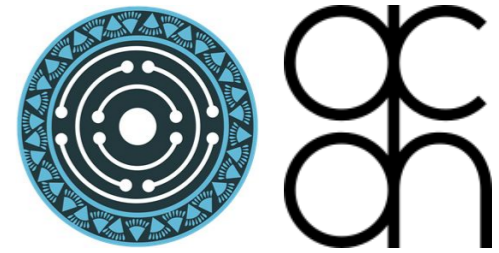


Parthenos ending - What got connected



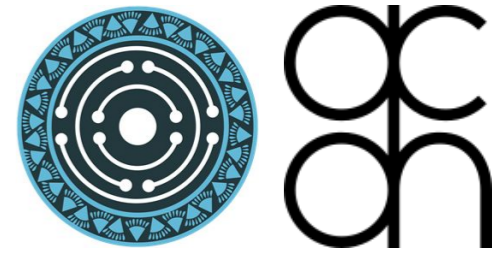
Matej Ďurčo, Klaus Illmayer, Stefan Resch
CLARIN Annual Conference, Leipzig, 2019-10-02

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



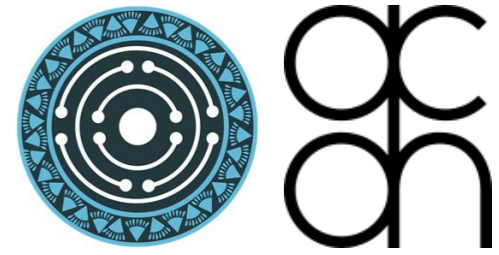
“Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies”. It is inspired by the name of Athena Parthenos, the Greek goddess of wisdom, inspiration and civilization.

- <http://www.parthenos-project.eu/>
- 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



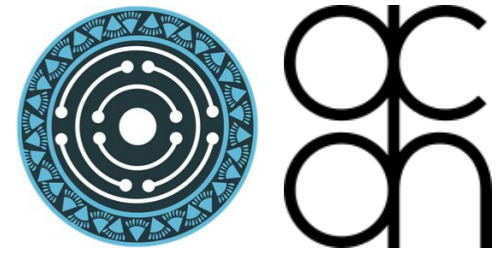
- sources

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

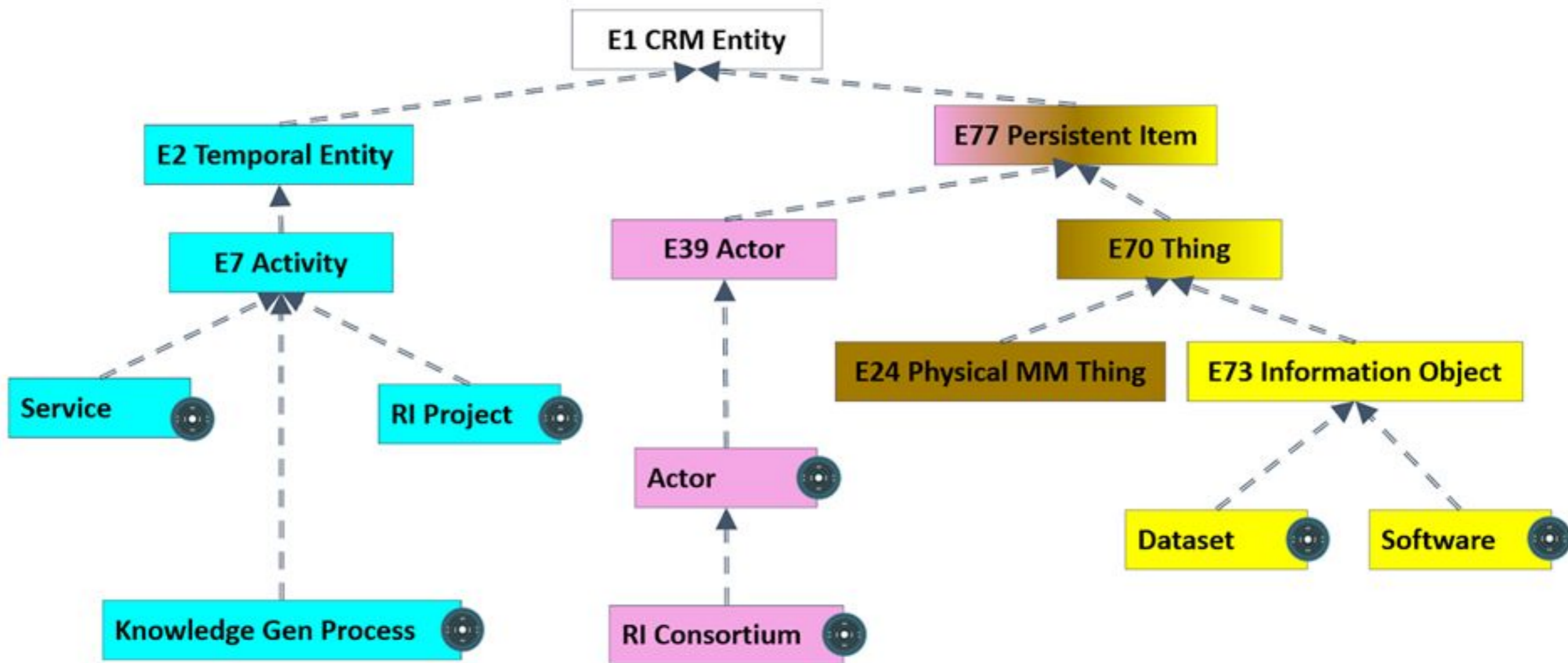
- target:

