1.Final publishable summary report

1.1.Executive summary

The Digital Research Infrastructure for the Arts and Humanities (DARIAH) will 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. The DARIAH infrastructure will connect 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; and, experiment and innovate in collaboration with other scholars.

The DARIAH infrastructure will be designed to strike the right balance between decentralisation and efficiency, empowering individual contributors 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. The construction phase of DARIAH, beginning 2012, will be organised around 4 Virtual Competency Centres (VCCs): VCC1 e-Infrastructure; VCC2 Research and Education; VCC3 Content Management; and VCC4 Advocacy and Outreach. Each VCC will have numerous contributors and each DARIAH partner will be able to contribute to multiple VCCs.

Throughout the Preparing DARIAH phase (Sept. 2008 - Feb. 2011) 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. Preparing DARIAH established a consortium committed to the infrastructure's objectives and has delivered an overall business plan, legal document stating the rights and obligations of DARIAH partners, and has secured sustainable financial support for the construction and initial operational phases of DARIAH. €4,000,000 per year, considered to be the minimum financial support necessary for the construction phase, have been secured for the first 3 years.

DARIAH will establish itself as a European Research Infrastructure Consortium (ERIC), and has encouraged participating countries to sign a Memorandum of Understanding (MoU) underlying their desire to participate in the infrastructure, thereby further strengthening Europe’s position as a centre of world class research. Germany, France, The Netherlands, Ireland, Greece and Croatia have already signed the MoU, with further signings from Denmark, Austria and Slovenia foreseen in the near future. Switzerland has committed itself to join the DARIAH ERIC as co-operating partner, and talks are under way with Lithuania and Norway in the hopes of securing their participation as well.

As DARIAH moves forward into the construction phase, it will continue to strive towards the provision of those components and services of a European research infrastructure in the A+H which will enhance scholarly research methods, and which will facilitate the publication and reuse of research data on an international level. Ultimately, DARIAH strives to be much more than the sum of each single national or individual contribution; it represent the next generation of research potential in the A+H.

1.2.Summary description of project context and objectives

The general objective of the Preparing DARIAH project is to tackle any and all remaining bottlenecks and be ready for the actual construction of DARIAH at the beginning of 2011. Crucial elements to that end have been identified by each work packages short after the kick- off meeting while preparing their operational plan and once again in December 2009 when the amendment for a 6 months prolongation had been officially accepted by the EC. Activities and internal objectives for the second reporting period have been defined and clearly scheduled during the regular management boards and the yearly consortium and General Assembly meetings. When needed, ad hoc committees have been set up.

Work on the strategy had one main action line: provide a clear and comprehensive overview of the state of the art, including:

- An overview and analysis of research and infrastructure activities in the digital humanities across Europe, with emphasis on the partner countries, making recommendations about the strategic direction for DARIAH in light of this analysis.

- Products and services provided by data organisations in all EU member states and the available technologies, tools and standards worldwide.

- Scoping the Primary functions of DARIAH (primitives), scope and prototype Collective Intelligence Resource Map (organisation of a workshop to this end)

- The strategic recommendations (strategic plan) for the next phases, set policies for the construction of DARIAH and develop policies and standards framework for the operation of DARIAH.

On the financial side, the objective was twofold. First, to develop an initial version of a cost model to see what the operation of an European research infrastructure would cost (minimum and maximum); and secondly, to organise roundtable symposia with relevant funding agencies to inform them on the possible funding scheme of DARIAH and receive their feedback on how they could contribute to the financial balance of DARIAH. The first cost model has been updated, a second funding agencies roundtable organised and the “White Paper for European funding scheme” as well as a “Plan of action for an European implementation of funding scheme” analysed.

In terms of governance, the objectives were to investigate governance models and identify the best governance and management structure for DARIAH in the Construction Phase, thereby defining the roles and responsibilities of the various components (Board of Directors, general assembly, etc). The governance model for the Construction Phase (structured around the DARIAH ERIC) has been designed, as has the operational plan. With input from all Preparatory Phase work packages, the DARIAH business plan has also been set up.

As part of the preparation of legal documents, the legal structure chosen during the first period, the European Research Infrastructure Consortium (ERIC), has been further developed. Drafts of the statutes were sent to the Ministries of the countries interesting in joining the Construction Phase, and their feedback has informed the statutes' final version. The first steps to prepare an application to the European Commission for the establishment of the DARIAH ERIC have been taken. The second goal of the legal WP for this period was to draft a ”User licence agreement and draft product and services contracts” and an “Accession form for future partners and Consortium agreement”. Both tasks have been finalised.

The technical work packages WP7 "Technical Reference Architecture" and WP8 "Conceptual Modelling" finalised the technical planning's in preparation of the DARIAH Construction Phase. This includes the completion of the technical demonstrators and prototypes, as well as the compilation of the technical roadmap - with the roadmap consisting of a reference architecture, functional specifications and an initial timeline with operational plan for the construction phase.

Development of several demonstrators and prototypes informed the planning of the technical infrastructure. This included the Arena and the TEI demonstrators, which were already specified in the proposal, started during the first project year, and completed in the second. In addition to them, several prototypes were identified as necessary, after the work in WP7 and WP8 unfolded, and they were also completed during the second project year:

- Arena-II demonstrator
- TEI demonstrator
- PID infrastructure
- OAI-ORE / ATOM federation
- Cloud/Grid interface
- General-purpose, grid-based VRE

The work on the demonstrators and infrastructure prototypes were shared between WP7 and WP8. While each team analysed different aspects of technical services and conceptual modelling, their findings eventually had to be combined into integrated systems.

The roadmap document provides a reference blueprint of key infrastructure services for the forthcoming DARIAH construction phase. It is the nature of digital infrastructure that it evolves continuously as its technological as well as organisational context changes. The infrastructure was designed to sustain such an evolution, and the planning documents reflect this openness that expects more input even as the construction phase is ongoing.

