



Digital Research Infrastructure for the Arts and Humanities

Enhancing and supporting digitally-enabled research
in the arts and humanities
across Europe

Information Brochure
Funders Round Table
22-23 June 2010, Paris

A project on the ESFRI Roadmap
European Strategy Forum in Research Infrastructures
<http://dariah.eu>



Table of Contents

1. DARIAH: A New Digital Research Paradigm.....	3
1.1. Introduction	3
1.2. Strategy	4
2. DARIAH's Services for Science.....	6
2.1. Virtual Competency Centres (VCC)	6
2.1.1. VCC1 - e-Infrastructure	9
2.1.2. VCC2 - Research and Education Liaison	11
2.1.3. VCC3 - Scholarly Content Management	13
2.1.4. VCC4 - Advocacy, Impact and Outreach	14
2.1.5. The DARIAH Coordination Office (DCO)	15
2.2. DARIAH Service Packages	16
2.3. DARIAH Technical Architecture	19
2.4. Scientific Cases.....	23
3. Implementation of DARIAH	25
3.1. Criteria for participation in a VCC.....	25
3.2. Legal Structure ERIC.....	26
3.3. Funding Model.....	31
Annex.....	37

1. DARIAH: A New Digital Research Paradigm

1.1. Introduction

The grand vision for the Digital Research Infrastructure for the Arts and Humanities (DARIAH) is to facilitate long-term access to, and use of, all European arts and humanities (A+H) digital research data. DARIAH will support research practitioners at all levels, from beginners through to those employing advanced techniques and methodologies.

Everybody interested in using digital means for A+H research is part of the DARIAH community of practice. The DARIAH infrastructure will be a connected network of people, information, tools, and methodologies for investigating, exploring and supporting work across the broad spectrum of the digital humanities.

Researchers will use DARIAH to:

- find and use a wide range of research data from across Europe
- exchange knowledge, expertise, methodologies, and practices across domains and disciplines
- ensure that they work to accepted standards and follow best practice
- experiment and innovate in collaboration with other scholars

The core strategy of DARIAH is to bring together national, regional, and local endeavours to form a cooperative infrastructure where complementarities and new challenges are clearly identified and acted upon.

The DARIAH network will be designed to be as decentralised as possible, empowering individual contributors (e.g. individual researchers; national centres; specialised/thematic centres) to work with and within the DARIAH community and shape its features to their needs. Each contribution of each contributor builds DARIAH, and all is linked together in DARIAH's architecture of participation. At the same time, however, collaboration across the borders of individual centres requires the usage of common technologies, e.g. for authentication or federation of archive contents. These considerations also have a place at the core of the DARIAH preparation plan.

The initial idea for DARIAH has grown into a clear vision shared by many. Since the commencement of the Preparing DARIAH phase in September 2008, DARIAH has been actively exploring its research-oriented objectives through a variety of means, including engagement with stakeholders, technical systems development, and explorations of models for governance and financing. The overall objective of the Preparing DARIAH phase is to be ready for the construction of DARIAH by 2011.

Preparing DARIAH specifically aims to:

- identify the components and services for a European research infrastructure in the A+H which will serve to enhance scholarly research methods
- facilitate the publication and reuse of research data on an international level
- establish a consortium committed to the objectives of DARIAH

- deliver an overall business plan, a legal document laying down the rights and obligations of different types of DARIAH partners, and secure sustainable financial support for the construction and initial operational phases of DARIAH

The work in the preparatory phase addresses coordination, strategic, financial, governance, logistical, legal, and technical issues, as well as management and dissemination activities to support this work.

DARIAH will be much more than the sum of each single national or individual contribution; it represents the next generation of research potential in the A+H.

1.2. Strategy

Mission Statement

The mission of DARIAH is to enhance and support digitally-enabled research across the humanities and arts. DARIAH aims to develop and maintain an infrastructure in support of ICT-based research practices and is working with communities of practice to:

- Explore and apply ICT-based methods and tools to enable new research questions to be asked and old questions to be posed in new ways
- Improve research opportunities and outcomes through linking distributed digital source materials of many kinds
- Exchange knowledge, expertise, methodologies and practices across domains and disciplines

Defining DARIAH: a Distinctive Role

What DARIAH is...

The DARIAH infrastructure is a connected network of people, expertise, information, knowledge, content, methods, tools and technologies for investigating, exploring and supporting work across the broad spectrum of the digital humanities. DARIAH focuses on research and knowledge creation practices with a particular emphasis on supporting scholarship in a rapidly changing digital environment. It is this focus on research data creation and use in a research context as opposed to a preservation context that marks out DARIAH as unique. DARIAH emphasises methods and epistemic practices rather than single disciplinary practices. Instead of solely providing access to resources, DARIAH focuses on creating virtual research environments that support the use of resources in a scholarly context.

What DARIAH is not...

DARIAH is not primarily a network of digital data centres, as is CESSDA, although it seeks to collaborate and partner with such centres and will act as a broker between researchers seeking preservation of their research data via trusted digital repositories. DARIAH is also not as singly discipline-focused as CLARIN, for example; instead DARIAH seeks to support all disciplines across the humanities, encouraging interdisciplinarity and the exploration and sharing of content, tools and methods. Furthermore, DARIAH is not a service seeking to provide integrated access to all cultural heritage materials, as is Europeana. DARIAH instead seeks to support the scholarly use of research data and the application of ICT-enabled tools and methods to support knowledge creation.

The Scientific Case: Adding Value, Ensuring Impact

The development and construction of DARIAH is driven by three key factors:

- The changing nature of research practice, knowledge creation, and information sharing across the humanities and arts
- The increasing volume of digitised source materials being made available by archives, libraries, museums and scholars, the highly distributed and dispersed nature of much of this digitised information, and the growing range of technologies and applications that offer opportunities for creating and sharing this information
- The emergence of ICT-enabled tools and methods that help scholars to analyse and interpret research data, to ask new questions of these source materials and to ask old questions in new ways

Research practice in the arts and humanities is about criticism and meaning, interpretation and re-interpretation, and about extracting meaning from often incomplete and fuzzy data. It requires researchers to seek out a wide range of primary and secondary sources, to organise and structure these, to analyse and interpret them, and to publish the results. In this era of pervasive broadband connectivity, the way in which these processes are undertaken is changing, and in some cases, the processes themselves are changing. Increasingly, research practitioners are using the power of the internet, new tools, and the range of digital information that is available to them to create their own personal network spaces, to digitally publish highly interactive, multimedia-themed collections (critical editions) of research information and knowledge, and to visualise and reconceptualise their interpretations and analysis. New forms of collaboration are also emerging as the tools available encourage and enable ‘web-working’ across the globe.

The nature of the source material on which research is based ranges from simple catalogues and bibliographies, through to detailed source materials in the form of printed text, video, audio, images, cultural 3D objects, maps, and so on. Digital surrogates of these materials are distributed across a wide range of different organizations including libraries, archives, museums, galleries, and cultural heritage organisations. However, although these digitised materials offer massive research potential, many remain entrenched in their ‘local’ context: they are described differently using varied metadata standards, may have used different technical standards in their creation, and are presented and published through a large number of disparate web sites, some simple, some highly complex.

We are also witnessing rapid changes in technology and in the nature of the digital environment in which we conduct research. The emergence of new methods and associated tools such as text analysis, text mining, geo-temporal visualisations, multi-spectral imaging, 3D object modelling, and many others, are offering new insights that were not possible before.

The goal for DARIAH is to establish a coherent and coordinated research infrastructure that addresses these three drivers, that can take advantage of all these new technologies have to offer, take account of the new ways of working and encourage new ones to develop, that can facilitate the creation and use of digital information, and that can ultimately foster new knowledge about our histories, our cultures, and our role in the society of today.

DARIAH Priorities: Enhancing Scholarship

The key strategic aim of DARIAH is to support researchers in the creation and use of research data and tools, and to apply and use ICT-enabled methods to analyse and interpret digital source materials.

The strategic objectives of DARIAH are to:

1. Support researchers to create, curate and preserve scholarly digital collections so that they may enhance their own scholarship and share these collections for others to reuse
2. Support researchers to use, analyse and interpret the increasing volume of research data made available by archives, libraries, and cultural heritage organisations across Europe and elsewhere so that they can take advantage of the significant investment in digitisation which continues to take place in these organisations
3. Support researchers to understand and use ICT-enabled methods and tools so that they can explore new questions and interpretations and add to the body of knowledge on European heritage, history and culture
4. Collaborate with archives, libraries and other cultural heritage organisations to provide virtual research environments that enable enhanced use of their collections for research, such as that in the EHRI (European Holocaust Research Infrastructure) project
5. Understand and promote the impact and value of digitally-enabled research and research infrastructures

2. DARIAH's Services for Science

2.1. Virtual Competency Centres (VCC)

As a result of detailed analyses of research practices and methods, it is planned to organise DARIAH as four Virtual Competency Centres (VCCs) and a Coordination Office. Each of the four VCCs is focused on one particular area of expertise: (1) e-Infrastructure, (2) Research and Education Liaison, (3) Scholarly Content, and (4) Advocacy. While centred on a specific area of expertise, the VCCs are at the same time cross-disciplinary, multi-institutional, and multi-national.

In practice, stakeholders such as individual researchers in the A+H, research networks, and A+H research centres will consult various VCCs for services. For example, a research network may consult with VCC Research when designing their research approaches, as well as with VCC Scholarly Content when defining their data and ensuring its sustainability. In its interaction with users, DARIAH aims to hide any internal complexities related to the VCC organisational structure and provide potential stakeholders a single point of access to the DARIAH services.



