



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

D1.5 Aggregation Report 1

Revision: Final
Date: 28/02/2015
Author(s): Andra Patterson (BL)

Abstract: This report documents the progress of Europeana Sounds aggregation activities during the first phase of task T1.4 Aggregation management, which began in August 2014. It describes the processes and achievements of Europeana Sounds data providers in aggregating their content and metadata, identifies the main challenges faced in aggregation and identifies solutions that ensure progress during and beyond this project.

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



Revision history

Version	Status	Name, organisation	Date	Changes
0.1	ToC	Andra Patterson, BL	16/01/2015	
1.0	Draft	Andra Patterson, BL	30/01/2015	
1.0	Draft	Andra Patterson, BL	13/02/2015	incorporates changes following feedback from external reviewers
1.1	Final draft	Andra Patterson, BL	24/02/2015	incorporates changes following feedback from project management board
1.2	Final	Richard Ranft, BL	27/02/2015	Layout, minor amendments

Review and approval

	Name, organisation	Date
Reviewed by	Cecile Devarenne, Europeana	13/02/2015
	Georg Eckes, Deutsches Filminstitut	13/02/2015
Approved by	Coordinator and PMB	27/02/2015

Distribution

No.	Date	Comment	Partner / WP
1	27/02/2015	Submitted to the European Commission	BL/WP7
2	27/02/2015	Posted on Europeana Pro website	BL/WP7
3	27/02/2015	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This document contains original unpublished work except where clearly indicated otherwise.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

Copyright © Members of the Europeana Sounds Consortium, 2014-2017. This work is licensed under the Creative Commons CC-BY License: <https://creativecommons.org/licenses/by/4.0/>.

Contents

Executive summary: D1.5 Aggregation Report 1	6
1 Introduction	6
1.1 Europeana Sounds data providers	6
1.2 WP1 tasks and documents providing context	7
1.3 Technical infrastructure	8
1.4 Timeline of activities and documents relating to aggregation	8
2 Ingestion objectives	10
2.1 Main stages of ingestion	10
2.1.1 Content selection	10
2.1.2 Metadata preparation	10
2.1.3 Metadata ingestion	11
2.1.4 Metadata curation	12
2.2 Performance indicators and targets	12
2.3 Metadata ingestion plan	13
2.3.1 Explanation of statistics	13
2.3.2 Statistics for each data provider	13
3 Progress during year 1, August 2014-January 2015	14
3.1 Comparison of targets with results	14
3.1.1 Performance indicators 1-2	15
3.1.2 Performance indicator 3	15
3.1.3 Progress achieved by each data provider on indicators 1-3	16
3.1.4 Performance indicator 4	19
3.1.5 Performance indicator 5	19
3.2 Summary of progress	19
4 Identification and assessment of ingestion challenges	20
4.1 Impact of technical requirements on selection of content	20
4.2 Changes to technical infrastructure	20
4.3 Rights and licensing	20
4.4 Development and maintenance of new genre concepts vocabulary	21
4.5 Timeframe for Year 1 aggregation	21
4.6 Documentation	22
5 Concluding observations and looking ahead	22
5.1 Concluding observations	22
5.2 Looking ahead	23
5.2.1 Training	23
5.2.2 Publication schedule	23
5.2.3 Changes to selection policy	24

6	Summary	24
7	References	25
8	Appendix A: Terminology	25

Executive summary: D1.5 Aggregation Report 1

This report documents the progress of Europeana Sounds aggregation activities during the first phase of task T1.4 *Aggregation management*, which began in August 2014. It describes the processes and achievements of Europeana Sounds data providers in aggregating their content and metadata, identifies the main challenges faced in aggregation and identifies solutions that ensure progress during and beyond this project.

This report complements earlier deliverables in Work Package 1 *Aggregation* (WP1), especially D1.1 *Content selection policy* [Ref 1] and D1.4 *EDM profile for sound* [Ref 2], which describe the content selection guidelines and the metadata schema used during ingestion.

Section 1 is an introduction providing the context of aggregation within the project, with an outline of WP1 tasks and documents, a brief description of the technical infrastructure developed by WP5 and a timeline of activities and documents relating to aggregation. Section 2 looks at the ingestion objectives, including the performance indicators and metadata ingestion plan for the project. In Section 3 the actual progress against performance indicators is presented. Ingestion challenges are identified and assessed in Section 4 followed by concluding observations and a brief outline of the training plan and the publication plan in Section 5.

1 Introduction

Europeana Sounds gathers together a large body of metadata for audio and audio-related material and makes it accessible through Europeana. The core tasks of WP1 Aggregation are:

- to ingest metadata describing 500,000 audio objects and 225,000 audio-related objects into Europeana
- to establish a best practice model for future aggregation [Ref 3: Part A page 6].

This report documents the progress of Europeana Sounds aggregation activities during the first phase of T1.4 *Aggregation management*, which began in August 2014. It describes the processes and achievements of Europeana Sounds data providers in aggregating their content and metadata, identifies the main challenges faced in aggregation and identifies solutions that ensure progress during and beyond this project.

This report complements earlier reports on WP1 deliverables, especially D1.1 *Content selection policy* [Ref 1] and D1.4 *EDM profile for sound* [Ref 2], which describe the content selection guidelines and the metadata schema used during ingestion.

1.1 Europeana Sounds data providers

The 18 data providers contributing to the project comprise a range of institutions from across Europe, including national libraries, research institutes, broadcasters, music libraries and media centres. These

data providers will become centres of expertise for other repositories, expanding the network and stimulating future aggregation.

Table 1: Europeana Sounds data providers

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

1.2 WP1 tasks and documents providing context

This brief outline of WP1 provides the context in which data providers are aggregating their content and metadata.