In addition to this, an overall objective was to create a well developed internal communication channel via the internal website, the wiki, the weeklies, the newsletters etc and to raise awareness of DARIAH among the outside world, disseminating the project results to relevant conferences and events and/or via the general website and the Newsletters wide mailing lists. The joint SDH (Supporting Digital humanities) /NEERI (Networking Event for Research Infrastructure) 2010 conferences organised with CLARIN in Autumn 2010 in Vienna were also forwarding that goal.

Next to this, a general wish was to expand the geographical coverage of DARIAH. To this end, potential new partners in countries where DARIAH was not yet represented have been actively approached.

1.3.Description of the main S&T results/foregrounds

1. Introduction

Over the last two decades the natural sciences, in partnership with national governments, have undertaken great efforts to create sustainable (eScience) research infrastructures[2] to address the opportunities and challenges created by the shift towards a "fourth paradigm" of data and information intensive methods of scientific discovery.[3] These efforts have been prompted by a vast

[2] Cyberinfrastructure Vision for 21st Centrury Discovery, National Science Foundation (2007)
[3] Hey, Tansley & Tolle (eds.), "The Fourth Paradigm: Data-Intensive Scientific Discovery" (2009)

increase in our technological capacity to collect research data from scientific instruments, to analyse them and share them with others, and on the other hand, a growing realization that the scientific disciplines are rapidly falling behind in the perhaps equally important tasks of adequately standardizing, curating and preserving such voluminous and complex data sets.[4]
More recently, the arts & humanities (A&H) community has also witnessed a tremendous increase in the volume of digitized cultural materials and a desire not just to offer access to them to the public but also analyze these as research data. This, in turn, has been coupled with growing interest on the part of A&H researchers in the use of innovative, digital research tools and methods as creative and interpretive approaches in their own right and a rising awareness of the importance of long-term preservation and curation of their research outcomes and practices.[5].
In response to these needs, the EU is funding DARIAH to prepare a sustainable research infrastructure for the arts and humanities for the member countries of the European Union. As with similar efforts underway in other countries at a national level (e.g. Project Bamboo in the United-States), DARIAH-EU seeks to allows properly curated and archives research data to be broadly accessible to the public, to facilitate collaboration and scholarly communication amongst researchers and institutions, facilitate the re-use and interoperability of tools and services and encourage the development of innovative new methods of scholarly practice.[6]

2. Services for Art & Humanities Research

Virtual Competency Centres (VCCs) and Service Packets

Internally, DARIAH is organized into four Virtual Competency Centres (VCCs) and a Coordination Office staffed and coordinated by its EU partners. 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 (further details on individual VCCs will be provided below).

Service Packets

In practice, it is not intended that individual A+H researchers, repositories, research centres or other stakeholders will need to discover on their own which of the various VCCs (or combination of different VCCs) should be consulted to provide them with a technical solution or expert consultation on a particular problem. Instead, DARIAH will provide its stakeholders with a single point of contact or portal. By this means the wide range of infrastructure and support services, which the VCCs can offer the DARIAH community, will be bundled into a smaller set of service packages targeting familiar and commonly requested activities. Overall, the individual service packages will be constituted as generically as necessary in order to highlight key DARIAH services to the largest possible number of A+H stakeholders. However, it is also intended to keep this structure flexible enough to allow it to address the evolving and growing needs of A+H stakeholders over time and the more particular needs of user groups within a specific research domain. For example, an A+H centre will be able to draw on the "curation" service package to draw on consultation, training and support services as well as archive-in-a-box reference software as well as storage and other technical services from the DARIAH e-infrastructure. In summary, it is intended that the service packets will function as easily comprehensible, single points of access for A+H stakeholders while hiding the technical and organizational details of the national VCC structure underneath.

---

[4] *Science*, Special Online Issue: Dealing with Data (2011)
[5] Unsworth (et al), "Our Cultural Commonwealth: The report of the American Council of Learned Societies (ACLS) Commission on Cyberinfrastructure for the Humanities and Social Sciences" (2006)
[6] In this respect, the priorities for DARIAH are parallel to those defined in the Humanities Cyberinfrastrcuture report of the ACLS in the United-States (Unsworth, 2006).

Virtual Competency Centers (VCCs)

The following section provides a high-level overview of the tasks and services of the individual VCCs during the construction phase of the DARIAH e-Infrastructure project and how they will interact with each other and with external stakeholders (largely by means of the service packet mechanism) in providing their services to the A+H community.

The distribution of responsibilities and organization of the individual VCCs described below was deliberately chosen to mirror the interests, experience and capabilities of the relevant host countries:

- **VCC1 - e-Infrastructure**
  ... to establish a shared technology platform for A+H research

- **VCC2 - Research and Education Liaison**
  ... to expose and share researchers' knowledge, methodologies and expertise

- **VCC3 - Scholarly Content Management**
  ... to expose and share scholarly content

- **VCC4 - Advocacy, Impact and Outreach**
  ... to interface to key influencers in/for A+H

- **DCO - DARIAH Coordination Office**
  ... to assume the overall responsibility and to ensure adequate operations across all DARIAH organisational units and partners

In the following section the tasks and responsibilities of the individual VCCs are described in more detail.

VCC1 - e-Infrastructure

... to establish a shared technology platform for A+H research. Primary target group: other VCCs, innovators and adopters of technical infrastructure.

Goals and Scope

The VCC1 e-Infrastructure establishes the technological basis for DARIAH as a trusted intermediary, in which community-developed data, tools and services can be preserved, shared and integrated with the larger A+H community.

