Appendix A: Inventory of enrichment efforts and tools initiated in the context of the Europeana Network

29/10/2015

Project Name	Type of enrichments	Tool for manual enrichment/annotation	Tool for automatic enrichment	Enrichment target created by the project
Europeana-related projects				
Europeana Operational Ingestion Process	**Simple string matching** Europeana currently applies semantic enrichment for enriching metadata properties that may reference contextual entities in EDM (e.g. Places, Agents, Time Spans, and Concepts). More information can be found at the main reference document for the enrichment framework, accessible at this location: https://docs.google.com/document/d/1JvjrWMTpMIH7WnuieNqcT0zpJAXUPo6x4uMBj1pEx0Y			

Europeana Photography(http://www.europeana-photography.eu/)	Manual enrichment: terms from their source terminologies were mapped to the terms from the Europeana Photography vocabulary. The thesaurus terms were mapped to specific targets schema elements.			The project created a multilingual thesaurus on photography. There are three concept schemes: keywords, photographic type/practices and photographic technique:http://bib.arts.kuleuven.be/photoVocabulary/en.html
PartagePlus (http://www.partage-plus.eu/)	Manual enrichment: terms from the thesaurus were mapped at the same time as mapping their data from their original format to a target schema. The thesaurus terms were mapped to specific targets schema elements.			The project developed a vocabulary about Art Nouveau comprising five concept schemes: object types, styles and periods, actors, materials and activities: http://partage.vocnet.org/
EuropeanaFashion (http://www.europeanafashion.eu/portal/home.html)	Manual enrichment: terms from the thesaurus were mapped at the same time as mapping their data from their original format to a target schema. The thesaurus terms were mapped to specific targets schema elements.	.		The project developed a vocabulary for fashion comprising five concept schemes type, material, colour, technique and subject: http://thesaurus.europeanafashion.eu/
Outcomes of the **Europeana Creative** (http://pro.europeana.eu/europeana-creative/) were reused as part of the **Europeana Sounds** (http://www.europeanasounds.eu/) project	**Manual enrichment:** terms from the thesaurus were mapped at the same time as mapping their data from their original format to a target schema. The thesaurus terms were mapped to specific targets schema elements. **Objects to objects linking using**	The Europeana Creative/Sounds Annotations API extends the User Generated Content service developed within the scope of Europeana Creative project. This implements the basic functionality for managing web annotations related to items in the Europeana metadata aggregation. The current implementation supports creation of annotations on		The Europeana Sounds project developed a vocabulary for music genres: http://data.europeana.eu/concept/soundgenres

	specific schema elements Semantic tagging	Europeana objects or images using simple text comments, tags or semantic tags. In Europeana Sounds more annotation and body target types will be supported using Pundit . The annotations are based on the Open Annotation data model. http://www.slideshare.net/gsergiu79/europeana-creative-annotation-api https://europeanadev.assembla.com/spaces/europeana-creative/wiki/Annotation_Service_and_Architecture		
Europeana Food&Drink (http://foodanddrinkeurope.eu/)develops a vocabulary on Food and Drinks. A first overview is available at http://www.slideshare.net/valexiev1/europeana-food-and-drink-classification-scheme	Manual enrichment: terms from the thesaurus were mapped at the same time as mapping their data from their original format to a target schema. The thesaurus terms were mapped to specific targets schema elements. Creation of semantic equivalences (co-referencing)			
Eagle (http://www.eagle-network.eu/)	Manual enrichment: terms from the thesaurus were mapped at the same time as mapping their data from their original format to a target schema.			The project developed at vocabulary for epigraphy available at http://www.eagle-network.eu/resources/vocabularies/

