PARTHENOS

Pooling Activities, Resources and Tools
for Heritage E-research Networking,
Optimization and Synergies

D2.4 REPORT ON THE ASSESSMENT OF INTEROPERABILITY, SERVICES AND TOOLS

PARTNER(s):
CLARIN (WP Leader)
CNR-OVI (Task Leader)
KNAW-NIOD,
MIBACT-ICCU,
OEAW,
SISMEL
With contributions from all PARTHENOS partners

DATE: 25th April 2018

HORIZON 2020 - INFRADEV-4-2014/2015:

Grant Agreement No. 654119

PARTHENOS

Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

Report on the Assessment of Interoperability, Services and Tools

Deliverable Number	D2.4
Dissemination Level	PUBLIC
Delivery date	26 April 2018
Status	Final
Author(s)	Emiliano Degl'Innocenti (CNR-OVI) Roberta Giacomi (SISMEL) Susanne Haaf (CLARIN-BBAW) Steven Krauwer (CLARIN) Klaus Illmayer (OEAW) Antonio Davide Madonna (MIBACT-ICCU) Maurizio Sanesi (SISMEL)
Deliverable Number	D2.4
Dissemination Level	PUBLIC
Delivery date	25 April 2018

Project Acronym	PARTHENOS
Project Full title	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies
Grant Agreement nr.	654119

Deliverable/Document Information

Deliverable nr./title	D2.4
Document title	Report on the Assessment of Interoperability, Services and Tools
Author(s)	Emiliano Degl'Innocenti (CNR-OVI, Task Leader)
	Roberta Giacomi (SISMEL)
	Susanne Haaf (CLARIN-BBAW)
	Steven Krauwer (CLARIN, WP Leader)
	Klaus Illmayer (OEAW)
	Antonio Davide Madonna (MIBACT-ICCU)
	Maurizio Sanesi (SISMEL)
Dissemination level/distribution	PUBLIC

Document History

Version/date	Changes/approval	Author/Approved by
V 0.1 22.11.2017	First draft of the document	Emiliano Degl'Innocenti
V 0.2 28.03.2018	2nd draft, evaluation of the global tools completed (1st revision)	M. Sanesi (SISMEL), R. Giacomi (SISMEL), K. Illmayer (OEAW), A. Joffres (CNRS), S. Haaf (CLARIN), A.D. Madonna (ICCU)
V 0.3 19.04.2018	Third draft, evaluation of the specialized services completed (1st revision)	C. Meghini (CNR)
V 0.4 23.04.2018	First complete draft completed	Emiliano Degl'Innocenti (CNR), Steven Krauwer (CLARIN)
V 0.5 24.04.2018	Review of draft, editing and formatting	S. Bassett (PIN)
Final 27.04.2018	Final Document	F. Niccolucci (PIN)

Contents

Executive Summary

This document is the deliverable D2.4 "Report on the Assessment of Interoperability, Services and Tools". It is part of the process to collect needs and requirements from the project constituencies (research communities, ESFRI and non-ESFRI Research Infrastructures, other projects), to organize them so that they can be used to feed the technical implementation of the project infrastructure. After a first collection of requirements documented in deliverable D2.1 "Report on User Requirements", published on 31 January 2016 and further updated on 20 October 2016, the content was taken up by WP5 (Interoperability and semantics) and WP6 (Services and tools). The role of WP5 was to define a common framework of knowledge representation and the related processes of use for the sector in order to integrate knowledge in an open scientific workflow. The role of WP6 was to provide a description of a cluster of tools integrated in the project, resulting in deliverable D6.2 "Report on services and tools", published on 28 April 2017.

The role of task 2.3, as described in the DoW and in consideration of deliverables D2.1 and D6.2:

> "Definition of interoperability and related service requirements: The task deals with the requirements about interoperability and related services, expressed by the research communities involved in the project, also considering services already activated and tools used by the research communities and their progress toward interoperability. It takes into account the different needs of the various communities and the resulting common area of action. It analyses prior work done within ESFRI and other integrating activities and describes user requirements and needs as regards interoperability, related services and common tools. It produces a requirement (internal) document to be used in WP5 and WP6 to be used as roadmap for the design and implementation of interoperability, services and tools and to be incorporated into D2.1.Further on in the project the task assesses the interoperability level achieved in WP5 as well as the tools and services implemented in WP6 and their usability, and if necessary proposes amendments."

The structure of the document reflects this aim: after a short introduction in which the interactions within the project (mainly with WP5 and WP6) and the methodology adopted

for the assessment are defined, the key elements that will be evaluated in the report are described:

Chapter A (Data) describes the PARTHENOS Dataspace with a particular focus on why it was decided to adopt a common application profile; although technically interoperability can be achieved with standardisation, agreement on information meaning and data semantics is required to provide a deeper integration level and to create a coherent digital ecosystem of data, services and tools.

Chapter B (Infrastructural Services and Tools) describes the tools integrated in the infrastructure, grouping them in:

- Tools enabling interoperability;
- VREs management;
- Social Networking Services.

Chapter C (General Purpose Services and Tools), deals with applications and tools providing access to global resources.

In **chapter D** (Domain Services and Tools) a group of experts coming from different research domains in the DH landscape, tested a cluster of services or combination of services (listed in table 2.1.2) meant to support specific scientific purposes related to various research fields. Such services are usually not provided by general-purpose (i.e.: global / not domain driven) digital tools or infrastructures.

Finally, **chapter E** (Systematic Gap Analysis) synthesizes the results of the previous chapters, representing the current level of adequacy of the PARTHENOS infrastructure with respect to the users' expectations, and proposing a set of amendments.

1. Introduction and structure of the document

Assessment is often strongly connected only to the concept of quality and has placed attention on the issues of control, verification and certification. Evaluation is a reflexive activity of research and analysis and does not simply mean stating the deviation from the rules and procedure, nor stiffly indicating what the best procedures are to follow. The assessment defines the reason for the correct behaviour of an infrastructure and is addressed to an organizational process with particular attention to the results actually achieved and to the response that these results provide to the needs originally intended to be satisfied. This assessment is an evaluation based on the collection and interpretation of information and is, therefore, configured as a research process.

This report describes and defines the assessment of the joint services and tools produced by Work Packages 5 and 6[1], comparing them with the requirements and needs expressed by the reference communities and collected by Work Package 2[2]. After defining the methodology and the approach used to carry out this evaluation, the report first analyses the services that guarantee the interoperability of project data, then evaluates the various services made available to the project starting from general purpose tools through to specialized and specific community services.

In the core part of the document, a systematic analysis is performed where the connection between the requested services and the offered tools is studied and amendments are proposed and recommended in order to make this binding more effective.

The aim of this report is to document the first iteration of the described process to align the platform to the PARTHENOS community needs. Further development work towards the production version of the PARTHENOS ecosystem will take into account the most relevant recommendations formulated in the following sections.

---

[1] Project's Work Packages organization http://www.parthenos-project.eu/activities-and-wps/
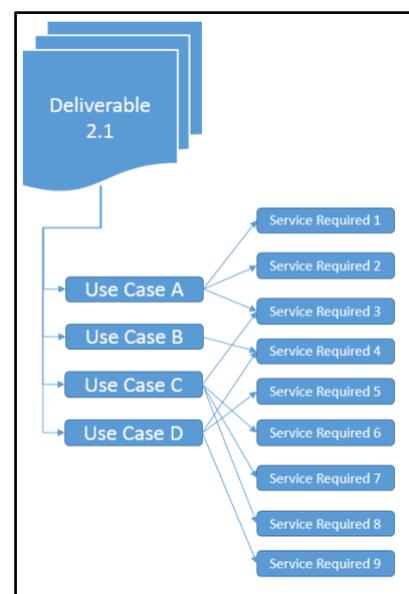 (last visited 27/04/2018)
[2] see above

1.1   State-of-the-art

The PARTHENOS infrastructure[3] provides access to a digital ecosystem with a variety of digital datasets and services, both generic and specialized for the needs of specific communities. At first, the Project constituencies collected needs from their reference communities, research infrastructures (both ESFRI [4]and non-ESFRI) and other projects through the analysis of existing reports. These needs have been organized and structured as requirements for the technical implementation phase, according to the project's global goals. Based on these requirements, knowledge bases and services have been built and integrated, to allow researchers both developing innovative research trends and supporting more traditional research paths with relevant digital tools and methods. To fully support current research practices and the development of novel workflows, a strong digital research infrastructure has been designed and deployed around two main software components: D-NET[5] and D-4Science[6]. While D-NET is providing a set of services for the construction of Aggregative Data Infrastructures, including data mediation, data mapping, data storage and indexing, data curation and enrichment, and data provision, D4Science is the framework allowing the creation and management of Virtual Research Environments[7] (VREs), where the services are running.

1.2   Interaction with WP2

The main input for users' needs in the PARTHENOS platform development has been the D2.1 Report on User Requirements[8], compiled by WP2. The Report, in addition to providing a set of global requirements, mainly focuses on the four communities involved in the project, i.e.: (i) History, (ii) Language Related Studies, (iii) Archaeology & Heritage Disciplines, (iv) Social Sciences. For each of these communities - through a selection of relevant use cases, showing key examples of the paths that a researcher usually follows to achieve results in his/her field – WP2 collected: (i)

---

[3] PARTHENOS Project official webpage http://www.parthenos-project.eu/
  (last visited 27/04/2018)
[4] ESFRI http://www.esfri.eu/ (last visited 27/04/2018)
[5] D-NET Lab http://www.d-net.research-infrastructures.eu (last visited 27/04/2018)
[6] D4Science Platform https://services.d4science.org/ (last visited 27/04/2018)
[7] D4Science VREs https://services.d4science.org/explore (last visited 27/04/2018)
[8] Report on User Requirements
https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a65b7331&appId=PPGMS (last visited 27/04/2018)

descriptions of research procedures; (ii) related scientific requirements. At the end of the above process, a set of requirements were distilled that are meant to highlight not only what a researcher needs to fulfil his/her research goals, but also to spot potential gaps or issues he/she can run into on the way to his/her final target. Furthermore, the aforementioned requirements were used as guidelines for the development and integration of the technical components of the PARTHENOS VREs ecosystem and, eventually, to check that its features are adequate enough to meet the researchers' needs.

1.3 Interaction with WP6

Among the PARTHENOS key features, there's the provision of a strongly integrated environment allowing a high level of interoperability for a selection of tools and datasets relevant for the researchers' community. The tools have been selected from among the ones (already available within the PARTHENOS community) matching two main criteria:

1. Tools that were used by a significantly wide community;

2. Tools explicitly requested during the gathering process of the PARTHENOS community's needs.
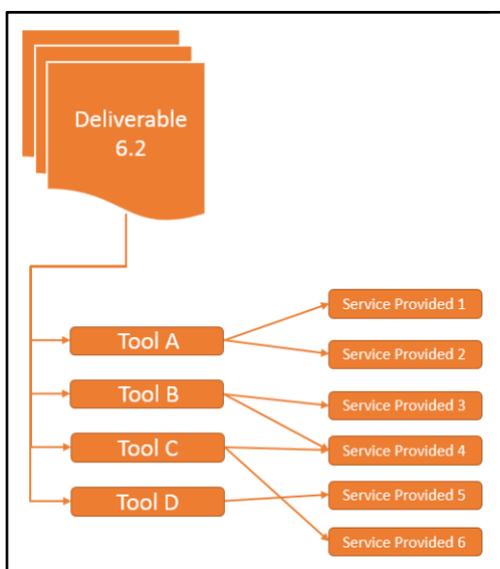
Work Package 6, whose motto is "Make Things happen", has the assignment of achieving interoperability between the selected tools and services and integrating them in the PARTHENOS infrastructure, enriching and linking together all the digital resources. To meet this target, Work Package 6 has developed deliverable 6.2[9] in which a set of tools and services candidates for integration was initially listed and described; the initial selection has been further consolidated and extended. A selection of specialized services from the revised list - already integrated in the PARTHENOS infrastructure at the moment of writing this document - represents the technical base of this assessment

---

[9] Report on services and tools http://www.parthenos-project.eu/Download/Deliverables/D6.2_Report_on_services_and_tools.pdf (last visited 27/04/2018)

2.   Assessment of interoperability

2.1 Approach adopted and methodology

The idea behind this report is to assess to what extent the requirements expressed by the PARTHENOS communities in terms of interoperability - as described in D2.1 - are satisfied by the services and tools actually integrated in the PARTHENOS infrastructure. For reasons of clarity and to make the report useful to plan future development actions for the PARTHENOS ecosystem, we adopted a taxonomic division of the services. The services will be analysed differently according to their purpose:

- infrastructural services: tools with particular focus on VRE management, Social networking services and data integration;
- global services: general-purpose tools (e.g. metadata inspection, data discovery and analysis);
- specialized services: tools supporting specific communities' needs. They will be further divided by community: (i) History, (ii) Language Related Studies, (iii) Archaeology & Heritage Disciplines (iv) Social Sciences .
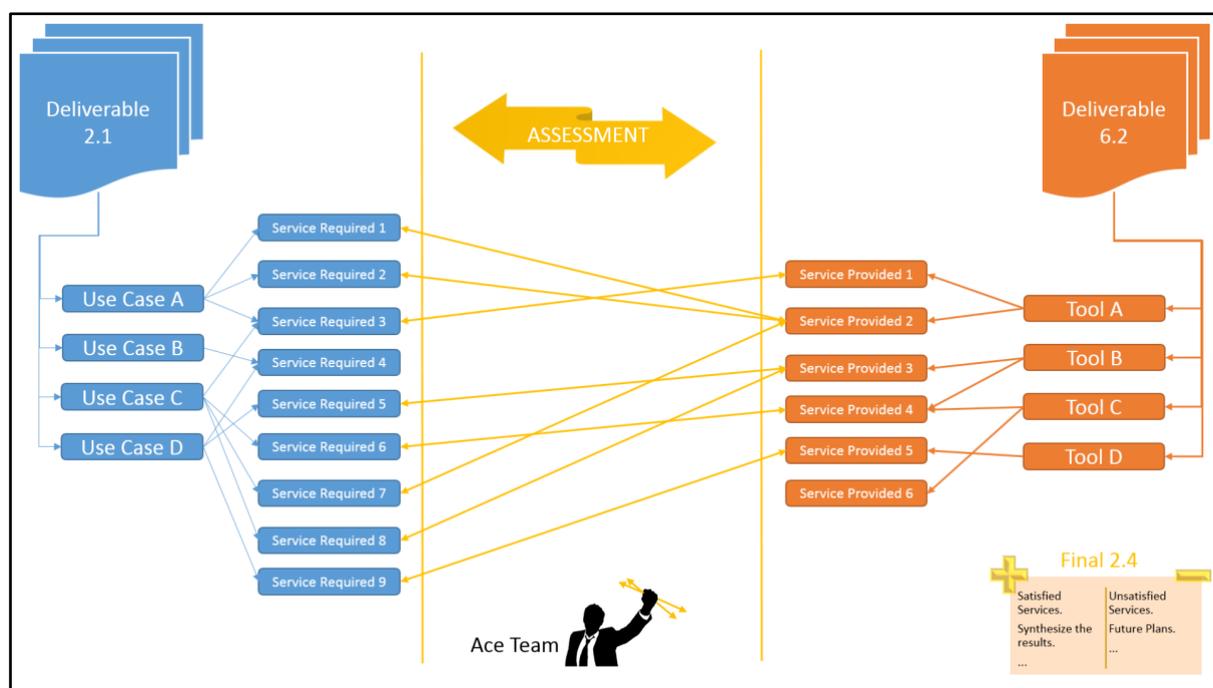

Figure 1. Matrix of WP2 requirements and WP6 outputs

To perform a good assessment it was important to consider different points of view in analysing and evaluating the same tool, and to adopt coherent indicators and metrics to

measure the performances of the different services we tested. Once the matrix of needs and requirements (extrapolated from WP2 deliverables) and services/features offered (from WP6 deliverables) has been established, a battery of tests has been performed by a team of highly trained experts (ACE)[10] coming both from the research and the ICT sectors. Eventually, a systematic gap analysis and a SWOT have been produced, to help better the understanding of (i) the level of adequacy of the PARTHENOS infrastructure with respect to the users' expectations and (ii) the opportunities and threats towards its further integrations and future developments.

2.1.1 KPIs used in the assessment

The quality model web used to assess the PARTHENOS resources is based on a selection of Key Performance Indicators (KPIs), describing the characteristics we considered to evaluate the success of a tool/service within a particular activity. Based on the methodology exposed in the Report on User Requirements and according to the outcomes of several PARTHENOS internal workshops - bringing together developers and researchers from different disciplines in the SSH domain - a list of indicators for the assessment was compiled:

Functionality: this parameter aims to measure how much a service/tool is adequate for the needs it wants to satisfy. A tool has to react and offer results in a satisfying way when used, avoiding potential off-topics or lack of results.

Scalability: when using a service/tool, a user has a set of more or less generic requests to be satisfied, depending on his expertise in a specific field. With respect to this, a resource has to be scalable, that is to say able to support the user in the most appropriate way, according to his level of knowledge.

Usability: an issue to consider, when dealing with services/tools, is that not all researchers are experts in computer science: a user should be able to find, manage, interact and share content using the tools provided by the infrastructure but - first and foremost - he has to clearly understand how to use them.

Layout/design: what the researcher sees of the whole mechanism of services/tools is a user interface. If this interface is easy to understand and well designed, according to the purpose the service aims to satisfy, the researcher will be more likely to use it, instead of preferring low hanging fruits: more traditional research procedures and/or old and - possibly - outdated tools he considers more adequate for achieving a task.

---

[10] ACE D4Science VRE https://services.d4science.org/group/ace/ (last visited 27/04/2018)

Accessibility: a tool also has to satisfy needs expressed from users with a variety of technical setups and/or persons with disabilities. There are guidelines and legislation (both at national and international level) to be followed depending on the service offered.

Documentation: this parameter is twofold. Primarily, if the service shares content with the user(s) (i.e. test data and/or other related files), it has to be documented and explained, so that the researcher is able to use it in the most appropriate way, according to his needs. Secondly, the service itself - at a technical level - has to be documented for sustainability purposes and to simplify maintenance, provide long term accessibility to resources and ease updates.

Language accuracy: often, tools that are tailored for a specific research domain are available in few languages. In PARTHENOS, researchers are coming from different areas of expertise across several EU countries, sharing content and resources in different languages. Multilingualism is a plus for tools that draw and share content with several European researchers.

Compatibility with project data: this parameter applies particularly to the PARTHENOS scenarios where communities and partners involved are invited to share data within the project: therefore, it has to be evaluated whether a tool is able or not to process this data, and at which level.

Security: this parameter - in PARTHENOS - is addressed by D4Science, the platform through which every service and tool in the project is shared and made accessible to users. Nevertheless, security is a crucial issue, both for research content shared via the platform, and/or personal data that may be needed to access a particular resource.

2.1.2 List of tools to be assessed

Service/Tool	Description	Responsible partner	Integration stage	User
CLARIN NLP tools	CLARIN NLP Tools are a lemmatizer for 24 different languages which and a service to create language models for this lemmatizer, as well as taggers, parsers, and named-entity recognizers for three different languages	CLARIN	Deployed	End User
D-NET Software Toolkit	Enabling framework for the realization and operation of aggregative metadata infrastructure	CNR-ISTI	Deployed	Infrastructure Admin
D4Science	D4Science is the framework allowing the creation and management of Virtual Research Environments (VREs), where the services are running.	CNR-ISTI	Deployed	All
Metadata Cleaner	Service for the harmonization of values according to controlled vocabularies	CNR-ISTI	Deployed: integrated with D-Net	Data Integrator
Metadata Inspector	GUI for the visualization of transformed metadata records and detection of "uncleaned" records	CNR-ISTI	Deployed: integrated with D-Net	Data Integrator
NERLiX	Set of tools to enable a service for Named Entity Recognition (NER)	OEAW-ACDH	Some of the tools (Stanbol wrapper) are deployed	End User
NERLiX	Set of tools to enable a service for Named Entity Recognition (NER)	OEAW-ACDH	Some of the tools (Stanbol wrapper) are deployed	End User
PARTHENOS Catalogue	The PARTHENOS Catalogue is a tool able to allow the project's partners to share their resources to make them easily searchable	CNR-ISTI	Deployed	All

Service/Tool	Description	Responsible partner	Integration stage	User
PARTHENOS Discovery Application	The PARTHENOS Discovery Application allows the end user to surf the projects resources using different navigation method. In addition to traditional textual research, in fact, it is possible to apply a series of filters, through the faceted search, which allow the user a better browsing experience and a better refining search	OEAW-ACDH	Deployed	End User
RUBRICA	RUBRICA aims to foster the interoperability and integration of various reference resources used in different disciplines. Starting from trusted knowledge bases (i.e.: databases, thesauri, authority lists etc.) researchers could create, merge, edit and reuse specialized reference resources, developed according to specific research purposes, without performing repetitive tasks on each resource	SISMEL, CNR-OVI	Deployed	End User
SPARQL endpoint	the SPARQL endpoint has been integrated to provide access - for collection managers - to the PARTHENOS contents.	CNR-ISTI	Deployed	End User
X3ML engine	Transformation engine capable of applying mappings defined via the 3M Editor.	FORTH	Deployed: integrated with D-Net	Data Integrator
X3ML toolkit	A set of open source components that assist the definition of mappings from XML to PARTHENOS Entities Model RDF for information integration.	FORTH	Deployed	Data Integrator

A. Data

3. The PARTHENOS Dataspace

3.1 Overview

A detailed overview on different aspects of data policies and related requirements, with a set of standardization solutions proposed by the PARTHENOS project can be found in deliverable D2.3 "Report on the Assessment of Data Policies and Standardization". The report gathered requirements and organized them to feed the technical implementation work, based on the needs already documented in other deliverables[11]. The report contains 2 focused assessments covering data policies and standardization and should be considered as the reference document as far as data related requirements are concerned. Therefore the scope of this section is not to double the work described above, providing additional evaluation, but to give a glimpse of the main characteristics of the context within which all the tools have been evaluated.

The development of modern technologies, in particular over the last two decades, has allowed the sharing of a critical amount of information which would have been previously unthinkable. Government agencies, research centres, GLAMs and other organizations have computerized their data [12] making them easily searchable for scholars and researchers. However, if on the one hand this opening up of data has triggered new opportunities, so on the other hand it made the creation of dedicated storage in which to share this information more urgent[13].

This need is particularly relevant for the digital humanities. Referring to these in a broader sense, they include several disciplines that have developed and codified specific standards within their fields (i.e. cataloguing standard or devoted metadata profile), that make it difficult to merge different formats of data and metadata.[14]

---

[11] D2.1 "Report on User Requirements"; D3.1 "Guidelines for Common Policies Implementation (draft); D4.1 "Standardization Survival Kit" and D4.2 "Report on Standardization"

[12] Cfr the neologism "datization" introduced by Viktor Mayer-Schönberger and Kenneth Cukier in their book Big Data, John Murray Publishers 2013

[13] Martin Hilbert *How much information is there in the "information society"?* Signifiance 9,4 (2012) https://doi.org/10.1111/j.1740-9713.2012.00584.x (last visited 27/04/2018)

[14] The standardization issue is extensively addressed by WP4 activities and deliverables.

Data spaces were born to meet this requirement. They, in fact, not only provide a common space for data (and metadata) produced by different institutions working in different fields, but they also help to address issues related to data integration, providing tool(s) able to present data and metadata in a homogeneous, reasonable and understandable manner for the end users.

