



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

MS6 Final tranche of content and metadata ready for ingestion

Revision: Final
Date: 03/08/2016
Authors: Tom Miles (British Library)

Abstract: This document details Milestone 6 for Europeana Sounds in month 30: the final tranche of metadata ready for ingestion. The document outlines the key performance indicators for Work Package 1 (Aggregation), the targets set for each data provider, difficulties encountered and action taken and detailed statistics of metadata transformed and published on Europeana.

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	15/06/2016	
0.2	1st draft	Tom Miles, BL	01/07/2016	
0.3	2nd draft	Tom Miles, BL	25/07/2016	Input from reviewers
0.4	Final draft	Tom Miles, BL	01/08/2016	Input from PMB
1.0	Final	Richard Ranft, BL	03/08/2016	Layout, minor changes

Review and approval

Action	Name, organisation	Date
Reviewed by	Treasa Harkin, ITMA	04/07/2016
Approved by	Coordinator and PMB	03/08/2016

Distribution

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

Application area

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

Statement of originality

This document contains original unpublished work except where clearly indicated otherwise.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

Project summary

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

- Double the number of audio items accessible through Europeana to over 1 million and improve geographical and thematic coverage by aggregating items with widespread popular appeal such as contemporary and classical music, traditional and folk music, the natural world, oral memory and languages and dialects.
- Add meaningful contextual knowledge and medium-specific metadata to 2 million items in Europeana's audio and audio-related collections, developing techniques for cross-media and cross-collection linking.
- Develop and validate audience specific sound channels and a distributed crowd-sourcing infrastructure for end-users that will improve Europeana's search facility, navigation and user experience. These can then be used for other communities and other media.
- Engage music publishers and rights holders in efforts to make more material accessible online through Europeana by resolving domain constraints and lack of access to commercially unviable (i.e. out-of-commerce) content.

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

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

Copyright notice

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

1	Executive summary: MS6	5
2	Introduction	5
2.1	Explanation of statistics	5
2.2	Key performance indicators	7
2.3	Targets for data providers	7
3	Progress from month 13 to month 30	8
3.1	How metadata is used in other work packages	8
3.2	Difficulties encountered and actions taken	8
3.3	Results	9
4	Next steps	12
5	Summary	12
6	References	12
	Appendix A: Terminology	13

1 Executive summary: MS6

This document records the aggregation contributions of all data providers to the Europeana Sounds project. By the due date of this milestone, the final tranche of metadata should be transformed on MINT and ready to be ingested by Europeana. The project has succeeded in meeting this milestone overall; the remaining six months of the project will be spent revising and adding to this metadata.

2 Introduction

2.1 Explanation of statistics

Before detailing the amount of metadata transformed by Europeana Sounds data providers, it is necessary to explain the various stages of processing and what is meant by transformation, publication, etc. This is explained in more detail in deliverable D1.5. *Aggregation report 1*.

Records transformed on MINT: the records have been correctly transformed into .xml files which follow the EDM Sounds format; mandatory fields have been completed and these files conform to the requirements of Europeana.

Records published on MINT: these are the aforementioned transformed records which are published and marked as waiting for harvest and published by Europeana.

Records published on Europeana: these records have been harvested from MINT and are now publicly searchable on Europeana.


Records and digital objects on Europeana: Some records on Europeana will link to more than one audio recording and, also, may link to other digital objects. Thus, the number of digital objects aggregated by Europeana Sounds is greater than the number of records on Europeana (see D1.7 *Aggregation report 2* and see Figures 1 and 2 below).

eana
ONS

Home Collections ▾ Explore ▾ Exhibitions ▾ Blog ▾

OUR SITES

Results / Item



Technical metadata

DOWNLOAD

Title Missa No. 2 in d-Moll and a 4 voci concertanti, 4 voci ripieni, 2 Violini, Viole. 2 Oboe (!) 2 Tromboni. 4 Trombe Timpani, fagotti, Violoncello e Basso, Organo M. D. C. | Salieri, Antonio [Komponist]

FIND OUT MORE
Europeana is displaying this item via
Osterreichische Nationalbibliothek - Austrian
National Library. Visit their site for more