Parthenos Entities Model (extension of CIDOC CRM)

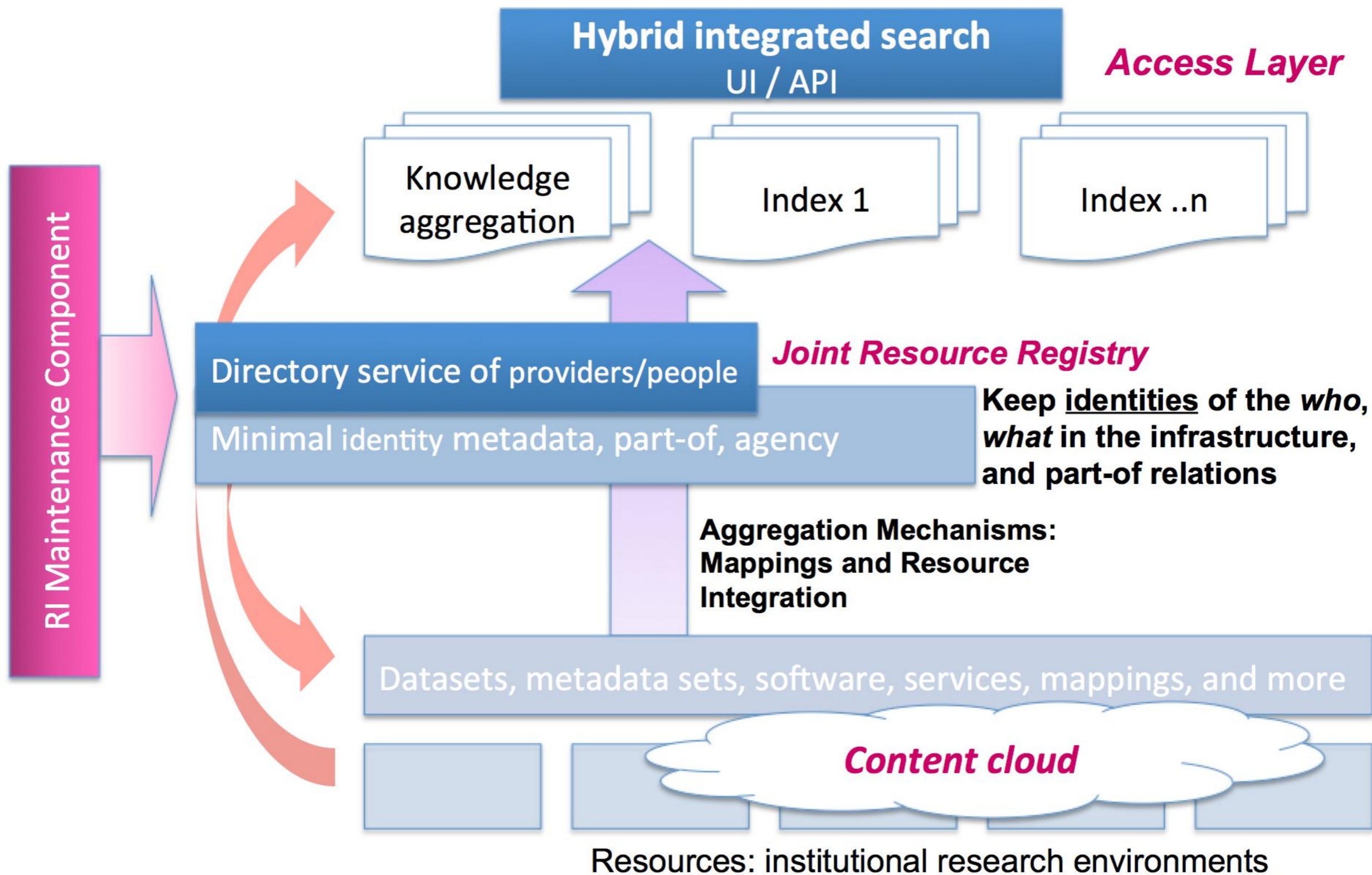
PARTHENOS Entities Model



- 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`

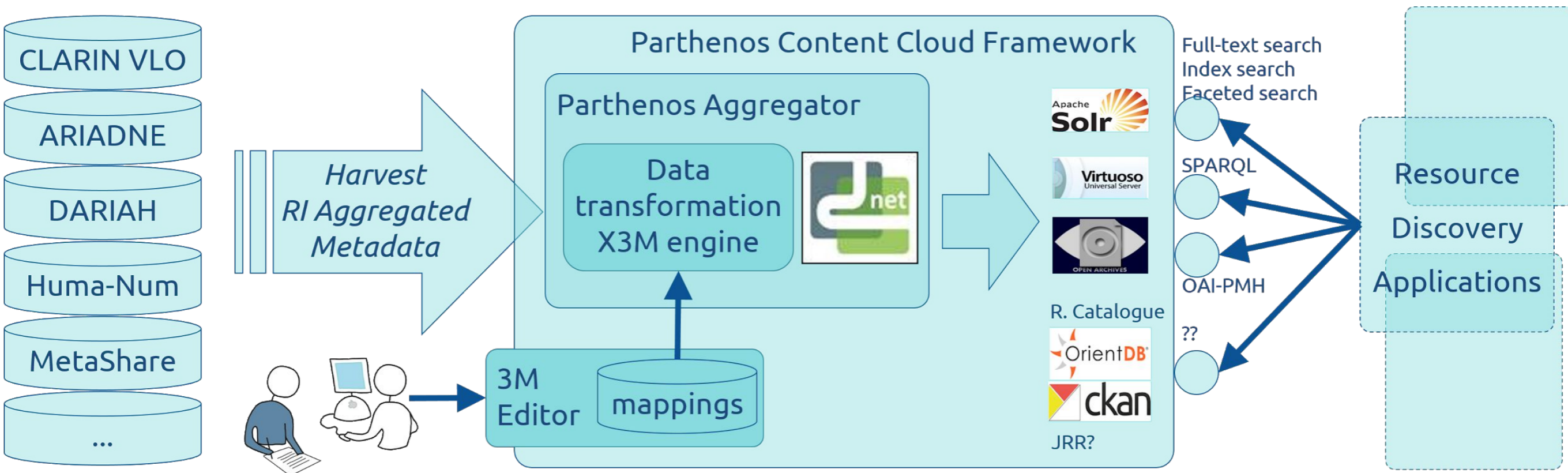


Conceptual Architecture

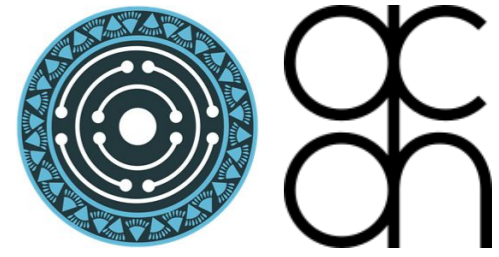


Mapping process

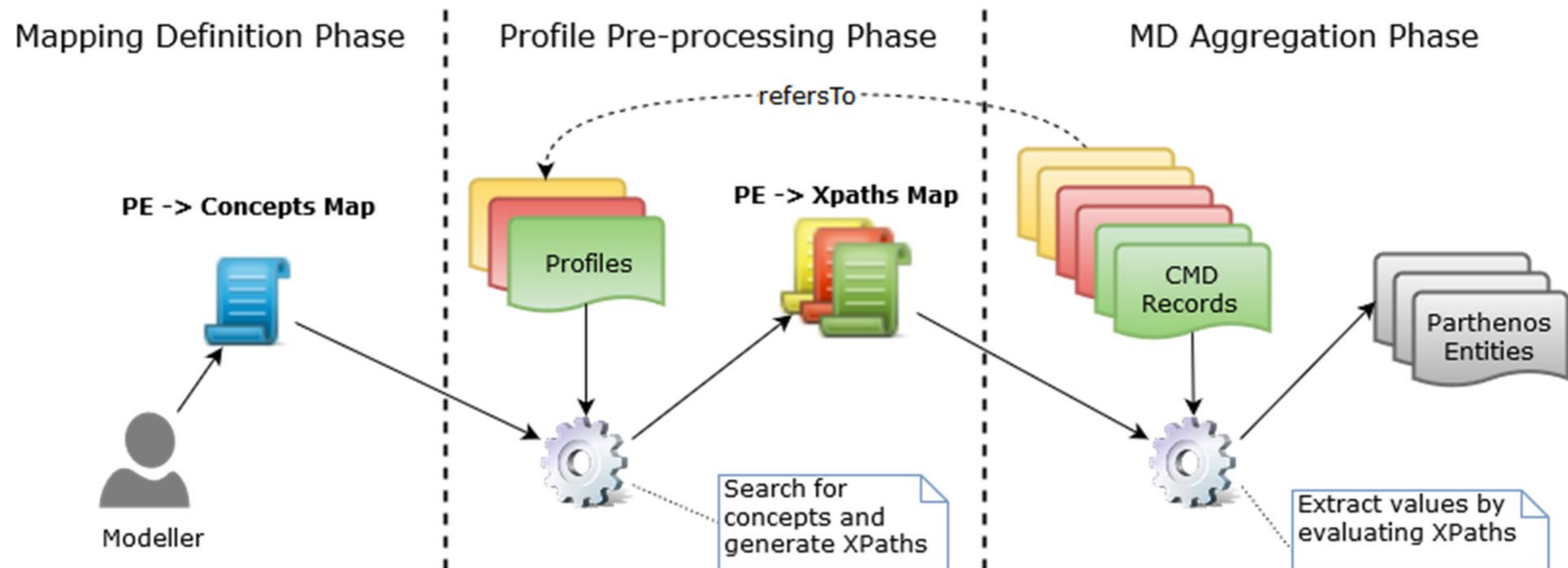
- [X3M engine](#) to define mappings (XML to RDF)
[integrated in d4science](#)
- metadata records from content providers to PEM