The infrastructure and interoperability services provided by the VCC can be broadly divisioned into core infrastructure services (such as Authentication and Authorization), reference software packages (such as a pre-packaged, archive-in-a-box solution for local repositories) and data federation services (to permit the development of interoperable, cross-repository web-services applications). In support of these efforts the VCC will provide the DARIAH developers with a centrally hosted service and user registry and a full-featured developer portal (code repository, bug tracking/ticketing system, and documentation).
In addition, the VCC will offer a range of technical consulting services (for example, on current open standards and best practices in digital preservation) and research demonstrators (software prototypes/pilots) in cooperation and consultation with the other VCCs as well as to individual A+H centres, institutions and researchers.

VCC2 - Research and Education Liaison

... to expose and share digitally-enabled A+H research methods, training, expertise and tools. Primary target group: individual A+H researchers and research networks

Goals and Scope

The goal of the Research and Education VCC is to understand A+H research practices and processes in the context of the services provided by DARIAH and to promote the use and application of digitally-enabled methods and tools, with a particular emphasis on promoting the interdisciplinary exchange of research data. It aims to address a range of people and interests, from established researchers to post-graduates to undergraduate students, as well as different disciplines and domains inside and outside of higher education. The VCC will contribute a knowledge base, which captures and links A+H methods, tools, and projects, and references current digital humanities curricula. In addition the VCC will offer a variety of programs and activities to engage with the A+H community via training and education programs, publications, workshops and seminars.

VCC3 - Scholarly Content Management

... to expose and share scholarly content. Primary target group: A+H data centres.

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 publication and reuse. The VCC will offer services and resources for the representation and management of scholarly data, as well as for the management of associated legal and organizational issues to a diverse target community including A+H data centres and research networks, libraries, publishers, digital humanities centres, and individual researchers.

The VCC will facilitate the identification and dissemination of existing digital assets by defining channels for reuse and exchange across communities and research infrastructures[7], providing reference data registries for the description of scholarly data (e.g. authority lists, registries, reference ontologies), collaborating with VCC1 e-Infrastructure to deploy necessary tools and registries and identifying relevant open standards, reference licenses, and best practice guidelines.

VCC4 - Advocacy, Impact and Outreach

... to interface to key influencers in/for A+H. Primary target group: funders, policy makers, industry, and others

Goals and Scope

This VCC will focus on (1) high level advocacy with key influencers in disciplines/industry who are in a position to assist DARIAH, (2) assessing the impact of DARIAH and measuring the 'added value' that it brings by facilitating the transfer of knowledge within and across disciplines, (3) outreach to wide groups of stakeholders outside the A+H community including industry, cultural tourism and publishing, and (4) ensuring the consistent and coordinated promotion and growth of the DARIAH partner networks.

The DARIAH Coordination Office (DCO)

[7] 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)

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

DCO Goals and Scope

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, external and internal communications). The initial legal seat of the DCO will be in Paris, France (CNRS), with a leadership arrangement between France, Germany and The Netherlands.

3. Technical Infrastructure

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 lightweight 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 Bamboo, CLARIN and other ESFRI initiatives).

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.

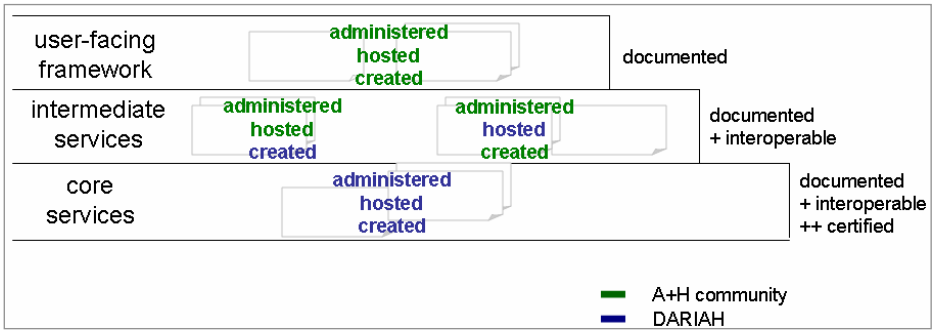


Fig. 4: Each tier may open up different organisational contexts for managing service components. Infrastructure services are created, hosted and administered by DARIAH ensuring reliability and scalability, whereas the A+H community is encouraged to contribute their own, potentially transient services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

a) 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.

b) 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.

c) 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.
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 Domesday[8]) 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
- **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

[8] BBC Domesday (http://www.domesday1986.com/)

Service Catalogue and Roadmap

This section provides an inventory and a timeline for the technical infrastructure services to be created during the DARIAH construction phase.[9] It presents the core building blocks of DARIAH and their role in the infrastructure, although not all of them are currently covered or fully specified. With the core infrastructure in place, further services will be added and adapted as new countries join DARIAH and novel technology developments emerge. For example, partner projects like the EHRI Holocaust Research Infrastructure add new requirements and opportunities to constructing DARIAH. More such dedicated research environments that build upon DARIAH are expected.

User-Facing Framework (UFF)

- *DARIAH Discover* - search across A+H collections, potentially enabling analysis and visualisation (e.g. geo-browsing, relation networks) for subsets of the objects

DARIAH in-a-box services

- *Archive-in-a-box* - reference list to relevant repository technologies
- *Research-Environment-in-a-box* - reference list to relevant research environment technologies

Infrastructure Service Environment

- *Collections Registry* - machine-readable registry for object sources. It fulfils a key role in the DARIAH federation framework.[10]
- *(Ad Hoc) Resource Registry* - machine-readable registry for orphaned objects
- *Metadata Registry* - management and versioning of metadata schemata to enable e.g. metadata schema mappings
- *Services Registry* - references and descriptions for relevant tools and services
- *DARIAH Reference User Registry* - interoperability for user references across DARIAH partner systems with local user management and authorisation; mainly used to enable trustworthy and meaningful provenance and linked to AAI
- *DARIAH Authority Mediation Service* - a framework for integrating divers authority databases, for example for "creators" (e.g. authors, artists), dictionaries, controlled vocabularies and other reference data

Core Infrastructure

- *Persistent Identifier Services (PID)* - PID service and meta-resolver across distinct PID schemata
- *Authentication, Authorization Infrastructure (AAI)* - single-sign-on (for arts and humanities users) across Europe
- *Infrastructure Management & Information Services* - statistics and monitoring of (infrastructure) services and data
- *Provenance Tools* - a framework for tracking and visualising provenance in digital objects as part of ensuring the integrity and authenticity of digital objects

[9] More detailed specifications, cost estimations, evaluations of existing technologies, etc can be found in DARIAH reports D8.1-D8.2

[10] Andreas Aschenbrenner, Tobias Blanke, Marc W. Küster, Wolfgang Pempe, „Towards an Open Repository Environment“, in: *Journal of Digital Information* (JoDI), Vol 11, No 1 (2010).

