Europeana Space – Spaces of possibility for the creative reuse of Europeana's content

CIP Best practice network - project number 621037

Deliverable number	D2.2
Title	The metadata processing unit

Due date	Month 12
Actual date of delivery to EC	31 March 2015

Included (indicate as appropriate)	Executive Summary	☒	Abstract	☐	Table of Contents	☒

Project Coordinator:

Coventry University

Prof. Sarah Whatley

Priority Street, Coventry CV1 5FB, UK

+44 (0) 797 4984304

E-mail: S.Whatley@coventry.ac.uk

Project WEB site address: http://www.europeana-space.eu

Context:

Partner responsible for deliverable	NTUA
Deliverable author(s)	Nasos Drosopoulos
Deliverable version number	1.0

Dissemination Level	
Public	☒
Restricted to other programme participants (including the Commission Services)	☐
Restricted to a group specified by the consortium (including the Commission Services)	☐
Confidential, only for members of the consortium (including the Commission Services)	☐

History:

Change log			
Version	**Date**	**Author**	**Reason for change**
0.1	03/12/2014	Nasos Drosopoulos, Vassilis Tzouvaras, Maria Symeonaki, Giorgos Stamou	First draft Requirement analysis Overview of existing services Architecture of the MPU Semantic publication, data enrichment & cleaning approaches

0.2	14/01/2015	Nasos Drosopoulos, Giorgos Marinellis Vassilis Tzouvaras	MPU section Integration in the Technical Space
0.3	13/02/2015	Nasos Drosopoulos, Alexandros Chortaras, Natasa Sofou	Data enrichment & cleaning section
0.4	19/02/2015	Nasos Drosopoulos	Prepared for Peer review
0.5	16/03/2015	Nasos Drosopoulos	Review feedback from Marinos Ioannides, CUT
1.0	30/03/2015	Nasos Drosopoulos	Incorporating comment from Tim Hammerton, Project Manager, COVUNI

Release approval			
Version	**Date**	**Name & organisation**	**Role**
1.0	31/03/2015	Tim Hammerton, COVUNI	Project Manager

Statement of originality:

This deliverable contains original unpublished work except where clearly indicated otherwise.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.

TABLE OF CONTENTS

EXECUTIVE SUMMARY

The Technical Space infrastructure of the Europeana Space project aims to offer to cultural institutions, professional users and software developers (both within and outside of the project) an intuitive platform for discovery and manipulation of cultural heritage resources so as to use and re-use them in their efforts to disseminate knowledge, reach new audiences, advance scholarly research and develop innovative applications. The Metadata Processing Unit (MPU) consists of a set of services for the establishment of metadata interoperability and the semantic enrichment of knowledge in the digital cultural heritage field. The resulting repository of metadata, annotations and enrichments will be accessible via traditional metadata delivery protocols (such as OAI-PMH), programmatic interfaces and semantic web technologies (such as SPARQL).

This document reports on the requirements regarding the management and publication of metadata collections, the specification and functionalities of the MPU and its APIs and, the release and documentation of the MPU services. The MPU is based on NTUA's MINT platform that supports different protocols for importing metadata and a user-friendly interface for implementing and applying crosswalks to formal metadata models such as EDM and LIDO. In addition, annotation services, group-edit tools and a semantic enrichment workflow engine allow for the evolution of knowledge around resources through the use of SKOS vocabularies, ontologies and Linked Data repositories.

In the framework of Task 2.3 (metadata processing unit) NTUA spent effort on the collection of requirements regarding metadata manipulation, on the deployment and maintenance of MINT, on small development tasks and planning for its integration with the Technical Space (mainly user management and publication) and, on supporting and training the pilot teams to ensure they can setup and complete their metadata processing workflows.

1 INTRODUCTION

The significance of cultural heritage and the constant generation of digital content attract more and more scientists, content holder institutions, educational organizations, business developers, creative industries and users from all backgrounds. These are looking for intuitive and innovative ways of aggregating, browsing, accessing, annotating, sharing, and in general using and leveraging the availability of cultural heritage repositories to achieve their goals. It is though those same features that denote the significance and richness of cultural heritage, such as diversity, growth and dispersion of sources, which maintain, or raise a wall that inhibit its transition and transformation.

In this environment, knowledge-based management and retrieval becomes a necessary evolution from simple syntactic data exchange between databases. In the process of aggregating heterogeneous metadata resources and publishing them for retrieval and creative re-use, networks such as Europeana and DPLA invest in technologies that achieve semantic data integration. Integration of diverse information is achieved through the use of formal ontologies, enabling reasoning services to offer powerful semantic search and navigation mechanisms. This, in turn, can empower the efforts towards re-using the content in a wide range of applications for the creative industry.