PARTHENOS, from this point of view, is a perfect application in which to test the data space potential against the wide landscape represented by the research communities involved. Furthermore, the heterogeneity of the institutions involved, and the consequent heterogeneity of their data, made it necessary to create a virtual space in which data and metadata could speak a "common language" and which is able to maintain the added value (i.e. to preserve the original characteristics and values) of the original information provided.

The heterogeneity of the data, however, is not the only reason why it was decided to adopt a common application profile. Another aspect that can't be underrated, concerns the differences between data and metadata and the information they contain. Taking for granted that literally "nothing is born meta", we adopted this widely used terminology to underline the difference between content (actual data sets) and their descriptions (necessary for discoverability purposes). In particular, the data, in many cases, presents sensitive information that, to be shared with other parties, requires a set of mediations (i.e.: policies) and conditions to be met. So, great attention has been paid to the information to be included in the minimal metadata schema.

For these reasons, it was decided to adopt the CIDOC Conceptual Reference Model[15]. In addition to the possibility to map the identified minimum metadata from the PARTHENOS community, this Conceptual Reference Model allowed to semantically connect resources coming from different fields, such as – for example - archaeology and linguistics, paving the way towards the creation of a rich and interoperable digital ecosystem[16].

See D4.1: https://goo.gl/VK62qM (last visited 27/04/2018); D4.2: https://goo.gl/pwCLYS (last visited 27/04/2018)

[15] CIDOC-CRM http://www.cidoc-crm.org/ (last visited 27/04/2018)

[16] E. Degl'Innocenti et al. *Bridging the Gap between Material and Immaterial Cultural Heritage in PARTHENOS VREs,* poster presented at the 2017 Digital Humanities Benelux Conference

The CIDOC, in fact, was born to provide "definition and a formal structure for describing the implicit and explicit concepts and relationships used in cultural heritage documentation". However, the possibility to build a common semantic framework was not the only reason for which the CIDOC was adopted. By its nature, it is an extensible language and can be customized according to specific needs, such as the CRMarchaeo extension, supporting the archaeological excavation process and all the various entities and activities related to it.

For this reason, the technical group worked, in the first part of the project, to define which information was fundamental (and common) to the different fields of digital humanities involved in PARTHENOS.

Thus, the elaboration of an extension of CIDOC-CRM, able to provide all the identified common fields during the analysis phase, was focused to optimize the searchability of the resources not only for projects partners and expert of digital humanities but also for researchers and end users. After the definition of this project's milestone, in order to achieve this goal, it was necessary to implement a series of high level entities, linking them semantically.

The identified entities, as described in the deliverable 5.1[17] were:
- **Projects**: a project is defined as an ongoing effort maintained by some group, specifically formed to pursue a certain aim over a given time span;
- **Services**: a service is defined as the continued, declared willingness and ability of an actor to execute - on demand by a client - certain activities of specific benefit to the client;
- **Digital objects**: a digital object is an information object represented as sets of bit sequences, which can have either a persistent or volatile nature;
- **Datasets**: a dataset is a set or collection of data, records or information that is kept as a persistent unit of information in the knowledge generation process;
- **Software**: a software is an artefact that can be executed on a computer to perform specific operations;

---

[17] Report on the common semantic framework http://www.parthenos-project.eu/Download/Deliverables/D5.1_Common_Semantic_Framework_Appendices.pdf (last visited 27/04/2018)

- **Actors**: an actor is either an institution, a team or an individual person that participates in the research infrastructure as partner providing data and/or services;
- **Knowledge generation processes**: a knowledge generation process represents the workflow, setup and processes involved in the production of specific datasets.

The final result of this preliminary analysis and the identification of the high level entities, led to the establishment of the PARTHENOS Entities Model (PEM)[18] extension which - as described in deliverable 5.1:

"...propose an ontological model and RDF schema to encode data of use in supporting the activities and aims of research infrastructures to pool and connect services, software, datasets and to enable users of such services to reach the actors and understand the knowledge generation processes which generated the offered datasets. Research infrastructures integrate highly heterogeneous resources for an often equally heterogeneous audience."

The actions described so far, contributed to address some of the most relevant issues related to the establishment and management of a heterogeneous data space such as the PARTHENOS one, nevertheless it was also necessary to improve the metadata quality. A controlled vocabulary was used to support the data normalization process enhancing classification. In order to solve this problem, the technical group of PARTHENOS customized a solution already adopted within the DARIAH[19] network: the Back Bone Thesaurus (BBT)[20], originally built to create a trusted reference resource able to cope with the terminological heterogeneity of the Digital Humanities. The BBT identified:

> "five basic requirements for the generation of sustainable and effective meta-vocabulary: the adoption of a semantic approach, a clear method to semantic division, creation of top level terms based on bottom up analysis of existing classificatory systems, open ended development of complete vocabulary including top terms and the ability to carry out this work as a distributed collective project."

Based on this vision, the PARTHENOS technical group started to work on the creation of a customized thesaurus, working towards the creation of a vocabulary interoperable with the PARTHENOS Entity Model and able to deal with the semantics of the cultural objects provided by the project's partners.

3.2 Requirements

In the following tables, we gathered and listed the requirements expressed by partners and communities in D2.1, with particular focus on the ones regarding data/metadata and their use within the project. Each table shows:

- the **partner** that expressed the requirement;
- the **actor**, that is who is performing the action described and needs a particular service;
- the **requirement**, that describes what is needed;
- the **explanation**, in order to have a clearer understanding of the requirement;
- when available, the **Use Cases ID** which the requirement is referred to, from D2.1.

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.2.: offering data definition, manipulation, and access languages
Explanation	The framework should provide a language for the definition of Enhanced Publications data models (EP-DMDL, EP Data Model Definition Language).

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.5.: Supporting data portability
Explanation	Support is needed for open standards for the representation of data

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.13.: Availability
Explanation	Requirements are needed which state that some information or resource can be used at any point in time when it is needed and its usage is authorized.
Use Cases ID	UR2

Partner	PIN
Actor	Data provider
Requirement from D2.1	3.8.1.1., 3.8.1.2.:Data accessibility
Explanation	The required data(sets) are available in an uncomplicated way. Content should be provided using the Creative Commons licence suite.
Use Cases ID	AR_03, AR_07, AR_04

Partner	PIN
Actor	Data consumer
Requirement from D2.1	3.8.1.4.: Metadata quality
Explanation	The available data(sets) are well described
Use Cases ID	AR_08, AR_09

Partner	PIN
Actor	Archive manager
Requirement from D2.1	3.8.1.5.: Data quality
Explanation	The available data(sets) are complete and well organised
Use Cases ID	AR_06

Partner	PIN
Actor	Data consumer
Requirement from D2.1	3.8.1.6.: International dimension
Explanation	Having easy access to international data(sets)
Use Cases ID	AR_03, AR_08

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.1.: Improvement of research data storage
Explanation	Fields of required improvements include protection of (dynamic) data during the research project phase; storage of (static) data after the research project phase; easy access to data stored; creation of sustainable data.
Use Cases ID	WRW1

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.2.: Enabling data re-use
Explanation	Ensuring free and easy access to research data so that it can be re-used by others.
Use Cases ID	WRW2

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.3.: Quality assurance of research data/datasets by the community
Explanation	Peer-reviewed descriptions of datasets, enabling comments by users (to be published with the dataset), open access availability and citation possibilities of datasets, code of conduct for researchers on data management and availability of datasets.
Use Cases ID	WRW3

Partner	CLARIN
Actor	Librarian, archivist
Requirement from D2.1	3.8.2.4.: Digitally available information on: access to historical sources, the existence of historical sources, their contents.
Explanation	The provision of the named information by libraries and archives and the sharing of these data should be augmented, so that researchers can be sure that they don't miss sources relevant for their research
Use Cases ID	Cen1

Partner	CLARIN
Actor	Librarian, archivist
Requirement from D2.1	3.8.2.5.: Ontologies, selection criteria
Explanation	Ontologies which reflect existing classifications and vocabularies used by researchers working on a certain topic (here: World War 1, Medieval Manuscripts); selection criteria for organising and displaying information.
Use Cases ID	Cen2

Partner	CLARIN
Actor	Librarian, archivist
Requirement from D2.1	3.8.2.17.: Optimization of data collection from corpora
Explanation	Time and learning curve are obstacles in learning the existent variety of query languages for average non-technical researchers; a way to cater for this is needed, either by providing a common query language or by ensuring interoperability at the level of query interpretation.
Use Cases ID	D3C-3

3.3 KPIs specific for data evaluation

For the PARTHENOS data space evaluation, the FAIR principles[21] were adopted as quality model[22]. The FAIR principles are guidelines for enhancing the re-use of scholarly data and to ensure a sustainable access to data. Their adoption also increases the ability of both machines and humans to find and re-use data. The guidelines consist of four core principles: (i) Findability, (ii) Accessibility, (iii) Interoperability, and (iv) Reusability. Therefore, in the FAIR Data approach, data should be:

---

[21] M.D. Wilkinson, M. Dumontier, B. Mons et al. *The FAIR Guiding Principles for scientific data management and stewardship* Scientific Data 3 (2016) doi:10.1038/sdata.2016.18

[22] See D2.3 Report on the assessment of data policies and standardization, cit.

- **Findable** – Easy to find by both humans and computer systems and based on mandatory description of the metadata that allow the discovery of relevant datasets;

- **Accessible** – Stored for long term such that they can be easily accessed and/or downloaded with well-defined licensing and access conditions (Open Access when possible), whether at the level of metadata, or at the level of the actual data content;

- **Interoperable** – Ready to be combined with other datasets by humans as well as computer systems;

- **Reusable** – Ready to be used for future research and to be processed further using computational methods.

B. Infrastructural Services and Tools

4. Tools enabling Interoperability

4.1 Overview

The PARTHENOS digital infrastructure has been designed and deployed on top of two main software components: D-NET and D-4Science. While D4Science allows the creation and the full management of Virtual Research Environments (VREs), with a focus on the social networking services, D-NET provides the machinery for the creation of information systems where organizations and researchers can find the means to integrate their data sources, establish uniform and valuable information spaces of object's metadata descriptions - including data mediation, data mapping, data storage and indexing, data curation and enrichment, and data provision. D-NET provides also a Metadata Inspector service and a Metadata Cleaner service: the former offering a low-level portal with search and browse features, through which a user might check his own data post-transformation and explore the XMLs of the entities generated from his initial XML records; the latter fostering the harmonization of values (normalization), according to controlled vocabularies and other reference resources.

4.1.1 X3ML Toolkit

Following the evolution of the semantic web, a common ontology is required in order to provide convergent data. In order to guarantee a uniform semantic data consistency, supporting both interdisciplinarity and interoperability efficiently, PARTHENOS provides an appropriate service that realizes the transformation of the source records to the target format (CIDOC-CRM, the PARTHENOS common semantic framework): 3M - the Mapping Memory Manager[23]. 3M is a package for managing mapping definition files. It's based on the FIMS[24] management system for the files administration and on the 3MEditor[25] for editing and viewing them.

Currently there are several components providing support for collection managers and other technical collaborators (infrastructure administrators, metadata providers etc.) to

---

[23] 3M Mapping Memory Manager https://mapping-d-parthenos.d4science.org/3M (last visited 27/04/2018)
[24] FIMS https://github.com/isl/FIMS (last visited 27/04/2018)
[25] 3MEditor https://github.com/isl/3MEditor (last visited 27/04/2018)

gather content into the PARTHENOS dataspace. The components already integrated in the platform are (see table 1):

- X3ML Toolkit;
- D-NET Software Toolkit;
- X3ML Engine;
- Metadata Cleaner;
- Metadata Inspector.

The X3ML Framework realizes the transformation of the source records to the target format: the engine takes as input the source data (currently in the form of an XML document), the description of the mappings in the X3ML mapping definition file and the URI generation policy file and is responsible for transforming the source document into a valid RDF document, which corresponds to the input XML file, with respect to the given mappings and policy.

4.1.2 DNET toolkit

The D-NET software toolkit is a general-purpose, service-oriented framework for the construction of customized Aggregative Data Infrastructures (ADIs). ADIs addresses the need - increasingly manifested by research communities - to operate over the integration of contents collected from several data sources (such as institutional repositories endowed with OAI-PMH[26] interfaces, or archives of research data). Typically, such infrastructures provide applications for:

- Managing a federation of input data sources (e.g., OAI-PMH, JDBC, FTP): data source registration and configuration;
- Aggregating data to form uniform Information Spaces: data collection, data conversion (mapping, transformation, cleaning, etc.), curation, and enrichment
- Providing the aggregated data to end-users or third-party consuming systems: web user interfaces for data discovery and standard APIs to access the data (e.g., OAI-PMH, HTTP Search API).

---

[26] OAI-PMH https://www.openarchives.org/pmh/ (last visited 27/04/2018)

D-NET provides a service-oriented framework where data infrastructures can be constructed in a LEGO-like approach, by selecting and properly combining into data processing workflows the D-NET services required for the implementation of research community requirements.

4.1.3 Metadata Inspector

The Metadata Inspector is a Web GUI integrated into D-NET that provides data curators with an overview of the information space, where they can search and browse records and verify the correctness of the transformation phase (e.g. no mapping mistakes or semantic inconsistencies, no records marked as 'invalid' by the Metadata Cleaner). Upon positive verification of the records in the information space, data curators can inform the PARTHENOS infrastructure administrators that the records can be published. Figures 2 and 3 show two screenshots of the Metadata Inspector: the main search form and an example of search results containing non-cleaned fields.

Figure 2. The main search form of the Metadata Inspector
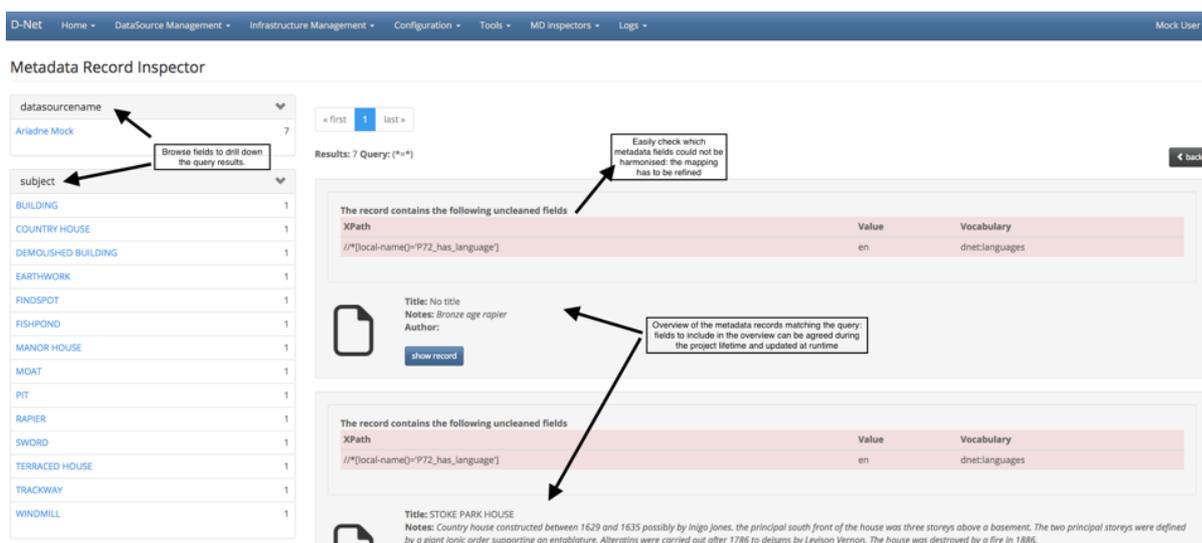
Figure 3. The Metadata Inspector shows metadata records with "uncleaned" fields

The Metadata Inspector features the following functionalities:

- **Generic search**: the main search form allows a data curator to search for metadata records containing specific values (All fields search box);

- **Advanced search**: the main search form allows a data curator to search for metadata records containing specific values in specific parts of the transformed XML records. If the data curator specifies different search criteria, those are considered to be combined with "AND" semantics. Currently, the following criteria can be used to perform an advanced search on RDF records compliant to the PARTHENOS Entities Model:

  - **Title**: the given value is matched against values in E35_Title/rdfs:label;
  - **Original Identifier**: the given value is matched against the identifier of the record as provided by the original data source;
  - **D-Net Identifier**: the given value is matched against the identifier of the record generated by the aggregator;
  - **URL**: the given value is matched against the rdf:about of an RDF resource;
  - **Cleaned records**: choose false to select RDF records containing values that could not be cleaned by the Cleaner service;
  - **Data Source**: select one of the aggregated data source;
  - **Data Source API**: one data source may provide metadata records via different endpoints and different protocols;
  - **Type**: select one of the available rdf:type.

If needed, the main search form can be configured with additional search fields. By clicking on "Search", the Metadata Inspector shows the list of RDF files matching the search criteria. An overview of each RDF file is given for reference purposes.

4.1.4 Metadata Cleaner

The Metadata Cleaner is a D-NET service that harmonises values in metadata records based on a set of cleaning rules. Currently, two types of cleaning rules are supported:

- **Groovy rule**: specifies an xpath and a groovy closure. The groovy closure is applied to the value of the XML element of a record identified by the given xpath. The element value is replaced with the new value;

- **Vocabulary rule**: specifies an xpath and the identifier of a D-Net thesaurus. A D-NET thesaurus consists of a vocabulary that is a list of authoritative terms together with associations between terms and their synonyms. Data curators – typically based on instructions from data providers and domain experts – are provided with user interfaces to create/remove vocabularies and edit them to add/remove new terms and their synonyms. Given a metadata format, the metadata cleaner service can be configured to associate the metadata fields to specific vocabularies. The service, provided records conforming to the metadata format, processes the records to clean field values according to the given associations between fields and vocabularies. Specifically, field values are replaced by a vocabulary term only if the value falls in the synonym list for the term. If no match is found, the field is marked as 'invalid'. The 'invalid' mark is exploited by the Metadata Inspector (see dedicated section below) to highlight non-cleaned records and suggest the update of D-NET thesauri or the update of the values in the input record. Currently in the PARTHENOS ADI, only vocabulary rules are used.

4.2 Requirements

In D2.1, the collection of requirements on interoperability was very articulated and touched upon several relevant topics. It included also requirements regarding infrastructural services and tools, that were taken in consideration according to the use cases expressed by WP2 partners. The following is a list of requirements regarding infrastructural services and tools:

Partner	KNAW-NIOD
Actor	Research Infrastructure admin
Requirement from D2.1	3.7.1.10: Data interoperability and data integration
Explanation	Being able to upload/harvest/integrate data from CHI into EHRI

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.8.: Support the integration of content
Explanation	Due to the heterogeneity of the content, a transformation and harmonization module is necessary in order to massage the incoming material and transform it in a homogeneous format, so that further operations can be performed on content without tackling again the peculiarities of each data source.

Partner	KNAW-NIOD
Actor	Research Infrastructure admin; access management
Requirement from D2.1	3.7.1.13.: Services to be able to authenticate and identify users and set authorization levels.

Partner	ICCU-MIBACT
Actor	Content Provider
Requirement from D2.1	3.8.1.7.: Metadata interoperability via mapping tool
Explanation	Enable an automatic mapping between different standards to ensure the best dissemination of research data, providing services of data checking, data preview and data enrichment.
Use Cases ID	MINT 01, MINT 02, MINT 03, MINT 04, MINT 05

Partner	ICCU-MIBACT
Actor	Content provider/ Data collection manager
Requirement from D2.1	3.8.1.8.: Metadata acquisition and interoperability via OAI-PMH repository
Explanation	Enable the acquisition of metadata from an OAI-PMH repository in a specified format providing services of metadata validation, reporting, update and managing of invalid metadata.
Use Cases ID	CI_01, CI_02, CI_03, CI_04, CI_05

4.2.1 Functional Use Cases

After the collection of use cases and requirements, we started a mapping activity to fit them into the PARTHENOS architecture. By mapping the use cases provided by partners against the functional ones – distilled by the technical team – we made sure that the cross-domain functionalities requested by the researchers were fully covered by the PARTHENOS architecture. This architecture allows the provision of services supporting actual research practices (i.e. actual knowledge generation processes), ensuring the scientific reliability of the contents and representing their provenance, thus preserving the chain of trust.

All the relevant information related to the PARTHENOS contents are collected, identified, described and connected in the registry that will also provide cross-domain search services on the registered entities. To match the above vision and provide input for the registry establishment and the PARTHENOS infrastructure implementation, all the requirements and use cases have been mapped against a set of 21 functional (i.e.: more abstract) use cases, implementing the following cross-domain functionalities:

- Entities registration
- Registered entities access
- Creation of domain specific VREs
- Use of domain specific VREs
- Metadata aggregation and export
- Resources curation

Registering people, services, data, metadata and software		
Manual registration entities in the PARTHENOS registry	People Services Data Metadata Software Research infrastructures	Web interface

Functional Use Cases
USE CASE REG_01: Manual registration of an entity in the PARTHENOS registry **ACTOR:** Research Infrastructure admin (RI admin), PARTHENOS curator **DESCRIPTION:** RI admin or a PARTHENOS curator wants to register a PARTHENOS entity in the registry to make it discoverable through the PARTHENOS infrastructure. The registration procedure should be performed via a graphical web interface that requests the information needed for the entity registration. Mandatory and optional fields requested by the graphical user interface vary based on the type of entity to be registered.

Aggregating and exporting metadata from RIs (AGGR)	
USE CASE AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud	
GOAL	Metadata aggregation and dissemination via OAI-PMH Check of metadata aggregated via the MINT mapping tool Data aggregated in MINT are enriched by external SKOS thesauri Automatic check provided by the harvester system during the ingestion process Discard invalid metadata Update metadata of a content provider

USE CASES from Archaeology, Heritage and Applied Disciplines	Aggregation of different metadata sets, providing them as one unified set (MINT_01) Checking metadata (MINT_03) Enriching metadata aggregated in MINT using external SKOS thesauri (MINT_05) Metadata validation (CI_02) Discard invalid metadata (CI_05) Repository update (CI_03)
USE CASE AGGR_02: Export metadata via standard protocols	
GOAL	Metadata aggregation and dissemination via OAI-PMH
USE CASES from Archaeology, Heritage and Applied Disciplines	Aggregation of different metadata sets, providing them as one unified set (MINT_02)

Curating PARTHENOS resources (CURA)	
USE CASE CURA_01: Subject coverage	
GOAL	Update metadata of a content provider
USE CASES from Archaeology, Heritage and Applied Disciplines	Repository update (CI_03)
USE CASE CURA_03: Invite curation	
GOAL	Inspect and enrich visual documents stored in one of the catalogues available in the PARTHENOS portal Data aggregated in MINT are enriched by external SKOS thesauri Update metadata of a content provider
USE CASES from Archaeology, Heritage and Applied Disciplines	Inspect and enrich visual media documents (AR_07) Enriching metadata aggregated in MINT using external SKOS thesauri (MINT_05) Repository update (CI_03)

5 VREs Management

5.1 Overview

A Virtual Research Environment (VRE) is a Web-based working environment that it is tailored to serve the needs of a Community of Practice, providing a whole array of commodities needed to accomplish the community's goal. The VRE is open and flexible with respect to the overall service offering and promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance, and attribution.

VRE Management includes the activities performed during the VRE operation in order to revise the behaviour of the application. It includes:

- the revision of the VRE user interface;
- the revision of the Information Space;
- the management of the VRE users and roles;
- the monitoring of the status of the resources partaking to the VRE.

