

Elevating Natural History Museums' Cultural Collections to the Linked Data Cloud

Giannis Skevakis, Konstantinos Makris, Polyxeni Arapi,
Stavros Christodoulakis

**Work carried out in the scope of
the **Natural Europe Project***



Technical University of Crete
Greece



Outline

- Motivation
- Natural Europe Architecture
- Methodology
- Natural Europe Semantic Layer
- Vocabularies
- The Natural Europe Ontology
- Transition to Europeana Data Model (EDM)
- Future Work

Motivation (1/3)

- Cultural heritage and biodiversity data are produced in a distributed, open fashion
- Data are:
 - syntactically and semantically heterogeneous
 - multilingual
- High quality content remains unexploited due to:
 - lack of interconnection and interoperability between the management systems of NHMs
 - lack of centralized access through a single point of reference

Motivation (2/3)

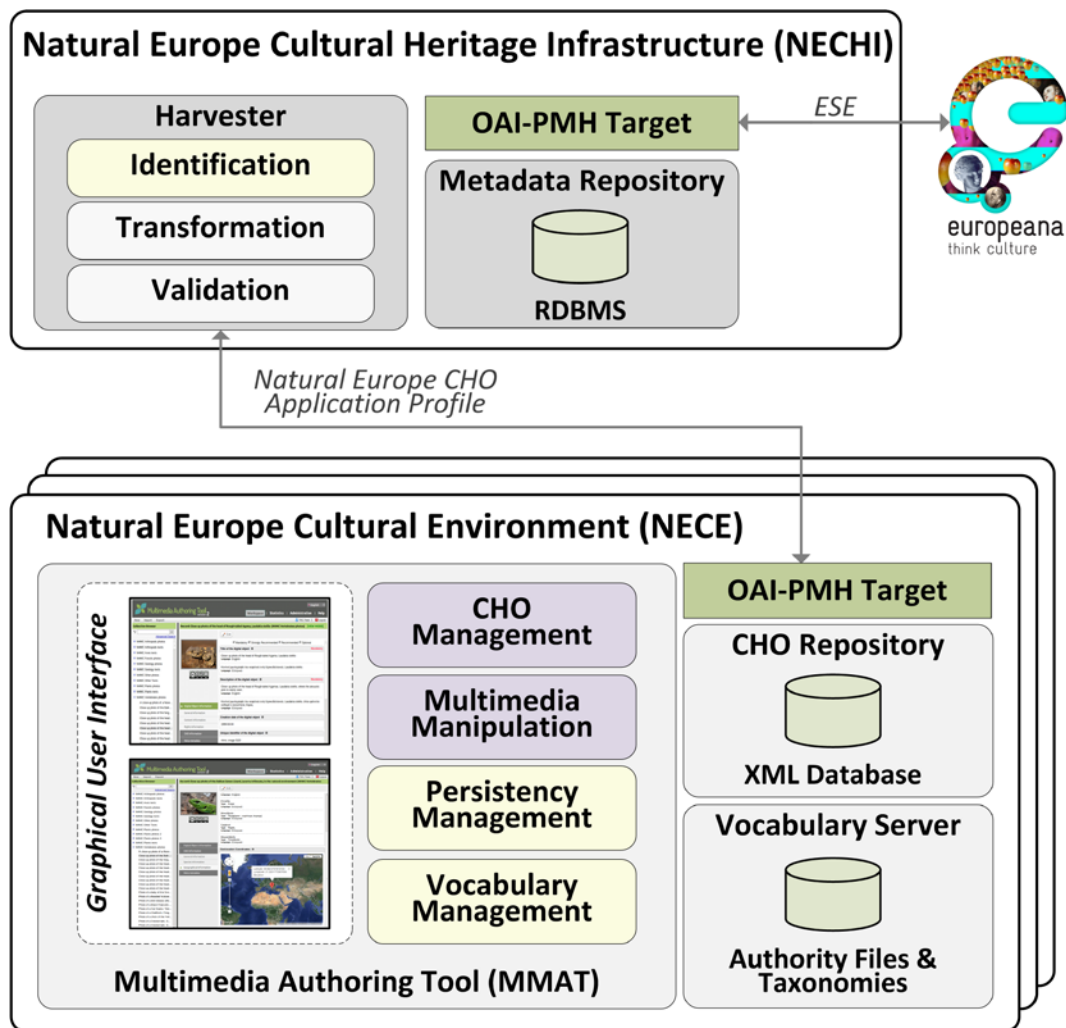
- Natural Europe
 - aims to improve the availability and relevance of natural history heritage content
- Federation of European Natural History Digital Libraries
- Tools and services that allow NHMs to:
 - **publish**, uniformly **describe** and **semantically annotate** their content
 - **interconnect** their digital libraries
 - **expose** metadata records to Europeana.eu and BioCASE