Europeana Space aims at the establishment of the Technical Space as a framework for storing, accessing and processing content and metadata. Its Metadata Processing Unit (MPU) offers aggregation services for the alignment and enrichment of DCH resources based on NTUA's MINT mapping tool. MINT implements an aggregation infrastructure offering a crosswalk mechanism to support subsequent critical activities:

- Assisting content providers to aggregate, transform and validate their metadata according to the Europeana Data Model.
- Allowing users to enrich metadata records (either their own, or metadata harvested from Europeana and other sources) using SKOS terminologies and ontologies.
- Linking metadata with internal and external resources, obtaining in that way additional information for the objects, places and people of interest.
- Publishing collections according to the data models and serializations required for their re-use.

Deliverable 2.1 - *Requirements for the creative use of Digital Cultural Resources; progress on collaboration towards Europeana Labs* - reported on requirement analysis, a process that complemented WP4 tasks that addressed pilot methodology, content sourcing and coordination (reported in D4.1 - *Pilots methodology and content sourcing* - and D4.2- *Pilots coordination: information on technical planning*), and which allowed for the identification and validation of requirements pertaining to the MPU (content sources, formats and licensing, metadata models, serializations and transformations, accessibility and expected operational scenarios). D2.1 also documents the architectural and technological choices for the Technical Space that will ensure the integration of different services and polyglot repositories.

The present report evaluates the requirements that directly affect the MPU in Section 2, and present its functionality in Section 3. Section 4 discusses the architectural and technological choices along with MPU's integration in the Technical Space, while Section 5 informs about the documentation and support for the MPU.

2 CULTURAL HERITAGE METADATA IN EUROPEANA SPACE

The Europeana Space project is developing Pilots (and subsequently hackathons) in six thematic areas in order to explore different scenarios of re-use of digital cultural content that is available through Europeana and other relevant sources. The pilots' initial planning on content sourcing and metadata manipulation, which is listed in the DoW, was further specified during the requirement analysis that is reported in D2.1, and is presented in detail through the work and deliverables of WP4 regarding methodology, content sourcing and technical planning. These results highlight that ease of access and manipulation of metadata resources, with the purpose of enabling interoperability between different repositories, is crucial for the successful implementation of identified use cases.

In particular, in terms of content the pilots are expected to use available content types as defined by Europeana, i.e. Image, Video, Text, Audio and to a lesser extent 3D resources (multimedia content). Content sourcing is performed predominantly through Europeana but also involves content curated by the pilots (from the regional, national and private archival collections of partners) while in the course of prototyping and incubation further identified sources will be addressed. These include aggregators such as the Digital Public Library of America and the Digital Commons Network, domain specific repositories such as the International Database for Artistic Research, the Archive of Digital Art and Europeana Early Photography, national aggregators, media repositories such as Beeld en Geluid's Open Images, Wikimedia Commons and Critical Commons. Pilots also expect to take advantage of widely used content delivery services such as Vimeo for videos, Flickr and Picassa for images.

Licensing of content and metadata is especially important in defining the features of the pilot applications as well as for taking the steps towards incubation and business development. In terms of accessing and managing license information using metadata, it is important that available information is included in a formal way while there are several use cases that target multiple licensed content and indicate the necessity of a license-based retrieval mechanism. This issue is tightly connected with the content retrieval mechanism and WP3's Content Space, while the MPU focuses on remediating any licensing information during metadata manipulation and enabling the alignment with sets of terms of use such as the Europeana Licensing and Content Re-use Frameworks.

In terms of data models, primary focus is given to the Europeana Data Model (EDM) as an aggregation schema, while domain specific models such as LIDO or EDM profiles (e.g. DM2E, Sounds) offer the required expressivity for specific collections and the application of further metadata generating services such as annotation. Required serializations for metadata records include XML, RDF/XML, N-triples and JSON-LD and there are three access scenarios, via HTTP download, HTTP APIs and, SPARQL endpoints. Vocabularies and terminologies that have been created or identified in domain aggregators are commonly used by respective content providers and should be used in the MPU in order to further homogenize access to different datasets.

Finally, the operational scenarios envisioned by pilots dictate the requirements regarding delivery of metadata and potential exposure of the MPU's services for usage by third-party applications. Applications envision a range of methods for populating their interfaces with cultural heritage resources, starting from curation and ingestion during development or deployment up to offering users live access to repositories. An important last point regarding the use of metadata in applications is round-tripping, when referring to pilots that expect to allow users to create metadata that content providers may wish to access in order to further enhance the original repositories.

