

The X3ML toolkit:

How to map excavation data to CIDOC CRM

Workshop: Semantic mapping of excavation data

15 June 2022

Maria Theodoridou & Vangelis Kritsotakis

ICS-FORTH

What means mapping of one schema to another

A sufficient specification for the **transformation** of each instance of a **source schema** into an instance of a **target schema** while **preserving** as much as possible its initial 'meaning'.

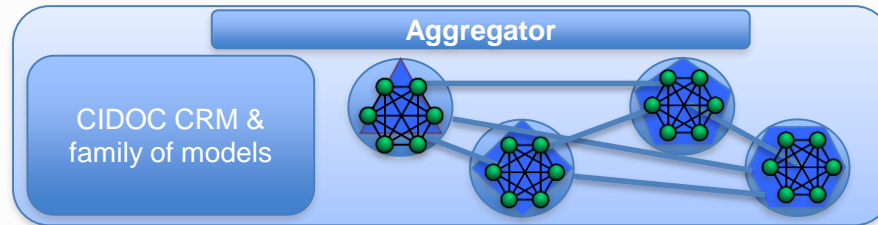
CIDOC CRM Approach: **target schema = CIDOC CRM + family of compatible models**

- interpretation of source schema as semantic model (nodes and links),
- mapping each element of that to an equivalent CIDOC CRM path,
- such that each instance of an element of the source semantic model can be converted into a valid instance of the CIDOC CRM with the same meaning.

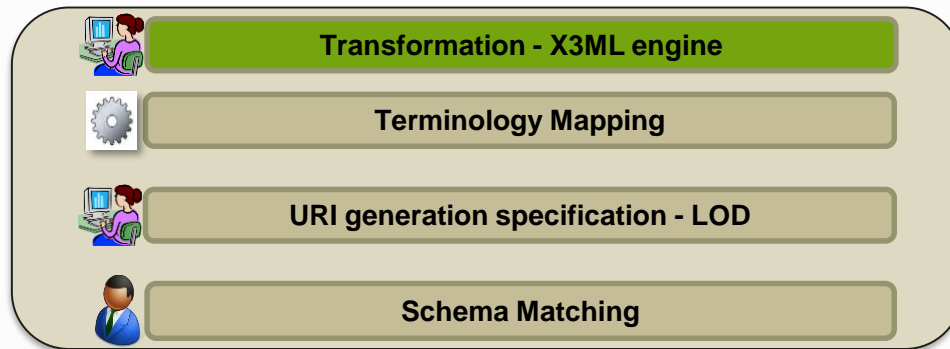


Data Transformation Workflow

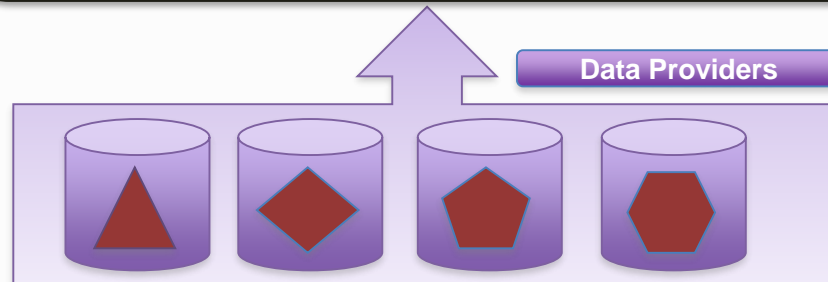
*Semantic Integration
& Interoperability*



