Parthenos ending
- What got connected

Matej Ďurčo, Klaus Illmayer, Stefan Resch
CLARIN Annual Conference, Leipzig, 2019-10-02
Agenda

● PARTHENOS Project
● PARTHENOS Data (WP5)
  ○ Sources
  ○ PARTHENOS Entities Model
● Mapping & aggregation process
  ○ CMDI - special treatment
  ○ Some numbers
  ○ Data quality issues
● PARTHENOS Discovery
PARTHENOS project

“Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies”. It is inspired by the name of Athena Parthenos, the Greek goddess of wisdom, inspiration and civilization.

- [http://www.parthenos-project.eu/](http://www.parthenos-project.eu/)
- 2015-04 - 2019-10
- 16 partners from 8 countries

Context of this presentation:

- WP5 - Interoperability and semantics
- WP6 - Service and tools
  - T6.5 Resource Discovery
PARTHENOS Data

● sources
  ○ ARIADNE
  ○ CENDARI
  ○ CLARIN
  ○ Cultura Italia
  ○ DARIAH-GR
  ○ EHRI
  ○ Huma-Num
  ○ LRE MAP
  ○ METASHARE

● target:
  Parthenos Entities Model (extension of CIDOC CRM)
PARTHENOS Entities Model

- extension of CIDOC CRM (cf. D5.5 Common Semantic Framework)
- 39 classes, 47 properties
- Main classes: crm:Actor, crm:Place
  pe:Service, pe:Software, pe:Dataset, pe:Project
Mapping process

- **X3M engine** to define mappings (XML to RDF) integrated in d4science
- metadata records from content providers to PEM
- lossy by design, just capture minimal metadata (of main entities + relations)
- **d-net infrastructure** for harvesting and transformation of the records
CMDI to PEM

- with manual XML/XSD to RDF mapping not feasible
- intermediate **programmatic** step for automatic generation of mappings based on concepts and profiles
What got collected

<table>
<thead>
<tr>
<th>Source</th>
<th>Actors</th>
<th>Datasets</th>
<th>Services</th>
<th>Software</th>
<th>Places</th>
<th>Project</th>
<th>Sum</th>
</tr>
</thead>
<tbody>
<tr>
<td>CLARIN</td>
<td>201,484</td>
<td>1,731,716</td>
<td>2,707,348</td>
<td>127</td>
<td>4,451</td>
<td>32</td>
<td>4,645,158</td>
</tr>
<tr>
<td>ARIADNE</td>
<td>9,554</td>
<td>1,997,459</td>
<td>1,013,915</td>
<td>1</td>
<td>956,619</td>
<td>0</td>
<td>3,977,548</td>
</tr>
<tr>
<td>CENDARI</td>
<td>129</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>6,660</td>
<td>0</td>
<td>6,790</td>
</tr>
<tr>
<td>Cultura Italia</td>
<td>6,817</td>
<td>82,831</td>
<td>0</td>
<td>0</td>
<td>559</td>
<td>0</td>
<td>90,207</td>
</tr>
<tr>
<td>DARIAH-GR</td>
<td>640</td>
<td>2,282</td>
<td>2,281</td>
<td>0</td>
<td>573</td>
<td>0</td>
<td>5,776</td>
</tr>
<tr>
<td>EHRI</td>
<td>26,052</td>
<td>45,365</td>
<td>754</td>
<td>0</td>
<td>7,587</td>
<td>0</td>
<td>79,758</td>
</tr>
<tr>
<td>Huma-Num</td>
<td>505</td>
<td>41,233</td>
<td>1</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>41,743</td>
</tr>
<tr>
<td>LRE MAP</td>
<td>4,805</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>16</td>
<td>0</td>
<td>4,822</td>
</tr>
<tr>
<td>METASHARE</td>
<td>1,485</td>
<td>5,136</td>
<td>3,218</td>
<td>6</td>
<td>87</td>
<td>0</td>
<td>9,932</td>
</tr>
<tr>
<td>PARTHENOS</td>
<td>253</td>
<td>223</td>
<td>30</td>
<td>16</td>
<td>10</td>
<td>10</td>
<td>542</td>
</tr>
<tr>
<td><strong>Sum</strong></td>
<td><strong>251,724</strong></td>
<td><strong>3,906,256</strong></td>
<td><strong>3,727,548</strong></td>
<td><strong>150</strong></td>
<td><strong>976,566</strong></td>
<td><strong>42</strong></td>
<td><strong>8,862,286</strong></td>
</tr>
</tbody>
</table>
Data quality issues