DARIAH Digital Research Infrastructure for the Arts and Humanities

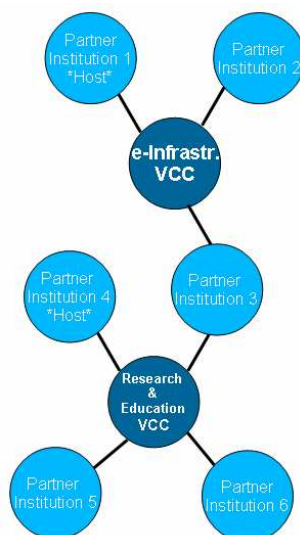


Figure 1: The organisational structure of DARIAH. Various DARIAH partners contribute to a single VCC, and each DARIAH partner may participate in multiple VCCs

This chapter presents how DARIAH will organise the different content-related issues in order to establish a research infrastructure for the A+H. In particular, it describes the tasks and services of the VCCs and how they interact with each other and with external stakeholders.

- **VCC1 - e-Infrastructure**
... to establish a shared technology platform for A+H research
primary target group: other VCCs, innovators and adopters of technical infrastructure
- **VCC2 - Research and Education Liaison**
... to expose and share researchers' knowledge, methodologies and expertise
primary target group: individual A+H researchers and research networks
- **VCC3 - Scholarly Content Management**
... to expose and share scholarly content
primary target group: A+H data centres
- **VCC4 - Advocacy, Impact and Outreach**
... to interface to key influencers in/for A+H
primary target group: funders, policy makers, industry, and others
- **DCO - DARIAH Coordination Office**
... to assume the overall responsibility and to ensure adequate operations across all DARIAH organisational units and partners

Each VCC may have numerous contributors and each DARIAH partner institution may contribute to multiple VCCs - hence *Virtual* Competency Centre. To coordinate the efforts within and between VCCs, for each VCC a **host** is selected according to the criteria defined below.

All VCCs interact with a multiplicity of stakeholders and user groups. For example, VCC3 interacts with all types of data contributors and related actors, including A+H research initiatives, data infrastructure initiatives, as well as A+H data centres. However, when it comes to the provision of a single point of contact for DARIAH services, each VCC has a **primary target group** (as noted in the bullet list above) for which it mediates between other VCCs, local institutions and other partners (cf. related report "DARIAH Service Catalogue - User Perspective").

The services provided by the VCCs can be separated in two types: **support services** and **infrastructure concepts and components**. In a nutshell, "support services" are ongoing tasks conducted by the VCCs, whereas "infrastructure concepts and components" are the products of those tasks. Users benefit from both: in a usual setting, users will address DARIAH in a specific situation to satisfy specific needs. DARIAH in turn will provide a package of support services and infrastructure from various VCCs and conducted by various partners, tailored to the specific situation of the user.

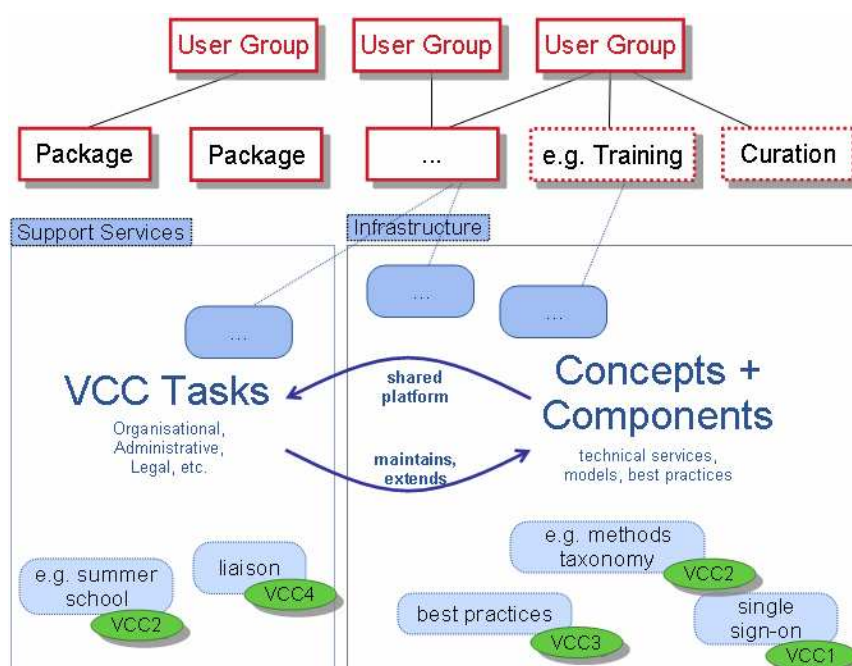


Figure 2: Matrix of DARIAH packages

Support services combine a range of organisational, administrative, legal and other support tasks for building and maintaining the infrastructure. During the DARIAH Construction Phase, they are responsible for creating specific infrastructure concepts and components, and they subsequently maintain and extend them during the Operational Phase. Support services need to be put into place and grow during the Construction Phase. Once they are stable, however, they may become chargeable services to the community and form part of a sustainable business model.

Infrastructure concepts and components, on the other hand, are physical or intellectual products established by the VCCs during the construction phase. A research infrastructure in the A+H is not only about technology; it is about people, information, knowledge, epistemic practices, communities of practice, boundary objects, content and data, and so on. A meaningful infrastructure will cater to and incorporate all of these elements. DARIAH comprises a system of infrastructure concepts and components, which together support the A+H research lifecycle for DARIAH users. These infrastructure components include best practices, standards and reference lists, in addition to technical services.

2.1.1. VCC1 - e-Infrastructure

... to establish a shared technology platform for A+H research

VCC1 Goals and Scope

The VCC1 e-Infrastructure establishes the technological basis for DARIAH to be a trusted intermediary, in which community-developed data and tools can be shared and integrated. VCC1 will:

- establish infrastructure services and standards which ensure interoperability across the whole DARIAH ecosystem
- support the creation and evolution of local data stores for the trustworthy management of research data (e.g. national data archives, special collections)
- foster the stability, openness and re-usability of scholarly tools and collaborative research environments (e.g. humanities centres, researchers in A+H)
- interact with institutional partners, A+H development groups, and related actors to ensure the quality, permanence and growth of e-infrastructure and technical services in the A+H

The VCC1 establishes a range of technical infrastructure components that are described in more detail in the following sections as well as in the chapter on DARIAH technical architecture (see chapter 2.3). Core infrastructure services are aimed to be interoperable with and to be created in collaboration with associated infrastructure initiatives (e.g. CLARIN and other ESFRI initiatives).

A+H Infrastructure Services

Rather than a monolithic, controlled infrastructure, DARIAH is intended to be a means of linking up people, services and data for research in the A+H. The infrastructure services to be developed and hosted by this task establish a framework and interoperability across the diverse, decentralised tools and services in the DARIAH ecosystem, as well as with other infrastructures, including ESFRI initiatives. Some of the core services required for this ecosystem to emerge include an overarching authentication and single-sign-on infrastructure, PID services, infrastructure management and others (described in chapter 2.3).

Reference Software Packages

Emerging A+H centres may employ the software packages to be created in this task to establish their technology platform, using a complete software stack or a smaller set of tools and services to make their services DARIAH-compliant. The DARIAH reference software will be built upon existing technologies and



addresses both the needs of research environments as well as the building of a trusted repository for preserving research data. Together, these two software stacks cater for the whole research lifecycle in active usage, preservation, and reuse of research data.

Preservation Infrastructure

This service supports and promotes A+H data centres in their efforts to ensure trustworthy preservation. DARIAH support is both technical and organisational, whereby VCC1 provides guidance regarding technical preservation infrastructure and VCC3 regarding preservation policies and standards. One of the key activities of this support service is to pool preservation experiences across existing A+H trusted repositories and to link them to the findings of the e-infrastructure and preservation communities (e.g. preservation registries, technology watch).

Data Federation and Interoperability

The DARIAH data backbone is designed to be open and decentralised. It differs from existing repository federations such as Europeana and DRIVER in various ways:

- The primary use case for Europeana and DRIVER is a federated search. DARIAH offers the federation mechanisms for other, research-oriented applications across diverse data sources, such as analysis, visualisation, and task management. Anybody can build respective applications on federated sources in the DARIAH network.
- DARIAH is decentralised. Any source can be added or excluded in specific application environments, and different subsets of data sources can co-exist.
- DARIAH is very light-weight. Its openness and flexibility offers the range of opportunities described above, but it also means that writing new applications that operate on the federated DARIAH backbone may require more effort than simply querying an existing database.

This activity is responsible for the infrastructure services needed to establish such a data backbone, and aims to provide a set of best practices and samples for building cross-repository interoperability channels (e.g. OAI-PMH, OAI-ORE, Linked Data).

Developer Community

DARIAH's architecture of participation encourages the A+H community at large to contribute their own data and tools. Besides technical standards and gateways, this activity aims to provide development and communication facilities (e.g. code repository, issue tracker) to trigger a collaborative community.

A+H Service Environment

The digital humanities are a large and heterogeneous community with a rich set of tools and methodologies. The previous activity (Developer Community) aims to harness the activities of the A+H community related to software development, acknowledging that many of those tools being created may be discontinued or superseded in short time periods. The current activity aims to identify and extend the lifespan of those services of more permanent value, and to make them available to the whole A+H community.

Some of the infrastructure components include semantic reference tools, such as dictionaries, thesauri, or geo-referencing services. VCC1 will create and host these reference tools; their content data, however, will be contributed by VCC2 and/or the A+H community.

In addition to hosting key reference services for the A+H community, this support service establishes best practice guidelines for ensuring the openness and interoperability of components in the A+H service environment.

A+H Research Demonstrators

This service creates user applications, tailored to the needs of a particular research context. These may be prototypes or even productive DARIAH-based services for demonstration purposes, and are built on behalf of and following the specifications of VCC2. Initially, two demonstrators will be built: a generic search service across all DARIAH resources with specialised functionalities on subsets of the data (e.g. faceted browsing, geo-visualisation), and a frontend tool to manage collections across heterogeneous sources.