A detailed look at the technical architecture of all the services is not feasible in the available space. We therefore drill-down into a few selected services only:[11]

The **Persistent Identifier Service (PID)** supports citability of research objects. It is not a single technical service, but rather links various system components with relevant policies. While there are numerous experiences on establishing PID services, DARIAH faces the specific challenges of (a) a huge amount of scientific data for which PIDs need to be minted[12], and (b) the need to weave together diverse PID schemas that are currently in place at arts and humanities data archives.

To tackle the diverse requirements of DARIAH partners and users, the technical architecture will build on (1) a PID service; (2) a meta-resolver to map different PID schemes; (3) usage licences for data objects; as well as (4) a model for pidgin metadata that comes along with the object and informs both humans as well as machines about the object.

The **Authority Mediation Service (AMS)** builds a network of reference data services, including library authority lists (e.g. VIAF[13]), online dictionaries and thesauri, and geo-referencing and place names databases. Reference data services can e.g. assist the description of digital objects through auto-completion (e.g. the author field), and ensure the semantic description of digital objects.

Building on the AMS, researchers can enrich references with research-relevant data. For example, the person data registry by BBAW (DARIAH-DE) will link to VIAF/PND, but will mainly enable researchers to collect evidence from various (potentially contradicting) sources. In other words, the AMS is aware of requirements by researchers, which are different to those of standard reference databases created by libraries and cultural heritage institutions

[11] For a detailed description of all the services, please refer to D 8.1.2 Technical Report and D8.2 Technical Roadmap.
[12] Note: Many existing PID schemas are primarily targeted at publications. Scientific data may pose different requirements to PIDs (e.g. identifying a specific version of a database entry or a specific sequence within a music recording) and a lot more PIDs are required for scientific data.
[13] Virtual International Authority File. http://www.viaf.org/

4. DARIAH At Work

In the recent period, DARIAH partners have been involved in various activities or projects that have already shown the capacity of the DARIAH partners to deliver support or services as for various user scenarios or user communities. Such activities include experimenting with advanced demonstrator technologies, partnering in scholarly digital humanities projects, developing tools and software and participating in major European cultural initiatives with very large research communities.
In the following section we present several examples from all four types of activity to provide a demonstration of the range of skills and competencies upon which DARIAH can begin to build immediately.
For ease of presentation the material in this section has been split into two parts: a) Tools, Infrastructure Services & Demonstrators, and b) Digital Humanities Use Case.

a) Tools, Infrastructure Services & Demonstrators

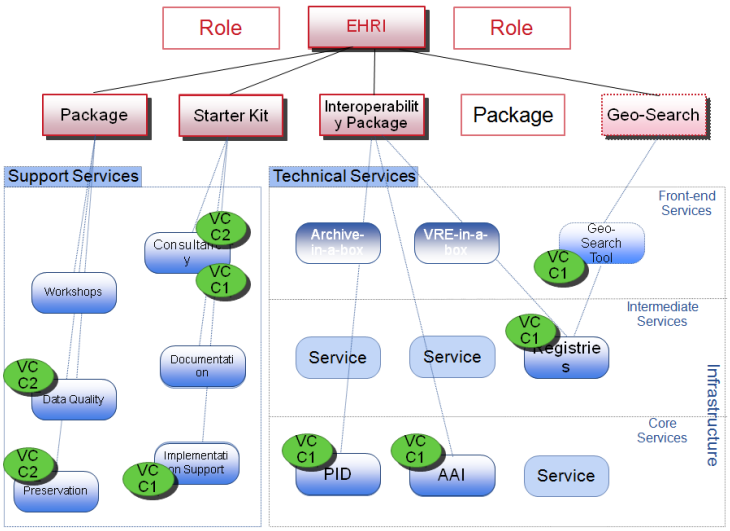
TextGrid: A collaborative research environment for textual sources
TextGrid is one of the first grid-based community projects in the humanities creating an infrastructure for the collaborative editing, annotation, analysis, and publication of specialist text resources. TextGrid represents the humanities in the German national grid initiative D-Grid, and provides a digital infrastructure, a collective network, and a comprehensive and extensible toolset for text scholars.
Technologically, its combination of grid and repositories, as well as services and tools with graphical interfaces establish an open environment that can be adapted to many use cases. In its core functionality, TextGrid focuses on (annotated) text as a data type since there is considerable demand in the community for processing text data.
TextGrid is particularly interesting for its openness. It avoids swamping the user with rules and requirements, yet still fosters participation and collaboration. For example interoperability can be achieved in a stepwise process and following an incentive system:
• Any data format can be uploaded, TextGrid ensures bit-preservation;
• Metadata facilitates data management and retrieval;
• By uploading XML-based texts, a series of services can be used on the data including streaming tools, an XML-editor, and other functionalities;
• For TEI encoded documents, TextGrid offers graphical editing, metadata extraction, and other functionalities;
• Defining a mapping to the TextGrid recommendation for a TEI core encoding allows interoperability on a semantic level.
One of the core goals of TextGrid lies in enhancing the re-usability of existing scholarly texts and services. For both areas—data and services—TextGrid offers various levels of integration: the lowest level offering a minimum barrier to participate, and the highest level offering maximum interoperability on a semantic level. Getting from lowest to highest is a stepwise process, and users are motivated and assisted for taking each step. In other words, interoperability is not given by design, but it is encouraged.
TextGrid is an example of a technology that supports core Digital Humanities of XML annotation in standard formats. It allows for collaborating and sharing of TEI resources as well as their dissemination.

