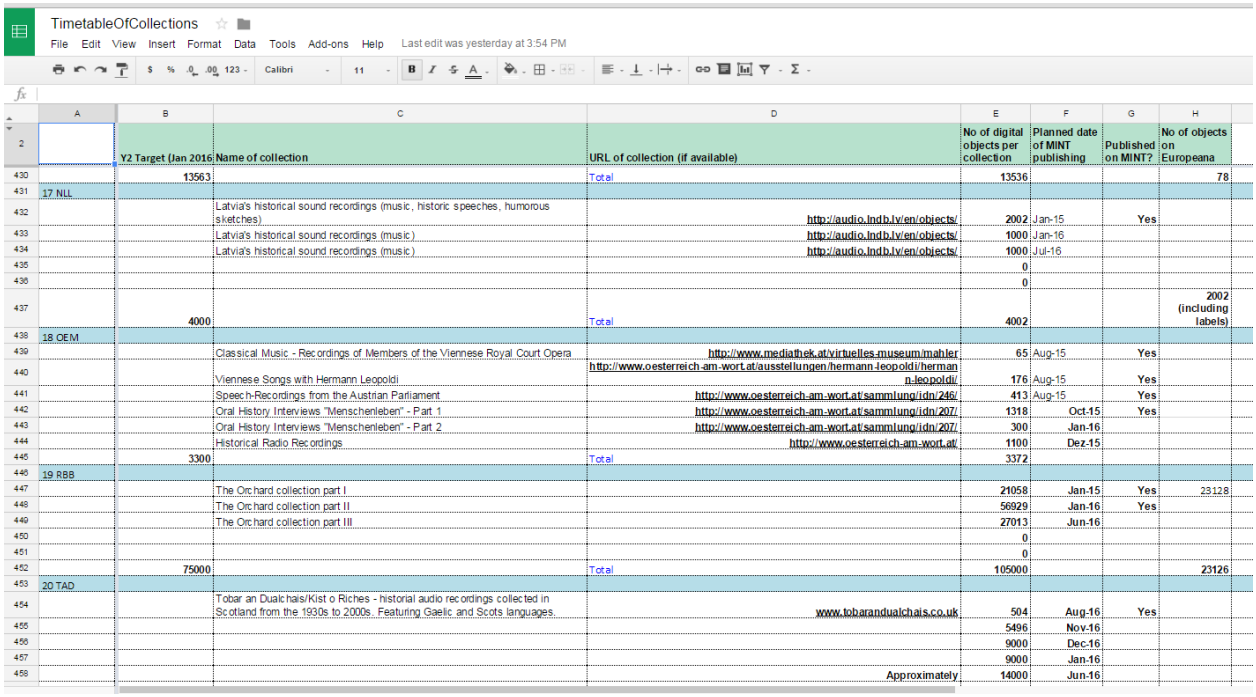


Figure 11: Year 2 progress after Q5, showing actual progress (blue line) against required rate of progress (black line) to meet KPI1

The curve, along which the initial publication figures were plotted, was the rate of progress required in order to meet the Year 2 targets from July 2015 to January 2016. However, the curve by no means predicted that the targets would be reached. Nor did the monthly publications of individual Data Providers predict a steady rate of progress, an exponential rate of progress or whether the providers would publish all their metadata in one go.

It was therefore necessary to be more aware of the Data Providers' intentions throughout the remainder of the year. At the MINT Training Workshop in June 2015, the WP1 lead asked Data Providers to plan, on a shared document, which datasets they intended to publish at what point in the year (fig.12). This was for several reasons:

- Data Providers often needed to carry out preparatory work on their datasets before uploading to MINT. A lack of publication on MINT did not mean a lack of progress
- The WP1 Lead was able to predict when more datasets were going to be published on MINT and whether this meant that KPIs would be met
- Data Providers were encouraged to plan their work for the entire year
- Data Providers were encouraged to see themselves as members of a project team rather than individual institutions.



TimetableOfCollections

File Edit View Insert Format Data Tools Add-ons Help Last edit was yesterday at 3:54 PM

	A	B	C	D	E	F	G	H
2		Y2 Target (Jan 2016	Name of collection	URL of collection (if available)	No of digital objects per collection	Planned date of MINT publishing	Published on MINT?	No of objects on Europeana
430		13563		Total	13536			78
431	17 NLL							
432			Latvia's historical sound recordings (music, historic speeches, humorous sketches)	http://audio.lndb.lv/en/objects/	2002	Jan-15	Yes	
433			Latvia's historical sound recordings (music)	http://audio.lndb.lv/en/objects/	1000	Jan-16		
434			Latvia's historical sound recordings (music)	http://audio.lndb.lv/en/objects/	1000	Jul-16		
435					0			
436					0			
437		4000		Total	4002			2002 (including labels)
438	18 OEM							
439			Classical Music - Recordings of Members of the Viennese Royal Court Opera	http://www.mediathek.at/virtuelles-museum/mahler	65	Aug-15	Yes	
440			Viennese Songs with Hermann Leopoldi	http://www.oesterreich-am-wort.at/ausstellungen/hermann-leopoldi/hermann-leopoldi/	176	Aug-15	Yes	
441			Speech-Recordings from the Austrian Parliament	http://www.oesterreich-am-wort.at/sammlung/idn/246/	413	Aug-15	Yes	
442			Oral History Interviews "Menschenleben" - Part 1	http://www.oesterreich-am-wort.at/sammlung/idn/207/	1318	Oct-15	Yes	
443			Oral History Interviews "Menschenleben" - Part 2	http://www.oesterreich-am-wort.at/sammlung/idn/207/	300	Jan-16		
444			Historical Radio Recordings	http://www.oesterreich-am-wort.at/	1100	Dez-15		
445		3300		Total	3372			
446	19 RBB							
447			The Orchard collection part I		21058	Jan-15	Yes	23128
448			The Orchard collection part II		56929	Jan-16	Yes	
449			The Orchard collection part III		27013	Jun-16		
450					0			
451					0			
452		75000		Total	105000			23126
453	20 TAD							
454			Tobar an Dualchais/Kist o Riches - historial audio recordings collected in Scotland from the 1930s to 2000s. Featuring Gaelic and Scots languages.	www.tobarandualchais.co.uk	504	Aug-16	Yes	
455					5496	Nov-16		
456					9000	Dec-16		
457					9000	Jan-16		
458				Approximately	14000	Jun-16		

Figure 12: a screenshot of the shared timetable showing Data Providers' planning of metadata publications

In the ensuing months, further statistics showed that the project was following the required rate of progress and, if Data Providers kept to their scheduled plans, the aggregation targets would be met (Figure 13).