2.1.2. VCC2 - Research and Education Liaison

... to expose and share researcher's knowledge, methodologies and expertise

VCC2 Goals and Scope

The VCC Research and Education Liaison aims to promote and support the use of research data and ICT methods and technologies, including the DARIAH infrastructure. It will act as the primary contact with the A+H research and teaching communities, providing the interface between the DARIAH research infrastructure and researchers undertaking basic, applied and practice-led research across the A+H. It will seek to understand A+H research practices and processes, and to understand and promote the use and application of ICT-enabled methods and tools, with a particular emphasis on interdisciplinary understanding and exchange.

VCC2 will encourage, support and enable research actors to use DARIAH tools and services in their research to assist them in asking new questions and to address old questions in new ways through the use of research data and the application of ICT methods. In so doing, it aims to address a range of people and interests, from established researchers to post-graduates to students, as well as different disciplines and domains within and outside of higher education.

In addition to the essential support services for designing and promoting the DARIAH infrastructure, and among other key concepts and components, this VCC contributes a knowledge base which captures and links A+H methods, tools, and projects, and references digital humanities curricula.

Understanding Research Practices

The A+H community needs to understand the changing nature of research practices in the digital environment. This VCC aims to capture the ongoing changes in research practices, and to support researchers to share expertise and knowledge in digital methods and practices. Specific actions under this task are:

- Identify research groups using ICT methods, tools and services and seek to understand their research, teaching, and learning practices and processes
- Support researchers to share and exchange their expertise (e.g. methods, tools), building on an ontology of methods and a registry of tools as a reference for users of ICT methods
- Develop a projects registry and case studies illustrating the use of ICT content and methods; information on expertise centres and groups
- Translate these practices and processes into specifications for upgrading and improving DARIAH services and systems

Training and Education Programme

This VCC will provide a training programme for researchers in the methods, tools, and approaches needed to engage with the digital environment, including DARIAH services, tools, and content. Services will include:

- DARIAH Summer School
- Online training and demonstrators to enable researchers to experiment with real life data and tools
- Collaborate with institutions providing undergraduate and postgraduate training in the digital humanities to embed DARIAH tools and services in their courses
- Registry of undergraduate and post-graduate courses

Community Engagement

Engagement with digital tools, methods and content is an emerging practice for a majority of A+H research practitioners. The best method of promoting engagement and increasing uptake of the large volume of research data being digitised across Europe and elsewhere, and in turn the research infrastructures designed to support the creation and use of this research data, is to support communities of researchers in coming together to learn from each other and express their needs and requirements. Services to this end will include:

- Advocacy and outreach to the A+H research and education communities to promote the digital humanities and DARIAH
- Engagement with professional bodies, learned societies, and national funding bodies
- Expert seminars
- A workshop series
- A publication and working paper series

To realise these services, DARIAH will develop a specific program termed ‘community engagement projects’. This program will be targeted at researchers who want to engage in developing new research methods or tools which would directly interact with DARIAH endeavours or contribute to the overall digital humanities landscape in Europe. In collaboration with European Science Foundation (ESF), which will take up the actual operational role in implementing the review process and project supervision, this program will also be an opportunity to correlate DARIAH activities with national funding agency priorities.

Virtual Research Environment

Since nearly every research project requires an amalgamation of specific types of scholarly data in combination with generic or specific tools for data querying, enrichment, and exploration, it is of utmost importance to provide facilities for the definition and implementation of a virtual research environment (VRE). Such environments are constructed upon the technical means provided by the e-infrastructure and based on a closed collaboration between scientists. Indeed, community engagement projects are of seminal importance in identifying possible reference VREs which can be further offered to a variety of research communities.

2.1.3. VCC3 - Scholarly Content Management

... to expose and share scholarly content

VCC3 Goals and Scope

The scholarly content management VCC will deal with the various stages of the scholarly content life cycle, from creation, curation, and dissemination, through to the pooling of scholarly digital resources and results for reuse. The VCC will offer services and resources for the representation and management of data, as well as for the management of associated legal and organizational issues. Thereby it aims to enhance data quality, preservation, and deep interoperability, as well as furthering a culture of data sharing in the A+H.

Among the key infrastructure concepts contributed by VCC3 are relevant standards, reference licenses, and best practice guidelines. Its products and support services address a diverse target community including A+H data centres and research networks, as well as individual researchers.

Curation

This activity supports A+H data centres by providing reference policies and guidelines, as well as by creating a virtual digital curation helpdesk for data management. Furthermore, it establishes a framework for evaluating A+H data centres as to their capacity to preserve A+H scholarly content over time, following such pre-existing guides as those established for trusted digital repositories¹.

Best Practices and Open Access

This activity will produce guidance and reference material in order to allow research communities to disseminate their research productions (publications and data) to a wide academic and societal audience:

- Provision of reference licenses (such as those offered by the creative commons framework) to optimize the trade-off between wider access and rights protection (in particular proper scholarly attribution)

¹ TRAC - Trustworthy Repositories Audit & Certification: Criteria and Checklist (<http://www.crl.edu/archiving-preservation/digital-archives/metrics-assessing-and-certifying>) or DRAMBORA - Digital Repository Audit Method Based On Risk Assessment (<http://www.dcc.ac.uk/resources/tools-and-applications/drambora>)



- When applicable, and in full allegiance to subsidiary principles, DARIAH will take up cross-national license negotiations to facilitate access to large collections or to optimize relations with complementary endeavours in the library domains
- Help users choose an optimal dissemination channel in collaboration with trusted digital repositories (Seal of Approval/Audit/Certification) that suit their requirements
- Work with communities to identify optimal policies according to the actual research fields and corresponding existing practices

Standardization

This activity aims to actively engage in the standards-building process, from consensus-building to eventual maintenance of the standard. It also aims to maintain reference information about relevant existing standardization frameworks and activities.

Reference Data Registries

This activity will facilitate the wide identification and dissemination of digital assets. It defines channels for data reuse and exchange across communities and research infrastructures², provides reference data registries for the description of scholarly data (e.g. authority lists, registries, reference ontologies), and collaborates with VCC1 e-Infrastructure for deploying relevant tools and registries.

Repository Support

This activity aims to support A+H data centres in building trusted digital repositories. In collaboration with VCC1 e-Infrastructure it aims to deploy easy-to-use repository environments that facilitate the inclusion of new partners' data holdings into the DARIAH infrastructure.

Enrich Digital Scholar Contents

In the digital humanities, data and document producers are increasingly exposing their data to be enriched and reused. As a result, locating and accessing data within this growing body of multilingual, multicultural, and multidisciplinary information is a challenge. Metadata quality is therefore an essential feature for ensuring deep and widespread scholar engagement. The use of the semantic RDF format will allow for the interconnection of heterogeneous data and the integration of new and valuable services for A+H.

2.1.4. VCC4 - Advocacy, Impact and Outreach

... to interface to key influencers in/for A+H.

VCC4 Goals and Scope

This VCC will focus on (1) high level advocacy, (2) assessing the impact of DARIAH and measuring the 'added value' that it brings, (3) outreach to wide groups of stakeholders, and (4-5) ensuring capacity and participation in DARIAH. 'Services' are not relevant to this VCC but tasks and activities are.

² Some of the mentioned activities could obviously benefit from joint activities between DARIAH and other initiatives in the humanities and social sciences domain (CLARIN, CESSDA, etc.) or beyond, with research infrastructures having to deal with similar data related challenges (GAVO, Elixir)

High-Level Advocacy

The targeted audience of this activity comprises key influencers in all roles and disciplines/industries who are in a position to promote, collaborate with, and finance DARIAH. DARIAH will engage with those audiences in managed formal and informal meetings, polished information packages and PR campaigns, as well as in the coordination of DARIAH's response and participation in, for example, funding calls.

Impact & Value

This activity assesses the impact of DARIAH and measures the 'added value' that it brings through quantitative and qualitative measures. It will focus on how DARIAH facilitates the transfer of knowledge and expertise between (a) the relevant disciplines within the humanities (themselves often silos) and (b) the humanities and other areas such as qualitative social sciences, computer science, information science, and bioinformatics.

Outreach

DARIAH seeks mutually beneficial relationships with a wide community of stakeholders which potentially feed into A+H research, including cultural tourism, industrial partners and publishers.

Ensuring Capacity in DARIAH

This activity ensures the consistency and growth of the DARIAH network of partners. To achieve this, it identifies

- contact points for potential national partners in Europe
- potential institutional partners
- potential industrial partners
- gaps in skills and resources within the DARIAH network

This activity also ensures the adequate incorporation of the skills of new DARIAH partners into relevant VCCs.

Ensuring Participation in DARIAH

In collaboration with VCC2 Research and Education, this activity aims to ensure the uptake of infrastructure-based research in A+H and to foster diversity in the DARIAH community. A+H is a very rich and heterogeneous area, and hence not all branches can be addressed at the same time. The DARIAH community will instead grow successively. This growth will primarily be driven by the community, but this activity will also work to ensure the openness of this growth along several dimensions, including disciplines, geography and skills (e.g. dealing with content types or methodologies).

2.1.5. The DARIAH Coordination Office (DCO)

...to assume the overall responsibility and to ensure adequate operations across all DARIAH organisational units and partners

5.1 DCO Goals and Scope

The initial legal seat of the DCO will be in Paris, France (CNRS), with a leadership arrangement between France and Germany. The DCO supports and integrates all levels of DARIAH, including the representation of all DARIAH partners (General Assembly, Scientific Board, Board of Directors, VCC Heads etc.). In its role as a coordinator, the DCO oversees the interactions with all DARIAH partners and boards and takes on a variety of vertical tasks (e.g. controlling, revision) and horizontal tasks (e.g. central services, overall financing, legal and tax requirements, transfer of skills and knowledge).