EHRI: Involving the Community
In October 2010, the European Holocaust Research Infrastructure (EHRI) was launched. EHRI aims to support the European Holocaust research community by initiating new levels of collaborative research through the development of innovative methodologies and transnational access to research

infrastructures and services. To this end, EHRI proposes to design and implement a Virtual Research Environment offering online access to a wide variety of disparate and dispersed key Holocaust archival materials and to a number of online tools to work with them. Building on integrating activities undertaken over the past decades by the 17 partners in the consortium and a large network of associate partners, EHRI sets out to transform the data available for Holocaust research around Europe and elsewhere into a cohesive corpus of resources.
The DARIAH expertise has been instrumental in planning and organising the technical service work for EHRI, in particular the requirements, data integration and virtual research environment work. DARIAH's involvement in all these key development fields will ensure that access to integrated archival material will follow best practices and European standards.
In the long-term, DARIAH offers to EHRI its expertise in ensuring data quality and access to its registry services. In detail, DARIAH will collaborate with the following EHRI services:

- Interoperability Service such as registries for collections and metadata
- Authentication and Authorisation Infrastructure
- Persistent identifiers PID

In terms of long-term remote data access, DARIAH will provide assistance with persistent identifier and single-sign-on services to EHRI.



This DARIAH/EHRI service collaboration demonstrates DARIAH's ability to support emerging digital research communities on all levels of their resource integration work

MIXED: Preserving content in the long run
MIXED is a digital preservation project. It uses a strategy of converting data to intermediate XML, and specifically for tabular data.
One of the obstacles to preserving data is software obsolescence. This stumbling block to preservation is usually tackled by either continuously migrating the data or by emulating the software tools. MIXED achieves preservation by migrating data to standard formats.
The strategy used with MIXED converts all datasets, upon ingest to the archive, into an intermediate, generic format. Upon dissemination of a dataset, it is converted from this generic format into a current vendor format of choice. It is likely that the intermediate format will also change, but at a much slower rate. The optimisation is that conversions are split into many contemporary conversions and a few time-bridging conversions. This is a much more manageable situation, and the complexity of bridging time can be dealt with by means of one well-defined format.
MIXED concentrates upon tabular data because the lack of standardisation is most keenly felt here, and there are several reasons for the acuteness of this feeling.

An XML schema, called M-XML, is used by MIXED in such a way that database and spreadsheets are expressed as valid M-XML documents. This format is a non-proprietary representation of the data. MIXED is an open source framework, which accepts converters as plug-ins. The converters allow conversion from existing vendor application formats to M-XML, and vice versa.
On ingest of tabular material into an archive, the data is converted to M-XML, and on dissemination, the data from M-XML is converted to any spreadsheet or database format the end user requires. Therefore it is possible to convert from one application format to another.

eSciDoc — a platform for digital scholarly resources

In the humanities, the preparation and enrichment of a digital source is part of the scholarly work. In particular many funded projects generate, as one of their results or simply as a by-product, digital editions of the reference documents that have been used in the course of the research activity. However, most of the research teams do not benefit from an in-house infrastructure for providing access to these digital documents, and even less for offering additional services allowing the rest of their research community to visualize them and express fine-grained searches.
This is the reason why DARIAH is integrating in its technical service a concept for managing digital documents which is articulated along two main lines:

- Taking-up of a generic repository platform, eSciDoc, which is developed and maintained within the Max-Planck Society
- Offering a dedicated service based on the eSciDoc platform and allowing the management of document encoded in conformance to the Text Encoding (TEI) guidelines.

eSciDoc is an e-Research environment developed specifically for use by scientific and scholarly communities to collaborate globally and interdisciplinary. It is an infrastructure encapsulating a Fedora Commons repository and implementing a broad range of services.
Its service-oriented architecture fosters the creation of autonomous services, which can be re-used independently from the rest of the infrastructure. eSciDoc provides a generic infrastructure and specialised solutions within the context of research questions. It integrates existing solutions and implements new ones.
The target audience of eSciDoc are research organisations, universities, institutes, and companies interested in e-Science-aware knowledge and information management. eSciDoc enables the user to publish, visualise, manage, and work with data artefacts or objects. Objects include both publication data and research data across disciplines. eSciDoc addresses aspects of data reliability, data quality, data curation and long-term preservation. It covers the whole lifecycle of objects, and supports semantic relations between objects.
The eSciDoc system is designed as a service-oriented architecture (SOA) implementing a scalable, reusable, and extensible service infrastructure. Application- and discipline-specific solutions can then be built on top of this infrastructure. Data resource access is provided via Fedora's own REST and SOAP interfaces.

TEI Repository Demonstrator

The purpose of the DARIAH TEI repository is to demonstrate the practical benefits of using TEI for the representation of digital resources of all kinds, but primarily of original source collections within the arts and humanities. As a community-focussed project, the TEI repository also aims to make it easy for humanities researchers to share TEI-encoded texts with others, and to compare their encoding practice with that of others in the TEI community.
The functions it provides are aimed primarily at humanities researchers with the following requirements. Registered users
• can upload and publish a TEI resource and associated materials to an online repository;
• can validate the TEI resources against the TEI-ALL Schema;
• can integrate the metadata provided with their resource into the repository database to facilitate cross-searching;

• can carry our a simple free-text and metadata search or combine it with more sophisticated XML-aware searching possibilities, and extract subsets ("collections") of documents;
• can generate an publishable XHTML version created dynamically from the TEI.