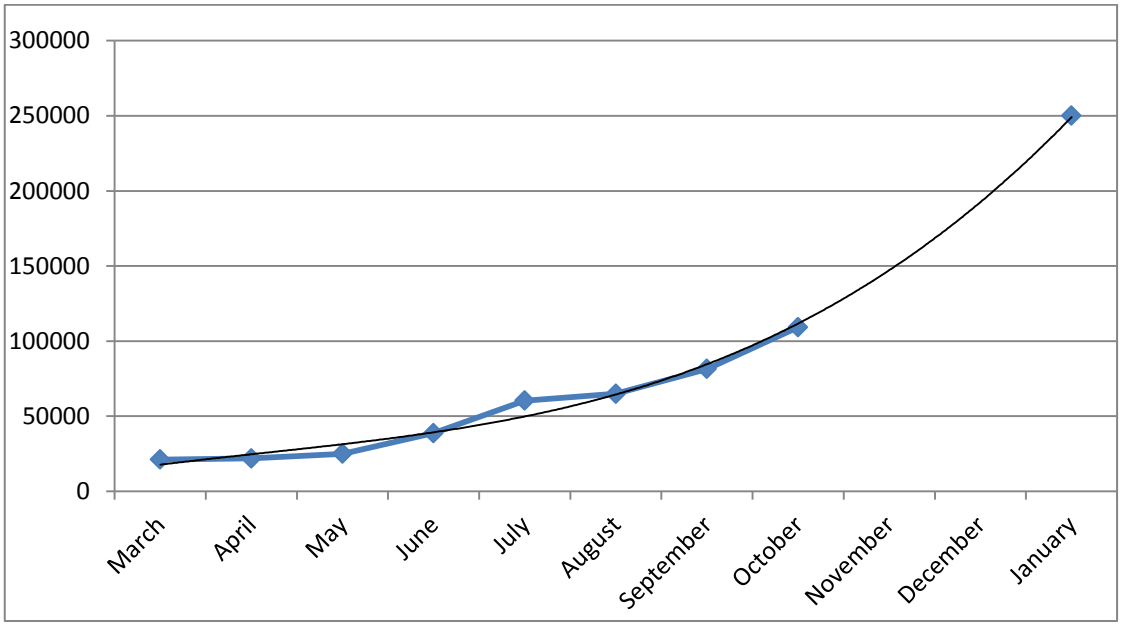


Figure 13: Aggregation progress after Q7, showing actual (blue line) against required rate (black line)

By the beginning of Q8, it was possible to predict with more confidence that the partners would achieve their overall key performance indicators; by the end of Year 2 it was known that the KPIs would be met, once several key datasets were published on MINT and subsequently harvested and ingested by Europeana (Figure 14).

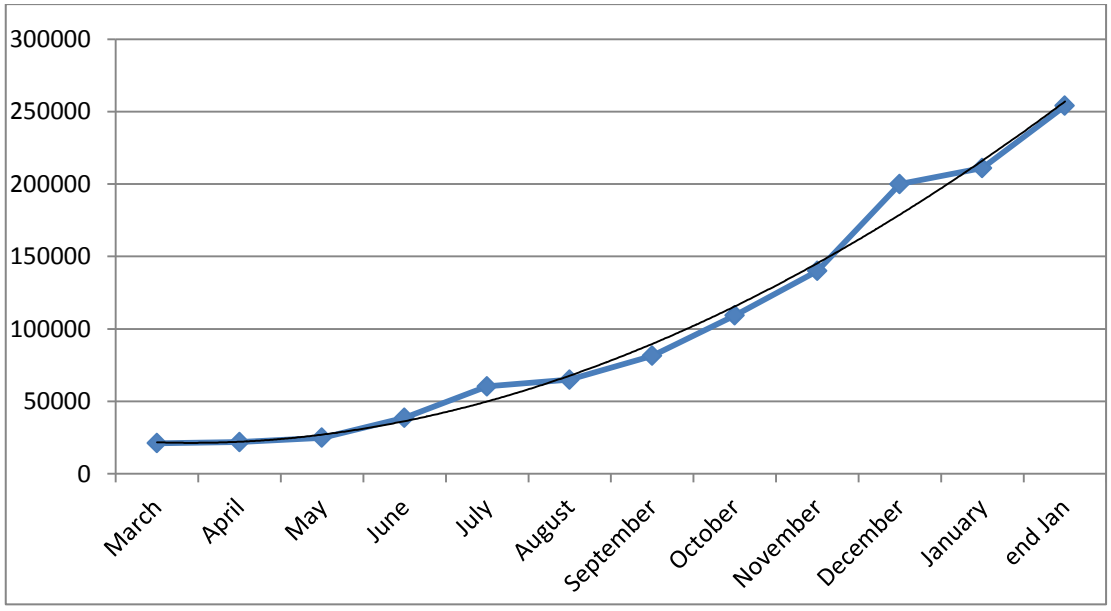


Figure 14: Aggregation progress at the end of Year 2 (Q8) showing actual progress (blue line) against required progress (black line)

3 Progress during year 2, February 2015-January 2016

3.1 Target and results

3.1.1 Performance indicator 1

Key Performance Indicator 1 target: 250,000 audio items aggregated by end of Year 2.

On the Europeana portal at the time of writing, there are 253,847 records for sound contributed by the Europeana Sounds project. These records link either to a sound directly on the isShownBy field or, link to the webpage from where the sound can be played via the institution's webpage on the isShownAt field.

Of these 253,847 items, 3,000 of them refer to the "etree" dataset, contributed by the Netherlands Institute of Sound & Vision (NISV), which is arranged hierarchically. That is, each item comprises several sounds, linked to in the hasView field, in addition to the sound linked to in the isShownBy field. In total, there are 44,500 sounds in the etree dataset accessible from Europeana.