3 METADATA INTEROPERABILITY AND DATA ENRICHMENT SERVICES

Aggregation using formal data models and the ability to access different serializations according to expected usage are needed to facilitate the pilots' design and development approaches. However, aggregating metadata records from different repositories through the application of crosswalks between two schemas may still create confusing display results, especially if some of the metadata was automatically generated or created without following a model's guidelines and best practices. The crosswalk itself implements mappings between metadata elements from different schemas, thus allowing interoperability on the syntactic level, but usually (depending on the differences in expressivity between the two models) without leveraging more expressive structures used to describe the content. Moreover, the use of different thesauri and terminologies without a formal connection between them hinders the ability to access the huge collection of resources that is being made available these days through a common interface. In that sense, the Technical Space and the MPU in particular aim to offer efficient, intuitive tools that enable semantic interoperability.

The Metadata Processing Unit is the main platform for aggregating and managing content providers' collections. It offers content providers the ability to aggregate, validate and transform their metadata to the Europeana Data Model and profiles, or any identified data model of interest. It then allows users to enrich metadata using SKOS terminologies and ontologies as well as to link to external data sources. The MPU is based on NTUA's MINT platform and is setup[1] for use by pilot development and content curation teams. The current set of functionalities include:

- Import using identified delivery protocols
- Visual mapping editor for crosswalk generation
- Transformation
- Schema Validation
- Data Cleaning
- Reconciliation with SKOS vocabularies
- Publication

MINT is being extended for the MPU in order to include a semantic layer for publication of RDF metadata to a semantic repository. This includes procedures regarding the generation of persistent URLs for all resources in metadata, linking to external sources and validation of the resulting dataset according to Linked Open Data requirements. The semantic layer enables further processing of metadata using semantic web technologies such as automatic and semi-automatic data enrichment tools. Furthermore, semantic publication enables the introduction of reasoning systems to take advantage of the available ontological knowledge.

3.1 AGGREGATION AND PUBLICATION WORKFLOW

The workflow engine allows for the implementation of aggregation and publication workflows for all the identified user scenarios. NTUA has broad experience in aggregation, having served

[1] http://mint-projects.image.ntua.gr/espace

a significant portion of the aggregation projects in the Europeana ecosystem during the last 8 years. MINT is also being used for Europeana's internal aggregation system and through this collaboration we have approached the definition of a common, formal data model for digital cultural heritage workflows. It allows for flexibility in the range of use cases it can support, addresses interoperability with third party services and, enables preservation of all required information that document the aggregation and publication activities.

3.2 METADATA INGESTION AND MANAGEMENT

The MPU allows for any data model as input when ingesting, while aggregation and publication can be performed using EDM and available profiles or more domain specific standards such as LIDO. Input record serialization can be CSV, XML or JSON with preview interfaces for raw data and available HTML renderings. Input repositories can be accessed using HTTP upload and download, the OAI-PMH protocol and HTTP API for the storage layer of the Technical Space (in cases that require manipulation of metadata collected in the Technical Space from external sources).

User and organization management allow for setting up aggregators and sub-aggregators, implementing parent-child organization supervising scenarios and detailed control over access and roles for users.

3.3 MAPPING EDITOR

Metadata mapping is the crucial step of the ingestion procedure. It formalizes the notion of a metadata crosswalk, hiding the technical details and permitting semantic equivalences to emerge as the centerpiece. It involves a user-friendly graphical environment (Figure 1 shows an example mapping opened in the editor) where interoperability is achieved by guiding users in the creation of mappings between input and target elements. User imports are not required to include the respective schema declaration. User's mapping actions are expressed through XSLT style sheets, i.e. a well-formed XML document conforming to the namespaces in XML recommendation. XSLT style sheets are stored and can be applied to any user data, exported and published as a well-defined, machine understandable crosswalk and, shared with other users to act as template for their mapping needs.