There are different access policies available for PARTHENOS VREs:

- VREs can be free to use (Open access policy), meaning that they are open to the public and users can enter by simply accepting their Terms of Use;
- VREs can also have Restricted access policy: in this case users can ask for access by clicking on the "Request Access" button. An authorised person or a group will examine the request and grant access;
- If a VRE follows a "Private" access policy for private to use, other users can join only upon invitation by the VRE owner.

The VRE layout characterises the arrangement of the VRE graphical user interface, specifying how the user interface main constituents are arranged graphically in tabs, sub-tabs, rows and columns.

The D4Science portal is based on the Liferay[27] portal and uses Liferay's base functionality to manage the layout of each VRE, which can be seen as an application conceived to

---

[27] Liferay portal http://www.liferay.com/ (last visited 27/04/2018)

serve the needs of a well defined pool of people, its users. In order to manage the pool of users of a VRE, the VRE Manager should use the Users Management Portlet, a user friendly interface through which:

- manage request for membership;
- revise the roles a user is entitled to play;
- remove users.

VRE Monitoring is the action through which the VRE Manager checks the status and quality of the resources partaking of the VRE. This activity is performed by using the underlying gCube Monitoring System.
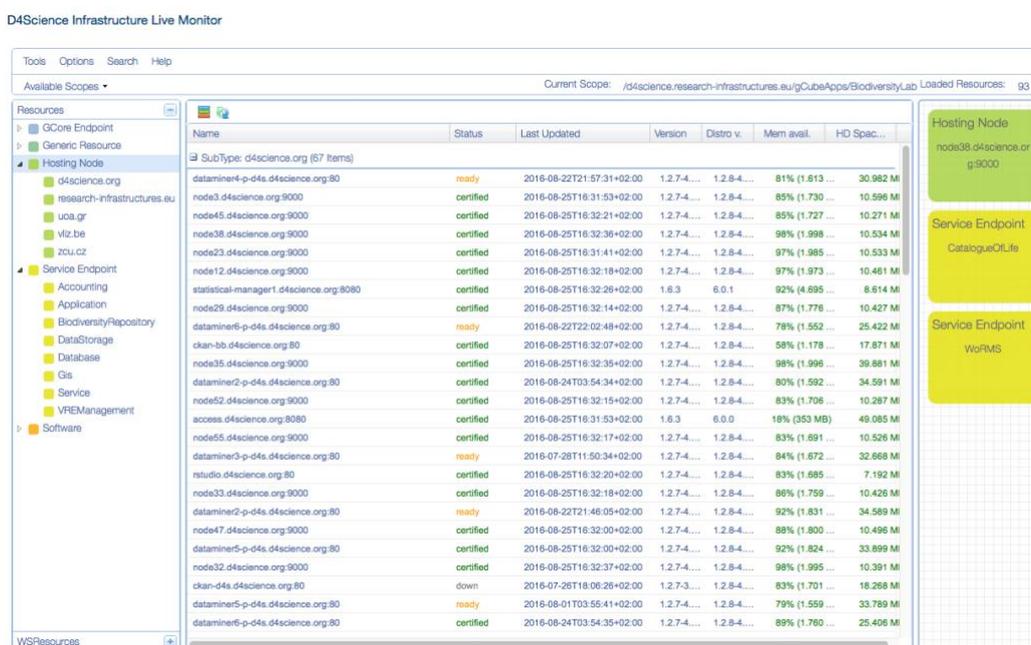
Figure 4. Example of a gCube Monitoring System view

5.2 Requirements

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.3.: Being able to use tools to crowdsource translation of (archival) documents

Partner	KNAW-DANS
Actor	VRE Manager

Requirement from D2.1	3.7.2.1.: Supporting different back-ends for data storage
Explanation	The framework should provide a storage management module for the configuration of the storage back-ends to be used. Depending on the functional requirements of the target Enhanced Publication Information System (EPIS), a type of back-end, may be preferable to another.

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.6.: Supporting the integration of heterogeneous data sources
Explanation	Data sources export different typologies of content according to different formats and via different protocols. EPMSs should support developers in the integration of such diverse content.

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.7.: Support the management of dynamic data sources
Explanation	Data source management functionality is needed to ease the administrative operations needed to take care of the dynamic nature of the data sources.

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.10.: Support the enrichment and curation of content
Explanation	To create high quality content it is needed to better the quality of the EPs and enrich the original content.

Partner	PIN
Actor	Archive manager
Requirement from D2.1	3.8.1.3.: Data accessibility
Explanation	The portal should ensure storing space and long-term availability of data.
Use Cases ID	AR_05, AR_06

5.2.1 Functional Use Cases

Setting up Virtual Research Environments		
Set up, user authentication, services integration in domain specific Virtual Research Environments	People Services Data Metadata Software Research infrastructures	Web interface
Functional Use Cases		
USE CASE VRESET_01: Set-up of a domain-specific VRE **ACTOR:** PARTHENOS infrastructure admin **DESCRIPTION:** The PARTHENOS infrastructure admin should be able to create and configure a new VRE to serve a specific user community represented in		

PARTHENOS. For example, there could be a VRE for linguistics, one for historians, one for social scientists, etc.

The PARTHENOS infrastructure admin configures a VRE by selecting datasets and services relevant to the VRE domain that are available in the PARTHENOS registry. He also invites users to join the VRE and configure their authorization levels for the newly created VRE.

USE CASE VRESET_02: Integration of services in the PARTHENOS infrastructure
ACTOR: Research infrastructure (RI) admin, Service Provider (SP), PARTHENOS Infrastructure (PI) admin.
DESCRIPTION: A research infrastructure offers a service and the infra admin would like to integrate it into one of the PARTHENOS domain-specific VREs. The RI admin mediates between PI admin and SP, to prepare the service for integration according to the instructions provided by the PI admin. Once integrated in the PARTHENOS infrastructure and registered in the PARTHENOS registry, the service can be activated for one or more domain-specific VREs.

USE CASE VRESET_03: VRE authentication and authorization
ACTOR: VRE manager
DESCRIPTION: A VRE manager wants to enable access to authenticated users only and configure the authorization levels required to perform actions on the VRE.

Using Virtual Research Environments		
Entities referencing, files deposition, resources sharing, dataset processing and results presentation	People Services Data Metadata Software Research infrastructures	Web interface
Functional Use Cases		
USE CASE VREUSE_04: Private and public sharing of resources deposited in the VRE workspace **ACTOR:** VRE user **DESCRIPTION:** A VRE user deposited a file in the VRE workspace (see USE CASE VREUSE_03: Deposition) and wants to share it with his/her collaborators. The VRE user should be able to share the file with other VRE users (collaborators registered in the VRE) or with external collaborators not registered in the VRE.		
USE CASE VREUSE_05: Process a dataset and publish results **ACTOR:** VRE user **DESCRIPTION:** A VRE user finds a dataset of interest for his/her research and wants to process it via a service. If the service is available in the PARTHENOS VRE the user belongs to (see USE CASE VRESET_02), the VRE user is allowed to start the processing and the VRE supports him/her in depositing (and optionally publishing) the results in the VRE (see USE CASES VREUSE_03 and VREUSE_04). If instead the service is available via one of the research infrastructures in the PARTHENOS ecosystem, the user will receive instructions on how to access the service. Results will be publishable and shareable according to USE CASES VREUSE_03 and VREUSE_04.		

6 Social Networking Services

6.1 Overview

The Social networking services complement the PARTHENOS infrastructure by promoting cooperation among users. Specifically, they allow posts sharing and discussions (i.e.: threads) creation functionalities, enabling users to comment, subscribe or re-share these resources. Attached to the content of a post, generally a text, one can add interesting links, also pointing to deposited files in the VRE workspace. Such facilities are implemented by benefitting from the underlying infrastructure to realise an elastic behaviour, i.e. the services realising them exploit the infrastructure to acquire computing capacities on demand. By leveraging social computing and cloud computing these services realise an integrated - web-based - working environment where scientists have at their fingertips what is needed to accomplish a scientific investigation.

6.2 Requirements

Partner	KNAW-NIOD
Actor	Research Infrastructure admin
Requirement from D2.1	3.7.1.11.: Being able to share and collaborate with other researchers on EHRI documents

Partner	KNAW-DANS
Actor	VRE Manager
Requirement from D2.1	3.7.2.4.: Enabling data sharing
Explanation	Need for supporting the export of content via different standard APIs and protocols to serve third-party applications.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.21.: Facilitate sharing, categorising, and indexing research questions and/or topics
Explanation	Users of EHRI would benefit from having access to these along with the sources selected to assist in answering a question or addressing a particular topic.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.22.: Facilitate sharing, categorising, and indexing additional information about a research project
Explanation	Users of EHRI would benefit from having access to these along with the sources selected to assist in answering a question or addressing a particular topic.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.23.: Facilitate sharing, categorising, and indexing notes and annotations on sources at various levels
Explanation	They are perceived as valuable for research

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.24.: Facilitate sharing, categorising, and indexing details (citations) of researchers' publications
Explanation	To assist in the 'chaining' process of moving from published works to other works, and to archival sources

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.25.: Facilitate sharing, categorising, and indexing of researcher bibliographies

Partner	ICCU-MIBACT
Actor	Data collection manager
Requirement from D2.1	3.8.1.9.: Metadata sharing by collection manager via OAI-PMH repository
Explanation	Enable the harvester to share data acquired by other content providers via OAI-PMH repository with different metadata formats
Use Cases ID	CI_06

6.2.1 Functional Use Cases

<table>
<tr>
<td colspan="3" align="center">Using Virtual Research Environments</td>
</tr>
<tr>
<td>Entities referencing, files deposition, resources sharing, dataset processing and results presentation</td>
<td>People

Services

Data

Metadata

Software

Research infrastructures</td>
<td>Web interface</td>
</tr>
<tr>
<td colspan="3" align="center">Functional Use Cases</td>
</tr>
<tr>
<td colspan="3">USE CASE VREUSE_04: Private and public sharing of resources deposited in the VRE workspace
ACTOR: VRE user
DESCRIPTION: A VRE user deposited a file in the VRE workspace (see USE CASE VREUSE_03: Deposition) and wants to share it with his/her collaborators. The VRE user should be able to share the file with other VRE users (collaborators registered in the VRE) or with external collaborators not registered in the VRE.</td>
</tr>
<tr>
<td colspan="3"></td>
</tr>
</table>

C. General Purpose Services and Tools

7. Global search tools

7.1 Overview

Given the huge amount of datasets that are expected to be available in the PARTHENOS data space, global search tools are a crucial element for finding useful datasets, relevant to address a given research question. Such a global search facility can rely on different approaches:

- an explorative way, by having a simple input field where search terms are entered. This gives back a list of resulting datasets, that is ideally sorted by relevance for a researcher.
- a complex search form, where several entities are chosen to deliver a specific search result set. An example is a search for datasets that are connected with persons that have a specific surname. Another condition could be a connection to a specific place.

Both approaches have their advantages and disadvantages. The explorative search can lead to surprising useful datasets but also to disappointing results. On the other hand, the complex search depends on a basic understanding of the relation between datasets and entities. On such a basis, it can deliver accurate results. Allowing both means of searching improves the search experience for users. It is important, anyway, that the usage of the search tools is clearly communicated and that the results are reproducible. In the case of the complex search, good documentation is needed, so that users can understand the mechanics of how datasets are connected with entities. Having a training module on doing well-formulated search would be a bonus. Finally, a cleverly designed user interface that supports users in the exploration of the PARTHENOS data space is of great value.

As of January 2018, a set of applications and tools providing access to global resources (i.e. not specialised services) has been deployed and integrated in the PARTHENOS infrastructure. Three applications are available for the end-user:

- The PARTHENOS **Catalogue**

- The PARTHENOS Discovery Application
- The SPARQL endpoint

The PARTHENOS Catalogue is a tool able to allow the project's partners to share their resources to make them easily searchable. The PARTHENOS Catalogue contains options for browsing according to resource types, groups providing resources, formats and keywords (tags). Similarly, faceted search in combination with a simple full-text search is possible. The faceted search additionally allows for filtering of locations and licenses. Currently, there are only prototypical data records available. However, these records already show which information will be available per record, i.e. the name of the resource, a description, tags for its classification, licensing information as well as information about author and creation of the record. The resource itself is linked. Different record types are expected, i.e. records describing procedures, actors, things, activities and resource types. Statistics are available granting insights on the total number of items, item revisions per week, to rated items, most edited items, largest groups, top tags, and users creating most items.
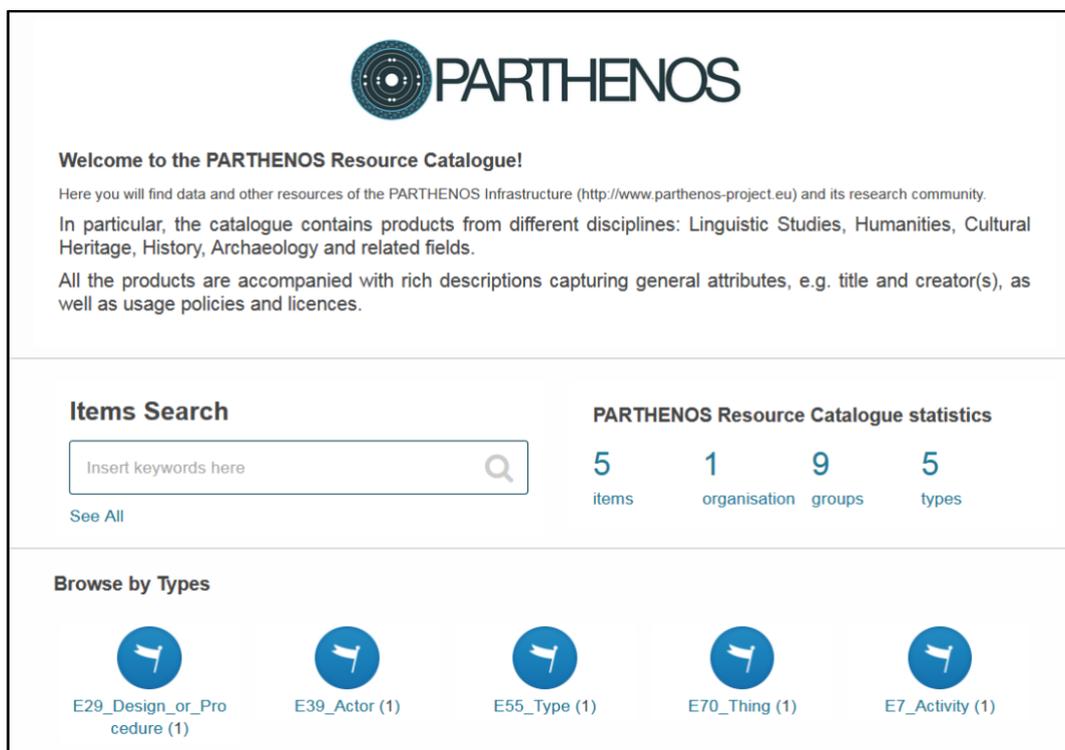
Figure 5. PARTHENOS Resource Catalogue front page

The registry adopted a common data model, the CIDOC-CRM and an extension realized for the project, the PARTHENOS Entities Model (PEM). In this way, both high-level

information (i.e. Project, services) and more detailed information on cultural objects (i.e. books) can be included in the registry, even if they come from very different disciplines in the landscape of the digital humanities. Moreover, it provides also the browsing of the resources with facets, so it is possible for the projects partners to easily surf the information they have provided.

The PARTHENOS Discovery Application allows the end user to surf the project resources using different navigation methods. In addition to traditional textual search method, it is possible to apply a series of filters, through the faceted search, which afford the user a better browsing experience and a more refined search. This system is widely adopted in the field of Digital Humanities because it is possible, for end users, to extract in a few steps the information they are looking for.

Finally, the SPARQL endpoint has been integrated to provide access - for collection managers - to the PARTHENOS contents. The SPARQL endpoint uses the RDF language and describes the relations between objects allowing the creation of customized queries based on the semantic structure of the data that can be downloaded in various formats.

There are a couple of other tools in development by Task 6.5 ("Resource Discovery") of PARTHENOS WP6. This tools will expand the scope of the global search and will help to discover useful resources for researchers. Two of these approaches are PAVLOV and the integration of ResearchSpace.

PAVLOV is a fork of the CLARIN Virtual Language Observatory (VLO)[28] that uses the PARTHENOS Entities to allow searching and discovering datasets in the PARTHENOS dataspace. It integrates a text search combined with a faceted search.

---

[28] CLARIN VLO: https://vlo.clarin.eu/ (last visited 27/04/2018)

Figure 6. Screenshot of a search in PAVLOV

PAVLOV will allow an easy and reproducible resource discovery that conforms to the PARTHENOS Entities Model. It aims to be a main entry point for researchers that are searching for resources in the PARTHENOS data space. Functionality, scalability, usability and layout/design of the global search will particularly benefit from PAVLOV. There will be also expert views, where experienced users can look into the details of a dataset, that is expressed in RDF. By integrating the RDFVisualizer there is also a tool at hand that will help to raise the compatibility of project data for users, and for tool developers.
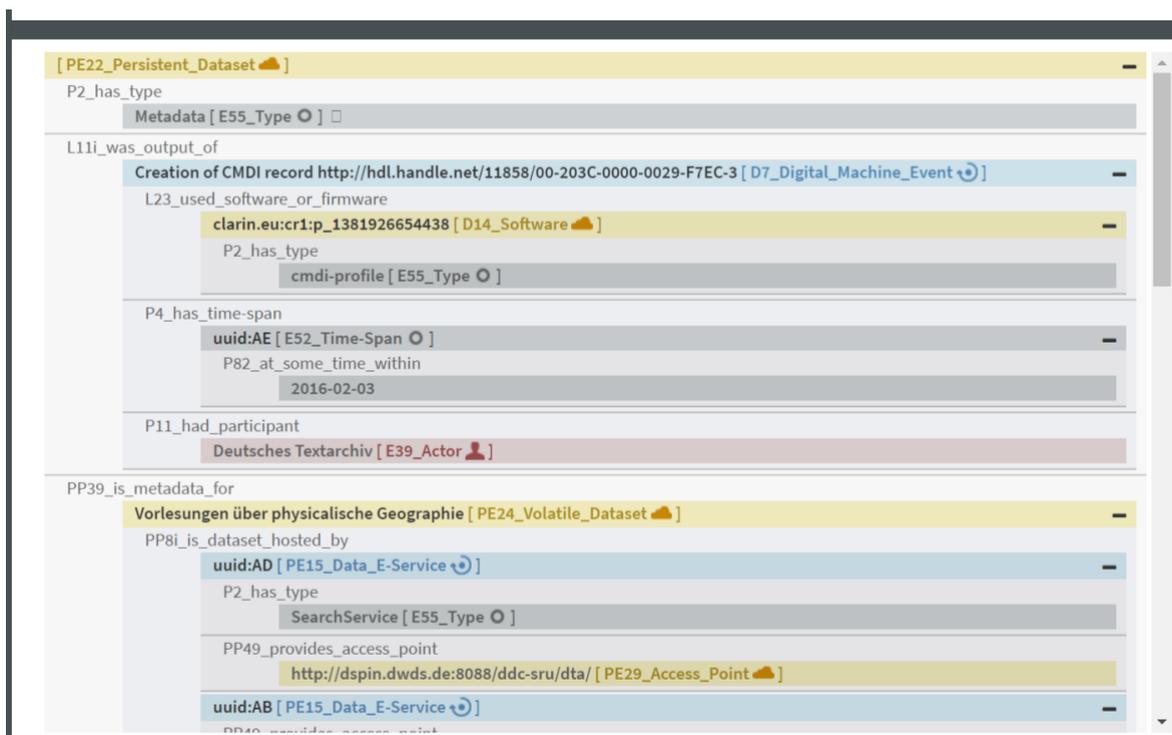
Figure 7. Screenshot of the RDFVisualizer integrated into PAVLOV

The second effort is the integration of ResearchSpace[29], a tool used by the British Museum. It allows the integration of Semantic Web technologies and to connect different entities, expressed in graphs and is, therefore, adaptable for the PARTHENOS Entities Model. ResearchSpace will allow the building of SPARQL queries in a user friendly query builder. This will lower the technical barrier for users for exploring the PARTHENOS data space and to do very specific global search, tailored to the needs of a researcher.

---

[29] Research Space: http://researchspace.org/ (last visited 27/04/2018)
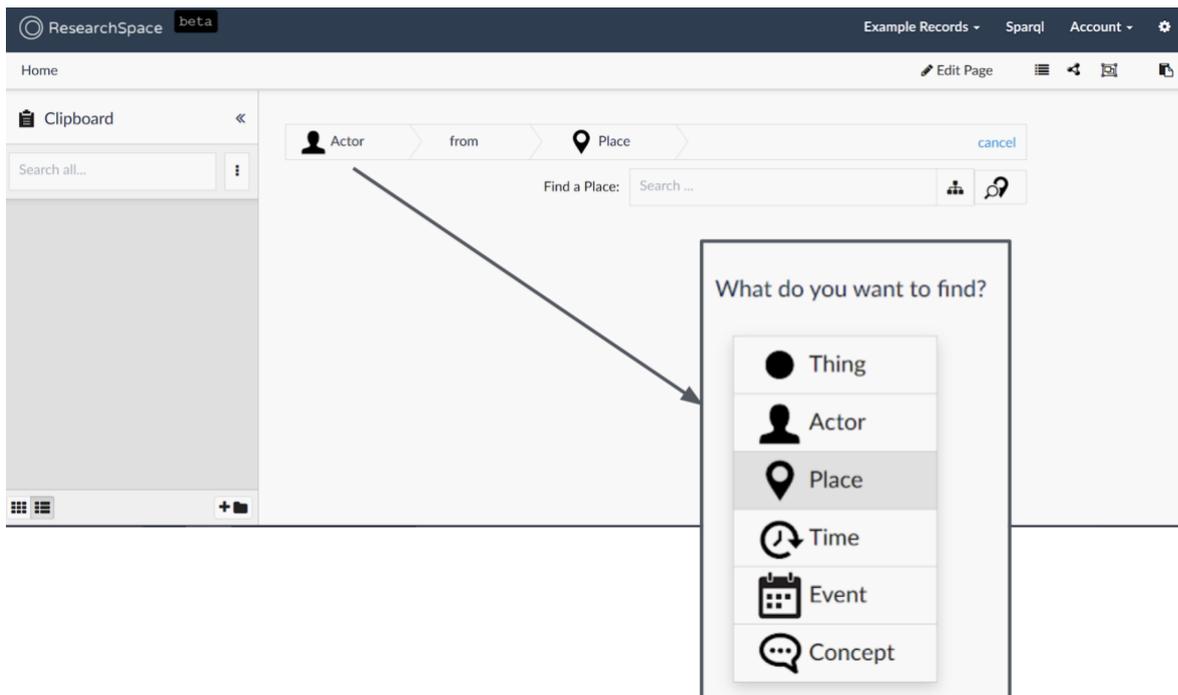
Figure 8. Combination of two screenshots: Finding resources in ResearchSpace

Deploying PAVLOV and integrating ResearchSpace in D4Science will increase the global search possibilities in the PARTHENOS data space. Hence, these tools are introduced in this deliverable, because a lot of the requirements for the global search from the different use cases will be implemented by the work of Task 6.5.