- lossy by design, just capture minimal metadata (of main entities + relations)
- [d-net infrastructure](#) for harvesting and transformation of the records



CMDI to PEM



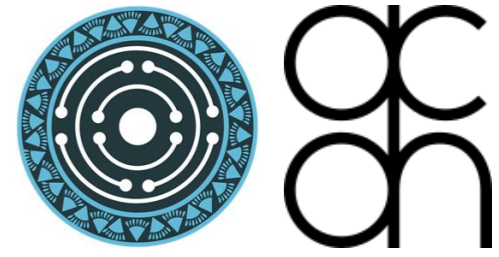
- with manual XML/XSD to RDF mapping not feasible
- intermediate programmatic step for automatic generation of mappings based on concepts and profiles



What got collected



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

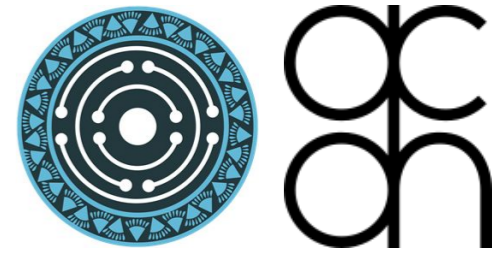


Data quality issues

- 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/blob/master/DataQuality.md)

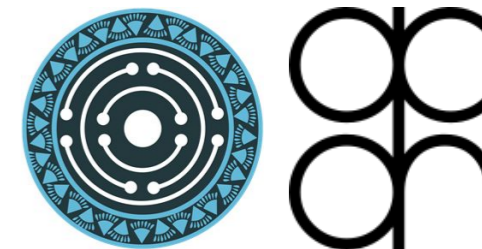
PARTHENOS Discovery - T6.5



- 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



<https://parthenos.acdh.oeaw.ac.at/>



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.



Help Feedback

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.)



