



EUROPEANA SOUNDS

Project Number: 620591

D1.3 Ontologies for Sound

Document Identifier: EuropeanaSounds-D1.3-Ontologies-for-sound-v1.2.docx

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

Date: 31/08/2014

Abstract

This document provides details on the ontologies (controlled vocabularies) selected and analysed for the Europeana Sounds Project, defined as Deliverable D1.3 in the Description of Work [REF 1] and builds on the document produced for Milestone *MS3 Initial ontologies selected* [REF 2]. It includes information about new SKOS vocabularies being built by the task group working on *T1.2 Ontologies* as well as information about existing linked open vocabularies and usage recommendations. This task is connected to task *T1.3 EDM profile* in which the Europeana Data Model is extended to accommodate metadata for audio and audio-related materials.

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

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



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



I. COPYRIGHT NOTICE

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

II. REVISIONS

Version	Status	Author	Partner	Date	Changes
0.1	ToC	Andra Patterson	BL	18/07/2014	
1.0	Draft	Andra Patterson & Christian Morbidoni, with input from Sergiu Gordea	BL NET7 AIT	06/06/2014	First draft
1.1	Second draft	Andra Patterson & Christian Morbidoni, with input from Sergiu Gordea & David Haskiya	BL NET7 AIT EF	15/08/2014	Integrated feedback from reviewers Antoine Isaac and Valentine Charles
1.2	Final	Andra Patterson	BL	29/08/2014	Final amendments

III. DELIVERY SLIP

	Name	Partner/WP	Date
Document Author	Andra Patterson andra.patterson@bl.uk	BL / WP1	28/08/2014
Reviewed by	Reviewers: Antoine Isaac Valentine Charles	EF / WP1 EF / WP1	15/08/2014
Approved by	Coordinator & PMB		29/08/2014

IV. DISTRIBUTION

No.	Date	Comment	Partner / WP
1	31/08/2014	Submitted to the European Commission	BL / WP7
2	31/08/2014	Posted on Europeana Pro website	BL / WP7
3	31/08/2014	Distributed to project consortium	BL / WP7

V. APPLICATION AREA

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

VI. DOCUMENT AMENDMENT PROCEDURE

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

VII. TERMINOLOGY

A complete project glossary is provided at the following page:

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

Further terms are defined below as required:

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

VIII. PROJECT SUMMARY

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

- Double the number of audio items accessible through Europeana to over 1 million and improve geographical and thematic coverage by aggregating items with widespread popular

appeal such as contemporary and classical music, traditional and folk music, the natural world, oral memory and languages and dialects.

- Add meaningful contextual knowledge and medium-specific metadata to 2 million items in Europeana's audio and audio-related collections, developing techniques for cross-media and cross-collection linking.
- Develop and validate audience specific sound channels and a distributed crowd-sourcing infrastructure for end-users that will improve Europeana's search facility, navigation and user experience. These can then be used for other communities and other media.
- Engage music publishers and rights holders in efforts to make more material accessible online through Europeana by resolving domain constraints and lack of access to commercially unviable (i.e. out-of-commerce) content.

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

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

IX. STATEMENT OF ORIGINALITY

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

X. EXECUTIVE SUMMARY: D1.3 ONTOLOGIES FOR SOUND

In this document we define controlled vocabularies that characterise audio objects. These vocabularies extend the Europeana ontology and the definition of the Europeana Data Model (EDM) profile for sound which is being developed in deliverable D1.4. The vocabularies selected and analysed for Europeana Sounds pertain to the following aspects of audio resources: Genre/Form, Mood, Subject, Place, Physical Carrier, Digital Format, Medium of Performance, Personal/Corporate Name, Role, and Work. The first section of this document outlines the working methods of the T1.2 Ontologies task group and provides technical information about searching for existing vocabularies, the selection criteria, and how the task group integrated, formalised and published the new SKOS vocabularies built for this project. The second section describes existing, open vocabularies and the new vocabularies created by this task under each aspect (Genre/Form, Mood, Subject etc.), with links, examples, explanatory text and usage recommendations. The third section provides a quick reference guide for data providers, who will make use of vocabularies when mapping to the Europeana Data Model (EDM) during ingestion. This task is connected to task T1.3 EDM profile in which the Europeana Data Model is extended to accommodate metadata for audio and audio-related material. Examples of how these vocabularies are expressed in the EDM will be included in D1.4 *EDM Profile for Sound*, due for completion at the end of September 2014.

TABLE OF CONTENTS

1	INTRODUCTION.....	8
1.1	General methodology.....	9
1.2	Analysis – Vocabulary selection and building criteria	9
1.3	Integration – linking	10
1.3.1	CSV as interchange format.....	10
1.3.2	Linking to Freebase and DBpedia	11
1.3.3	Search APIs.....	13
1.3.4	Hierarchical Structuring	15
1.3.5	Spreadsheet representation	15
1.4	Formalisation & Publishing.....	16
1.4.1	Creating the SKOS representation.....	16
1.4.2	Visualisation and documentation.....	17
1.4.3	SKOS repository and maintenance	17
1.4.4	Summary of tools used	18
2	SELECTED VOCABULARIES AND USAGE RECOMMENDATIONS.....	19
2.1	Genre/Form vocabularies.....	19
2.1.1	Europeana Sounds Broad Genre	19
2.1.2	Library of Congress Genre / Form	20
2.1.3	FAST	21
2.1.4	EBU genres vocabulary	23
2.1.5	DISMARC (dmGENRE)	25
2.1.6	DBpedia.....	26
2.1.7	Freebase.....	27
2.1.8	New genre / form vocabularies developed in the task	27
2.1.9	Usage recommendations: genres.....	38
2.2	Mood vocabularies	39
2.3	Subject vocabularies.....	40
2.3.1	LCSH.....	41
2.3.2	FAST	41
2.3.3	DBpedia.....	42
2.3.4	RAMEU.....	42
2.3.5	Places as subjects	43
2.3.6	Personal and corporate names as subjects	43
2.3.7	Works as subjects.....	43
2.3.8	Usage recommendations: subjects	43
2.4	Vocabularies for places.....	43
2.4.1	GeoNames.....	44
2.4.2	Getty Thesaurus of Geographic Names Online.....	44
2.4.3	Place names in LC	44
2.4.4	Geographic names in FAST.....	45
2.4.5	Geographic names in VIAF.....	45
2.4.6	Usage recommendations: places	46
2.5	Vocabularies for physical audio carriers	46
2.5.1	RDA carrier types.....	46

2.5.2	LC Carriers Scheme	48
2.5.3	DISMARC (dmFormats)	48
2.5.4	New Europeana Carrier Types scheme	50
2.5.5	Usage recommendations: physical carriers	52
2.6	Vocabularies for digital formats	53
2.6.1	RDA encoding formats	53
2.6.2	Format descriptions	54
2.6.3	Usage recommendations: digital formats	55
2.7	Vocabularies for medium of performance	55
2.7.1	LCMPT	55
2.7.2	Freebase instruments	56
2.7.3	MIMO-VOC	57
2.7.4	RAMEAU	57
2.7.5	Usage recommendations: medium of performance	58
2.8	Vocabularies for personal and corporate names	58
2.8.1	VIAF	58
2.8.2	ISNI	59
2.8.3	LC Name Authorities	59
2.8.4	RISM (Répertoire International des Sources Musicales)	60
2.8.5	Usage recommendations: personal and corporate names	60
2.9	Vocabularies for roles	60
2.9.1	RDA relationship designators	61
2.9.2	MARC code list for relators	61
2.9.3	DISMARC (dmsvAgentRoles)	61
2.9.4	Usage recommendations: roles	62
2.10	Vocabularies for works	62
2.10.1	VIAF	62
2.10.2	LC Name/titles and Title authorities	63
2.10.3	Recommendations: works	63
3	SUMMARY OF RECOMMENDATIONS	64
4	APPENDIX	67
4.1	Appendix I. EDM Profile for Sounds - Collection of Use Cases template	67
5	REFERENCES	70

1 INTRODUCTION

Using shared, controlled vocabularies in metadata records, especially multilingual vocabularies, enables users to search and navigate across different metadata sets, enhancing discoverability. For Europeana Sounds task *T1.2 Ontologies* we focused on enhancing ontologies by using controlled vocabularies with standardised semantic representation (SKOS)¹. Task T1.2 identifies mandatory and recommended vocabularies for data providers (multilingual when this information is available) to use in their metadata for audio and audio-related objects.

The T1.2 task group has investigated a core set of vocabularies that are available for re-use and model the main aspects of audio resources: Genre/Form, Mood, Subject, Place, Physical Carrier, Digital Format, Medium of Performance, Personal/Corporate Name, Role, and Work. This was achieved by researching existing vocabularies in diverse formats and evaluating them against the needs of the project. The primary focus was targeting vocabularies that are already available in SKOS format, or vocabularies that are open for re-use and convertible to RDF/SKOS². We were able to reuse existing vocabularies for almost all aspects of audio resources. However for music genres there was no single preferred SKOS vocabulary in existence and it was necessary to build a new one by combining, matching and refining music genre vocabularies from different sources. Additionally, SKOS vocabularies for non-music genres and for audio carriers are also under construction, to assist data providers in providing full metadata.

The task is connected with task *T1.3 EDM profile*, in which an EDM profile is being developed to incorporate metadata for audio content. For this reason the *EDM Profile for Sounds - Collection of Use Cases* template [Appendix I] was designed by the T1.3 task group, with input from the T1.2 task group, to collect insights, use cases and data samples from data providers. Disseminated to data providers in April 2014, the template included questions about the current use of vocabularies by data providers and asked for their needs for metadata enrichment. Analysis of the use cases and data samples developed our understanding of the vocabularies already in use by data providers, and has guided the investigations of vocabularies by the T1.2 task group.

Section 1.1 of this document provides an overview of the general methodology and tools used during this task. Section 2 lists and summarises features of the selected vocabularies for the different aspects of audio resources that were addressed by the task group, and provides usage recommendations. In Section 3 we summarise the recommendations of the task group in a table which serves as a quick reference guide for data providers.

¹ SKOS: Simple Knowledge Organization System. <http://www.w3.org/2004/02/skos/>. In a semantic web context, "ontologies" are artefacts that govern the structure and semantics of (meta)data, closer to what is investigated in task T1.3, which defines an EDM profile for sound.

² RDF: Resource Description Framework <http://www.w3.org/RDF/>

1.1 General methodology

Task T1.2 Ontologies began in April 2014, the task group comprising people from Net7 (Task Lead), British Library (BL), Rundfunk Berlin-Brandenburg (RBB), National Technical University of Athens (NTUA), Austrian Institute of Technology (AIT) and The Language Archive of the Max Planck Gesellschaft TLA. One of the first tasks was to establish the relationship between this task and task T1.3 EDM profile, as any vocabularies selected under T1.2 need to be accommodated in the EDM profile for sound.

The use cases collected from data providers in April-May 2014 proved to be a good starting point for investigating vocabularies for digital audio objects, as they provided concrete examples of vocabularies already used by data providers in their metadata. Information about different vocabularies was first collected in a spreadsheet, which was then expanded into a shared working document. The task group began by discussing vocabularies in group Skype calls. This led to the formation of smaller groups with more targeted expertise, who worked on investigating specific types of vocabularies, such as musical genre, audio carriers/formats, musical instruments, etc. Data providers contributed valuable expertise to these investigations and many of the vocabularies we focus on in this document are already known to data providers. Discussions about the technical aspects of SKOS vocabulary construction were centred around the creation of Europeana SKOS vocabularies for music genres and non-music genres.

In the following section we summarise the steps we performed to select and analyse existing vocabularies and build new vocabularies.

1.2 Analysis – Vocabulary selection and building criteria

In selecting relevant vocabularies to recommend or to use as starting point for new vocabularies, we considered the following characteristics:

- *Size and structure:* The vocabularies will be used both to support data providers in mapping their metadata and to drive end-user navigation and search. We favoured vocabularies with hierarchical structures where each concept has a relatively small number of sub-concepts. However, a trade-off has to be considered between the number of children associated with a parent concept and the depth of the hierarchy. Too many levels could be confusing and difficult to navigate. Some of the vocabularies investigated (e.g. DISMARC³, DBpedia⁴) lack a complete structure and not all the concepts are consistently linked to parent and child concepts.
- *Documentation:* We favoured vocabularies with user-friendly, understandable labels and where the master language was English, as this provides the best foundation for linking

³ <http://www.dismarc.org/>

⁴ <http://dbpedia.org/About>

concepts to linked open datasets such as Freebase⁵ and DBpedia. The main goal of this alignment is to enable the retrieval of multilingual labels and descriptions and in regard to this we found Freebase to be more complete in terms of number of languages supported.

- *Relations between terms.* The primary relationships between terms within the SKOS representation are given by the hierarchical organisation of the vocabulary though the specification of the *broader* and *narrower* concepts. This provides a fast overview of the meaning of a vocabulary term while enabling the implementation of advanced search and browse functionality. However, terms that are too broad or too narrow have limited value for search engines⁶. Regarding the proper understanding and usage of vocabulary terms, it is particularly important to exploit multilingual definitions and usage examples that are available in the semantic web, such as in Freebase and DBpedia. Where available, exact matches of the concepts were identified within these semantic web resources; in other cases, we matched terms to broader concepts.
- *Access.* The way a vocabulary is made accessible can make a difference to the ease of automatic and manual linking and to extending the vocabulary or extracting sub-hierarchies, as was done in the Music genres schema.

The availability of a dump (e.g. of RDF files) is useful, but we favoured vocabularies with a proper linked data interface so that links to concepts could be dereferenced in a standard way, and with APIs for searching and querying, as these are important for working with the vocabularies. In the case of music genres we used Freebase APIs and the DBpedia SPARQL endpoint to gather relevant data to analyse and use as input to the Europeana Music Genre Vocabulary. Unfortunately many vocabularies do not have querying APIs and only provide dumps or a linked data interface, which is not enough to meaningfully analyse a vocabulary and semi-automatically extract useful information. In cases where we derived new SKOS schemas from existing ones, we linked as much as possible to “original concepts” to enable future schema updates.

1.3 Integration – linking

1.3.1 CSV as interchange format

The technical activities in this task have been performed using a variety of tools. For creating the new Music Genre Vocabulary, the use of relational databases (Microsoft Access) and Java code⁷ provided a convenient way of retrieving data from the starting vocabularies DISMARC and Freebase. Some operations were easily achieved using Open Refine⁸, such as assigning URLs to concepts, mapping them into a SKOS compliant structure and exporting in RDF format. As most of the operations were done using table-like data we found it convenient to adopt CSV⁹ as an interchange

⁵ <https://www.freebase.com/>

⁶ <http://en.wikipedia.org/wiki/Tf%E2%80%93idf>

⁷ <https://github.com/gsergiu/music-genres/>

⁸ <http://openrefine.org/>

⁹ Comma-separated Variables

format between different tools, as it is simple to use and widely supported, for example by Open Refine, Excel, Access and Google spreadsheets.

1.3.2 Linking to Freebase and DBpedia

Both Freebase and DBpedia knowledge bases are built based on Wikipedia articles, which are edited by the crowd in multiple languages. Consequently, different descriptions are not always consistent with each other as the content of the articles in multiple languages is provided by different users, in different moments in time, being influenced by their cultural background (e.g. see the description of Classical Music¹⁰). Additionally, the categorisation of the concepts is not complete: concepts can often be assigned to more media types (e.g. see Ballad¹¹ or Ballet¹²), but this list is not guaranteed to be complete (e.g. see Tarantella¹³). Such inconsistencies are an important limitation in the effectiveness of searching and in the automatic identification of matching concepts with different knowledge sources. Because of this, it is necessary for domain experts to verify and validate the vocabularies.

In building the Music Genre Vocabulary it was particularly important to link the DISMARC vocabulary that is already used in Europeana with the rich descriptions available in linked open data repositories. Some of the DISMARC music genres are mentioned in Wikipedia but do not have a Wikipedia article, and consequently they are not recognized as topics in Freebase and DBpedia (see Chakri¹⁴ for example). In such cases, references to a particular concept need to be manually searched in Wikipedia, with the referencing categories collected as broader matches for the concept (i.e. Music of Kashmir¹⁵ in the case of Chakri).

Using a reconciliation service in Open Refine

One method that can be used to link concepts semi-automatically in a vocabulary with corresponding entities in Freebase and DBpedia is to use Open Refine in combination with a “reconciliation service”.

We used this approach to establish links from the Audio Carriers, Digital Formats and Radio vocabularies. If the vocabulary is in CSV (or Excel) format, the first step is to load it into Open Refine by creating a new project. Open Refine supports reconciliation services, which are http APIs that when given a natural language name, attempt to find matching entities in a given dataset.

¹⁰ Classical music description in Freebase <https://www.freebase.com/m/Oggq0m>

¹¹ Ballad concept description in Freebase <https://www.freebase.com/m/O1gjw>

¹² Ballet concept description <https://www.freebase.com/m/Od6n1>

¹³ Tarantella concept description in Freebase <https://www.freebase.com/m/O1pkpt>

¹⁴ Chakri mentioned in Music of Kashmir http://en.wikipedia.org/wiki/Kashmiri_music#Chakri

¹⁵ http://dbpedia.org/resource/Music_of_Kashmir,_Jammu_and_Ladakh

The standard reconciliation service in Open Refine used to be the service provided by Freebase¹⁶. Linking to Freebase was considered very important as Freebase contains more multilingual names and descriptions than DBpedia. Unfortunately the Freebase reconciliation service is no longer supported by Open Refine (it is no longer maintained by Google). For this reason we matched concepts against DBpedia and then derived (following an owl:sameAs link) the URL of the corresponding entity in Freebase. Our approach was as follows:

1. Create a reconciliation service on top of the public DBpedia SPARQL endpoint¹⁷. This can be done by means of the RDF extension by choosing “Add reconciliation service” then “Based on SPARQL endpoint”, and then typing the address of the desired SPARQL endpoint (DBpedia in our case). This is shown in the following screenshot:

Add SPARQL-based reconciliation service

Name:
A human readable name

Endpoint details

Endpoint URL:

Graph URI:
Leave empty to use the default graph

Type:
This determines the syntax that will be used for search

Label properties

Select properties that are used to label resources in the endpoint. These properties will be used to match resources:

rdfs:label skos:prefLabel dcterms:title dc:title
 foaf:name
 Other...

2. Choose the column where the names of the concepts are collected, and start automatic reconciliation against the newly created service.

¹⁶ Freebase reconciliation service <http://wiki.freebase.com/wiki/Reconciliation>

¹⁷ DBpedia SPARQL endpoint, <http://dbpedia.org/sparql>

3. Manually analyse the results and confirm matches. In cases where no match was found for a given concept we searched DBpedia manually and chose an appropriate match.

Tip: to search DBpedia sometimes it is convenient to search en.wikipedia.org and then, once the desired entity is found, derive the DBpedia URL of the corresponding entity (e.g. the corresponding entity of <https://en.wikipedia.org/wiki/Speech> will be <http://dbpedia.org/resource/Speech>)

4. Once the links to DBpedia were established we queried the DBpedia SPARQL endpoint to get the corresponding URL in Freebase (if any). As DBpedia keep such information a SPARQL query like the following would return the Freebase URL:

```
select distinct ?def where {
  {
    DBPEDIA_URL <http://dbpedia.org/ontology/abstract>
    ?def.
    FILTER (LANG(?def) = "" || LANGMATCHES (LANG(?def),
"en"))
  } UNION {
    DBPEDIA_URL
    <http://dbpedia.org/ontology/wikiPageRedirects> ?x.
    ?x <http://dbpedia.org/ontology/abstract> ?def.
    FILTER (LANG(?def) = "" || LANGMATCHES (LANG(?def),
"en"))
  }
}
```

Tip: sometimes matched entities in DBPedia simply redirect to other equivalent entities. The second query (after the UNION) addresses this issue.

To automatically perform such a query for all the DBpedia matched entities we used the “Add column by fetching URL” option of Open Refine. In this way one call to the DBpedia endpoint for each row is performed automatically and the result is stored in a new column.

More information on how to use the Open Refine reconciliation service can be found at <http://freemetadata.org/reconciliation/>.