WP1 tasks T1.1, T1.2 and T1.3 and WP3 task T3.1 provided foundation documents to assist with content selection, with preparation of metadata and with aggregation:

- D1.1 *Content selection policy* [Ref 1]
- D1.2/D3.1 *Rights labelling guidelines & Guidelines for contributing audio content to Europeana* [Ref 4]
- D1.3 *Ontologies for sound* [Ref 5]
- D1.4 *EDM profile for sound* [Ref 2]

Links to these documents and further information about metadata requirements, the aggregation process, the aggregation schedule and the first metadata ingestion training session have been provided in the *Europeana Sounds Manual for Data Providers* [Ref 6].

Task T1.4 *Aggregation management* began in month 7 of the project. This task involves managing the community of data providers, disseminating and implementing standards and workflows and ensuring that all project partners work effectively towards a common goal [Ref 3: DOW Part A, page 7]. All 18 data providers are included in this task, preparing their metadata and aggregating their content and metadata. Task T1.5 *Training and support in content selection policy and aggregation workflow* began in parallel with task T1.4. This task ensures that training and ongoing support is made available to all data providers for the lifetime of the project.

1.3 Technical infrastructure

The technical infrastructure required for aggregation has been developed in WP5 *Technical infrastructure*. This is the metadata ingestion platform (MINT), which has been customised to meet the needs of audio material, in accordance with the specifications in D1.4 *EDM profile for sound*. An initial instance of this customisation was put in place for the first metadata training session, held on 23-24 October 2014. This MINT instance was tested and refined during November and December, with the live version for Europeana Sounds made available to data providers on 19 December 2014 and the publication function made available on 3 February 2015.

Links to MINT documentation have been provided in the *Europeana Sounds Manual for Data Providers*.

1.4 Timeline of activities and documents relating to aggregation

The table below provides a timeline of the main activities undertaken and deliverables completed that relate to aggregation of content and metadata for Europeana Sounds. Full preparation of the first tranche of metadata for ingestion could not be completed until data providers had received training in the Europeana ingestion process, EDM and the EDM Profile for Sounds, and in mapping metadata using the MINT ingestion tool.

Table 2: Activities relating to aggregation of content and metadata

Date	Activities, deliverables, etc.	Partners
31 Jul 2014	D1.1 <i>Content selection policy</i> completed	BL, CNRS, DNB, ONB, RBB, TLA
Aug 2014	Select content and prepare first tranche of metadata for ingestion	All data providers
29 Aug 2014	D1.2/3.1 <i>Rights labelling guidelines & Guidelines for contributing audio content into Europeana</i> completed	KL, BL, EF and all data providers
29 Aug 2014	D1.3 <i>Ontologies for sound</i> completed	NET7, BL, RBB, NTUA, AIT, TLA
Sep 2014	Continue to select content and prepare first tranche of metadata for ingestion	All data providers
Sep 2014	Plan logistics for first metadata ingestion training session	NTUA, BL, EF
Sep 2014	Prepare training material for first metadata ingestion training session	NTUA, EF, BL

Date	Activities, deliverables, etc.	Partners
30 Sep 2014	D1.4 EDM profile for sound completed	EF, NTUA, BL, DNB, RBB, NISV, BNF and members of the wider Europeana Tech community
Oct 2014	Continue to select content and prepare first tranche of metadata for ingestion	All data providers
Oct 2014	Prepare training material for first metadata ingestion training session	NTUA, EF, BL
Oct 2014	Prepare Europeana Sounds MINT instance for training, with SKOS vocabulary	NTUA, EF
17 Oct 2014	<i>Europeana Sounds Manual for Data Providers</i> completed	BL
23-24 Oct 2014	First metadata ingestion training session held at NTUA, Athens	NTUA, EF, BL, all data providers
27 Oct 2014	Begin to map metadata to EDM in MINT (test environment)	All data providers
Nov 2014	Continue to select content and prepare first tranche of metadata for ingestion	All data providers
Nov 2014	Continue to map metadata to EDM in MINT (test environment)	All data providers
Nov 2014	Test EDM Sounds MINT instance	NTUA, BL, EF
Nov 2014	Refine EDM Sounds MINT instance	NTUA
Nov 2014	Revise genre vocabulary	EF, BL
Dec 2014	Continue to select content and prepare first tranche of metadata for ingestion	All data providers
Dec 2014	Continue to map metadata to EDM in MINT (test environment)	All data providers
Dec 2014	Test EDM Sounds MINT instance	NTUA, BL, EF
Dec 2014	Refine EDM Sounds MINT instance	NTUA
Dec 2014	Implement revised genre vocabulary	EF, NTUA
19 Dec 2014	EDM Sounds MINT instance ready	NTUA
19 Dec 2014	Begin to map metadata to EDM Sounds in MINT (live environment)	All data providers
Jan 2015	Continue to select content and prepare first tranche of metadata for ingestion	All data providers
Jan 2015	Continue to map metadata to EDM Sounds in MINT (live environment)	All data providers
Jan 2015	Begin to transform datasets in MINT and receive feedback from EF & NTUA	All data providers
03 Feb 2015	Publication function available in EDM Sounds MINT instance	NTUA

2 Ingestion objectives

This section provides information on the project's ingestion objectives and about the activities that have taken place to meet these objectives since task T1.4 Aggregation management began in August 2014.

The broad ingestion objectives for the project are:

- to ingest metadata describing 500,000 audio objects and 225,000 audio-related objects into Europeana
- to establish a best practice model for future aggregation [Ref 3: DOW Part A, page 6]

2.1 Main stages of ingestion

The work undertaken in task T1.4 Aggregation management takes into account the four main stages of ingestion: content selection, metadata preparation, metadata ingestion and metadata curation. This section outlines activities relating to T1.4, showing the association with each of the stages of ingestion.

2.1.1 Content selection

From August 2014, following the completion of D1.1 *Content selection policy* [Ref 1] at the end of July 2014, data providers began selecting content, according to the D1.1 guidelines and according to the figures in Table 0 Underlying content (DoW, Part B, pages 22-27).

