The ARIADNE project

Advanced Research Infrastructure for Archaeological Data Networking in Europe

Franco Niccolucci – PIN
Project Coordinator

Facing the future, Berlin 21-22 November 2013

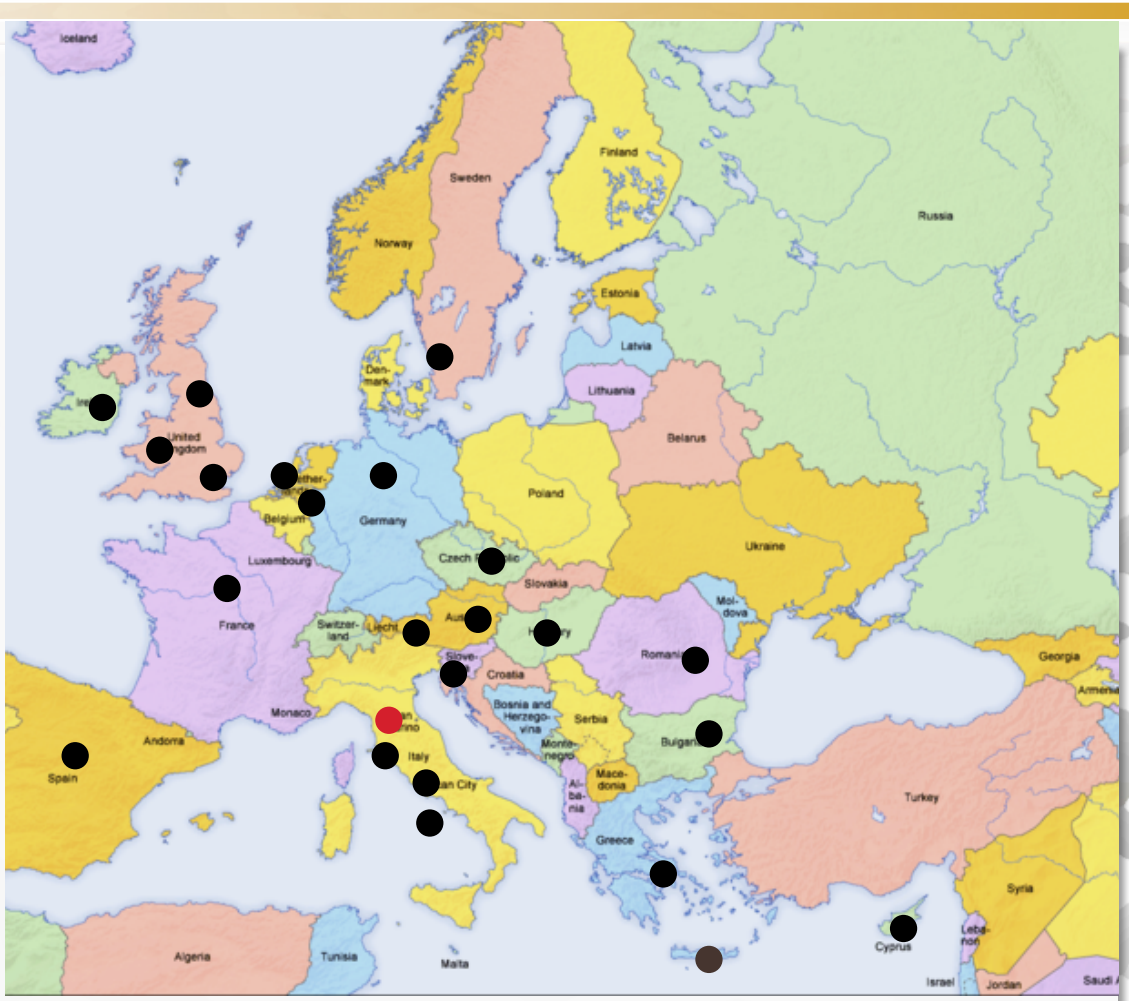
ARIADNE is funded by the European Commission's Seventh Framework Programme

What is ARIADNE

- ARIADNE is a EU-funded project aimed at integrating the European Research Infrastructures on archaeological datasets
- Its overall goal is to overcome the fragmentation of archaeological data repositories and to foster a culture of archaeological data sharing and re-using
- Start date: 1st February 2013
- www.ariadne-infrastructure.eu

The ARIADNE partnership

- Coordination
  - PIN
  - UoY-ADS
- 24 partners
- 17 countries



ARIADNE focus

- ARIADNE is focused on archaeology and, in general, on tangible heritage
- Its disciplinary scope has very specific methods and needs for integration: it deals with 'things'
- There are common interests with digital humanities (texts) and history, so we are collaborating with DARIAH on content, methods and tools
- Very important overlap with conservation and restoration, so we are establishing strong collaborations with projects in this domain