1.3.3 Search APIs

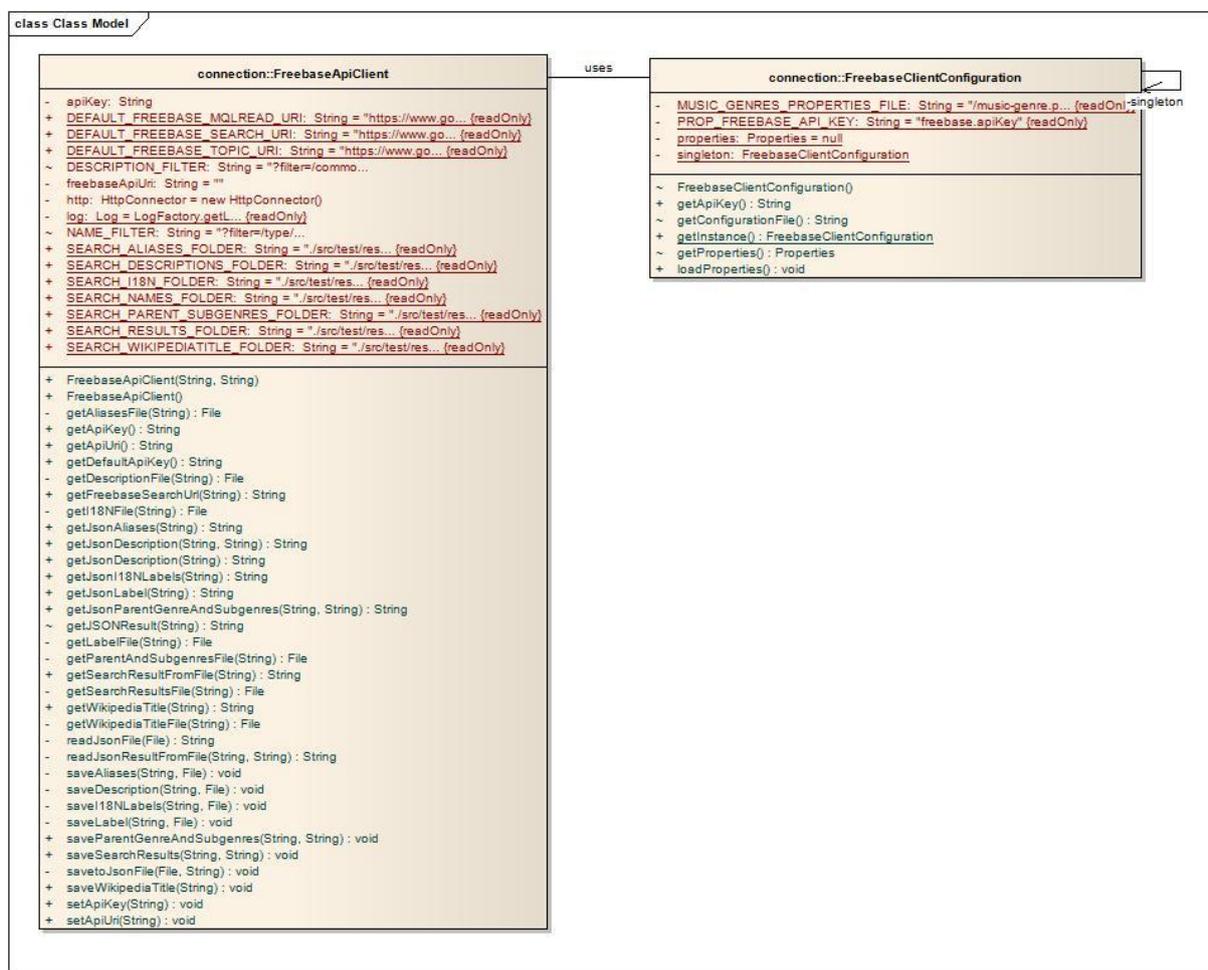
In Freebase, Composition Type and Music Genre are two different classes of concepts. In many cases composition types are also music genres, but they are not always annotated correctly. Similarly, many popular music genres are primarily known as dance genres (see description of Csardas¹⁸ or

¹⁸ <https://www.freebase.com/m/024kwt>

Charleston¹⁹), and these are often not marked as music genres in Freebase. Similar cases are easily found in DBpedia as well. By using the search API offered by Freebase, a set of candidate concepts can be preselected, allowing the domain experts to easily identify the matching concepts.

After identification of the appropriate concepts in Freebase, the descriptions, DBpedia references, alternative and multilingual labels are collected using the client API developed specially for this purpose (in the case of exact matches), which is shown below:

FreebaseApiClient Class Diagram



¹⁹ <https://www.freebase.com/m/0g965>

1.3.4 Hierarchical Structuring

The SKOS model that was chosen to represent the sound vocabularies is able to model graph structures. The organisation of music genres in linked data repositories is also graph-based. However, the hierarchical structuring (i.e. tree structure) is more appropriate for the implementation of search and browsing functionality. There are multiple criteria²⁰ that are often used to classify music genres, including the triptych of art/popular/traditional, time periods, regional or national origins, fused origins, instrumentations, social functions (e.g. dancing, sacred, ballroom) etc. We follow the same approach and combine these criteria for defining an intuitive structure for the Music Genres Vocabulary which has defined 7 top level categories: Art music, Popular music, Traditional (world) music, Children’s music, Military music, Religious music and Fusion music. For each of these categories, different criteria are appropriate for creating subdivisions. Within the category of art music, “classical” creates a very well-defined subgroup, while “traditional music” can be best classified by regional origins. “Popular music” includes several major trends (such as folk, rock, country, blues, jazz, electronic, etc.) which are also used as primary stylistic origins for fusion genres. “Military music” and “Children’s music” do not have such a large number of subgenres. “Sacred music” might be subdivided by each religion and its specific rituals.

Often, it is easy to identify the appropriate parent category for a particular genre, given its description. However the classification according to this criteria is either not explicitly stated in Freebase/DBpedia, or it is ambiguous (for example, a list of stylistic origins is provided, but the main one is not explicitly marked). Therefore, human interaction is required for defining an appropriate structure for the vocabulary. Information or evidence of structures available in DBpedia, such as parent genre, subgenres, stylistic origins, fusion genres, and instrumentation, provide a good basis for determining the classification.

1.3.5 Spreadsheet representation

As described in the following section, the SKOS representation of the Music Genre Vocabulary has been derived from a spreadsheet representation. The current version of the spreadsheet, serialised as comma separated values (CSV), can be found in the following google drive folder:

<https://drive.google.com/#folders/0B975aUJtXAm4S1NzbVE2cm85MTQ>

File: **Music-Genres.csv**

The main columns in this file are:

- **skos_id** – the unique identifier for each genre in the vocabulary. We derived the identifier from the English labels. This field was used to create unique URIs for SKOS concepts.
- **Label** - the preferred label for genres. This was mapped to skos:prefLabel in the SKOS representation.
- **description** – the English description for each genre, extracted from Freebase and manually checked. This was mapped to skos:definition in the SKOS representation.

²⁰ See http://en.wikipedia.org/wiki/Music_genre

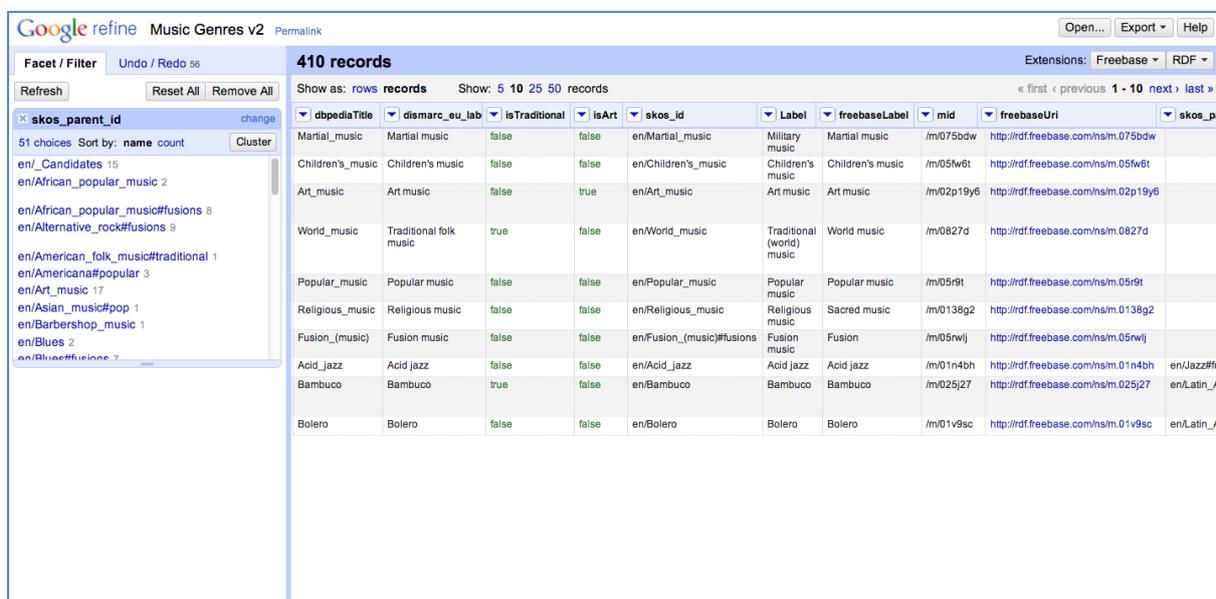
- **skos_parent_id** – the identifier of the parent genre for each genre. This was used to derive skos:broader relation among concepts in the SKOS representation.
- **dbpedia_genre** – the URL of the DBpedia resource corresponding to each genre. This was used to add skos:exactMatch relations in the SKOS representation.
- **freebaseUri** – the URL of the Linked Data entities in Freebase that best match each genre. This was used to derive skos:closeMatch relations in the SKOS representation. As not all the concepts had a precise match in Freebase, we chose to be conservative and use skos:closeMatch instead of skos:exactMatch.
- **broaderMatch** – the URLs of DBpedia resources that represent music genres broader than the ones in the SKOS vocabulary. This was used to derive skos:broadMatch relations. Broad matches were added for each genre that did not have a precise match in DBpedia.

1.4 Formalisation & Publishing

1.4.1 Creating the SKOS representation

The SKOS representation was derived from the tabular data created following the previous steps using the RDF extension²¹ of Open Refine²² (formerly Google refine), one of the most used open-source data curation tools. To do this we imported the tabular data in the form of Comma Separated Values (Excel format can also be used) into Open Refine.

The user interface of the tool, showing how the tabular data is visualised in Open Refine:



dbpediaTitle	dismarc_eu_lab	isTraditional	isArt	skos_id	Label	freebaseLabel	mid	freebaseUri	skos_pa
Martial_music	Martial music	false	false	en/Martial_music	Martial music	Martial music	/m075bdw	http://rdf.freebase.com/ns/m.075bdw	
Children's_music	Children's music	false	false	en/Children's_music	Children's music	Children's music	/m05fw6t	http://rdf.freebase.com/ns/m.05fw6t	
Art_music	Art music	false	true	en/Art_music	Art music	Art music	/m02p19y6	http://rdf.freebase.com/ns/m.02p19y6	
World_music	Traditional folk music	true	false	en/World_music	Traditional (world) music	World music	/m0827d	http://rdf.freebase.com/ns/m.0827d	
Popular_music	Popular music	false	false	en/Popular_music	Popular music	Popular music	/m059t	http://rdf.freebase.com/ns/m.059t	
Religious_music	Religious music	false	false	en/Religious_music	Religious music	Sacred music	/m0138g2	http://rdf.freebase.com/ns/m.0138g2	
Fusion_(music)	Fusion music	false	false	en/Fusion_(music)#fusions	Fusion music	Fusion	/m05nwj	http://rdf.freebase.com/ns/m.05nwj	
Acid_jazz	Acid jazz	false	false	en/Acid_jazz	Acid jazz	Acid jazz	/m01n4bh	http://rdf.freebase.com/ns/m.01n4bh	en/Jazz#Fusion
Bambuco	Bambuco	true	false	en/Bambuco	Bambuco	Bambuco	/m025j27	http://rdf.freebase.com/ns/m.025j27	en/Latin_American_music
Bolero	Bolero	false	false	en/Bolero	Bolero	Bolero	/m01v9sc	http://rdf.freebase.com/ns/m.01v9sc	en/Latin_American_music

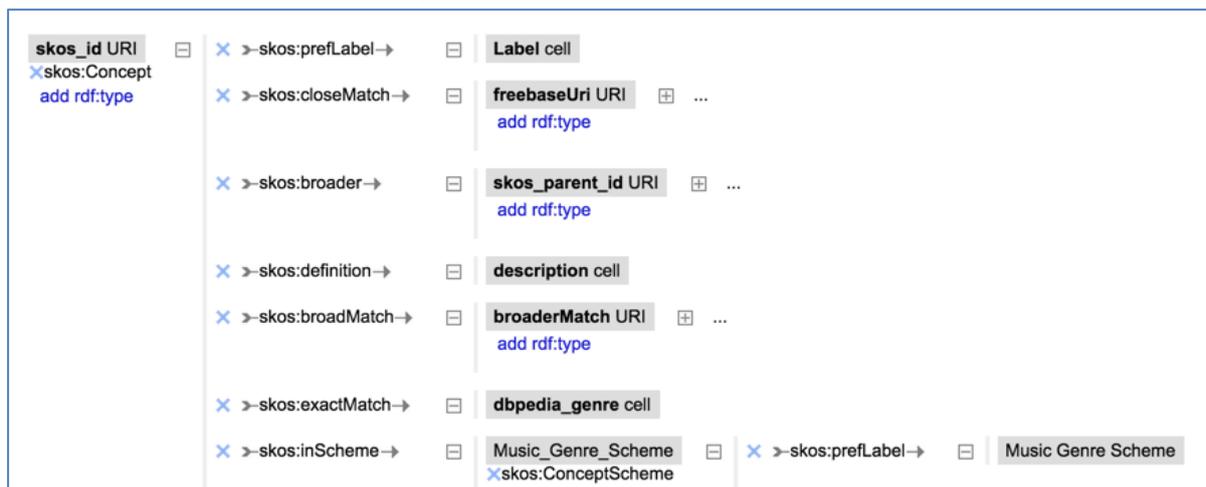
RDF representation of the tables in Open Refine can be generated by first installing the RDF extension and then editing the RDF template. The RDF template is a configuration that allows us to

²¹ Open Refine RDF extension, <http://refine.deri.ie/>

²² Open Refine, <http://openrefine.org/>

map each column with a given triple, where the value in a cell can be used as subject or object. This general methodology can be used to produce a variety of RDF representations. In our case we imported the SKOS namespace into the template and we used SKOS properties and classes to model output triples.

The RDF export template created for the Music Genre vocabulary:



The Open Refine projects containing the spreadsheet representations and the RDF export template used to produce the SKOS representation of the developed vocabularies can be found at the following google drive folder:

<https://drive.google.com/#folders/0BzFrvHHWI4bZZE5NUXNTVHdSNDA>

1.4.2 Visualisation and documentation

There are currently few tools that create simple and readable visualisations of a SKOS vocabulary. We have chosen SKOSPlay²³, an open-source online tool that can be used to generate interactive hierarchical views and to create PDF documentation for each concept included in a vocabulary.

1.4.3 SKOS repository and maintenance

OpenSKOS²⁴ was chosen as a SKOS vocabularies repository as it is already in use by the Europeana Foundation²⁵. An open-source solution, it supports storage and consumption of SKOS vocabularies via REST-API and a web interface to allow domain experts to edit the vocabularies. The REST-API facilitates integration with other applications, for example in the end-user facing portal channels which will be developed in WP4. Vocabularies produced by this task will be hosted at the OpenSKOS instance provided and maintained by Europeana.

²³ SKOSPlay, <http://labs.sparna.fr/skos-play>

²⁴ OpenSkos: <http://openskos.org/>

²⁵ Used for a World War I vocabulary in *Europeana 1914-1918*

1.4.4 Summary of tools used

The following table summarises the different tools we used to achieve specific goals, such as extracting, converting or visualising SKOS vocabularies:

Table 1: Summary of tools

Tool	Description
Freebase API https://developers.google.com/freebase/	Search Freebase for appropriate concepts
DBpedia API https://github.com/dbpedia/lookup	Search DBpedia for appropriate concepts
Json-toCSV http://www.convertcsv.com/json-to-csv.htm	Convert json (e.g. Freebase data) to .csv representation
SkosPlay http://labs.sparna.fr/skos-play/	A web tool to visualise SKOS vocabularies
Open Refine http://openrefine.org/	A data curation tool that includes some useful features in this context: RDFisation and semi-automatic reconciliation (linking) with Freebase, DBpedia and others

2 SELECTED VOCABULARIES AND USAGE RECOMMENDATIONS

2.1 Genre/Form vocabularies

Genre and form concepts describe what an object *is*, rather than what it is about. Genre concepts reflect the style of works, for example “Punk”, “Classical music”. Form concepts reflect the structure or purpose of works, for example “Sonatas”, “Portraits”, “Interviews”, “Military music”. Some vocabularies combine genre and form concepts into a single concept, such as “Fantasy radio programs” (“Fantasy”=genre + “radio programs”=form).

2.1.1 Europeana Sounds Broad Genre

The T1.2 task group has defined a set of broad genre concepts based on the broad categories outlined in the Europeana Sounds *D1.1 Content selection policy*²⁶. In the table below, the D1.1 broad categories are in the left-hand column while the right-hand column shows the corresponding Europeana Sounds Broad Genre concept which data providers will add to their metadata during ingestion, allowing Europeana Sounds material to be grouped into broad genre categories. Although these concepts are currently expressed only with English terms, it is envisaged that the concepts can be expressed with multilingual terms through automatic enrichment to Freebase and DBpedia.