Therefore the total number of sounds aggregated by Europeana Sounds in Year 2 is in excess of 295,000.

Status: KPI1 met.

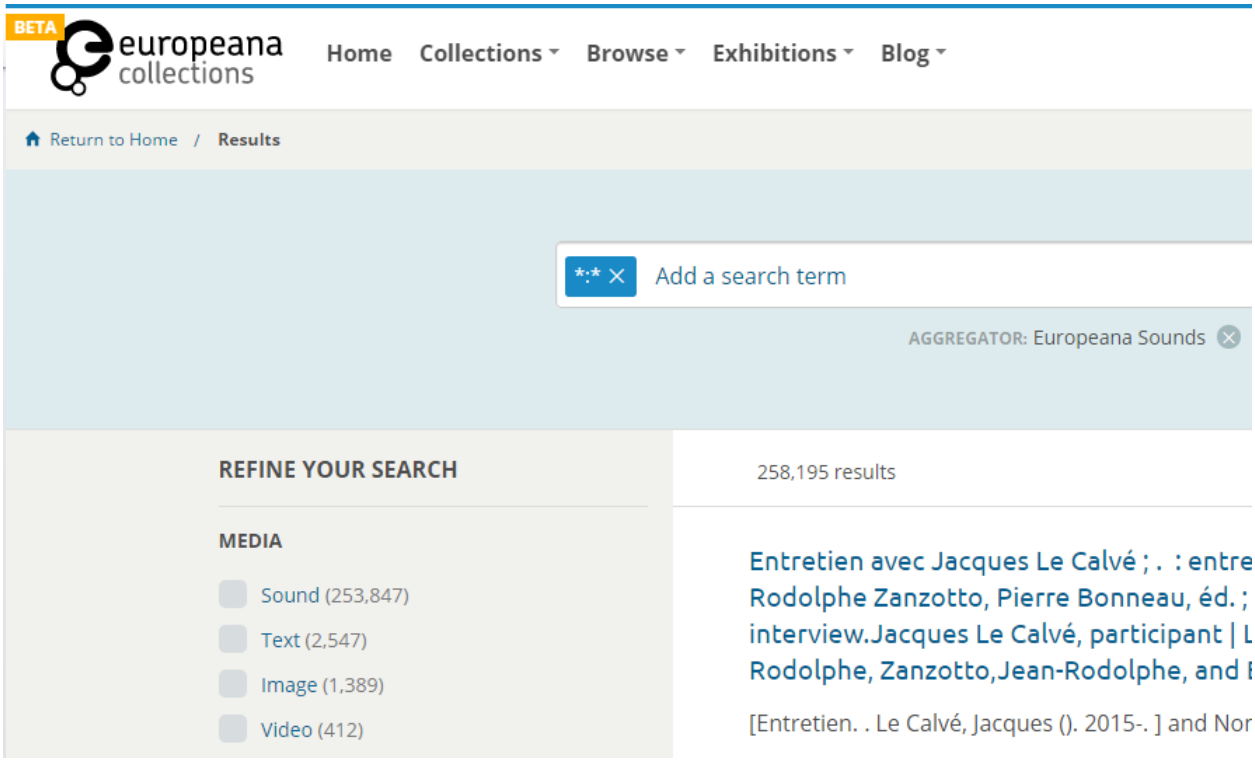


Figure 15: screenshot of records aggregated on Europeana by Europeana Sounds

3.1.2 Performance indicator 2

Key Performance Indicator 2 target: 90,000 other items aggregated by end of Year 2.

On the Europeana portal, there are 2,547 text records, 1,389 image records and 412 video records, aggregated through the Europeana Sounds project. Of these 2,547 texts, 1,691 are from the Austrian National Library (ONB). These records are generally of musical manuscripts and include autograph scores that link back to the ONB's website on the isShownAt field, from where the digitised images of the entire scores can be seen. This is the entire contribution from ONB and it represents 152,977 audio-related images aggregated by Europeana Sounds.

There are also media resources which enhance sounds and images. As outlined in 2.3.1, these can be photographs of contributors, musical transcriptions of recordings, or some other complementary resource. A search of the Europeana Sounds portal gives 13,738 records as having these resources in their metadata.

Finally, there are records which have unique objects that are used as thumbnails for sound recordings. It should be noted that this figure excludes the waveforms supplied by the British Library (BL) and CNRS. Most of these objects are photographs of the disc labels which provide important information about recordings that cannot be replicated through text description.

In total, therefore (see table 6), the number of audio-related items aggregated by Europeana Sounds is in excess of 185,000.

Status: KPI2 met.

3.1.3 Performance indicator 3

Key Performance Indicator 3 target: 40,000 items freely available for re-use aggregated by end of Year 2.

As defined in the glossary (section 8) below, items freely available for re-use are digital objects that are available for re-use with minimal or no conditions, specifically those objects labelled Public Domain, CC0, CC-BY and CC-BY-SA.

All 152,977 images from the Austrian National Library, aggregated by Europeana Sounds, are in the public domain, and therefore can be re-used. In addition to this, there are 18,649 items which can be reused with attribution.

For the remainder of the records on Europeana Sounds, 16,907 can only be used with restrictions and 220,948 require permission in order to be reused.

The total number of reusable items is 171,626.

Status: KPI3 met.

3.1.4 Performance indicator 4

Key Performance Indicator 4: 100% partners to have made use of new EDM profile by end of Year 2

Status: KPI4 met.

3.1.5 Performance indicator 5

Key Performance Indicator 5: 66% of partners to have made use of training sessions by end of Year 2

All consortium partners have now attended either the first or second MINT training course in Athens in October 2014 [Ref 8] and June 2015 [Ref 9].

Status: KPI5 met.

3.2 Summary of progress

The project has met or exceeded all its year 2 KPIs (table 4).

Table 4: WP1 KPIs and actual progress in Year 2 (compare table 3)

Indicator no	Relating to objective /result	Indicator name	Expected progress Year 2	Actual progress Year 2
1	Aggregation (WP1)	Number of audio items aggregated	250,000	295,000
2	Aggregation (WP1)	Number of other items aggregated	90,000	185,000
3	Aggregation (WP1)	Number of items freely available for re-use	40,000	171,626
4	Aggregation (WP1)	Number of Data Providers using new EDM profile [ref 7]	100%	100%
5	Aggregation (WP1, WP5)	Number of consortium partners to have made use of training sessions	66%	100%

Most Data Providers have met or exceeded their individual targets; however some have not met them and will need a more focused approach in order to reach their final targets by the end of July 2016 (Milestone 6).