Figure 1: an example of a metadata record with multiple images

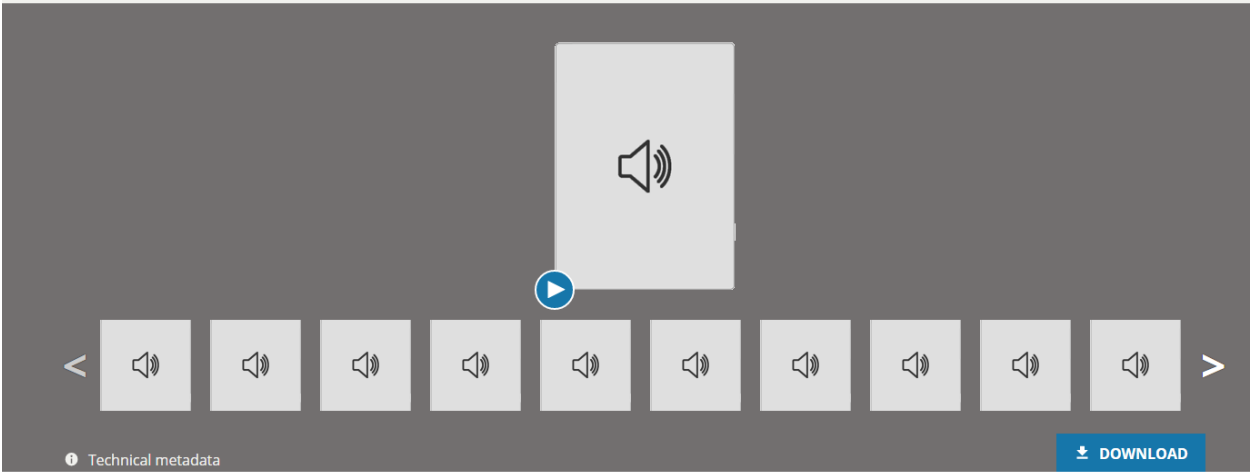
http://www.europeana.eu/portal/en/record/2059216/data_item_onb_sounds_AL00608496.html

europeana
collections

Home Collections ▾ Explore ▾ Exhibitions ▾ Blog ▾

OUR SITES

Home / Results / Item



Technical metadata

DOWNLOAD

Title Zwan Live at Alter Schlachthof on 2003-06-05 | Zwan

2003-06-05 - Alter Schlachthof

Description Venue: Alter Schlachthof and cd1: 01: mary star of the sea 02: declarations of faith 03: el sol 04: for your love 05: lyric 06: endless summer 07: jesus, i [lyte] -> god's gonna set this world on fire (tease)

FIND OUT MORE
Europeana is displaying this item via
Netherlands Institute For Sound And Vision. Visit their site for more information.

CAN I USE IT?

Figure.2: an example of a metadata record with multiple audio files

http://www.europeana.eu/portal/en/record/2059210/data_sounds_zwan2003_06_05_flac16.html

2.2 Key performance indicators

The Key Performance Indicators for aggregation for each year were set at the start of the project. The KPIs for Work Package 1 were met for each year (see also D1.5 and D1.7 Aggregation Reports for Years 1 and 2 respectively).

Although the KPIs are measured for Month 36, these have already been met or are about to be met.

Table 1: KPIs for Work Package 1 for end of Year 3.

No	Relating to objective /result	Indicator name	Expected progress Year 3	Actual progress Month 30 in Year 3	Current Status
1	Aggregation (WP1)	Number of audio items aggregated	500,000	552,865 on Europeana	Exceeded
2	Aggregation (WP1)	Number of other items aggregated	225,000	247,525 on MINT; 210769 on Europeana	Exceeded
3	Aggregation (WP1)	Number of items freely available for re-use	90,000	177,248 on Europeana	Exceeded
4	Aggregation (WP1)	Number of Data Providers using new EDM profile	100%	100%	Met
5	Aggregation (WP1, WP5)	Number of consortium partners to have made use of training sessions	100%	100%	Met

2.3 Targets for data providers

As outlined in the Description of Work for Europeana Sounds (page 75), each data provider has committed to providing a certain number of audio and audio-related objects. The data providers' combined targets total 544,000. Section 3.3 outlines how each data provider has performed against their individual targets.