Motivation (3/3)

- Semantic Approach
 - exploitation of data by semantic web applications
 - expose the content to linked data communities
- The Semantic Web standards provide a basis on which interoperable systems can be built
 - RDF(S), SKOS, SPARQL, and OWL
 - published datasets make the reusability of existing data possible

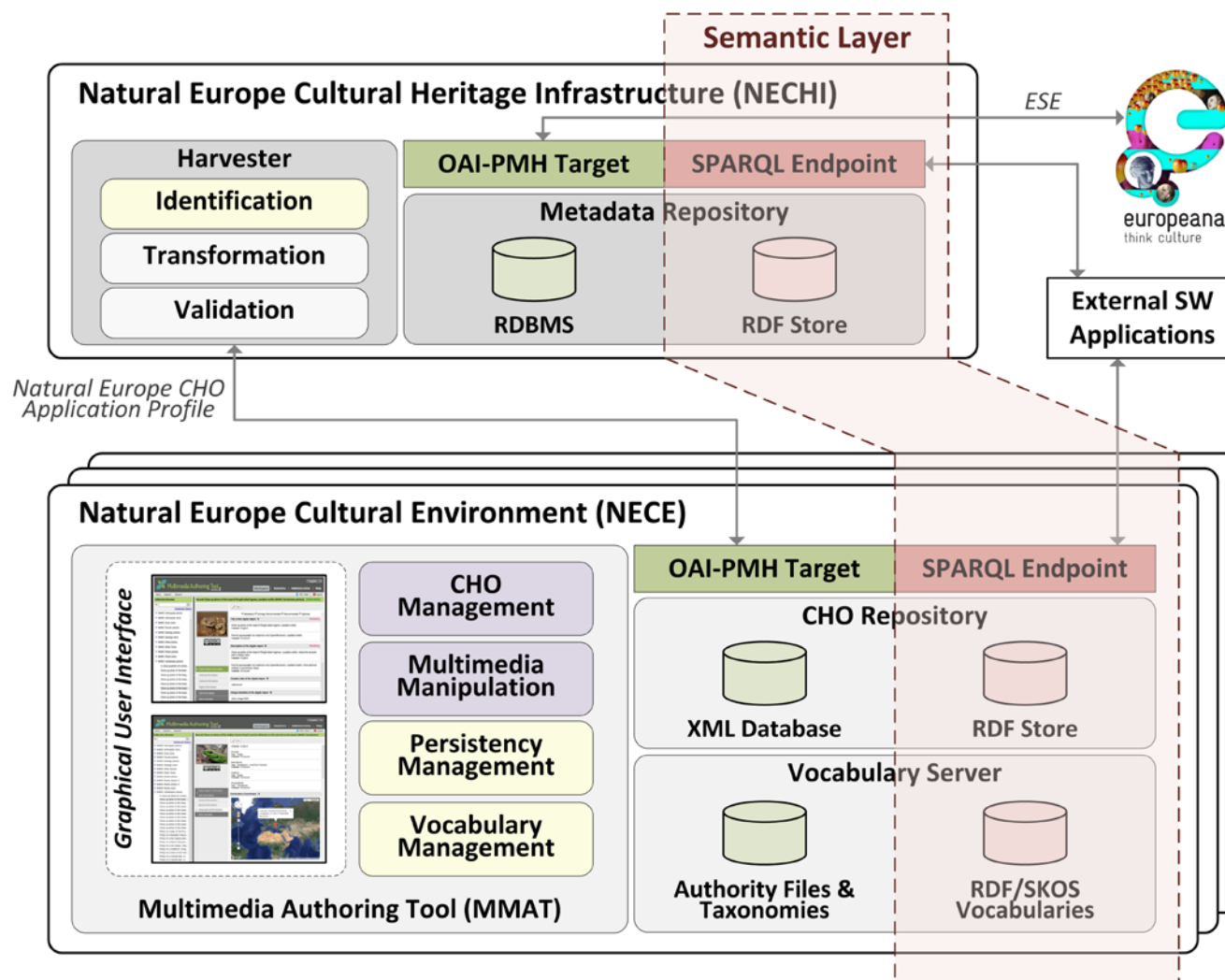
Natural Europe Architecture



Methodology

- Create the [Natural Europe Ontology](#)
 - introduce semantics
- [Enrichment](#) of the Natural Europe metadata
- [Convert](#) XML to RDF
- [Publish](#) data to the Linked Data cloud
 - SPARQL endpoint
 - Resolvable URIs

