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

MS30 Final assessment on aggregation toolset

Revision: Final

Date: 30/11/2016

Authors: Vassilis Tzouvaras, NTUA

Abstract: This document reports on the final assessment of the aggregation toolset available for Europeana Sounds, delivered by NTUA within Work Package WP5. It serves as an update of MS27, where the user evaluation of the MINT tool is reported. All data providers have used the MINT tool and managed to deliver quality metadata and content to Europeana in the EDM sound profile using the SKOS vocabularies created in the project. The ingestion helpdesk has supported all providers for tackling complex or simple actions.

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

Revision history

Version	Status	Name, organisation	Date	Changes
0.1	ToC	Natasa Sofou, NTUA	23/11/2016	
0.2	1st draft	Vassilis Tzouvaras, NTUA	24/11/2016	
0.4	Final draft	Vassilis Tzouvaras, NTUA	29/11/2016	
1.0	Final	Laura Miles & Richard Ranft, BL	30/11/2016	Layout, minor changes

Review and approval

Action	Name, organisation	Date
Reviewed by	Sergiu Gordea, AIT	28/11/2016
Approved by	Coordinator and PMB	30/11/2016

Distribution

No.	Date	Comment	Partner / WP
1	30/11/2016	Submitted to the European Commission	BL/WP7
2	30/11/2016	Posted on Europeana Pro website	BL/WP7
3	30/11/2016	Distributed to project consortium	BL/WP7

Application area

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

Statement of originality

This document contains original unpublished work except where clearly indicated otherwise.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

Project summary

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

- Double the number of audio items accessible through Europeana to over 1 million and improve geographical and thematic coverage by aggregating items with widespread popular appeal such as contemporary and classical music, traditional and folk music, the natural world, oral memory and languages and dialects.

- Add meaningful contextual knowledge and medium-specific metadata to 2 million items in Europeana's audio and audio-related collections, developing techniques for cross-media and cross-collection linking.

- Develop and validate audience specific sound channels and a distributed crowd-sourcing infrastructure for end-users that will improve Europeana's search facility, navigation and user experience. These can then be used for other communities and other media.

- Engage music publishers and rights holders in efforts to make more material accessible online through Europeana by resolving domain constraints and lack of access to commercially unviable (i.e. out-of-commerce) content.

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

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

Copyright notice

Contents

Executive summary: MS30 Final assessment on aggregation toolset

The aim of Work Package WP5 *Technical Infrastructure* is to enable metadata aggregation by extending and enhancing the existing Europeana aggregation infrastructure. This milestone reports on the aggregation infrastructure prototype available, as delivered by NTUA.

The proposed aggregation mechanism extends Europeana's infrastructure toolset, allowing for the ingestion of semi-structured data and offering the ability to align and take advantage of well-defined, machine understandable schemes in an intuitive manner. The delivered prototype reported in this milestone document offers aggregation services that cover registration and metadata import, mapping and transformation, and publication as part of WP5, reported in MS23 *Revised aggregation design available.* The basic aggregation functionality within European Sounds is covered by MINT ingestion services. Technical specifications regarding implementation details, installation instructions, platform access and usage have been reported in MS24 *Aggregation infrastructure prototype available*. In this report, we give an update on the aggregations mechanism version 2, and the changes that have been applied to previous versions (as described in MS24).

1   Introduction

The aim of Work Package WP5 is primarily to enable metadata aggregation by providing an aggregation infrastructure that extends and enhances the existing Europeana infrastructure. In interaction with WP1 (Aggregation), EDM sounds has been deployed as an anchor, to which various data providers can be attached and become, at least partly, interoperable, while controlled vocabularies as defined in T1.2: *Ontologies* were converted to SKOS and used in the enrichment processes.

In the Europeana Sounds framework the basic aggregation functionality is covered by MINT ingestion services, as already described in MS24 *Aggregation infrastructure prototype available*. MINT services compose a web-based platform designed and developed to facilitate aggregation initiatives for cultural heritage content and metadata in Europe.

2   Aggregation toolkit

The MINT platform serves as the basic aggregation mechanism in the Europeana Sounds framework. It offers a user and organisation management system that allows the deployment and operation of different aggregation schemes (thematic or cross-domain, international, national or regional) and corresponding access rights.

The main role of the MINT ingestion platform in the Sounds project is to enable users to:

- Provide metadata records in a range of "source" formats.

- Convert metadata to selected target schema (currently EDM, EDM sounds profile).