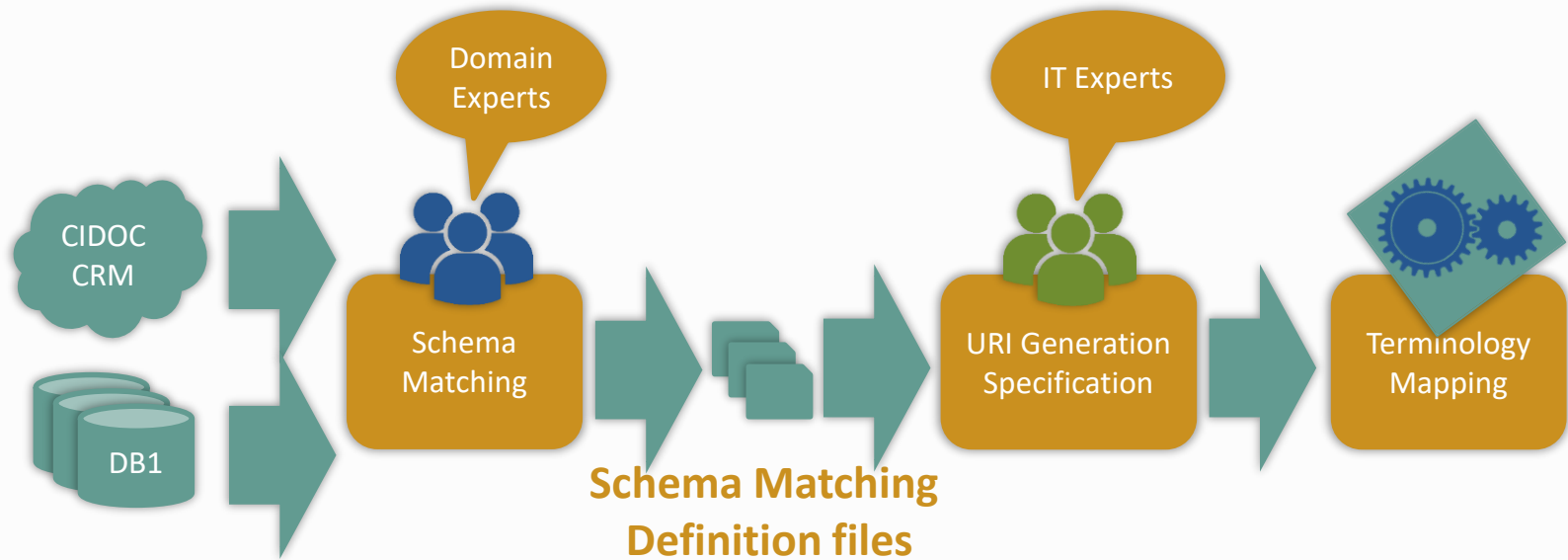
*Data Transformation
Tools*



*Heterogeneous
Data Collections*



Data Transformation Workflow



A set of small, open source, microservices that follow the SYNERGY Reference Model. They are designed with open interfaces and they can be easily customized and adapted to complex environments. The X3ML Toolkit consists of a set of software components that assist the data provisioning process for information integration.

Key components:

- X3ML Mapping Definition Language,
- 3M Mapping Memory Manager,
- X3ML Engine
- RDF Visualiser

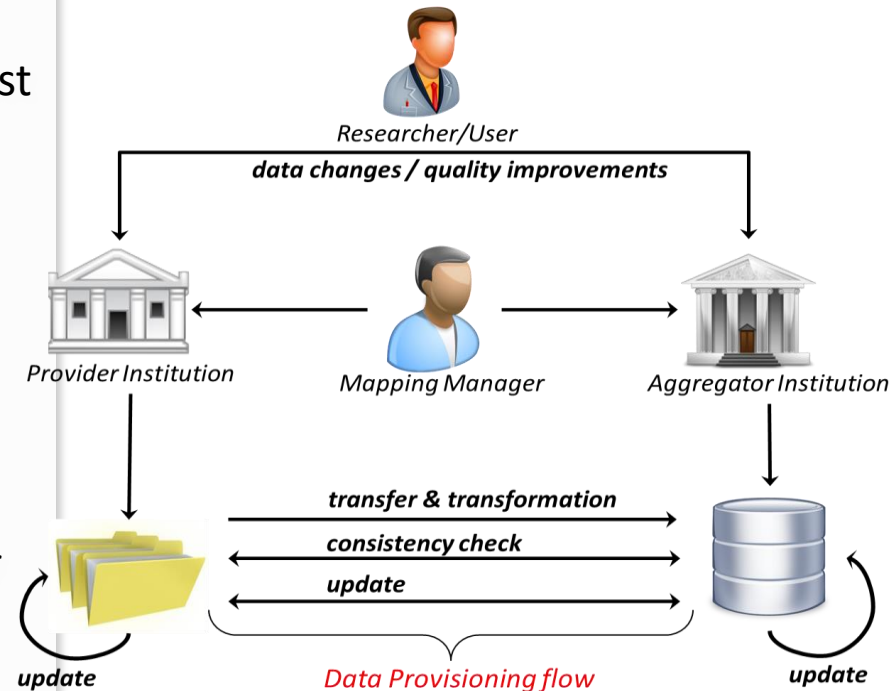
The SYNERGY Reference Model is a rich and comprehensive Reference Model for a better practice of data provisioning and aggregation processes, primarily in the cultural heritage sector, but also for e-science.