3 Progress from month 13 to month 30

3.1 How metadata is used in other work packages

Records with musical themes and a direct link to media can be included in the Music Collections facet of Europeana. In addition, these records could be used for WP2 Data enrichment and included in material that could be tagged using the platforms Pundit (from Net7) and WITH (from NTUA).

As the project has developed, data providers have responded to the need for direct links, either by revising datasets to include the URLs of digital objects (see figure 1 and 2); or, where this is not possible, by supplying the URLs of embedded digital objects which can be streamed but not downloaded. This has ensured that more material is available for data enrichment such as tagging and edit-a-thons (WP2); and more music-themed material is available in Music Collections (WP4). This work has been carried out working closely with licensing (WP3). During the lifetime of the project the material aggregated by WP1 has been highlighted to publicise the project through tweets, Facebook posts and blog postings (WP6).

3.2 Difficulties encountered and actions taken

Issue 1: The British Library has transformed the metadata for 79,000 sounds and 36,000 images; therefore it has met its targets for the project as outlined in the DOW. However, some datasets cannot be published on Europeana immediately as they need to be made accessible to a worldwide audience.

- Some traditional and folk music collections are still awaiting permission from their collectors (or estates thereof) for these to be accessible worldwide, rather than to UK Higher Education and Further Education communities only, as at present.
- It is likely that the BL's collection of pre-1962 commercial recordings, comprising around 2,600 recordings, will be seen as infringing copyright if accessible within the US. For this reason, these recordings can only be streamed from within the European Union.

Action taken: The above datasets are transformed on MINT and will be published on Europeana when the content is freely accessible throughout the world. If the content cannot be made accessible worldwide, the BL will source alternative material.

Issue 2: In the Description of Work, CNRS had declared that 20,000 of its records that would be available for Europeana would need clearance in order to be accessed worldwide. These records cannot be made accessible worldwide due to rights restrictions. Therefore, there was the risk that CNRS could only supply a total of around 17,000 instead of 37,000 records.

Action taken: CNRS has been able to submit extra records between CNRS/CREM and CNRS/MMSH. In addition to this, other institutions (CNRS/CRESSON and CNRS/LARHRA) will be providing another 10,000. This would bring the total sounds submitted to 32,000 - which is 87% of the total as outlined in the DOW. As some other data providers are supplying more metadata than originally targeted, this would not jeopardise the KPIs of the project as a whole. The PMB agreed this new content was a reasonable substitution for the original.

Issue 3: FCSH has also encountered problems in sourcing worldwide accessible content. A large collection was considered to be suitable for publication on Europeana; however, it became apparent that due to rights restrictions this collection could not be made freely accessible worldwide.

Action taken: FCSH has been looking at other sources of recordings, including recordings from the Fado Museum. Around 6,400 audio items have now been transformed on MINT, which is encouraging progress. However, a further 33,500 audio items will need to be found; and, also, a further 2,500 non-audio items. Although FCSH believes it will make up the shortfall between month 30 and month 36, this matter is now being looked at carefully by the PMB and a decision will need to be taken on how best to proceed.

Issue 4: Although NISV has been able to deliver a large amount of extra audio to the project in addition to audio, video and images, it has been difficult to source the 4,000 texts the institution also committed to, due to an unresponsive external partner.

Action taken: These 4,000 texts represent 4% of NISV's entire contribution to the project. Over the remaining six months NISV will explore alternative sources if the current external partner is not able to provide this material.

3.3 Results

The table below details the progress of each provider against their targets.