Table 2: Broad Categories and Genres

D1.1 Broad Category	Europeana Sounds Broad Genre
Music recordings Includes: Classical/Art music from western and non-western cultures; Traditional/Folk music; Popular music	Music
Spoken word recordings Includes: Oral memories; Languages & dialects; Spoken word performances	Spoken word
Radio recordings Includes: Any recorded sound broadcast on radio	Radio

²⁶ D1.1 Content Selection Policy <http://pro.europeana.eu/documents/2011409/70e86a37-7608-4f37-9507-1b6f6152cd42> [REF 3]

Environment recordings Includes: Natural sounds; Soundscapes	Environment
Audio-related material Includes: Music scores; Texts such as interview transcripts; Videos; Images	Use one or more of the genres above, as appropriate

2.1.2 Library of Congress Genre / Form

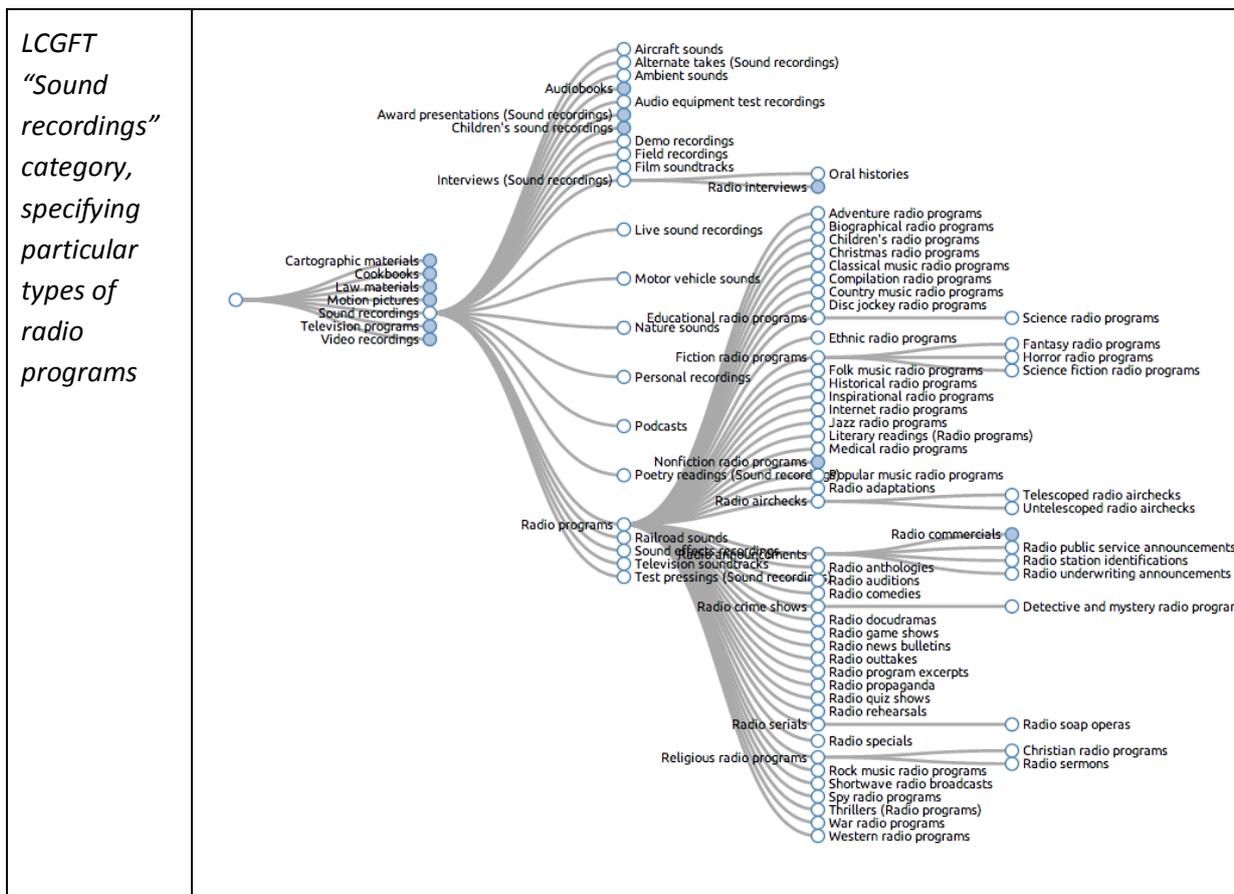
Availability	Search and browse: http://id.loc.gov/authorities/genreForms.html Download: http://id.loc.gov/download/
Terms of use	Free access, and link to data at granular level
Languages	English only
Format	SKOS (example: http://id.loc.gov/authorities/genreForms/gf2011026243)

The *Library of Congress Genre/Form* SKOS vocabulary provides an extensive thesaurus of English language genre/form concepts. These concepts are currently under development, either undergoing conversion from Library of Congress Subject Headings²⁷ (LCSH) to LCGFT, or with additional genre/form concepts being constructed, as appropriate. Some categories have been completed, such as moving images (films, television programs, and video recordings), spoken-word recordings (including radio programs), legal materials, and cartographic materials. Other categories are under development, such as music and religious materials. For these categories there are working documents available online, such as a list of candidate music genre/form terms for discussion²⁸. Due to the developmental nature of this vocabulary, it is currently necessary to use LCSH in tandem with LCGFT, as this is the major source of concepts being converted to LCGFT.

The “Sound recordings” category in LCGFT contains a number of sound types, some of which (e.g. “Radio programs”) are developed to a high level of detail, while others (e.g. “Nature sounds”) are less extensive. Sub-categories of this vocabulary could be used by data providers to extend the Europeana Sounds broad genre concepts “Music”, “Spoken word”, “Radio” and “Environment”.

²⁷ <http://id.loc.gov/authorities/subjects.html>

²⁸ <http://www.loc.gov/catdir/cpsd/musicterms.pdf>

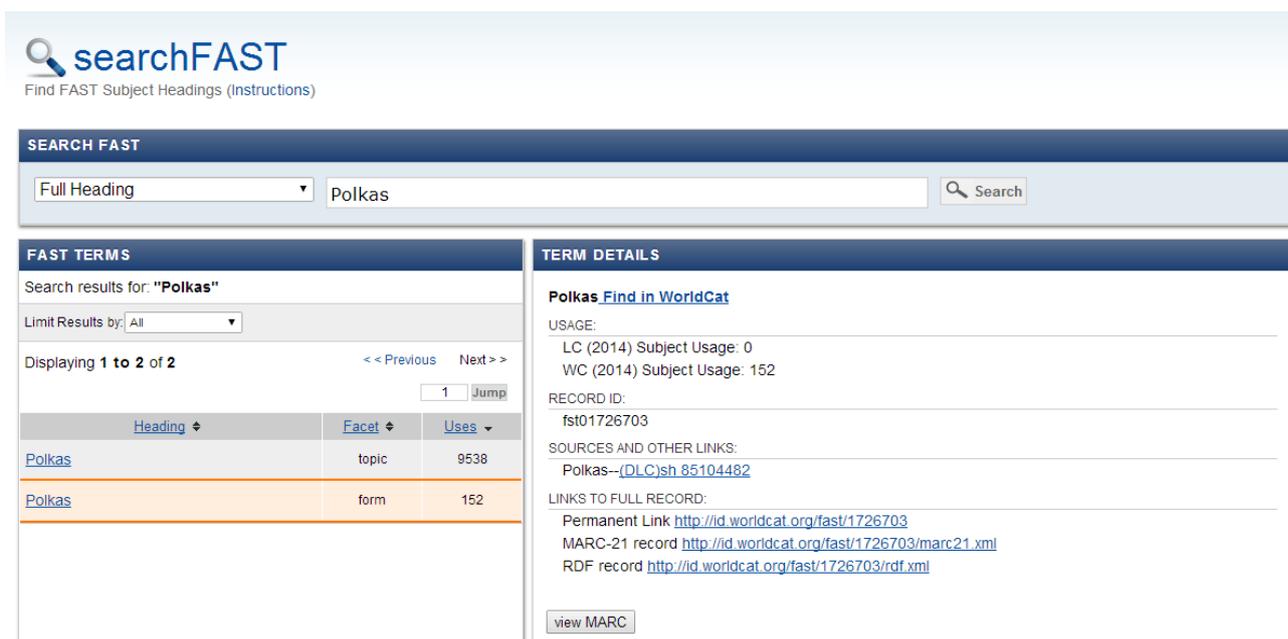


2.1.3 FAST

Availability	Search: http://fast.oclc.org/searchfast/ Download: http://www.oclc.org/research/activities/fast/download.html
Terms of use	Available under the Open Data Commons Attribution License (ODC-By) v1.0
Languages	English
Format	SKOS

Genre/Form concepts are included in FAST (Faceted Application of Subject Terminology), an enumerative, faceted subject heading schema derived from the Library of Congress Subject Headings (LCSH), with a simplified syntax²⁹. Like LCGFT (2.1.2 above), genre/form concepts in FAST are currently under development, with concepts undergoing conversion from subject to genre/form or with additional genre/form concepts being constructed, as appropriate. An advantage of FAST over LCGFT is that it provides a single search mechanism with clear search results. However, like LCGFT, there are concepts that could be used as genre/form which have not yet been established as separate genre/form concepts.

FAST search result showing topic and form concepts



The screenshot shows the searchFAST interface. At the top, there is a search bar with the text 'Polkas' and a search button. Below the search bar, there are two main sections: 'FAST TERMS' and 'TERM DETAILS'.

FAST TERMS

Search results for: "Polkas"

Limit Results by: All

Displaying 1 to 2 of 2

Heading	Facet	Uses
Polkas	topic	9538
Polkas	form	152

TERM DETAILS

Polkas [Find in WorldCat](#)

USAGE:

LC (2014) Subject Usage: 0
WC (2014) Subject Usage: 152

RECORD ID:

fst01726703

SOURCES AND OTHER LINKS:

[Polkas--\(DLC\)sh 85104482](#)

LINKS TO FULL RECORD:

Permanent Link <http://id.worldcat.org/fast/1726703>
MARC-21 record <http://id.worldcat.org/fast/1726703/marc21.xml>
RDF record <http://id.worldcat.org/fast/1726703/rdf.xml>

[view MARC](#)

²⁹ <http://www.oclc.org/research/activities/fast.html?urlm=159754>

FAST search result showing topic concept that could also be a form concept


Find FAST Subject Headings (Instructions)

Full Heading

sonatas

FAST TERMS

Search results for: "sonatas"

Limit Results by: All

Displaying 1 to 1 of 1

<< Previous Next >>

1 Jump

Heading	Facet	Uses
Sonatas	topic	2879

TERM DETAILS

Sonatas [Find in WorldCat](#)

USED FOR:
Sonatinas

USAGE:
LC (2014) Subject Usage: 17
WC (2014) Subject Usage: 2,859 (2,879)

RECORD ID:
fst01125801

SOURCES AND OTHER LINKS:
[Sonatas--\(DLC\)sh 85124818](#)

LINKS TO FULL RECORD:
Permanent Link <http://id.worldcat.org/fast/1125801>
MARC-21 record <http://id.worldcat.org/fast/1125801/marc21.xml>
RDF record <http://id.worldcat.org/fast/1125801/rdf.xml>

2.1.4 EBU genres vocabulary

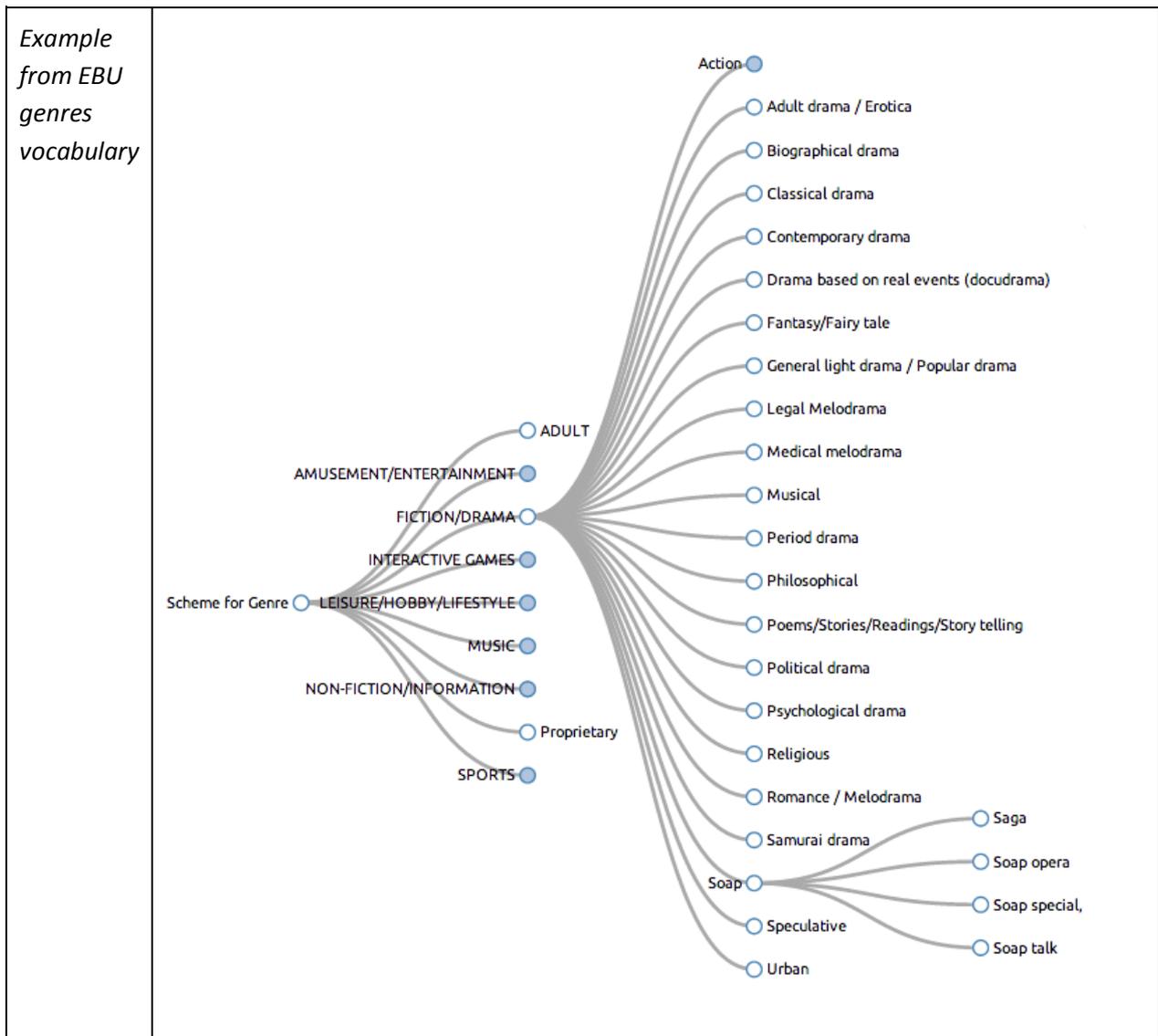
Availability	View: http://www.ebu.ch/metadata/cs/web/ebu_ContentGenreCS_Mapping_p.xml.html Info: https://tech.ebu.ch/MetadataReferenceData
Terms of use	Free to use, licensed under CC BY SA
Languages	English
Format	SKOS at http://www.ebu.ch/metadata/ontologies/skos/ebu_ContentGenreCS.rdf

The EBU³⁰ (European Broadcasting Union) genres vocabulary provides an extensive classification of content genres including broadcast audio and entertainment. The LCGFT and FAST vocabularies described in sections 2.1.2 and 2.1.3 provide concepts that are sometimes a combination of genre

³⁰ <http://www3.ebu.ch/home>

and form, such as “Horror radio programs” (“Horror”=genre + “radio programs”=form). The EBU scheme comprises genre terms only which means that the scheme can be used to characterise a variety of objects, such as television programs, audio recordings, movie clips, etc.

Compared to other more extensive and “neutral” vocabularies like LCGFT and FAST, the EBU scheme provides a coherent and domain specific taxonomy aimed at categorising recorded broadcast programs according to their genre/topic. It provides a possible end-user hierarchical access to data (e.g. a specific broadcasting channel). FAST is easier to search and is available as Linked Data, but is less structured than the EBU scheme.



2.1.5 DISMARC (dmGENRE)

Availability	Search: To browse genres, select Genres tab at http://www.dismarc.org/index.php?form=browse&db=0#tabPaneBrowsecriteria1406892805.993 Download: http://www.dismarc.org/index.php?form=admin.thesauri&task=showThesaurus&thesaurus=dmGenres
Terms of use	
Languages	English
Format	CSV

DISMARC³¹ (Discovering Music Archives) has aggregated music audio and related material for Europeana. It uses a vocabulary of genres comprising concepts supplied by DISMARC data providers. The genre vocabulary can be supplied in a SKOS version. To analyse this vocabulary, the task group created a spreadsheet containing the DISMARC dmGenre vocabulary³².

Non-music genres in DISMARC

Providing a top level classification for sound recordings with respect to what produced the original sound (e.g. a human being, an animal, a machine) is a common choice in sound archives (British Library, Library of Congress, etc.). In DISMARC, top level categories follow the pattern as shown below in the top level categories “Nature” and “Animal”. However, further modelling tends to be domain dependent and highlight specific aspects:

Sound effects / Nature / Catastrophes / Volcanoes

or

Sound effects / Animals / Finnish birds

Sometimes, very specific information is organised in a rather subjective way in DISMARC. This is due to the fact that DISMARC has been iteratively enriched by data providers to model their specific needs. For example:

Sound effects / International / Africa / Traffic, transport

³¹ <http://www.dismarc.org/info/>

³² <https://docs.google.com/spreadsheets/d/10ilXZ6XxNEnzVOiy2aRxThHiUcRuACuyLN66BWtfQU/edit>

Music genres in DISMARC

Similarly to non-music genres, music genres in DISMARC are built using a bottom-up approach. On the positive side, this reflects the needs of the data providers using DISMARC but in the negative side the resulting vocabulary lacks a proper hierarchical structure. The vocabulary doesn't provide distinct categorization according to the criteria defined in section 1.3.5 Hierarchical structuring, and it doesn't provide a definition of the listed terms.

The set of genres in DISMARC has been used as a starting point for the creation of the new music genre vocabulary, as this is already used in Europeana by the Dismarc content providers. The vocabulary terms were linked to Freebase and DBpedia and placed into a proper hierarchical structure by using the knowledge available as linked data.

2.1.6 DBpedia

Availability	Search API: https://github.com/dbpedia/lookup Linked Data (example resource): http://dbpedia.org/resource/Rock_music SPARQL: http://dbpedia.org/sparql
Terms of use	
Languages	Multilanguage. Languages are not uniform across resources.
Format	RDF

DBpedia is a crowd-sourced community effort to extract structured information from Wikipedia and make this information available on the Web. The resulting data currently consists of several billion RDF triples and covers domains such as geographic information, people, companies, online communities, films, music, books and scientific publications. The English version of the DBpedia knowledge base currently describes 4.0 million things, out of which 372,000 are creative works including 116,000 music albums. DBpedia resources are multilingual, although the available languages are not uniformly distributed among resources.

Music genres on DBpedia are found under the class term "music genre"³³. The music genre concepts were extracted by the task group into a spreadsheet³⁴ in a flat structure for investigative purposes.

³³<http://dbpedia.org/describe/?url=http%3A%2F%2Fdbpedia.org%2Fontology%2FMusicGenre&graph=http%3A%2F%2Fdbpedia.org>

They were later extracted with the full structure to enable the concepts to be merged with music genre concepts from Freebase and DISMARC to form the Europeana Sounds Music Genre Thesaurus.

2.1.7 Freebase

Availability	Search & download links using APIs: https://developers.google.com/freebase/
Terms of use	CC-BY license (http://www.freebase.com/policies/attribution)
Languages	Multilingual
Format	JSON, RDF

The Freebase music commons contains recording artists, albums, and songs. Data found here is a combination of information sourced from MusicBrainz and Wikipedia - further information sources will be integrated in the future.

The MQL endpoint³⁵ provides access to this information, which is categorised into music genres that the task group put into a spreadsheet³⁶ for investigative purposes. They were later extracted with the full structure to enable the concepts to be merged with music genre concepts from DBpedia and DISMARC to form the Europeana Sounds Music Genre Thesaurus.

2.1.8 New genre / form vocabularies developed in the task

Note about sustainability

Vocabularies produced in this task will be loaded into a repository. The OpenSKOS instance at Europeana has been chosen for this purpose as it provides storage, online/API access and editing functionalities. Vocabularies will be represented in SKOS using the RDF model and will be published with an open license where possible. They will be freely extendable and reusable by SKOS compliant editors. We will provide source files and open source code that can be a starting point for others to reproduce all or part of the data workflow. As this work is still in progress it is not yet possible to provide links to the repositories in this document.

³⁴ https://docs.google.com/spreadsheets/d/1TUttX_GVoCZspT0UagrtToQMLkehFdIEKVqEo2ZMjNM/edit

³⁵ <http://wiki.freebase.com/wiki/MQL>

³⁶ <https://docs.google.com/spreadsheets/d/1v1xTDOCEdyabPmpWT3YswnbMqxJNwd6il6qfzHGMVM/edit#gid=1265737710>

2.1.8.1 Music genre / form scheme

After investigating existing controlled vocabularies that include music genre concepts, the task group found that most of these are not easy for data providers to search, as it is hard to identify music genres in large vocabularies covering many topics, such as LCGFT and FAST. To facilitate the addition of music genres by data providers during ingestion, a new SKOS thesaurus is under construction that links, combines, matches, structures and refines music genres from DBpedia, Freebase and DISMARC. Music-related portals were consulted for identifying the needs from the music retrieval perspective, including Music Brainz³⁷, Music by Mood³⁸, and Amazon Digital Music Store³⁹.

At the time of writing, the task group is analysing the first version of the vocabulary, comparing music concepts with concepts in LCGFT, LCSH and FAST to identify differences, gaps and issues for discussion. The outcome of these discussions will determine the development of the draft vocabulary. Issues include whether to use singular or plural terms as preferred terms, for example “Sonata” or “Sonatas” and whether headings require qualification, for example “Jungle” or “Jungle (Music)”, “Dance” or “Dance music”.

Current status and links

The RDF/SKOS files and the documentation of the vocabulary were saved on Google Drive⁴⁰.