The Natural Europe Semantic Layer



The Natural Europe Ontology

■ Natural Europe CHO Application Profile (XML)

- *Cultural Heritage Object (CHO) information*
- *Digital Object information*
- *Meta-metadata information*
- *Collection information*

■ Natural Europe Ontology (OWL-DL)

- exploiting the class and property axioms, enabling the inference of new knowledge out of the existing data
- *CHO, CHO collection, specimen, species, observation, multimedia object, person, and organization, etc.*
- Object aggregations

Vocabularies

▣ GeoNames

- ▣ Geographic information
- ▣ 10 million geographical names
- ▣ Mapped to several other well-known ontologies, like linkedgeodata.org

▣ Dbpedia

- ▣ Knowledge base describing over 3.6 million things (persons, places, species)
- ▣ Extracted mainly from Wikipedia

▣ Catalogue of Life (CoL)

- ▣ Comprehensive catalogue of all known species

▣ Uniprot

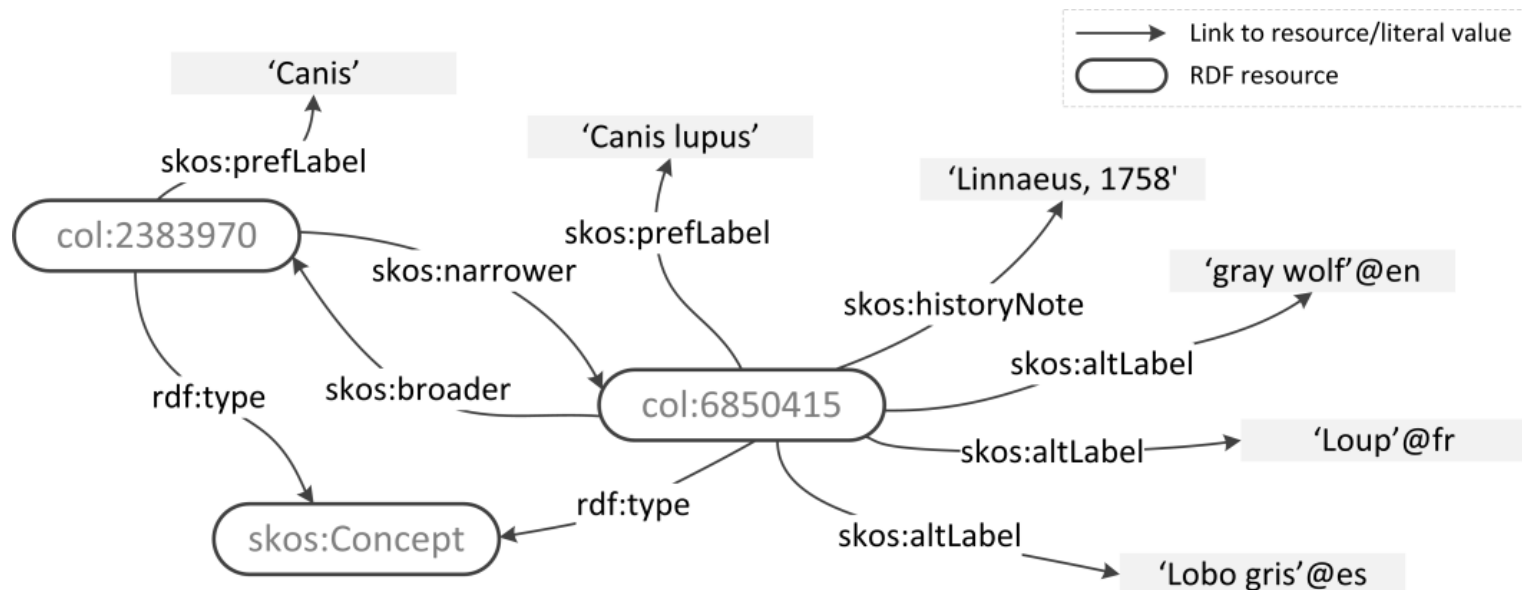
- ▣ Protein sequence, species taxonomy, literature citations and keywords