Project activities

- Networking activities
  - Community building: involving researchers and institutions in sharing data and establishing together guidelines
  - Standardization and good practices
- Trans-National Access
  - Training
  - Access to on-line resources
- Research activities
  - Knowledge organization
  - Data management
  - New or improved tools to extract information
  - Advances in methodology

Progress in the initial nine months

<u>Networking</u>

- Set up users' needs survey and SIGs
- Collected detailed information on partners' datasets

<u>TNA</u>

- Outlined the training program for 2014
  - Summer schools

Progress in the initial nine months

Joint Research

- Started the design of integration
  - Mapping metadata schemas to CIDOC-CRM
- Working on an extension of CIDOC-CRM suitable for archaeological documentation
  - Draft proposal for excavation data in preparation
- Creation of the ACDM (ARIADNE Catalog Data Model)

Overview of partners' datasets

- **DBMS**: RDBMS, MySQL, Access
- **Dataset**: repository of digital objects with the same structure
- **Collection**: set of text files/images stored in a hierarchical system
- **Multimedia**: 3D, images, videos
- **Sparse files**: mostly grey literature, no (common) schema
- **GIS**: information layers, coordinates

Content includes:

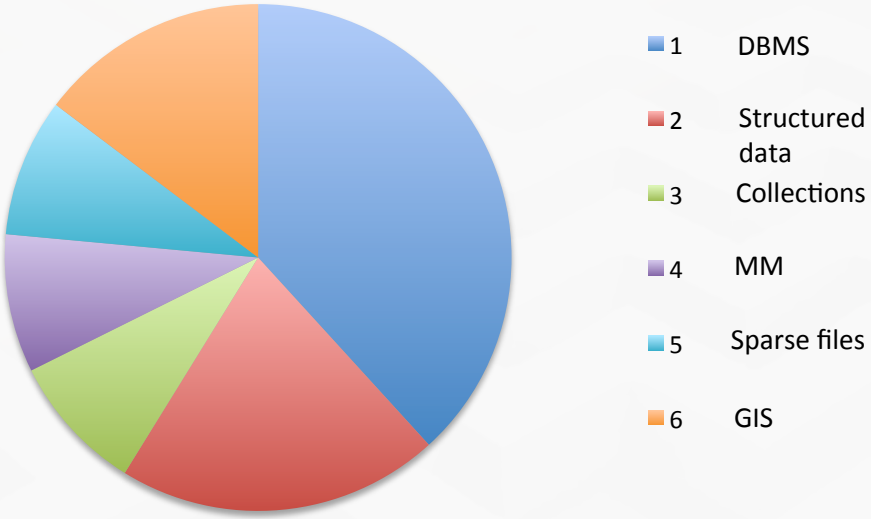
- Excavation data
- Monuments and sites
- Scientific analyses

	DBMS	Structured data	Collections	MM	Sparse files	GIS
ZRC-SAZU	X					X
OEAW	X				X	
DISCOVERY	X	X				X
ARHEO		X				
INRAP	X	X				
ARUP-CAS	X					
NIAM-BAS	X					
SND	X					X
ADS	X	X	X	X	X	X
DANS	X	X	X		X	
MIBAC-ICCU	X	X		X		X
MNM-NOK	X					
ATHENA RTC-CETI	X					
AIAC			X			
Cyi-STARC		X		X		
DAI	X					

ARIADNE in numbers

- 20 countries
- 24 languages
- 1,500,000+ database records
- 40,000+ grey literature files

38% DBMS
20% Structured data (datasets)
9% Collections
9% Multimedia
9% Sparse files
15% GIS