The main reasons for this were:

- Unexpected extra work required in preparing metadata
- Unexpected rights issues meaning that certain datasets couldn't be made available
- Lack of familiarity in operating MINT

Data providers who have not reached their targets are working with WP1 lead on contingency plans to ensure they meet the Year 2 shortfall and, also, meet the remainder of their aggregation targets by milestone 6 at the end of July 2016.

3.2.1 Aggregation on MINT

Table 5 provides details of:

- records imported to MINT by each provider (in either .csv or XML form)
- records transformed to the EDM/EDM Sounds profile
- records prepared for publication – these records are the ones ready for harvesting and ingestion by Europeana

Table 5: Aggregation of metadata on MINT by each Data Provider

Data Provider	Records imported to MINT	Records aggregated (transformed to the EDM/EDM Sounds profile)	Records published on MINT
01 BL	43,730	81,128	77,627
02 NISV	60,794	60,416	6,310
06 BNF	15,117	14,616	0
10 CNRS	119,641	156,160	83,937
11 DIZI	47,042	82,252	29,814
12 DNB	5,008	1,052	1,052
13 FMS	1,923	2,826	3,302
14 ICCU	45,217	45,177	90,354
15 ITMA	3,495	6,958	6,958
16 TLA	2,061	4,122	4,122
17 NLL	3,710	5,375	3,368
18 OEM	7,927	15,808	15,808
19 RBB	101,402	198,090	155,974
20 TAD	62,684	97,900	51,548
21 SB	2,099	4,015	3,978
22 ONB	5,073	5,073	3,382
23 FCSH	1,154	1,348	408
24 CCE	14,600	29,153	29,048
Total	**542,677**	**811,469**	**566,990**

It should be noted that the "records published on MINT" include metadata both in EDM and EDM Sounds format; this means that the number of records published on MINT is roughly twice the number eventually published on Europeana (see Table 6: contributions of Data Providers to Europeana with individual Year 2 targets).

3.2.2 Publications on Europeana

Table 6 provides details of:

- the number of metadata records published on Europeana, for audio and non-audio
- the number of digital objects described by these metadata records
- the individual Year 2 targets for each Data Provider, set internally by the project

Table 6: contributions of Data Providers to Europeana with individual Year 2 targets

Data Provider	Audio records published on Europeana	No. of audio objects on Europeana	Year 2 target for audio items on Europeana	No. of text, image and video records on Europeana	No. of text, image and video objects published Europeana	Year 2 target for text, image and video objects on Europeana
01 BL	38,710	38,710	39,000	564	5,072	35,885
02 NISV	3,055	44,500	40,000	100	100	0
06 BNF	3,731	3,731	5,100	0	3,731	5,000
10 CNRS	9,674	9,674	26,000	0	0	26,000
11 DIZI	10,234	10,234	10,000	0	0	0
12 DNB	526	526	511	0	0	0
13 FMS	1,233	1,233	200	0	3,570	0
14 ICCU	44,581	44,581	22,000	0	16,749	0
15 ITMA	1,582	1,582	1,000	1,878	1,878	2,400
16 TLA	6,328	6,328	13,563	0	0	0
17 NLL	1,684	1,684	2,000	0	1,684	2,000
18 OEM	7,670	7,670	3,300	0	0	0
19 RBB	84,064	84,064	75,000	0	0	0
20 TAD	25,270	25,270	25,334	0	0	0
21 SB	1,989	1,989	1,900	0	0	0
22 ONB	0	0	0	1,691	152,977	98,000

23 FCSH	89	89	22,000	115	115	2,700
24 CCE	13,427	13,427	9,500	0	0	0
Total	**253,847**	**295,292**	**296,408**	**4,348**	**185,876**	**171,985**

Figure 16: actual achievement against individual targets (audio) and Figure 17: actual achievement against individual targets (non-audio) show how Data Providers have performed against their Year 2 targets for audio and non-audio items respectively.