Files:

Music-Genres.rdf - The RDF/SKOS representation of the complete music genres hierarchy with links to Freebase and DBpedia entities.

Music-Genres-i18n.rdf - In addition to the previous vocabulary, this includes alternative labels and multilingual labels extracted from Freebase.

Music-Genres-Schema-Documentation.pdf - The complete documentation of the vocabulary (created with SKOSPlay⁴¹).

The following diagram lists the top level categories and shows the sub-hierarchies of the concepts:

³⁷ <http://musicbrainz.org/>

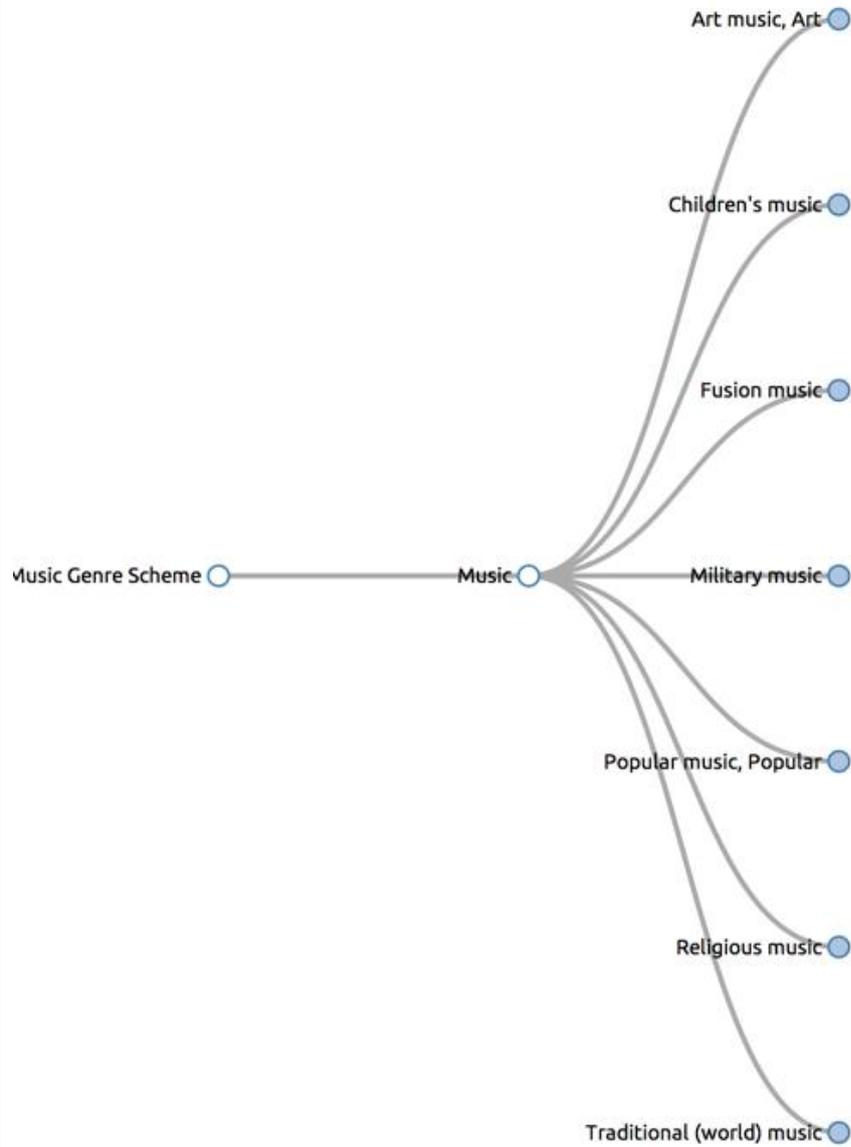
³⁸ <http://musicbymood.aishakaliel.com/>

³⁹ http://www.amazon.com/MP3-Music-Download/b/ref=sd_allcat_digi_str?ie=UTF8&node=163856011

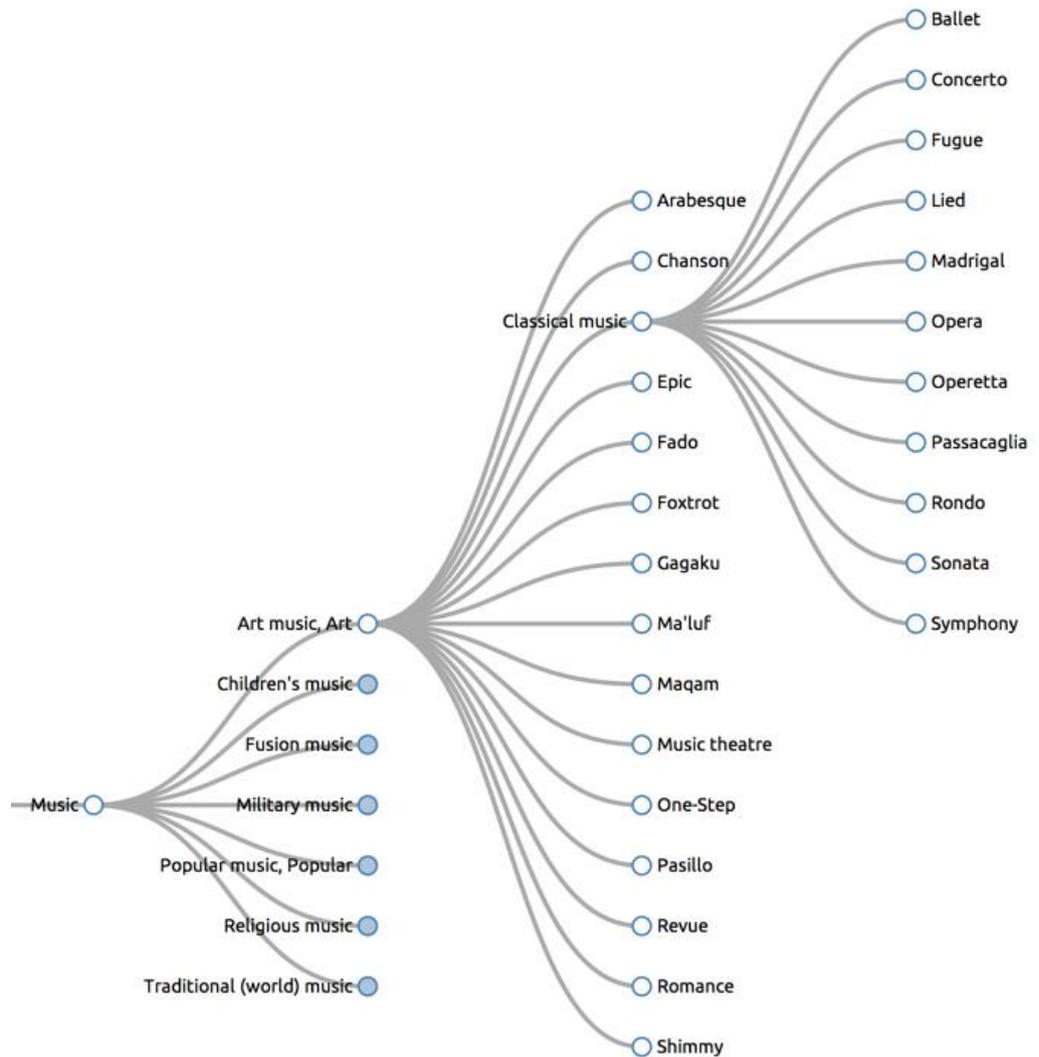
⁴⁰ <https://drive.google.com/#folders/OB975aUJtXAm4S1NzbVE2cm85MTQ>

⁴¹ <http://labs.sparna.fr/skos-play>

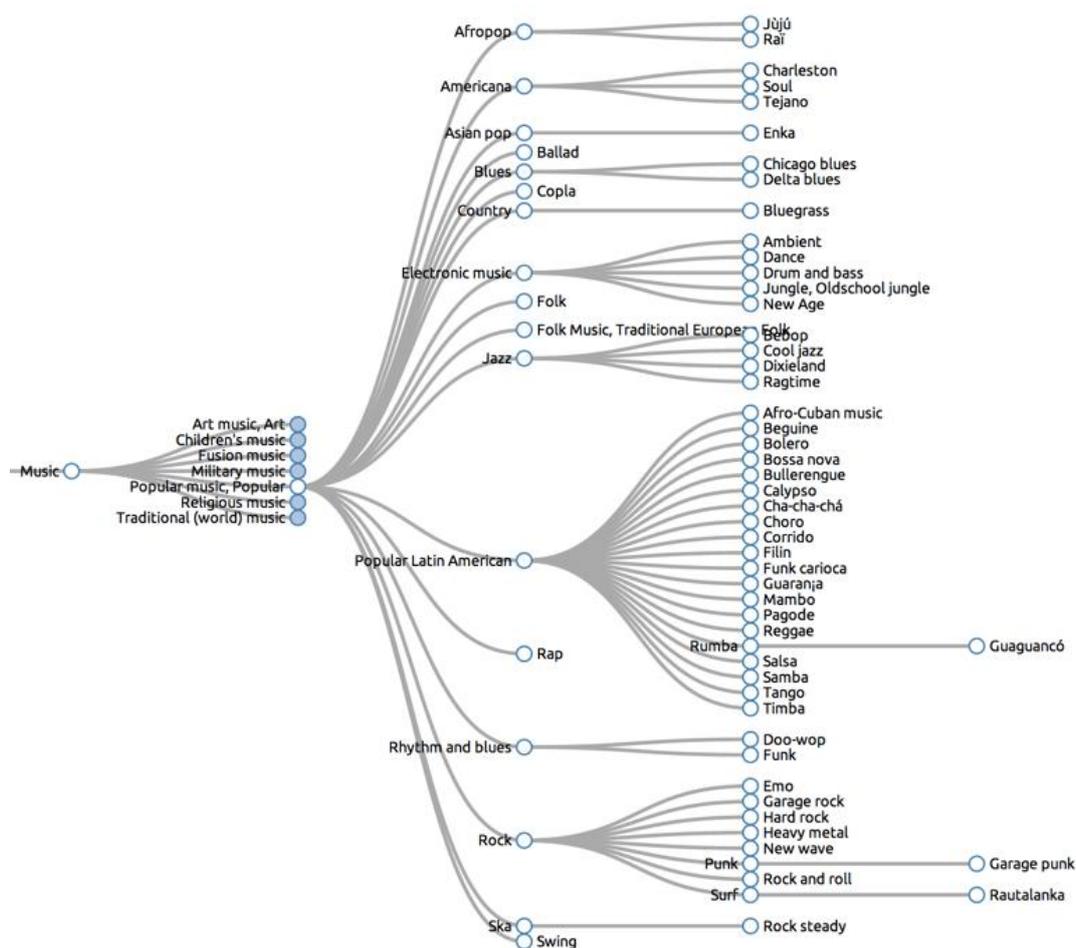
*Top level
categories in
the Music
Genre scheme*



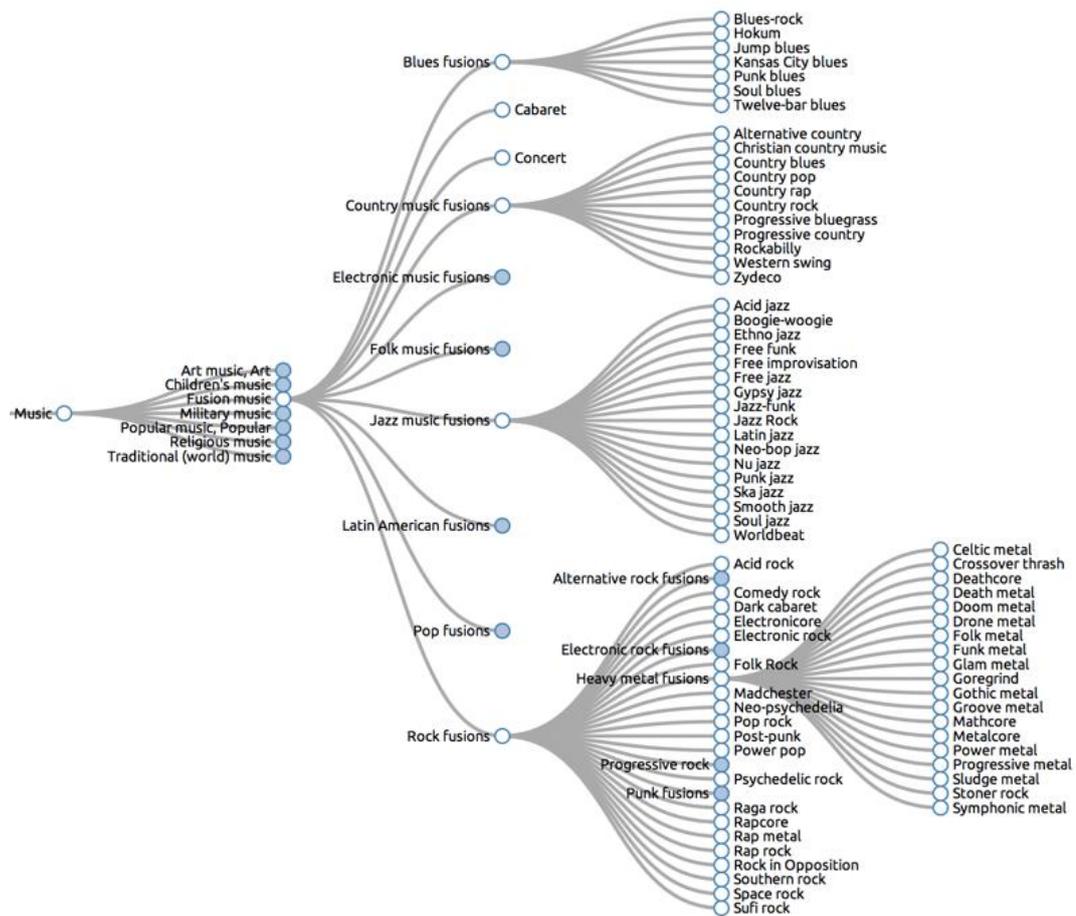
The Art Music hierarchy in the Music Genre scheme

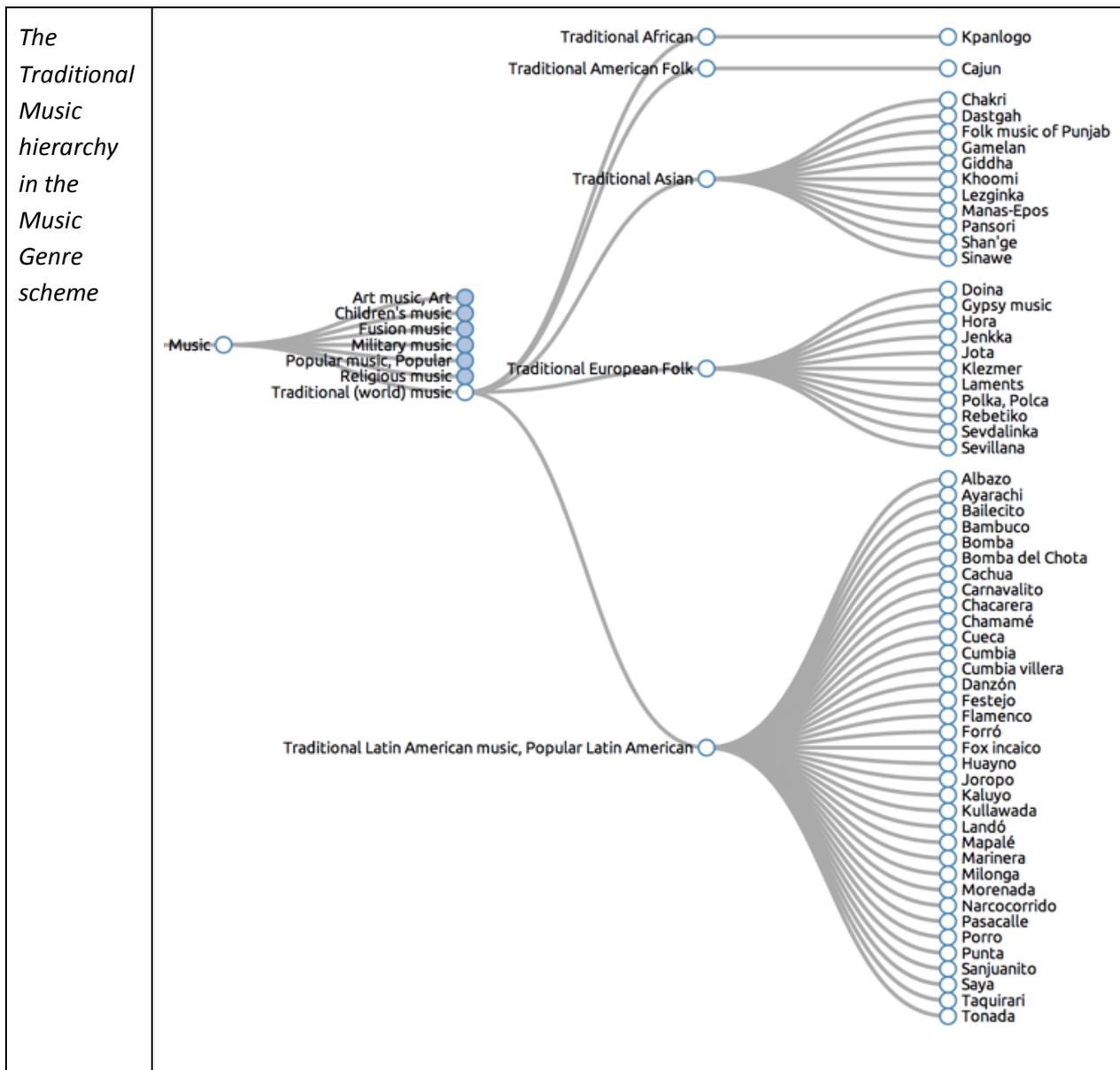


*The
Popular
Music
hierarchy
in the
Music
Genre
scheme*

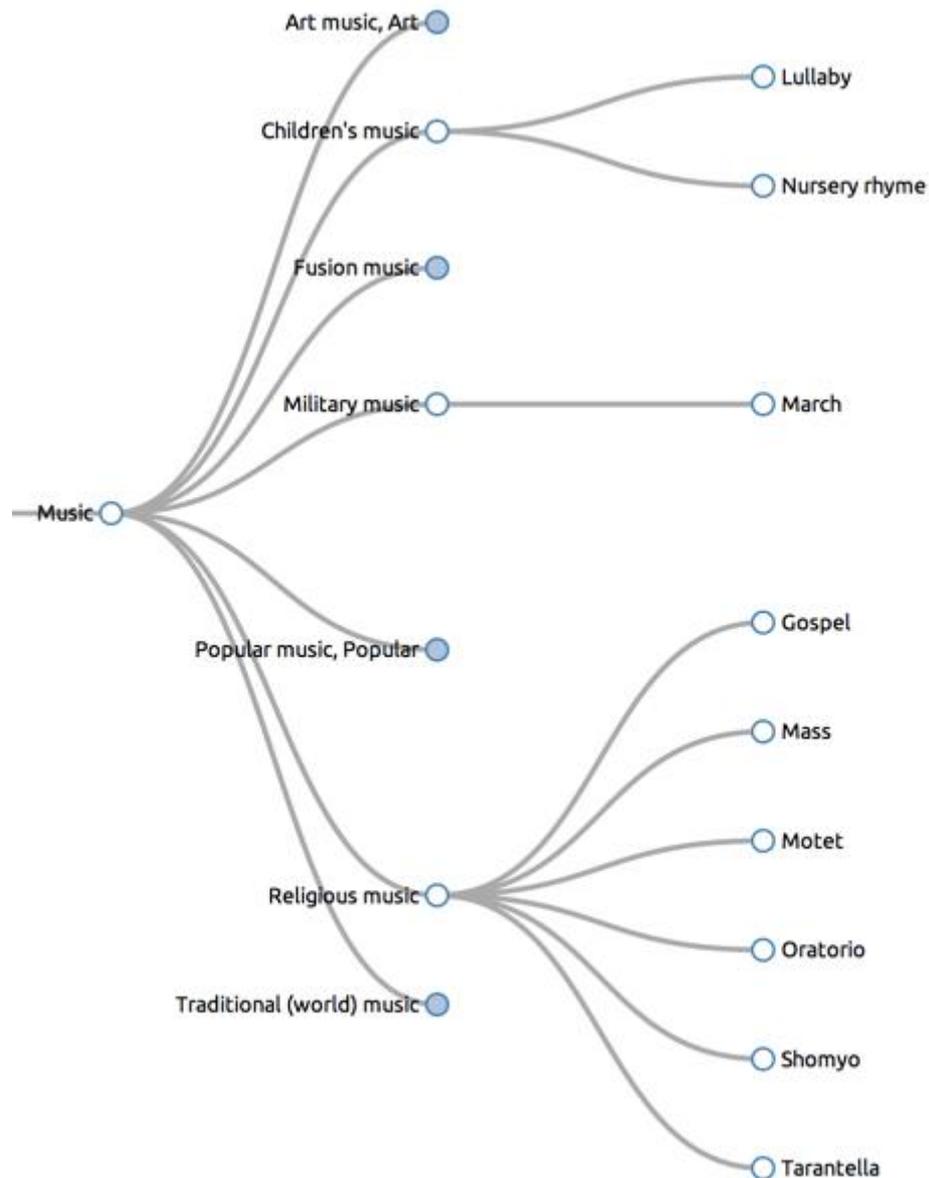


*The
Fusion
Music
hierarchy
in the
Music
Genre
scheme*





The Religious, Military and Children's music hierarchies in the Music Genre scheme



2.1.8.2 Generic sounds scheme

The Generic Sounds Scheme is a simple SKOS vocabulary created by taking into account the top level categories of content specified in T1.1 Content selection policy. Its main purpose is to categorise sounds that are not classified as music. Each concept was linked to Freebase and DBpedia in order to make multilingual labels (and possibly concept definitions) available.