- Missing data
  - Missing in source
  - Incomplete mapping
- Missing labels
- Erroneous mappings
  - Repeating of the same information
  - Wrong types
- Literal referring to entities

=> [github:/acdh-oeaw(pd-labs/.../DataQuality.md](https://github.com/acdh-oeaw/pd-labs/.../DataQuality.md)
PARTHENOS Discovery - T6.5

• extensive evaluation of existing solutions for resource discovery and visualisation (D6.4)

• [metaphactory](#) / [ResearchSpace](#)
  • dual licensing + support by metaphacts
  • large number of customizable HTML5 components
  • working against any SPARQL 1.1 endpoint
  • customisation via templating + fields
  • based on [Fundamental Categories and Relationships](#) for intuitive querying
PARTHENOS Discovery

Your gateway to the PARTHENOS Entities Dataspace. Metadata from all partner Research Infrastructures is aggregated here under a common semantic model, allowing for unprecedented interdisciplinary view on these heterogeneous datasets.

Explore/Search

PARTHENOS Discovery application offers multiple options to explore the dataset.
You can search by keyword, by semantic relations, by sources or via the underlying model.

- by keyword
  Keyword search with faceting over selected properties

- by semantic relations
  Advanced semantic search using query builder

- by sources
  Dataset divided by the contribution from individual sources

- by model
  Explore the data based on the underlying ontology the PARTHENOS Entities Model

Additionally you can dive into the dataset through one of the main types: actors, datasets, services, software, projects and places.
Either through keyword search or by exploring the entities in the model.
(The numbers indicate how many instances of given type (including subclasses) are there, distinguishing between explicitly created instances and all instances.)
Highlights

Main Entities
Browse through the manually crafted dataset of top-level entities like project partners, services and datasets, for which descriptions did not previously exist.

Showcases
This is a hand curated list of entities-detail views and other specialized pages to highlight specific aspects of the application or the dataset.

Images
Browse through records featuring an image representation.

Places
Browse through geo-referenced records. (Only available for records with geo-coordinates).

PARTHENOS Discovery Lab
Here are a few more advanced options to explore and experiment with the dataset.

SPARQL
User interface to run any SPARQL query on the underlying dataset.

Statistics
A suite of aggregation queries offering an overview over various aspects of the dataspace.

Data Notebooks
Connect to the dataset via a Jupyter Notebook and do your own processing.

Curation
Collection of custom pages to run experiments on or to showcase data quality aspects of the dataset.
Search/Explore
different options to dive into the dataset:
• keyword search
• search by semantic relations
• explore by sources
• explore via model
# Custom detail views

## PARTHENOS Project Consortium

### URI

http://parthenos.d4science.org/handle/Parthenos/REG/Actor/D1953717-1D39-3A6B-B8FC-6092C1DC5C402

### View in RDF visualizer

[View in RDF visualizer](http://parthenos.d4science.org/handle/Parthenos/REG/Actor/D1953717-1D39-3A6B-B8FC-6092C1DC5C402)

### Type

RI Consortium, Actor, Team, Legal Body

### Description

The PARTHENOS Project Consortium is a pan-European collective of research partners with Research Infrastructures from across the domains of archaeology, humanities, cultural heritage and linguistics inter alia, forming its core membership. The purpose of the consortium is to research, implement and disseminate common solutions to RI problems and cross-disciplinary access to resources.