Synergy Reference Model

A rich and comprehensive **Reference Model** for a better practice of data provisioning and aggregation processes.

- Initiated from CIDOC CRM Special Interest Group
- Aims at identifying, supporting or managing the process when a provider and an aggregator agree:
 - *To transfer data from the provider to the aggregator*
 - *To transform their format to the homogeneous format of the aggregator*
 - *To curate the semantic consistency of source and target data*
 - *To maintain the transferred data up-to-date with respect to changes on both sides*



X3ML Mapping Definition Language

X3ML is a declarative, XML based language which describes schema mappings in such a way that they can be collaboratively created and discussed by experts.

Emphasis is on establishing a **standardized mapping description** which lends itself to collaboration and the building of a **mapping memory** to accumulate knowledge and experience.

Key Features:

- It provides a declarative way for describing schema mappings
- Decoupled from the URI and values generation process
- Mappings are described using XML serialization



<https://github.com/isl/x3ml/blob/master/docs/x3ml-language.md>



X3ML Mapping Definition Constructs

X3ML supports **1:N mappings** and uses the following special constructs:

- **Intermediate nodes** used to represent the mapping of a simple source path to a complex target path.
- **Additional expression nodes** used to assign additional attributes to an entity.
- **Constant expression nodes** used to assign constant attributes to an entity.
- **Conditional statements** within the target node and target relation support checks for existence and equality of values and can be combined into Boolean expressions.
- **“Same as” variable** used to identify a specific node instance for a given input record that is generated once but is used in a number of locations in the mapping.
- **Join operator (==)** used in the source path to denote relational database joins.
- **Info and comment blocks** throughout the mapping specification bridge the gap between human author and machine executor.



3M: Mapping Memory Manager

3M is a web tool aiming in assisting end-users to easily complete the mapping definition process, by providing the appropriate environment which offers rich and complex functionality hidden under a simple and friendly user interface. Data experts can transform their internal structured data and other associated contextual knowledge to other schemata, including, the CIDOC CRM.

An **online version** of the system can be found here (however credentials should be requested in order to access it):

<https://demos.isl.ics.forth.gr/3m/Projects>

For an **easy deployment on your machine** please use the Docker image:

<https://gitlab.isl.ics.forth.gr/cc/3m-docker>



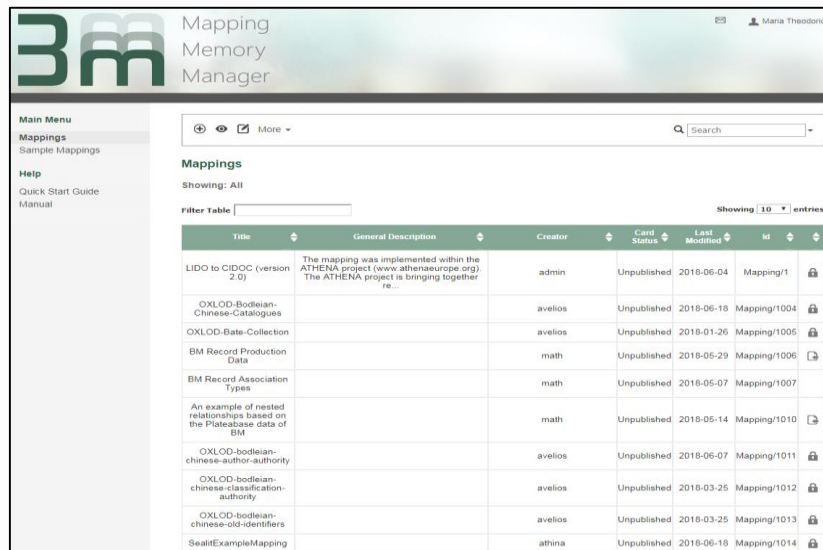
• • •

The screenshot displays the ARADebug Excavation m... interface. At the top, there's a header with the application name and a navigation bar. Below the header, a list of targets is shown, including 'Source: ARADebug Excavation m...', 'Target: ARADebug Excavation m...', and 'Target: ARADebug Excavation m...'. A large blue watermark with the text '1st Release 2021' is overlaid on the right side of the image.