	The thesaurus terms were mapped to specific targets schema elements.			
MIMO(http://www.mimo-international.com/MIMO/)**:** aggregated LIDO metadata records from 20 museums in 7 languages	Strings are matched to URIs from the MIMO vocabulary multilingual variants			The project created a vocabulary describing music instruments and persons at http://www.mimo-db.eu/InstrumentsMakers/ and http://data.mimo-db.eu/
ASSETS	Linking between objects			
Europeana Cloud(http://pro.europeana.eu/europeana-cloud)	The project is working on the improvement of the current Europeana enrichment framework and will Implement different models for connecting content based on semantic similarity of full text or metadata;			
PATHS http://www.paths-project.eu/): (the ultimate goal of the PATHS project was to develop new ways to explore Cultural Heritage information. The project analyzed techniques for automatically enrich Europeana items with semantic information.	Schema elements relationships.	The project enriches items with a) links to relevant vocabularies, b) links to related Wikipedia pages and c) links among similar items. Being a research project, there was no actual enrichment and thus no enrichment workflow. Instead, the project gathered a sample of Europeana items (3 million items), enriched them, and evaluated the results.	The PATHS project proposed to use Web Annotation by making a profile /extension to the model to represent additional information such as confidence and the target vocabulary being used. http://www.paths-project.eu/eng/content/download/5342/40580/version/6/file/edmPATHS.pdf Another approach comes from the Apache Stanbol called the Fusepool Annotation Model which also designed an extension to	

			Web Annotation to easily represent different usage scenarios such as language annotation, entity mention / linking, topic classification and NIF integration. https://github.com/fusepoolP3/overall-architecture/blob/master/wp3/fp-anno-model/fp-anno-model.md	
LoCloud (http://www.locloud.eu/) is supporting small and medium-sized institutions in making their content and metadata available to Europeana, by exploring the potential of cloud computing Technologies.	**Simple string matching** In the MORe enrichment API: Users can create simple rules in the form: if subject contains X or keyword contains Y then add subject collection K. The subject collections are collections of subject terms that refer to a specific theme and are created out of 29 standard SKOS vocabularies that are available in MORe.		**LoCloud APIs** MORe enrichment API (as part of the MORe API) allows to apply an enrichment plan on a packaged dataset. LoCloud microservices APIs (these are per service and they can be found here) Semi manual enrichment as part of the LoCloud geocoding services *Background link service*: analyzes the textual elements of Cultural Heritage items (title, description, subject, etc), and links the CH item to related pages in DBpedia. The service uses Dbpedia Spotlight v0.7 as a backbone (statistical back-end), with Spanish and English models. The service can be used through it API, as described at http://support.locloud.eu/Metadata%20enrichment%20API%20technical%20documentation

Vocabulary matching service: matches entities from text based | |

			on SKOS vocabularies. The service performs simple lookup into a list of vocabularies as provided by the user. The API is described **at** http://support.locloud.eu/Metadata%20enrichment%20API%20technical%20documentation	

ARROW (http://www.arrow-net.eu/) intellectual property rights infrastructure			The European Library performs a high precision matching of authors against the VIAF dataset. The European Library performs named entity recognition in metadata. In this context, it is only applied to the recognition of person names from statements of responsibility. Look at Supporting Rights Clearance for Digitisation Projects with the ARROW Service.and Use of Authorities Open Data in the ARROW Rights Infrastructure	
The European Library(http://www.theeuropeanlibrary.org/tel4/)	The European Library performs enrichment on: **Names** **Geographic places:** The European Library is currently performing linking and enrichment of place names and geographic coordinates. It covers mainly places of publication and places as subject of the digital objects. **Subjects headings:** The European Library exploits the results of the project Multilingual Access to Subjects (MACS)		Within its aggregation infrastructure, The European Library is currently performing linking and enrichment of place names and geographic coordinates. It covers mainly places of publication and places as subject of the digital objects. The linking and enrichment is done against the Geonames ontology. The method applied is not publicly documented, but is is based in previous work on the Geoparser developed in Europeana Connect	
AthenaPlus (http://www.athenaplus.eu/)	Manual enrichment: terms from the source terminologies were mapped to the terms from the AthenaPlus vocabularies.	The enrichment is performed using Terminology Management Platform (http://www.culture-terminology.org/) for vocabulary		