Metadata schemas

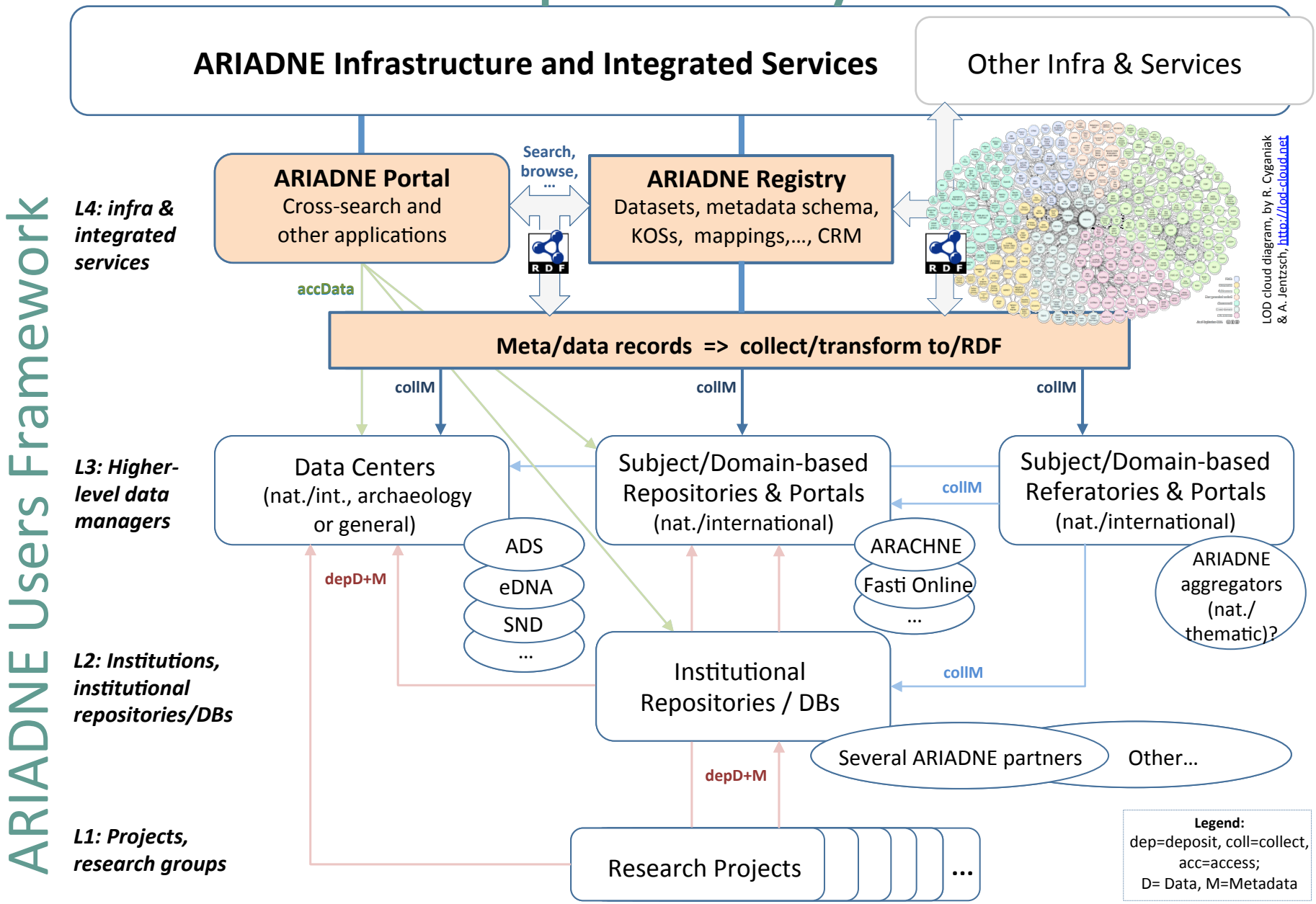
Eight partners (SND, KNAW-DANS, DISCOVERY, MiBAC-ICCU, INRAP, ADS, CYI-STARC) have adopted formal metadata standards for their datasets. The metadata standards reported are:

- DDI, DataCite, MARC/UNIMARC, TriDAS, Dublin Core application profiles, INSPIRE, ISO 11915, CARARE, LIDO, CIDOC-CRM.

Ten partners (ZRC SAZU, MiBAC-ICCU, ADS, AIAC, MNM-NOK, CYI-STARC, ARUP-CAS, ATHENA RC, NIAM-BAS) have developed proprietary metadata schemas for some of their datasets. However, all these can be mapped onto CIDOC-CRM.

Four partners (DISCOVERY, INRAP, ARHEO, OAEW) reported some datasets for which a metadata definition is not currently available but could be derived from the database structure.