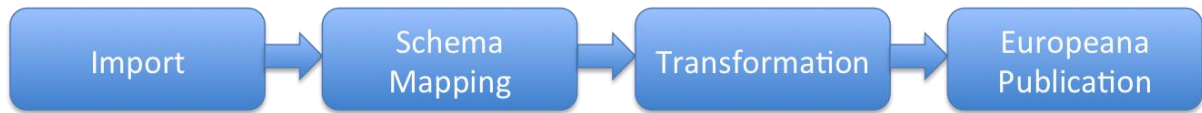- Monitor the progresses of content provision.

Figure 1: MINT ingestion workflow

3   Basic functionality

3.1   Mapping module

The mapping of metadata into the desired schema is the basic functionality of the MINT ingestion platform. A visual mapping editor for the XSL language is used. The mapping of metadata schemes is performed following a drag-and-drop procedure, by selecting items form input schema area and dropping them in mapping area. Input operations are translated into the corresponding XSLT code. The mapping editor supports string manipulation functions for input elements, structural element mappings, constant or controlled value assignment (i.e. from target schema enumerations), conditional mappings and value mappings between input and target value lists. By setting preview options users have the ability to preview XML code of import and transformed items.
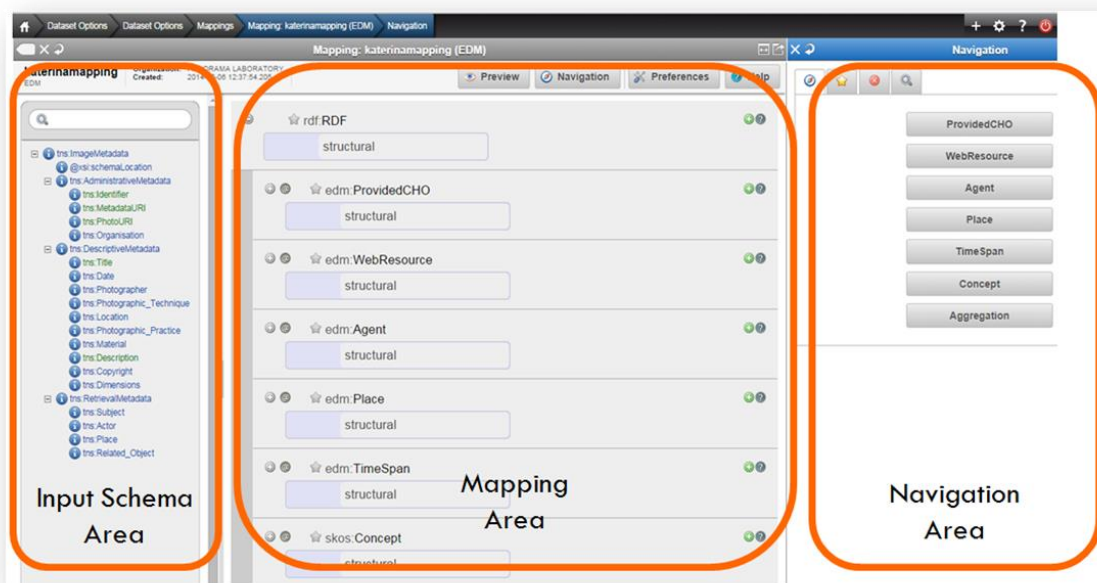

Figure 2: MINT browsing schema

3.2   EDM Sounds Profile update

The EDM sounds profile specification was designed and released under T1.3: *EDM profile* and in D1.4: *EDM profile for sound* (http://pro.europeana.eu/files/Europeana_Professional/EuropeanaTech/EuropeanaTech_taskforces/EDMSound//TF_Report_EDM_Profile_Sound_301214.pdf).