2.1.2 Metadata preparation

Preliminary information on content and metadata requirements

Preliminary information regarding content and metadata requirements was included in D1.1 *Content selection policy* enabling data providers to begin preparing the first tranche of content and metadata for ingestion, establishing the correct rights statement for the objects from the *Europeana Available Rights Statements*¹ and ensuring the inclusion of permanent URIs to link to the digital objects described by the metadata. Metadata quality issues were highlighted in a session in the Metadata ingestion training session 1 (see section 2.1.3).

Rights clearance and rights labelling

Most data providers attended a rights workshop run by WP3 *Licensing guidelines* on 12 June 2014 in Copenhagen, where preliminary discussions on rights clearance issues took place. Following the completion of D1.2/D3.1 *Rights labelling guidelines & Guidelines for contributing audio content to Europeana* [Ref 4] at the end of August 2014, data providers had more information about rights that they could use in selecting content and in preparing their metadata.

¹ <http://pro.europeana.eu/available-rights-statements>

EDM Profile for Sound

Deliverable D1.4 *EDM profile for sound* [Ref 2] was completed at the end of September 2014. This provided data providers with documentation on the metadata schema to which they would be mapping their source metadata during the project. Data providers received training on the general *Europeana Data Model*² and on the *EDM profile for sound* as part of Metadata ingestion training session 1 (see section 2.1.3).

2.1.3 Metadata ingestion

It was necessary for data providers to gain an understanding of the processes and tools associated with ingestion to Europeana before metadata preparation could be completed and ingestion could begin.

Metadata Ingestion Training Session 1

Final preparation of the first tranche of metadata for ingestion could not be achieved until data providers had received training in the Europeana Sounds *Metadata ingestion plan*, the Europeana ingestion process, EDM and *EDM profile for sound*, and in mapping and ingesting metadata using the Metadata Ingestion Tool (MINT). This training took place on 23-24 October 2014 at NTUA in Athens, ahead of schedule and with all data providers attending.



**Cécile Devarenne (EF) presenting at Metadata Ingestion Training Session 1
(Photo: Andra Patterson CC-BY)**

Following on from the training, data providers were better able to assess their metadata against the requirements for ingestion to Europeana and plan how to make adjustments to their metadata, either by editing source metadata or by planning enrichments that could be made during ingestion in MINT.

It is not possible to gain a full understanding of metadata requirements until each data provider has mapped and transformed some metadata using MINT, and has received feedback about their datasets. This process began on 19 December 2014 and data providers are currently receiving valuable feedback

² <http://pro.europeana.eu/edm-documentation>

on their transformed metadata which will enable them to make further refinements to enhance the quality of their metadata.

2.1.4 Metadata curation

Once metadata has been ingested into Europeana it can be further enriched by using MINT to normalise metadata and to add controlled vocabulary concepts. Data providers will receive training on normalisation of metadata during the 2nd metadata ingestion training workshop, scheduled to take place in July 2015.

Task T1.2 *Ontologies* developed new vocabularies for carriers and genres and these are presented in D1.3 *Ontologies for sound* [Ref 5], which was delivered at the end of August 2014. These vocabularies have been implemented as Europeana SKOS³ vocabularies which have been made available to data providers in MINT. The genre vocabulary has been refined and will be expanded as the project progresses.

2.2 Performance indicators and targets

The performance indicators for the project as a whole that relate directly to aggregation are indicators 1-5 in table B3.2b2 Performance monitoring table in the DoW [Ref 3: DOW, Part B, page 91], reproduced here as Table 3.

Table 3: WP1 Performance targets

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, WP5)	Number of consortium partners to have made use of training sessions	33%	66%	100%

Table 0 Underlying content in the DoW [Ref 3: DoW, Part B, pages 22-27] provides information about how many digital objects (items) each data provider will make available as their contribution towards these targets, along with brief information regarding the type of content. Fuller information regarding content selection guidelines and the content selected for Europeana Sounds by each data provider is provided in D1.1 *Content selection policy* [Ref 1].

³ <http://www.w3.org/2004/02/skos/>

2.3 Metadata ingestion plan

The selection guidelines and figures from D1.1 *Content selection policy*, figures from the Performance monitoring table (DoW Part B, page 91) and information from the *Europeana Sounds rights and metadata ingestion survey* [Ref 7] provide a framework for the metadata ingestion plan. This plan takes into account the needs of data providers and the needs of the Europeana Ingestion Team for scheduling. It also takes into account the training plan for data providers. To enable assessment against indicators 1-3 in the *Performance monitoring table* the metadata ingestion plan divides the metadata ingestion into 3 sets. This deliverable focuses on progress during Year 1, which is the same as Metadata set 1 in Table 4:

Table 4: Metadata sets 1-3

Metadata set	Digital objects - audio	Digital objects – audio-related	Total digital objects	Digital objects freely available for re-use (subset of total)
1	50,000	30,000	80,000	10,000
2	200,000	60,000	260,000	30,000
3	250,000	135,000	385,000	50,000
All	500,000	225,000	725,000	90,000

2.3.1 Explanation of statistics

For Europeana Sounds, one metadata record may provide access to multiple digital objects. The figures in Table 4 show the number of digital objects that will be made accessible by Europeana Sounds metadata, rather than showing the number of metadata records. Accordingly, data providers have been asked to count the number of digital objects made accessible through their metadata (counted in the same way as in Table 0 in the DoW) and also to count the number of metadata records they have prepared for ingestion.

The definition of digital objects “freely available for re-use” can be found in the Europeana Glossary ⁴:

Freely Re-usable Content: 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.

2.3.2 Statistics for each data provider

Target figures have been established for each data provider for each metadata set. These figures are recorded in a master spreadsheet monitored by the British Library. Each data provider has an individual spreadsheet that feeds into the master spreadsheet, showing target figures and actual figures achieved.