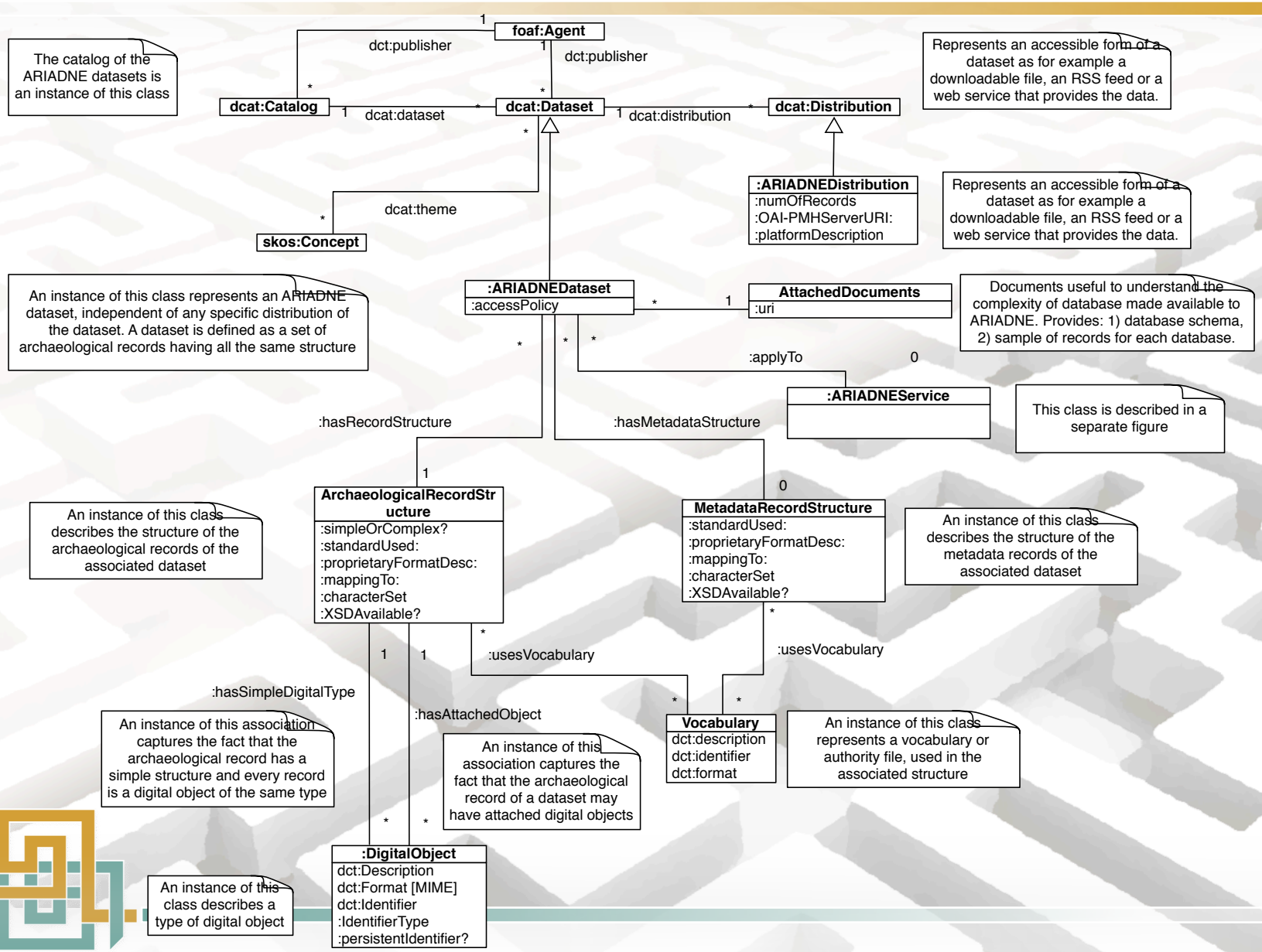
Interoperability Framework



The ACDM

- The ARIADNE Catalog Data Model (ACDM) aims at describing datasets, services, and resources in the archaeological domain
  - So far it addresses databases, collections, thesauri, with plans for covering all relevant resource types
- Based on DCAT (W3C recomm.) and other widespread ontologies
- Data collected into a registry, using a data acquisition tool
- To be used internally to support integration design
- Conceived to became a publicly available service and to be offered/ extended to other domains (e.g. conservation/restoration)