▣ GEMET

- ▣ Multilingual thesaurus

SKOSification of Catalogue of Life

- Annual checklist, available on the Catalogue Of Life web site.
 - using D2R Server
- Mappings between SQL and RDF



Metadata Enrichment

- Data in Relational/XML databases is rarely connected to external data, due to:
 - structure of the storage
 - most of these have been created long before the introduction of Open Data

- Linked to external vocabularies/thesaurus by exploiting their services
 - Spatial Information → Geonames
 - Species Information → COL / Uniprot
 - General Information → DBpedia
 - Keywords → GEMET

Example Natural Europe Record

```
<record xmlns="http://www.natural-europe.eu/nhm/aip/">
  <objectUri>http://nhmc.natural-europe.eu/12dda2d5</objectUri>
  <contextUri>http://www.nhmc.uoc.gr/museum/40319</contextUri>
  <contentType>http://purl.org/dc/dcmitype/Image</contentType>
  <scientificName xml:lang="la">Canis lupus</scientificName>
  <classification xml:lang="la" annotation="FAMILIA">Canidae</classification>
  <classification xml:lang="la" annotation="ORDO">Carnivora</classification>
  <commonName xml:lang="el">Λύκος</commonName>
  <commonName xml:lang="en">Wolf</commonName>
  <title xml:lang="en">Wolf, Canis lupus</title>
  <title xml:lang="el">Λύκος, Canis lupus</title>
  <creator xml:lang="en" annotation="Photographer">Trichas, A.</creator>
  <creator xml:lang="el" annotation="Φωτογράφος">Τριχάς, Α.</creator>
  <subject xml:lang="en">Mammals</subject>
  <subject xml:lang="el">Θηλαστικά</subject>
  <description xml:lang="en">Photo of wolves in forest diorama in the Paranesti NHM.</description>
  <contributor xml:lang="en" annotation="Curator">Lymberakis P.</contributor>
  <contributor xml:lang="el" annotation="Εφορος">Λυμπεράκης, Π.</contributor>
  <identifier>nhmc.image.40319</identifier>
  <spatial xml:lang="en">Greece</spatial>
  <spatial xml:lang="el">Ελλάδα</spatial>
  <geolocation latitude="35.296227084320144" longitude="23.91901402771254"/>
</record>
```

Enriched Natural Europe Record

```
<record xmlns="http://www.natural-europe.eu/nhm/aip/">  
  <objectUri>http://nhmc.natural-europe.eu/12dda2d5</objectUri>  
  <contextUri>http://www.nhmc.uoc.gr/museum/40319</contextUri>  
  <contentType>http://purl.org/dc/dcmitype/Image</contentType>  
  <licenseUri>http://creativecommons.org/licenses/by-nc-nd/3.0</licenseUri>  
  <scientificName>http://www.catalogueoflife.org/col/6850415</scientificName>  
  <commonName xml:lang="el">Λύκος</commonName>  
  <commonName xml:lang="en">Wolf</commonName>  
  <title xml:lang="en">Wolf, Canis lupus</title>  
  <title xml:lang="el">Λύκος, Canis lupus</title>  
  <creator>http://nhmc.natural-europe.eu/persons/158ggse7</creator>  
  <subject>http://www.eionet.europa.eu/gemet/concept/4982</subject>  
  <description xml:lang="en">Photo of wolves in forest diorama in the Paranesti NHM.</description>  
  <contributor>http://nhmc.natural-europe.eu/persons/1dg5hhd7</contributor>  
  <type xml:lang="en">Preserved specimen</type>  
  <identifier>nhmc.image.40319</identifier>  
  <alternative xml:lang="en">Photo of Canis lupus</alternative>  
  <spatial>http://www.geonames.org/390903</spatial>  
  <geolocation latitude="35.296227084320144" longitude="23.91901402771254"/>  
  <relation>http://live.dbpedia.org/page/Gray_wolf</relation>  
  <relation>http://live.dbpedia.org/page/Carl_Linnaeus</relation>  
</record>
```