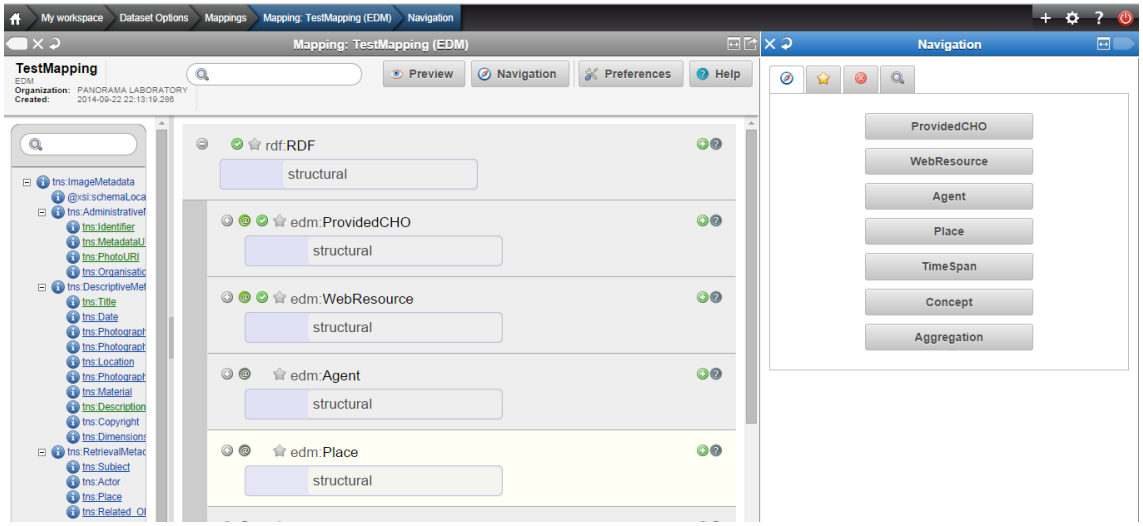

Figure 1. MINT's mapping editor

The user interface of the mapping editor is schema aware regarding the target data model and enables or restricts certain operations accordingly, based on constraints for elements in the target XSD. For example, when an element can be repeated then an appropriate button appears to indicate and implement its duplication. Several advanced mapping features of the language are accessible to the user through actions on the interface, including:

- String manipulation functions for input elements.
- N-1 mappings with the option between concatenation and element repetition.
- Structural element mappings.
- Constant or controlled value assignment.
- Conditional mappings (with a complex condition editor).
- Value mappings editor (for input and target element value lists).
- Custom XSL code inclusion.

Preview interfaces present the steps of the aggregation such as the current input xml record, the XSLT code of mappings, and the transformed record in the target schema, subsequent transformations from the target schema to other models of interest, and available html renderings of each xml record.

3.4 DATASET INDEXING AND STATISTICS

MINT indexes datasets and calculates statistics to facilitate the mapping process and offer querying/filtering for services such as the group-edit. The Statistics panel lists the XPaths (i.e. elements) of an imported dataset together with their distinct values count and the average length of their values, while by clicking on an element you can browse its values.

Indexes for XML datasets are configured automatically while manual configuration interfaces allow for more elaborative queries. The user can filter the items appearing on the Item Browser by searching over a dataset in different ways:

1. Search everything: Finds all items with a field that has a value containing the string provided by the user.
2. Search label: Find all items whose label contains the string provided by the user.
3. Search field: Find all items that contain an element whose value matches the string entered by the user.
4. Search using the SOLR query syntax.

3.5 DATASET VALIDATION, TRANSFORMATION AND VERSIONING

MINT offers schema validation for formal data models (such as EDM and profiles, domain models like LIDO), based on the XSD and/or Schematron rules, and validation for serialization formats and containers (CSV, XML, ZIP).

It uses an XML processor to transform datasets between schemas using the mappings (XSLT) produced by the editor or uploaded from external tools. It also contains a set of established, formalized crosswalks between data models that can be applied (e.g. LIDO to EDM mapping).

The different versions of a dataset, produced by transformations or data manipulation, are stored and can be accessed individually.

The MPU maintains the separation of metadata records based on their grouping in datasets when importing. Some if its services, such as the group edit, allow a user to exclude items from a dataset after transformations (e.g. tagging or enrichment). Merging of datasets (input or

transformed) is not allowed natively in MINT, which exposes them as such to the Technical Space. The UI and collection management system of the latter will eventually allow for more flexible management of datasets, such as the ability to split, merge and combine.

3.6 METADATA ANNOTATION AND CLEANING

Annotations are a way to further transform datasets, either by editing a single item or by filtering a dataset and batch editing the resulting set. In both cases the changes are saved in the database and as such, when clicking on a Transformation the user will see the latest (annotated) version of the items. The user can restore the dataset to its original state, just after the respective transformation was applied by reversing annotations.

MINT includes a single item annotator as well as a group-edit tool. The former (Figure 2) enables users to edit structured XML content, rendered in an intuitive data entry environment based on MINT's mapping tool UI components. This gives non-technical users the ability to create structured content including semantics and structure definitions based on an XSD. The interface can switch between a simple view, in which each record is rendered fully as a tree and, a complex view where pre-defined bookmarks allow the grouping of specific groups of elements in a way that is more intuitive for the user. The latter allows for setting up of predefined element groups, of bookmarks and in general for highlighting the elements of importance for the purpose of helping content providers, while the former allows an experienced (both in the usage of the mapping editor as well as in the target data model of interest) user to navigate and use the full definition of the target schema for his annotation.