The top level categories specified in deliverable *D1.1 Content Selection Policy* could be enriched with more specific classifications. However, there is no single way of deriving such sub-concept hierarchies, as this largely depends on the context or aspect one is interested in. For example, “Natural sounds” could be grouped by provenance, or by involved natural elements (e.g. water, fire, etc.), and in many other ways. For this reason the present vocabulary does not contain further classifications of the top level concepts and instead we encourage the use of external vocabularies such as LCGFT, and genre concepts from LCSH, FAST, Freebase and DBpedia.

Two sub-vocabularies have also been developed as part of this task, which could be integrated into the Generic Sounds Scheme:

- **Radio Vocabulary**, which uses the LCGFT schema to model specific kinds of radio programs, and Freebase and DBpedia for other top level categories.
- **Soundscapes Vocabulary**, where soundscapes recordings are characterised by the kind of environment recorded.

Note: Given the heterogeneity of the domain, there might be several cases where a single sound recording falls into more than one category. For example a radio recording could reproduce the sounds of a natural environment or of an animal. While there is no constraint preventing sounds being categorised under multiple branches of the tree at the same time, this decision will have to take into account end-users’ needs in terms of ease of browsing and search.

Current status and links

The RDF/SKOS files and the documentation of the vocabulary were saved on Google Drive⁴².

Files:

Non-Music-Top-Concepts.rdf - RDF/SKOS representation of the Generic Sounds Scheme.

Radio-Schema.rdf - RDF/SKOS representation of the Radio programs hierarchy.

Soundscapes.rdf - A simple vocabulary to categorise soundscapes recordings according to the kind of environment recorded (derived from Wikipedia).

Non-Music-Complete.rdf - An integrated vocabulary including all the concepts from the previous three vocabularies.

The complete **documentation** of the vocabulary (created with *SKOSPlay*⁴³) was saved on Google Drive⁴⁴.

⁴² <https://drive.google.com/#folders/OBzFrvHHW14bZUWowZmJFbm9pZUK>

⁴³ <http://labs.sparna.fr/skos-play>

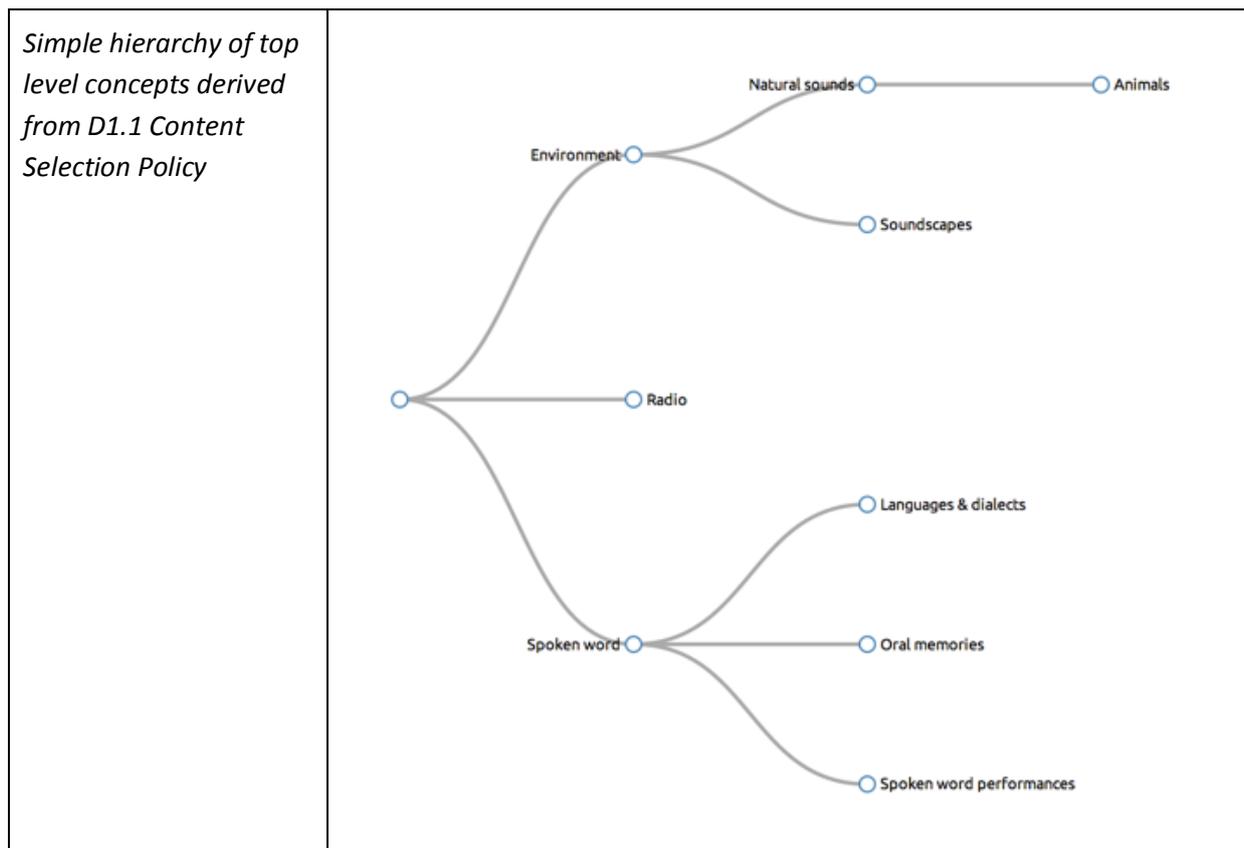
⁴⁴ <https://drive.google.com/file/d/OBzFrvHHW14bZY04zN0NmeHhkNVE/edit?usp=sharing>

A note about incorporating the geographical ‘aspects’ of a sound

Existing vocabularies and end-user oriented classifications, such as British Library Sounds⁴⁵, often use the geographical provenance of an audio resource as a classification means. The DISMARC genres hierarchy contains some geographic categories under the “International” category that clearly refer to the “provenance” or the location of the recorded sound. Under each geographic category (e.g. Europe, South America), the vocabulary simply repeats more general categories, already modelled at upper level (e.g. Nature, Animals, Music). Rather than integrating place names into a genre hierarchy we recommend the use of place name vocabularies such as Geonames, LCSH and FAST in conjunction with appropriate EDM properties, to model the geographic aspects of a sound, such as the place where the sound was recorded, the place where an animal lives, the nationality of a voice. Vocabularies for places are outlined in section 2.4 in this document.

Top level concepts

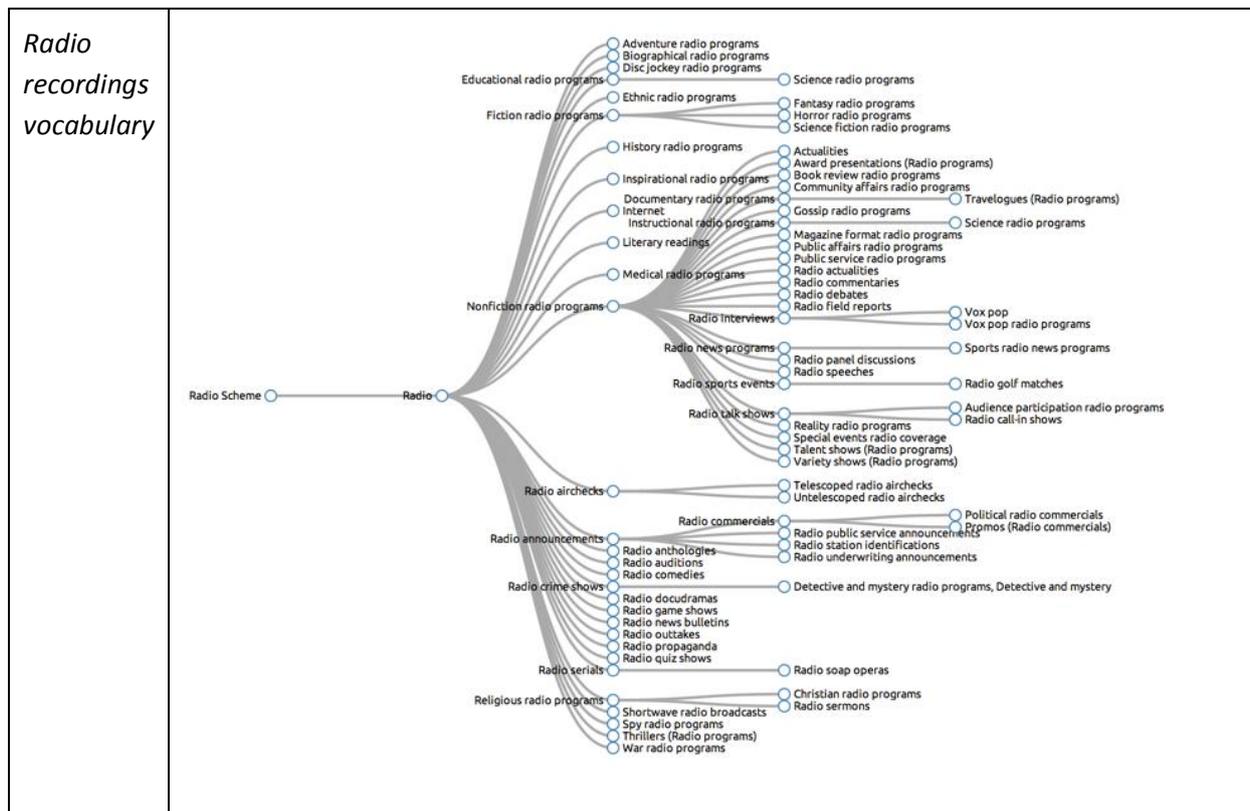
The section that follows lists the top level categories and provides illustrations of the hierarchy of sub-concepts. These concepts have been derived from deliverable *D1.1 Content selection policy*, as shown below.



⁴⁵ <http://sounds.bl.uk/>

Radio vocabulary

This new taxonomy for describing radio programs recordings was created by extracting relevant concepts from the Library of Congress Genre/Form vocabulary⁴⁶ and linking these to related concepts in DBpedia and Freebase, where multilingual labels and descriptions can be found.



Soundscapes vocabulary

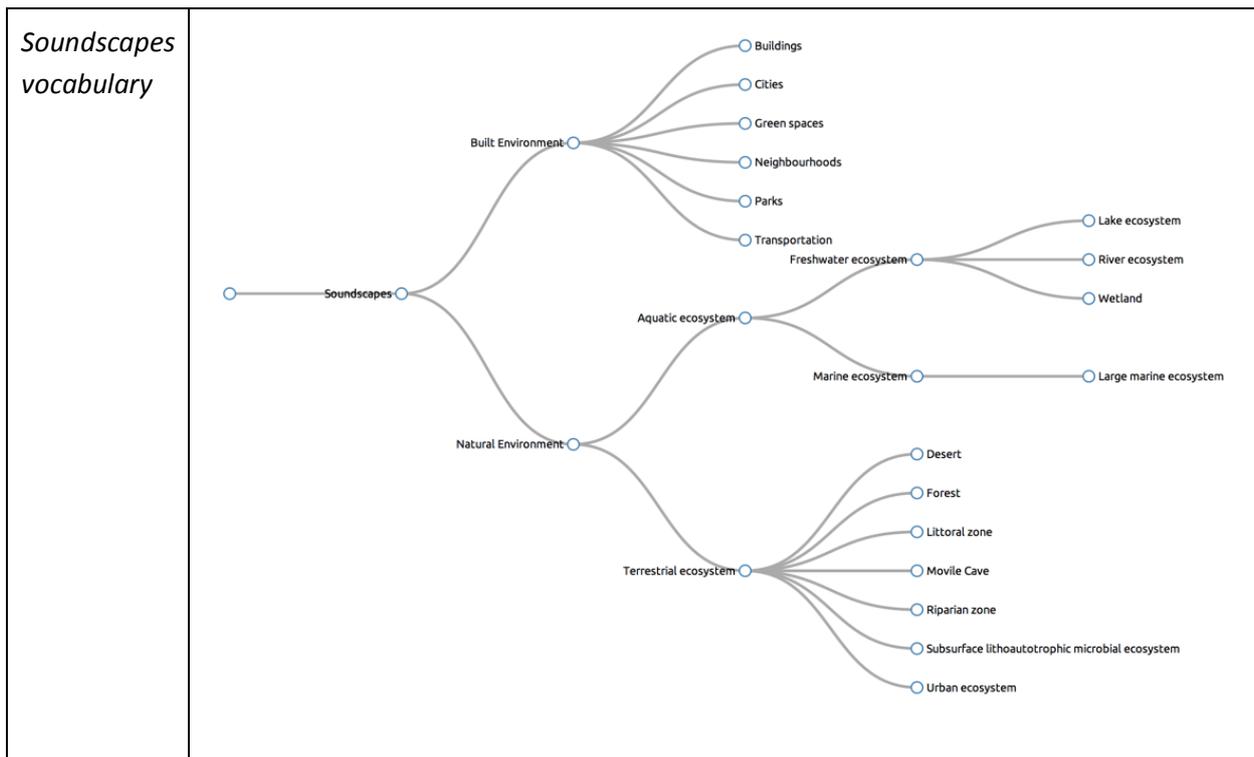
According to Wikipedia, a soundscape is a “sound or combination of sounds that forms or arises from an immersive environment”⁴⁷. The proposed top level categories in this new vocabulary distinguish between a natural environment and a built environment. The sub-categories shown here, which may require further expansion, were derived by manually selecting relevant Wikipedia pages *Built environment*⁴⁸ and *Ecosystem types*⁴⁹.

⁴⁶ <http://id.loc.gov/authorities/genreForms.html>

⁴⁷ <https://en.wikipedia.org/wiki/Soundscape> [REF 39]

⁴⁸ https://en.wikipedia.org/wiki/Built_environment [REF 40]

⁴⁹ <https://en.wikipedia.org/wiki/Ecosystem#types> [REF 41]



2.1.9 Usage recommendations: genres

The Europeana Sounds DoW specifies that it is mandatory to include at least one genre term in each metadata record⁵⁰. The use of consistent, multilingual genres in descriptive metadata is particularly important as it enables users to refine their searches quickly and enables them to find similar material across the Europeana portal.

Broad genre/form concept

In each metadata record, it is **mandatory** to record one or more Europeana Sounds broad genre/form concepts. Data providers will add one or more of the following broad concepts when mapping their data to EDM during the ingestion process:

- Music
- Spoken word
- Radio
- Environment

⁵⁰ C-GA including Annexe I (“Description of Work”), Part A, p.7, under T1.2
<http://pro.europeana.eu/documents/2011409/8d0e9833-4608-494e-af77-681e68f8a8c8> [REF 1]

Specific genre/form concept

In each metadata record, it is recommended that data providers include specific genre/form concepts from any suitable controlled vocabulary. Data providers may already have these in their metadata before mapping to EDM or they may add genres during mapping. Suggested sources are:

- *Europeana Music Genre/Form Vocabulary*
- *Europeana Non-Music Genre/Form Vocabulary*
 - This includes the *Radio Recordings Vocabulary* and the *Soundscapes Vocabulary*
- Another established thesaurus, e.g. LC Genre/Form⁵¹, FAST form concepts⁵², EBU⁵³, or appropriate genre concepts established as topics in LCSH⁵⁴ and FAST. When choosing, consider that:
 - FAST and LCSH are aligned to a large extent, they are available as linked data and are searchable online.
 - FAST and LCSH both have a neutral structure designed to target a wide range of users. Domain-specific vocabularies like EBU on the other hand provide targeted user-oriented taxonomies, e.g. for recorded/broadcast programs consumers.

2.2 Mood vocabularies

Although mood concepts were not present in data providers' use cases, motivated by the presence of concepts for mood in systems such as Spotify, the task group investigated the possibility of defining concepts to describe the mood of an audio object, especially music. Various sources of mood concepts were analysed by the task group. Although it would be possible to derive a simple SKOS taxonomy from Parrott's Classification of Emotions⁵⁵ and link concepts to DBpedia and Freebase to obtain multilingual labels, the task group decided that assigning mood concepts would be a task more appropriate to user annotations interactions, if deemed desirable by task T2.1 *Crowdsourcing* in WP2. A field to accommodate this would need to be added to the EDM Profile for Sound at a later date.

⁵¹ <http://id.loc.gov/authorities/genreForms.html> [REF 38]

⁵² <http://fast.oclc.org/searchfast/> [REF 42]

⁵³ http://www.ebu.ch/metadata/cs/web/ebu_ContentGenreCS_Mapping_p.xml.html [REF 43]

⁵⁴ <http://id.loc.gov/authorities/subjects.html> [REF 21]

⁵⁵ Parrott, W. (2001), *Emotions in Social Psychology*, Psychology Press, Philadelphia. [REF 44]

Parrott's Classification of Emotions

Primary emotion	Secondary emotion	Tertiary emotions
Love	Affection	Adoration, Affection, Love, Fondness, Liking, Attraction, Caring, Tenderness, Compassion, Sentimentality
	Lust	Arousal, Desire, Lust, Passion, Infatuation
	Longing	Longing
Joy	Cheerfulness	Amusement, Bliss, Cheerfulness, Gaiety, Glee, Jolliness, Joviality, Joy, Delight, Enjoyment, Gladness, Happiness, Jubilation, Elation, Satisfaction, Ecstasy, Euphoria
	Zest	Enthusiasm, Zeal, Zest, Excitement, Thrill, Exhilaration
	Contentment	Contentment, Pleasure
	Pride	Pride, Triumph
	Optimism	Eagerness, Hope, Optimism
	Enthrallment	Enthrallment, Rapture
	Relief	Relief
Surprise	Surprise	Amazement, Surprise, Astonishment
Anger	Irritation	Aggravation, Irritation, Agitation, Annoyance, Grouchiness, Grumpiness
	Exasperation	Exasperation, Frustration
	Rage	Anger, Rage, Outrage, Fury, Wrath, Hostility, Ferocity, Bitterness, Hate, Loathing, Scorn, Spite, Vengefulness, Dislike, Resentment
	Disgust	Disgust, Revulsion, Contempt
	Envy	Envy, Jealousy
	Torment	Torment
Sadness	Suffering	Agony, Suffering, Hurt, Anguish
	Sadness	Depression, Despair, Hopelessness, Gloom, Glumness, Sadness, Unhappiness, Grief, Sorrow, Woe, Misery, Melancholy
	Disappointment	Dismay, Disappointment, Displeasure
	Shame	Guilt, Shame, Regret, Remorse
	Neglect	Alienation, Isolation, Neglect, Loneliness, Rejection, Homesickness, Defeat, Dejection, Insecurity, Embarrassment, Humiliation, Insult
	Sympathy	Pity, Sympathy
Fear	Horror	Alarm, Shock, Fear, Fright, Horror, Terror, Panic, Hysteria, Mortification
	Nervousness	Anxiety, Nervousness, Tenseness, Uneasiness, Apprehension, Worry, Distress, Dread

2.3 Subject vocabularies

The subjects of sounds cover a broad range of concepts. For example the subject of a song or of a speech can be about anything: a concept, a person, a mood, a place, another song or speech, etc. Therefore, creating a new controlled vocabulary for Europeana Sounds would be impractical. Instead, we provide examples of linked data sources that are appropriate to be used for enrichment purposes. Linking to well-known and/or authoritative resources across datasets is key to enabling features such as cross-collection search. Freebase and DBpedia provide multilingual labels which could be used to enable multilingual search.