Overall Coordination

- Management of the entire organisation
- Communication with the different groups within the organisation
- Interface and risk management
- Reporting and controlling

Organisational Coordination

- Budgeting and taxation
- Legislation
- Setting up an internal regular update system on financial matters
- Personnel administration and payroll accounting

Communications

- Website
- Internal communications: wiki, mailinglists
- External communications: newsletter and promotional material

2.2. DARIAH Service Packages

DARIAH service packages combine the tasks and concepts from distinct VCCs into a single package, thereby providing users with a single point of access. This chapter describes some of the packages and how they are tailored to specific target groups and activities, such as packages for "collective intelligence" and "data curation".

The DARIAH infrastructure offers a range of services to its users. DARIAH services are geared to (a) accelerate their current work, (b) foster collaboration between them, and (c) support the sustainability of any activities in the A+H ecosystem.

The packages described in this chapter are single points of access to DARIAH services. They highlight only some of the key services offered by DARIAH, and they may grow or evolve over time. As such, this description is an illustration of how the DARIAH infrastructure benefits specific users, rather than a comprehensive specification of its tools and services.

Each package links to distinct organisational entities in DARIAH (both VCCs and institutional partners). Packages are as generic as possible, but can be adapted to specific user groups with regard to a specific research domain, or the role of the user. They aim to function as single points of access for users to hide any organisational or technical intricacies from the user.

Collective Intelligence

DARIAH supports the collaboration of researchers to e.g. re-use data, co-develop tools, share experiences, and discuss methodologies. Collective intelligence emerges from the Knowledge Base that collates information about projects, tools, and methods from disparate user groups, as well as the gateways that connect the heterogeneous resources into a single environment.

Overall responsibility: VCC2

Target user group: researcher; individual researchers and research projects from diverse domains

Tasks and responsibilities:

- VCC2 - maps practices in the humanities, and populates the knowledge base with information about research projects and methodologies
- VCC1 - establishes the technical infrastructure for a A+H Service Environment as well as data interoperability gateways, and encourages the co-development of research tools in a developer community
- VCC3 - creates best practices and interoperability standards for sharing and enriching scholarly content
- VCC4 - evaluates the impact of DARIAH's activities towards collective intelligence and takes care of the enhancement of the knowledge base by integrating new communities
- A+H research centres, and research projects - share experiences, discuss methodologies, expose data, and contribute tools

Accelerator

As new research activities commence their work, they often need to invest considerable efforts in establishing the concepts and technical environment suited to their specific goals and context. DARIAH accelerates this initialisation period, offering standards and best practices as guidance for research activities, technical infrastructure and tool recommendations upon which bespoke research environments can be built, and offering training and advice to ensure speedy uptake.

Overall responsibility: VCC2

Target user group: researcher; research projects and networks from various domains

Tasks and responsibilities:

- VCC2 - mediates between researchers and the technical infrastructure; fosters the linkage of research activities within and across domains

- VCC1 - establishes technical infrastructure hosted by DARIAH, and also collates software packages and brokers tools from other research initiatives to be deployed by research projects and institutions
- VCC3 - creates best practices and interoperability standards for sharing and enriching research data
- VCC4 - evaluates the impact of DARIAH's activities towards collective intelligence

Data Curation

Foster trusted data curation for owners of digital assets, and promote interoperability and preservation from early stages in the research lifecycle. While DARIAH does not conduct preservation activities or even house a long-term repository, it encourages institutional partners to establish trusted services for preservation. DARIAH supports those institutional preservation services, evaluates them, and mediates between users and the network of trusted repositories.

Overall responsibility: VCC3

Target user group: researcher (data provider); both, individual researchers as well as A+H projects

Tasks and responsibilities:

- A+H Data Centre - transparently document curation service policies (selection criteria, retention criteria, risk level, costs)
- A+H Data Centre - offer organisational and technical means for the ingest of digital assets
- VCC3 - ensure conceptual interoperability across the curation products at DARIAH Centres
- VCC3 - Policy comparison between curation products at A+H Data Centres; and mediate between data owners and the Centres
- VCC3 - Audit Centres for compliance with DARIAH standards
- VCC2,3 - Promote standards and best practices to foster data curation (including preservation actions) already at early stages of the research lifecycle
- VCC1 - Support the creation of technical preservation infrastructure at institutions, and fosters technical interoperability and convergence between them
- VCC1 - Ensure interaction with the international preservation community

Building Capacities

Support new A+H Research Centres in establishing their technical infrastructure as efficiently as possible and support existing A+H Research Centres in adoption of DARIAH infrastructure and standards.

Overall responsibility: VCC3

Target user group: A+H Research Centres; both, existing and emerging Centres

Tasks and responsibilities:

- VCC1,3 - offer technical standards (VCC1) and conceptual standards (VCC3) fostering quality and interoperability across DARIAH centres
- VCC1 - establishes customisable software packages for data management (repositories) and for building research environments

- VCC3 - offers advice on building trusted services for preservation, including any preservation services or repository packages (as offered by VCC1)

Technology Uptake and Evolution

Harmonise technology stacks used in A+H, and integrate novel technologies into the A+H.

Technical infrastructure in the digital humanities has multiple conceivable origins and numerous influences on their deployment and application. DARIAH aims to transfer existing infrastructure technologies into the A+H where useful, liaising with technical research and infrastructure initiatives from e.g. the grid, preservation, and semantic technologies communities. At the same time, DARIAH aims to capture valuable A+H services, extract them from context, and make them available in a generic way to other A+H research activities.

Overall responsibility: VCC1

Target user group: other VCCs, including via the other VCCs also A+H research centres and research activities

Tasks and responsibilities:

- VCC1 - design and extend the DARIAH technical infrastructure
- VCC1 - liaise with related research and infrastructure initiatives, e.g. preservation projects such as PLANETS or the Digital Curation Centre in Germany; European e-Science infrastructure such as EGI, and others
- VCC2,3 - identify gaps and requirements for A+H systems; respectively identify opportunities for adopting specialised services and offering them in a more generic way to other domains

Advocacy and Liaison

DARIAH is embedded in a large ecosystem of related initiatives, due to their goals, technologies, organisational contexts, or due to other reasons. Existing resources and experiences need to be tapped to enhance the quality and the sustainability of DARIAH. Interaction with related initiatives is therefore occurring on all levels.

Overall responsibility: VCC4

Target user group: all VCCs

Tasks and responsibilities:

- VCC1 - liaise with related infrastructure projects
- VCC2 - liaise with A+H research projects, interest groups, etc.
- VCC3 - liaise with A+H data centres and bodies for data interoperability and standardisation
- VCC4 - high-level advocacy at funders, policy makers, industry, et al.

2.3. DARIAH Technical Architecture

The DARIAH research infrastructure is an open, collaborative environment that enables research in the A+H by linking data, functionalities, and people. Its "architecture of participation" accommodates A+H data

centres, research networks and researchers that are widely independent, stem from multiple backgrounds, interact with DARIAH following diverse goals, and employ various entry-points into DARIAH. Linking this diversity, DARIAH aims for a very light-weight and decentralised infrastructure that can be fit to each stakeholder's situation. Rather than a single technical solution, DARIAH may be many, according to community activities and willingness to collaborate.

The DARIAH technical architecture is built of three horizontal tiers, as well as vertical interoperability frameworks for both data and services. In each of those aspects and for every component, DARIAH seeks a broad interest base and collaborations. In particular, core infrastructure services may be created in close interaction with affiliated initiatives (including CLARIN and other ESFRI initiatives).

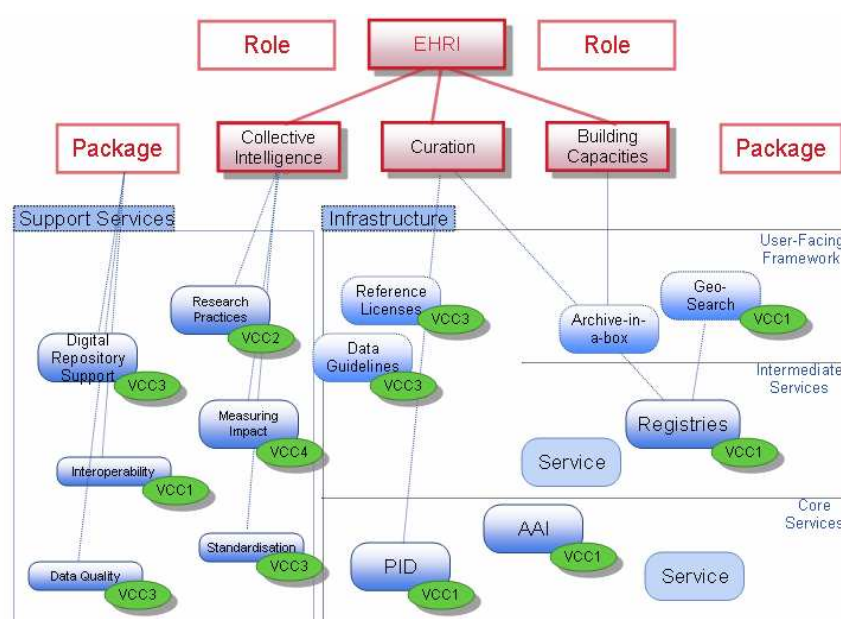


Figure 3: Composing various services and infrastructure components to establish a tailored research environment, in this example for EHRI, the European Holocaust Research Infrastructure

Three-Tier Architecture Model

The DARIAH technical infrastructure is built as a loosely-coupled service-oriented architecture with three structural tiers in its architecture model: (a) the user-facing framework, (b) infrastructure service environment, and (c) core infrastructure. It also describes how services can move up and down these horizontal tiers, to enable an architecture of participation that is open to contributions and evolves over time.