Using the aforementioned specification NTUA implemented the actual schema in XSD, which was then deployed in MINT, in order to enable users to map their metadata into EDM sounds profile schema, which is the sound enriched version of EDM used as intermediate schema before publishing to Europeana. The EDM sounds profile XSLT is provided as complementary material.
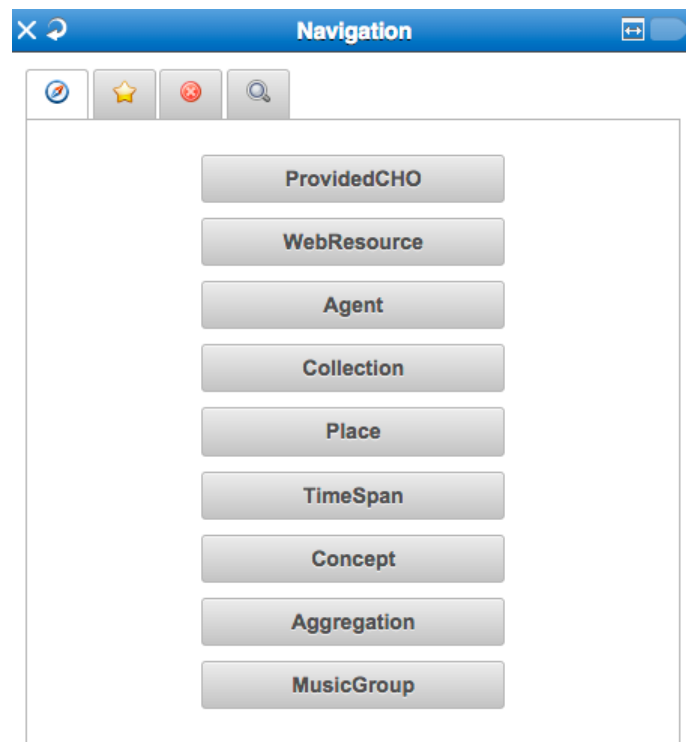

Figure 3: Top classes of the EDM profile for Sounds

3.3   SKOS vocabularies integrations

The vocabularies designed and delivered within WP1 have been converted to SKOS representation and the resulted thesaurus (which can be thought to be composed of all sounds vocabularies). This is integrated into the MINT platform in order to enable data providers to align their metadata with the elements of the specific thesaurus.
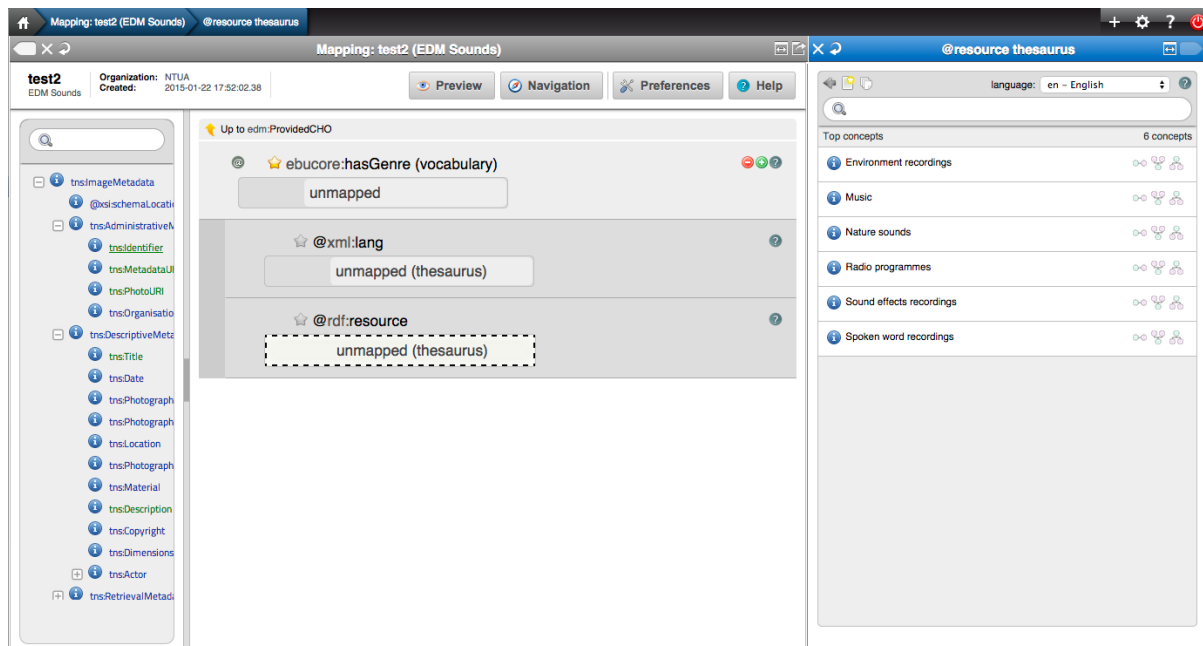
Figure 4: Genres thesaurus in MINT

3.4   OAI publication

The main objective of the publication task is the delivery of high quality metadata to Europeana. MINT enables crosswalks between EDM Sounds profile and EDM schema through a user-friendly interface. In addition, it provides validation services for both EDM Sounds and EDM metadata models together with preview interfaces, through which providers can a-priori check how would their metadata look like when published on Europeana, thus ensuring high quality metadata.