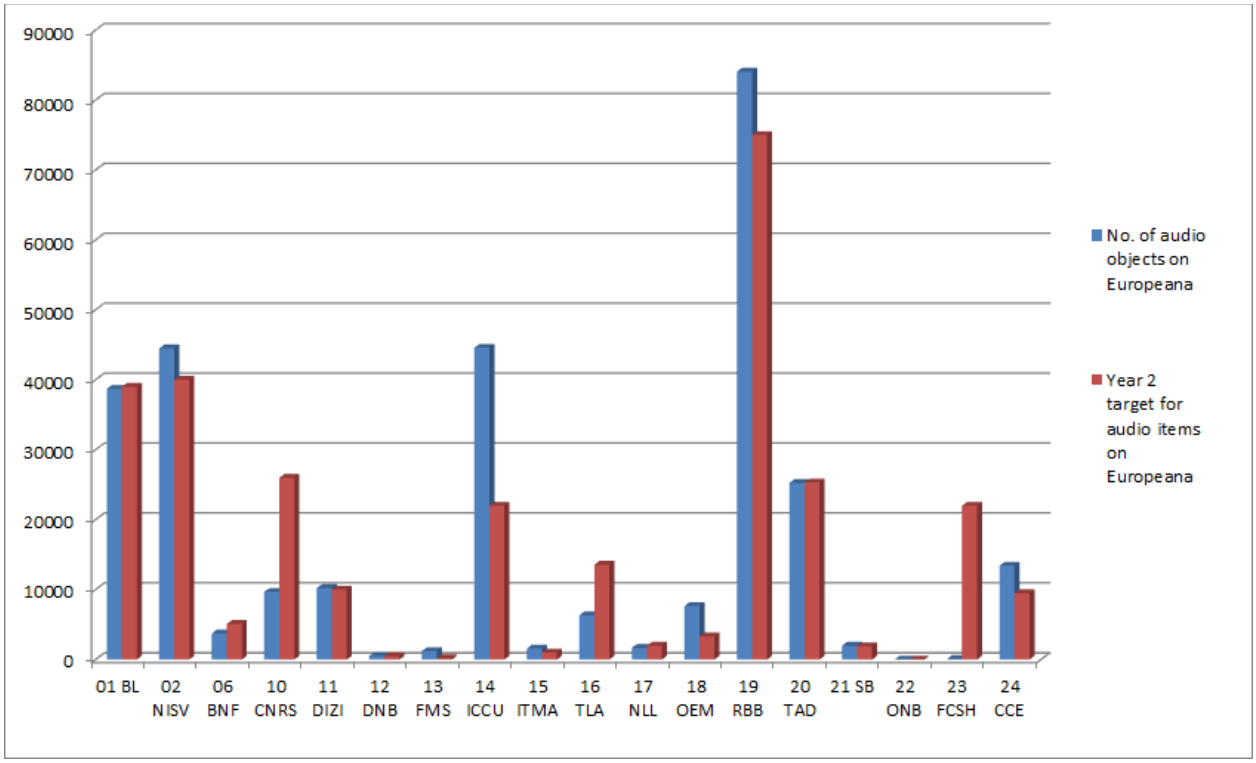


Figure 16: actual achievement against individual targets (audio)

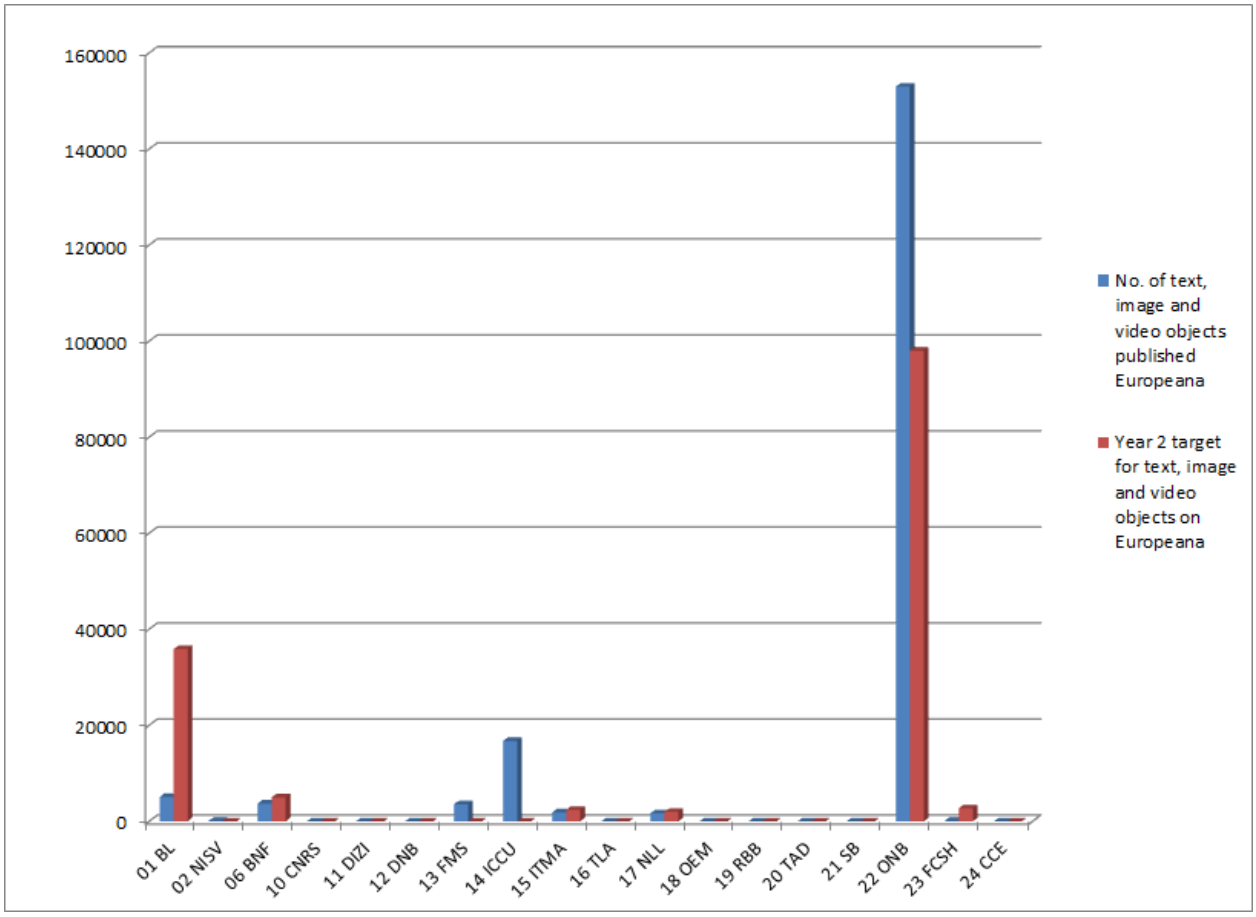


Figure 17: actual achievement against individual targets (non-audio)

4 Identification and assessment of ingestion challenges

4.1 Impact of technical requirements on selection of content

4.1.1 Directly linked digital objects

A mandatory element of the Europeana Music Collections (developed in WP4 of this project) has been that all metadata featuring on it must contain a link to the CHO's digital object in the isShownBy field, as well as or in place of a link to the providing institution's webpage in which the digital object is embedded. The reason for doing this is to provide a more positive user experience for the Europeana visitor, because it allows the visitor to consult the digital object directly on the Music Collection.

The added advantages for directly linked digital objects is that there is more scope for enriching data automatically when digital objects are machine-accessible. For example, the technical characteristics of sounds can be analysed for Music Information Retrieval (developed in WP2 of this project).

Therefore, providing a significant amount of directly-linked objects on Europeana will be instrumental in reaching several of the project's key performance indicators.