Table 2: Summary of aggregation for each data provider

Part ner #	Data Provider	Records uploaded on MINT	Records transformed on MINT	Records published on MINT	Records published on Europeana	Digital objects transformed on MINT	Digital objects on Europeana	Target (from DOW)	Type of file	Remaining digital objects requiring transformation on MINT	Records still to be published on Europeana
1	BL	80,608	80,608	66,888	55,851	80,608	57,191	79,000	Audio	0	14,000
		1,417	1,417	1,417	1,340	38,756	17,000	36,000	Images	0	19,000
2	NISV	35,166	31,508	13,439	13,327	169,980	166,000	86,170	Audio	0	0
					100	100	100	50	Video	0	0
						10,000		10,000	Images	0	10,000
								4,000	Text	4,000	4,000
6	BNF	36,219	26,039	4,540	4,540	4,540	4,540	10,000	Audio	5,460	5,460
						4,540	4,540	10,000	Images	5,460	5460
10	CNRS							37,600	Audio		37,600
		11,193	18,307	18,307	10,644	18,307	10,644		Audio		
		5,331	4,601	4,650	4,268	4,268	4,268		Audio		
		1,057	753	753	753	753	753		Audio		
11	DIZI	57,276	77,218	14,907	10,234	10,234	10,234	10,000	Audio	0	0
12	DNB	7,849	526	526	526	526	526	500	Audio	0	0
13	FMS	6,635	3,644	4,120	3,296	3,296	3,296	2,500	Audio	0	0
						3,300	3,300	3,300	Other	0	0
14	ICCU	83,310	83,122	83,122	83,122	83,122	83,122	80,000	Audio	0	0
						23,270	23,270	0	(images)		
15	ITMA	5,948	6,008	4,707	1,588	1,588	1,588	1,200	Audio	0	0
					3,508	3,508	3,508	3,600	Other	0	0
16	TLA	2,421	2,175	2,175	7,539	17,454	7,539	13,563	Audio	0	6,000
					2,054	2,054	2,054	0	(images)		0
					1,745	1,745	1,745	0	(Video)		0
17	NLL	5,794	4,759	2,071	2,071	2,071	2,071	2,000	Audio	0	0

						2,071	2,071	2,000	Images	0	0
18	OEM	17,654	17,631	17,631	17,497	17,497	17,497	12,000	Audio	0	0
19	RBB	134,956	153,657	111,252	119,585	119,585	119,585	105,000	Audio	0	0
20	TAD	115,556	76,258	52,282	38,235	38,235	38,235	38,000	Audio	0	0
21	SB	14,352	12,080	12,064	12,050	12,050	12,050	11,900	Audio	0	0
22	ONB		6,764	1,691	1,691	152,977	152,977	150,000	Images	0	0
23	FCSH		6,454	0	0	6,454	0	39,000	Audio	32,546	32,546
				204	204	204	204	2,700	Text	2,496	2,496
						5,000			(images)	0	
24	CCE		14,537	14,481	13,426	14,481	13,426	15,800	Audio/vi deo	1,400	1,400
Total audio						590,568	552,565	528,433			
Total other						247,525	210,769	224,980			

4 Next steps

- Ensuring that all MINT transformations are published on Europeana as quickly as possible;
- Revising records where necessary, aiming to provide the best possible user experience;
- Sourcing additional content where possible;
- Working with Associate Partners as they use MINT to publish further material on Europeana.

5 Summary

Key performance indicators

Overall, the data providers have now exceeded the Key Performance Indicators 1, 2 and 3 for the final year of the project:

KPI 1 target: 500,000 audio recordings

Number on Europeana: 552,565

Number transformed on MINT: 590,568

KPI 2 target: 225,000 non-audio digital objects

Number on Europeana: 210,769

Number transformed on MINT: 247,525

KPI 3 target: 90,000 freely re-usable objects

Number on Europeana: 177,248.

The total amount of digital objects transformed on MINT (590,568) now exceeds the total amount as set out in the DOW (528,433). 13 out of the 18 data providers have now met or exceeded their overall targets as set out in the DOW; 16 out of the 18 have now reached at least 90% of their targets.

6 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	D1.7 Aggregation Report 2 http://pro.europeana.eu/files/Europeana_Professional/Projects/Project_list/Europeana_Sounds/Deliverables/europeanasons-d1.7-aggregation-report-2-v1.3.pdf

Appendix A: Terminology

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

Additional terms are defined below:

Term	Definition
APEX	Archives Portal Europe network of excellence
EC-GA	Grant Agreement (including Annex I, the Description of Work) signed with the European Commission
EDM	Europeana Data Model
PC	Project Coordinator
PI	Performance Indicator
PMB	Project Management Board
TEL	The European Library
WP	Work Package