⁴ <http://pro.europeana.eu/glossary>

	A	B	C	D	E	F	G	H	I	J	K
1	Europeana Sounds T1.4 Reporting		Organisation: 01 BL			Contact: Andra Patterson andra.patterson@bl.uk					
2	Please edit the orange cells!										
3	Data set		Audio				Non-audio				Total
4			Metadata records	Digital objects re-use=Y *	Digital objects re-use=N *	Subtotal audio	Metadata records	Digital objects re-use=Y *	Digital objects re-use=N *	Subtotal non-audio	
5	Set 1	21/11/2014				0				0	0
6		19/12/2014				0				0	0
7		23/01/2015				0				0	0
8											
9		Target total		300	2000	2300		15000	175	15175	17475
10		Actual total	0	0	0	0	0	0	0	0	0
11											
12	Set 2	20/02/2015				0				0	0
13		20/03/2015				0				0	0
14		24/04/2015				0				0	0
15		22/05/2015				0				0	0
16		19/06/2015				0				0	0
17		24/07/2015				0				0	0
18		21/08/2015				0				0	0
19		18/09/2015				0				0	0
20		23/10/2015				0				0	0
21		20/11/2015				0				0	0
22		18/12/2015				0				0	0
23		22/01/2016				0				0	0
24											
25		Target total		0	36700	36700		6000	14710	20710	57410
26		Actual total	0	0	0	0	0	0	0	0	0
27											
28	Set 3	19/02/2016				0				0	0
29		18/03/2016				0				0	0
30		22/04/2016				0				0	0
31		20/05/2016				0				0	0
32		17/06/2016				0				0	0
33		22/07/2016				0				0	0
34											
35		Target total		0	40000	40000		0	115	115	40115
36		Actual total	0	0	0	0	0	0	0	0	0
37											
38		Target total all sets		300	78700	79000		21000	15000	36000	115000
39		Actual total all sets	0	0	0	0	0	0	0	0	0
40											
41	* Note: "Available for re-use" = PDM, CC0, CC-BY or CC-BY-SA										

Figure 1: Example of a data provider's statistics spreadsheet

The information used to establish these figures was supplied by data providers before they had a full understanding of the Europeana Data Model or of rights issues, and before they had received any training on metadata ingestion using MINT. For this reason, the target figures for Metadata set 1 are set slightly above the target total in the Performance monitoring table [Ref 3: DOW Part B, page 91], to allow for adjustments that may be necessary when data providers have a full understanding of what is required. This full understanding cannot be achieved until each data provider has mapped and ingested some metadata using MINT, and has received feedback about their metadata.

3 Progress during year 1, August 2014-January 2015

3.1 Comparison of targets with results

The table below compares expected progress as defined by the performance indicators with actual progress during Year 1, August 2014-January 2015.

Table 5: Comparison of targets and actual progress

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	50%	89%
5	Aggregation (WP1, WP5)	Number of consortium partners to have made use of training sessions	33%	100%

3.1.1 Performance indicators 1-2

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*⁵ has been determined
- links to digital objects are present and working, and the objects are accessible worldwide, without a password

As can be seen in Table 5 above, actual progress on the number of items aggregated has exceeded the target for performance indicator 1 by 5% and has exceeded the target for performance indicator 2 by 6%.

3.1.2 Performance indicator 3

Actual progress on the number of items freely available for re-use has exceeded the target for performance indicator 3 by 296%. In line with the Europeana glossary, “freely available for re-use” is defined as digital objects with rights statements of Public Domain, CC0, CC-BY or CC-BY-SA. Although this figure could change as data providers become more familiar with the application of *Europeana Available Rights Statements*, the target is easily achievable even if some rights statements need adjusting.

⁵ <http://pro.europeana.eu/available-rights-statements>

3.1.3 Progress achieved by each data provider on indicators 1-3

The table below shows the progress made by each data provider on the number of audio and non-audio items aggregated and the number of items freely available for re-use. The progress shown here is in accordance with the *Metadata ingestion plan* (see section 2.3).

Table 6: Progress achieved by each data provider on indicators 1-3

Data provider	Audio				Non-audio				Total
	Metadata records ready	Digital objects Reuse=Y	Digital objects Reuse=N	Subtotal audio digital objects	Metadata records ready	Digital objects Reuse=Y	Digital objects Reuse=N	Subtotal non-audio digital objects	Digital objects audio + non-audio set 1
01 BL	6307	0	6307	6307	324	26700	0	26700	33007
02 NISV	0	0	0	0	100	570	30	600	600
06 BNF	3000	6000	0	6000	0	0	0	0	6000
10 CNRS	2522	855	1652	2507	0	0	0	0	2507
11 DIZI	500	500	0	500	0	0	0	0	500
12 DNB	12	12	0	12	0	0	0	0	12
13 FMS	36	0	37	37	0	0	69	69	106
14 ICCU	0	0	0	0	0	0	0	0	0
15 ITMA	312	108	204	312	1388	754	634	1388	1700
16 TLA	43	0	43	43	0	0	0	0	43
17 NLL	1016	1016	0	1016	0	1016	0	1016	2032
18 OEM	645	0	645	645	0	0	0	0	645
19 RBB	21058	0	21058	21058	0	0	0	0	21058
20 TAD	11968	0	11968	11968	0	0	0	0	11968
21 SB	131	131	0	131	0	0	0	0	131
22 ONB	0	0	0	0	0	0	0	0	0
23 FCSH	115	0	115	115	1909	1909	0	1909	2024
24 CCE	1788	0	1788	1788	0	0	0	0	1788
Total	49453	8622	43817	52439	3721	30949	733	31682	84121
Target total (PI1-2)				50000				30000	80000
% target				105%				106%	105%
Total ready for re-use		8622				30949			39571
Target total for re-use (PI3)									10000
% target re-use		86%				310%			396%

Table 7: Brief description of Year 1 items aggregated

Data provider: 01 BL - The British Library

Description: Audio objects: Oral history, wildlife and environment recordings, including oral history recordings of people associated with Glyndebourne Opera, the George Ewart Evans collection of recordings documenting rural life in the UK in the late 19th and early 20th centuries, a collection of interviews offering insights into the lives and opinions of British people in the 1980s, the Opie collection of children's games & songs, sounds of water, weather and wildlife. Printed music: Vocal and instrumental music published in the 16th-century, including motets, madrigals, songs, lute music.