3.5   Group edit functionality for metadata cleaning and normalization

In order to facilitate advance metadata cleaning and normalization functionalities, a number of operations were defined under the group edit module:

- The creation of an element on each selected record, which can be filled with a user defined value.

- The deletion of an element based on conditional criteria from all the selected records.

- Updating of a value from existing elements.

The group edit module is integrated into MINT platform. More details about it can be found in MS25 *Sounds SKOS ontology and normalization and cleaning module beta*.
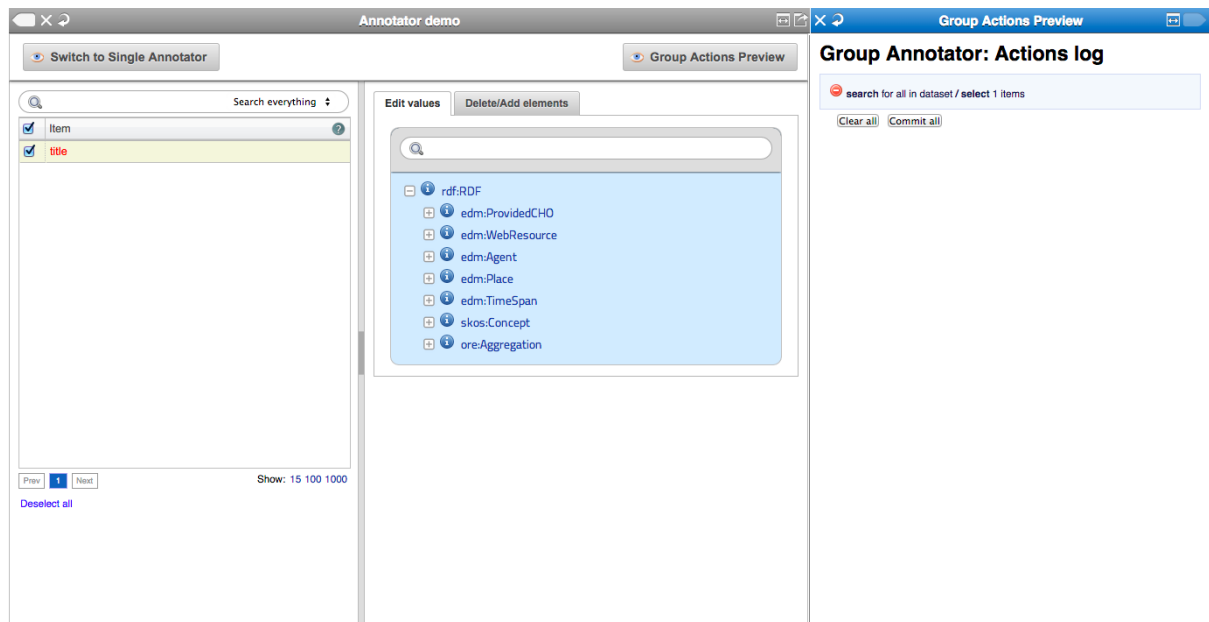
Figure 5: Group Annotator

4 Assessment

The final assessment is mostly positive. MINT has been used by all data providers very successfully to deliver quality metadata and content to Europeana. With a few exceptions, users easily created crosswalks from their proprietary schema in the EDM sounds profile using the SKOS vocabularies created in WP1. Only in a few cases, users had to use the advanced mapping functionality and the support of the helpdesk to create complex crosswalks. During all phases of the ingestion process, a detailed preview and statistics services were available to help users create the correct mappings. The helpdesk supports (NTUA for the technical issues, British Library and Europeana for the schema issues) aided users for complex or simple actions. A detailed reporting service was available to British Library to monitor the process and warn providers that may not have been meeting targets as anticipated.

Appendix A: Terminology

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

Additional terms are defined below:

Term	Definition
EC-GA	Grant Agreement (including Annex I, the Description of Work) signed with the European Commission
EDM	Euroepana Data Model
GA	General Assembly
MINT	Metadata Interoperability Platform
NTUA	National Technical University of Athens
OAI	Open Archives Initiative
PC	Project Coordinator
PI	Performance Indicator
PMB	Project Management Board
SKOS	Simple Knowledge Organization System
TEL	The European Library
UAP	User Advisory Panel
WP	Work Package
XML	Extensible Markup Language
XSLT	Extensible Stylesheet Language Transformations
XSD	XML Schema Definition