However, in the Description of Work for Europeana Sounds, it was not a mandatory requirement for Data Providers to supply directly linked objects. At the start of the project, Data Providers made a choice:

- to supply a link to their webpage in the “isShownAt” field
- to supply a direct link to their digital object in the “isShownBy” field
- both of the above

Discussions followed about how Data Providers could make more directly linked content available. The subject was discussed further at the Europeana Sounds’ plenary session in Lisbon (January 2016). There were several reasons given for not providing directly linked media:

- Legal: due to licensing restrictions, the institution’s digital objects were not accessible except when presented on the Data Provider's own platform. (For example, most of the content at the British Library). One advantage of providing content in this way is that more can be given access within a single portal than when shared more widely through other portals such as Europeana.
- Technical: due to the way in which the institution’s website is set up, it would be complicated and costly to redesign the technical solution to allow direct links to digital objects
- Data Providers did not realise both isShownAt and isShownBy fields could be used together in a single recordand thought they had to choose between the two; more work would be required to revise datasets by adding links in the isShownBy fields.
- Data Providers and/or rights owners wish users to find content on Europeana and then visit their own website where more contextual information is given.

However, there is an understanding that, where possible, direct links should be used as this will enhance the user’s visit to Europeana and not require too many clicks to arrive at the content they seek - especially if this content is reusable. More work on this topic will take place in Year 3 of the project.

4.2 Changes to technical infrastructure

4.2.1 Europeana publication timetable

Up until the end of 2015, Data Providers were required to publish their datasets in instalments, by the 21st of each month. These datasets were harvested by Europeana at the end of each month; more processing was then required and this meant that the datasets might be published on Europeana at least a month after the work was completed on MINT. As a result of this time lag, it was difficult to review and revise datasets on Europeana. Towards the end of the year this publication process became far swifter and this meant that the number of MINT publications - and, also, alterations of previous publications - could be doubled to twice a month. Therefore, Data Providers could review their work a week after publishing on MINT and any revisions could take place swiftly.

4.3 Rights and licensing

All content aggregated by Europeana Sounds needed to be rights cleared to be streamed or downloaded from the internet worldwide. Data providers have been working on ensuring that certain previously

restricted datasets will be made available; so, for example, datasets will be accessible without the visitor requiring a password.

When mapping data in MINT, it is mandatory for Data Providers to select a relevant license (public domain, CC0, CC-BY, etc.) from a drop-down list. These options create links to the Creative Commons website so that it is clear to the user what the license of a sound recording or other media is and how, if at all, the content can be reused.

Some digital media, aggregated onto Europeana, might only be accessible by being streamable but not downloadable; however, through contacting rights holders it may be possible to change this and thus provide a direct link from Europeana to the digital media (the benefits of this are discussed in 4.1.1).

4.4 Development and maintenance of new genre concepts vocabulary

The development and maintenance of a new genre concepts vocabulary was outlined in Deliverable D1.3 *Ontologies for Sound*. The two-tiered genre concept is now a mandatory field in MINT: Data Providers must select a genre and sub-genre for datasets from a drop-list in order for MINT to transform into EDM Sounds.

In collaboration with WP2 *Enrichment* Data Providers have been working with other consortium partners to see how metadata could be enhanced automatically or semi-automatically. For example, Cultuurlink [Ref 3] (an online toolkit developed by NISV) will examine a dataset containing names of musical instruments. Cultuurlink will then link these names with instrument names from the Musical Instrument Museums Online (MIMO) project [Ref 4]. The user can confirm or reject these links. Where links are accepted, music recordings can be linked more closely with the images from the MIMO project. Similarly, Pundit (developed by Net7 [Ref 5]) can offer the user the possibility of annotating records on Europeana and providing more contextual information. This is discussed in more detail in D2.4 [Ref 6].

Year 3 will consider whether it is possible to maintain a deeper music vocabulary on MINT. Doing so would make it simpler to apply musical terms to different datasets.

4.5 Timeframe for Year 3 aggregation

The next six months will require Data Providers to publish the remainder of their required metadata, by month 30 (see 5.3.2). There will also be scope for Data Providers to revise their metadata – this activity can continue to the end of the project.

4.6 Tracking progress

Quarterly, narrative reports were compiled by the WP1 Lead throughout Year 2. Data Providers were invited to outline what they had achieved at the end of each quarter, detailing highlights, difficulties encountered and, where necessary, giving reasons for not achieving planned results. Data Providers also outlined what work they planned for the following three months.

5 Lessons learned and looking ahead

5.1 Lessons learnt

5.1.1 Europeana portal search options

The technical development of the Europeana portal needs to continue so that the metadata can be navigated more easily. Some of the statistics provided in this document - such as the number of unique digital resources linked in the hasView field - were only obtained through searching Solr at the Europeana offices, rather than searching the Europeana portal.

5.1.2 Linking of audio and audio-related material

There is scope for more matchings between audio-related material and audio. Music recordings could be linked more closely with musical instruments. A search on Europeana may not bring up the records most closely related to each other. For example, Figure 18 shows a recording of an ennanga, aggregated by Europeana Sounds. Figure 19 is of a record aggregated by Musical Instrument Museums Online (MIMO) --- this shows what the instrument looks like.

But some records only have the generic name of an instrument, such as bow harps or arched harps and not the specific name. So, there is scope for linking related Europeana records, confirming names and types of musical instruments and thereby enriching records from Europeana Sounds and elsewhere. Enrichment of data (through crowdsourcing events such as editathons, etc), as part of Work Package 2, could open exciting opportunities for Europeana Data Providers and users alike.