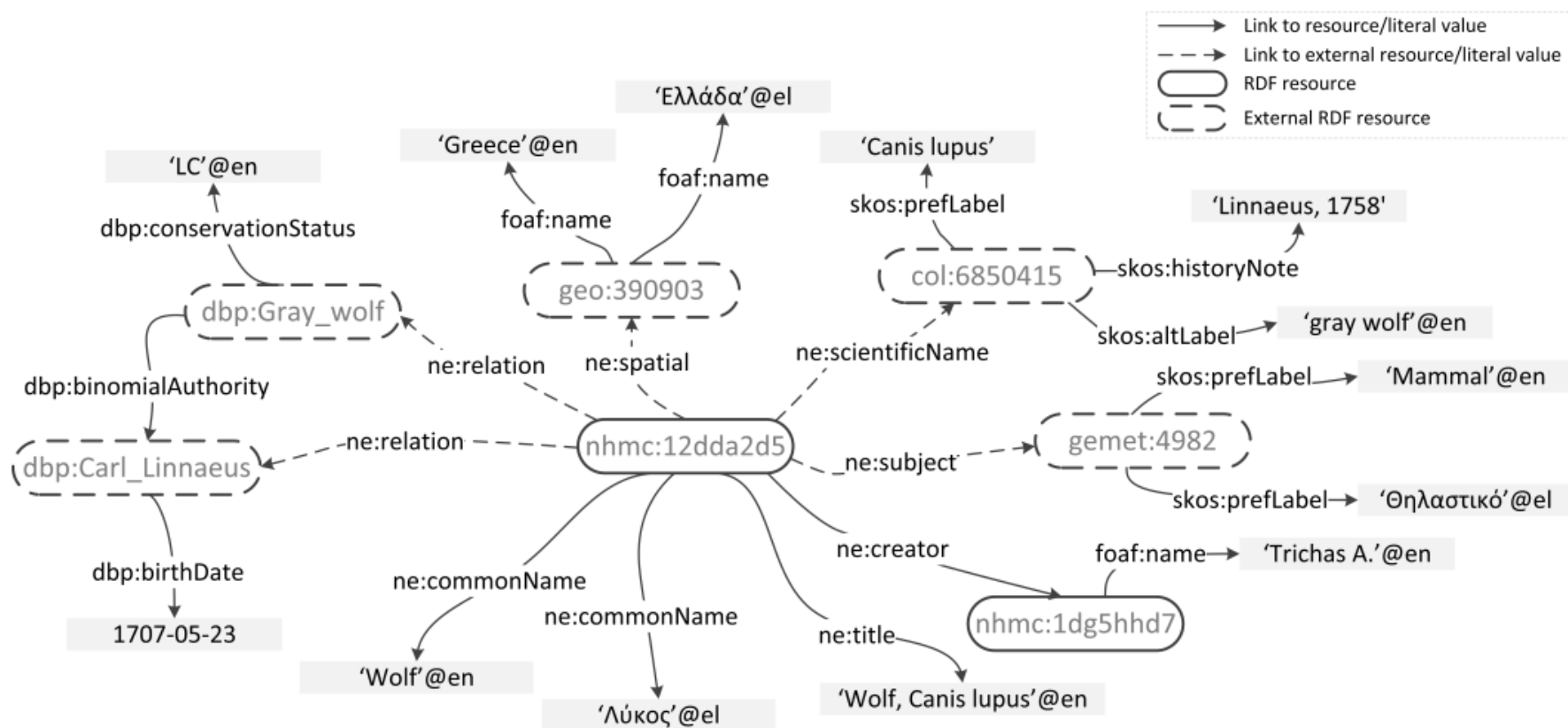
semantic linkage

semantic enrichment

Conversion of Metadata to RDF

- XML to RDF data conversion has been performed through automatic transformation processes
 - Intermediate format
- Identification module providing unique identifiers
- The final RDF data have been persisted in an RDF store
 - SPARQL endpoint
 - Browse interface
 - Faceted search

Natural Europe RDF example



Connection to the Linked Data Cloud

Example: Consider two museums that have described a specimen of a gray wolf (*canis lupus*)

- CHOs are connected to the SKOS *Concept* “*Canis lupus*”
- At least two resources of the class *Specimen* are linked to “*Canis lupus*”
- We can utilize the relation between these two specimens using:
 - the SPARQL endpoint in the federal node
 - federated query of SPARQL 1.1 specification

Transition to EDM

- EDM adheres to the modeling principles that underpin the approach of the Web of Data
 - no fixed schema
 - aggregations of similar objects, allowing different schemas to be attached
- Europeana recently provided an EDM ingestion mechanism
 - our approach ensures that the data aggregation will be supported with minimum effort

Future Work

- Support **live update** on the triples after each change on the data
- **Ontology-based mediator system**
 - integration of the Natural Europe federated nodes with cultural heritage and RDF data providers, using different metadata schemas
 - mediator schema
 - retrieval of up-to date triples
 - SPARQL-RW Framework, accessing federated RDF data sources complying to different Ontology Schemas

Thank you!