Actor

🔍 search | ? model
| 42136 / 251722 |



Dataset

🔍 search | ? model
| 3893440 / 3906344 |



Service

🔍 search | ? model
| 789 / 3741383 |



Software

🔍 search | ? model
| 144 / 150 |



Project

🔍 search | ? model
| 39 / 42 |



Place

🔍 search | ? model
| 138568 / 976568 |



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

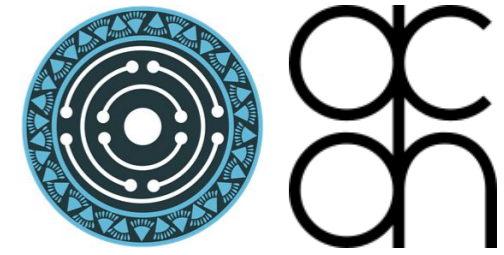
Connect to the dataset via a Jupyter Notebook and do your own processing.



Curation

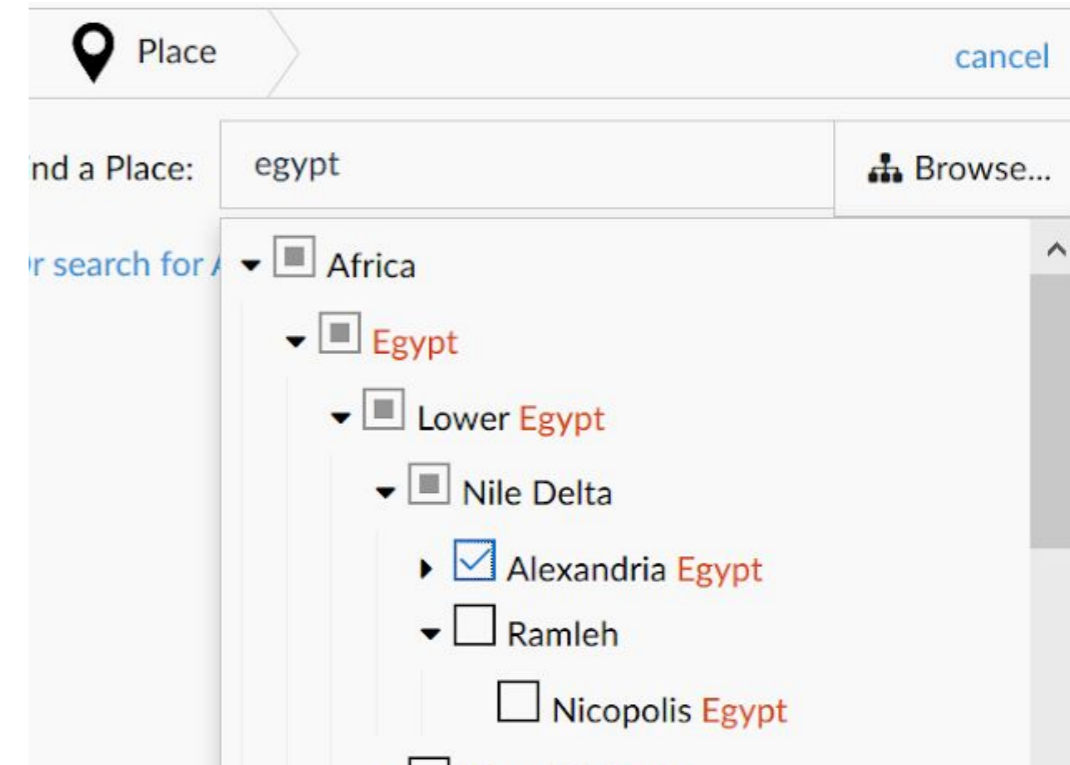
Collection of custom pages to run experiments on or to showcase data quality aspects of the dataset.

Search/Explore



different options to dive into the dataset:

- keyword search
- search by semantic relations
- explore by sources
- explore via model



Find: Actors HASCURATED PARTHENOS Policy Wizard Dataset



Filter sidebar with a 'Hide Filter' button and three expandable filter categories:

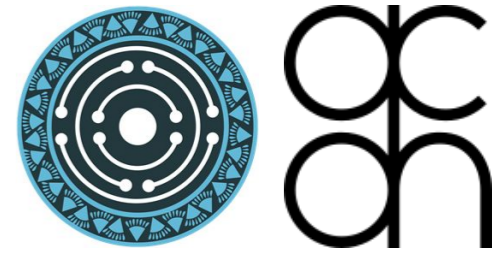
- hasCurated
- hosts
- has member

found 1 matches

Filter Results

Subject	Type
PARTHENOS Project Consortium	RI Consortium, Actor, Team, Legal Body

Custom detail views



PARTHENOS Project Consortium



URI	http://parthenos.d4science.org/handle/Parthenos/REG/Actor/D1953717-1D39-3A6B-B8FC-6092C1DCC402
	View in RDF visualizer
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)