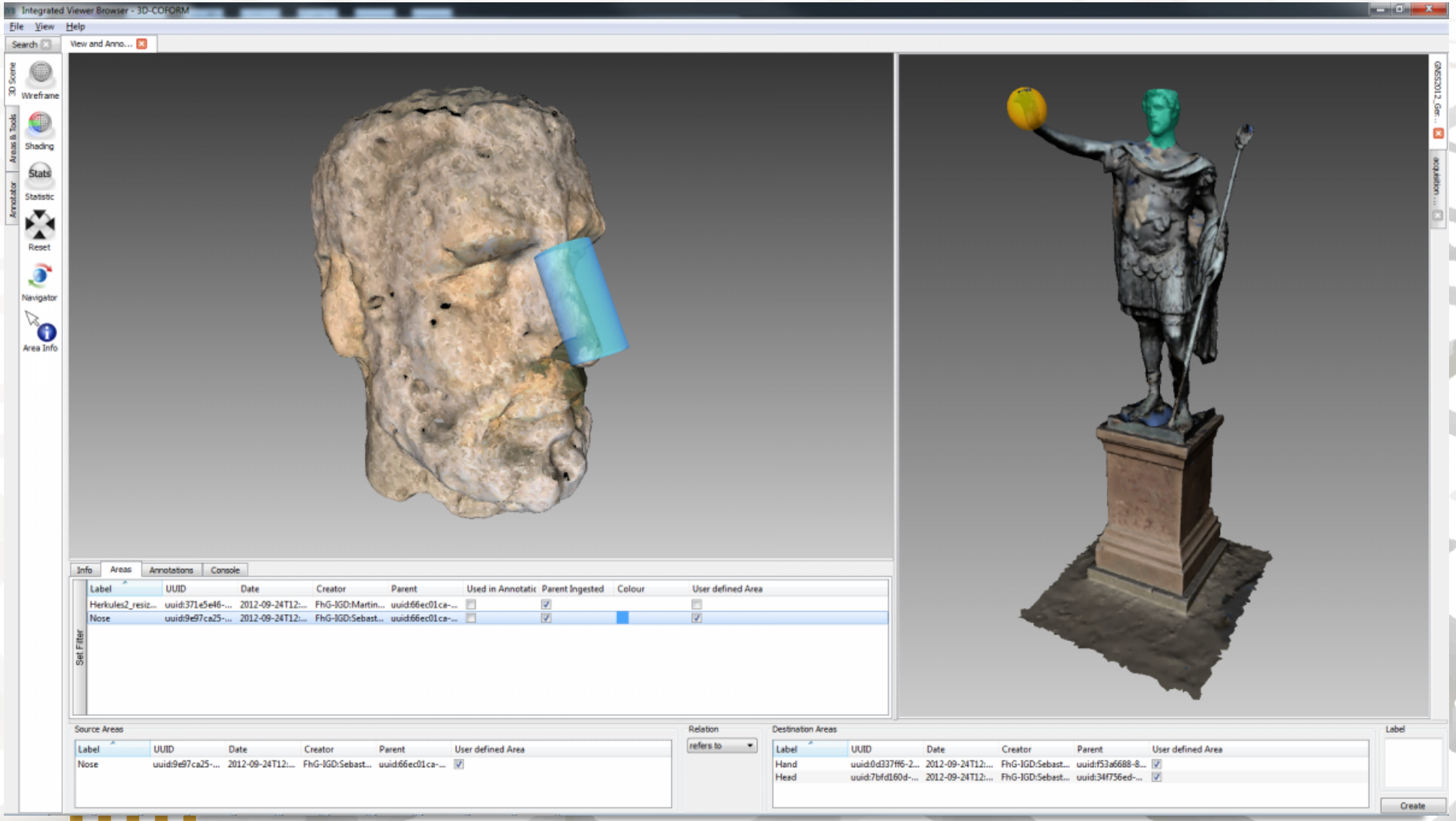
ACDM Model


foaf:Agent
dct:publisher
dct:publisher
The catalog of the ARIADNE datasets is an instance of this class
dcat:Catalog
dcat:dataset
dcat:Dataset
dcat:distribution
dcat:Distribution
Represents an accessible form of a dataset as for example a downloadable file, an RSS feed or a web service that provides the data.
dcat:theme
skos:Concept
:ARIADNEDistribution
:numOfRecords
:OAI-PMHServerURI:
:platformDescription
Represents an accessible form of a dataset as for example a downloadable file, an RSS feed or a web service that provides the data.
An instance of this class represents an ARIADNE dataset, independent of any specific distribution of the dataset. A dataset is defined as a set of archaeological records having all the same structure
:ARIADNEDataset
:accessPolicy
AttachedDocuments
:uri
Documents useful to understand the complexity of database made available to ARIADNE. Provides: 1) database schema, 2) sample of records for each database.
:applyTo
:ARIADNEService
This class is described in a separate figure
:hasRecordStructure
:hasMetadataStructure
An instance of this class describes the structure of the archaeological records of the associated dataset
ArchaeologicalRecordStructure
:simpleOrComplex?
:standardUsed:
:proprietaryFormatDesc:
:mappingTo:
:characterSet
:XSDAvailable?
MetadataRecordStructure
:standardUsed:
:proprietaryFormatDesc:
:mappingTo:
:characterSet
:XSDAvailable?
An instance of this class describes the structure of the metadata records of the associated dataset
:usesVocabulary
:usesVocabulary
:hasSimpleDigitalType
:hasAttachedObject
An instance of this association captures the fact that the archaeological record has a simple structure and every record is a digital object of the same type
An instance of this association captures the fact that the archaeological record of a dataset may have attached digital objects
Vocabulary
dct:description
dct:identifier
dct:format
An instance of this class represents a vocabulary or authority file, used in the associated structure
:DigitalObject
dct:Description
dct:Format [MIME]
dct:Identifier
:IdentifierType
:persistentIdentifier?
An instance of this class describes a type of digital object