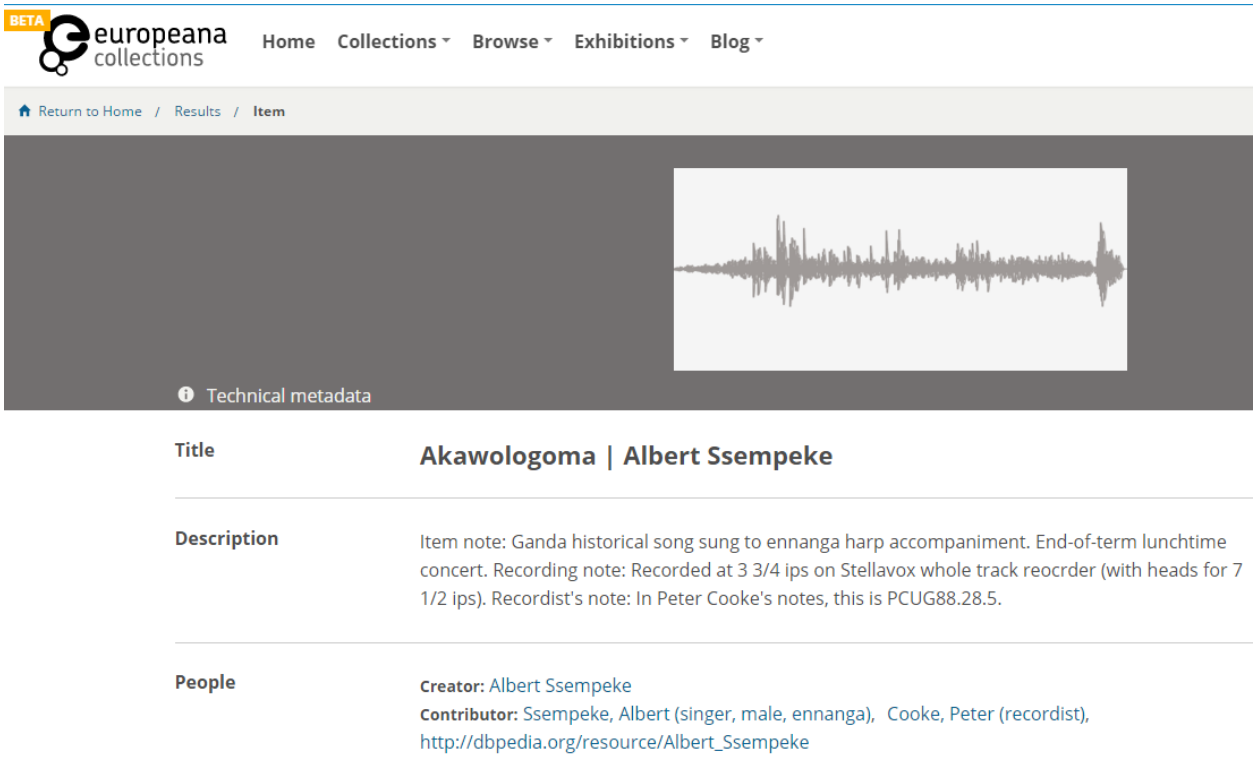


Figure 18: Europeana Sounds record of an ennanga recording from British Library

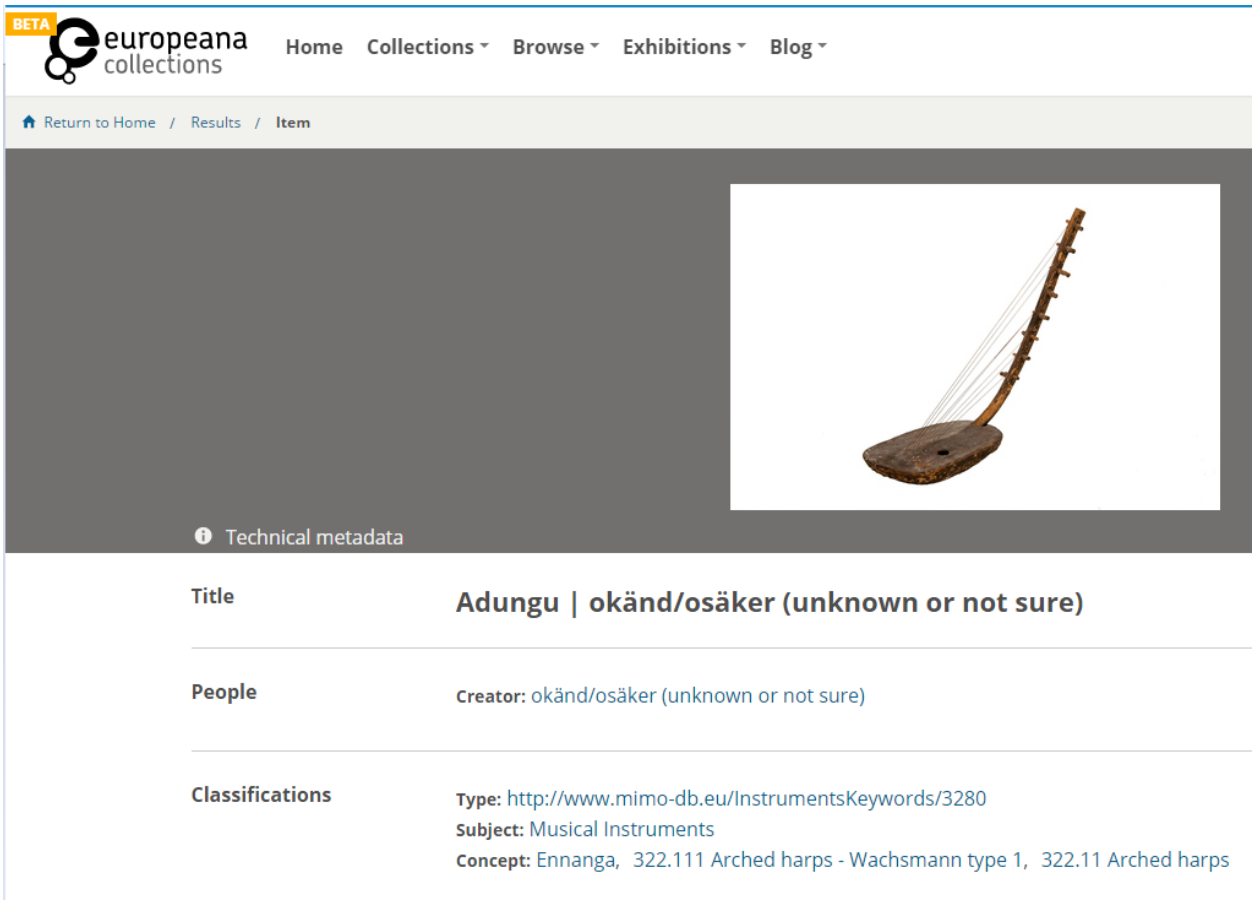


Figure 19: Europeana record from MIMO showing image of an ennanga

5.2 Looking ahead

5.2.1 Training

The third and final MINT training workshop will held be in Dublin, 28th – 29th April 2016, hosted by Comhaltas Ceoltóirí Éireann (CCE) at Clasac, CCE's music venue. Although this will take place when there are only three months of aggregation remaining before milestone MS6 (when all metadata is due to be published on MINT before 31st July 2016), this will be a vital workshop as collections will need to be revised and enhanced in collaboration with Data Enrichment (WP2), Licensing (WP3) and Music Collections (WP4).