7.2 Requirements

The collection of requirements on interoperability, services, and tools which were stated in D2.1, included – among others – specific requirements for search services and for the access of resources. These were described from the perspective of the different disciplines under consideration. In the following list the respective requirements (numbers refer to the entry number in D2.1, Ch. 3) are presented:

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.1.: Visualize search results
Explanation	Being able to visualize objects and location

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.2.: Visualize search results
Explanation	Being able to visualize search paths

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.4.: Visualize search results
Explanation	Being able to map archive location and type in a tool

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.5.: Visualize search results
Explanation	The researcher can understand and display the spatial or chronological relationships between documents - Geotime visualisation tool

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.6.: Visualize search results
Explanation	The researcher can understand and present the spatial and chronological relationships between documents - 3D-visualization

Partner	KNAW-NIOD
Actor	Researcher/archivist/librarian as VRE manager
Requirement from D2.1	3.7.1.7.: Citation of dataset
Explanation	Researcher can use CENDARI material in a presentation or publication without having to figure out citation format

Partner	KNAW-NIOD
Actor	Research Infrastructure admin
Requirement from D2.1	3.7.1.9.: Enable (federated) search services
Explanation	Being able to search (keywords, persons, location/geographical information, events, time/dates...)

Partner	KNAW-NIOD
Actor	3.7.1.12.: Research Infrastructure admin
Requirement from D2.1	Being able to present data in an online portal

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.12.: Confidentiality
Explanation	Requirements are needed which state that some sensitive information may not be disclosed to unauthorized parties.
Use Cases ID	UR1

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.14.: Reliability
Explanation	Requirements are needed which constrain the software to operate as expected over long periods of time.
Use Cases ID	UR3

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.15.: Accuracy
Explanation	Requirements are needed which constrain the state of the information processed by the software to reflect the state of the corresponding physical information in the environment accurately.
Use Cases ID	UR4

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.16.: Usability

Explanation	For human interaction, usability requirements are needed which prescribe input/output formats and user dialogues to fit the abstractions, abilities and expectations of the target users.
Use Cases ID	UR5

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.17.: Architectural
Explanation	Requirements are needed which impose structural constraints on the software-to-be to fit its environment
Use Cases ID	UR6

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.18.: Archival information on collections
Explanation	Providing as much information as possible about archives that hold collections of interest helps researchers to be prepared

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.19.: Archival information on how archives manage and describe the holdings.
Explanation	Providing this information helps researchers to be prepared for working in an archive

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.20.: As much archival information as possible on archival holdings
Explanation	Providing this information helps to enable the researchers to undertake an initial assessment of the value of the archival holdings for their research.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.26.: Tool for finding sources
Explanation	Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.27.: Tool for organizing resource
Explanation	Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. This allows the user to take dynamic notes, organize them in useful ways, and link her/his research to data in the CENDARI dataspace.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.28. : Tool for displaying own research
Explanation	Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community. This not only displays the user's research in provoking ways, but can also reveal connections and patterns that may inform the conclusions of his/her research or guide further research.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.29.: Tool for accessing existing data resources (published or not)
Explanation	Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.30.: Tool for productively connecting with other researchers
Explanation	Studies reveal that functions facilitating early research (for example, finding, organizing, and displaying sources) are the most used and sought after in the scholarly community.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.31.: Tools to search and browse at a general level
Explanation	To find institutions and collections and other information.

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.32.: Tools to search and browse at a more detailed level
Explanation	To find detailed information on a given subject/work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.33.: Provide indication of the language and the place in which works were written (origin/provenance)
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.34.: Provide the year of composition of a work
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.35.: Show the availability of printed editions of a work
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.36.: Show manuscripts related to a work
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.37.: Show author, place and time of translations of a work
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.38.: Show availability of digital objects related to a work.
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.39.: Show a bibliography
Explanation	To help the user in their research work

Partner	KNAW-DANS
Actor	Content Provider
Requirement from D2.1	3.7.2.40.: Advanced tools for research and discovery.
Explanation	To enable new forms of research and discovery.

Partner	CLARIN
Actor	Librarian, archivist
Requirement from D2.1	3.8.2.6.: Visibility of an institution
Explanation	A website about an institution (library or archive) is not enough; it is also necessary to provide information on the collections or individual sources (digitally) available within the respective information; information should be available in different languages (esp. English)
Use Cases ID	Cen3

Partner	CLARIN
Actor	Researcher; data manager/archivist/librarian
Requirement from D2.1	3.8.2.7.: Mapping archival networks by location
Explanation	Being able to map archival networks by location of the archives.
Use Cases ID	Cen3

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.9.: Guidance on IPR (intellectual property rights) clearance
Explanation	Necessity to avoid IPR problems in collecting and sharing data; guidance on what constitutes an "IPR-free" resource, and what the (various levels of) freedom imply
Use Cases ID	814-3

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.10.: Workflow guidelines and workflow sharing
Explanation	Guidelines for researchers to create workflows, tools to support the creation of workflows and the ability to share workflows with other researchers
Use Cases ID	814-3

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.11.: Different search methods
Explanation	Depending on goals and circumstances, simple text search ("Google-style"), advanced search (e.g. with access to metadata fields and/or to the results of text analysis) and exploratory search (by browsing) are all needed.
Use Cases ID	814-5

Partner	CLARIN
Actor	Novice and expert researcher
Requirement from D2.1	3.8.2.12.: Functionalities that help with the provision of metadata
Explanation	The needs of researchers and purposes for data creation vary and require different metadata fields to be filled in upon artefact record creation (or entering them into storage); different pre-formatted metadata profiles are the way to capture the similarities and provide for dissimilarities; tools may help to extract metadata for a text semi-automatically
Use Cases ID	814-6/CPrep2

Partner	CLARIN
Actor	Researcher, student, instructor
Requirement from D2.1	3.8.2.13.: Data analysis/visualisation toolkits, with clearly defined input structures (in terms of standardized formats)
Explanation	Search and analysis results need to be displayed in a manageable way.
Use Cases ID	814-7/CPrep5

Partner	CLARIN
Actor	Researcher
Requirement from D2.1	3.8.2.14.: Flexible tools to analyse various data sets linguistically
Explanation	Pre-processing (spelling normalization, stemming, lemmatization), lexical and syntactic analysis (NER, shallow parsing etc.), information search/content extraction tools (e.g. corpus query functionalities), tools for translation and comparative corpus studies, tools for analysing corpora of speech and visual resources (e.g. for the annotation of gestures, prosody etc.)
Use Cases ID	CPrep3/D3C-2

Partner	CLARIN
Actor	Researcher, developer
Requirement from D2.1	3.8.2.15.: Standardized/compatible formats
Explanation	The usage of standardized (input/output) formats; the possibility to convert between formats (for texts, audio or visual data) in general as well as specifically the possibility to convert from non-standard to standard formats.
Use Cases ID	CPrep4

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.1.: Possibility to compare the text of an author with works by different authors.
Explanation	Preparation: search/browse for content and create a virtual collection with found works; create relationships; compare the works like in, document the comparison, add this document to the collection. Usage: search/browse this collection presentation as graph presentation with map/timeline; make annotations; add comments; possibly modify the virtual collection or add new documents (further preparation).
Use Cases ID	SISMEL_01, SISMEL_03, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	The researcher wants to know when the manuscripts have been produced.
Explanation	Focus on range of time based; search/browsing; presentation with map and timeline.
Use Cases ID	SISMEL_01, SISMEL_04, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	Get a list of all the manuscripts containing a certain text.
Explanation	Do faceted browsing on manuscripts and institutions; find content, not yet considered in editions: assumed that all editions declare a relation "is_edition_of" to the original; search all content, from this person, which don't declare this relationship.
Use Cases ID	SISMEL_01, SISMEL_04

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.4.: Find out how many collections of unedited texts are held in European libraries
Explanation	Faceted browsing on ...; find works which are not part of a relationship "is_edition of".
Use Cases ID	SISMEL_01, SISMEL_04

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.5.: Search for philosophical concepts.
Explanation	Faceted browsing on ...; and search for works annotated with keyword "philosophical", about diffusion of works create relationships between philosophical concepts and named authorities; so that users can navigate through the graph of related content.
Use Cases ID	SISMEL_02

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.6.: Search for quotations in any century texts.
Explanation	Combination of faceted browsing and search on works of the 15th century which are annotated with, for example, Aristotle or have a relationship to Aristotle; create a virtual collection with the results of the search; compare findings; collaborate on this task; create relationships to authorities, e.g. people, places, events, etc., so that users can navigate through the graph of related content to identify which preachers have used Aristotle's texts and florilegia.
Use Cases ID	SISMEL_02, SISMEL_04, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.7.: To find something out about any author.
Explanation	Search the repository, possibly prepare a virtual collection, make notes, create saved search, make annotations, build relationships, creation of bibliography, list of related collections, content, archives/libraries, list of groups or users, list of topics or research areas, list of research questions – provide general information about an author or artist. Something like a summary page, with personal information (biography, ...)
Use Cases ID	SISMEL_01, SISMEL_03, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.8.: Know in which manuscripts and repository the texts are held.
Explanation	More focussed on special questions: which manuscripts held by which institution faceted browsing based on: archives, collection, in the boundaries of countries, language/translation, date, range of time, ... present the diffusion of work, possibly with a map and timeline.
Use Cases ID	SISMEL_01 to SISMEL_06

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.9.: Find in which language the works are written.
Explanation	With focus on languages faceted browsing on ... and languages, the search and browsing can be supported by language related metadata and annotations.
Use Cases ID	SISMEL_01, SISMEL_02, SISMEL_03, SISMEL_04, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.10.: Harvest biographical information on ... (any Author).
Explanation	Support a possibility for semantic annotation of content: automatic or manual named entity recognition (person, organization, place, date, event, do faceted browsing on this information, present this information with a map and timeline.
Use Cases ID	SISMEL_01, SISMEL_03, SISMEL_05

Partner	SISMEL
Actor	Researcher
Requirement from D2.1	3.8.3.11.: To know what kind of texts are transmitted with an author's work.
Explanation	Follow relationships; see what works are related to an individual's work; do faceted browsing on country, region, range of time, etc. to understand the cultural context
Use Cases ID	SISMEL_01 to SISMEL_06

Hence, the requirements collected in D2.1 with regard to search for and access of resources, concern four overall domains:

- search options within a catalogue;
- the information provided for retrieval;
- aspects of access granted to the resources, and
- of data import into the catalogue.

As for the search options, a prominent feature requested was the possibility not only to search but also to browse resources, e.g. by a faceted search, with proposals for various possible concepts to be the entry points of such browsing. Of course, a simple and advanced metadata search was also requested. Further requirements concerned the handling of search results, e.g. their display, organization and further processing.

Some of the issues named above do not concern the application side of the platform but rather the level of detail of the data provided and retrieved (e.g. the desideratum for elaborate information about collections and holdings of archives, information on possible research questions for the presented material and on research output created based on it). Other issues will have to be addressed by tools that allow for work on the actual data, like the proposed NERLiX application (see D6.2 and infra, the devoted section). Here, the issues of data processing, handling various languages, long-term data preservation, (again) user management and working collaboratively may have to be addressed.

7.2.1 Functional Use Cases

<table>
<tr><td colspan="2" align="center">**Accessing Entities registered in PARTHENOS (ACCESS)**</td></tr>
<tr><td colspan="2">ACCESS_01: Search and browse the PARTHENOS registry

ACCESS_02: Search and browse the PARTHENOS content cloud across several research Infrastructures</td></tr>
<tr><td>Method</td><td>● Faceted search on metadata records
● Text search on metadata records</td></tr>
<tr><td>Goal</td><td>● Find digital language resources (e.g. language corpora)
● Find non-digitized material existent in archives</td></tr>
<tr><td>Use Cases from Language Studies</td><td>● Search for non-digitized material in archives with help of Metadata (OEAW_04)
● Search/browse available metadata about language resources using a combination of the faceted and text search (OEAW_01)
● Detection of relevant data sets for corpora (CLARIN_01)</td></tr>
<tr><td colspan="2">ACCESS_03: Retrieval/access of metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud</td></tr>
<tr><td>Method</td><td>● Access of full, elaborate metadata records</td></tr>
<tr><td>Goal</td><td>● Get detailed information on resources</td></tr>
<tr><td>Use Cases from Language Studies</td><td>● Gather metadata for data sets relevant to the research goal (CLARIN_01)
● Get detailed information about resource (preview found resources) (OEAW_02)</td></tr>
</table>

7.2.1 Functional Use Cases

ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	
Method	● Access resources
Goal	● View and reuse the actual resources
Use Cases from Language Studies	● Gathering of the relevant data sets (CLARIN_01) ● Access resource from search results (OEAW_03) ● Take the accessed resource and use it for research (OEAW_03) ● Selection of corpus items; creation of the primary corpus (CLARIN_01) ● Detection of relevant data sets (CLARIN_01)

Using Virtual Research Environments (VREUSE)	
VREUSE_02: Reference entities of the PARTHENOS registry in VRE posts	
Method	● Refer to and comment on resources
Goal	● Share information on and about resources
Use Cases from Language Studies	● Reference and comment non-digitized Material (OEAW_05)
VREUSE_03: Deposition	
Method	● Upload material into a VRE
Goal	● Store resources
Use Cases from Language Studies	● Inclusion of the relevant data into the VRE (CLARIN_01) ● Upload private digitized material in a private VRE (OEAW_05)

VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	
Method	● Provide resources via a VRE
Goal	● Share resources, work collaboratively
Use Cases from Language Studies	● Data analysis within a shared working platform (CLARIN_01) ● Publication of results and provision of the corpora (CLARIN_01) ● Share private digitized material with selected researchers (OEAW_05)
VREUSE_05: Process a dataset and publish results	
Method	● Data analysis and publication
Goal	● Work with data and share results
Use Cases from Language Studies	● Corpus analysis and comparison (CLARIN_01) ● Data analysis within a shared working platform (CLARIN_01)

Using Virtual Research Environments (VREUSE)	
UC_VREUSE_03: Deposition	
Goal	deposit some of the data produced in a PARTHENOS compatible archive Preview of metadata aggregated via the mapping tool
Use Cases from Archaeology, Heritage and Applied Disciplines	Deposit data (AR_04) Preview metadata in html format (MINT_04)

UC_VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	
Goal	Creation of an OAI-PMH repository to share ingested metadata
Use Cases from Archaeology, Heritage and Applied Disciplines	Sharing ingested metadata (CI_06)

UC_VREUSE_05: Process a dataset and publish results	
Goal	Inspect and enrich visual documents stored in one of the catalogues available in the PARTHENOS portal
Use Cases from Archaeology, Heritage and Applied Disciplines	inspect and enrich visual media documents (AR_07)

<table>
<tr><td colspan="2" align="center">**Accessing Entities registered in PARTHENOS (ACCESS)**</td></tr>
<tr><td colspan="2">UC_ACCESS_01: Search and browse the PARTHENOS registry
UC_ACCESS_02: Search and browse the PARTHENOS content cloud across several research Infrastructures</td></tr>
<tr><td>Goal</td><td>Get a list of institutions holding archaeological information concerning excavations, objects, periods
See a preview of data available to allow a user determining the relevance of the data for her/his research
Discover tools or best practices to achieve a certain goal</td></tr>
<tr><td>Use Cases from Archaeology, Heritage and Applied Disciplines</td><td>Search/browse archaeological information concerning excavations, objects, periods (AR_01)
Possibility to get a preview of available data (AR_02)
Search and access the services registry (AR_05)</td></tr>
<tr><td colspan="2">UC_ACCESS_03: Retrieval/access of metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud</td></tr>
<tr><td>Goal</td><td>Access archaeological collections to compare information on a specific research field
Access information about metadata schemas and ontologies used for archiving archaeological resources
Retrieve information about collections and datasets according to specific terms from a vocabulary or retrieve a location from a gazetteer</td></tr>
<tr><td>Use Cases from Archaeology, Heritage and Applied Disciplines</td><td>Access archaeological collections (AR_03)
Access information about a metadata format (AR_08)
Retrieve information from vocabularies and gazetteers (AR_09)</td></tr>
<tr><td colspan="2">UC_ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud</td></tr>
<tr><td>Goal</td><td>Metadata acquisition of a data provider</td></tr>
<tr><td>Use Cases from Archaeology, Heritage and Applied Disciplines</td><td>Metadata harvesting (CI_01)</td></tr>
</table>

D. Domain Services and Tools

8. Specialized Services and Tools

8.1 Overview

With the term specialized services we identify that branch of services - or combination of services - that are optimised for scientific high-level purposes, based on specific domain driven research questions and paths. Such services are usually not provided by general-purpose (i.e.: global / not domain driven) digital tools or infrastructures.

Driving the creation of new technologies, methodologies, and information systems, various communities within the digital humanities - through the collection of use cases - have identified the need to integrate in the PARTHENOS infrastructure a selected cluster of tools and specialized services which, due to the interdisciplinary nature of the project, will be expanded, made available and usable to this wide research area.

8.2 Methodology

As regards the methodology for the evaluation of specialized services, a selective strategy - based on the matching of the requested services and services offered directly in the VREs of D4Science containing Domain specific tools (PARTHENOS_LAB VRE, Rubrica VRE and NERLiX VRE), was adopted. These research environments contain tools that provide high-profile services that met specific domain oriented requirements.

It should be noted that, unlike the "symmetric" methodology used to make the overall assessment - in which the deliverable D6.2 and D2.1 have the same weight - in this assessment on the specialized services the evaluation has as its origin in the requests formulated by the research communities involved in PARTHENOS, gathered an organized by WP2.

For this reason and to make clear the origin of each scientific need that has been considered, in this chapter the requested services are classified by reference community.

8.3 Tools and services integration levels

There are different ways of integrating a specific tool within D4Science. The different types of integration, and accordingly authorizations, significantly change the processes that can be performed by both users and the service manager. D4Science, and consequently the PARTHENOS project, uses as authorization protocol the gCube Authorisation framework. It is a token based authorisation system, compliant with the Attribute-based access control (ABAC), that defines an access control paradigm whereby access rights are granted to users through the use of policies which combine attributes together.

For each Virtual Research Environment (VRE), users can exploit three types of tokens:

- **Personal Token**: the personal token has to be used for any programmatic interaction with the services;
- **Qualified Token**: this is a token associated with a mnemonic label.
- **Application Token**: this is a token associated with an application identifier. Operations performed with this token are accounted to the application and not to the user profile. The VRE Manager is informed.

Any token is valid only in the VRE in which it was created.

This said, there are 3 levels of integration available for Community Services in D4Science:
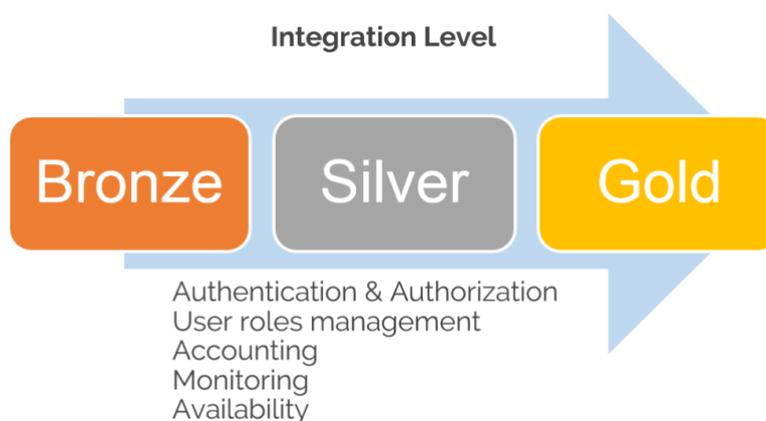
Figure 9. Integration Levels

The highest level of integration is the Gold one. The infrastructure provides the Community Service with authentication, authorization, monitoring and accounting. However the Service must be written according to predetermined policies and run over an authorised Web Container (SmartGears).

The second level of integration is Silver: the Service will run on any platform and use a User Token. The user can contact the D4Science tool and can perform operations on behalf of other users.

In the lowest level of integration, Bronze, the Community Service will run on any platform and use an Application Token. The user can contact D4Science services but cannot perform operations on behalf of the users.

8.4 Tools

PARTHENOS aims to serve the Humanities at large, and may also be relevant for some Social Sciences and other neighbouring disciplines. Although sometimes treated as one community (e.g. in the context of European research infrastructure consortia), this is actually a very broad and heterogeneous group, and PARTHENOS as a cluster of research infrastructure initiatives within this broad domain cannot serve all of them equally, but will have to prioritize. To identify the core services to focus on, we adopted a bottom-up approach, starting with the communities represented in PARTHENOS, grouped using the following labels:

- History (in a broad sense: including Medieval Studies, Recent History, Art History, Epigraphy, etc.);
- Language-related Studies (including Literature, Linguistics, Philology, Language Technology, etc.);
- Archaeology, Heritage & Applied Disciplines (including Cultural Heritage, Archives, Libraries, Museums, Preservation / Conservation experts, Digital curation / edition / publishing, etc.);
- Social Sciences (in a broad sense: Sociology, Political Science, Geography, Anthropology, Cultural Studies etc.).

History, Language Studies and Heritage and Applied Disciplines can be considered the highest priority for PARTHENOS, since the Social Sciences community is much less strongly represented in the project.

In what follows we give an assessment of the tools addressing requirements coming from these broad groups, referenced in the tables.

8.4.1 RUBRICA - Reference Resources Integration plAtform

8.4.1.1 Overview

During the scientific research process, the idea of relying on a set of reference resources or tools - according to specific domain needs – possibly available in different formats (i.e. digital or printed) denotes the need to add meaningful and trusted (i.e. with a clear provenance and a reliable chain of trust) contextual information. Even if the process of reliance on these resources and assets available in the digital ecosystem is now consolidated, new needs are emerging. Recent research[30] has noted that the proliferation of online information sources has radically changed both the context in which academic reference resources can be found and how they are used.

In this panorama of heterogeneous information stands PARTHENOS. One of the PARTHENOS pivotal goals is the integration of datasets and services available in the Humanities and Social Sciences digital ecosystem. Within the PARTHENOS project, a virtual research environment, RUBRICA (Reference Resources Integration plAtform), is being developed. RUBRICA aims to foster the interoperability and integration of various reference resources used in different disciplines. Starting from trusted knowledge bases (i.e.: databases, thesauri, authority lists etc.) researchers can create, merge, edit and reuse specialized reference resources, developed according to specific research purposes, without performing repetitive tasks on each resource. RUBRICA also allows sharing this knowledge base with other users thanks to the IT infrastructure on which PARTHENOS is based.

8.4.1.2 Scope

The scope of RUBRICA is to prove that by the integration of different reference resources (both generic and specialized) in a custom Virtual Research Environment (VRE), users are able to create, merge, modify and share a specialized knowledge base created according to the most up to date research needs. RUBRICA approaches to digital resources and tools that support significant information needs, rethinking the basics of the reference resource production workflow.

---

[30] Oxford University Press, *How academic users understand, discover, and utilize reference resources* https://global.oup.com/academic/content/pdf/navigatingresearch.pdf (last visited 27/04/2018)

8.4.1.3 Methodology

Service and data integration, as well as interdisciplinarity, are increasingly relevant for the work of individual researchers, at the point that these are also becoming common features of scientific research and thus a basic requirement for a well designed RI.