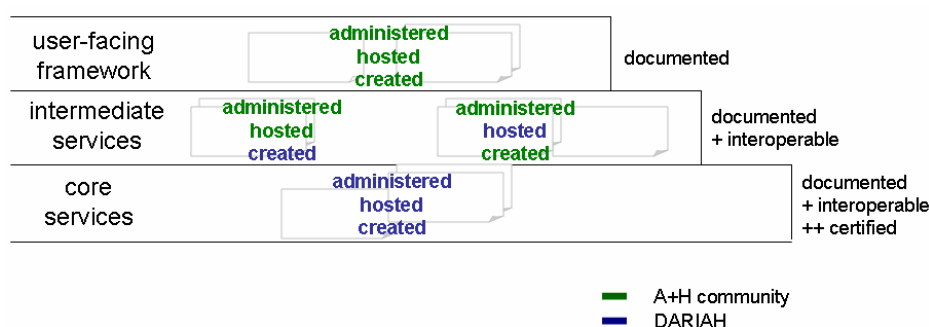


Figure 4: Each tier may open up different organisational contexts for managing service components. Infrastructure services are created, hosted and administered by DARIAH ensuring reliability and scalability, whereas the A+H community is encouraged to contribute their own, potentially transient services towards the higher-level tiers

User-Facing Framework (UFF)

The UFF accommodates a collection of end-user tools contributed by research projects or third parties. At a minimum, components in the DARIAH UFF tier need to be well-documented to facilitate reuse. Beyond mere documentation, tools and services ideally comply with the DARIAH service framework to foster interoperability with other DARIAH components. Other than that, there is no central control of development efforts in the UFF tier; collaborations are encouraged within the open DARIAH developer community.

Infrastructure Service Environment

Reference services fill the infrastructure service environment with life by offering actual research-relevant content for reference and reuse. For example, authority data on authors and other persons, thesauri, dictionaries from various epochs, and other reference data are often essential for research initiatives, yet are outside their scope. Shared reference services that offer data for reuse and perhaps mechanisms to contribute new data are hence infrastructure components for ensuring quality and efficiency in A+H research, as well as focus points for collaboration.

Core Infrastructure

The core layer includes services that serve to sustain the DARIAH infrastructure and establish coherent operation across the open DARIAH environment. Services such as a Persistent Identifier (PID) resolver and Authentication and Authorisation Infrastructure (AAI) are essential for enabling interoperability across the heterogeneous data sources and decentralised services in the DARIAH ecosystem. Other components in the core tier offer statistics and monitoring for ensuring stability and evolution in the DARIAH infrastructure despite its decentralised and open nature.

'In-a-box' Services

These are currently two special DARIAH-created solutions aimed at A+H institutions who wish to create their own new digital archives or wish to build a digital research environment for their institution's research

community. Both ‘In-a-box’ solutions combine software that is installed and administered at the institution and ‘connects’ to the DARIAH central infrastructure services.

Interoperability Frameworks for Data and Services

Linking diversity is at the core of DARIAH’s philosophy. Disciplines in the humanities differ greatly with regard to their resources – their data, tools and methodologies. Moreover, innovation is sometimes associated with introducing variations into their data, tools, or methodologies, thereby reinforcing heterogeneity even within a single discipline. Through linking this diversity DARIAH aims to build bridges, and enable researchers from different disciplines or cultural backgrounds to collaborate on the same material, and to share their diverse perspectives and methodologies. A prerequisite to benefit from this opportunity, however, is interoperability between the diverse resources in DARIAH without enforcing specific formats. In other words, DARIAH aims to mediate between heterogeneous resources, and even though interoperability guidelines are optional, their implementation opens up additional opportunities such as increased visibility, collaboration, and the applicability of advanced techniques.

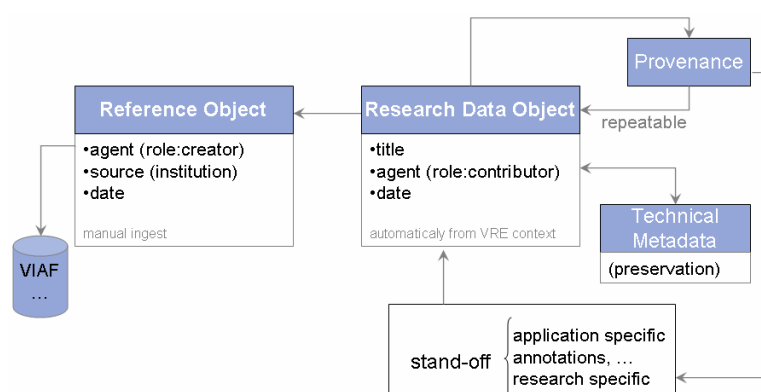


Figure 5: Schematics of the DARIAH content model. To facilitate the role as a mediator between decentralised sources with potentially differing object semantics, the DARIAH metadata approach follows, among others, the conceptual foundations developed by Dublin Core³

Among the interoperability channels in DARIAH are digital objects and the data sources that contain them, as well as services and research environments.

- **Research objects:** Content models in DARIAH distinguish between the reference object (e.g. a sculpture by Michelangelo, a paper manuscript by the archaeologist XY, the born digital BBC Domesday⁴) and the digital research data object used for scholarly work. An object’s content model may have modules that are specialised to specific domains or applications, and it may be distributed to various systems

³ Stuart Weibel. The Evolving Metadata Architecture for the World Wide Web: Bringing Together the Semantics, Structure and Syntax of Resource Description. 1997 (<http://www.dl.slis.tsukuba.ac.jp/ISDL97/proceedings/weibe.html>) and Thomas Baker: A Grammar of Dublin Core. D-Lib Magazine, October 2000 (<http://www.dlib.org/dlib/october00/baker/10baker.html>)

⁴ BBC Domesday (<http://www.domesday1986.com/>)

- **Data sources:** Potential sources may include large institutional archives as well as homepages of individuals. Both the technical protocols for federating these data sources and their policies and concepts need to be shared to ensure effective interoperability
- **Services and tools:** To achieve service interoperability across tiers, service providers, and scholarly domains, issues need to be resolved including the underlying technical paradigm (e.g. SOAP, REST), passing on user identification, and ensuring provenance
- **Research environments:** Ideally, research environments are tailored to the specific needs of a domain or a research question, and they potentially combine various services and tools that may be reused between different research environments

2.4. Scientific Cases

What would be a scientific infrastructure without scientists? What are indeed the expectations of scientists with regards such an infrastructure as DARIAH? Should DARIAH take care of scientists who already have a high awareness of e-science issues? Alternatively, should DARIAH consider how it could work for the humanist scholarly community at large?

The following hypothetical scientific cases set forth examples of scientific work that could benefit from the services that have been defined within the DARIAH organisational scheme.

Providing Access to Data

After three years, the nationally-funded project eXplore has collected, digitised and encoded all the manuscripts of a famous nineteenth century author. The partners have used their extensive scholarly knowledge to record the sources and to compile a digital corpus of the corresponding documents according the best e-humanities practices. They have stored all documents on a private server running at one partner's institution, with access limited to project members. Aware of the necessity to provide access to their results to their research community, they now look for support to identify the best environment for doing so. This environment must offer wide accessibility to and reusability of their scientific results (data, research methods, publications), with the possibility of using additional tools developed by colleagues worldwide. The project also needs advice on the best licensing scheme to which they could bind their data. After a few weeks of joint meetings with DARIAH experts, their data is harvested and integrated in the scholarly network of genetic editions, enabling all scholars worldwide to seamlessly retrieve, query and enrich them under a simple Creative Commons (CC-BY) scheme.

Training Future e-Researchers

Patrick is a professor in a university with extensive expertise in the humanities and digital scholarship, having participated in several national and European initiatives in this domain. He has regularly faced the difficulty of finding adequately-trained students that are able to become quickly involved in his projects, both for compiling the digital sources and for carrying out research on the data.

He thus wants to put together a digital humanities curriculum across several departments in his university and connect it to similar international endeavours. After interacting with DARIAH partners, he starts



adopting reference e-humanities curricula and takes part in the definition of the European Master's and PhD programs in digital scholarship.

Getting Started in Digital Humanities

Anne has just received a four year grant to digitise and study a corpus of late eighteenth century manuscripts in French and German with the goal of studying literary interculturality during this period. As an experienced scholar, she has a very precise vision of the corpus she needs as well as of the features she wants to study and would wish to have encoded in her data. However, she has very little experience in digital research and would like to receive specific guidance so that she may later instruct students and computer scientists employed within her project as to how to proceed.

Anne and two of her best students attend a workshop organized by DARIAH at which she learns of best practices for her project, and also meets with other European scholars carrying out similar projects.

The Solitary Humanist

Antonio has worked for several decades on collecting the works, notes and data of a particular archaeologist. Over the years he has scanned most documents and then transcribed their textual contents in a word processor. He has identified that the corpus as it stands cannot be reliably preserved and further exploited by the research community. Having heard of new methods for managing digital assets, he consults with the DARIAH *virtual curation centre* for guidance on the best way forward.

The virtual curation centre helps him convert his files into a simple but standard TEI XML structure. He is then able to upload all documents to the TEI demonstrator and thereby make them accessible to the archaeological community.

The Digital Postgraduate

Daniela is a postgraduate researcher (PhD candidate) in material culture at a Greek university. She holds a first degree in classical archaeology and a Master's in anthropology. Daniela is an innovator. She uses ICT tools very efficiently and considers them vital for her research.

She is currently researching material culture and its relationship to the perception of space and landscape in an area of northern Greece. Her topic lies within the areas of archaeology and cultural anthropology; therefore, Daniela's sources include artifacts, interviews with local people, and extensive visual material. As her work is largely interdisciplinary in nature, there is a propensity for making fortuitous discoveries while looking for something unrelated; serendipity is very important in her work. This approach means that she needs to login to different services simultaneously - even login multiple times a day if disconnected - and she must keep multiple open windows in her browser. Furthermore, she must be able to evaluate the authenticity and value of any item she discovers.