Minimal metadata

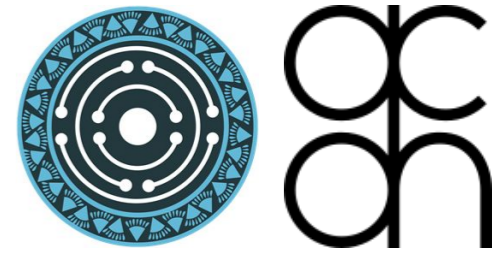
[Source \(Raw triples\)](#)

[Viewer](#)

[Knowledge Graph](#)

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	Sheena Bassett , Hella Hollander , Laurent Romary , Bente Maegaard , George Bruseker , CLARIN-EU , EHRI Consortium , DARIAH EU , ARIADNE Consortium , DARIAH-GR Consortium , CENDARI Consortium , Huma-num Consortium
"Website"@	http://www.parthenos-project.eu/

Knowledge Graph



PARTHENOS Project Consortium

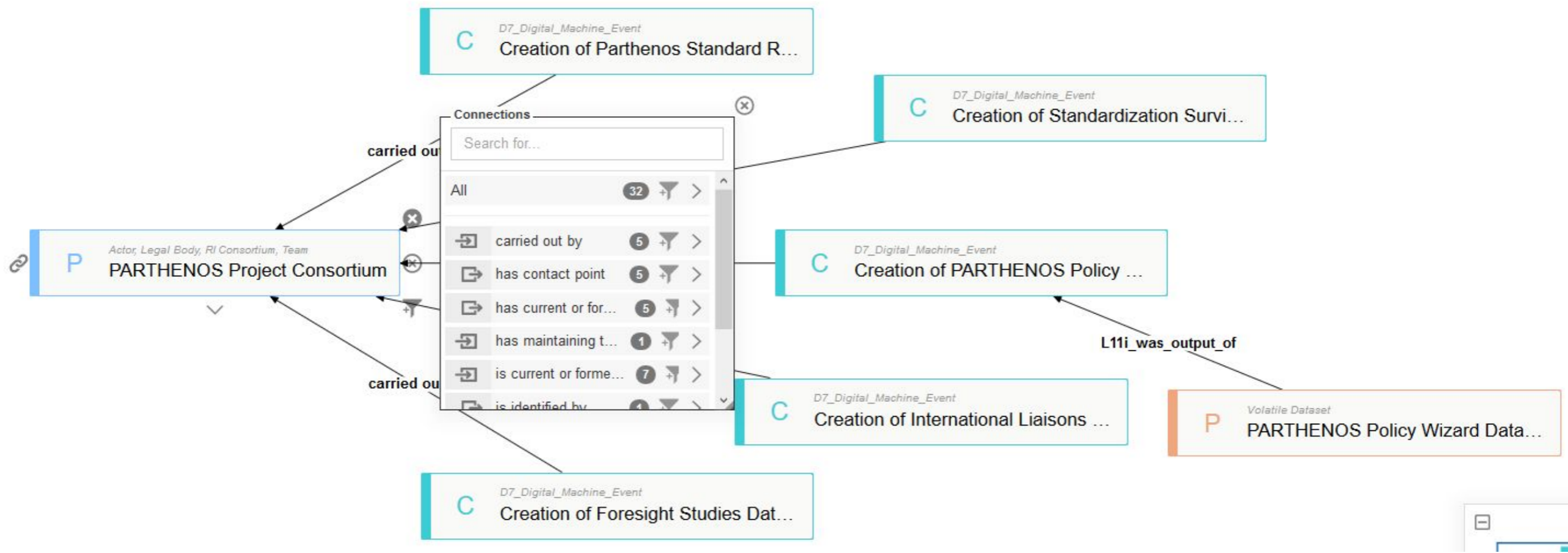
URI <http://parthenos.d4science.org/handle/Parthenos/REG/Actor/D1953717-1D39-3A6B-B8FC-6092C1DCC402>
[View in RDF visualizer](#)