### Source

PARTHENOS (http://parthenos.d4science.org/handle/api________::parthenos::__::parthenos::topLevel)

<table>
<thead>
<tr>
<th>Minimal metadata</th>
<th>Source (Raw triples)</th>
<th>Viewer</th>
<th>Knowledge Graph</th>
</tr>
</thead>
</table>

### Appellation

PARTHENOS Project Consortium

### Addresses

PIN, Servizi Didattici e scientifici per l'Università di Firenze Piazza G. Ciardi, 25, 59100 Prato (PO), Italy

### Phone

00 39 0574 602578

### Provides Services

PARTHENOS Policy Wizard, Parthenos Reference Resources Management, Standardization Survival Kit, PARTHENOS International Liaisons

### has created

Foresight Studies Dataset, Parthenos Standard Reference Resources, International Liaisons Contact information Dataset, Standardization Survival Kit Dataset, PARTHENOS Policy Wizard Dataset

### Has Member


### "Website"

http://www.parthenos-project.eu/
Knowledge Graph

PARTHENOS Project Consortium


Type: RI Consortium, Actor, Team, Legal Body
Description: The PARTHENOS Project Consortium is a pan-European collective of research partners with Research Infrastructures from across the domains of archaeology, humanities, cultural heritage and linguistics. The purpose of the consortium is to research, implement and disseminate common solutions to RI problems and cross-disciplinary access to resources.
Source: PARTHENOS (http://parthenos.d4science.org/handle/apl________:parthenos___:parthenos:topLevel)

Minimal metadata  Source (Raw triples)  Viewer  Knowledge Graph

Diagram showing connections and events related to the PARTHENOS Project Consortium, including:
- Creation of Parthenos Standard Resources
- Creation of Standardization Surveys
- Creation of PARTHENOS Policy
- Creation of International Liaisons
- Creation of Foresight Studies Dataset
- L11 was output of
- Votainus Database: PARTHENOS Policy Wizard Data

Diagram elements are labeled with event types and descriptions, indicating the various activities and connections within the consortium.
Ontology View

PE12_Data_Curating_Service

URI: http://parthenos.d4science.org/CRMext/CRMpe.rdf/PE12_Data_Curating_Service

Type: Class, Class

- **Subclasses**
  - Filter Subclasses
    - PE17_Curated_Data_E-Service

- **Superclasses**
  - Filter Superclasses
    - PE10_Digital_Curating_Service
    - PE3_Curating_Service
    - PE1_Service
    - Activity
    - Event

- **CRM Entity**
  - Temporal Entity
  - Period
  - Event
  - Activity
  - PE1_Service
  - PE3_Curating_Service
  - PE10_Digital_Curating_Service
  - PE12_Data_Curating_Service
  - PE17_Curated_Data_E-Service

- **Spacetime Volume**
  - Period
  - Event
  - Activity
  - PE1_Service
  - PE3_Curating_Service
  - PE10_Digital_Curating_Service
  - PE12_Data_Curating_Service
  - PE17_Curated_Data_E-Service

- **PE15_Data_E-Service**
## Ontology View

**Summary usage statistics:** `subject.class property object.class`

<table>
<thead>
<tr>
<th>Predicate</th>
<th>Object class</th>
<th>Count of triples</th>
</tr>
</thead>
<tbody>
<tr>
<td>type</td>
<td>Class</td>
<td>2322</td>
</tr>
<tr>
<td>is indexed by</td>
<td>Volatile Dataset</td>
<td>1602</td>
</tr>
<tr>
<td>has dataset snapshot</td>
<td>Persistent Dataset</td>
<td>1214</td>
</tr>
<tr>
<td>is identified by</td>
<td>Appellation</td>
<td>3117</td>
</tr>
<tr>
<td>is identified by</td>
<td>Identifier</td>
<td>3361</td>
</tr>
<tr>
<td>has type</td>
<td>Type</td>
<td>574</td>
</tr>
<tr>
<td>is dataset hosted by</td>
<td>Data E-Service</td>
<td>1615</td>
</tr>
</tbody>
</table>
Places

Explore georeferenced Places.