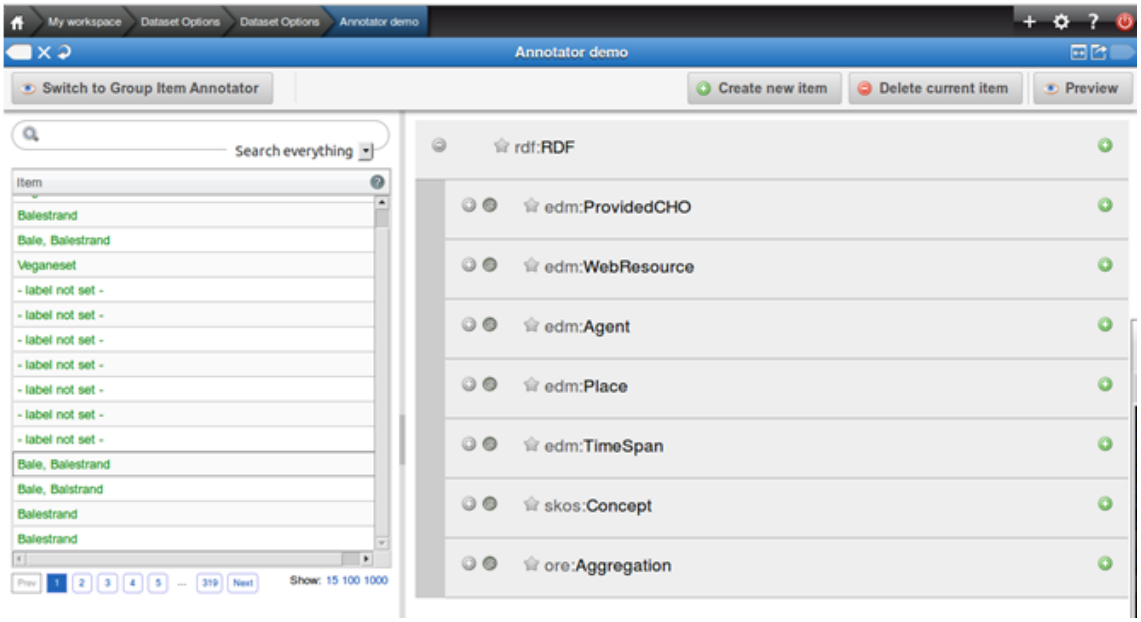

Figure 2. XML metadata annotator

The group edit service (Figure 3) allows for data cleaning processes (adding, deleting and replacing element values) based on filtering datasets. The search capability is used to define groups of records based on specific criteria while allowing to deselect and exclude records from the result set. Based on this set generation, a number of operations are supported - on a schema level for the addition, deletion and update of elements with specific values (free text or controlled) and, on a data level for the application of conditional edits, string manipulation or numerical functions on already existing values. These operations aim at correcting values for a specific item or across a dataset, switching between formats (e.g. for dates), and perform

value mapping between input terminologies and target schema's associated controlled vocabularies or thesauri. For the latter it supports SKOS and offers an intuitive interface for navigating SKOS hierarchies.

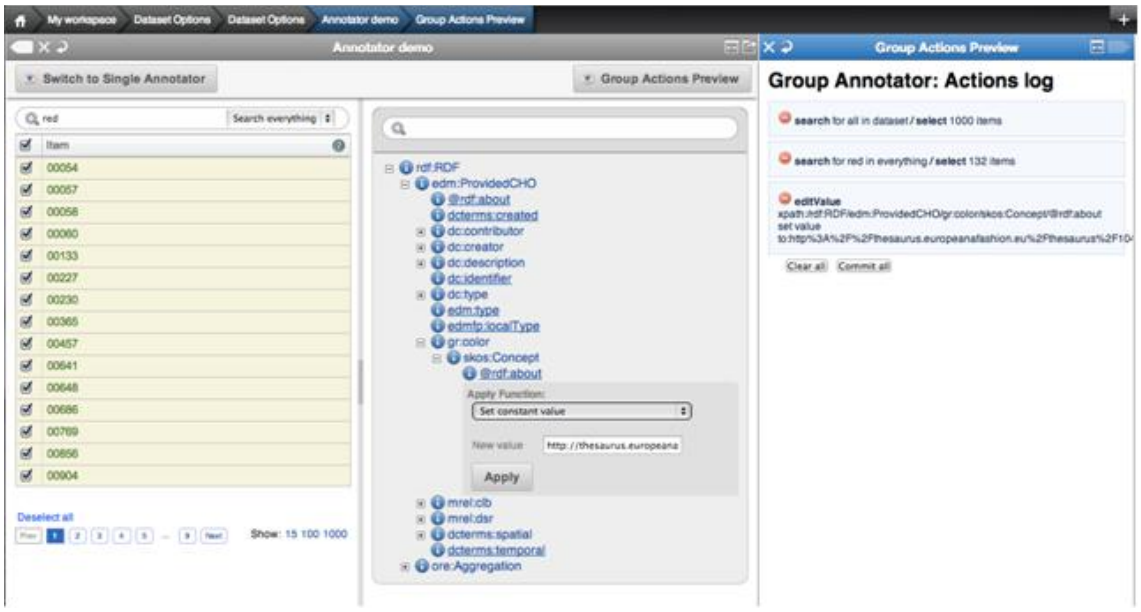

Figure 3. Group edit

It is able to offer pre-configured group edit processes that implement specific tasks such as tagging a (sub-) set of items for a specific element (e.g. place or color).