Based on the above assumptions, RUBRICA is a D4Science Virtual Research Environment allowing the user to create new reference resources as the result of a selection and merging process of multiple resources.
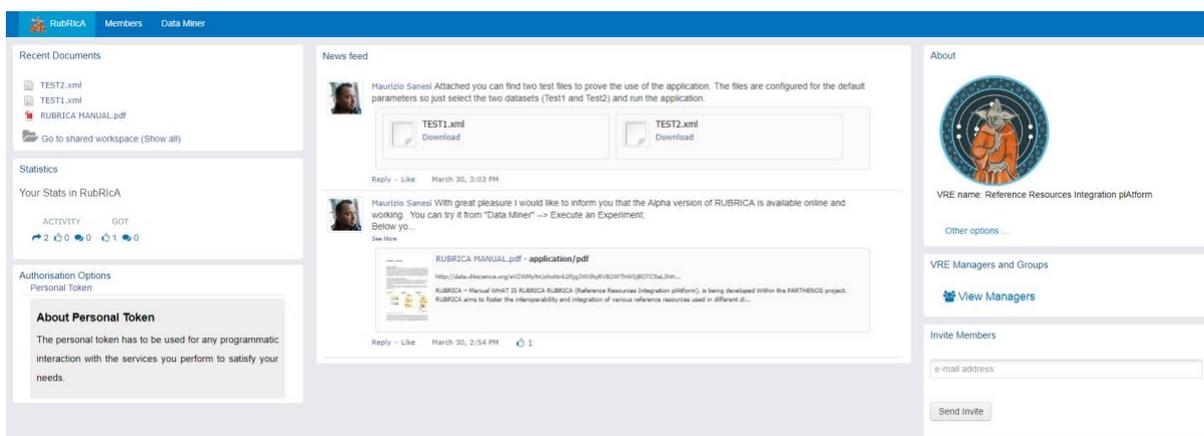
Figure 10. RUBRICA VRE Landing Page

Following the evolution of the semantic web in which anyone can contribute to the creation of extraordinary knowledge graphs, RUBRICA produces and operates on RDF resources. A common ontology is required in order to provide convergent data, so to have subclasses describing a specific topic under the same taxonomic name. In the RUBRICA environment each resource should be mapped on the CIDOC Conceptual Reference Model (CRM) common semantic framework. In case the initial reference resources are not already modeled on a common model, the required mapping is made possible thanks to the 3M Mapping Tool.

8.4.1.4 Outcomes

The expected outcome of RUBRICA is a Virtual Research Environment, integrated in the PARTHENOS project infrastructure, that could possibly be further extended to other domains in the SSH landscape and beyond.
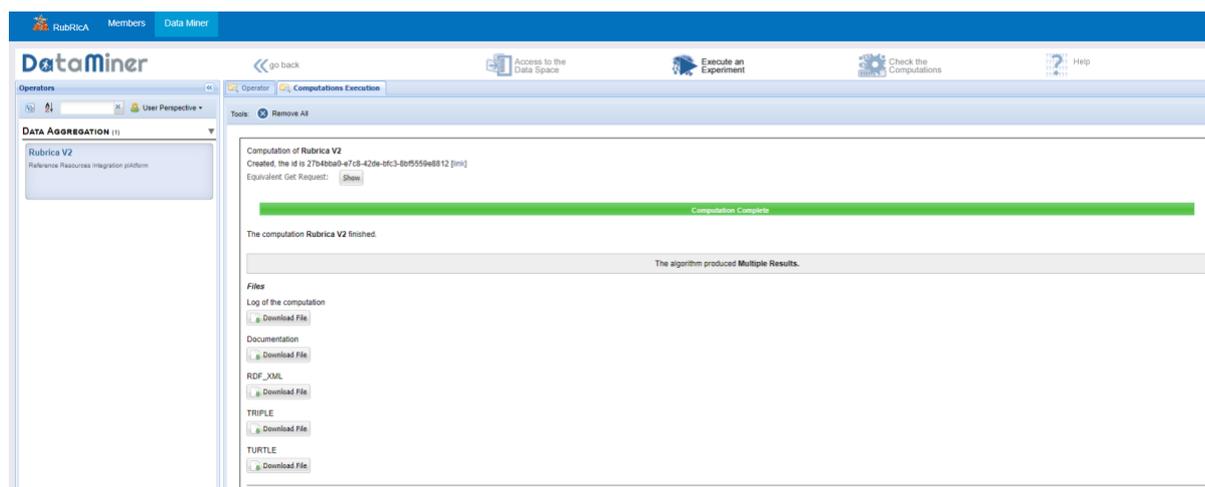
Figure 11. RUBRICA VRE Output download page

RUBRICA, in its final iteration, in addition to integrating and creating improved reference resources, will implement data editing (i.e. coverage check), browsing and will allow results to be exported in a standard format (i.e. XML/RDF), for storage or further reuse. RUBRICA will also provide a descriptive document showing the analytics and process parameters that were used during the creation of the new reference resource. The VRE will be accessible via the D4Science PARTHENOS Gateway.

8.4.1.5 Satisfied Requirements

RUBRICA guarantees the satisfaction of some requirements made explicit by the Historians community. The nature of the core algorithm allows aggregation of data and metadata about the resources, enabling RUBRICA to maintain a high level quality control, satisfying the requirement exposed in the use cases 3.6.3.5. SISMEL_01 to 3.6.3.10. SISMEL_06 and 3.6.1.10. MINT_01, about the disambiguation of names of persons and places, documents (i.e.: manuscripts shelfmarks), titles of texts/works. RUBRICA also satisfies requirements expressed in the Prosopography scenario provided for the development of the Standardization Survival Toolkit.

In addition to the previous specialized requirements, RUBRICA produces a summary documentation, the dataset resulting from the aggregation process in three different formats approved by standard protocols (RDF-XML, TTL (Turtle) and N-Triple) satisfying the general use case AGGR_02 (Export metadata via standard protocols).

From an infrastructural point of view the integration process of RUBRICA (Gold level) and its subsequent publication within a specific VRE, also meets other requirements that are more directed to the role of the administrator or data manager, specifically the use cases

VRESET_01 (Set-up of a domain-specific VRE), VRESET_02 (Integration of services in the PARTHENOS infrastructure), VRESET_03 (VRE authentication and authorization) and VREUSE_04 (Private and public sharing of resources deposited in the VRE workspace).

KPI	Assessment
Functionality	RUBRICA provides the integration of different reference resources in a custom Virtual Research Environment. Users are able to create, merge, modify and share a specialized knowledge base created according to the most modern research needs. The user is informed in D4Science when the process finished. The output file is available in the PARTHENOS user data space as well as download links. The functionality of the tool is adequate towards the expressed needs.
Scalability	RUBRICA runs with two RDF document and various paramenters as custom input. This settings and the possibility of the VRE manager to give the dataminer a custom computing power, satisfy the scalability of the tool.
Usability	The user can select files directly from the available VREs and enter the parameters following the instructions found inside the VRE DIRECTORY. This guarantees a good usability of the tool.
Layout/design	RUBRICA core engine is integrated into the D4Science layout interface. Users that already worked with D4Science, should recognize immediately how the tool is working.
Accessibility	This is addressed by the D4Science interface. The tool itself has no additional accessibility support.
Documentation	In RUBRICA VRE is available a Manual providing information about the tool (what is RUBRICA, What is a reference resource, input paramenters, Output, Data grouping and expected results)RUBRICA will also provide a descriptive document showing the analytics and process parameters that were used during the creation of the new reference resource. The VRE will be accessible via the D4Science PARTHENOS Gateway .
Language accuracy	Currently RUBRICA manuals and documentation are available for English. Future steps see the implementation of Spanish, French, German and italian version

Compatibility with project data	RUBRICA is integrated in the D4Science Infrastructure. The tool ca run files both from user personal dataspace and other official VRE dataspaces
Security	This is addressed by the D4Science infrastructure. The tool itself has no additional security layers

Chapter of D2.1	Title in D2.1	Satisfied Requirement

		Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.1. Use cases from Archaeology, Heritage and applied disciplines	3.6.1.10. MINT_01	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.5. SISMEL_01	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.6. SISMEL_02	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.7. SISMEL_03	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.8. SISMEL_04	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.9. SISMEL_05	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.3. Use cases from Studies of the Past	3.6.3.10. SISMEL_06	Disambiguation of names of persons and places, documents (i.e. manuscripts shelfmarks), titles of texts/works
3.6.1 + 3.6.3	AGGR_02	Export metadata via standard protocols
3.6.2 + 3.6.3 + 3.7.2	VRESET_01	Set-up of a domain-specific VRE
3.6.1 + 3.6.2 + 3.6.3 + 3.7.1 + 3.7.2	VRESET_02	Integration of services in the PARTHENOS infrastructure
3.7.1 + 3.7.2	VRESET_03	VRE authentication and authorization
3.6.1 + 3.6.2 + 3.7.1 +3.7.2	VREUSE_04	Private and public sharing of resources deposited in the VRE workspace

Extra (Scenario provided for the construction of the Standard Survival Toolkit (SSK)	Prosopography scenario	Need of a tool providing interoperability and integration of various reference resources used in different disciplines.

8.4.2 NERLiX

8.4.2.1 Overview

NERLiX stands for Named Entity Recognition and Entity Linking and Extraction. As a service in PARTHENOS it covers different steps for the enrichment of texts and/or the extraction of (meta)data from text. Named Entity Recognition (NER) is frequently used in Digital Humanities when it comes to text analysis. When working with a lot of text documents or with a long text document, NER can help to give a brief overview on topics in a text document without reading it in detail. Also, NER can extract such topics out of a bunch of text documents, that can be further processed, e.g. with Social Network Analysis (SNA). There are a lot more of use cases where NER is the starting point of the analysis.

In principle, there are many possibilities which kind of entities to recognize in a text document. Usually this covers persons, places, organizations or different kinds of unique concepts. The primary condition for doing an NER is the availability of an authority file, from where these entities can be fetched, e.g. DBPedia. NERLiX covers this requirement by building up on Apache Stanbol[31]. Apache Stanbol is a Semantic Engine, where enhancement chains are defined, that integrate authority files. Apache Stanbol then extracts entities from a text, based on an enhancement chain.

8.4.2.2 Scope

The idea behind NERLiX is a more complex one, than the here described single NER tool. In it's final state, NERLiX should allow a seamless workflow for users, beginning with the selection of useful resources and resulting in re-using this resources in an enriched state. Raising thus the quality of the data and supporting a research data life cycle. This bigger scope of NERLiX is described in PARTHENOS Deliverable 6.2, subchapter 6 (Title: "Use Case 1 - NERLiX")[32].

---

[31] Apache Stanbol: https://stanbol.apache.org/ (last visited 27/04/2018)
[32] PARTHENOS D6.2 Report on services and tools: http://www.PARTHENOS-project.eu/Download/Deliverables/D6.2_Report_on_services_and_tools.pdf (last visited 27/04/2018)

Currently, some of the tools for the workflow are in place. Especially the NER tool that is based on Apache Stanbol. It is already available in the PARTHENOS VREs and is called "Stanbolwrapper". Other tools like the visualization of the NER results are in development and available in a beta state.

8.4.2.3 Methodology

The methodology of "Stanbolwrapper" is described as follows. Currently, it is a precondition that a text document is either uploaded or otherwise created (e.g. as a result of another tool in the PARTHENOS VREs) in the PARTHENOS data space in order to be processed with "Stanbolwrapper". This text document needs to be a plain text document. In the future, there should be a seamless workflow, where users can choose that a text document found with the PARTHENOS discovery app can be pushed directly to the NERLiX tools. That means that there is no need of a manual upload (the manual upload stays as an option). It is also intended to allow more complex data formats for the input.

After choosing the "InFile" (i.e. the text document to be processed), there are two other options: EnhancementChain and OutputFormat. Different OutputFormat are currently implemented: JSON, RDF/XML, RDF/JSON, Turtle, and N-Triples. The choice depends on the further processing of the results with other tools. It is possible to get an RDF description and to put this into a triple store for further analysis. The EnhancementChain relies on Apache Stanbol and indicates which authority files to use and therefore, which named entities to recognise:

- DBPEDIA: uses as an authority file DBpedia[33] for the NER, thus involving entities from Wikipedia, covering most of the mentioned named entities like persons, organizations, places, etc.

As sometimes only the recognition of places are of interest, the other EnhancementChains are pointing to GeoNames[34]:

- COUNTRIES: uses Geonames, recognises only countries
- CITIES: uses Geonames, recognises only cities
- LOCATIONS: uses Geonames, recognises all kind of places

The EnhancementChains will be expanded based on articulated needs of the user communities.

---

[33] DBpedia: http://wiki.dbpedia.org/ (last visited 27/04/2018)
[34] GeoNames: http://www.geonames.org/ (last visited 27/04/2018)
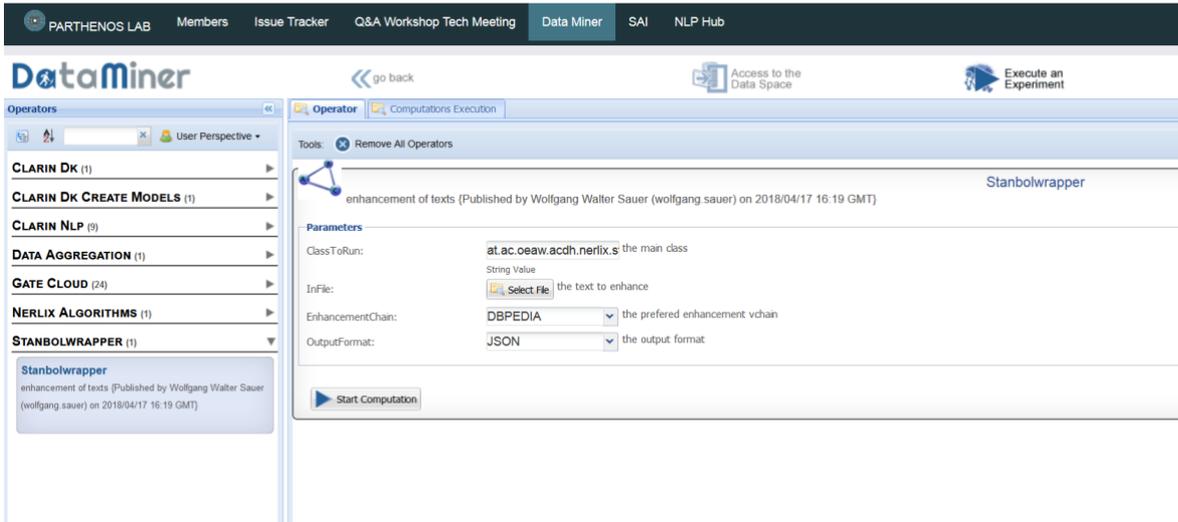
Figure 12. Choosing the parameters for the Stanbolwrapper in the PARTHENOS Lab VRE

Depending on the size of the text document, the processing can take some time. The text is sent to an Apache Stanbol instance and the entities are recognized based on the chosen EnhancementChain.
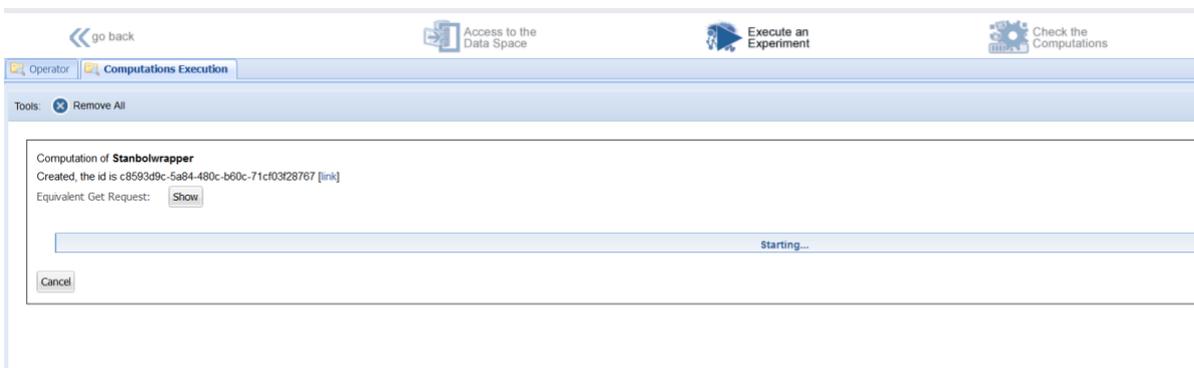
Figure 13. A running NER process of the Stanbolwrapper in the PARTHENOS Lab VRE

After finishing the NER process, a log of the computation is created that gives some insight into the background process. The OutFile is available in the chosen OutputFormat and is saved into the PARTHENOS dataspace of the user. It can also be downloaded directly. In the future, there should be more dissemination services available, especially a visualizer of the results (such a tool is in development and is called Distanbol).
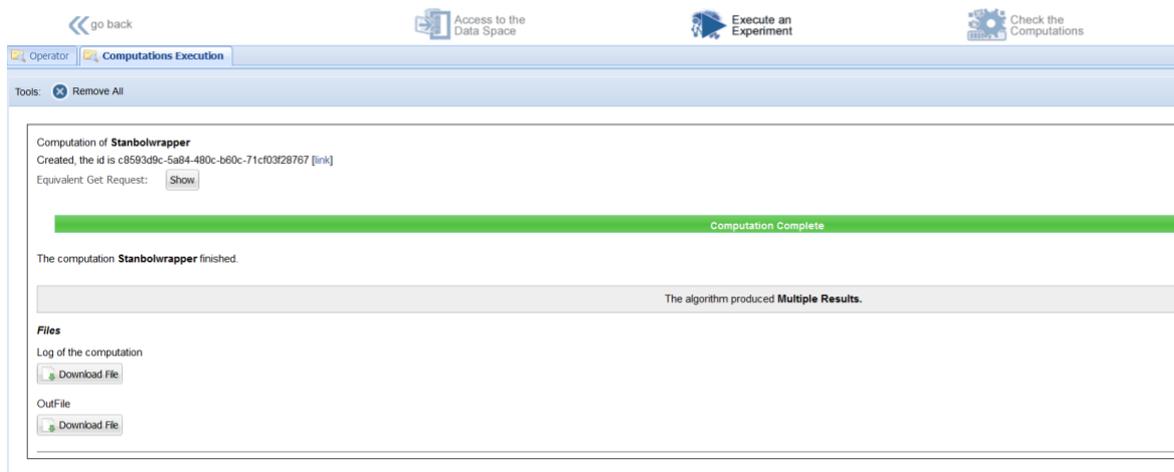
Figure 14. Screenshot: Output after finishing the NER process of the Stanbolwrapper in the PARTHENOS Lab VRE

The next steps depend on the aim of the user. It could be possible to convert the result file into TEI or to make some analysis on the basis of the extracted entities (e.g. showing mentioned places and connecting them to a timeline). For the finalized NERLiX workflow, there will be different options for the user for what to do with the results. These options will be available in the output window.

8.4.2.4 Outcomes

Stanbolwrapper is integrated into the NERLiX workflow, but it is also possible to use Stanbolwrapper alone. The outcome of Stanbolwrapper is a document in the chosen output format and is based on the Apache Stanbol data structure. The data structure of the output is described in the documentation of Apache Stanbol [35], in detail in the documentation of the Apache Stanbol enhancer[36].

The output document can be used for further processing or analysis. It is meant to push the output to other tools that make sense out of the processed data. This data are the recognized entities that are bound to persistent identifiers from the chosen authority file(s). Stanbolwrapper is an important part of the bigger NERLiX use case. The integration of Stanbolwrapper into the PARTHENOS VRE is an important step for the other parts of the NERLiX use case. The already working export of a result document can be used for different other use cases. Distanbol, a tool in development, will be helpful in visualizing the

---

[35] Documentation of Apache Stanbol: https://stanbol.apache.org/docs/trunk/ (last visited 27/04/2018)
[36] Documentation of Apache Stanbol Enhancer:
https://stanbol.apache.org/docs/trunk/enhancementusage.html (last visited 27/04/2018)

results of Stanbolwrapper. There are other tools in evaluation, that can handle the results of Stanbolwrapper. Especially when the results are pushed into a triple store or into the resource discovery of PARTHENOS, they will become usable for many other researchers and, therefore, enabling a research data life cycle and helping to raise data quality.

8.4.2.5 Satisfied Requirements

The following requirements are based on PARTHENOS Deliverable 2.1 (D2.1): Report on User Requirements:

Chapter of D2.1	Title in D2.1	Satisfied Requirement
0.2.3. Language Related Studies	Language Related Studies	It is mentioned, that NER tools are not only important for linguistic studies, but are also relevant for other sciences. NERLiX approach is to enable NER for all research communities that are part of PARTHENOS
2.1.1.5. (see also 3.6.3.10.) History	Historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period	A NER service is a precondition for this use case. NERLiX can be used to cover this precondition
2.1.2.1. Language Related Studies	Natural Language Processing Expert wants to test her tool for semantic annotation on an available digital edition of historical texts	A NER tool is mentioned in this use case. This could be NERLiX. Even more, if NERLiX allows TEI as input format, this use case can be covered by NERLiX.
2.1.2.2. Language Related Studies	Create annotated digital edition	Performing NER is a main part of the scenario. NERLiX can cover this.
3.6.2.6. Language Related Studies	Corpus-based Analysis of Historical Newspapers	NER is mentioned in the user story, could be covered by NERLiX.
3.8.2.8. 3.8.2.14. Language Related Studies 3.8.3.10. History	Needs for NER tools	These three requirements need NER tools. NERLiX can be a NER service covering these requirements.
4.4. -	User study about digital approaches at University of Copenhagen	NER as a sophisticated data processing method, that would be useful. Can be covered by NERLiX.

Another requirement of D2.1 is the need for standards and best practices for performing NER (see: Deliverable 2.1, chapter 2.2, UC#7). NERLiX can deliver a stable reference

point, where best practices for NER can be developed around the tool workflow and where standards for output formats can be established.

Assessment of Stanbolwrapper against the KPIs listed in section 2.1.1:

KPI	Assessment
Functionality	After choosing a plain text file as input and starting the NER process, the tool immediately reacts and performs the NER. The progress of the process is shown in percentages. The user is informed in D4Science when the process finished. The output file is available in the PARTHENOS user data space as well as download link and conforms to the Apache Stanbol data model. The functionality of the tool is adequate for the expressed needs.
Scalability	As long as only one text document is involved, the scalability of the tool is satisfied. When it comes to more than one text document, there is the need for a batch process. Such a batch process is currently not available and should be integrated. If the currently missing integration of the tool in the resource discovery is solved, it is expected that the scalability of the tool is sufficient.
Usability	As there is until now not a workflow to get data from the resource discovery directly to tools in the VREs, a user needs to understand, how to get data into the user dataspace. This is handled by the D4Science infrastructure, where Stanbolwrapper does not have an influence. As long as this part is not solved, usability is only guaranteed for users that do have a basic understanding of D4Science. When it comes to integrate a new enhancement chain for the Stanbolwrapper, the usability is currently unclear. For the further development of the tool, it is recommended to integrate a workflow that allows to enable additional enhancement chains.
Layout/design	Stanbolwrapper is integrated into the D4Science layout interface. Users that already worked with D4Science, should recognize immediately how the tool works. Nevertheless the D4Science layout interface could benefit from some investment into a more easy to understandable layout, especially for users that are not used to D4Science. Stanbolwrapper itself should have a more clever layout, where the different parameters are better explained and more prominently highlighted. Also, unchangeable parameters could be hidden in an expert view. It would be also great to have some kind of preference settings, e.g. to define the favoured output format.