Data provider: 02 NISV - Nederlands Instituut Voor Beeld en Geluid

Description: Video objects: Dutch newsreels in which music is the main subject or plays a ceremonial role. There are videos in which [music instruments from the olden days are being played](#), [items about music boxes](#), [a group of people playing accordion](#), [famous violinist and conductor Yehudi Menuhin at 14 years of age](#), the making of [guitars](#), [violins](#) and [accordions](#). Furthermore, some videos in which music plays a ceremonial role are part of the collection as well: [the remembrance of Pieter Bruegel](#) or [the installation of a mayor in the Netherlands](#) are examples of this category.

Data provider: 06 BNF - Bibliothèque nationale de France

Description: Audio objects : Archives of the spoken word (Ferdinand Brunot 1911-1914), Archives of the spoken word (Jean Poirot 1920-1924), Archives of the spoken Word / Musée de la Parole et du Geste (Hubert Pernot 1924-1930), Exposition coloniale internationale de Paris (1931), World music, Classical music and Opera, speeches from French politicians during WWI, Fonds Gilles Deleuze, Fonds Fred Deux.

Data provider: 10 CNRS - Centre national de la recherche scientifique

Description: From CREM: Worldwide traditional music and spoken word recordings made during the Universal Exhibition In Paris 1900 by Docteur Léon Azoulay, a french linguist of the Anthropological Society of Paris.

From MMSH: Recordings from the Mediterranean area. Some are from the Provençal Alps (botanical anthropology, customs, use of wood, recipes, etc.); others are from the Bouches-du-Rhône, especially in Marseille, Arles and Aix-en-Provence, about oral history of Armenian memory, women and politics, sociability in Provence, life and working organisations of fishermen and about football supporters and players; four collections offer ethnologists' field recordings in Iran, Yemen, Sahara, Romania, Italy; four oral histories about the 1954-1962 Algerian War from the viewpoint of Harki.

Data provider: 11 DIZI - UAB DIZI

Description: Audio objects: Lithuanian folksongs and laments, folk and folk dance music from field recording sessions in Lithuania.

Data provider: 12 DNB - Deutsche Nationalbibliothek

Description: Audio objects: music recordings, classical and popular music, from the shellac era in the 20th century.

Data provider: 13 FMS - Syllogos Oi Filoi Ths Mousikhs / Σύλλογος Φίλων της Μουσικής

Description: Audio, sheet music, lyrics and photos deriving from the research programme "Recording, study and promotion of the Thracian heritage in music and dance", conducted by Music Library of Greece of the Friends of Music Society. All files are unique and representative of the music and dance tradition of the acritic region of Thrace and Eastern Macedonia, in Greece.

Data provider: 14 ICCU - Istituto Centrale per il Catalogo Unico delle biblioteche italiane e per le informazioni bibliografiche

Description: ICCU will aggregate metadata in Years 2-3

Data provider: 15 ITMA - Irish Traditional Music Archive / Taisce Cheol Dúchais Éireann

Description: Videos of Irish traditional dance, music and song recorded by ITMA staff at the Frankie Kennedy Winter School, the Willie Clancy Summer School, Sean-Nós Cois Life, The Con Curtin Festival, Cruinniú na bhFlúit and in the ITMA studio. Audio recordings of Irish traditional music and song from the Breandán Breathnach collection, the Hugh Shields collection and the Jimmy McBride collection. Public domain recordings from cylinder, 78 and LP and recordings made by ITMA staff. Images featuring performers of Irish traditional music, artefacts, woodcuts from ballad sheets and publications relating to Riverdance. Text objects of public domain books, manuscripts, ballad sheets, a 1960s journal and Shamrock, Rose and Thistle, the publication that grew out of the Hugh Shields

collection. Also included are PDFs of song words from the Inishowen Song Project, which grew out of the Jimmy McBride Collection, and scores from Ryan's Mammoth Collection.

Data provider: 16 TLA - Max Planck Gesellschaft

Description: Audio collection of recordings made of the Popoloc language. The recordings of the Popoloc language were made and analysed by Annette Veerman-Leichsenring during several field work periods from 1980 till 1990 which has resulted in Gramática del popoloca de Metzontla. The recordings hold different aspects of the language, ranging from verbs, adjectives to nasal vocals and numbers. The Popolocan language family belongs to the Otomanguan phylum and is composed of Popoloc(a) (± 11,000 speakers), Chocho (± 1,000 speakers), Ixcatec (no more than 10 speakers) and more distantly related Mazatec (± 100,000 speakers) which is not included in the digitised recordings. All four languages are spoken in Mexico, in the southern part of the State of Puebla, the northeastern part of the State of Oaxaca and in an adjacent fringe in the State of Veracruz.

Data provider: 17 NLL - Latvijas Nacionālā Bibliotēka

Description: Audio objects: Well-known opera arias (mostly sung in Latvian), Latvian folksongs and choral music, Latvian folk music, folk dances, some Western classical music, Latvian classical music, Latvian classical choir music, popular music of the 1930s, comic songs and sketches, some oral history recordings and radio broadcasts, mostly in Latvian. Images: labels of the shellac records.

Data provider: 18 OEM - Österreichische Mediathek

Description: Audio objects: Classical music (opera recordings, shellac era), music and cabaret (Jewish composers and performers, shellac era) and recordings from the Austrian Parliament (radio recordings, 1950s and 1960s).

Data provider: 19 RBB - Rundfunk Berlin-Brandenburg

Description: Audio objects: 21,000 30-second excerpts of a variety of in-commerce recordings, including Classical (Berlioz, Mozart, Telemann, etc.), Electronic (House, Dance, Techno), HipHop, Latin, New Age, Pop, Reggae, World Music and more. All 'independent' releases. Each track comes with a genre description in two levels (main-genre, sub-genre). Images: related CD-covers.