Type RI Consortium, Actor, Team, Legal Body

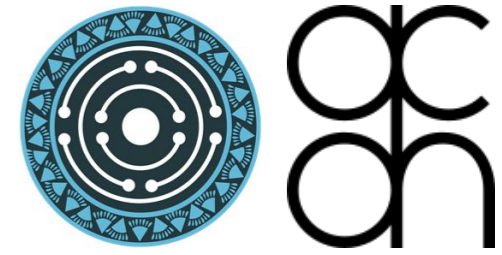
Description The PARTHENOS Project Consortium is a pan-European collective of research partners with Research Infrastructures from across the domains of archaeology, humanities, cultural heritage and linguistics. The purpose of the consortium is to research, implement and disseminate common solutions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

Source PARTHENOS (<http://parthenos.d4science.org/handle/api/parthenos::parthenos::topLevel>)

[Minimal metadata](#) [Source \(Raw triples\)](#) [Viewer](#) [Knowledge Graph](#)



Ontology View



V - Parthenos

Ontologies / Parthenos / PE12_Data_Curating_Service

PE12_Data_Curating_Service

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

Type: Class, Class

Taxonomy

Properties

Individuals

Subclasses

Filter Subclasses

PE17_Curated_Data_E-Service	direct
---	--------

Superclasses

Filter Superclasses

PE10_Digital_Curating_Service	direct
PE3_Curating_Service	inferred
PE1 Service	inferred
Activity	inferred
Event	inferred

« 1 2 »

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

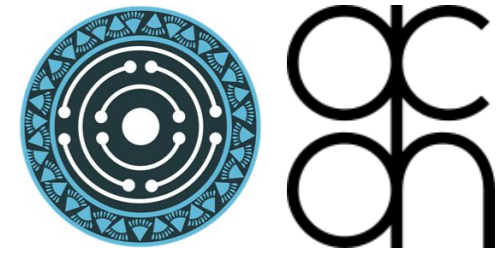
PE10_Digital_Curating_Service

PE12_Data_Curating_Service

PE17_Curated_Data_E-Service

PE15_Data_E-Service

Ontology View



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

[Home](#) / [Assets](#) / [Ontologies](#) / Volatile Dataset  Edit Page



[Taxonomy](#)

[Properties](#)

Usage

[Instances](#)

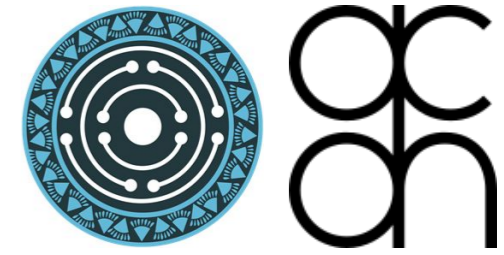
[Type](#)

Overview with which properties and other classes **Volatile Dataset** is being used:

Filter Results

Predicate	Object class	Count of triples
type	Class	2322
is indexed by	Volatile Dataset	1602
has dataset snapshot	Persistent Dataset	1214
is identified by	Appellation	3117
is identified by	Identifier	3361
has type	Type	574
is dataset hosted by	Data E-Service	1615

Places



Places

Explore georeferenced Places.

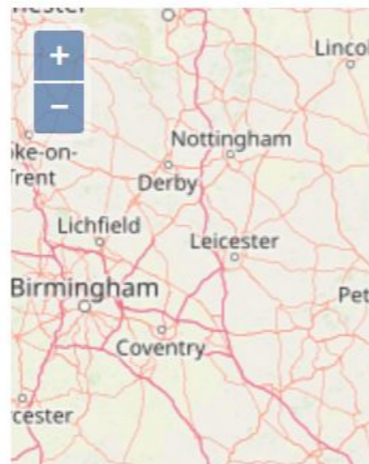
rom

Regions

Results

found 5 matches

Subject	Coordinates
Lageweg (opgravingscode LA'07), Romeostraat (opgravingscode ROM'07), Amersfoort, Utrecht	POINT(5.40807343 52.15614322)
Lageweg (opgravingscode LA'07), Romeostraat (opgravingscode ROM'07), Amersfoort, Utrecht	POINT(5.40607349 52.16089817)
Lorry Park, Rom Valley Way, Romford	POINT(0.18623499572277 51.569087982178)
ROM VALLEY WAY LORRY PARK	POINT(0.17863899469376 51.561134338379)
ROM VALLEY WAY OLDCHURCH PARK	POINT(0.17863899469376 51.561134338379)
KomaĔrom (Hungary)	