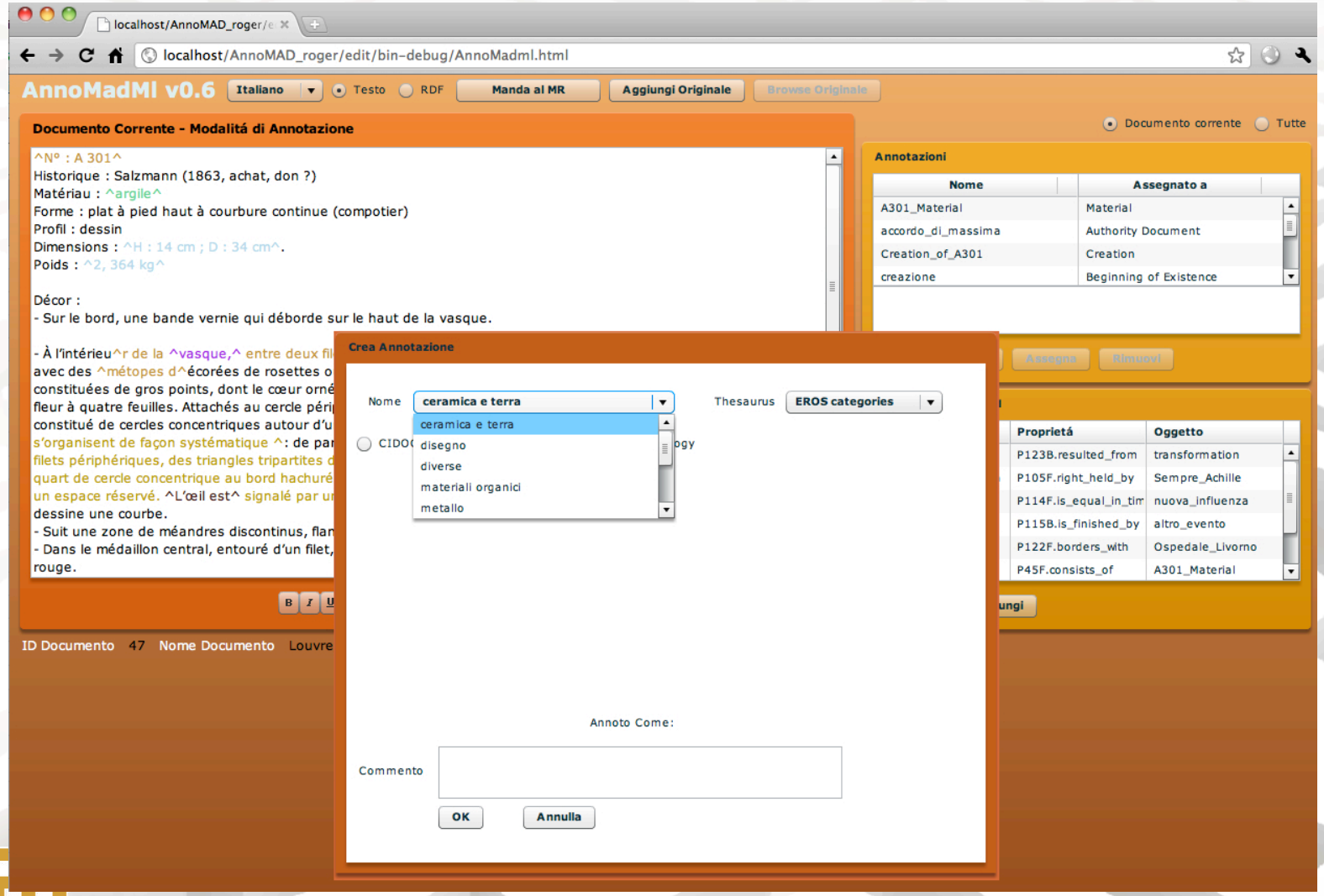
ARIADNE Required Services

- Services for resource discovery
- Services integrating datasets on a geographic base, enabling access according to geographic criteria
- Services integrating datasets on a temporal base, enabling access according to temporal criteria and using a complex timeline
- Services implementing interoperability among different datasets
- Metadata management (semantic repositories with advanced query and retrieval functionalities)
- Repository infrastructures and ingestion services
- Visualization services for
  - Images
  - Videos
  - 3D Single Objects & Large Scenes
- Services will re-use, as far as possible, concepts/tools created in other projects