Data provider: 20 TAD - Sabhal Mòr Ostaig

Description: Audio objects: 12,000 tracks mostly recorded in Scotland from the 1930s to 2000, in Scottish Gaelic, Scots and English. The tracks contain spoken word, singing, poetry, excerpts, radio programmes, 'Mòd' recordings (an annual national music contest celebrating Gaelic culture) and music. The material comes from three collections - The Scottish Studies Archive at the University of Edinburgh, the BBC Gaelic Department and The National Trust for Scotland's Canna Collection. The project is known as Kist O Riches in Scots and Tobar an Dualchais in Gaelic, meaning 'The chest/well of heritage' and the tracks are publicly available at www.tobarandualchais.co.uk.

Data provider: 21 SB - Statsbiblioteket

Description: Audio objects: 127 restored audio files from the Ruben Collection, which contains the oldest Danish sound recordings, dating from 1889 to 1897. Consul-General Gottfried Ruben heard the improved Edison Phonograph at the Paris World Fair in 1889, and during the fall of 1889 he established the company "Edisons Fonograph-Compagniet" in Copenhagen, thus becoming the agent for the company in Scandinavia. In collaboration with his business partner, optician Cornelius Knudsen, he recorded several of the most famous Danish singers and actors of the era. Recordings of Peter Schram, Augusta Lütken, Louise Phister, Otto Zinck, Frederik Jensen, Emil Oluf Poulsen and Niels Juel Simonsen have all been preserved, along with numerous recordings made with the intention of demonstrating the technology for customers. Of these recordings approximately 150 wax cylinders are still intact and they now form the Ruben Collection.

Data provider: 22 ONB - Österreichische Nationalbibliothek

Description: ONB will aggregate metadata in Years 2-3

Data provider: 23 FCSH - Faculdade de Ciências Sociais e Humanas da Universidade Nova de Lisboa

Description: Collection of oral memories from different regions of Portugal in which interviewees talk about local daily life, their traditions, superstitions and songs. Collection of oral traditions, such as proverbs, rhymes, popular poems and tongue twisters. Some records are also related to the making and playing of traditional musical instruments such as the “adufe”, the “ronca” and the Portuguese bagpipe. All these collections are part of the MemoriaMedia project (<http://www.memoriamedia.net/>).

Data provider: 24 CCE - Comhaltas Ceoltóirí Éireann

Description: Audio objects: Irish traditional music and song of various artists recorded by Comhaltas Ceoltóirí Éireann in the later half of the 20th century for commercial release. Irish traditional music, song and spoken word recorded or collected in reel to reel format from various sources by Séamus Mac Mathúna for Comhaltas Ceoltóirí Éireann in the mid and later half of the 20th century.

3.1.4 Performance indicator 4

The number of data providers using the EDM profile for sounds is 16 out of 18, or 89%, which is 39% above target.

There are two Europeana Sounds data providers who have not been able to contribute to aggregation in Year 1 and who will begin to aggregate content and metadata in Year 2.

ICCU are undergoing a major transformation of their online catalogue, which means that although they are able to select content for the project they are currently unable to harvest and supply their metadata. This issue has been integrated into the metadata ingestion plan and ICCU will begin to process metadata for Europeana Sounds in Year 2.

ONB have prepared a large amount of content (107,665 digital objects) and metadata (1280 records) that has not been included in Year 1 while an issue around the rights labelling of their music scores is resolved (see section 4.3). Discussions are taking place between ONB, KL, EF and BL to resolve this issue and the *Metadata ingestion plan* has been adjusted to move ONB content and metadata to Years 2 and 3.

3.1.5 Performance indicator 5

The number of data providers to have made use of training sessions is 100%, which is 66% above the target for Year 1. This is due to a decision made by WP1 and WP5 that it would be advantageous to train all data providers together to promote an environment in which everyone could learn from each other's questions. We will monitor the effectiveness of this in the months ahead, particularly for ICCU and ONB who, for the reasons stated above, have not yet made use of their training to ingest metadata in MINT.

3.2 Summary of progress

All targets for performance indicators 1-5 in the Performance monitoring table [Ref 3: DOW Part B, page 91] have been exceeded.

4 Identification and assessment of ingestion challenges

4.1 Impact of technical requirements on selection of content

In selecting content for Europeana Sounds we have ensured that a wide variety of audio and audio-related content will be aggregated. We have favoured inclusion rather than imposing strict criteria for inclusion, apart from the need to meet metadata and other technical requirements. The driver for this is that we need to aggregate a large quantity of audio and audio-related material in order to establish best practice in terms of the application of the metadata profile, rights clearance activities, rights labelling and technical requirements for material made available for re-use, which will make it easier for future data providers to aggregate these types of material for Europeana.

It is difficult for data providers to be certain about the content and metadata they can provide to a project before they are fully aware of all the requirements involved. For the future it would be useful for prospective data providers to gain a better understanding of these requirements before committing to a project. To this end, later in this project we should run a workshop on best practice for a wider pool of potential data providers.

4.2 Changes to technical infrastructure

As described in section 3.1.4 of this report, ICCU have not been able to contribute in Year 1 and will begin to aggregate content and metadata in Year 2. Like ICCU, other data providers may face changes to technical infrastructure within their organisations which impact on their ability to aggregate content and metadata during the project. Usually, as is the case with ICCU, these changes are foreseen and can be integrated into the project's aggregation plan.

We acknowledge that on occasion the full impact of changes made to technical infrastructure may not be anticipated. This may have an impact on a data provider's ability to aggregate content and metadata at a particular point in time so it is vital that we keep lines of communication open so that issues such as this are raised and resolved.

4.3 Rights and licensing

ONB have raised an issue with the WP1 and WP3 Leads regarding assigning the "Free Access No Re-use" rights statement to their "Public Domain" digitised music scores. Very aware of the wording of the Consortium Agreement and the Data Exchange Agreement, ONB were concerned that there could be a conflict between these two documents and ONB's institutional strategy and practice.