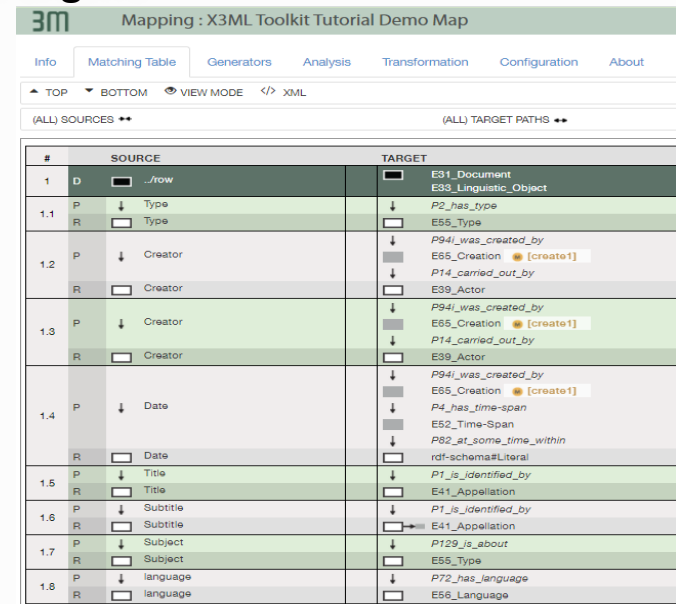


3M Editor

- Enables the creation of mapping definitions (X3ML) between source and target schemata
- Supports guided mappings by analyzing source resources and target schemata
- Provides user space and mapping storage
- Transforms data (in RDF format) using X3ML Engine



Title	General Description	Creator	Card Status	Last Modified	Id
LIDO to CIDOC (version 2.0)	The mapping was implemented within the ATHENA project (www.athenaeurope.org). The ATHENA project is bringing together re...	admin	Unpublished	2018-06-04	Mapping/1
OXLOD-Bodleian-Chinese-Catalogues		avelos	Unpublished	2018-06-18	Mapping/1004
OXLOD-Bate-Collection		avelos	Unpublished	2018-01-26	Mapping/1005
BM Record Production Data		math	Unpublished	2018-05-29	Mapping/1006
BM Record Association Types		math	Unpublished	2018-05-07	Mapping/1007
An example of nested relationships based on the Platebase data of BM		math	Unpublished	2018-05-14	Mapping/1010
OXLOD-bodleian-chinese-author-authority		avelos	Unpublished	2018-06-07	Mapping/1011
OXLOD-bodleian-chinese-classification-authority		avelos	Unpublished	2018-03-25	Mapping/1012
OXLOD-bodleian-chinese-old-identifiers		avelos	Unpublished	2018-03-25	Mapping/1013
SealExampleMapping		athina	Unpublished	2018-06-18	Mapping/1014



#	SOURCE	TARGET
1	D /row	E31_Document E33_Linguistic_Object
1.1	P Type R Type	P2_has_type E55_Type
1.2	P Creator R Creator	P94i_was_created_by E65_Creation [create1] P14_carried_out_by E39_Actor
1.3	P Creator R Creator	P94i_was_created_by E65_Creation [create1] P14_carried_out_by E39_Actor
1.4	P Date R Date	P94i_was_created_by E65_Creation [create1] P4_has_time-span E52_Time-Span P82_at_some_time_within rdf-schema#Literal
1.5	P Title R Title	P1_is_identified_by E41_Appellation
1.6	P Subtitle R Subtitle	P1_is_identified_by E41_Appellation
1.7	P Subject R Subject	P129_is_about E55_Type
1.8	P language R language	P72_has_language E56_Language



New 3M



- Powered over new technologies (React, Java Spring, Material Design, etc.);
- Better performance stability over the old 3M;
- New improved and modern UI, which is configurable in several aspects;
- Quick and easy initialization of new mappings through a wizard of a few steps;
- Prevents users from defining the same namespaces with different prefixes;
- Provides validation of all selected values (uncovering conflicts);
- Provides a way of keeping track of the progress been made, based on the source input values coverage;
- Better supports the RDF Visualizer by providing a list of the instances by class to choose;



New 3M



- **Allows users to organize mappings by labelling them under categories;**
- **Encapsulates a generator definition manager;**
 - Allows constructing new definitions and using them;
 - Allows adding further definitions from a generator policy file;
 - Allows exporting the definitions used to a new generator policy file;
 - Provides steps for recovering issues related to missing definitions within a mapping;
- **Native additional element support (additional within additional);**
- **Named graph support within Mapping, Domain and Link Elements;**
- **Allows multiple users to edit the same mapping simultaneously, by handling the synchronization among the different user views.**