2.3.1 LCSH

Availability	Search & download: http://id.loc.gov/authorities/subjects.html
Terms of use	
Languages	English
Format	Various formats available

The Library of Congress Subject Headings (LCSH) provides a large vocabulary of topic concepts in English. At the time of writing LCSH also includes many genre concepts which have not yet been converted to LCGFT (see section 2.1.2).

2.3.2 FAST

Availability	Search: http://fast.oclc.org/searchfast/ Download: http://www.oclc.org/research/activities/fast/download.html?urlm=159755
Terms of use	
Languages	English
Format	SKOS

FAST (Faceted Application of Subject Terminology) is an enumerative, faceted subject heading schema derived from the Library of Congress Subject Headings⁵⁶ (LCSH). The FAST linked data authorities are available for download in SKOS format.

⁵⁶ <http://www.oclc.org/research/activities/fast.html?urlm=159754>

2.3.3 DBpedia

Availability	Search API: https://github.com/dbpedia/lookup SPARQL: http://dbpedia.org/sparql
Terms of use	
Languages	Multilanguage. Languages are not uniform across resources.
Format	RDF: http://dbpedia.org/describe/?url=http%3A%2F%2Fdbpedia.org%2Fresource%2FCategory%3ASmoking&sid=248109)

For general information about DBpedia, see the preceding section on Genre. DBpedia subjects are multilingual and have the attribute type “Concept”. Place names are included as concepts, for example “London”⁵⁷.

2.3.4 RAMEU

Availability	Search: Select “Autorités matière” in the first drop-down box “Choisir un critère” at http://catalogue.bnf.fr/jsp/recherche_autorites_rameau.jsp?nouvelleRecherche=O&host=catalogue
Terms of use	
Languages	French
Format	SKOS

Rameau⁵⁸ comprises the authority files produced by the Bibliothèque nationale de France. It has been put into SKOS in the context of the data.bnf.fr⁵⁹ project and it is now available as linked open data. It contains an encyclopaedic subject vocabulary, not specifically for music. The concepts are in

⁵⁷ <http://dbpedia.org/describe/?url=http%3A%2F%2Fdbpedia.org%2Fresource%2FCategory%3ALondon&sid=248109&urilookup=1>

⁵⁸ <http://rameau.bnf.fr/>

⁵⁹ <http://data.bnf.fr>

French. RAMEAU is part of the MACS project along with LCSH and SWD which can serve as a basis for multilingual search⁶⁰.

2.3.5 Places as subjects

Please see section 2.4 for vocabularies for place names. Place name vocabularies will need to be used when a resource is about a place, for example:

Title: I love Paris in springtime

Subject: Paris (France)

2.3.6 Personal and corporate names as subjects

Please see section 2.8 for vocabularies for personal and corporate names. Name vocabularies will need to be used when a resource is about a person or corporate body, for example:

Title: Ballad of John and Yoko

Subject: Lennon, John, 1940-1980

Subject: Ono, Yōko

2.3.7 Works as subjects

Please see section 2.10 for vocabularies for works. These will need to be used when a resource is about a work, for example

Title: And the band played Waltzing Matilda.

Subject: Waltzing Matilda (Song)

2.3.8 Usage recommendations: subjects

Many objects, for example a sound recording of a string quartet, will have no subject. For Europeana Sounds we recommend that subjects are added to descriptive metadata when this is appropriate to describe the object. Use concepts from an established subject thesaurus such as LCSH, FAST, RAMEAU, Schlagwortnormdatei (SWD), or from another type of vocabulary (places, names, works) when a resource is about these concepts.

2.4 Vocabularies for places

Place names model geographic aspects of a sound, such as the place where a sound was recorded, the place where an animal lives, or the nationality of a voice.

⁶⁰ <http://www.cs.vu.nl/STITCH/rameau/>

2.4.1 GeoNames

Availability	Search: http://www.geonames.org/
Terms of use	Creative Commons Attribution 3.0 License
Languages	Multilingual
Format	SKOS: http://www.geonames.org/2524810/monte-etna.html

Geographic names in GeoNames include coordinates and are multilingual, enabling spatial search and cross-lingual retrieval in keyword searches. The names are in a simple form, in the language of the country of the place: Lithuania; Pisa; Montparnasse; Monte Etna.

2.4.2 Getty Thesaurus of Geographic Names Online

Availability	Search: http://www.getty.edu/research/tools/vocabularies/tgn/
Terms of use	All rights reserved
Languages	Multilingual
Format	SKOS: http://www.getty.edu/vow/TGNFullDisplay?find=london&place=&nation=&english=Y&subjectid=7011781)

Geographic names in the Getty Thesaurus include coordinates and are multilingual, enabling spatial search and cross-lingual retrieval in keyword searches. They are in the following style: Lithuania (nation), Pisa (inhabited place), Montparnasse (neighborhood), Mount Etna (volcano).

2.4.3 Place names in LC

Availability	Search: http://id.loc.gov/authorities/names.html Search: http://id.loc.gov/authorities/subjects.html Search: http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First
---------------------	-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Terms of use	
Languages	English
Format	SKOS: http://id.loc.gov/authorities/names/n79034879)

Place names in LC are established according to RDA/AACR2, in the following style: Lithuania ; Pisa (Italy) ; Montparnasse (Paris, France) ; Etna (Mount, Italy). Some place names are algorithmically matched with GeoNames, and include coordinates.

2.4.4 Geographic names in FAST

Availability	Search: http://fast.oclc.org/searchfast/ Download: http://www.oclc.org/research/activities/fast/download.html?urlm=159755
Terms of use	Open Data Commons Attribution License (ODC-By) v1.0
Languages	English
Format	SKOS: http://id.worldcat.org/fast/1204823

Geographic names in FAST have links to LC and VIAF forms of the same concept. They are formulated in a more straight-forward style than LC and VIAF: Lithuania ; Italy--Pisa ; France--Paris--Montparnasse ; Italy--Mount Etna.

2.4.5 Geographic names in VIAF

Availability	Search: http://viaf.org/
Terms of use	Available under the Open Data Commons Attribution License (ODC-BY)
Languages	Multilingual
Format	SKOS: http://viaf.org/viaf/204007255

Geographic names in VIAF follow the same RDA/AACR2 format as LC names: Lithuania ; Pisa (Italy) ; Montparnasse (Paris, France) ; Etna (Mount, Italy).

2.4.6 Usage recommendations: places

We recommend that place names are included in metadata when this is appropriate. Use concepts from an established vocabulary of place names, such as Geonames, Getty, LC, FAST or VIAF. Prefer vocabularies that include coordinates, such as GeoNames, some LC.

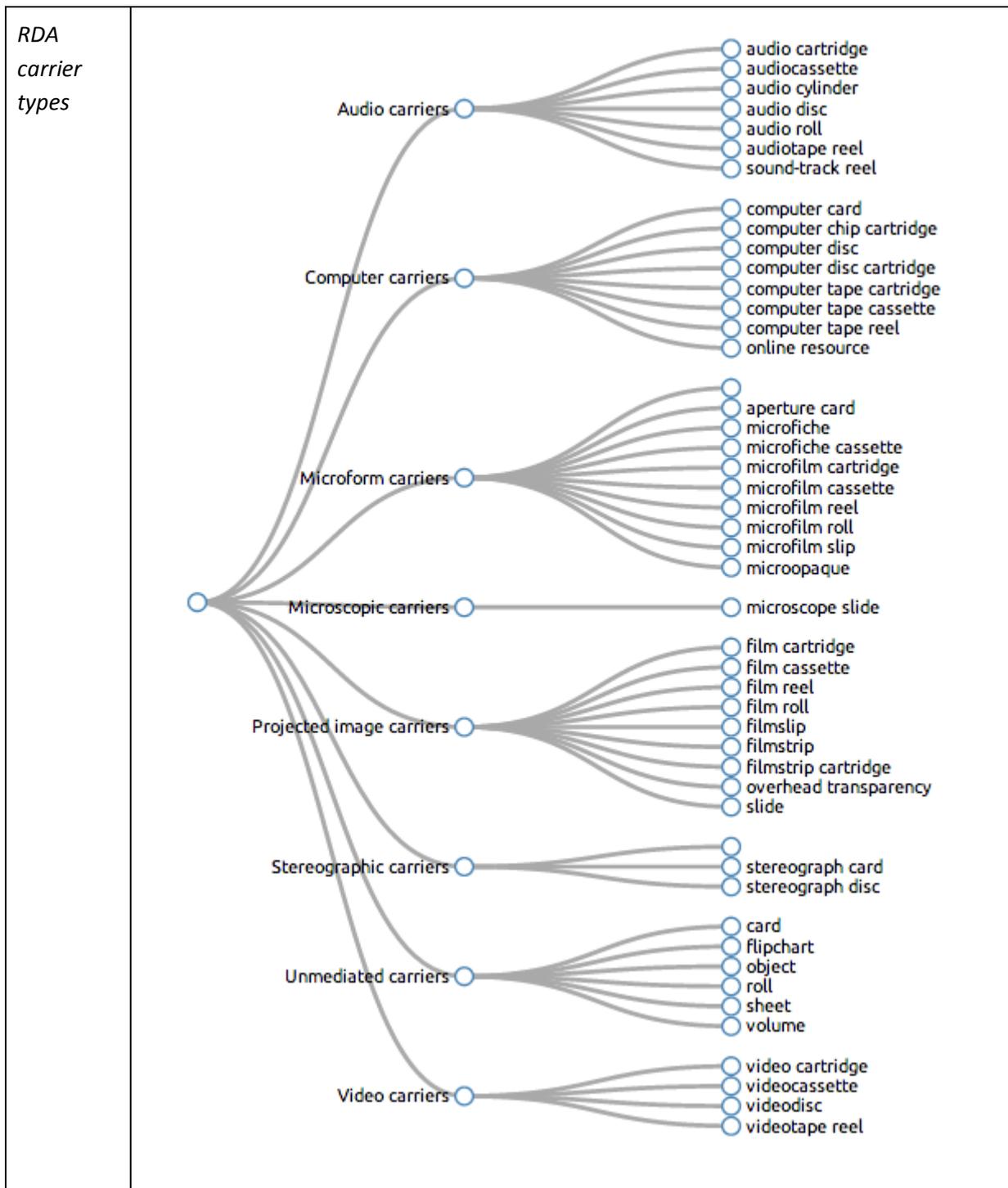
2.5 Vocabularies for physical audio carriers

2.5.1 RDA carrier types

Availability	Search: http://www.loc.gov/standards/valuelist/rdacarrier.html Download: http://metadataregistry.org/conceptprop/search?concept_term=carrier&commit=Search+Vocabularies
Terms of use	
Languages	English
Format	SKOS

This vocabulary, maintained by the Metadata Management Associates is represented and published as SKOS and hosted at the Open Metadata Registry⁶¹. The taxonomy defined in the vocabulary includes all the main physical carriers arranged into macro-categories. Among the macro-categories, **Audio Carriers** is the most relevant to Europeana Sounds. However, as media content in the project also includes text and videos, the whole hierarchy has been considered.

⁶¹ <http://metadataregistry.org/agent/show/id/67.html>



2.5.2 LC Carriers Scheme

Availability	Search & Download: http://id.loc.gov/vocabulary/carriers.html
Terms of use	
Languages	English
Format	SKOS

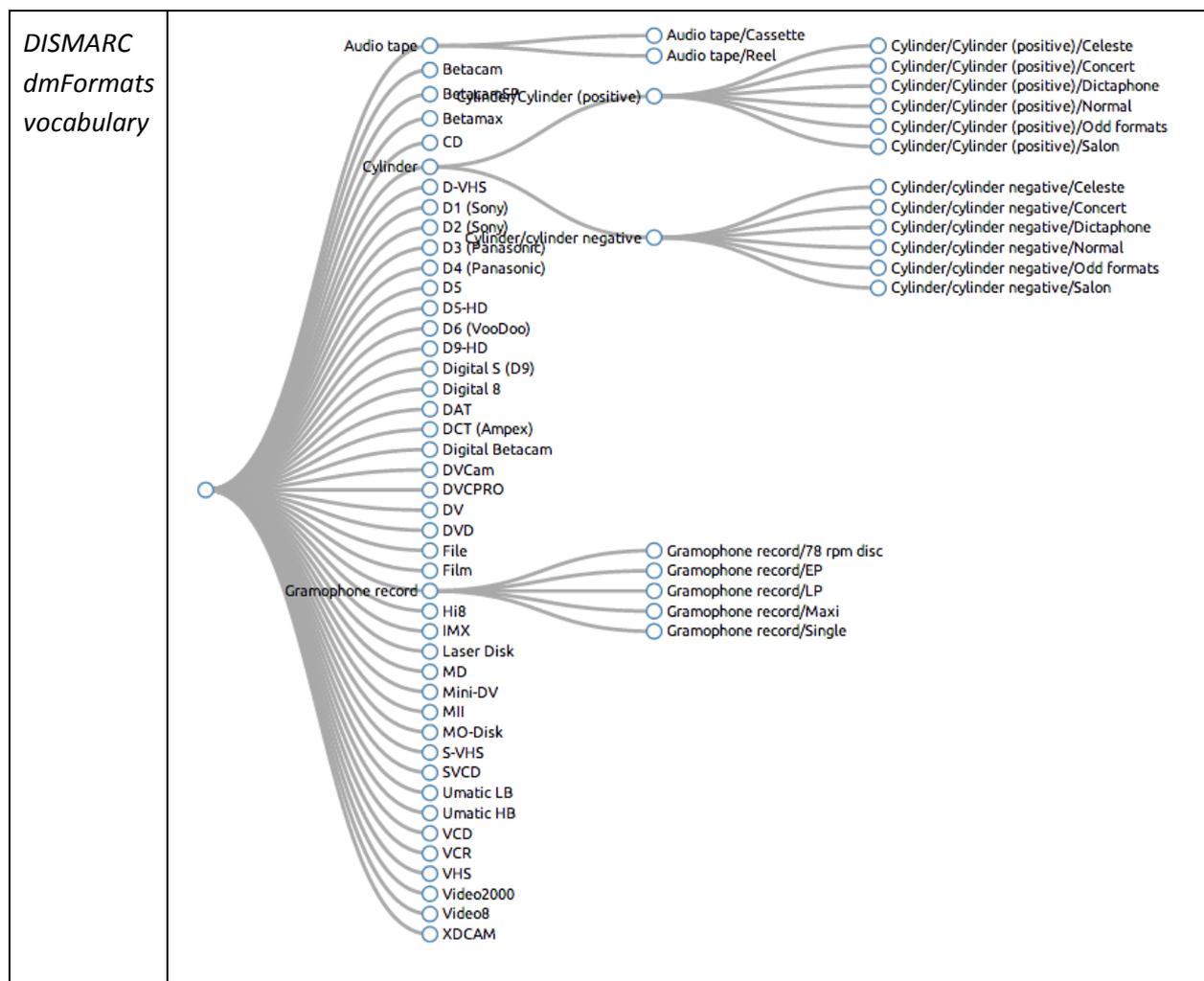
The Library of Congress Carriers Scheme is derived from a controlled list of coded values representing carrier types principally used in RDA cataloguing. It is published in SKOS format, and, if compared to the RDA Carrier Types vocabulary previously described, there is no significant difference: all the items are present in both vocabularies with the very same prefLabel. However, the Carrier Scheme is modelled as a flat list and lacks the general categorizations found in the RDA Carrier Types vocabulary.

2.5.3 DISMARC (dmFormats)

Availability	Search: Select Formats tab at http://www.dismarc.org/index.php?form=browse&db=0#tabPaneBrowsecriteria1406892805.993
Terms of use	
Languages	English
Format	CSV

The vocabulary currently used in DISMARC⁶² to classify resources accordingly to their formats has been built in a bottom-up fashion by content providers of the DISMARC platform and includes a number of specific carriers and recording formats. Compared with the RDA Carrier Types vocabulary, it is more specific and includes instances of top level categories that are clearly defined in the RDA Carrier Types vocabulary, but DISMARC lacks a precise hierarchy.

⁶² <http://www.dismarc.org/>

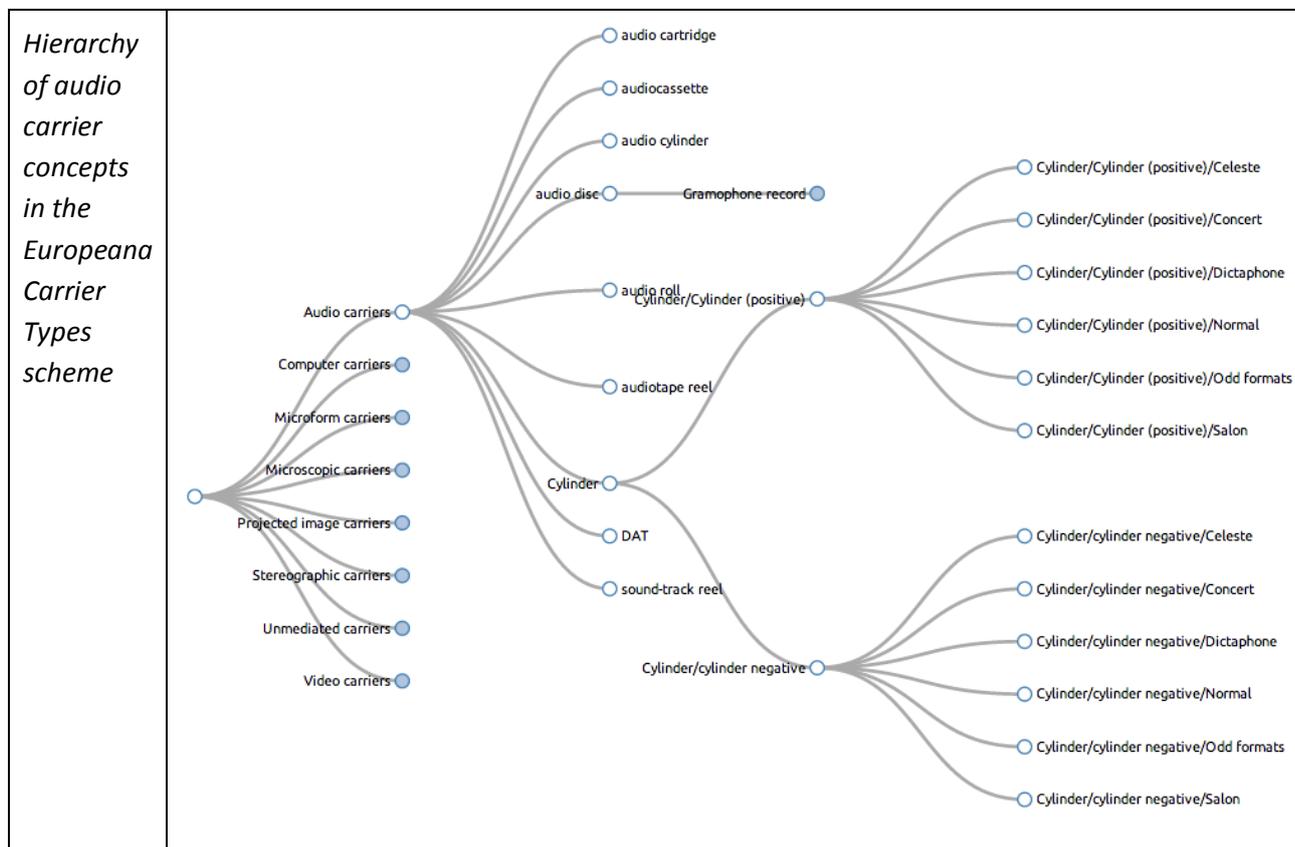


The vocabulary used in DISMARC contains a number of very specific items (e.g. Umatic LD, Betacam, etc.) that are not modelled by the RDA carrier types vocabulary. The task group considers that this level of detail could be useful for data providers as well as users of the Europeana portal.