3.7 DATA ENRICHMENT AND LINKING

MINT has introduced in the Europeana Fashion[2] project a more elaborate data enrichment and linking tool, in which (mostly) free text data entered by the various providers were processed in order to extract semantic information, enrich the objects in which they refer and link them with entities of the Fashion Ontology. The starting point for the design of the ontology was the Fashion Thesaurus, which provides the main categorization of fashion objects along with a quite rich vocabulary for the essential object categories (more details can be found in the project's deliverables[3] [4].

The tool uses a workflow engine and respective interface for setting up the required processing steps, where the content provider or an expert can design appropriate workflows and combine available native or third party services (text processing, translation services, entity extraction, natural language processing etc.) into a chain of steps that appropriately enhances an input record or dataset. This approach becomes important given that most of the metadata properties allow free text descriptions and controlled vocabulary values (URIs), something that on one hand allows for easy mapping when no controlled vocabularies are used in the input data, but on the other prevents homogenization and further semantic processing of datasets by not enforcing specific terminology for a given element.

[2] http://www.europeanafashion.eu/

[3] http://cordis.europa.eu/docs/projects/cnect/7/297167/080/deliverables/002-EuropeanaFashionDeliverable33aEnrichmentToolReportv1.pdf

[4] http://cordis.europa.eu/docs/projects/cnect/7/297167/080/deliverables/002-EuropeanaFashionDeliverable23EuropeanaFashionThesaurusv1.pdf

While a controlled vocabulary value is a resource with a certain URI that has a well-defined meaning (which is defined in the context of the vocabulary or ontology in which it is included) and can be used in conjunction with semantic technologies to allow e.g. semantic query answering, a free text value lacks any semantics and can be used in a limited way only in conjunction with traditional keyword search (i.e. string matching). The aim of this enrichment process is thus to take a free text description and compute a set of concepts (in particular URI's of a domain Thesaurus or of other external commonly used vocabularies, that characterize some aspects of the content of the respective free text). The enrichment process is thus at the same time a linking process of the content to external resources of the Linked Data cloud.

This service will be generalized for the Europeana Space Technical Space in order to use available published vocabularies from selected sources instead of only those specifically set up or connected to a target schema. A set of predefined workflows (e.g. extract entities from the description element of an EDM record and look up terms using published open vocabularies[5] or knowledge bases such as DBPedia) will be provided for quick usage and integration of the service in a typical setup, while an advanced user shall be able to define specialized enrichment tasks when needed.

The setup of the workflow consists of developing a customized enrichment/linking process for each provider, taking into account the special data form used by the provider. The whole enrichment process consists of a series of smaller processing steps whose initial input is a set of item metadata properties values (containing several free text descriptions) for each item, and the final output is the same set of metadata properties values, augmented with a set of URI's that have been added as values for the metadata properties for which some relevant vocabulary values could be detected. To facilitate the implementation and view the enrichment process in a more general and adjustable way (independently of each provider), it is split into several self-contained sub-processes:

- Data Retrieval, by using an RDF store and appropriate SPARQL.
- Concept Label Generation**,** to create the multi-lingual concept categories that can be used for enrichment (e.g. material, techniques, colors).
- Thesaurus-based Enrichment, being the main component that applies NLP and regular expression string matching techniques to detect occurrences of the concepts in metadata values.
- DBPedia-based Linking, to link textual values using an external vocabulary, in particular the DBpedia ontology.
- Enriched/Linked Data Generation, to update the original resources with the results.

[5] lov.okfn.org/

4 THE METADATA PROCESSING UNIT IN THE TECHNICAL SPACE

The MPU is an integral part of the Technical Space architecture (Figure 4), used to establish interoperability between the various formats for metadata ingested, harvested or discovered using the platform. Apart from its main functionalities that were described in the previous section, it is used for metadata storage in XML and implements the important modules of publishing to Europeana (via OAI-PMH) and to the semantic store. Finally, it serves several of the interfaces required for metadata access (OAI-PMH, lucene-based indexing, XML records API) as well as a scalable processing infrastructure for XML.