Thanks to the technical environment developed in the VCC e-Infrastructure of DARIAH, Daniela benefits from both a single sign-on environment for accessing all the necessary digital assets, as well as a virtual portfolio where she can gather all selected sources. The metadata and provenance data associated with an item allows her to evaluate the authenticity and value of an asset, and to link to related items (e.g. other

sources for the same asset, other formats, research addressing that asset) once she has found something of interest. From this portfolio, she can publish geographical views on her data, which she exchanges and discusses with other colleagues in Europe.

Expectations of the Infrastructure

Beyond organisational aspects, the central objective of a research infrastructure is to provide direct support to scholars and scholarly communities so that they can conduct their research more efficiently. In this respect, we can draw out a few expectations which have guided the definition of DARIAH's mission:

- Scholars expect e-research to provide them with a **reliable** place from which their digital assets can be easily **retrievable** and **reusable** by them or by others, and ultimately **cited** within the scholarly works of their community.
- They want to benefit from **reliable** and **coherent guidance** with regard to the purpose of their digital humanities endeavours. Technical issues are thereby made understandable in relation to their research objectives.
- They want to perceive the technical infrastructure as providing all the necessary **basic components and tools** so that they may focus on added-value technical developments.
- They believe that digital scholarship cannot be carried out in isolation but as part of a **community endeavour** where people can share their experience and competence.
- Above all, they want to see the infrastructure as a **facilitator for their research**, and wish to support the principle of wider access to data pooled together across research organisations as the way towards better and more reliable research in their field.

3. Implementation of DARIAH

3.1. Criteria for participation in a VCC

The Selection Procedure

DARIAH potential partners will be asked to submit a proposal requesting membership in the DARIAH construction phase as either a VCC head or a VCC contributor. Potential partners may request membership as a contributor in more than one VCC, but may not propose to head more than one VCC. The proposals must be submitted to the DCO of DARIAH and will be reviewed by external experts and members of the DARIAH Scientific Board, who together will form a selection committee to make recommendations based on an assessment as to what degree the proposal meets the criteria outlined below. The proposals and the recommendations of the selection committee will be presented to the DARIAH General Assembly who will make the final decision about each individual proposal, documenting it and providing feedback to the proposers.

General and Additional Criteria

A set of criteria is outlined here to enable a judgement to be made on the suitability of a partner to contribute to or head a VCC. A common set of criteria will apply to all those wishing to contribute to a VCC,

with an additional set of criteria for those who wish to head a VCC. We propose that each application for membership be assessed against its ability to meet the criteria with a range from 1-5 (1 being the lowest score and 5 the highest).

Criteria for Contributors	Additional Criteria for Heading a VCC
<p>Ability to make a financial and in-kind contribution to the VCC</p> <p>Possesses the expertise and skills required for the VCC services and activities (as defined for each VCC)</p> <p>Knowledge and understanding of A+H research practices and infrastructure requirements</p> <p>Knowledge and understanding of European needs and requirements (as opposed to national needs and requirements)</p> <p>Institutional approval and support</p>	<p>Proven management and financial skills</p> <p>Significant national activity demonstrated by:</p> <ul style="list-style-type: none"> ▪ Funding streams ▪ Project communities <p>International standing demonstrated by:</p> <ul style="list-style-type: none"> ▪ Publication record ▪ Volume/ nature of grants and projects ▪ Conference papers and keynote collaborations <p>Significant previous experience directly relevant to the VCC</p> <p>Institutional setting demonstrates a concentration of activity in the area</p>

3.2. Legal Structure ERIC

What is a European Infrastructure Consortium?

The strategic vision of the European Union is to make Europe the best place for researchers and in this way create a competitive advantage. In the process of realising of this goal, the European Commission carried out a study on the issue of the suitability of the existing national, European and international legal forms for the creation and operation of research infrastructures on a pan-European level⁵. Based on this study, the European Commission proposed to the Council the creation of a new legal form, that of the European Research Infrastructure Consortium (ERIC)⁶, in order to facilitate the goals of the European Research Area. The Council accepted this proposal.

An ERIC is a legal person under European law which is recognised in all EU member states without the need to adhere to formalities in each and every one. It has been modelled after the paradigm of an international organisation, utilising only its the most advantageous aspects. This means that states and other intergovernmental organisations can participate directly in the bodies of the ERIC and that the ERIC can take advantage of tax exemptions as well as a separate and simpler procurement procedure.

⁵ SEC (2008) 2778, pp. 16ff

⁶ Council Regulation (EC) No. 723/2009 of 25 June 2009 on the Community legal framework for a European Research Infrastructure Consortium (ERIC), OJ L 206, 8.8.2009, pp. 1-8

Its establishment is not subject to an international treaty or to ratification procedures in the national parliaments. It is only necessary to make an application to the Commission to initiate an ERIC. The ERIC will be established immediately following the publication of the affirmative decision of the Commission in the EU Official Journal.

What are the Advantages of an ERIC?

The main advantage of an ERIC is its separate legal personality in the entire territory of EU without the need for separate formalities in different countries. This is especially crucial for distributed infrastructures, such as DARIAH. The ERIC legal form provides the opportunity to operate in different locations under a simple, single form and organisation.

Another advantage of the ERIC legal form is the enjoyment of tax exemptions, namely VAT and excise tax exemption, along with exemption from any other tax included in the agreement between its members. This means that the cost for the procurement of the necessary material and equipment for the construction and operation of the ERIC will be significantly lower than in any other legal form. In the same framework, each ERIC is entitled to create its own procurement policy, which need not abide by the provisions of the Public Procurement Directive⁷. This means that procurement of goods and services can be very easy and fast.

DARIAH ERIC

DARIAH proposes to create the DARIAH ERIC at the end of the Preparatory Phase. The DARIAH ERIC will undertake the construction and operation of a distributed infrastructure. The proposed membership scheme and organisational structure is relatively light-weight while still encompassing the major, obligatory bodies and the extra bodies necessary for the optimal operation of DARIAH in close contact with the community and in strict adherence to scientific standards.

DARIAH ERIC Membership Scheme

All interested parties for participation in DARIAH will have the option of applying for membership as one of the following types: a) Members, b) Observers and c) Co-Operating Partners. The first two options are reserved for states which may delegate their membership and representation to research organisations and universities, while the last option is open to institutions from non-participating states.

a) Members

Full membership will allow for the participation in DARIAH to the fullest extent possible. Each member will have the right to use all tools and services offered by DARIAH as described above in the VCC section of this document. In the case of chargeable services⁸, members will be entitled to reduced fees. Moreover, members will have the ability to influence the development of the infrastructure since the majority of the personnel will come from the members themselves. Most importantly, members are entitled to participate and vote in the General Assembly, which is the supreme body of the ERIC and takes all important decisions

⁷ Directive 2004/18/EC of the European Parliament and of the Council of 31 March 2004 on the coordination of procedures for the award of public works contracts, public supply contracts and public service contracts, OJ L 134, 30.04.2004, pp. 114-240

⁸ In principle all DARIAH tools and services shall be offered for free, but some services and/or tools may be offered for a fee

regarding the operation and the future of the ERIC. Furthermore, members will have the option to head one VCC. Naturally, full participation means full financial obligations toward the budget for the construction and operation of DARIAH.

b) Observers

An observer will be fully integrated into the construction and operation of the DARIAH ERIC. An observer will have the right to use all tools and services. An observer will also have the option to participate in the development of these tools. However, such influence will be limited, since an observer will not be able to head a VCC but only participate in its operations. Furthermore, an observer will only be allowed to sit as a guest in the General Assembly.

c) Co-Operating Partners

Only institutions from countries not participating in DARIAH as member or observer can be accepted as a co-operating partner. The level of participation of co-operating partners will be relatively low. The co-operating partners will work together with one or more VCCs with regard to specific tasks agreed to by the relevant VCC(s) and approved by the General Assembly. A co-operating partner will not be required to contribute to the budget, but will have to bear its own costs for the co-operation with DARIAH.

DARIAH Bodies

DARIAH, as a legal person, is required to have its own bodies that will express the will of the organisation and represent it in its external relations and before any court or other authorities.

a) The General Assembly

The General Assembly provides the forum where members are able to participate in the decision making procedure. The General Assembly is the competent organ to decide on, among other issues, the budget and the composition of the other bodies. All other bodies and administrative units are obliged to report to the General Assembly, the supreme body of the ERIC.

The main powers of the General Assembly will be to: a) approve the budget; b) approve the financial reports and the annual report of the activities; c) elect or dismiss the members of the Board of Directors; d) accept new members, observers, co-operating partners; e) elect or dismiss members of the Scientific Board; f) approve the creation, amendment or dissolution of VCCs; g) amend the statute; h) decide its internal rules and procedures; i) expel members; j) dissolve the DARIAH ERIC; and, k) act on any other issue that no other body is explicitly authorised to decide upon.

Each member will be entitled to one vote in the General Assembly. The nature of DARIAH and the proposed funding scheme, as described further below, is in line with all major research organisations in Europe. This justifies the one vote per member scheme.

b) The Board of Directors

The Board of Directors will be the executive body of the organisation, entrusted with the task of managing the day-to-day operations of the DARIAH ERIC and to legally represent the DARIAH ERIC in its external

relations and before national authorities. It will comprise three members with equal powers, with a three year term each. In this way, a mixture of experienced directors with fresh ideas will be ensured. The main task of the Board will be to implement the decisions of the General Assembly.