The DISMARC vocabularies come in the form of CSV files, where each item is identified by a unique identifier. A first version in SKOS has been produced as an intermediate result by using Open Refine RDF mapping. The SPARQL-based reconciliation service has been used for linking DISMARC to DBpedia; this is useful in order to retrieve multilingual labels at a later stage, when these are present. While there was a small number of items automatically matched (this is due to the possible variants of format names), after a manual refinement it has been possible to link 60% of the items to matching entities in DBpedia.

2.5.4 New Europeana Carrier Types scheme

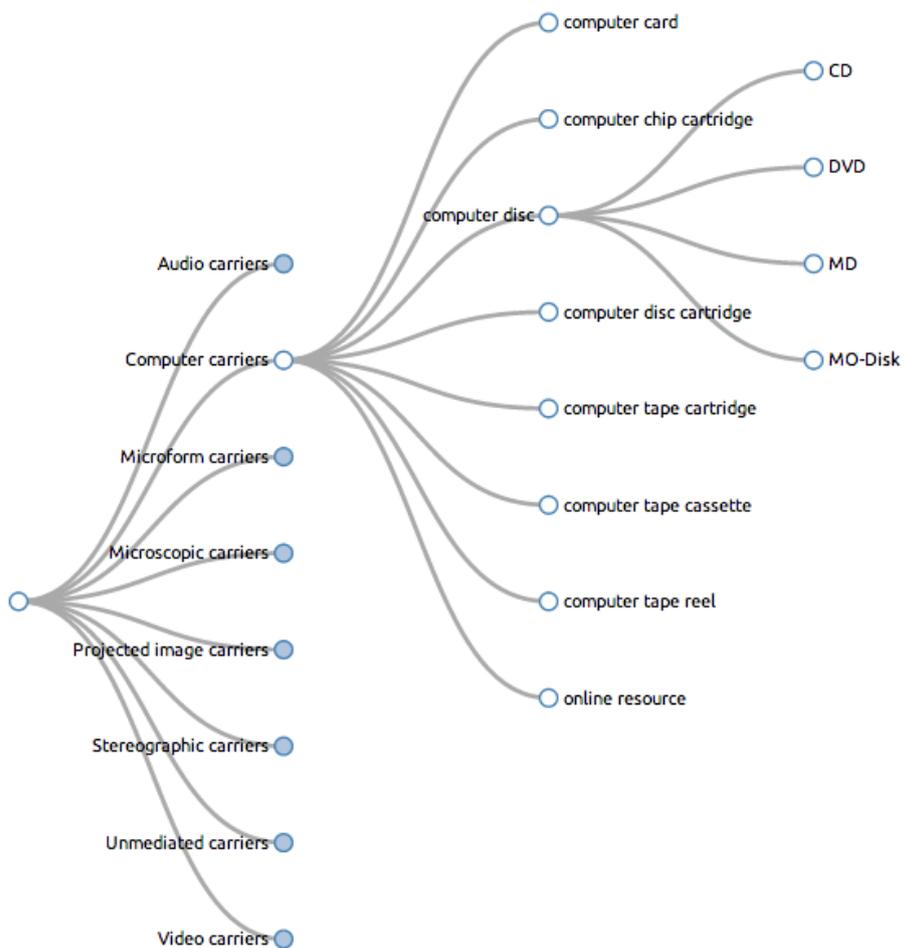
A new SKOS vocabulary has been produced by extending the RDA Carrier vocabulary with more specific carrier types extracted from the DISMARC dm:Formats vocabulary. The integrated SKOS vocabulary including the complete RDA Carrier Types scheme and the DISMARC extracted concepts is available as an RDF file⁶³ and all concepts have been documented in a pdf⁶⁴.

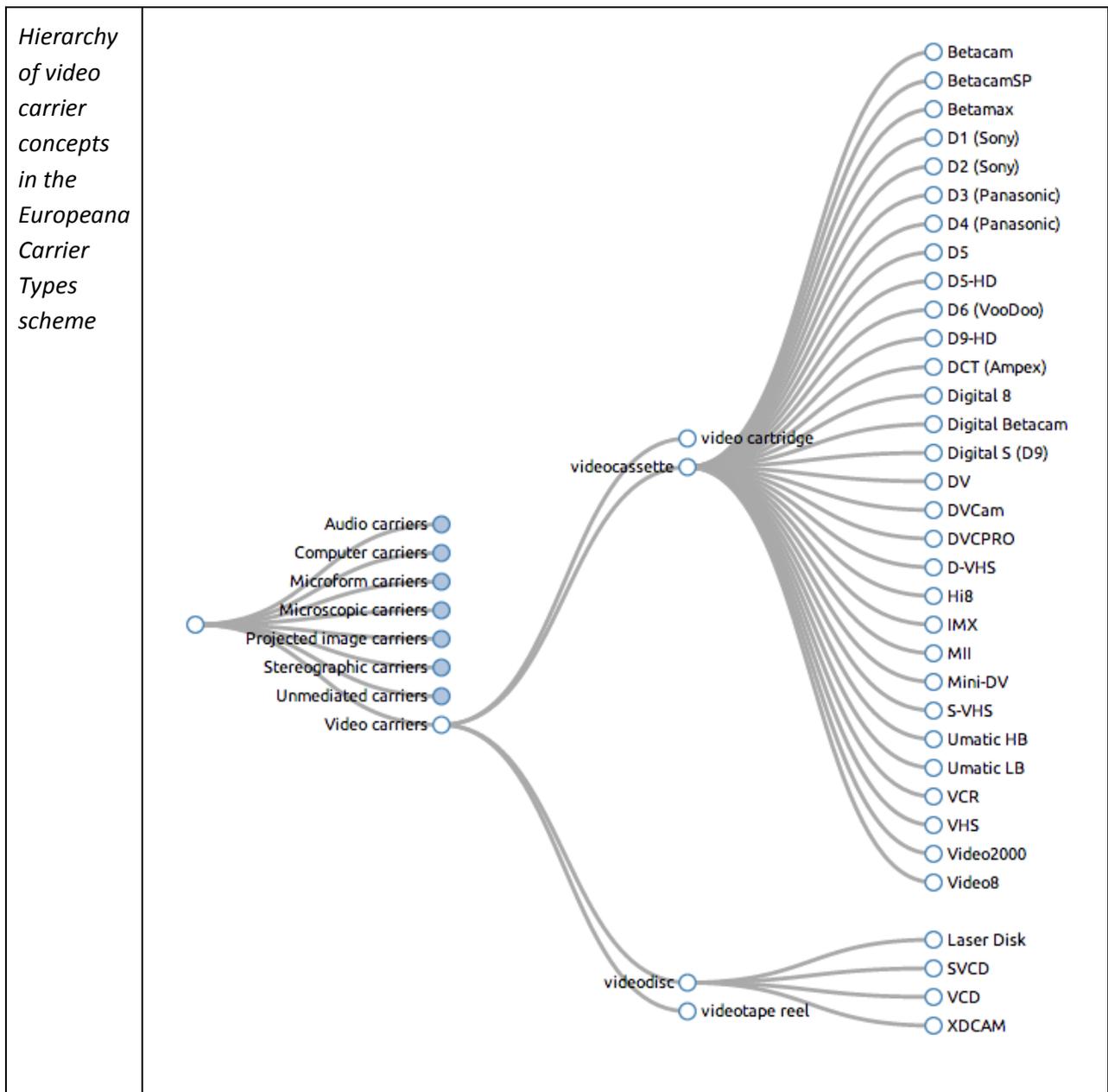


⁶³ <https://drive.google.com/file/d/0BzFrvHHW14bZeDNKOXIkNk1uemM/>

⁶⁴ <https://drive.google.com/file/d/0BzFrvHHW14bZM09aOWY4eU5LVIU/>

*Hierarchy
of
computer
carrier
concepts in
the
Europeana
Carrier
Types
scheme*





2.5.5 Usage recommendations: physical carriers

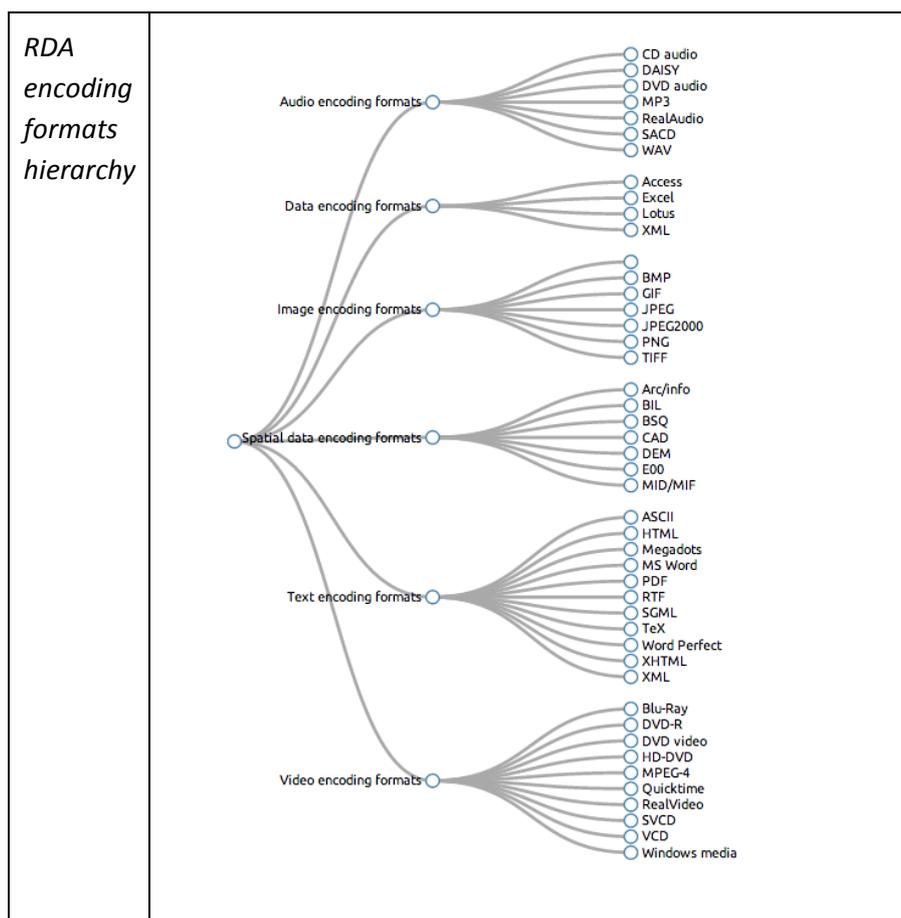
Use concepts from an established schema such as RDA Carrier types, LC Carriers Scheme or the new Europeana Carrier Types scheme to specify the physical carrier of the object described.

2.6 Vocabularies for digital formats

2.6.1 RDA encoding formats

Availability	Search: http://metadataregistry.org/concept/list/vocabulary_id/87.html Download: http://metadataregistry.org/vocabulary/show/id/87.html
Terms of use	
Languages	English
Format	SKOS

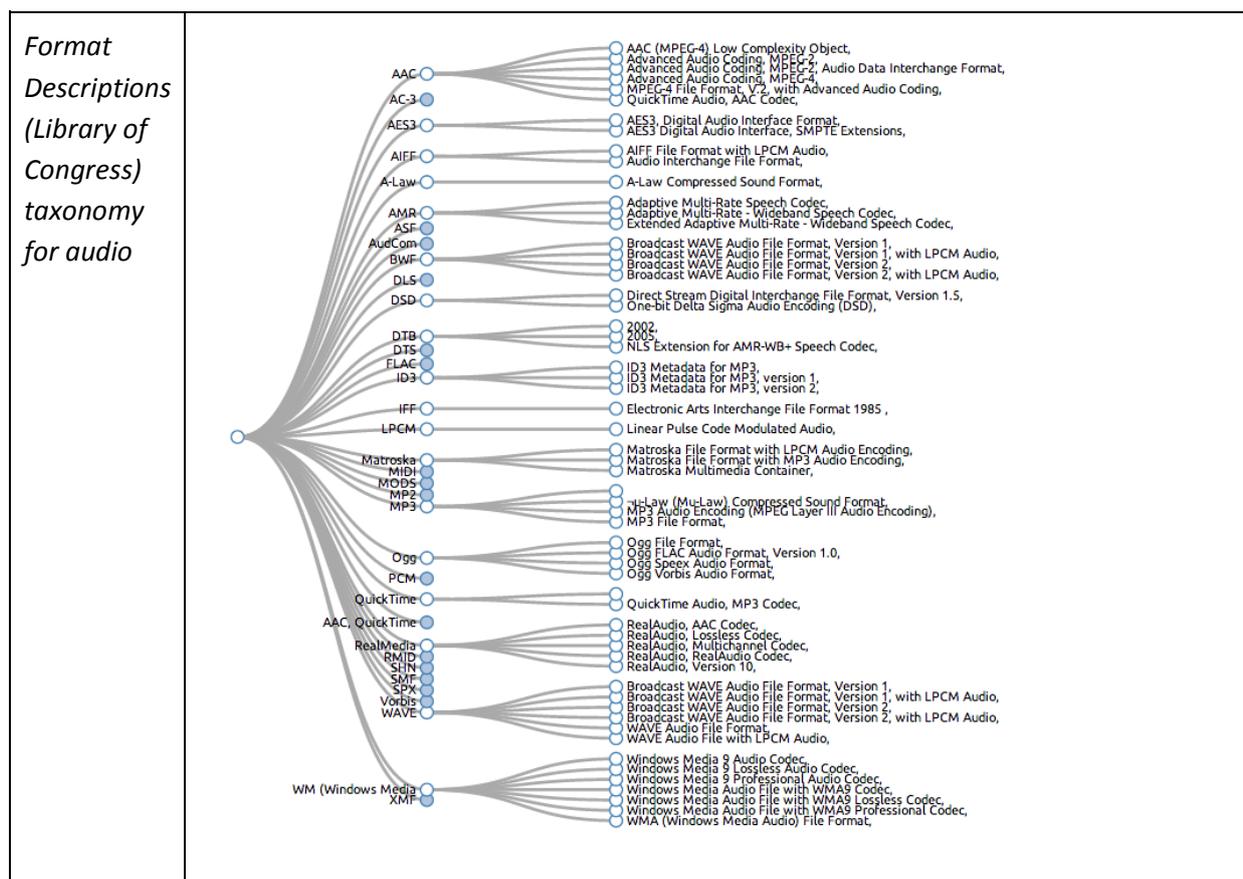
The RDA encoding formats vocabulary lists the main encoding formats organised in a simple hierarchy.



2.6.2 Format descriptions

Availability	Search: http://www.digitalpreservation.gov/formats/fdd/descriptions.shtml Download: http://www.digitalpreservation.gov/formats/fdd/fdd_xml_info.shtml
Terms of use	
Languages	English
Format	SKOS

The Format descriptions vocabulary published by the Library of Congress goes into more depth if compared to the RDA encoding formats vocabulary. The format descriptions are available as XML. Although the task group created a draft SKOS version of the vocabulary by taking into account only those formats relating to audio, the vocabulary includes all kinds of digital files.



2.6.3 Usage recommendations: digital formats

Use concepts from an established thesaurus of digital formats such as *RDA encoding formats* or *LC format descriptions*, to specify the format/encoding of a digital audio or media file.

2.7 Vocabularies for medium of performance

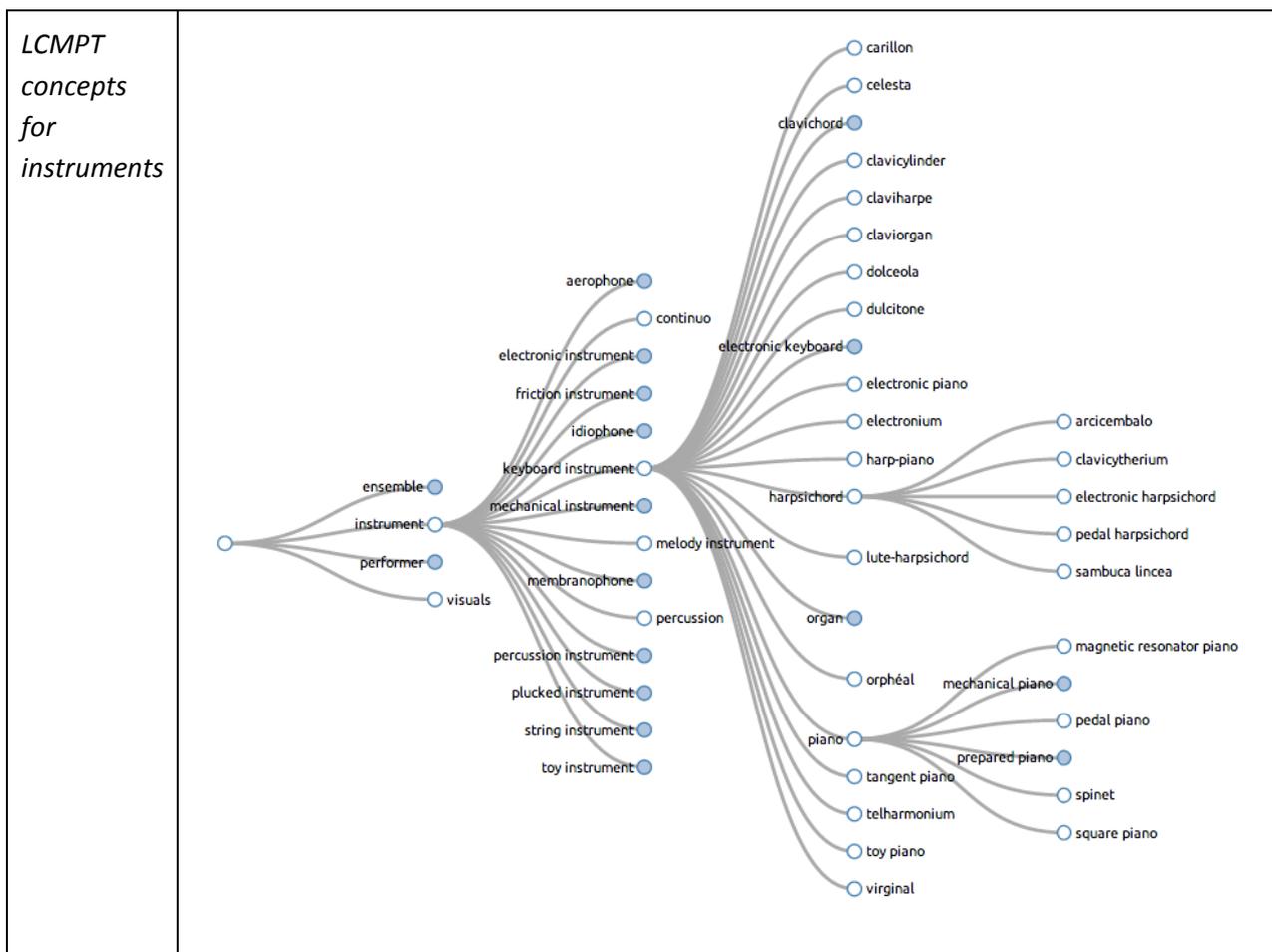
Recording the medium of performance of a musical work enables users to retrieve music by a particular instrument or instrumental group, voice or vocal group, or other medium of performance the searcher may specify.