5.2.2 Publication schedule

The final six months of metadata aggregation will be timetabled in the same way as the last nine months: Data Providers will timetable their collections on a shared document, detailing the number of items in each dataset and a collection level description of the dataset. This information is of use to the other work package leads, as it informs them which collection could be publicised (WP6), could feature in the Music Collections (WP4) or could be used for data enrichment experiments (WP2).

Table 5: publication schedule for Data Providers, Q9-Q10

Data Provider	Total audio required	No. of audio objects on Europeana	Remainder of audio required	Total text, image and video content required	No. of text, image or video objects on Europeana	Remainder of text, image and video content
01 BL	79,000	38,710	40,290	36,000	564	35,436
02 NISV	86,170	44,500	41,670	14,050	100	13,950
06 BNF	10,000	3,731	6,269	10,000	3,731	6,269
10 CNRS	37,600	9,674	27,926			
11 DIZI	10,000	10,234	-234			
12 DNB	511	526	-15			
13 FMS	2,500	1,233	1,267	6,630	3,570	
14 ICCU	80,000	44,581	35,419			
15 ITMA	1,200	1,582	-382	3,600	1,878	1,722
16 TLA	13,563	6,328	7,235			
17 NLL	2,000	1,684	316	2,000	1,684	316
18 OEM	12,000	2,945	9,055			
19 RBB	105,000	84,064	20,936			
20 TAD	38,000	25,270	12,730			
21 SB	11,900	1,989	9,911			
22 ONB				150,000	152,977	-2,977
23 FCSH	39,000	89	38,911	2,700		2,700
24 CCE	15,800	13,427	2,373			

5.2.3 Associate Partners

During the second year of the project, the consortium issued a call for Associate Partners. These Partners would not receive project funding but would receive training using MINT and would contribute content that would be aggregated by European Sounds. The first two of these partners are Koç University and the Dublin Institute of Technology. Two other potential Associate Partners are finalising agreements with the consortium.

Soundscape of Istanbul, Koç University, Istanbul, Turkey

The Soundscape of Istanbul is a research project conducted by Pınar Çevikayak Yelmi during her doctoral studies (2012-2016) at Koç University, supervised by Assoc. Prof. Nina Ergin, Assoc. Prof. Sertaç Kakı and Assoc. Prof. Asım Evren Yantaç. The project deals with the urban culture of the city of Istanbul and the sounds of everyday traditions. Sounds are integral parts of any urban environment and our daily lives. Sounds also constitute a significant aspect of cultural heritage and sonic values are crucial for cultural memory and for the sustainability of cultural identity; thus, they deserve to be protected as intangible cultural heritage elements. The Soundscape of Istanbul project aims to collect representative sounds and protect them in a publicly accessible archive, Koç University Suna Kıraç Library Digital Collections. Access the collection.

In order to expand this sound collection, the Soundsslike project was initiated by Pınar Çevikayak Yelmi which is a crowdsourced sound map that encourages people to record and upload sounds that they consider worth protecting. The intention of the Soundsslike project is to draw people's attention to urban sounds, to make them aware of their uniqueness and to create a platform in which anyone can share cultural sounds.

Dublin Institute of Technology (DIT)

The Dublin Institute of Technology (DIT) is one of Ireland's largest and most innovative university-level institutions combining the academic excellence of a traditional university with professional, career-orientated learning in a student and staff community of *c* 22,000. The DIT Conservatory of Music & Drama, which was founded by Dublin's city council in 1890, provides the conditions for musicians to develop practical, academic and research skills and facilitates cross-fertilisation of artistic and technological activity thereby enabling innovative technological, economic, social and cultural progress. Our special collections focus on the music *in* and *of* Ireland, supporting both internal and national research, and feature the historical audio collections of Frank & Bernadette Caruana and six Grammy Award winner Derek Bell.

6 Summary

The second year built on the development and methodologies and initial aggregation work carried out in Year 1: 50,000 records were ready to be published on Europeana and agreed genres were being incorporated into Data Providers' work.

The aggregation in Year 2 of Europeana Sounds has moved forwards dramatically: the metadata from Year 1 has now been published on Europeana alongside further material - over 250,000 sounds and 170,000 audio-related, digital objects. Work Package 1 of the project has therefore met all five of its Y2 key performance indicators. This has taken place alongside the development of the Europeana portal, the development of Europeana's Music Collections and, also, the streamlining of Europeana's publishing strategy. This means that the landscape of the project has changed considerably.

Although the Data Providers have been spread over the continent and have only met twice a year, they have worked as a team online through Basecamp, contributing to discussions, answering each other's queries wherever possible, rather than waiting for the MINT team to respond. A community has been built up with a growing archive of questions and answers. This should be useful for the new Associate Partners joining the project.

7 References

Ref 1	D1.5 Aggregation Report 1 http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D1.5-AggregationReport1-v1.2.pdf
Ref 2	Available rights statements (Europeana pro) http://pro.europeana.eu/page/available-rights-statements
Ref 3	Cultuurlink http://cultuurlink.beeldengeluid.nl/app/#/about
Ref 4	Musical Instrument Museums Online (MIMO) http://www.mimo-international.com/MIMO/
Ref 5	Pundit http://thepund.it/
Ref 6	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 7	D1.4 EDM profile for sound http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/EuropeanaSounds-D1.4-EDM-profile-for-sound.pdf
Ref 8	Europeana Sounds' first MINT training workshop, Athens, October 2014 http://www.europeanasounds.eu/news/aggregation-underway
Ref 9	Europeana Sounds' second MINT training workshop, Athens, June 2015 http://www.europeanasounds.eu/news/europeana-sounds-travels-to-athens

8 Appendix A: Terminology

A project glossary is provided at: http://pro.europeana.eu/web/guest/glossary.

Additional terms are defined below:

Term	Definition
AB	Advisory Board
APEX	Archives Portal Europe network of excellence
CHO	Cultural Heritage Organisation
EC-GA	Grant Agreement (including Annex I, the Description of Work) signed with the European Commission
PI	Performance Indicator
PMB	Project Management Board
TEL	The European Library
UAP	User Advisory Panel
WP	Work Package