Discussions are taking place between ONB, KL, EF and BL to resolve this issue and the metadata ingestion plan has been adjusted to move aggregation of ONB content and metadata to Years 2 and 3.

An issue has been raised regarding ambiguous wording in the general *EDM mapping guidelines*⁶, where in ore:Aggregation edm:rights an example is provided that does not reflect best practice in rights labelling. This had been discussed and the information in this documentation will be updated by EF.

4.4 Development and maintenance of new genre concepts vocabulary

An issue was raised by Europeana regarding the ongoing development and maintenance of the new genre concepts vocabulary that was developed for Europeana Sounds under task T1.2 *Ontologies*. It was decided that rather than implementing and trying to develop and maintain the large genre vocabulary that had been created, two levels of this vocabulary would be put in place for use by data providers so that development and maintenance can be tested, beginning on a small scale and increasing as the project progresses⁷.

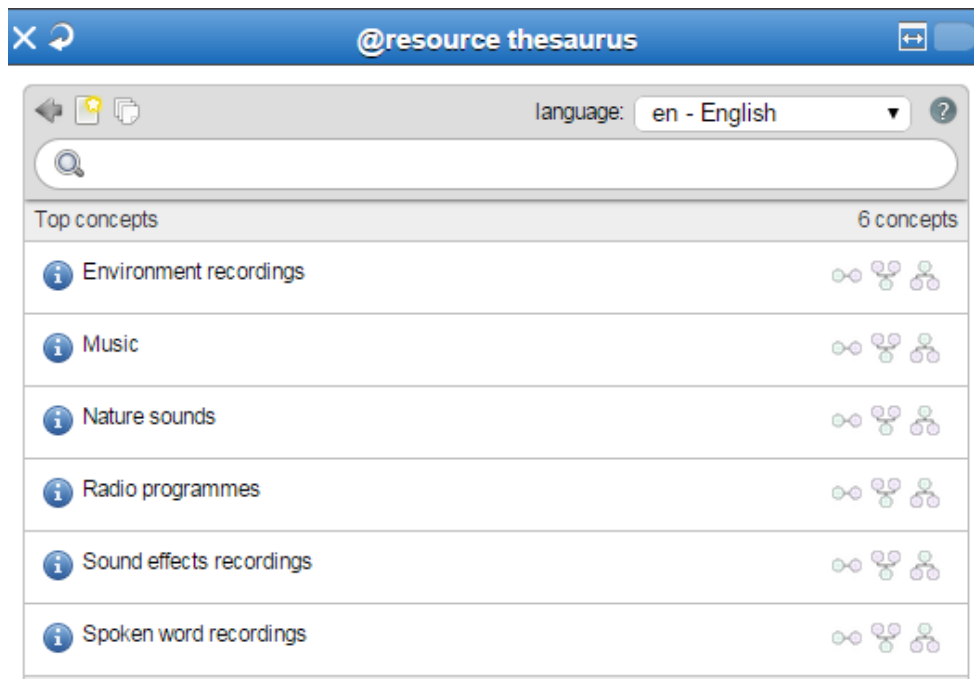


Figure 2: Top level concepts of the new genre vocabulary, shown in MINT

4.5 Timeframe for Year 1 aggregation

Data providers have been selecting content and preparing metadata since August 2014. Metadata ingestion training session 1 took place in late October and data providers were not in a position to fully understand the requirements of the *EDM profile for sound* [Ref 2] until this training had taken place. As a result of the training it has been necessary for some data providers to revise the metadata they had prepared.

During Metadata training session 1, data providers learned how to use the MINT platform to map and ingest their metadata. Following the training, they were able to practice this in a test environment using

⁶<http://pro.europeana.eu/edm-documentation>

⁷ For an example, see

<http://skos.europeana.eu/api/concept?id=http://data.europeana.eu/concept/soundgenres/Music.html>

the general EDM profile rather than the EDM Sounds profile, which was not available for live use until 19 December 2014, just before Christmas, leaving data providers only 5 weeks in Year 1 to gain familiarity with and use the properties in this complex profile.

Completion of WP5 deliverable D5.1 *Report on the evaluation of the aggregation mechanism* has been moved from month 12 to month 14 of the project to allow data providers a little more time in using the aggregation mechanism before undertaking an evaluation.

For future projects it would be better to have more time assigned between the completion of a new EDM profile and its implementation on MINT, to allow time for testing and refining the implementation. It would also be better to train data providers on a fully-tested live implementation rather than on a test implementation. This would give data providers a better foundation on which to aggregate their content and metadata

4.6 Documentation

There is a large amount of complex documentation for data providers to understand and use in their aggregation work. From time to time a particular aspect of aggregation needs to be highlighted and explained in more detail.

The project's WP1 Basecamp is used for this, with some key issues also documented in the *Europeana Sounds Manual for Data Providers* [Ref 6]. This manual is a working document that can be updated throughout the project, providing key information in one place, with links to relevant documents.

5 Concluding observations and looking ahead

5.1 Concluding observations

Although there has been a short timeframe in which to select and prepare content and metadata, to learn about EDM, the *EDM profile for sound* [Ref 2] and how to map and transform metadata in MINT, data providers have been successful in achieving the performance targets for Year 1. They have extended their knowledge about rights issues and on assigning correct rights statements for their content. They are using MINT to map their metadata to the *EDM profile for sound* and they are receiving feedback on their transformed datasets, sharing interesting issues through the WP1 Basecamp.

We have implemented a solution for the issue relating to maintaining the new genre concepts vocabulary and we discussed further development of the vocabulary during the plenary project meeting in February 2015. We look forward with interest to the outcome of T5.2 *Aggregation infrastructure evaluation*, to be presented in deliverable D5.1 *Report on the evaluation of the aggregation mechanism*, which will be completed at the end of March 2015. Kennisland will meet with ONB to discuss the issue regarding which rights statement will be assigned to ONB's music scores. We have set up methods of providing access to documentation for WP1 and we will continually monitor their effectiveness.