found 5 matches

<table>
<thead>
<tr>
<th>Subject</th>
<th>Coordinates</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lageweg (opgravingscode LA'07), Romeostraat (opgravingscode ROM'07), Amersfoort, Utrecht</td>
<td>POINT(5.40807343 52.15614322)</td>
</tr>
<tr>
<td>Lageweg (opgravingscode LA'07), Romeostraat (opgravingscode ROM'07), Amersfoort, Utrecht</td>
<td>POINT(5.40607349 52.16089817)</td>
</tr>
<tr>
<td>Lorry Park, Rom Valley Way, Romford</td>
<td>POINT(0.18623499572277 51.569087982178)</td>
</tr>
<tr>
<td>ROM VALLEY WAY LORRY PARK</td>
<td>POINT(0.17863899469376 51.561134338379)</td>
</tr>
<tr>
<td>ROM VALLEY WAY OLDCHURCH PARK</td>
<td>POINT(0.17863899469376 51.561134338379)</td>
</tr>
<tr>
<td>KomaErom (Hungary)</td>
<td>POINT(5.40807343 52.15614322)</td>
</tr>
</tbody>
</table>

Lorry Park, Rom Valley Way, Romford

URI
http://parthenos.d4science.org/handle/Ariadne/AriadnePortal/r1?se2923m0k

Type
Place

Source
ARIADNE (http://parthenos.d4science.org/handle/api_________::ariadne::ADS)

Minimal metadata

<table>
<thead>
<tr>
<th>has coordinates</th>
<th>falls within</th>
<th>is Subject of</th>
</tr>
</thead>
<tbody>
<tr>
<td>POINT(0.18623499572277 51.569087982178)</td>
<td>ENGLAND</td>
<td>Lorry Park, Rom Valley Way, Romford. Archaeological Desktop Study</td>
</tr>
</tbody>
</table>
Images

Museo del mare di Monopoli
Museo archeologico regionale di Caltanissetta e zona Giallo Gibil
Museo dell'aviazione
MUPAC - Museo dei paesaggi di terra e di fiume di Colonnato
Museo capitolare di Rutigliano "S. Maria della Colonna e S. Nicola"
Museo del picapreta
Museo della figurina
Raccolta d'arte Fondazione Cassa di Risparmio di Termini Imerese
Museo civico Ugento
Museo etnografico "G. Pitre"

Rocca Melfi
Area archeologica di Santa Cristina
Mostra permanente - "Il museo della Città romana di Cisternica"
Museo Ugo Carà
Walsermuseum
Mudas museum - Museo diocesano di Arezzo
MET - Museo degli usi e costumi della gente di Romagna
Museo della Resistenza di Bologna
Museo dei minerali e della mineraria
Museo sacri e regimento "Giovanni Forcioni" Aeronautica

Museo specializzato del treno in miniatura
Museo provinciale di storia naturale
Museo Fragonard da Tusa
Museo zoologico Cambria
Museo raccolta di fisica in collezione stabile
Scuola di disegno "Pietro Bazzoni"
Cà la Ghironda
Castello Visconteo di San Vito
Museo del Novecento di Milano
Sala storica della resistenza

Museo dell'aviazione

URI
http://parthenos.d4science.org/handle/Culturaitalia/ICCU/Actor
/EF7B3448-27A9-385F-889A-7FF92AB8C5A0

View in RDF visualizer

**Type**

**Description**

Auxiliary resources

- **p:Docs**
- **pd-labs**
  - scripts for automatic access of PARTHENOS data (Python notebooks)
  - give feedback on Parthenos Discovery
  - track data quality issues
- **sparqLaborer**
  - python application to automatically run a set of SPARQL-queries
  - store the results in various formats, either locally as CSV files, or as Google spreadsheets
  - features query parametrisation and recursive query execution
Conclusions & Outlook

• We came a long way, …
  … but not to the envisaged goal

• PARTHENOS Discovery will become part of PARTHENOS VRE
• Expression of heterogeneous mappings in a sound conceptual model
  as base for further harmonisation

• Good basis for curation
  representation of Datasets, Services, Software, Actors
  => take up in SSHOC / Marketplace or other EOSC-related activities
  related to semantic interoperability

• Inventory of relevant vocabularies
  (D5.4 Report on Reference Resources)
  => use in CMDI curation?