Accessibility	This is addressed by the D4Science interface. The tool itself has no additional accessibility support.
Documentation	There is currently in general a lack of documentation of the tools. It seems, that the D4Science interface needs to address this issue. For the NERLiX use case, there is a workflow and technical documentation in PARTHENOS D6.2, but this is not integrated in the Stanbolwrapper. Integrating the documentation in the interface of the tools in D4Science should get a priority, otherwise it is not easy for researchers to fully use the tools.
Language accuracy	Currently, the NER is available for English and for German. By adding new enhancement chains, more languages can be covered. It is recommended to add a language option in the start page of Stanbolwrapper, so that users can get a clear picture of supported languages. It should be also possible to call for the implementation of currently unsupported languages for the NER process.
Compatibility with project data	As Stanbolwrapper is built into the D4Science infrastructure, the tool can run data from the PARTHENOS data space. Nevertheless it needs a way to get data from the discovery tools directly into the VRE tools, something that needs to be solved by the D4Science infrastructure.
Security	This is addressed by the D4Science infrastructure. The tool itself has no additional security layers

There is a new tool integrated into D4Science called "NLP Hub", that is not part of this assessment. It also offers NER, but it is not built directly into the D4Science infrastructure (therefore it is currently not possible to load data from the PARTHENOS data space and to store results into the PARTHENOS data space). It deals instead with some of the gaps from the KPIs assessment of Stanbolwrapper, especially in regard to being able to choose from a couple of languages and to finely granulate which entities to the NER process should recognise.

Figure 15. Screenshot: Start page of the NLP Hub

It could be useful to combine the effort of NLP Hub with the one from NERLiX, to give users more options for doing their NER process.

8.4.3 CLARIN Natural Language Processing tools

8.4.3.1 Overview

Text-based studies of different Digital Humanities' disciplines (i.e. not only language-related studies, but also literary, social or historical studies, etc.) heavily rely on digitized textual resources, i.e. digital texts and corpora representing various domains, points in time, events, languages, text types, etc. Sharing and providing digital resources in order to allow their use and reuse in different scholarly contexts is just one aspect of dealing with this situation where increasingly large amounts of digitised textual sources are needed for quantitative research. Quantitative studies on texts may also benefit greatly from Natural Language Processing (NLP) on the respective texts before further analysis. For instance, historical texts may contain a variety of spelling variants which may be mapped on one normalised spelling; individual word forms (tokens) represent lemmas and are parts of their paradigms of inflected forms; lemmas are instances of Parts-of-Speech; combinations of word forms may build syntactical structures; and so on. Depending on the research question, scholars may not be specifically interested in the particular word forms of a text but rather in what these word forms represent on different language levels. Thus, tools, provided by language-related studies, to linguistically analyse texts of various languages

may be of interest not only to linguists, but also to scholars of other Humanities' disciplines. These potential users have to be enabled to apply NLP tools and services to their texts and corpora, without being NLP experts themselves. With the integration of NLP services from the CLARIN infrastructure, PARTHENOS addresses this requirement of Digital Humanities' studies.

8.4.3.2 Scope

As of today, NLP tools from different domains have been prototypically included in the PARTHENOS D4Science VRE. These are a lemmatizer for 24 different languages which has been developed at the Center for Language and Technology (Center for SprogTeknologi, CST), partner of CLARIN-Denmark (CLARIN-DK), a service to create language models for this lemmatizer, also provided buy CLARIN-DK, as well as taggers, parsers, and named-entity recognizers for three different languages provided by the CLARIN-Poland (CLARIN-PL) Language Technology Center. These tools have been made accessible by the PARTHENOS Lab Data Miner.

8.4.3.3 Functionalities

For the lemmatization task, the CST Lemmatizer of CLARIN-DK was included[37]. With this tool the analysis step of lemmatization may be carried out for texts from 24 different languages. The languages supported are: Bulgarian, Danish, English, Estonian, Farsi, French, Greek, Icelandic, Italian, Latin, Macedonian, Dutch, Polish, Portuguese, Romanian, Russian, Slovakian, Serbian, Slovenian, Spanish, Czech, German, Ukrainian, and Hungarian. Lemmatization is the process of mapping an inflected word form (token) to its lexical base form (lemma), e.g. word forms run, runs, ran would be mapped to lemma run. The CST lemmatizer is one component of a tool set provided by CLARIN-DK. It performs the lemmatization task based on, first, a set of affix rules that was compiled via a supervised learning step, and second, optionally, on a dictionary to map word forms to base forms. It allows for plain or XML text as input.

The CLARIN-DK Create Models service can be used to train individual language models for the PCRF-seq2seq lemmatizer[38] based on training material uploaded by the user. For further analysis steps, tools hosted by the CLARIN-PL platform[39] were integrated which allow for the analysis of text in three different languages: English, German and Polish.

---

[37] CST lemmatizer https://github.com/kuhumcst/cstlemma (last visited 27/04/2018)
[38] PCRF-seq2seq lemmatizer https://github.com/UKPLab/coling2016-pcrf-seq2seq (last visited 27/04/2018)
[39] CLARIN-PL platform http://ws.clarin-pl.eu/ (last visited 27/04/2018)

The next analysis step after lemmatization may be part-of-speech (POS) tagging. In this step information about the word class is added to each token. For the POS-analysis step, tagging tools were integrated from the CLARIN-PL Language Technology Center for three languages: English, German, Polish. Thus, within the PARTHENOS Lab, the POS tagging task can be carried out for English texts, with the tagger of Python's NLTK (Natural Language Toolkit)[40], for the Polish Language by usage of the Wrocław CRF morpho-syntactic tagger (Wcrft2)[41], and for German texts with the tagger of the Spacy tool set[42].

Furthermore, the functionality of syntactical parsing is available via the PARTHENOS Lab. By this analysis step syntactical structures of texts and text-parts (noun/verb phrases, verb dependencies, …) are identified. For the English and German language the Spacy parser was integrated, for dependency parsing on Polish texts the Maltparser is now provided via the PARTHENOS Lab[43].

Finally, named-entity recognition (NER) tools have been integrated prototypically for the named languages. These tools comprise the NER of Spacy for English and German texts as well as the Liner2 NER tool for the Polish language. Potentially needed pre- and post-processing steps were also included according to the CLARIN-PL tool chains (e.g. text format conversion and morpho-syntactic analysis as pre-processing for the German NER).

8.4.3.4 Outcomes

The NLP services described above were integrated in the so-called DataMiner, a component of the PARTHENOS Lab. Due to the integration of NLP services from the CLARIN-PL suite, users of PARTHENOS are now able to perform various NLP tasks for the English, German, and Polish language as well as lemmatization for a wide range of languages in the PARTHENOS Lab on D4Science. The design of the interface and functionalities presumably has the potential of being prototypical for the integration of further similar services for other languages.

In the current state of development, users may upload to their private repositories on D4Science data they want to analyse and access these data as input for the respective NLP services from the DataMiner. In the course of the current assessment, raw text was used as input data. The output format for all POS taggers and Named-Entity-Recognizers

---

[40] Natural Language toolkit http://ws.clarin-pl.eu/tagerEnNLTK.shtml?en (last visited 27/04/2018)
[41] Wrocław CRF morpho-syntactic tagger (Wcrft2) http://ws.clarin-pl.eu/tager.shtml?en (last visited 27/04/2018)
[42] Spacy tool set http://ws.clarin-pl.eu/tagerDe.shtml?en (last visited 27/04/2018)
[43] PARTHENOS Lab http://ws.clarin-pl.eu/parser.shtml?en; http://maltparser.org/ (last visited 27/04/2018)

is the CCL format, a derivative of XCES[44]. CCL is an XML format for linguistic inline annotation. The CCL files provided by the DataMiner after text analysis contain:

- sentence information, i.e. tokens of each sentence grouped into an element <sentence>
- analysis information on token level
- element <tok> with children <orth> (word form), <lex> (lemma and POS information), and <ann> (other annotations)
- subelements of <lex>: <base> (lemma), <ctag> (part-of-speech)
- different attribute values of @chan in <ann>

As output for the attribute values of @chan different vocabularies are used by the different integrated services. Here, documentation on the correct interpretation of attribute values should be provided.

Output of the Parsers are .tsv documents according to the CoNLL-U format. This format is described in detail on the Universal Dependencies platform[45]. Information provided by the CoNLL-U format includes: word, lemma, token information, POS tags according to the Universal POS tagset and to language specific tagsets, morphological features (here: only for Polish texts), information on the head in dependency trees as well as on dependency relations according to the Universal Dependency Relations set[46] and information on missing space (here: only for English and German texts). Sentence boundary information is provided for English and German texts. The lemmatization produced as output a .tsv document, containing in each line the original token and its corresponding lemma. All analyses which were run in one session are stored as such and may be re-addressed during that session. The results are also automatically saved in the user's private workspace.

As a conclusion, we can state that the integration of the respective tools was successful. Texts were analysed successfully by usage of the DataMiner CLARIN NLP services with a decent processing time. The interface is not fully self-explanatory, but users can get to re-usable results with a reasonable learning curve. What would help immensely is short

---

[44] Documented in detail on http://nlp.pwr.wroc.pl/redmine/projects/corpus2/wiki/CCL_format (last visited 27/04/2018)
[45] Universal Dependencies platform http://universaldependencies.org/format.html (last visited 27/04/2018)
[46] Universal Dependency Relations set http://universaldependencies.org/u/dep/index.html (last visited 27/04/2018)

documentation on the most important components and features. The services and formats re-used by the DataMiner are all documented on the platforms of their origin. However, users shouldn't be obliged to gather the relevant information themselves on other platforms, especially since various services are integrated here at once and information on these services is scattered across several platforms. E.g. there already exists a screencast explaining the integration and usage of the services integrated from CLARIN-PL which should be openly provided on the platform. Also, further information on the integration to the DataMiner and possible consequences of that step for the users, should be provided. It might also help users to have tooltips or short explanations, e.g. on possible input formats.

8.4.3.5 Satisfied Requirements

First and foremost, CLARIN NLP within PARTHENOS DataMiner addresses requirements of data analysis, as e.g. expressed in use cases 3.6.2.3. OEAW_03, 3.6.2.6. CLARIN_01 and 3.6.3.5. SISMEL_01:

Use Case	Subscenario
3.6.2.3. OEAW_03 Language Related Studies	Take the accessed resource and use it for research
3.6.2.6. CLARIN_01 Language Related Studies	4d: Corpus analysis and comparison
3.6.3.5. SISMEL_01 History	4: Compare Text of Cassiodorus with others author's work
3.6.2.6. CLARIN_01 Language Related Studies	4c. Selection and inclusion of services for corpus analysis (wrt. linguistic, lexical, topical, etc. features) and corpus comparison

From an infrastructural perspective, two scenarios are addressed which were relevant in D2.1:

- VRESET_02: Integration of services in a domain-specific VRE
- VREUSE_05: Process a dataset and publish results

In addition, the fact that various CLARIN NLP services are now integrated in a PARTHENOS Lab VRE significantly augments the usability of these services. So, usage may now benefit from the working space that allows for easy publication and sharing of resources:

Use Case	Subscenario
3.6.2.6. CLARIN_01 Language Related Studies	4a. Creation of a working platform (VRE)
3.7.2.23. KNAW-DANS Social Sciences	Facilitate sharing, categorising, and indexing notes and annotations on sources at various levels
3.6.2.5. OEAW_05 Language Related Studies	Share private digitized material with selected researchers
3.6.2.6. CLARIN_01 Language Related Studies	4. Data analysis within a shared working platform
	5. Publication of results and provision of the corpora

In this respect, the following tasks are addressed here:

- VRESET_01: Set-up of a domain-specific VRE
- VREUSE_04: Private and public sharing of resources deposited in the VRE work-space

As for the KPIs and KPRs, mentioned above in Ch. 2.1.1, the following can be said for the CLARIN NLP integration in the PARTHENOS DataMiner.

KPI	Assessment
Functionality	The services are stable and provide extensive results in decent runtimes.
Scalability	The services can be launched on the user's data without primary programming skills or algorithmic knowledge in the field of NLP. The analysis data gained is diverse and may be useful for various research scenarios. Users gain output in standardized, documented and commonly known formats, hence allowing for their unproblematic interpretation and subsequent use.
Usability	Usability suffers from lack of documentation on design and clearer instructions about the use of the tool. Could be augmented by clear and targeted documentation as well as by the addition of tooltips and short descriptions at different steps in the analysis process.

Layout/design	for users who are familiar with the D4Science platform, once they know about the proper input formats, usage of the interface is straightforward. However, up till now, the CLARIN NLP set lacks documentation on its proper usage which may not be fully compensated by documentation on certain components of the CLARIN NLP which is scattered on different platforms.
Accessibility	This is addressed by the D4Science interface. The tool itself has no additional accessibility support.
Documentation	There is lack of documentation on the @chan vocabularies used as well as on the language-specific POS tagsets, which would be highly beneficial to the interpretation of results, especially to users from fields other than language-related studies. The lack of documentation is affecting also layout/design and usability.
Language accuracy	About language accuracy the integration is on the right track. The CLARIN-DK lemmatizer, as integrated into the DataMiner environment, supports 24 languages. Up till now, the services from NLP CLARIN which are integrated in the DataMiner are suitable for three different languages, Polish, English and German. Apart from a multilingual tagger, the CLARIN-PL platform, which these services were gained from, does not provide tools for other languages. However, it might be considered to include other NLP services for further languages, that are hosted by other CLARIN centres, into the CLARIN NLP suite of the PARTHENOS DataMiner.
Compatibility with project data	Users can upload and share data which can then be accessed from the DataMiner to be analysed by NLP CLARIN services. So, the service provided here is generally designed to be compatible with project data. Raw text data (UTF-8) can be processed without problems and will be analysed from scratch, that is, it is not necessary to preprocess the raw text material to be able to conduct the analysis. To evaluate the extent of compatibility with possible other formats, though, more information on supported input formats would be necessary.
Security	There are no further security issues apart from the normal usage D4Science allows for (sharing data, enhancing and storing data, etc.). Analysis results are stored automatically in the user's private workspace.

E. Systematic gap analysis

9. Users' needs vs. available tools

In this section we synthesized the results of the previous chapters representing the level of adequacy of the PARTHENOS infrastructure with respect to the users' expectations using a matrix composed of (i) the needs and requirements and (ii) the services/features offered. A SWOT diagram was also produced to support the next phase of development and integration of the PARTHENOS platform.

9.1 Infrastructural needs

Partner: PIN	
Infrastructure: DARIAH	
Requirement ID (Use Case): AR_06	
Brief description: An archive manager wants to prepare and register a new collection	
Service Required: Prepare a collection to be added to the collections held in the registry	**Service Provided by D4Science Platform:** REG_01: Manual registration of an entity in the PARTHENOS registry

Partner: OEAW	
Infrastructure: DARIAH, CLARIN	
Requirement ID (Use Case): OEAW_06	
Brief description: Development of a Framework as a dynamic environment of standards able to support the provision of language-(web)service interoperability.	
Service Required: Creation of specific VRE	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE

Partner: MIBACT-ICCU

Infrastructure: DARIAH

Requirement ID (Use Case): CI_03

Brief description: Repository update

Service Required: Update metadata of a content provider	**Service Provided by D4Science Platform:** REG_02: A research infrastructure joins PARTHENOS and integrates its registry; AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud; CURA_01: Subject coverage; CURA_03: Invite curation

Partner: MIBACT-ICCU

Infrastructure: DARIAH

Requirement ID (Use Case): CI_04

Brief description: Reporting system

Service Required: The system sending an email when the ingestion process is finished.	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS Infrastructure

Partner: MIBACT-ICCU

Infrastructure: DARIAH

Requirement ID (Use Case): CI_05

Brief description: Discard invalid metadata

Service Required: All the invalid metadata records are not ingested.	**Service Provided by D4Science Platform:** AGGR_01: Aggregate resource metadata from research Infrastructures into the PARTHENOS content cloud

Partner: KNAW-DANS

Infrastructure: DARIAH, CLARIN

Requirement ID: 3.7.1.27.KNAW-DANS_21

Brief description: Users of EHRI would benefit from having access to these along with the sources selected to assist in answering a question or addressing a particular topic.

Service Required: Facilitate sharing, categorising, and indexing research questions and/or topics	**Service Provided by D4Science Platform:** CURA_02: Invite new content providers

Partner: KNAW-DANS

Infrastructure: DARIAH, CLARIN

Requirement ID: 3.7.1.36.KNAW-DANS_30

Brief description: Studies reveal that functions facilitating early research, for example, finding, organizing, and displaying sources, are the most used and sought after in the scholarly community

Service Required: Tool for productively connecting with other researchers	**Service Provided by D4Science Platform:** VREUSE_01: Communication between users within VRE

Partner: KNAW-NIOD

Infrastructure: DARIAH

Requirement ID: NIOD_13

Brief description: Services to be able to authenticate and identify users and set authorization levels

Service Required: User authentication and authorization	**Service Provided by D4Science Platform:** VRESET_03: VRE authentication and authorization

Partner: : OEAW

Infrastructure: DARIAH, CLARIN

Requirement ID (Use Case): OEAW_05

Brief description: A researcher wants to discuss the found material with other researchers

Service Required: Share private digitized material with selected researchers	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace

Partner: : MIBACT-ICCU

Infrastructure: DARIAH,

Requirement ID (Use Case): CI_06

Brief description: Sharing ingested metadata

Service Required: Creation of an OAI-PMH repository to share ingested metadata in CulturaItalia.	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace

Partner: KNAW_NIOD

Infrastructure: DARIAH,

Requirement ID: NIOD_11

Brief description: Being able to share and collaborate with other researchers on EHRI documents

Service Required: See description	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace

9.2 Global needs (global tools)

For the on going work, with regard to functionalities of the catalogue, we propose to factor in the work that has been done in the context of existing services for certain Humanities' communities in other EU infrastructure projects, e.g. CLARIN's Virtual Language to Observatory (VLO). A large amount of the desiderata named by the community seems have been an issue for and were addressed by these services. For instance, the VLO already handles facetted search (based e.g. on keywords, language, availability, collections, organizations or country). There are options to perform simple, facetted and advanced searches for resources and to combine these search options, as well as to display and bookmark resources or to save searches. For each resource that is found, other, probably similarly interesting resources are displayed so that users may easily find related resources and collections. Data is harvested from international providers and various sources. Terms of availability are displayed and may be taken as a basis for filtering. The importer supports different formats, especially once these are modelled in CMDI profiles, and there is a huge, ongoing effort on harmonizing the resources included in the platform. The PARTHENOS project may thus profit from experiences gained with the development and design of the VLO and other, similar services, which might thus be used as reference points for the further development of the PARTHENOS Registry.

While the front-end is fairly self-explanatory, usability of the platform might be considerably enhanced by the provision of documentation, e.g. on how users may provide their records to the platform, on the supported formats, on features of the query language, and on the design of the catalogue in general (e.g. ranking of results, handling of relations between resources/resource records, ...).

Partner: PIN

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): AR_01

Brief description: An archaeologist wants to search/browse available data

Service Required: Get a list of institutions holding archaeological information concerning excavations, objects, periods	**Service Provided by D4Science Platform:** ACCESS_01: Search and browse the PARTHENOS registry	**Matched:** YES

Partner: PIN

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): AR_02

Brief description: An archaeologist wants to have a data preview

Service Required: See a preview of data available to allow a user determining the relevance of the data for his research	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: PIN

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): AR_03

Brief description: An archaeologist wants to access collections

Service Required: Access collections to compare information about burial site from the Iron Age with burial practice elsewhere in Europe.	**Service Provided by D4Science Platform:** ACCESS_03: Retrieval/ access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	**Matched:** YES

Partner: PIN		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): AR_04		
Brief description: An archaeologist wants to deposit data		
Service Required: Deposit some of the data produced in a PARTHENOS compatible archive.	**Service Provided by D4Science Platform:** VREUSE_03: Deposition	**Matched:** YES

Partner: PIN		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): AR_05		
Brief description: An archive manager wants to search and access the services registry.		
Service Required: Discover tools or best practices to achieve a certain goal	**Service Provided by D4Science Platform:** ACCESS_01: Search and browse the PARTHENOS registry	**Matched:** YES

Partner: PIN		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): AR_07		
Brief description: An archaeologist wants to inspect and enrich visual media documents		
Service Required: Inspect and enrich one of the visual documents stored in one of the catalogues available in the portal	**Service Provided by D4Science Platform:** VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: PIN

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): AR_08

Brief description: An archaeologist wants to access information about a metadata format

Service Required: Access information about metadata schemas and ontologies used for archiving archaeological resources	Service Provided by D4Science Platform: ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	Matched: YES

Partner: PIN

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): AR_09

Brief description: An archaeologist wants to retrieve information from vocabularies and gazetteers

Service Required: Retrieve information about collections and datasets according to specific terms from a vocabulary or retrieve a location from a gazetteer.	Service Provided by D4Science Platform: ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	Matched: YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_01		
Brief description: A research team has been established to produce a digital edition of Cassiodorus' Institutiones		
Service Required: 1: Get List of all the mss containing «Institutiones»	**Service Provided by D4Science Platform:** REG_01: Manual registration of an entity in the PARTHENOS registry	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_01		
Brief description: A research team has been established to produce a digital edition of Cassiodorus' Institutiones		
Service Required: 2: Search timeline Information about the Manuscripts	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_01		
Brief description: A research team has been established to produce a digital edition of Cassiodorus' Institutiones		
Service Required: 3: Search information about places, dates and bibliography	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_01		
Brief description: A research team has been established to produce a digital edition of Cassiodorus' Institutiones		
Service Required: 4: Compare Text of Cassiodorus with others author's work	**Service Provided by D4Science Platform:** VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_02		
Brief description: The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity.		
Service Required: 1: A researcher wants to find out how many collections of unedited sermons are held in European libraries	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_02		
Brief description: The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity.		
Service Required: 2: A researcher wants to publish a critical edition of 15th century Italian collections of sermons held in Italian libraries.	**Service Provided by D4Science Platform:** VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: SISMEL

Community: History

Requirement ID (Use Case): SISMEL_02

Brief description: The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity.

Service Required: 4: A researcher wants to search for Aristotle's quotations in 15th century sermons	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIS	**Matched:** YES

Partner: SISMEL

Community: History

Requirement ID (Use Case): SISMEL_02

Brief description: The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity.

Service Required: 5: A researcher wants to search for Plato's quotations in 15th century sermons	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL

Community: History

Requirement ID (Use Case): SISMEL_02

Brief description: The goal of the research is to find out how many Italian 15th century preachers employed philosophy in their preaching activity.