The initial implementation of the TEI repository uses the eSciDoc platform. It also uses the full spectrum of eSciDoc services to support service development and deployment. As everything is open source, new services can be easily added and existing ones can be amended.

Arena 2 Demonstrator

The objective of the ARENA2 demonstrator is to migrate ARENA into a sustainable environment by adding service logic and exposing its resources as autonomous services in a Service Oriented Architecture (SAO) over selected partner data centres.
This demonstrator will exhibit added value in terms of the ability to sustain applications from cultural heritage and arts and humanities research beyond the lifespan of this particular project.
The basic architecture adopted follows the 'Publish-Find-Bind' approach. Services publish themselves to a registry as being in accordance with a web service specification. These services are then found and bound to by a client. Key to creating a SOA implementation of the ARENA2 service is the specification of how services should communicate. This required the creation of the ARENA Gateway Service Specification document (AGSS). The Arena Gateway Service Specification itself consists of a WSDL (Web Services Description Language) document.
In this case the specification is the ARENA Gateway Service Specification, the client is the ARENA2 portal, the services are either compliant monument inventory services or 'wrapped' services based on legacy protocols such as z39.50 or OAI PMH and the registry is an instance of a Universal Description Discovery and Integration registry, the ARENA UDDI registry.
The existing ARENA2 prototype uses a very simple query building interface allowing the user to set values for the 'Where, What, When' elements of the service specification to build up a query. During the process of creating the service specification for each of the 'Where, What, When' elements a universally agreed and available controlled list would have to be selected to of a schema to which each of the diverse data sources could be mapped.
• WHERE – the search interface will have an open layers based geospatial selection interface;
• WHAT – mapping of the MI records to the top-level terms of the poly-hierarchical English Heritage Thesaurus of Monument Types (TMT);
• WHEN – the Forum for Information Standards in Heritage (FISH, UK) maintain the Manual and Data Standard for Monument Inventories (MIDAS) controlled vocabularies.
The service version of the ARENA2 portal integrates a number of design features that allow more intuitive and meaningful searching in comparison to the existing online prototype. The interface styling itself has benefited from extensive user testing of a similar design paradigm, taking place in relation to a separate project undertaken at the ADS.

General Purpose VRE Demonstrator
A VRE is considered a collaborative digital environment that facilitates the integration of information resources and tools for supporting research activities. The Centre for e-Research-based TEXTvre using the TextGrid environment project is concerned with the institutional integration of VREs in the specialised domain of Digital Humanities, specifically the creation of XML-based resources.
This VRE demonstrator aims to find new ways of integrating and organising the heterogeneous and often unstructured digital resources used in humanities research, including advanced search and browse services. Standardisation is unlikely to solve all issues raised in linking up humanities data, for several reasons:

- There is a great deal of legacy data in diverse and often obsolete formats;

- Training users in the application of a standard may incur a significant investment of time and money, which is not always available;
- Standards are generally developed within particular disciplines or domains, such as inscriptions, whereas research is often inter-disciplinary, making use of varied materials, and incorporating data conforming to different standards.

This demonstrator is based on use cases that were identified during the earlier research activities of the JISC ENGAGE project LaQuAT. LaQuAT investigated how to integrate scattered, heterogeneous and autonomous data resources relating to ancient texts, mainly databases but also including XML documents.
The starting point for this demonstrator was D4Science, a production-level infrastructure serving mainly scientific communities, which is not biased towards any particular discipline. gCube, on which the infrastructure is based, is a distributed, extensible system designed to support the full life-cycle of modern research, with particular emphasis on application-level requirements for information and knowledge.
The demonstrator will address a research scenario involving the following stages:

- Document-centric and text-centric search; creation of virtual collections;
- Creation of annotations and links;
- Generation of research reports.

Experimenting technology
During its preparatory phase, the technical partners in DARIAH indentified and assess some of the core technologies that have to be deployed to offer optimal services. Among these, two central technologies, which are central for most eInfrastructures have been experimented:

- Mechanisms for uniquely referencing digital assets (PID – persistent identifiers) so that scientist can make a precise reference to a digital source or component thereof, be confident that this reference will be supported in the long term;
- Implementing a model for the management of heterogeneous sources so that they can be pooled together and used in the most seamless way.

Experimenting PID solutions
Permanent storage and access to digital material requires a more durable referencing method than currently employed by the Internet.
One mechanism that is widely used to deal with this problem of resource location changing is called Persistent Identifiers (PIDs). In short, PIDs are given to any resource or object that needs to be permanently identifiable. Once a PID is minted for a resource, it is tied to this resource for an indefinite period, and any reference using this PID will always refer to the resource it is tied to. When a researcher cites an article or dataset in his (hardcopy) thesis, he needs to be assured that the citation itself will always lead to the original resource he has used.
The PID experiment has been carried out to investigate in what way PIDs can help DARIAH design and construct its research infrastructure. To this end, a prototype PID system has been implemented that demonstrates the following scenarios:

- Forward clients to the actual location of the resource, based on the HTTP protocol, regardless of the PID's origin;
- Enable users to refer to parts of resources;
- Enable users to refer to particular representations of resources.

The architecture of the system accounts for the fact that, although this implementation will be based on HTTP, it should also be capable of working with other protocols. In other words, the architecture has been made flexible enough to adapt to different protocols and delivery mechanisms.

Integrating heterogeneous archives using OAI-ORE
In this infrastructure experiment, which directly targets the interoperability layer, we explored building a highly flexible repository federation for research data on the basis of loosely-coupled services and formats. We examined a prototype for the federation of grid-based TextGrid repository with an iRODS/Fedora repository, which caters for data analysis (i.e. XQuery capabilities on XML/TEI objects) across repositories, and other conceivable applications.
Research questions:

- Modelling in OAI-ORE and related standards. Transfer Fedora / OAI-ORE, automatic vs. manual, internal vs. external, etc.
- Disaggregating object models. Linking with external agents (Europeana, DRIVER).
- Federation of Repositories through OAI-ORE (virtualisation on the semantic/repository layer, as opposed to the storage layer—we look at both layers separately).