Lorry Park, Rom Valley Way, Romford
POINT(0.186234995722 51.569087982178)
ROM VALLEY WAY
LORRY PARK
POINT(0.178638994693 51.561134338379)



Lorry Park, Rom Valley Way, Romford

URI

<http://parthenos.d4science.org/handle/Ariadne/AriadnePortal/r17se2923m0k>

[View in RDF visualizer](#)

Type

Place

Source

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

Minimal metadata

Map

Source (Raw triples)

Viewer

Knowledge Graph

has coordinates

POINT(0.18623499572277 51.569087982178)

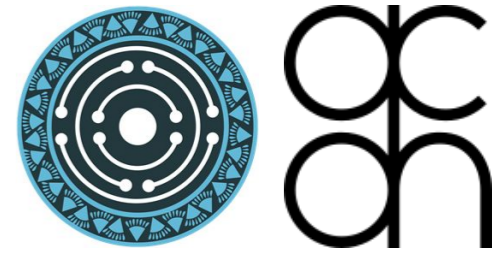
falls within

ENGLAND

is Subject of

[Lorry Park, Rom Valley Way, Romford. Archaeological Desktop Study](#)

Images



Museo del mare di Monopoli	Museo archeologico regionale di Caltanissetta e zona Gibil Gabib	Museo dell'aviazione	MUPAC - Museo dei paesaggi di terra e di fiume di Colomo	Museo capitolare di Rutigliano "S. Maria della Colonna e S. Nicola"	Museo del picapedra	Museo della figurina	Raccolta d'arte Fondazione Cassa di Risnarmio di Terni e	Museo civico Ugento	Museo etnografico "G. Pitre"
Rocca Meli Lupi	Area archeologica di Santa Cristina	Mostra permanente - "Il museo della Città romana di Claterna"	Museo Ugo Carà	Walsermuseum	Mudas museum - Museo diocesano di Arezzo	MET - Museo degli usi e costumi della gente di Romaena	Museo della Resistenza di Bologna	Museo dei minerali e della miniera	Museo sacratio reggimento "Giovani Fascisti" Africa
Museo specializzato del treno in miniatura	Museo provinciale di storia naturale	Museo Fra Gianmaria da Tusa	Museo zoologico Cambria	Museo raccolta di fisica (o collezione storico-	Scuola di disegno "Giuseppe Bottoli"	Cà la Ghironda	Castello Visconti di San Vito	Museo del Novecento di Milano	Sala storica della resistenza
Museo naturalistico paleontologico	Teatro comunale di Galeata	Museo di storia naturale di Parma	Museo Nazionale dell'Ebraismo Italiano della Shoah - ME						

Museo dell'aviazione

URI

<http://parthenos.d4science.org/handle/Culturalitalia/ICCUCI/Actor>

[/EF783448-27A9-3B5F-8B9A-7FF92AB8C5A0](#)

[View in RDF visualizer](#)

Type

Actor

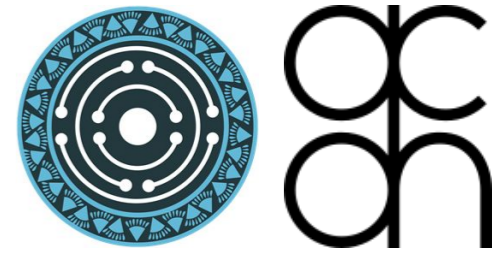
Description

Progettato sulla fine degli anni Ottanta da un gruppo di ex ufficiali dell'Aeronautica Militare Italiana, il Parco-Museo è stato inaugurato nel 1994. Tra i più importanti di questo genere in Italia, copre un'area di oltre sessantamila metri quadrati. Qui sono collocati rari esemplari di velivoli: dal Sukoi 7, unico esistente in occidente, al Mig 23 della nota serie russa, unico in Italia. Non meno raro è il Saiman 202 del 1932. Il

Feedback

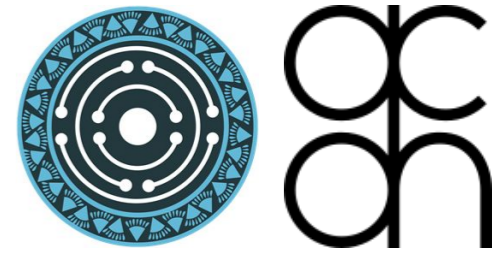


Auxiliary resources



- [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?