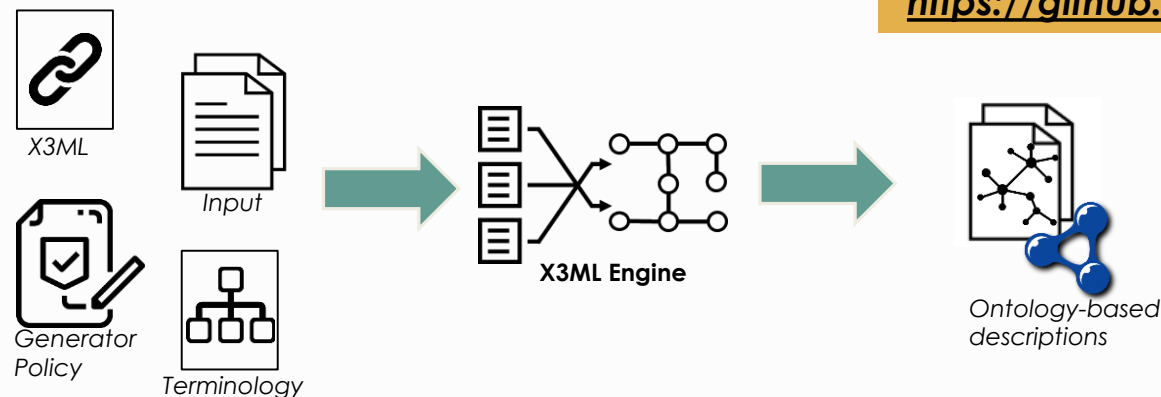


X3ML engine: A Transformation Tool

X3ML Engine is a tool that realizes the transformation of data resources to a target format with respect to an X3ML Mapping definition.

Main principles:

- Simplicity by design
- Transparency in terms of expected output
- Re-use of standards and technologies as much as possible
- Facilitating the instance matching process



X3ML engine: A Transformation Tool – cont'd

- X3ML Engine has been designed by FORTH.
 - The initial development has been carried out by DELVING B.V. under the support and contribution of FORTH (*until version 1.3 – March 2015*)
 - FORTH took over the full development of X3ML Engine since March 2015.
- 24 Releases in total (*Latest one: version 1.9.4 August 2020*)
- Available as: API, executable (console-based & GUI), service



<https://github.com/isl/x3ml>

```
C:\Repositories\Github\X3ML\target>java -jar x3ml-engine-1.9.4-SNAPSHOT-exe.jar
usage: x3ml -xml <input records> -x3ml <mapping file> hello
Options
  -a,--assocTable <arg> export the contents of the association table in XML format
  -f,--format <arg>      Output format. Options:
                        --format application/rdf+xml (default)
                        --format application/n-triples
                        --format application/trig
                        --format text/turtle
  -i,--input <arg>       XML input records.
                        Option A-single file: --input input.xml
                        Option B-multiple files (comma-sep): --input input1.xml,input2.xml,input3.xml
                        Option C-folder: --input F_folder_path
                        Option D-URL: --input @input_url
                        Option E-multiple URLs: --input @input_url1,input_url2,input_url3
                        Option F-stdin: --input @
                        merge the contents of the association table with the RDF output
  -m,--mergeAssocWithRDF The value policy file: --policy policy.xml
                        reports the progress of the transformations
                        the SKOS taxonomy
  -o,--output <arg>      Option A-single file: --terms @skos.terms.nt
                        Option B-URL: --terms @skos.terms.url
                        Create a test UUID generator of the given size.
                        Default is GUID from operating system
  -p,--policy <arg>      X3ML mapping definition.
                        Option A-single file: --x3ml mapping.x3ml
                        Option B-multiple files (comma-sep): --x3ml mappings1.x3ml,mappings2.x3ml
                        Option C-URL: --x3ml @mappings_url
                        Option D-stdin: --x3ml @
  -r,--reportProgress
  -t,--terms <arg>
  -u,--uuidTestSize <arg>
  -x,--x3ml <arg>

Missing required options: i, x
```



RDF Visualizer

- Visualizes and enables browsing of RDF data
 - Schema independent
 - Supports configuration rules to improve the visual layout of particular classes properties
 - displays information of high-density in one screen
- Supports browsing RDF data from
 - Triplestores (e.g. Virtuoso, Blazegraph)
 - Files (e.g. in TURTLE format)

1. Enter the URI of a resource

Enter subject

Choose a Template:



2. Visualize and browse over its connections

https://www.dainst.org/COIN/100044

Choose a Template:

100044 [CRM: E22_Man-Made_Object]

P2_has_type
coins [CRM: E55_Type]