Integrated Viewer-Browser Tool



Text Annotation Tool



Mapping Tool (PIN, beta)

3D-COFORM Legacy Mapping & Exporting Tool

Legacy Data Schema | Load | Mapped Elements (CIDOC-CRM) | Namespace | Thesauri

mus_part_location_history_data_fitness
mus_part_location_history_data_normal
mus_part_location_history_data_normal
mus_part_location_history_data_normal
mus_part_location_history_data_normal
mus_part_location_history_data_normal
mus_part_location_history_data_note
mus_part_location_history_data_user
mus_part_location_history_date
mus_part_location_history_user
mus_part_location_normal
mus_part_location_normal_free
mus_part_location_normal_th_h
mus_part_location_normal_th_i
mus_part_location_normal_val
mus_part_location_note
mus_part_location_user
mus_part_name
mus_part_name_currency
mus_part_name_currency_th_h
mus_part_name_currency_th_i
mus_part_name_currency_val
mus_part_name_field
mus_part_name_field_th_h
mus_part_name_field_th_i
mus_part_name_field_val
mus_part_name_language
mus_part_name_level
mus_part_name_level_th_h
mus_part_name_level_th_i

mus_obj_images_association_val: E3.Condition_State => P5B.forms_part_of => E3.Condition_State
mus_part_description: E31.Document => P70F.documents => E24.Physical_Man-Made_Thing
mus_part_location_note: E17.Type_Assignment => P41F.classified => E14.Condition_Assessment

CIDOC-CRM Ontology Browser

Use | Choose Entity | depicts * P62F | Element for Thesaurus ...

All Classes | Upper Class

E24.Physical_Man-Made_Thing (Use)
- E22.Man-Made_Object (Use)
- - E84.Information_Carrier (Use)
- E78.Collection (Use)
- E25.Man-Made_Feature (Use)

Selected Property:

P62F.depicts (Use)