It is proposed that the elected directors will be active researchers, employed as members of the DARIAH Board of Directors at 50% full time equivalent. The half position will secure the employment of active researchers, since they will not be requested to abandon their research activities. The employment of active researchers as members of the Board of Directors will ensure a continuing, active and close connection of DARIAH management with the academic community it is going to serve.

c) The Scientific Board

The Scientific Board will be entrusted with the scientific overview of the DARIAH ERIC. The Board will consist of qualified individuals such as scholars, software developers, IT experts and experts in other disciplines. They will have the task of evaluating the work and the operations of DARIAH and each VCC. The Scientific Board will submit an annual report to the General Assembly in which its findings on the scientific evaluation of the infrastructure will be included. The Scientific Board will propose to the Board of Directors and the General Assembly any action it deems necessary for the development of DARIAH, including the creation of new VCCs, the amendment of the scope and tasks of each VCC, and the dissolution of a VCC if necessary. Naturally, it will have a significant role in the selection of the VCC head institutions.

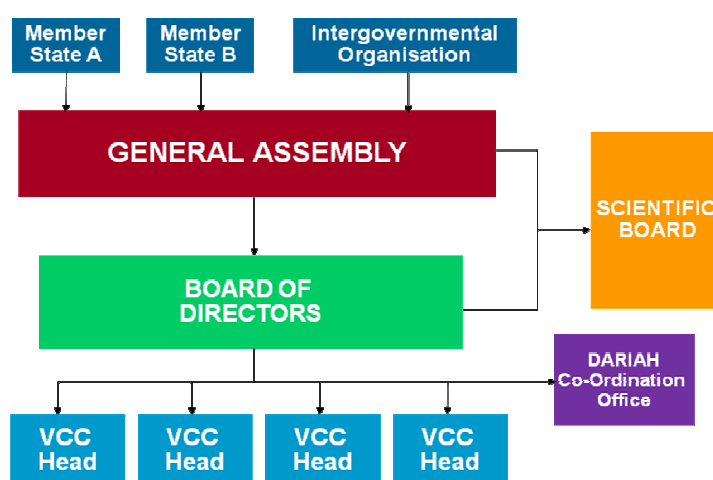


Figure 6: DARIAH ERIC Organisational Chart

DARIAH Policies

a) Access and Data Policy

All DARIAH tools and services will, in principle, be offered for free. This does not mean that some services, such as helpdesks, custom software development or summer courses cannot be provided for a fee. However, this fee will be reduced for participating countries and institutions, although not necessarily at the same rate for all levels of participation. Any software, data or publication created in the framework of DARIAH, either by DARIAH personnel or its users, should be made freely available under an open access/open source licence or its equivalent..

b) Employment Policy

DARIAH should be an equal opportunity employer according to the applicable EU and national legislation.

Any personnel employed by partner institutions and seconded to DARIAH will be subject to the employment policies of the employing organisation.

c) Procurement Policy

The DARIAH ERIC should adopt a simplified, transparent and competitive procurement policy which adheres to the principle of “best value for money”.

The following table presents the basic procedure to be followed according to the total value of the goods and/or services to be procured.

Amount of Procurement	Minimum Quotations
€ 500,00 - € 2.500,00	1 written quotation
€ 2.500,01 - € 10.000,00	2 written quotations
€ 10.000,01 - € 40.000,00	3 written quotations
€ 40.000,01 - € 200.000,00	4 written quotations
More than € 200.000,00	full tendering procedures according to the Public Procurement Directive

It goes without saying that if the procurement is going to be done by a partner institution, a member on behalf of DARIAH, or with the purpose of offering the procured goods and/or services to DARIAH as an in-kind contribution, such procurement shall be subject to the public procurement legislation of the member state.

3.3. Funding Model

In general, humanities and social sciences are very inexpensive in comparison with science disciplines. To take the example of the eight ESFRI projects dealing with physical sciences and engineering, the estimated costs per year amount to six billion Euro. Energy project budgets approach two billion Euros annually.

During the Construction Phase, DARIAH will cost about six million Euro per year, including an allocation of funds specifically for engaging in research projects. One-third of the cost will go to Community Engagement Projects (see chapter 2.1.2) to promote innovative research practices.

DARIAH for its operation, but also for its funding shall rely heavily on the national roadmaps. DARIAH's aim is to utilise and integrate already existing programmes and projects of the national digital humanities roadmap, if a Member already has one, and encourage the creation of a national roadmap, where there is not one. This means that DARIAH will not render existing investment on projects and infrastructure redundant.

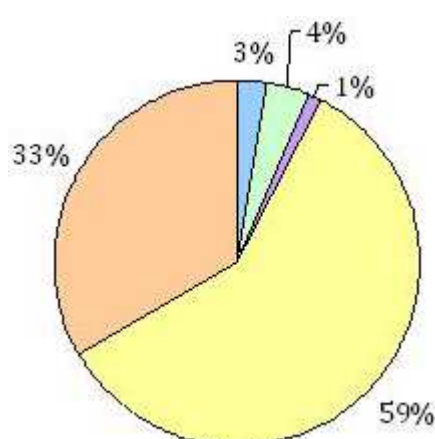
DARIAH will create an infrastructure based on the co-ordinated use, integration and development of the already existing roadmaps, creating thus an infrastructure for the digital humanities which in fact is more than just the sum of the existing projects.

Expenditure

DARIAH has already identified the main cost categories and the amount needed for its successful onset. The following tables show the amounts envisaged for each cost category.

Asset	Amount
Board of Directors	150.000 €
DARIAH Coordination Office	230.000 €
Travel Budget	70.000 €
VCC Costs	3.550.000 €
TOTAL	4.000.000 €
Community Engagement Projects (CEP) via the European Science Foundation	2.000.000 €
GRAND TOTAL	6.000.000 €

Table 1: Minimum budget required for the DARIAH Construction Phase



VCC	Amount	Scale
e-Infrastructure	1.630.000 €	46%
Research & Education Liaison	670.000 €	19%
Scholarly Content Management	1.000.000 €	28%
Advocacy, Impact & Outreach	250.000 €	7%
TOTAL	3.550.000 €	100%

Table 2: Distribution of the budget among the different VCCs

Public Funding Model

DARIAH prefers a mixed funding model, which is a combination of contributions both in cash and in-kind. This model strikes the correct balance between the necessary (cash) funds for the independent operation of DARIAH, and the need to not overburden the member states and respond to their sensitivity regarding the expenditure of taxpayers' money abroad (in-kind).

In this model, a percentage of the contribution to the budget will be required in cash. This amount has been calculated at a minimum of 10%. This amount is needed in order to enable DARIAH to operate the DARIAH Co-Ordination Office. The DCO will function independently of any influence of the member states so that its employees and officers will not have any conflicts of interest.

Cash Contributions

One part of the contribution shall necessarily be in cash. Each member shall be able to decide the funds that will be contributed in cash. However, the cash contribution must in minimum be 10% of the total contribution. This minimum amount is necessary for the remuneration of the DARIAH Co-Ordination Office personnel and the Board of Directors; both will need to operate independently of any influence of any single member, thus being directly employed and paid by DARIAH secures that they will not have any conflict of interests. The cash contribution is also expected to cover other costs, such as the travel budget and costs related to the Scientific Board.

In-kind Contributions

The in-kind contribution represents an amount of money invested nationally, albeit in a co-ordinated way, based on the needs of DARIAH and the decisions of the General Assembly, where the national funding agencies participate. In this way the money invested nationally help the member to fulfil its obligations against DARIAH and at the same time promote the advancement of the national infrastructure and serve the national roadmap for the arts and humanities. These national investments represent the funding for the continued operation of national projects and partners and any additional costs for the necessary actions for their scalability into nodes and components of the DARIAH European Infrastructure.

This means that all members will be able to utilise existing infrastructure and projects or services already in operation and connect them to the wider DARIAH infrastructure as in-kind contributions. In this way



DARIAH ensures that the money already invested in national roadmaps will not be lost. Furthermore, this connection of DARIAH and national roadmaps signifies that DARIAH does not aim to operate in a vacuum, re-establishing a roadmap of its own, but on the contrary that it aims to create a European roadmap based on the interests and priorities of its members.

In-kind contributions, in order to be acceptable, shall have to be approved by the General Assembly as necessary for the operation of DARIAH. In this way it is ensured that DARIAH develops in a co-ordinated way, creating in the end an effective infrastructure, which is something more than the sum of its components. Furthermore this co-ordination by the General Assembly ensures that DARIAH receives the in-kind contributions it actually needs, securing that there are going to be no gaps hindering its operations.

In order to establish the value of each member's in-kind contribution, methods, criteria and a procedure will be set. The proposal is to use national rates of the contributing country and fair market value, depending of the type of contribution. Therefore, an FTE should be evaluated based on the national rates of the member that has is contributing such FTE as an in-kind contribution, whereas the contribution of a room for a meeting shall be evaluated based on the fair market value for renting such a room, etc. This value shall be properly recorded in the books and accounts of DARIAH and shall count towards the total contribution of the member.