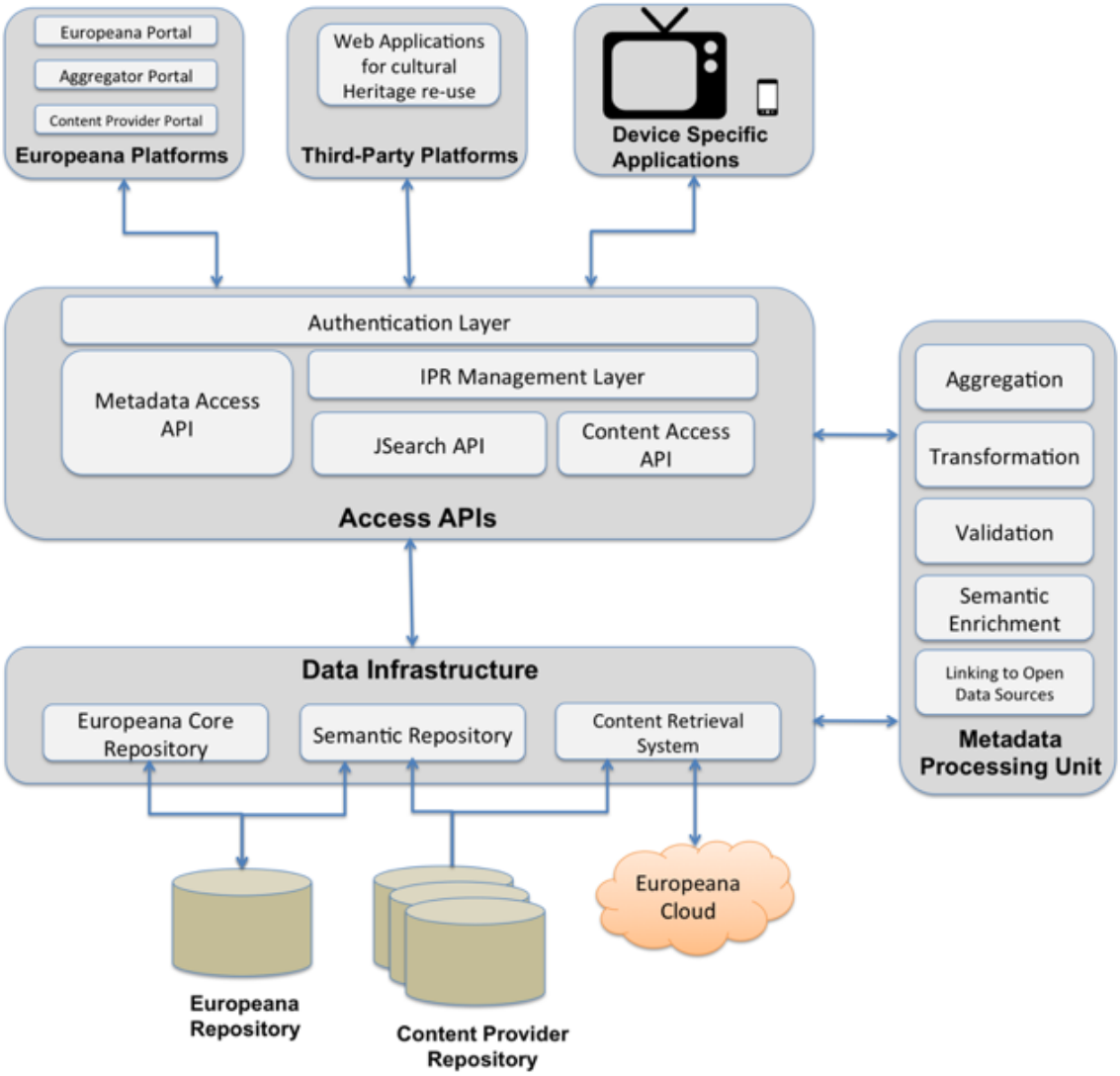

Figure 4. High-level architecture for the Technical Space (from the Description of Work)

4.1 STORAGE FOR XML AND SEMANTIC WEB SERIALIZATIONS

The storage layer of the Metadata Processing Unit offers an XML repository in which metadata taken from Europeana or from Content Providers' repositories can be stored, as well as the versions produced after operations by the MPU services. It stores all versions of a dataset used in a workflow in XML serialization, along with formal transformations (XSLT) between them.

With the introduction of the new workflow data model (see section 3.1), file storage is decoupled from housekeeping data that reside in a relational database (PostrgesDB). For XML and JSON files, an appropriate NoSQL system has been integrated (MongoDB, used for XML files based on which MINT's OAI-PMH server is implemented).

Published datasets can also be available through a semantic store, an industrial-strength repository following W3C recommendations for RDF serializations and the SPARQL 1.1 Query Language. Having reviewed and tested several solutions for RDF storage such as graph databases and triplestores, among them *Bigdata*, *OWLIM*, *4Store, Neo4j, SHARD, Dydra* and *Sesame*, MINT now uses the *Apache Jena* TDB store and *Fuseki* SPARQL server. The setup provides a high performance RDF store and a server that provides REST-style SPARQL HTTP Update, SPARQL Query, and SPARQL Update. It uses the SOH (SPARQL Over HTTP) set of scripts for working with SPARQL 1.1.