2.7.1 LCMPT

Availability	Search & download: http://id.loc.gov/authorities/performanceMediums.html
Terms of use	
Languages	English
Format	SKOS

The Library of Congress Medium of Performance Thesaurus⁶⁵ (LCMPT) provides English language terminology to describe the instruments, voices, etc., used in the performance of musical works. The core concepts in LCMPT are based chiefly on existing LC subject headings, but some additional concepts that do not already appear in LCSH have also been included. The thesaurus has as its broadest concepts: **ensemble**, **instrument** and **performer**. Each of the other concepts is hierarchically subordinate to one or more of these concepts as demonstrated below. The **performer** category is less developed and includes general concepts only.

⁶⁵ <http://id.loc.gov/authorities/performanceMediums>



2.7.2 Freebase instruments

Availability	Search & download: Using Freebase APIs - https://developers.google.com/freebase
Terms of use	
Languages	Multilingual
Format	JSON

Instrument names are included in the Freebase music commons and can be extracted using the Freebase APIs.

2.7.3 MIMO-VOC

Availability	Search: http://www.mimo-international.com/MIMO/instrument-families.aspx Download: http://www.mimo-international.com/MIMO/doc/keywords.xml Download: http://www.mimo-international.com/MIMO/doc/thesaurus_0.csv
Terms of use	
Languages	English, French, German, Italian, Dutch, Swedish
Format	SKOS

A multilingual (English, French, German, Italian, Dutch, Swedish) SKOS vocabulary for musical instruments, MIMO-VOC⁶⁶, has already been developed for Europeana as part of the MIMO Project⁶⁷ (Musical Instrument Museums Online). The vocabulary is based on orchestral terms from a version of the Hornbostel and Sachs classification⁶⁸ revised by Jeremy Montagu in 2009, then further enhanced for MIMO with the addition of vocabulary for electronic instruments.

2.7.4 RAMEAU

Availability	Search: Select “Autorités matière” in the first drop-down box “Choisir un critère” at http://catalogue.bnf.fr/jsp/recherche_autorites_rameau.jsp?nouvelleRecherche=O&host=catalogue
Terms of use	
Languages	French, English,
Format	SKOS: http://catalogue.bnf.fr/ark:/12148/cb119367821/PUBLIC

For a general description of RAMEAU, see section 2.3.4.

Since 2013, musical instrument terms have been added to RAMEAU by the French research program *Labex : Past in the present*, specifically for the project “Sources for Ethnomusicology” carried out by

⁶⁶ <http://labs.europeana.eu/apps/MIMO-VOC/>

⁶⁷ <http://www.mimo-international.com/MIMO/>

⁶⁸ <http://en.wikipedia.org/wiki/Hornbostel%E2%80%93Sachs>

the Bibliothèque nationale de France, the Centre national de la recherche scientifique and Musée du quai Branly. It contains a vocabulary for music instruments in French, linked with Library of Congress Subject Headings (English) and the Hornbostel and Sachs classification.

2.7.5 Usage recommendations: medium of performance

For musical objects, include terms for Medium of Performance from an established thesaurus.

2.8 *Vocabularies for personal and corporate names*

Descriptive metadata for audio and audio-related material will often contain names of composers, performers, conductors, collectors, publishers, etc. The metadata coming from 18 data providers will contain names constructed according to various standards such as Library of Congress Authorities⁶⁹, authority files of individual libraries and archives such as the Deutsche Nationalbibliothek (DNB) and the Bibliothèque nationale de France (BNF) as well as specialist databases such as Répertoire International des Sources Musicales⁷⁰ (RISM). Creating a new controlled vocabulary of names for Europeana Sounds would be impractical. Instead, we will indicate a set of Linked Data sources that could be matched with DBpedia and Freebase for enrichment purposes. Linking to well-known and/or authoritative resources across collections and providers allows users to search and navigate across collections.

2.8.1 VIAF

Availability	Search: http://viaf.org/
Terms of use	Available under the Open Data Commons Attribution License (ODC-BY)
Languages	Multilingual
Format	

The Virtual International Authority File⁷¹ (VIAF) is an international service designed to provide convenient access to the world's major name authority files. Its creators envision the VIAF as a building block for the Semantic Web to enable switching of the displayed form of names for persons to the preferred language and script of the Web user. VIAF began as a joint project with the Library of Congress (LC), the Deutsche Nationalbibliothek (DNB), the Bibliothèque nationale de France (BNF) and OCLC. It has, over the past decade, become a cooperative effort involving an expanding number

⁶⁹ <http://authorities.loc.gov/>

⁷⁰ <http://www.rism.info/en/home.html>

⁷¹ <http://www.oclc.org/viaf/>

of other national libraries and other agencies. At the beginning of 2012, contributors⁷² include 20 agencies from 16 countries.

VIAF⁷³ (Virtual International Authority File) provides multi-lingual instances of names. Each authority contains a VIAF ID (e.g. VIAF ID: 71319254 (Personal)) and a permalink (e.g. Permalink: <http://viaf.org/viaf/71319254>).

2.8.2 ISNI

Availability	Search: http://www.isni.org/search
Terms of use	
Languages	Multilingual
Format	html, xml and RD/xml

ISNI⁷⁴ (International Standard Name Identifier) is the ISO certified global standard number for identifying the millions of contributors to creative works and those active in their distribution, including researchers, inventors, writers, artists, visual creators, performers, producers, publishers, aggregators, etc.

Linked data is part of a strategy to make ISNIs freely available and widely diffused. ISNI metadata will soon be available in RDF triples which will be embedded in the public web pages and will be available via the persistent URI and the SRU search API. Each assigned ISNI is accessible by a persistent URI. e.g. the form <http://isni.org/isni/0000000121224298>

2.8.3 LC Name Authorities

Availability	Search: http://id.loc.gov/authorities/names.html or http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First
Terms of use	
Languages	English
Format	Various formats available http://id.loc.gov/authorities/names.html

LC Names is a large authority file of names established according to RDA/AACR2.

⁷² <http://www.oclc.org/viaf/contributors.en.html>

⁷³ <http://viaf.org/>

⁷⁴ <http://www.isni.org/>

2.8.4 RISM (Répertoire International des Sources Musicales)

Availability	Search: http://www.rism.info/en/community/development/personal-names-search.html#c2519 Download: https://opac.rism.info/index.php?id=8&L=1
Terms of use	Licensed under a Creative Commons Attribution 3.0 Unported License CC-BY
Languages	
Format	MARCXML or RDF

The International Inventory of Musical Sources - Répertoire International des Sources Musicales (RISM) is a multinational, non-profit joint venture which aims for comprehensive documentation of extant musical sources worldwide. These primary sources are manuscripts or printed music, writings on music theory, and libretti, housed in libraries, archives, monasteries, schools and private collections.

2.8.5 Usage recommendations: personal and corporate names

Include names in descriptive metadata when this is appropriate to describe the relationship between a person or corporate body and an object. Provide a name as a text string using an established thesaurus or a reference to a VIAF or ISNI permalink and/or ID.

e.g.

Mozart, Wolfgang Amadeus, 1756-1791.

or

<http://viaf.org/viaf/32197206>

2.9 Vocabularies for roles

Describing the role, function or relationship that an agent has with an object help users to understand the relationship between a name and a resource, thus enhancing the user's interpretation of a resource. This is particularly important for sound recordings where a person could have multiple roles in the same resource, such as conductor and performer, or where a person who is a conductor of some resources is a performer in other resources. The inclusion of roles with names means these can be displayed in search results, enabling users to identify a name undertaking a particular role; for example:

Schiff, Andrés, **performer**

Schiff, Andrés, **conductor**

2.9.1 RDA relationship designators

Availability	http://metadataregistry.org/schema/show/id/4.html
Terms of use	
Languages	English
Format	Various formats available

RDA (Resource Description & Access) relationship designators defining the relationship between a name and a resource are defined at <http://www.rdaregistry.info/Elements/a/>.

2.9.2 MARC code list for relators

Availability	http://www.loc.gov/marc/relators/relacode.html
Terms of use	
Languages	English
Format	MARC

MARC relator code list for "roles" e.g. director, musician, composer, collector, etc. is at <http://www.loc.gov/marc/relators/relacode.html>

2.9.3 DISMARC (dmsvAgentRoles)

Availability	http://www.dismarc.org/index.php?form=admin.thesauri&task=showThesaurus&thesaurus=dmsvAgentRoles
Terms of use	
Languages	English
Format	CSV

The vocabulary used in DISMARC to describe roles comes in the form of a CSV file and contains unique identifiers. Similarly to other DISMARC vocabularies, this vocabulary can be published in RDF/SKOS.

2.9.4 Usage recommendations: roles

Include a term from an established vocabulary to describe the role, function or relationship that an agent has with a resource.

2.10 Vocabularies for works

To facilitate navigation between different objects containing the same work, we are working with T1.3 EDM Profile to investigate how best to link different vocabularies used by data providers to describe works. In this document we focus on vocabularies for musical works as this will be the most common navigation required for Europeana Sounds.

There are many existing controlled vocabularies identifying musical works, including Library of Congress Authorities⁷⁵, authority files of individual libraries and archives such as the Deutsche Nationalbibliothek (DNB) and the Bibliothèque nationale de France (BNF). Although coded identifiers for musical works are available through International Standard Musical Work Codes⁷⁶ (ISWC), responses from the Europeana Sounds *Rights and Metadata Ingestion Survey* carried out in April 2014 show that they are not used by data providers.

Authorities for works are formulated using a preferred or uniform title for the work, preceded by the name of the creator, if there is a known creator, e.g.

Title authority (no known composer): Londonderry air

Name/title authority (known composer): Handel, George Frideric, 1685-1759. Messiah

2.10.1 VIAF

Availability	Search: http://viaf.org/ REST API: https://platform.worldcat.org/api-explorer/VIAF
Terms of use	Available under the Open Data Commons Attribution License (ODC-BY)
Languages	Multilingual
Format	Different formats including XML and RDF

⁷⁵ <http://authorities.loc.gov/>

⁷⁶ <http://www.iswc.org/>

For general information on VIAF, see section 2.8.1. VIAF (Virtual International Authority File) provides multi-lingual instances of some works, including musical works. Each authority contains a VIAF ID (e.g. VIAF ID: 179836400 (Work)) and a permalink (e.g. Permalink: <http://viaf.org/viaf/179836400>).

2.10.2 LC Name/titles and Title authorities

Availability	Search: http://id.loc.gov/authorities/names.html or http://authorities.loc.gov/cgi-bin/Pwebrecon.cgi?DB=local&PAGE=First
Terms of use	
Languages	English
Format	Various formats available http://id.loc.gov/authorities/names.html

Name/title authorities are established in the LC Name Authority file and can be expressed as text strings or as URIs, e.g.

Text string: Handel, George Frideric, 1685-1759. Messiah

URI: <http://id.loc.gov/authorities/names/n82066674>

2.10.3 Recommendations: works

Include controlled vocabulary headings for musical works in metadata for western classical music, using any established authority file. Provide a name/title or title text string (as appropriate) using an established thesaurus, such as LC Authorities or a reference to a URI

e.g.

Mozart, Wolfgang Amadeus, 1756-1791. Symphonies, K. 550, G minor

or

<http://viaf.org/viaf/179424998>

3 SUMMARY OF RECOMMENDATIONS

The purpose of task T1.2 was to identify controlled vocabularies for audio and audio-related objects and recommend the usage of these vocabularies for Europeana Sounds. In this document we have identified and made recommendations pertaining to the following aspects of audio resources: Genre/Form, Mood, Subject, Place, Physical Carrier, Digital Format, Medium of Performance, Personal/Corporate Name, Role, and Work.

Of these types of vocabularies, for Europeana Sounds it will be mandatory for at least one Broad genre term to be included in each metadata record, enabling users to search and navigate across metadata sets from different sources, enhancing discoverability of audio resources.

The table below defines the usage (mandatory or recommended) of vocabularies for Europeana Sounds. Examples of how these vocabularies are expressed in the Europeana Data Model (EDM) will be included in D1.4 EDM Profile for Sound, due for completion at the end of September 2014.

Table 3: Usage of vocabularies for Europeana Sounds

Type	Usage	Vocabularies
Broad genre/form	Mandatory for data providers to add at least one broad genre/form concept to each Europeana Sounds metadata record, when mapping to EDM during ingestion	Use one or more of the following broad concepts: <ul style="list-style-type: none"> • Music • Spoken word • Radio • Environment
Specific genre/form	Recommended to include one or more specific genres to each metadata record	Use term(s) from these sources: <ul style="list-style-type: none"> • <i>Europeana Music Genre/Form Vocabulary</i> • <i>Europeana Non-Music Genre Vocabulary</i> • An established thesaurus, e.g. <ul style="list-style-type: none"> ○ LC Genre/Form ○ FAST ○ EBU
Moods	Not generally recommended. Could be added via user annotations	N/A

Subjects	Recommended to include when appropriate	Use an established thesaurus, e.g. LCSH FAST RAMEAU
Places	Recommended to include	Use an established thesaurus, e.g. GeoNames Getty LC Names FAST VIAF
Audio carriers	Recommended to include	Use term(s) from the <i>Europeana Carriers Vocabulary</i> [name may change], or from an established thesaurus, e.g. RDA Carrier Types LC Carriers Scheme
Digital formats	Recommended to include	Use an established thesaurus, e.g. LC format descriptions RDA encoding formats
Medium of performance	Recommended to include for music materials	Use an established thesaurus, e.g. LCMPT MIMO-VOC RAMEAU
Names	Recommended to include	Provide a name text string using an established thesaurus, e.g. LC Names or a reference to a VIAF or ISNI ID e.g. Mozart, Wolfgang Amadeus, 1756-1791 or http://viaf.org/viaf/32197206 or http://isni-url.oclc.nl/isni/0000000121269154
Roles	Recommended to include	Use an established thesaurus, e.g. RDA relationship designators MARC code list for relators DISMARC dmsvAgentRoles

Works	Recommended to include for western classical music	Provide a name/title or title text string (as appropriate) using an established thesaurus, e.g. LC Names or a reference to a VIAF ID e.g. Mozart, Wolfgang Amadeus, 1756-1791. Symphonies, K. 550, G minor or http://viaf.org/viaf/179424998
-------	----------------------------------------------------	------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

4 APPENDIX

4.1 Appendix I. EDM Profile for Sounds - Collection of Use Cases template



EDM profile for Sounds
Collection of use cases

The Task Force on EDM profile for Sounds aims at creating an EDM profile for describing sounds materials. The first step in the process is the collection of use cases. These cases will allow the Task Force to scope its activities and define clear requirements that will form the basis of the EDM profile for Sounds.

The current fields for the use case templates are:

1) Title of the use case(s)

2) Owner's case_name and name of the represented institution

1



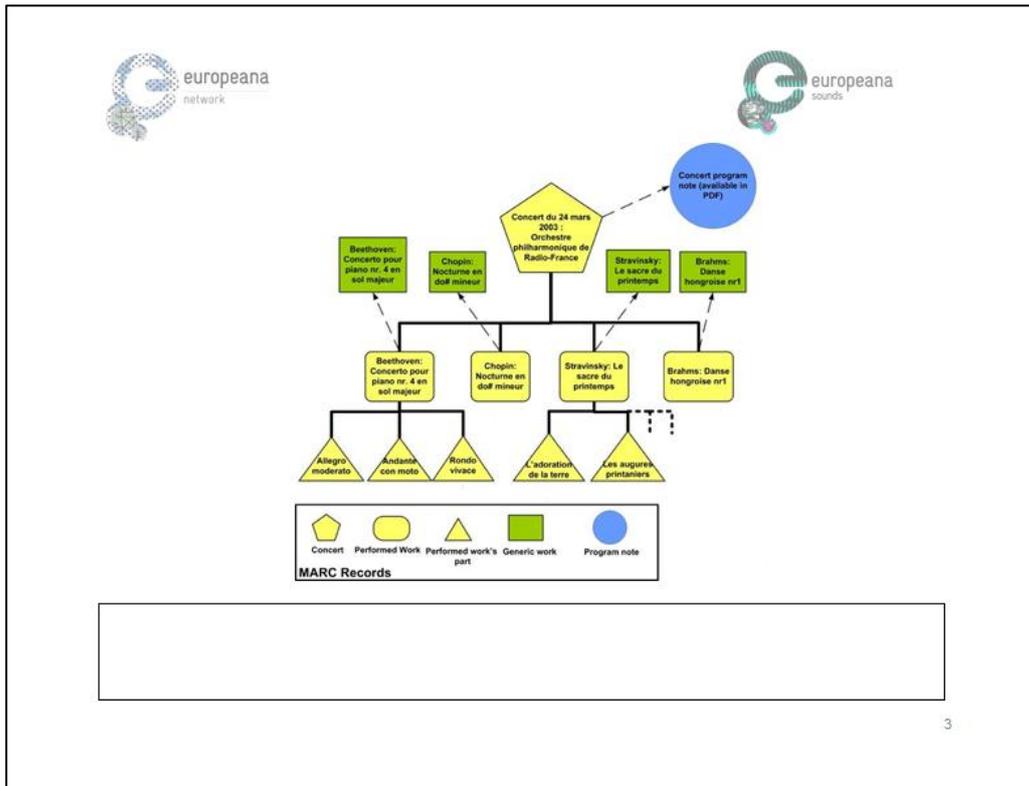
Use case(s):

3) Provide a general description of your use case(s). You can also add to your use case(s) images, diagrams if needed.
A use should contain information on:

- what is the type of sound objects you provide *Europeana* (vinyle, recording, music sheet...)
- Are these objects composed of different elements? If yes, what is the type of the relationship between an object and its different components or between the components themselves?
- Do the objects have digital representations? And if yes, of what type (same recording in different formats, video, images...)?

Example from the report of the Task Force on Hierarchical Objects <http://nro.europeana.eu/web/network/europeana-tech/wiki/Main:Taskforce-on-hierarchical-objects> written by **Rodolphe Bailly, Cité de la Musique**.
At the **Cité de la Musique**, a concert recording is documented as a hierarchy of records.
The first, top level record describes the "concert". It contains general information about the recording of the concert itself, such as title, date, location and the most important performers and composers. The other records describe the "constituents" of the concert, i.e. the musical works performed during the concert. Each of these constituent, situated at a lower level of the hierarchy, contains detailed information about the performed musical works (for instance the complete list of performers). If needed, each of the performed musical work record may also have constituent records, such as specific parts in a symphony.
This concert recording can be represented as follow:

2



3



Data sample illustrating the use case

4) Each use case should be illustrated by (a) data sample(s).
 Upload your data sample(s) at <https://drive.google.com/folderview?id=0B4N3jJ2PToabWU1scDVZGVKVEU&usp=sharing> and indicate the name of the file(s) in this questionnaire for reference.

5) What metadata format is used in the samples(s)?

6) Describe how you use this metadata format to support the described use case(s) (can include details on specific metadata elements, data quality aspects...)

4



Vocabularies

7) Are you using specific vocabularies (e.g. thesauri) to describe information such as subjects, genres in your metadata?
If not go to question 11.

8) If yes, how are you using them? *In which metadata fields?*

9) What is the language coverage of these thesauri/vocabularies?

+

10) How are the thesauri/vocabularies represented (e.g. a database, an XML file, SKOS, etc.)? If possible provide an example of (a part) of the thesauri/vocabularies you use (e.g. a fragment of XML, etc.).

5



Copy and paste an example in the questionnaire or upload your vocabulary sample at
<https://drive.google.com/folderview?id=0B4N3j2PT0abWU1scDVFZGVKVEU&usp=sharing> and indicate the name of the file(s).

Enrichment

11) If you are not using thesauri/vocabularies, would you like to see your metadata enriched with references to thesauri/vocabularies?

+

12) Which metadata fields would you like to see enriched?

6

5 REFERENCES

Ref 1	EC-GA including Annexe I (“Description of Work”) http://pro.europeana.eu/documents/2011409/8d0e9833-4608-494e-af77-681e68f8a8c8
Ref 2	MS3 Initial ontologies selected http://pro.europeana.eu/documents/2011409/0bb14e34-6bd6-4a93-8f20-fb64137fdfe7
Ref 3	D1.1 Content Selection Policy http://pro.europeana.eu/documents/2011409/70e86a37-7608-4f37-9507-1b6f6152cd42