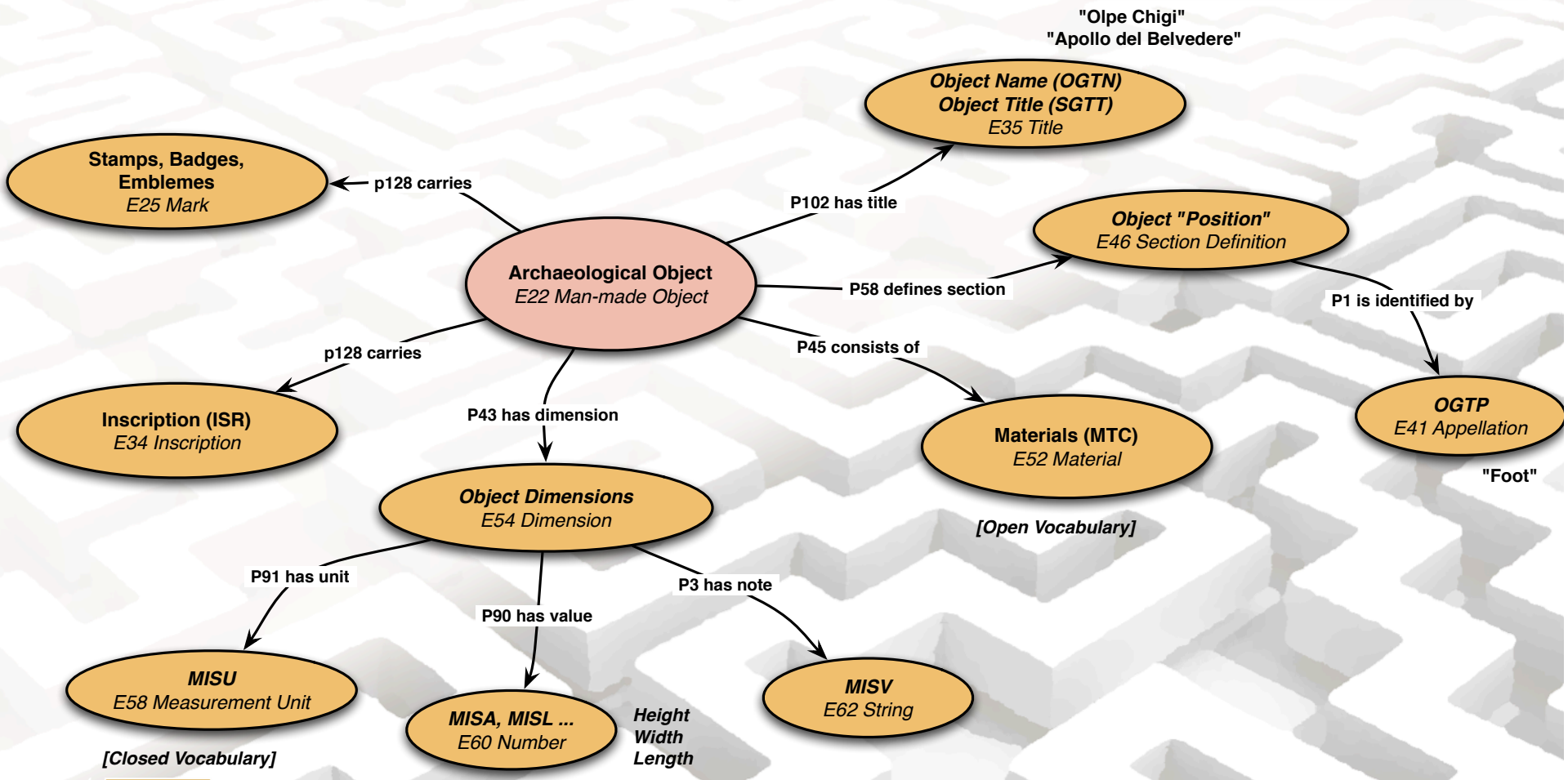
---- E46.Section_Definition (Use)
-- E39.Actor (Use)
--- E21.Person (Use)
--- E74.Group (Use)
---- E40.Legal_Body (Use)
-- E70.Thing (Use)
--- E71.Man-Made_Thing (Use)
---- E24.Physical_Man-Made_Thing (Use)
----- E22.Man-Made_Object (Use)
------ E84.Information_Carrier (Use)
----- E78.Collection (Use)
----- E25.Man-Made_Feature (Use)
---- E28.Conceptual_Object (Use)
----- E73.Information_Object (Use)
------ E33.Linguistic_Object (Use)
------- E34.Inscription (Use)
------- E35.Title (Use)
------ E36.Visual_Item (Use)
------- E37.Mark (Use)
-------- E34.Inscription (Use)
------- E38.Image (Use)
------ E29.Design_or_Procedure (Use)

Mapping Element

mus_obj_images_field => E84.Information_Carrier => P62F.depicts => E22.Man-Made_Object

Extend | Del. Step | Clear | Accept

Mapping


"Olpe Chigi"
"Apollo del Belvedere"
Object Name (OGTN)
Object Title (SGTT)
E35 Title
Stamps, Badges, Emblemes
E25 Mark
p128 carries
P102 has title
Object "Position"
E46 Section Definition
Archaeological Object
E22 Man-made Object
P58 defines section
P1 is identified by
p128 carries
P45 consists of
Inscription (ISR)
E34 Inscription
P43 has dimension
Materials (MTC)
E52 Material
OGTP
E41 Appellation
"Foot"
Object Dimensions
E54 Dimension
[Open Vocabulary]
P91 has unit
P3 has note
P90 has value
MISU
E58 Measurement Unit
MISA, MISL ...
E60 Number
Height
Width
Length
MISV
E62 String
[Closed Vocabulary]

Challenges

- Languages
  - Most data are (and must remain) in their original languages
- Traditions
  - The way data are understood and captured are different according to schools and scientific traditions
- Fragmentation
  - There is a multitude of micro-archives that contain valuable information and are at risk of loss
- Technology
  - Advances in the use of technology (e.g. 3D, scientific analyses) are creating 'big data' archives, not properly managed with the currently available tools
- Virtualization
  - Virtual tools need to be fully integrated in the research methodology of the discipline

ARIADNE is a project funded by the European Commission under the Community's Seventh Framework Programme, contract no. FP7-INFRASTRUCTURES-2012-1-313193.

Contact: niccolucci@unifi.it

www.ariadne-infrastructure.eu

European Commission

SEVENTH FRAMEWORK PROGRAMME