WP2 will also investigate the use of different indexing techniques for RDF data; using the built-in engine of the repository, the Technical Space should offer the ability to query using materialized views.

The transformation between XML and semantic web serializations for datasets that are represented using RDF vocabularies are performed by the system, transparently to the user. Finally the user will be able to access input and published datasets in a similar coherent manner by both the Technical Space interface and the MPU, and switch to use the MPU services in an integrated way.

4.2 MPU TECHNICAL SPECIFICATIONS

It is written in JAVA, JSP, HTML and Javascript. It uses PostgreSQL as an object-relational database with Hibernate as the data persistence framework, and mongoDB as a document-oriented database. MINT is also reusing other open source development frameworks and libraries according to specific deployments and customizations. MINT source code versions are released (some versions available at https://github.com/mint-ntua) under a free software license (GNU Affero GPL).

4.3 PROCESSING INFRASTRUCTURE

XML processors (Apache Xerces, SAXON, Nux) are used for validation and transformation tasks as well as for the visualization of XML and XSLT. For issues of scalability with respect to the amount of data and concurrent heavy processing tasks, parts of the services are multi-threaded or use specific queue processing mechanisms.

Specifically, the processing infrastructure (MINT PI) serves as a scalable mechanism for structured data processing that can be deployed on cloud infrastructures (prototyped in the INDICATE project). It is built using RabbitMQ in its core and employs two distinct queue patterns, an RPC Queue pattern which is used for cases where the client desires to block while the processing is executed and also awaits for a response in a pre-defined format, and a Working Queue pattern which is used for non blocking processing where the client submits the data for processing and does not wait for a response. The first case is mainly used for the implementation of specific commands, e.g. for implementing the cleaning or deletion of a repository, while the second case is used for bulk processing of raw data, e.g. data transformation and enrichment of records.

The Technical Space will introduce a new Processing Infrastructure that will be also available to the MPU and current work involves its integration and application where it can offer higher performance and elasticity. It is being implemented using Akka (http://akka.io/), a toolkit and runtime for building highly concurrent, distributed, and resilient message-driven applications

on the JVM. NTUA is also following the specification and development of Europeana Cloud's Data processing service (DPS) that will be used for the transformation of metadata records.

5 DEPLOYMENT, SUPPORT AND DOCUMENTATION

WP2 has deployed the metadata processing unit, enabling the appropriate, formalized sourcing of records and resources from content providers, Europeana and other identified 3rd party repositories and, the eventually required manipulations in terms of data and its various models and serializations. The MPU can publish finalized datasets to the Technical Space platform, making them available for reuse by the content provider or other users according to permissions. In the next iteration, the MPU will be integrated in the released Technical Space platform to complement its data infrastructure and services. Current plans include a single sign-on solution, the execution of some workflows through the Technical Space UI (when there is no need for using the MPU's specialized services such as the visual mapping editor) and the control of the metadata repository functions (upload, transform using stored crosswalks, publish/unpublish) by the Technical Space UI.

WP2 organizes technical workshops and tutorials during plenary meetings to introduce, train and familiarize Pilot/hackathon/demonstrator development teams with the features and usage of the MPU. It also offers technical assistance to Pilot/hackathon/demonstrator development teams, aiming to inform and guide providers in the use of the MPU as well as to set up custom data enrichment workflows as those have been presented in Section 3.7. So far the museum and photography pilot teams have used the MPU for the preparation and remediation of their content, but other are expected to follow over the coming months. The former mapped their metadata to the proprietary data model used by one of the Pilot applications, and added information not originally provided (mainly in the form of constant values that apply to all records) and, the latter used the editor to form and include in the metadata the new URLs for the high quality images that are hosted in the content space and, published and made available the collection used for an exhibition in the Technical Space.

Full documentation of MINT is available online (http://mint-wordpress.image.ntua.gr/mint-end-user-documentation/). It includes a description of the platform and its functionalities, offering step-by-step tutorials and screencasts for all the required tasks. In addition, a tutorial is available for understanding and mapping to the Europeana Data Model (tutorials also available for the LIDO and CARARE schemas) according to its guidelines, accompanied by screencasts and exercises to improve understanding of the system and underlying principles of cultural heritage aggregations. The videos that accompany the documentation can be found on the dedicated YouTube channel[6], also grouped in playlists[7] according to the tutorial sections.

[6] https://www.youtube.com/channel/UCqYT3GhkABWcwbWayGoLSRA

[7] https://www.youtube.com/channel/UCqYT3GhkABWcwbWayGoLSRA/playlists