		creation and MINT for metadata enrichment		
Europeana 1914-1918 (http://www.europeana1914-1918.eu/en)	Subject data was aligned with a subset of LCSH subjects related to WW1.			A small SKOS vocabulary related to First World War was created
DM2E (http://dm2e.eu/)	DM2E has systematically aligned entities from its providers to DBpedia, using Silk.		The project used the Silk Framework to perform the alignment.	
Europeana Connect (http://pro.europeana.eu/project/europeanaconnect)		The project developed Annotorius which is an image annotation tool for the web.	The project used the AMALGAME Matching Tool Amalgame (AMsterdam ALignment GenerAtion MEtatool) to find, evaluate and manage vocabulary alignments. ● Workflow Automation ● Semantic data Layer First Operational Version Fuzzy Alignment Tool The Fuzzy Alignment Tool is used for computing the strength of each alignment. ● Europeana and linked cultural heritage data by Antoine Isaac, VU University ICTdelta, Rotterdam, the Netherlands, 18 March 2010 Tools for multilingual mapping Multilingual mapping of controlled	Summary of the vocabularies investigated in the project Europeana Connect Semantic Layer Methodology for semi-automatic matching Intermediary results of the efforts to (semi-)automatically derive alignments between Europeana vocabularies Recommendation on how to obtain SKOS vocabularies by content providers Semantic Data Layer for semantic processing available (final) Semantic Data Layer extended to external semantic resources Mapped vocabularies and metadata schemes available Multilingual mapping of schemes and vocabularies

			vocabularies has the purpose of providing multilingual search and browsing capabilities in the Europeana portal - Multilingual Mapping of Controlled Vocabularies - Languages and Vocabularies for Selection - EuroVoc Conference 2010: Multilingual Access to Online Content - The Europeana Experience - Multilingual Mapping of Schemes and Vocabularies - Final Report	
EUScreen (http://www.euscreen.eu/): has published its aggregated metadata as LOD. EUScreenXL: A Linked Open Data Pilot - similar to one of the EUscreen project - will be implemented.			A brief overview of the pilot is given here while a more detailed description of the overall workflow and enrichment can be found at Linking Europe's Television Heritage	
Europeana Film Gateway (EFG) (http://www.europeanfilmgateway.eu/)		EFG metadata editor It allows for editing a pre-defined set of data fields of an individual record http://labs.europeana.eu/apps/efg-		

		metadata-editor		
Europeana Inside (http://www.europeana-inside.eu/home/index.html)		Europeana Inside didn't do enrichment itself, but investigated the possibility to re-ingest the enrichment made by Europeana using the Europeana API. Part of that workflow was that Data providers could view and select the kind of enrichment's they found useful for ingesting back into their own collection management system (or a layer above).		
PACKED In 2013 - 2014 the project worked about persistent identification of artworks in 7 Flemish museums and 3 art-organisation. manageable and searchable.	The project identified contextual data about artworks such as creator (via VIAF, RKDartist, ODIS and Wikidata) and object names (AAT). We also enriched these entities with data from external authorities (different name forms, birth and death dates of creators etc). The goal of this project was to see if persistent identification of artworks and identification and enrichment of contextual data via external authorities will help cluster different information on the same artworks when aggregating them and thus make the data better			
Projects formally outside the Europeana family				
Pelagios (http://pelagios.org)	Contextualisation of objects with entities of the same type	The workflow differs by content type, but in general is NER (on		

The Pelagios 3 project is enriching text and map image objects with gazetteer URIs.		plaintext), and automatic gazetteer matching step (for NER'd text, CSV and image transcriptions), and user interfaces for verification and correction. The workflow is supported with an Open Source Annotation tool developed in the project Recogito. It allows users to annotate places in maps and other geospatial documents.		
DCERT (http://www.seco.tkk.fi/projects/dcert/) (Dynamic Configurable Entity Recognition from Text) is a project at the Semantic Computing Research Group at Aalto University. The SeCo group also has a long history of creating and experimenting with various enrichment workflows in projects such as MuseumFinland, CultureSampo, SmartMuseum and BookSampo)				
FREME (http://www.freme-project.eu/)(Open Framework of E-	e-Link based on NLP and Linked Open Vocabularies e-Entity based on Named Entity Recognition and linked entity	Terminology API: "Terminology Annotation Showcase", using the standard Internationalisation Tag Set 2.0 (ITS 2.0) recommended		

Services for Multilingual and Semantic Enrichment of Digital Content)	datasets in LOD e-Terminology for terminology management and terminology annotation, based on cloud terminology services Tilde Terminology e-Translation, based on machine translation, in the cloud e-Publishing, using authoring environment and standardized publication formats	by W3C. Tilde Terminology API can be found at http://term.tilde.com/ and the description of the API at http://term.tilde.com/Content/api_spec.pdf		