Our achievements in year 1 demonstrate that there has been an excellent start to aggregation for Europeana Sounds.

5.2 Looking ahead

During Years 2 and 3 data providers will continue to select content and will continue to prepare, map and transform their metadata. They will extend their knowledge about the *EDM profile for sound* and MINT through further training sessions and by sharing issues, questions and answers.

5.2.1 Training

An “Ingestion clinic” was held at the Europeana Sounds Plenary Meeting at the Bibliothèque nationale de France, 10-11 February 2015. Representatives from the Europeana Foundation and NTUA joined the WP1 Lead in answering questions from data providers on topics such as the *EDM profile for sound*, MINT and the ingestion schedule. A second Metadata Ingestion Training Session is planned for July 2015. All data providers will attend this session, in which they will learn advanced techniques for enriching their metadata.

5.2.2 Publication schedule

Publication of metadata to Europeana is a two-stage process. The first stage of publication is carried out by the data providers themselves after they have mapped and transformed their metadata in MINT. The second stage of publication involves a re-processing of the metadata by the Europeana Ingestion Team. Following this, the metadata is visible on the Europeana portal. Table 8 shows the draft publication schedule for publication of metadata and associated provision of access to digital objects.

Table 8: Draft publication schedule

Batch	Metadata records (approximate*)	Digital objects made available	Date published by data provider	Month published by Europeana
1	47,280	72,500	21 Mar 2015	May 2015
2	47,280	72,500	21 May 2015	Jun 2015
3	47,280	72,500	21 Jul 2015	Sep 2015
4	47,280	72,500	21 Sep 2015	Oct 2015
5	47,280	72,500	21 Nov 2015	Dec 2016
6	47,280	72,500	21 Jan 2016	Feb 2016
7	47,280	72,500	21 Mar 2016	Apr 2016
8	47,280	72,500	21 May 2016	Jun 2016
9	47,280	72,500	21 Jul 2016	Sep 2016
10	47,280	72,500	21 Sep 2016	Oct 2016

Total	472,800	725,000	(Project ends Jan 2017. Final improvements to data to be made between Oct 2016-Jan 2017)
--------------	----------------	----------------	--

*Approximate figures are provided for metadata records as the performance indicator targets are for the number of digital objects made available rather than for the number of metadata records ingested.

5.2.3 Changes to selection policy

Responses to the WP1/WP3 *Europeana Sounds Rights and Metadata Ingestion Survey* [Ref 7] show that data providers have selected a good range of material for inclusion in the project, in accordance with the *Description of Work* and in accordance with the guidelines in D1.1 *Content selection policy*. There are no recommendations for changes to the selection policy at this point in time. This will be reviewed in D1.7 *Aggregation report 2*, due for completion in February 2016.

6 Summary

The first phase of aggregation for Europeana Sounds began in August 2014 and this report documents the progress of aggregation activities up until the end of January 2015. It describes the processes and achievements of Europeana Sounds data providers in aggregating their content and metadata, identifies the main challenges faced in aggregation and identifies solutions that ensure progress during and beyond this project.

In Section 1 we provided the context of aggregation within the project, with an outline of relevant associated tasks and documents, a brief description of the technical infrastructure developed by WP5 and a timeline of activities and documents relating to aggregation. In Section 2 we presented the ingestion objectives, including the performance indicators and metadata ingestion plan for the project.

Section 3 outlined the actual progress made by the project as a whole and by each data provider contributing to the project. In this section we showed that actual progress on the number of items aggregated has exceeded the target for the number of audio items aggregated by 5%, has exceeded the target for the number of other (non-audio) items aggregated by 6% and has exceeded the target for the number of items freely available for re-use by 296%. We saw that 89% of data providers are using the EDM Profile for Sound and that 100% of data providers have made use of training sessions. To summarise actual progress, all performance indicators for Year 1 of the project have been achieved or exceeded.

Ingestion challenges were identified and assessed in Section 4, including challenges relating to the impact of technical requirements on the selection of content, the impact of changes to technical infrastructure within organisations, rights and licensing issues, the ongoing development and maintenance of the new genre concepts vocabulary, the short timeframe for Year 1 aggregation, and challenges relating to documentation.

In Section 5 we concluded that there has been an excellent start to aggregation for Europeana Sounds and we outlined the training and publication plans which will ensure our continued success in Years 2 and 3 of the project.

7 References

Ref 1	D1.1 Content Selection Policy http://pro.europeana.eu/documents/2011409/70e86a37-7608-4f37-9507-1b6f6152cd42
Ref 2	D1.4 EDM profile for sound http://pro.europeana.eu/documents/2011409/5c845fc2-dcf7-46c2-a66d-6aea9dac1c0c
Ref 3	Europeana Sounds "Description of Work" http://pro.europeana.eu/documents/2011409/8d0e9833-4608-494e-af77-681e68f8a8c8
Ref 4	D1.2/D3.1 Rights labelling guidelines & Guidelines for contributing audio content to Europeana http://pro.europeana.eu/documents/2011409/b4f9281a-b72b-400f-94c8-bedfeae2fbb6
Ref 5	D1.3 Ontologies for sound http://pro.europeana.eu/documents/2011409/ed7f27af-65ff-4ea3-9d00-8345c31489cf
Ref 6	Europeana Sounds Manual for Data Providers https://docs.google.com/document/d/19vjibRkMAbIC04_-1l_59bPBSQBTAmDSMgSJ3Mp00hk/edit
Ref 7	Europeana Sounds rights and metadata ingestion survey http://pro.europeana.eu/documents/2011409/dc33e84d-6dca-4178-9dd2-aab41878c2b8

8 Appendix A: Terminology

A project glossary is provided at: <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
BL	British Library
EC-GA	Grant Agreement (including Annex I, the Description of Work) signed with the European Commission
EDM	Europeana Data Model
PI	Performance Indicator
PMB	Project Management Board
TEL	The European Library
WP	Work Package