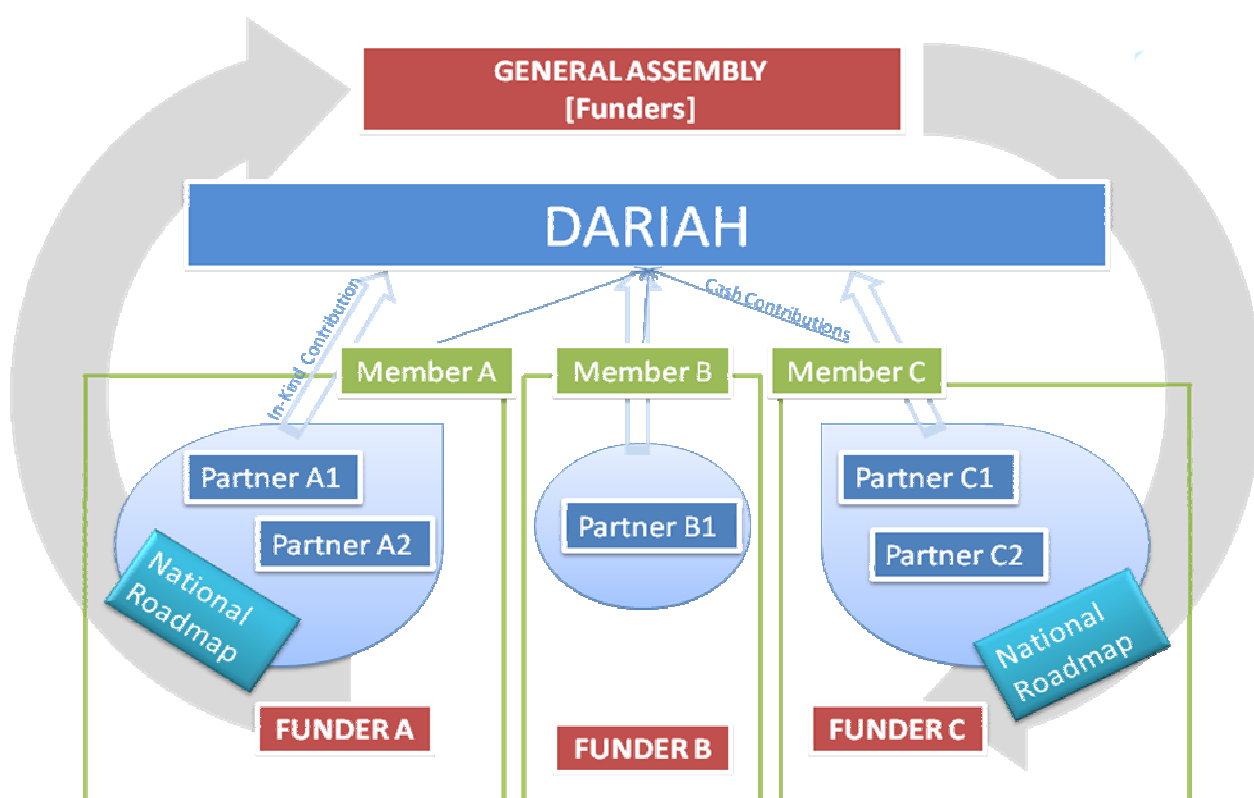


Figure 7: Funding flow

Method for Calculation

For calculating the fees owed by each member, DARIAH proposes to use an index which, generally speaking, accurately represents the ability of each member to contribute to the budget of DARIAH. Three different indexes have been proposed: a) Gross Domestic Product, b) Gross National Income and c) Research and Development Expenditure.

The Gross Domestic Product (GDP), published by the World Bank⁹ is an index which measures the wealth produced within the borders of a state, irrespective of the attributes of the natural or legal person producing it (i.e. national or alien).

The Gross National Income index (GNI), also published by the World Bank¹⁰, expresses the wealth produced within the national borders of a state, plus the wealth produced by nationals abroad, minus the wealth produced by aliens inside the national borders.

The Research and Development Expenditure (R&D Expenditure)¹¹ expresses the amount of funds spent in the domain of research and innovation in each state. This index is not standalone; it can be expressed in various forms such as net amount, as a percentage of the GDP, or as appropriations of the budget.

From these three proposed indexes, the Gross Domestic Product index seems the most appropriate to utilise in order to calculate the fees payable to the DARIAH ERIC by its members. The GDP is used by major international research organisations such as the European Molecular Biology Laboratory (EMBL)¹², the European Organisation for Nuclear Research (CERN)¹³ and the European Space Agency (ESA)¹⁴. In comparison to the GNI or the R&D Expenditure, the GDP measurement minimises volatility of wealth production, with the exception of rare situations like major financial crises, and is not dependent upon private spending or political decisions.

Scale of Contribution

The scale of contribution, which refers to the percentage of the total budget that each member will have to pay as fees, is calculated based on the ratio of the member's GDP to the total combined GDP of all members.

$$\text{Scale of Contribution} = \text{GDP of Member} / \text{Combined GDP of all Members}$$

The fees payable by each member will be calculated based on the total budget expenditure, after having deducted the amount that the observers (see below) will contribute.

$$\text{Fees} = \text{Scale of Contribution} \times (\text{Total Budget} - \text{Observer Fees})$$

⁹ <http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf>

¹⁰ <http://siteresources.worldbank.org/DATASTATISTICS/Resources/GNI.pdf>

¹¹ See Eurostat, Science, Technology and Innovation in Europe, Eurostat Pocketbooks, Luxembourg 2009

¹² Article VII, Agreement establishing the European Molecular Biology Conference (<http://www.embl.org>)

¹³ Article VII, Convention for the establishment of a European Organisation for Nuclear Research (<http://council.web.cern.ch>)

¹⁴ http://www.esa.int/SPECIALS/About_ESA/SEMNO4FVL2F_0.html

The calculated amount includes both the cash and in-kind portions of the contribution, since in-kind contributions will be considered part of the budget.

State	GDP	Scale of Contribution	Total Contribution	In Cash (10%)	In Kind (90%)
State A	\$ 3.652.825	44,94%	1.797.645 €	179.765 €	1.617.881 €
State B	\$ 2.853.063	35,10%	1.404.063 €	140.406 €	1.263.657 €
State C	\$ 860.336	10,58%	423.393 €	42.339 €	381.053 €
State D	\$ 480.021	5,91%	236.230 €	23.623 €	212.607 €
State E	\$ 281.776	3,47%	138.669 €	13.867 €	124.802 €
TOTAL	\$ 8.128.021	100,00%	4.000.000 €	400.000 €	3.600.000 €

Table 3: Example of contributions for five countries for an annual budget of 4.000.000 € without observers

Observer Fees

Observers will have a limited contribution to the budget. Their contribution will also be calculated based on the GDP of the observer and will be half of what the observer would have to pay if they were a full member.

The calculation of the fees in the case of an observer will need two steps. In the first step, the fees will be calculated as if all participating countries were members. These results will be divided by two to determine the observer fees. In the next step, the fees of the members will be calculated using only the full members, but deducting the fees paid by the observers from the final budget.

Member	GDP	Scale of Contribution	Total Contribution	In Cash (10%)	In Kind (90%)
State A (Member)	\$ 1.600.000	28%	1.120.841 €	112.084 €	1.008.757 €
State B (Member)	\$ 2.000.000	35%	1.401.051 €	140.105 €	1.260.946 €
State C (Observer)	\$ 910.000	16%	637.478 €	63.748 €	573.730 €
State D (Member)	\$ 1.000.000	18%	700.525 €	70.053 €	630.473 €
State E (Member)	\$ 200.000	4%	140.105 €	14.011 €	126.095 €
TOTAL	\$ 5.710.000	100%	4.000.000 €	400.000 €	3.600.000 €

Table 4: Step 1: Calculation of the contributions as if all countries were members for a budget of 4.000.000

Member	GDP	Scale of Contribution	Total Contribution	In Cash (10%)	In Kind (90%)
State A	\$ 1.600.000	33%	1.227.087 €	122.709 €	1.104.378 €
State B	\$ 2.000.000	42%	1.533.859 €	153.386 €	1.380.473 €
State D	\$ 1.000.000	21%	766.929 €	76.693 €	690.236 €
State E	\$ 200.000	4%	153.386 €	15.339 €	138.047 €
TOTAL	\$ 4.800.000	100%	3.681.261 €	368.126 €	3.313.135 €
State C (Observer)			318.739 €	31.874 €	286.865 €
GRAND TOTAL			4.000.000 €	400.000 €	3.600.000 €

Table 5: Step 2: Final calculation of the fees for a budget of 4.000.000, with one observer

Budgetary Procedure

The competent organ to decide the budget will be the General Assembly, which will convene at least once per year for budget-related discussions. The General Assembly will have the duty to examine the budgetary proposal of the Board of Directors which shall describe in detail all envisaged expenditures. The head of each VCC, together with the DARIAH financial officer, will aid the Board of Directors in the preparation of the budgetary proposal. The heads of each VCC will be in the best position to document the needs of their respective VCC.

In the budgetary proposal, the Board of Directors will provide a calculation of the fees of each member and observer (if applicable). However, the members of the General Assembly will have the right to adjust the fees in order to accommodate political criteria and decisions.

Annex

I)	DARIAH Partners	38
----	-----------------------	----



DARIAH Partners

DARIAH was founded by 14 partners from 10 countries who joined forces in 2008, resulting in a wide-spread expertise for ongoing joint research, cutting-edge tool and system design, and organisational planning.

During the Preparatory Phase, full membership is only open to the current 14 partners, but the expansion process has begun by adding interested organizations as associate partners.

Founding Partners

Academy of Athens	Greece
Archaeology Data Service	United Kingdom
Centre for eResearch	United Kingdom
Centre National de la Recherche Scientifique	France
Data Archiving and Networked Services	Netherlands
Digital Curation Unit, Athena Research Centre	Greece
European University Cyprus	Cyprus
Institute for Contemporary History	Slovenia
Irish Research Council for the Humanities and Social Sciences	Ireland
Max Planck Gesellschaft, Max Planck Digital Library	Germany
Ruđer Bošković Institute	Croatia
Oxford University Computing Services	United Kingdom
University of Copenhagen, Department of Scandinavian Research	Denmark
University of Goettingen (Goettingen State and University Library)	Germany

Associate Partners

Fondazione Rinascimento Digitale (Digital Renaissance Foundation)	Italy
The Heritage Laboratory, Spanish National Research Council	Spain
HUMLab, Umeå University	Sweden
Institute for Corpus Linguistic and Text Technology, Austrian Academy of Sciences	Austria
Institute of Mathematics and Informatics, Humanities Informatics Department, Bulgarian Academy of Sciences	Bulgaria
Swedish National Data Service	Sweden

As DARIAH grows and takes shape, it will introduce new partner institutions from Europe and beyond. Potential associate partners are encouraged to contact DARIAH with an expression of interest.