P43_has_dimension
urn:uuid:fa92ac2c-3085-407a-8fa4-182ebb319ad5 [CRM: E54_Dimension]
weight of coin 100044 [CRM: E54_Dimension]
diameter of coin 100044 [CRM: E54_Dimension]

P52_has_current_owner
German Archaeological Institute [CRM: E40_Legal_Body]

P45_consists_of
Bronze [CRM: core#Concept]

P3_has_note
Datierung: ca. 215-Anfang 2. Jh. v. Chr. Westermarck 1991, 14; BMC (Mysien) S. 121-122, 75-83

P50_has_current_keeper
German Archaeological Institute [CRM: E40_Legal_Body]

P108l_was_produced_by
production of coin 100044 [CRM: E12_Production]
P7_took_place_at
Pergamon [CRM: E53_Place]
type
http://www.cidoc-crm.org/cidoc-crm/E12_Production
P17_was_motivated_by
Attaliden [CRM: E39_Actor]
P10_falls_within
hellenistisch [CRM: E4_Period]
P1_is_identified_by
100044 [CRM: E41_Appellation]

<https://demos.isl.ics.forth.gr/RDFV-Demo/>



RDF Visualizer – cont'd

Functionalities:

- Display images and image galleries
- Expand collapse big texts or big sets of results
- Display labels if available, URI on label hover
- Properties can be prioritized by user
- Select a specific URI and open it as root in new tab, copy it on clipboard or retrieve the path of its sub-graph
- Mark same instances

The screenshot displays the RDF Visualizer interface for a Hei-tiki object. The main panel shows a list of properties and their values, including 'P48_has_preferred_identifier' (EOC10881) and 'P3_has_note' (Tiki, neck ornament made of nephrite). A pop-up window titled 'Image Gallery' shows a green nephrite tiki mask with red eyes. Another pop-up window shows a context menu with options like 'Open Uri in new tab', 'Copy Uri', and 'Get path of selected item'.



Example

```
<?xml version="1.0" encoding="UTF-8"?>
<ExcavationUnits>
  <ExcavationUnit ExcavationUnitID="1006" SiteTrenchID="T1">
    <ExcavationInfo>
      <ExcavationWork>τεχνητό επίπεδο/arbitrary level</ExcavationWork>
      <ExcavationTechnique>κασμαδάκι μυστρί/hoes trowel</ExcavationTechnique>
      <ExcavationUnitRecorderID>A.N.& M.K.</ExcavationUnitRecorderID>
      <ExcavationUnitRecordingDate>11/08/2000</ExcavationUnitRecordingDate>
    </ExcavationInfo>
    <AnimalBone id="1"/>
    <AnimalBone id="2"/>
    <AnimalBone id="3"/>
    <Find id="1"/>
    <Find id="2"/>
  </ExcavationUnit>
</ExcavationUnits>
```



Example (cont'd)

```
<AnimalBone id="1">
  <description>Animal limb and head bones were scattered over the planum.</description>
  <BoneLimbWeight>3.0</BoneLimbWeight>
  <BoneHeadWeight>4.0</BoneHeadWeight>
  <AnimalBoneSpecies>goat</AnimalBoneSpecies>
</AnimalBone>
<AnimalBone id="2">
  <description>Animal head bones.</description>
  <BoneHeadWeight>5.0</BoneHeadWeight>
  <AnimalBoneSpecies>dog</AnimalBoneSpecies>
</AnimalBone>
<AnimalBone id="3">
  <description>Animal limb bones.</description>
  <BoneLimbWeight>2.0</BoneLimbWeight>
  <AnimalBoneSpecies>cat</AnimalBoneSpecies>
</AnimalBone>
```



Example (cont'd)

```
<Find id="1">
  <FindDescr>A crushed pot, bottom and top found separately</FindDescr>
  <FindType>pot</FindType>
  <FindMaterial>Clay</FindMaterial>
</Find>
<Find id="2">
  <FindDescr>Finger-Ring with D-shaped Profile</FindDescr>
  <FindType>ring</FindType>
  <FindMaterial>Silver</FindMaterial>
</Find>
```



THANK YOU!

ARIADNEplus is a project funded by the European Commission under the H2020 Programme, contract no. H2020-INFRAIA-2018-1-823914.

The views and opinions expressed in this presentation are the sole responsibility of the authors and do not necessarily reflect the views of the European Commission.

Contact: maria@ics.forth.gr
vkrits@ics.forth.gr

www.ariadne-infrastructure.eu