Service Required: 6: A researcher wants to search for classical, literary and philosophical quotations in late medieval sermons	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_03		
Brief description: A researcher is interested in finding information about Ramon Llull works and textual tradition		
Service Required: 1: A researcher wants to find all the works of Llull in which he employs the Ars combinatoria.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_03		
Brief description: A researcher is interested in finding information about Ramon Llull works and textual tradition		
Service Required: 2: He wants to find how many Llullian works are held in German Libraries.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_03		
Brief description: A researcher is interested in finding information about Ramon Llull works and textual tradition		
Service Required: 3: He wants to know how many Latin and Catalan works Llull wrote.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL **Community:** History **Requirement ID (Use Case):** SISMEL_03 **Brief description:** A researcher is interested in finding information about Ramon Llull works and textual tradition		
Service Required: 4: He wants to know when Llull travelled to Italy.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL **Community:** History **Requirement ID (Use Case):** SISMEL_03 **Brief description:** A researcher is interested in finding information about Ramon Llull works and textual tradition		
Service Required: 5: He wants to know what other works are copied with the Liber de amico et amato.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL **Community:** History **Requirement ID (Use Case):** SISMEL_04 **Brief description:** Tracking of the circulation of the legend of Barlaam and Josaphat		
Service Required: 1: Access to all existent editions of the manuscript in the Gallica, BVMM, e-codices, manuscriptorium databases.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_04		
Brief description: Tracking of the circulation of the legend of Barlaam and Josaphat		
Service Required: 2: Access to all existent editions of the prints from Incunabula Short-Title Catalogue (ISTC) and MEI (Material Evidence in Incunabula).	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_04		
Brief description: Tracking of the circulation of the legend of Barlaam and Josaphat		
Service Required: 3: Assess how many medieval and renaissance manuscripts of this legend survive today in our libraries using the META-OPAC CERL Portal to access a wide number of electronic catalogues of manuscripts.	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_05		
Brief description: A researcher wants to know who was the writer of the first western treatise on how retard old age, to collect his possible sources and reconstruct the diffusion of his work.		
Service Required: 1: Trace the manuscripts in Latin or in French thanks to the search engine TRAME	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_05		
Brief description: A researcher wants to know who was the writer of the first western treatise on how retard old age, to collect his possible sources and reconstruct the diffusion of his work.		
Service Required: 4: Undertake a systematic research from manuscripts to print, using the Incunabula Short-Title Catalogue (ISTC), the ERC program on "The 15th-century Book Trade", the English Short Title Catalogue and the digital collections of libraries	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: SISMEL

Community: History

Requirement ID (Use Case): SISMEL_06

Brief description: A historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period.

Service Required: 1: Survey all existent editions of Donatus	Service Provided by D4Science Platform: ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	Matched: YES

Partner: SISMEL

Community: History

Requirement ID (Use Case): SISMEL_06

Brief description: A historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period.

Service Required: 2: Assess the 15th and 16th-century use of these editions	Service Provided by D4Science Platform: ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	Matched: YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_06		
Brief description: A historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period.		
Service Required: 3: Assess how many medieval and renaissance manuscripts of this work survive today in our libraries using the META-OPAC CERL Portal to access a wide number of electronic catalogues of manuscripts	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: SISMEL		
Community: History		
Requirement ID (Use Case): SISMEL_06		
Brief description: A historian wants to track the dissemination of a given author's works during the Medieval and Early Modern period.		
Service Required: 4: Assess the presence of this work in catalogues of medieval libraries in Europe, to understand the popularity and circulation of this work in the medieval and early modern period by using TRAME tool.	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_01		
Brief description: Being able to visualize objects and location		
Service Required: Visualize search results	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_02		
Brief description: Being able to visualize search paths		
Service Required: Visualize search results	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_03		
Brief description: Being able to use tools to crowdsource translation of (archival) documents		
Service Required: See description	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_04		
Brief description: Being able to map archive location and type in a tool		
Service Required: Visualize search results	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_05		
Brief description: The researcher can understand and display the spatial or chronological relationships between documents		
Service Required: Visualize search results	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	Matched: YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_06		
Brief description: The researcher can understand and present the spatial and chronological relationships between documents		
Service Required: Visualize search results	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD		
Community: History		
Requirement ID: NIOD_07		
Brief description: Researcher can use CENDARI material in a presentation or publication without having to figure out citation format		
Service Required: Citation of dataset	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-NIOD

Community: History

Requirement ID: NIOD_08

Brief description: Create a schedule/calendar for how much time the researcher will need for each archive, when the archives are open, and their contact details, national/religious holidays when they will be closed, etc. Link the results of my searches (archives I want to visit and when) to real-world information for planning (calendars, travel websites for airline, fares and train reservations, hotels, etc.)

Service Required: Planning of research	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	Matched: YES

Partner: KNAW-NIOD

Community: History

Requirement ID: NIOD_09

Brief description: Being able to search (keywords, persons, location/geographical information, events, time/dates ...)

Service Required: Enable (federated) search services	Service Provided by D4Science Platform: ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	Matched: YES

Partner: KNAW-NIOD

Community: History

Requirement ID: NIOD_10

Brief description: Being able to upload/harvest/integrate data from CHI into EHRI

Service Required: See description	Service Provided by D4Science Platform: REG_02: A research infra joins PARTHENOS and integrate its registry	Matched: YES

Partner: KNAW-NIOD

Community: History

Requirement ID: NIOD_12

Brief description: Being able to present data in an online portal

Service Required: See description	Service Provided by D4Science Platform: VREUSE_05: Process a dataset and publish results	Matched: YES

Partner: OEAW

Community: Language-related Studies

Requirement ID (Use Case): OEAW_01

Brief description: A researcher wants to search/browse available metadata about language resources using a combination of the faceted and text search.

Service Required: Search for resources	Service Provided by D4Science Platform: ACCESS_01: Search and browse the PARTHENOS registry	Matched: YES

Partner: OEAW		
Community: Language-related Studies		
Requirement ID (Use Case): OEAW_01		
Brief description: A researcher wants to search/browse available metadata about language resources using a combination of the faceted and text search.		
Service Required: Search for resources	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: OEAW		
Community: Language-related Studies		
Requirement ID (Use Case): OEAW_02		
Brief description: A researcher wants to preview found resources.		
Service Required: Show metadata of search results	**Service Provided by D4Science Platform:** ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	**Matched:** YES

Partner: OEAW		
Community: Language-related Studies		
Requirement ID (Use Case): OEAW_03		
Brief description: A researcher wants to access a resource.		
Service Required: Access resource from search results	**Service Provided by D4Science Platform:** ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	**Matched:** YES

Partner: OEAW		
Community: Language-related Studies		
Requirement ID (Use Case): OEAW_04		
Brief description: A researcher wants to find non-digitised material in physical archives.		
Service Required: Search for material with help of metadata	**Service Provided by D4Science Platform:** ACCESS_01: Search and browse the PARTHENOS registry	**Matched:** YES

Partner: OEAW		
Community: Language-related Studies		
Requirement ID (Use Case): OEAW_04		
Brief description: A researcher finds non-digitised material in a physical archive		
Service Required: Find new material on location	**Service Provided by D4Science Platform:** REG_01: Manual registration of an entity in the PARTHENOS registry	**Matched:** YES

Partner: OEAW **Community:** Language-related Studies **Requirement ID (Use Case):** OEAW_05 **Brief description:** A researcher wants to reference and comment non-digitized material.		
Service Required: Reference and comment non-digitized material	**Service Provided by D4Science Platform:** VREUSE_02: Reference entities of the PARTHENOS registry or resources of the PARTHENOS content cloud in VRE posts	**Matched:** YES

Partner: OEAW **Community:** Language-related Studies **Requirement ID (Use Case):** OEAW_05 **Brief description:** A researcher wants to make notes on a private digitized selection of the found material		
Service Required: Upload private digitized material in a private VRE	**Service Provided by D4Science Platform:** VREUSE_03: Deposition	**Matched:** YES

Partner: TCD Community: History Requirement ID (Use Case): 3.6.1.16.TCD01 Brief description: Researcher wishes to extract data on battlefield transportation from Europeana 1914-18 via the Europeana RESTful API and make notes on selected outputs of the API call in the CENDARI NTE in order to create a set of metadata for broad analysis		
Service Required: Extract data via API	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: TCD		
Community: History		
Requirement ID (Use Case): 3.6.1.16.TCD01		
Brief description: Researcher wishes to extract data on battlefield transportation from Europeana 1914-18 via the Europeana RESTful API and make notes on selected outputs of the API call in the CENDARI NTE in order to create a set of metadata for broad analysis		
Service Required: import (Previous) into note taking environment	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: TCD		
Community: History		
Requirement ID (Use Case): 3.6.1.17.TCD02		
Brief description: Researcher wants to gather testimonies from refugees following the Hungarian Revolution of 1956. Wants to conduct a search on the Europeana Portal to look for digital content that she can download and analyse		
Service Required: Extract data via platform portal	**Service Provided by D4Science Platform:** ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	**Matched:** YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: Search for content	Service Provided by D4Science Platform: ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	Matched: YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: Extract content	Service Provided by D4Science Platform: ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	Matched: YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: integrate TRAME search portal	Service Provided by D4Science Platform: REG_01: Manual registration of an entity in the PARTHENOS registry	Matched: YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: Use Trame search portal	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure	Matched: YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: Download the data	Service Provided by D4Science Platform: AGGR_02: Export metadata via standard protocols	Matched: YES

Partner: TCD

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD03

Brief description: Lecturer wants to find materials for workshop with group of History Postgrads looking into Medieval attitudes to women. Uses TRAME to search for documents that show accounts of women. She will use these in a workshop looking at sentiment analysis tools, but needs to be able to download the data.

Service Required: Use the Sentiment analysis tool	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure	Matched: YES

Partner: MIBACT-ICCU		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): MINT_03		
Brief description: A user wants to check the metadata		
Service Required: A user checks the metadata aggregated via the MINT mapping tool.	**Service Provided by D4Science Platform:** AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud	**Matched:** YES

Partner: MIBACT-ICCU		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): MINT_04		
Brief description: A user wants to preview metadata in html format		
Service Required: A user wants to preview metadata in html format	**Service Provided by D4Science Platform:** VREUSE_03: Deposition	**Matched:** YES

Partner: MIBACT-ICCU		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): MINT_05		
Brief description: A user wants to enrich metadata aggregated in MINT using external SKOS thesauri		
Service Required: A user wants to enrich metadata aggregated in MINT using external SKOS thesauri.	**Service Provided by D4Science Platform:** AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud	**Matched:** YES

Partner: MIBACT-ICCU		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): CI_01		
Brief description: Metadata harvesting		
Service Required: Metadata acquisition of a data provider within CulturaItalia; metadata are shown in the Portal.	**Service Provided by D4Science Platform:** ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	**Matched:** YES

Partner: MIBACT-ICCU		
Community: Cultural Heritage, applied disciplines, and Archaeology		
Requirement ID (Use Case): CI_02		
Brief description: Metadata validation		
Service Required: The harvester system provides an automatic check during ingestion process.	**Service Provided by D4Science Platform:** AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 1. Determination of the research question and goal	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 2. Selection of the NP (and NP issues) of primary interest; creation of the primary corpus	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 3. Detection, selection and, if necessary, amendment of relevant data for comparison, from (a) other NPs, (b) other documents than NPs	**Service Provided by D4Science Platform:**	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 3a. Detection of relevant data sets	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 3b. Gathering of the metadata for the relevant data sets	**Service Provided by D4Science Platform:** ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 3c. Gathering of the relevant data sets	**Service Provided by D4Science Platform:** ACCESS_04: Retrieval/access of resources from the PARTHENOS content cloud	**Matched:** YES

Partner: CLARIN-BBAW

Community: Language-related Studies

Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)

Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies

Service Required: 4a. Creation of a working platform (VRE)	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE	**Matched:** YES

Partner: CLARIN-BBAW

Community: Language-related Studies

Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)

Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies

Service Required: 4b. Inclusion of the relevant data into the VRE	**Service Provided by D4Science Platform:** VREUSE_03: Deposition	**Matched:** YES

Partner: CLARIN-BBAW		
Community: Language-related Studies		
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)		
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies		
Service Required: 5. Publication of results and provision of the corpora	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.7 KNAW-DANS_01		
Brief description: The framework should provide a storage management module for the configuration of the storage back-ends to be used. Depending on the functional requirements of the target Enhanced Publication Information System (EPIS), a type of back-end, may be preferable to another.		
Service Required: Supporting different back-ends for data storage	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.8. KNAW-DANS_02

Brief description: The framework should provide a language for the definition of EP (enriched publications) data models (EP-DMDL, EP Data Model Definition Language)

Service Required: Offering data definition, manipulation, and access languages	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.9. KNAW-DANS_03

Brief description: Making manipulation of resources possible whose types are defined in the EP (enriched publications) data model (EP-DSML, EP Domain Specific Manipulation Language).

Service Required: Able to operate on EP instances (compliant to the defined EP data model) with a dedicated domain-specific language	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.10. KNAW-DANS_04

Brief description: Need for supporting the export of content via different standard APIs and protocols to serve third-party applications.

Service Required: Enabling data sharing	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.11. KNAW-DANS_05		
Brief description: Support is needed for open standards for the representation of data		
Service Required: Supporting data portability	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.13. KNAW-DANS_07		
Brief description: Data source management functionality is needed to ease the administrative operations needed to take care of the dynamic nature of the data sources.		
Service Required: Support the management of dynamic data sources.	**Service Provided by D4Science Platform:** AGGR_01: Aggregate resource metadata from research infrastructures into the PARTHENOS content cloud; REG_02: A research infra joins PARTHENOS and integrate its registry; CURA_01: Subject coverage	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.17.KNAW-DANS_11

Brief description: Based on the requirements of existing EPISs

Service Required: Supporting the addition of new domain-specific functionalities	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure	Matched: YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.18.KNAW-DANS_12

Brief description: Requirements are needed which state that some sensitive information may not be disclosed to unauthorized parties.

Service Required: Confidentiality	Service Provided by D4Science Platform: VRESET_03: VRE authentication and authorization	Matched: YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.19.KNAW-DANS_13

Brief description: Requirements are needed which state that some information or resource can be used at any point in time when it is needed and its usage is authorized.

Service Required: Availability	Service Provided by D4Science Platform: VRESET_03: VRE authentication and authorization	Matched: YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.20.KNAW-DANS_14		
Brief description: Requirements are needed which constrain the software to operate as expected over long periods of time.		
Service Required: Reliability	**Service Provided by D4Science Platform:**	**Matched:** NO

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.21.KNAW-DANS_15		
Brief description: Requirements are needed which constrain the state of the information processed by the software to reflect the state of the corresponding physical information in the environment accurately.		
Service Required: Accuracy	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure; CURA_03: Invite curation	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.22.KNAW-DANS_16		
Brief description: For human interaction, usability requirements are needed which pre-scribe input/output formats and user dialogues to fit the abstractions, abilities and expectations of the target users.		
Service Required: Usability	**Service Provided by D4Science Platform:** VRESET_02: Integration of services in the PARTHENOS infrastructure	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.23.KNAW-DANS_17

Brief description: Requirements are needed which impose structural constraints on the software-to-be to fit its environment

Service Required: Architectural	Service Provided by D4Science Platform: VRESET_02: Integration of services in the PARTHENOS infrastructure	Matched: YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.25.KNAW-DANS_19

Brief description: Providing this information helps researchers to be prepared for working in an archive

Service Required: Archival information on how archives manage and describe the holdings	Service Provided by D4Science Platform: ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud	Matched: YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.26.KNAW-DANS_20		
Brief description: Providing this information helps to enable the researchers to undertake an initial assessment of the value of the archival holdings for their research		
Service Required: As much archival information as possible on archival holdings	**Service Provided by D4Science Platform:** ACCESS_03: Retrieval/access metadata about an entity of the PARTHENOS registry or a resource in the PARTHENOS content cloud; CURA_03: Invite curation	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.28.KNAW-DANS_22		
Brief description: Users of EHRI would benefit from having access to these along with the sources selected to assist in answering a question or addressing a particular topic.		
Service Required: Facilitate sharing, categorising, and indexing additional information about a research project	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE; CURA_03: Invite curation	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.29.KNAW-DANS_23		
Brief description: They are perceived as valuable for research		
Service Required: Facilitate sharing, categorising, and indexing notes and annotations on sources at various levels	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE; CURA_03: Invite curation	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.30.KNAW-DANS_24

Brief description: To assist in the 'chaining' process of moving from published works to other works, and to archival sources

Service Required: Facilitate sharing, categorising, and indexing details (citations) of researchers; publications	Service Provided by D4Science Platform: VRESET_01: Set-up of a domain-specific VRE; CURA_03: Invite curation	Matched: YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.31.KNAW-DANS_25

Brief description: [Not available]

Service Required: Facilitate sharing, categorising, and indexing of re-searcher bibliographies	Service Provided by D4Science Platform: VRESET_01: Set-up of a domain-specific VRE; CURA_03: Invite curation	Matched: YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.32.KNAW-DANS_26

Brief description: Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community.

Service Required: Tool for finding sources	Service Provided by D4Science Platform: ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	Matched: YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.33.KNAW-DANS_27		
Brief description: Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community. This allows the user to take dynamic notes, organize them in useful ways, and link her/his research to data in the CENDARI data space.		
Service Required: Tool for organising resource	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE; CURA_03: Invite curation	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.34.KNAW-DANS_28		
Brief description: Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community. This not only displays the user's research in provoking ways, but can also reveal connections and patterns that may inform the conclusions of his/her research or guide further research		
Service Required: Tool for displaying own research	**Service Provided by D4Science Platform:** VRESET_01: Set-up of a domain-specific VRE; VRESET_02: Integration of services in the PARTHENOS infrastructure; VREUSE_05: Process a dataset and publish results	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.35.KNAW-DANS_29

Brief description: Studies reveal that functions facilitating early research for example, finding, organizing, and displaying sources are the most used and sought after in the scholarly community.

Service Required: Tool for accessing existing data resources (published)	**Service Provided by D4Science Platform:** VREUSE_04: Private and public sharing of resources deposited in the VRE workspace	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.37.KNAW-DANS_31

Brief description: To find institutions and collections and other information

Service Required: Tools to search and browse at a general level	**Service Provided by D4Science Platform:** ACCESS_01: Search and browse the PARTHENOS registry	**Matched:** YES

Partner: KNAW-DANS

Community: History

Requirement ID: 3.7.1.38.KNAW-DANS_32

Brief description: To find detailed information on a given subject/work

Service Required: Tools to search and browse at a more detailed level	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.39.KNAW-DANS_33		
Brief description: To help the user in his research work		
Service Required: Provide indication of the language and the place in which works were written (origin/provenance)	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.40. KNAW-DANS_34		
Brief description: To help the user in his research work		
Service Required: Provide the year of composition of a work	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.41.KNAW-DANS_35		
Brief description: To help the user in his research work		
Service Required: Show the availability of printed editions of a work	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.42.KNAW-DANS_36		
Brief description: To help the user in his research work		
Service Required: Show manuscripts related to a work	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several research infrastructures	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.43. KNAW-DANS_37		
Brief description: To help the user in his research work		
Service Required: Show author, place and time of translations of a work	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.44.KNAW-DANS_38		
Brief description: To help the user in his research work		
Service Required: Show availability of digital objects related to a work	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

Partner: KNAW-DANS		
Community: History		
Requirement ID: 3.7.1.45.KNAW-DANS_39		
Brief description: To help the user in his research work		
Service Required: Show a bibliography	**Service Provided by D4Science Platform:** ACCESS_02: Search and browse the PARTHENOS content cloud across several RIs	**Matched:** YES

9.3 Specialized Services (PARTHENOS Lab)

9.3.1 History

Partner: KNAW-DANS	
Infrastructure: DARIAH, CLARIN	
Community: History	
Requirement ID: 3.7.1.16.KNAW-DANS_10	
Brief description: To create a high quality content it is needed to better the quality of the EPs (enhanced publications) and enrich the original content	
Service Required: Support the enrichment and curation of content	**Specialized tools providing the service:** NERLiX, CLARIN NER, CLARIN PARSER, CLARIN TAGGER, CLARIN LEMMATIZER

Partner: KNAW-DANS	
Infrastructure: DARIAH, CLARIN	
Community: History	
Requirement ID: 3.7.1.24.KNAW-DANS_18	
Brief description: Providing as much information as possible about archives that hold collections of interest helps researchers to be prepared	
Service Required: Archival information on collections	**Specialized tools providing the service:** NERLiX, CLARIN NER, CLARIN PARSER, CLARIN TAGGER, CLARIN LEMMATIZER

Partner: KNAW-DANS

Infrastructure: DARIAH, CLARIN

Community: History

Requirement ID: 3.7.1.46.KNAW-DANS_40

Brief description: To enable new forms of research and discovery

Service Required: Advanced tools for research and discovery	**Specialized tools providing the service:** RUBRICA, NERLIX, CLARIN NER, CLARIN PARSER, CLARIN TAGGER, CLARIN LEMMATIZER

Partner: TCD

Infrastructure: DARIAH

Community: History

Requirement ID (Use Case): 3.6.1.17.TCD04

Brief description: Researcher wants to structure their multiple datasets in order to make it interoperable, and uses the FlareNet guidelines to determine appropriate standards

Service Required: Preparation of multiple datasets	**Specialized tools providing the service:** RUBRICA

9.3.2 Language Related Studies

Partner: OEAW

Infrastructure: DARIAH, CLARIN

Community: Language-related Studies

Requirement ID (Use Case): OEAW_03

Brief description: A researcher wants to use a resource for research

Service Required: Take the accessed resource and use it for research	**Specialized tools providing the service:** RUBRICA, NERLiX, CLARIN NER, CLARIN PARSER, CLARIN TAGGER, CLARIN LEMMATIZER

Partner: CLARIN-BBAW
Infrastructure: CLARIN
Community: Language-related Studies
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies

Service Required: 4c. Selection and inclusion of services for corpus analysis (wrt linguistic, lexical, topical, etc. features) and corpus comparison	**Specialized tools providing the service:** RUBRICA

Partner: CLARIN-BBAW
Infrastructure: CLARIN
Community: Language-related Studies
Requirement ID (Use Case): Corpus-based Analysis of Historical Newspapers (Label: Hist. NPs)
Brief description: Comparative analysis for different newspapers (NPs) with regard to specialities in language, to ways of reporting on certain topics, events or discourses, to political tendencies

Service Required: 4d. Corpus analysis and comparison	**Specialized tools providing the service:** RUBRICA, NERLiX, CLARIN NER, CLARIN PARSER, CLARIN TAGGER, CLARIN LEMMATIZER

9.3.3 Archaeology, Heritage and Applied Disciplines

Partner: MIBACT-ICCU

Infrastructure: DARIAH

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): MINT_01

Brief description: A user, acting as a content provider, is willing to aggregate a number of different metadata sets, providing them as one unified set.

| **Service Required:** A content provider aggregates its metadata (according to a specific data model) and disseminates them via OAI-PMH. | **Specialized tools providing the service:** RUBRICA |

Partner: MIBACT-ICCU

Infrastructure: DARIAH

Community: Cultural Heritage, applied disciplines, and Archaeology

Requirement ID (Use Case): MINT_02

Brief description: A user that manages an OAI-PMH repository, acting as a content provider, is willing to aggregate a number of different metadata, providing it as one unified set.

| **Service Required:** A content provider aggregates their metadata (according to a specific standard) and disseminates them via OAI-PMH. | **Specialized tools providing the service:** RUBRICA |

9.4 Value proposition

The overall mission of the PARTHENOS infrastructure is to provide sustainable access to a digital ecosystem with a variety of digital datasets and services, to a broad range of communities in the areas of Cultural Heritage, Humanities and Social Sciences (abbreviated CHHSS). In this section we will focus on the digital data and services created and maintained in the PARTHENOS Project under the auspices of WP5 and WP6, and try to describe the elements that, together, constitute the PARTHENOS Value Proposition for this part of the work done in the project.

In order to do that, we will first briefly describe the main target audience(s) of these activities, as different stakeholders and audiences may have different interests and expectations, and may attach different values to what we offer. We will then for each of them summarize the main benefits that PARTHENOS has to offer, and then initiate the discussion on the crucial question which bodies or organisations will ensure the sustainability of what will be offered to the communities after completion of the PARTHENOS Project in 2019.