Cloud API to the Grid
This experiment analysed patterns for mixing and merging infrastructures. In particular it looked at "repositories" as they overarch scientific infrastructure and interactive applications. In an analysis covering a series of experiments, we find an optimal setup in the combination of grid and web technologies through a REST-based interface, which opens up a variety of novel architectural patterns. This combines two contexts and usage patterns: infrastructure for large-scale scientific applications on the one hand, and open environments for interactivity and user-generated content and services on the other.
We connected grid and web environments and developed a RESTful abstraction upon the Storage Resource Manager (SRM). SRM is a versatile, pivotal grid standard, and it is highly interlinked with its environment on an operating system level. As an interface between a grid node (SRM) and a web server (the repository), we hence looked for a lightweight interface that is capable of translating between the two worlds. Existing cloud services are a premier model for translating between infrastructure and the web. Despite their simplicity, REST-based protocols satisfy all the needs of the web community.
Our experimental implementation therefore re-engineered the REST API of the Amazon S3 storage service using Python WSGI (Web Server Gateway Interface). The advantages of such a loosely-coupled, HTTP/REST-based architecture are manifold: The interface between the repository and the cloud-like service is obviously very light-weight. Due to the loosely-coupled architectural paradigm, the interdependencies between infrastructure and application (in this case: the repository) are minimised and the two can evolve separately.
A generic repository storage API and a decoupled architecture pattern like this enables other services to tie into the system environment. Multiple repositories can build on a single storage, and even specialised services e.g. for format conversion or other administrative tasks are conceivable to work directly at the level of the S3 API. Administrative workflows triggered by the repository, yet executed on the storage level may boost overall scalability of the system environment considerably. Moreover, this loosely-coupled approach may trigger the creation of low- level repository services and hence a variety of agents interacting in an open repository ecosystem.

b) Digital Humanities Use Cases

Philology / Cultural Studies
Ulrike Czeitschner, Karlheinz Mörth and Claudia Resch, Austrian Academy of Sciences
Ulrike Czeitschner, Karlheinz Mörth and Claudia Resch are investigating texts that belong to the so-called Dance of Death and memento mori genres. The Dance of Death project is based at the Austrian Academy of Sciences' Institute of Corpus Linguistics and Text Technology (ILCTT). These texts date back to the Baroque era, in particular the years from 1650 until 1750, and were written to

admonish readers to live a life of virtue in order to be prepared for death at any time. By employing a variety of methodologies and contexts from corpus linguistics, art history and literary studies, the text is able to provide insights into the way of life and the perspectives on death from the period and serve as a rich source for the study of popular imagery.
By offering a coherent, shared research environment for images and texts, DARIAH allows researchers from various disciplines to create a collaborative edition of this corpus. In addition DARIAH can offer means to ensure standards based, long-term digital preservation of the data, while at the same time embedding it in a publication environment that permits ease of access, on-going collaboration and reuse.

Musicology
Joachim Veit (Detmold), Gerhard Allroggen (Hamburg), Raffaele Viglianti (London), and Frank Ziegler (Berlin)
A team of music researchers from across Germany are building a comprehensive corpus of Carl Maria von Weber's works, including music, letters, diary entries, and other resources in a comprehensive corpus. This will allow research into the artistic networks and influences into von Weber's work, as well as his impact beyond his time and historical context.
The team has already invested much effort into creating music-oriented manuscript description standards (called MEI - Music Encoding Initiative) and is actively using and contributing to the virtual research environment TextGrid to create a truly collaborative edition of the corpus. In addition, the team is developing dedicated tools and methodologies to address their new research questions.
DARIAH will support these efforts and provide the researchers with a broad audience for the standardisation and dissemination of their work. In addition, since TextGrid is already part of the DARIAH network, the researchers can be assured that similar initiatives will benefit from their work by reusing their tools and data, and by linking the von Weber corpus to other related corpora.

Philology / Onomastics
Peder Gammeltoft, Institute of Name Research, University of Copenhagen, Denmark
Prof. Gammeltoft is interested in researches into Danish place names, general onomastic theory (the study of names) and the historical record of place names in the Scandinavian Viking Age colonies in particular. One project of recent focus is DigDag, a digital atlas of the Danish administrative boundary units that addresses their history, archaeology, place-names, statistics and geography in a uniform research infrastructure. DigDag is both an online database for individual researchers and a GIS (geographic information system) platform linked to a standard technical interface (API), allowing other tools and services to access and query the database remotely.
Even well-established digital database projects such as DigDag can benefit greatly from working closely with DARIAH. By ensuring that its technical web interfaces (APIs) conform to DARIAH recommended standards, the DigDag project can augment its geographic maps with additional layers of information (such as language distribution data) from cooperating DARIAH collections. This in turn can allow researchers to carry out more complex and sophisticated geospatial analyses in the DigDag research environment than were previously possible.

5. Implementation and Governance

1. 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**
Ability to make a financial and in-kind contribution to the VCC Possesses the expertise and skills required for the VCC services and activities (as defined for each VCC) Knowledge and understanding of A+H research practices and infrastructure requirements Knowledge and understanding of European needs and requirements (as opposed to national needs and requirements) Institutional approval and support	Proven management and financial skills Significant national activity demonstrated by: ▪ Funding streams ▪ Project communities International standing demonstrated by: ▪ Publication record ▪ Volume/ nature of grants and projects ▪ Conference papers and keynote collaborations Significant previous experience directly relevant to the VCC Institutional setting demonstrates a concentration of activity in the area

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 level[14]. 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)[15], in order to facilitate the goals of the European Research Area. The Council accepted this proposal.

[14] SEC (2008) 2778, pp. 16ff

[15] 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

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 it's 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.

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 Directive[16]. 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 lightweight 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 services[17], 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

[16] 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

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

importantly, members are entitled to participate and vote in the General Assembly, which is the supreme body of the ERIC and takes all the important decisions 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.

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.

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 **0**) 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.

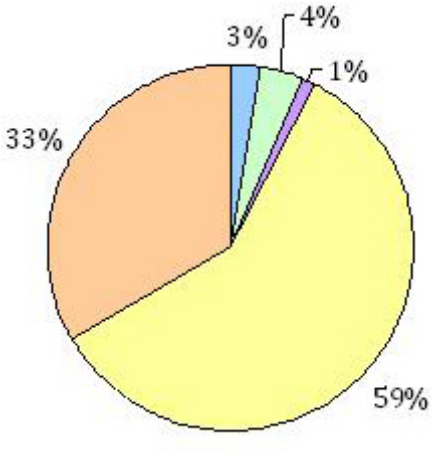
DARIAH will create an infrastructure based on the co-ordinated use, integration and development of the already existing roadmaps, creating thus an infrastructure for the digital humanities, which in fact is more than just the sum of the existing projects.

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 Coordi at on 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 €
GR ND TOTAL	**6.000.000 €**

Table 1: Minimum budget required for he 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.

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 Bank[18] 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 Bank[19], 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) [20] 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)[21], the European Organisation for Nuclear Research (CERN)[22] and the European Space Agency (ESA)[23]. 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.

Scale of Contribution = GDP of Member / Combined GDP if 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.

---

[18] http://siteresources.worldbank.org/DATASTATISTICS/Resources/GDP.pdf
[19] http://siteresources.worldbank.org/DATASTATISTICS/Resources/GNI.pdf
[20] See Eurostat, Science, Technology and Innovation in Europe, Eurostat Pocketbooks, Luxemburg 2009
[21] Article VII, Agreement establishing the European Molecular Biology Conference (http://www.embc.org)
[22] Article VII, Convention for the establishment of a European Organisation for Nuclear Research (http://council.web.cern.ch)
[23] http://www.esa.int/SPECIALS/About_ESA/SEMNQ4FVL2F_0.html

Fees = Scale of Contribution x (Total Budget – Observer Fees)

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.

1.4.Potential impact

The work undertaken during the preparatory phase has provided a comprehensive understanding of the landscape of the digital humanities/e-humanities. DARIAH has built and expanded upon its relationships with key research centres and digital humanities projects which share, in whole or in part, the DARIAH goal of bringing ICT methods and tools to its community of researchers. Foremost, DARIAH has striven to create a sense of European consciousness around research infrastructures for the arts and humanities.

For example, DARIAH and several other infrastructure initiatives including CLARIN, CentreNet, Project Bamboo, the Alliance of Digital Humanities Organizations, and TextGrid, among others, came together to form CHAIN, the Coalition of Humanities and Arts Infrastructures and Networks. In addition to creating ample opportunities for dialogue among its partner projects, DARIAH's engagement with CHAIN will help to ensure that present, proposed, and future activities are interdependent, complementary and oriented towards working together to overcome barriers, and to create a shared environment where technology services can interoperate and be sustained, thus enabling new forms of research. CHAIN is just one example of many in which DARIAH has shown that it believes strongly in collaborative engagement with its cohort projects and initiatives. With the European perspective at the centre of these initiatives, DARIAH helps to cultivate and catalyse the socio-economic advantages for the continent.

DARIAH has cultivated relationships with a wide range of projects and initiatives having similar or overlapping goals and targets. Some of these relationships have led to the establishment of more formal cooperation in the form of key proposals. For example, DARIAH will work closely with the European Holocaust Research Infrastructure (EHRI) to share its knowledge of research infrastructure development, architecture and usage. Some EHRI partners are also integrally involved in DARIAH,

and there are expected to be significant overlaps between the two projects in terms of technological and strategic approaches to research infrastructure development. The aim is to create a cohesive body of integrated research materials that will be made available online to the public.

Similarly, DARIAH will also work closely with the Collaborative European Digital and Archival Resource Infrastructure (CENDARI) project in its efforts to create a powerful platform for delivering and manipulating historical data in a transnational fashion, overcoming the national and institutional data silos that now exist, and raising the level of access for scholars to archives that may already have advanced digitisation programmes as well as to those that do not.

The work undertaken in DARIAH's technical work packages (WPs 7 and 8) has provided a comprehensive and solid understanding of the technological approaches, which will guide DARIAH during the construction phase and beyond. One of the foremost aspects of this work comes from WP8's work on process modelling of scholarly activity, based on a combination of scholarly information behaviour literature survey with qualitative research on how arts and humanities researchers interact with information. This work contributes to the development of a process model of scholarly information work, and also feeds into the specification of an object model for information resources and scholarly objects pertinent to humanities research, an important factor in determining digital infrastructure requirements for scholarship. DARIAH recognises the vital role of the scholarly community in informing DARIAHs development work. This is the community that DARIAH will serve when operational, and therefore a comprehensive understanding of the community's scholarly processes is vital to ensuring that DARIAH delivers value and usefulness to its users. This will in turn have a positive impact on the European arts and humanities research area in general.

Through its advocacy and dissemination work, DARIAH has inspired the inclusion of humanities research infrastructure development on national roadmaps in several European countries, among them Austria, Germany, The Netherlands and France. Further, DARIAH has encouraged the funding and development of national infrastructure initiatives that will feed into and support DARIAH during the construction and operational phases.

All of these action demonstrate the positive impact of DARIAH on the landscape of digitally-oriented art and humanities research in Europe, and shows the way towards continued collaboration and development work with stakeholder from across the broad spectrum of arts and humanities research. DARIAH will continue to leverage its broad mandate in the support of further progress in this area.

1.5.Address of the project public website and relevant contact details.
www.dariah.eu
info@dariah.eu