9.4.1 Audiences and stakeholders

Our ultimate target audience consists of researchers and practitioners in the broad field of Humanities and Social Sciences – possibly extending to Heritage Science, but as the WP2 work in the first year has shown, their requirements and expectations vary widely per discipline, and often even within disciplines. For each of them, the ideal Value Proposition may in part overlap and in part look different. Some tools and facilities may be indispensable for everybody but remain totally invisible to end users.

We will therefore follow the same subdivision as in the earlier sections of this and look at:

- General Collaboration facilities
- General purpose services and tools

and specialized services for:

- historical studies
- language related studies
- archaeology, heritage and applied disciplines
- social sciences.

Even if the individual researchers and practitioners are our main audience our Value Proposition will also need to address organized thematic communities that may or may not

come with their own pre-existing infrastructure facilities, as they are the ones that in the end will have to decide whether or not to take on the PARTHENOS heritage, be it by adopting the facilities as users or be it by taking responsibility for the future operation and maintenance of all or parts of the PARTHENOS Infrastructure. We will use the term stakeholders to refer to them.

9.4.2 General collaboration facilities

For individual users, their institutions and other stakeholders, the following collaboration facilities are offered:

- Collaboration space: the PARTHENOS Dataspace, based on CIDOC-CRM, which not only provides a common space for data (and metadata) produced by different institutions of different fields, but also overcomes the problems related to data integration, providing tool(s) able to present data and metadata in a homogeneous and reasonable manner for the end users.

- Interoperability tools: D-NET provides a set of services and tools for the creation of information systems where organizations and researchers can find the tools to integrate their data sources to form uniform and valuable information spaces of object metadata descriptions, including data mediation, data mapping, data storage and indexing, data curation and enrichment, and data provision. D-NET provides also a Metadata Inspector service and a Metadata Cleaner service.

- VRE management: A Virtual Research Environment (VRE) is a Web-based working environment that it is tailored to serve the needs of a Community of Practice. It is expected to provide a community of practice with the whole array of commodities needed to accomplish the community's goal; The VRE is open and flexible with respect to the overall service offering and lifetime and promotes fine-grained controlled sharing of both intermediate and final research results by guaranteeing ownership, provenance, and attribution.

- Social Network Services: The Social networking services complement the PARTHENOS infrastructure offering by promoting the cooperation among users. Specifically, they allow to share posts and start discussions, enabling users to comment, subscribe or re-share these posts. Attached to the content of a post, generally a text, one can add interesting links, also pointing to deposited files in the VRE workspace.

9.4.3 General purpose services and tools

Given the huge amount of datasets that are expected to be available in the PARTHENOS data space, global search tools are a crucial element for finding useful datasets to a research question. A set of applications and tools providing access to global resources (i.e. not specialised services) has been deployed in the PARTHENOS infrastructure. Three applications are available for the end-user:

- The PARTHENOS Catalogue: The PARTHENOS Catalogue is a tool able to allow the project's partners to share their resources to make them easily searchable. The PARTHENOS Catalogue contains options for browsing according to resource types, groups providing resources, formats and keywords (tags);

- The PARTHENOS Discovery Application allows the end user to surf the projects resources using different navigation method. In addition to traditional textual research, in fact, it is possible to apply a series of filters, through the faceted search, which allow the user a better browsing experience and a more refined search. This system is widely adopted in the field of Digital Humanities because it is possible, for end users, to extract in few steps the information they are looking for;

- The SPARQL endpoint: The SPARQL endpoint has been integrated to provide access - for collection managers - to the PARTHENOS contents. The SPARQL endpoint uses the RDF language and describes the relations between objects allowing the creation of customized queries based on the semantic structure of the data that can be downloaded in various format.

9.4.4 Integration of domain and community specific services and tools

Driving the creation of new technologies, methodologies, and information systems, communities of digital humanities, through the collection of use cases, have identified the need to integrate in the PARTHENOS infrastructure a selected cluster of tools and specialized services which, due to the interdisciplinary nature of the project, will be expanded, available and usable in the different communities.

Three modalities of integration of services are possible in the D4Science platform:

- The highest level of integration is Gold. The infrastructure provides the Community Service with authentication, authorization, monitoring and

accounting. However, the Service must be written according to predetermined policies and run over an authorised Web Container (SmartGears).

- The second level of integration is Silver: the Service will run on any platform and use a User Token. The user can contact D4Science tool and can perform operations on behalf of other users.

- In the lowest level of integration, Bronze, the Community Service will run on any platform and use an Application Token. The user can contact D4Science services but cannot perform operations on behalf of the users.

Silver and Bronze Integration are excellent means to make services hosted by other infrastructures available to the users of the PARTHENOS platform and vice versa, thus being an important enabler for cross-discipline research.

9.4.5 Domain specific tools and services already integrated in PARTHENOS

- RUBRICA: Within the PARTHENOS project a virtual research environment, RUBRICA (Reference Resources Integration plAtform), is being developed. RUBRICA aims to foster the interoperability and integration of various reference resources used in different disciplines. Starting from trusted knowledge bases (i.e.: databases, thesauri, authority lists etc.) researchers could create, merge, edit and reuse specialized reference resources, developed according to specific research purposes, without performing repetitive tasks on each resource. RUBRICA also allows to share this knowledge base with other users thanks to the IT structure on which PARTHENOS is based.

- NERLiX: NERLiX stands for Named Entity Recognition and Entity Linking and Extraction. As a service in PARTHENOS it covers different steps for the enrichment of texts and/or the extraction of (meta)data from text. Named Entity Recognition (NER) is frequently used in Digital Humanities when it comes to text analysis. When working with a lot of text documents or with a long text document, NER can help to give a brief overview on topics in a text document without reading it in detail. Also, NER can extract such topics out of a bunch of text documents, that can be further processed, e.g. with Social Network Analysis (SNA). There are a lot more of use cases where NER is the starting point of the analysis.

145

- NLP tools: NLP tools from different domains have been prototypically included in the PARTHENOS D4Science VRE. These are a lemmatizer for 24 different languages which has been developed at the Center for Language and Technology (Center for SprogTeknologi, CST), partner of CLARIN-Denmark (CLARIN-DK), a service to create language models for this lemmatizer, as well as taggers, parsers, and named-entity recognizers for three different languages provided by CLARIN-Poland (CLARIN-PL). These tools have been made accessible by the PARTHENOS Lab Data Miner.

9.4.6 What value from whom and for whom?

In order for a Value Proposition to make sense, there has to be an audience of potentially interested users and it has to be backed by some entity or group of entities that are prepared and able to secure the operation, maintenance and further development of its offerings in a sustainable way over a significant period.

There is no doubt that the tools and services currently being offered by PARTHENOS can be extremely valuable for the community of SSHs and – possibly - also for researchers in the Heritage Science sector, practitioners (and students!) at large, as well as for communities that have no or limited infrastructure facilities of their own, and, as things stand, the volume of available services and tools will only increase in the period between now and the end of the project.

New or existing research or research infrastructure projects might very well benefit from the fact that adopting an existing platform may save a lot of human and financial resources and leave more room to concentrate on content matters and research questions, but business models for this will have to be elaborated.

For communities that have their own technical research infrastructures to rely on, or for organisations that have as their main mission to operate infrastructures on a more permanent basis (such as CLARIN, DARIAH, CESSDA, SHARE, ESS and E-RIHS) there is certainly a significant added value in the fact that the PARTHENOS platform allows for cross-linking of services, and exchanges of data to support cross-discipline research, but it is by no means sure that they would consider abandoning the platforms that they have already created for themselves and move their operations to the PARTHENOS platform. But the situation may be different for different stakeholders.

In order to ensure that the investments made by the community in PARTHENOS and its technical platform do not get lost, the main questions to be addressed in the very near future at the project management level are, of course:

- What happens to the PARTHENOS heritage once the project is over;
- Will there be funding available for further operation and development of the platform;
- Will there be possibilities (financial and organisational) for PARTHENOS to be 'lifted' to the EOSC;
- Will there be sister infrastructures or other players who will be interested in 'adopting' (and paying for) all or parts of the platform (rather than building their own from scratch, or continuing to use their own), and what would be appropriate business models.

9.5 Conclusions & Next Steps

Based on the previous chapters, a SWOT diagram has been elaborated to spot the Strengths, Weaknesses, Opportunities and Threats for the PARTHENOS infrastructure so far, and guide the next phase of its development, towards the production version.

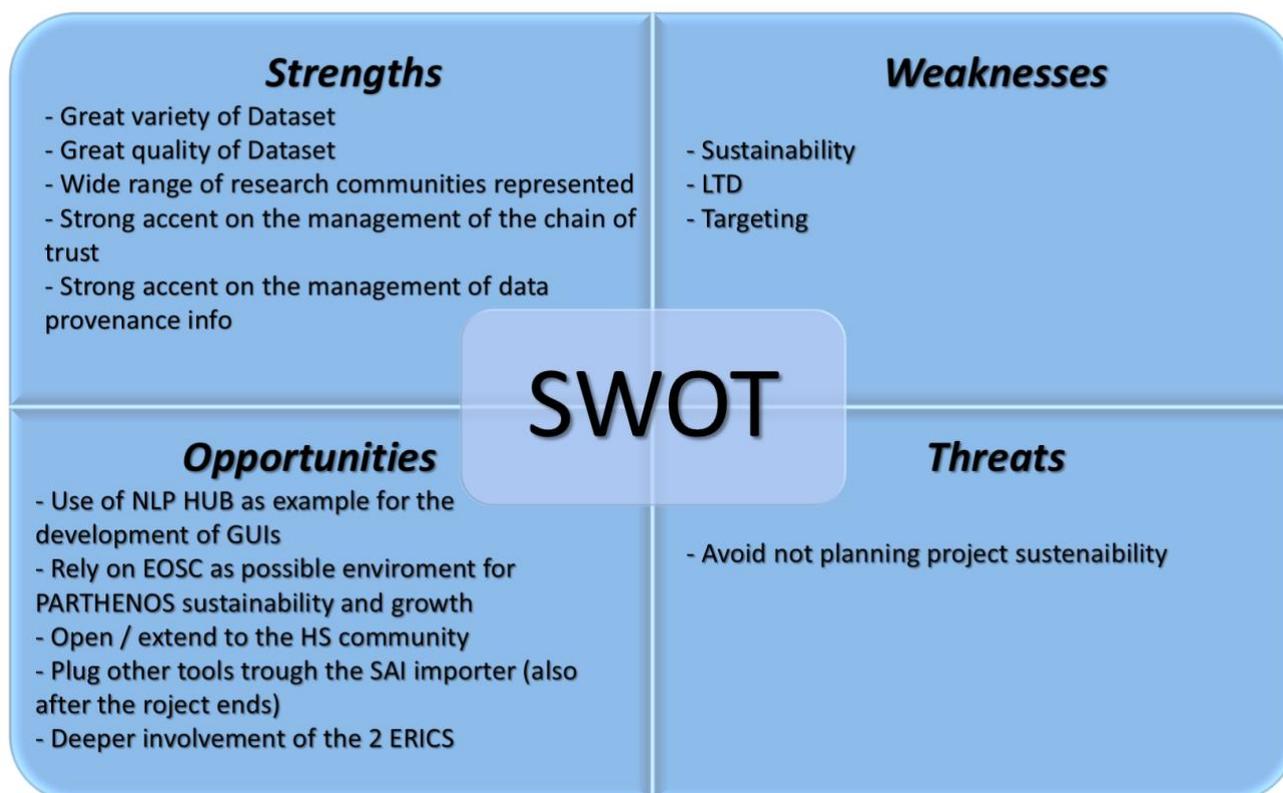We considered strong points for PARTHENOS to be:

- the vast availability of datasets coming from different research domains;
- the overall quality of the data gathered by the participating infrastructures;
- the attention payed to modeling and managing the data chain of trust and provenance model.

On the other hand, we also spotted a few points where a focused set of activities during the final development period could bring relevant improvements to the whole system (see the attached SWOT analysis, summarizing the assessment's most relevant findings):

- the need to develop a coherent and consistent set of GUIs to improve the usability and user experience;
- the need to develop a suitable sustainability plan for the resources and the platforms, also involving the 2 ERICs in the S&CI field;
- the need to address Long Term Preservation issues for the resources produced and gathered;

- the need to better target the system towards the scientific communities.

Strengths
- Great variety of Dataset
- Great quality of Dataset
- Wide range of research communities represented
- Strong accent on the management of the chain of trust
- Strong accent on the management of data provenance info

Weaknesses
- Sustainability
- LTD
- Targeting

Opportunities
- Use of NLP HUB as example for the development of GUIs
- Rely on EOSC as possible enviroment for PARTHENOS sustainability and growth
- Open / extend to the HS community
- Plug other tools trough the SAI importer (also after the roject ends)
- Deeper involvement of the 2 ERICS

Threats
- Avoid not planning project sustenaibility

SWOT

Concerning the possible threats for the PARTHENOS infrastructure, the key issue is the risk of not exploiting its full potential due to the lack of resources to sustain its maintenance and growth, both from the technical and scientific point of view.

To address the weaknesses and mitigate the risks, we formulated a series of opportunities (i.e.: recommendations) to be seized, in particular:

- use the best example already available in the platform to harmonize the GUIs of the services: a sound candidate could be the NLP hub;
- rely on the EU initiatives in the field to lift the PARTHENOS community in the upcoming EOSC community, also as a possible environment to support its sustainability;
- capitalize on the existing links with the heritage science community to extend the scope of the infrastructure;
- leverage on the extensibility of the PARTHENOS platform to plug-in new tools and services extending the SAI importer capabilities;

- foster the deeper involvement of key international players in the field (i.e. ERICs and other relevant actors) to secure to PARTHENOS a relevant role as mediator and integrator.

This document has been developed not only to be a qualitative assessment of the integration work carried out within the project so far, but also to show the great opportunities created within the infrastructure. In fact, PARTHENOS is not only a step forward in terms of satisfying the many requirements expressed in the initial phase of the project but is offering a solid and structured setting for the effective development of new research, analysis and knowledge dissemination methods.

9.6 Abbreviations

Abbreviations		
Abbreviation	Extended name	Link
3M	3M Mapping Memory Manager	https://mapping-d-parthenos.d4science.org/3M
BBT	Back Bone Thesaurus	http://83.212.168.219/DariahCrete/sites/default/files/dariah_bbt_v_1.1_new_title.pdf
CIDOC CRM	CIDOC Conceptual Reference Model	http://www.cidoc-crm.org/
ESFRI	European Strategy Forum on Research Infrastructures	http://www.esfri.eu/
FIMS	Framework for Interoperable Media Services	https://github.com/isl/FIMS
ICT	Information and Communications Technology	
JSON	JavaScript Object Notation	https://www.json.org/
KPI	Key Performance Indicators	
N-TRIPLE	Triple format	https://www.w3.org/TR/n-triples/
NER	Named Entity Recognition	
NERLiX	Named Entity Recognition and Entity Linking and Extraction	https://services.d4science.org/group/nerlix
NLP	Natural-language processing	
PARTHENOS	Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies	http://www.parthenos-project.eu/
PEM	PARTHENOS Entity Model	
POS	Part of Speech	
RDF	Resource Description Framework	https://www.w3.org/RDF/
RI	Research Infrastructure	
RUBRICA	Reference Resources Integration plAtform	https://services.d4science.org/group/rubrica

SNA	Social Network Analysis	
SSH	Social Sciences & Humanities	https://ec.europa.eu/programmes/horizon2020/en/area/social-sciences-humanities
TTL	Turtle format	https://www.w3.org/TR/turtle/
UTF-8	Unicode Transformation Format, 8 bit	http://www.utf-8.com/
VLO	CLARIN Virtual Language Observatory (VLO)	https://vlo.clarin.eu/
VRE	Virtual Research Environments	https://services.d4science.org/explore
WP	Work Package	http://www.parthenos-project.eu/activities-and-wps/
XML	Extensible Markup Language	https://www.w3.org/XML/
GLAMs	Galleries, Libraries, Archives and Museums	
URL	Uniform Resource Locator	https://tools.ietf.org/html/rfc1738
URI	Uniform Resource Identifier	https://datatracker.ietf.org/doc/rfc3986/
API	Application Programming Interface	

9.7 Reference documents and services endpoints

Reference documents	
D2.4 shared folder on GDocs	https://goo.gl/k6LsT1
List of tools from D6.2 + addendum (use sheets 1 and 5)	https://goo.gl/EzR7xz
PARTHENOS integrated services document	https://goo.gl/FQorp6
Deliverable 2.3	https://goo.gl/QAazPZ
Deliverable 6.2	https://goo.gl/y1tSpN
D-NET aggregation services document	https://goo.gl/DtdgMV
Services endpoints	
PARTHENOS Catalogue (end-user)	https://goo.gl/PwAc7u
PARTHENOS Virtuoso endpoint (collection mgrs)	https://goo.gl/mCBuqv
Discovery App (end-user)	https://goo.gl/BMKuut
VRE links	
https://services.d4science.org/group/parthenos_lab	PARTHENOS Lab
https://services.d4science.org/group/NERLiX	NERLiX
https://services.d4science.org/group/rubrica	RUBRICA
https://services.d4science.org/group/ace	ACE Group
To see the tools available in the VRE: DataMiner -> Execute an Experiment. Documentation: https://wiki.gcube-system.org/gcube/DataMiner_Manager	

9.8 Bibliography

3MEditor https://github.com/isl/3MEditor

3M Mapping Memory Manager https://mapping-d-parthenos.d4science.org/3M

Apache Stanbol: https://stanbol.apache.org/

Bardi Alessia, Bruseker George, Ďurčo Matej, Kemps.Snijders Marc, Lorenzini Matteo, Theodoridou Maria .2017. *D6.2 Report on services and tools.* http://www.parthenos-project.eu/Download/Deliverables/D6.2_Report_on_services_and_tools.pdf

Bruseker George, Doerr Martin, Theodoridou Maria. 2017. D5.1 Report on the common semantic framework.

http://www.parthenos-project.eu/Download/Deliverables/D5.1_Common_Semantic_Framework_Appendices.pdf

CCL format http://nlp.pwr.wroc.pl/redmine/projects/corpus2/wiki/CCL_format

CLARIN-PL platform http://ws.clarin-pl.eu/

CLARIN VLO: https://vlo.clarin.eu/

CRF morpho-syntactic tagger (Wcrft2) http://ws.clarin-pl.eu/tager.shtml?en

CST Lemmatizer https://github.com/kuhumcst/cstlemma

D4Science Platform https://services.d4science.org/

D4Science VREs https://services.d4science.org/explore

DARIAH EU Webpage https://www.dariah.eu/

DARIAH Back Bone Thesaurus

http://83.212.168.219/DariahCrete/sites/default/files/dariah_bbt_v_1.1_new_title.pdf

DBpedia: http://wiki.dbpedia.org/

Degl'Innocenti Emiliano, Giacomi Roberta, Sanesi Maurizio *Bridging the Gap between Material and Immaterial Cultural Heritage in PARTHENOS VREs,* poster presented at the 2017 Digital Humanities Benelux Conference.

D-NET Lab. http://www.d-net.research-infrastructures.eu

Documentation of Apache Stanbol: https://stanbol.apache.org/docs/trunk/

Documentation of Apache Stanbol Enhancer:

https://stanbol.apache.org/docs/trunk/enhancementusage.html

Drude Sebastian, di Giorgio Sara, Ronzino Paola, Links Petra, van Nispen Annelies, Verbrugge Karolien, Degl'Innocenti Emiliano, Stiller Juliane, Oltersdorf Jenny, Spiecker Claus. 2016. *D2.1 Report on User Requirements.*

https://ec.europa.eu/research/participants/documents/downloadPublic?documentIds=080166e5a65b7331&appId=PPGMS

ESFRI http://www.esfri.eu/

FIMS https://github.com/isl/FIMS

GeoNames http://www.geonames.org/

Hilbert Martin. 2012. *How much information is there in the "information society"?*
Signifiance 9,4. https://doi.org/10.1111/j.1740-9713.2012.00584.x

Hollander Hella, Morselli Francesca, Admiraal Femmy, Conrad Anders, Trippel Thorsten, Zeldenrust Douwe, Ronzino Paola, Di Giorgio Sara, Madonna Antonio Davide, Hedges Mark. 2017. *D3.1 Guidelines for Common Policies Implementation.*
http://data.d4science.org/THhhK2NEVTBtM2p0WXE5NVNaZnRoTTBGanRaYmwyNHVHb
WJQNStIS0N6Yz0

Krauwer Steven, Verbrugge Karolien, Di Giorgio Sara, Spiecker Claus, Chatzimichail Christos, Degl'Innocenti Emiliano, Gautier Hélène, Giacomi Roberta, Henriksen Lina, Illmayer Klaus, Madonna Antonio Davide, Sanesi Maurizio. D2.3 Report on the assessment of data policies and standardization.
http://data.d4science.org/MmdnamhTVE91RlFTRG9EaXRacjRCdUVWQVVJVjFMSmtHb
WJQNStIS0N6Yz0

Liferay http://www.liferay.com/

Maltparser http://ws.clarin-pl.eu/parser.shtml?en; http://maltparser.org/

Mayer-Schönberger Viktor, Cukier Kenneth. 2013. *Big Data*. John Murray Publishers.

Natural Language Toolkit http://ws.clarin-pl.eu/tagerEnNLTK.shtml?en

Navigating research. How academic users understand, discover, and utilize reference resources, Oxford University press
https://global.oup.com/academic/content/pdf/navigatingresearch.pdf

OAI-PMH https://www.openarchives.org/pmh/

PARTHENOS Project official webpage http://www.parthenos-project.eu/

PARTHENOS Project's Work Packages organization. http://www.parthenos-project.eu/activities-and-wps/

PCRF-seq2seq lemmatizer https://github.com/UKPLab/coling2016-pcrf-seq2seq

Research Space: http://researchspace.org/

Romary Laurent, Banski Piotr, Bowers Jack, Degl'Innocenti Emiliano, Ďurčo Matej, Giacomi Roberta, Illmayer Klaus, Joffres Adeline, Khan Fahad, Khemakhem Mohamed, Larrousse Nicolas, Litke Antonis, Monachini Monica, van Nispen Annelies, Ogrodniczuk Maciej, Papadakis Nikolaos, Pastore Graziella, Pernes Stefan, Puren Marie, Riondet

Charles, Sanz Mikel, Sanesi Maurizio, Siozos Panayiotis, de Valk Reinier. 2017. *D4.1 Standardization Survival Kit.* https://goo.gl/VK62qM

Romary Laurent, Banski Piotr, Bowers Jack, Degl'Innocenti Emiliano, Ďurčo Matej, Giacomi Roberta, Illmayer Klaus, Joffres Adeline, Khan Fahad, Khemakhem Mohamed, Larrousse Nicolas, Litke Antonis, Monachini Monica, van Nispen Annelies, Ogrodniczuk Maciej, Papadakis Nikolaos, Pastore Graziella, Pernes Stefan, Puren Marie, Riondet Charles, Sanz Mikel, Sanesi Maurizio, Siozos Panayiotis, de Valk Reinier. 2017. *D4.2 Report on Standardization.* https://goo.gl/VK62qM

Spacy tool set http://ws.clarin-pl.eu/tagerDe.shtml?en

Universal Dependencies platform http://universaldependencies.org/format.html

Universal Dependency Relations set http://universaldependencies.org/u/dep/index.html

Wilkinson M.D., Dumontier M., Mons B. et al. 2016. *The FAIR Guiding Principles for scientific data management and stewardship* Scientific 3 Data doi:10.1038/sdata.2016.18