DARIAH-DE

Funded under: 01UG1110A until N

Report to the Technical Advisory Board (TAB)
DARIAH-DE and CLARIN-D

February 17th, 2015

Project title:
DARIAH-DE – Aufbau von Forschungsinfrastrukturen für die e-Humanities
DARIAH-DE – Digital Research Infrastructure for the Arts and Humanities

Project period: **Reporting period**:
01.03.2011–29.02.2016   01.03.2011–01.02.2015

DARIAH-DE

Project coordination: Göttingen State and University Library

Funding: Federal Ministry of Education and Research (BMBF) project reference number 01UG1110A – N

Table of Contents

Introduction

DARIAH-DE supports digitally-enabled research and teaching in the arts and humanities. The project is developing a research infrastructure, offers tools, core services and access to research data as well as materials for research and education in the Digital Humanities (DH).

DARIAH-DE is the German national contribution to the European research infrastructure "DARIAH-EU Digital Research Infrastructure for the Arts and Humanities".

Currently, 20 partner institutions from the fields of information technology as well as the arts and humanities are involved in DARIAH-DE, including universities, data centers, disciplinary institutions, two libraries, one academy, one commercial partner, one NGO, and one association:.

- Berlin-Brandenburgische Akademie der Wissenschaften (BBAW)
- DAASI International GmbH (DAASI)
- Deutsches Archäologisches Institut (DAI)
- Technische Universität Darmstadt – Interdisziplinäre Arbeitsgruppe Digital humanities (Germanistische Computerphilologie / Philosophie / Ubiquitous Knowledge Processing) (TUD)
- Universität Detmold/Paderborn – Musikwissenschaftliches Seminar (DT/PB)
- Göttingen Centre for Digital Humanities (GCDH)
- Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen (GWDG)
- Verband Digitale Geisteswissenschaften im deutschsprachigen Raum (DHd-Verband)
- Herzog August Bibliothek Wolfenbüttel (HAB)
- Universität zu Köln – Historisch-Kulturwissenschaftliche Informationsverarbeitung (HKI)
- Leibniz-Institut für Europäische Geschichte (IEG)
- Jülich Supercomputing Centre (JSC)
- Karlsruher Institut für Technologie (KIT)
- Otto-Friedrich-Universität Bamberg – Fakultät für Wirtschaftsinformatik und Angewandte Informatik (MInf-BA)
- Open Knowledge Foundation e.V. Germany (OKF)
- Max-Planck-Institut für Wissenschaftsgeschichte (MPIWG) – Max Planck Digital Library (MPDL)
- Rechenzentrum Garching der Max-Planck-Gesellschaft (RZG)
- Salomon Ludwig Steinheim Institut für deutsch-jüdische Studien (STI)
- Niedersächsische Staats- und Universitätsbibliothek Göttingen (SUB)
- Universität Würzburg – Institut für deutsche Philologie (UWÜ)

The central mission of DARIAH-DE is to enable the interoperability of tools and research data in the Digital Humanities field. Following internationally valid and accepted standards and policies, the project aims at ensuring their long-term preservation and future use.

DARIAH-DE also supports and advises researchers as well as research projects in planning humanities research initiatives within a digital environment.

Effective ways of handling digital resources, concepts, and methods in the Digital Humanities must be introduced into training and instruction for humanities researchers at all educational and career levels. In close consultation with disciplinary communities, existing study and training courses are being coordinated, made more visible, and, if necessary, developed more fully. Moreover, DARIAH-DE is developing individual qualification modules, such as international workshops for experts dealing with specific themes.

In order to establish digital research in the arts and humanities, it is necessary to enhance knowledge of digital research methods and practices. The use and application of these processes is supported by special tools and services that are designed, adapted, and made available as a basic infrastructure within the context of DARIAH-DE.

In order to emphasize the added value of Digital Humanities methods, services, and tools, discipline-specific requirements in the form of concrete research questions have been identified. Based on these specifications, individual solutions in the form of "domain specific services" have been developed to demonstrate both specific methods and the overall potential for research in the Digital Humanities. The plan is to integrate more services, developed from affiliated partners, into the infrastructure to ensure a preservation and long term perspective for them.

Research data form an essential basis for scientific work during the entire research process, from collecting and analyzing data to publication and the subsequent use by third parties. Unrestricted access is one requirement that is of central importance for working with research data. DARIAH-DE evaluates, discusses, and recommends standards for data, metadata, licensing and tools as well as for procedures and organizational structures.

This infrastructure will enable researchers to carry out research in an increasingly digital environment, across disciplines and institutions in collaborative ways and towards sustainable results.

Below is an organizational diagram that is reflected in the structure of this report of the specialized technical infrastructure and the interlacing of the four DARIAH-DE core working fields: research, training & education, research data and infrastructure.
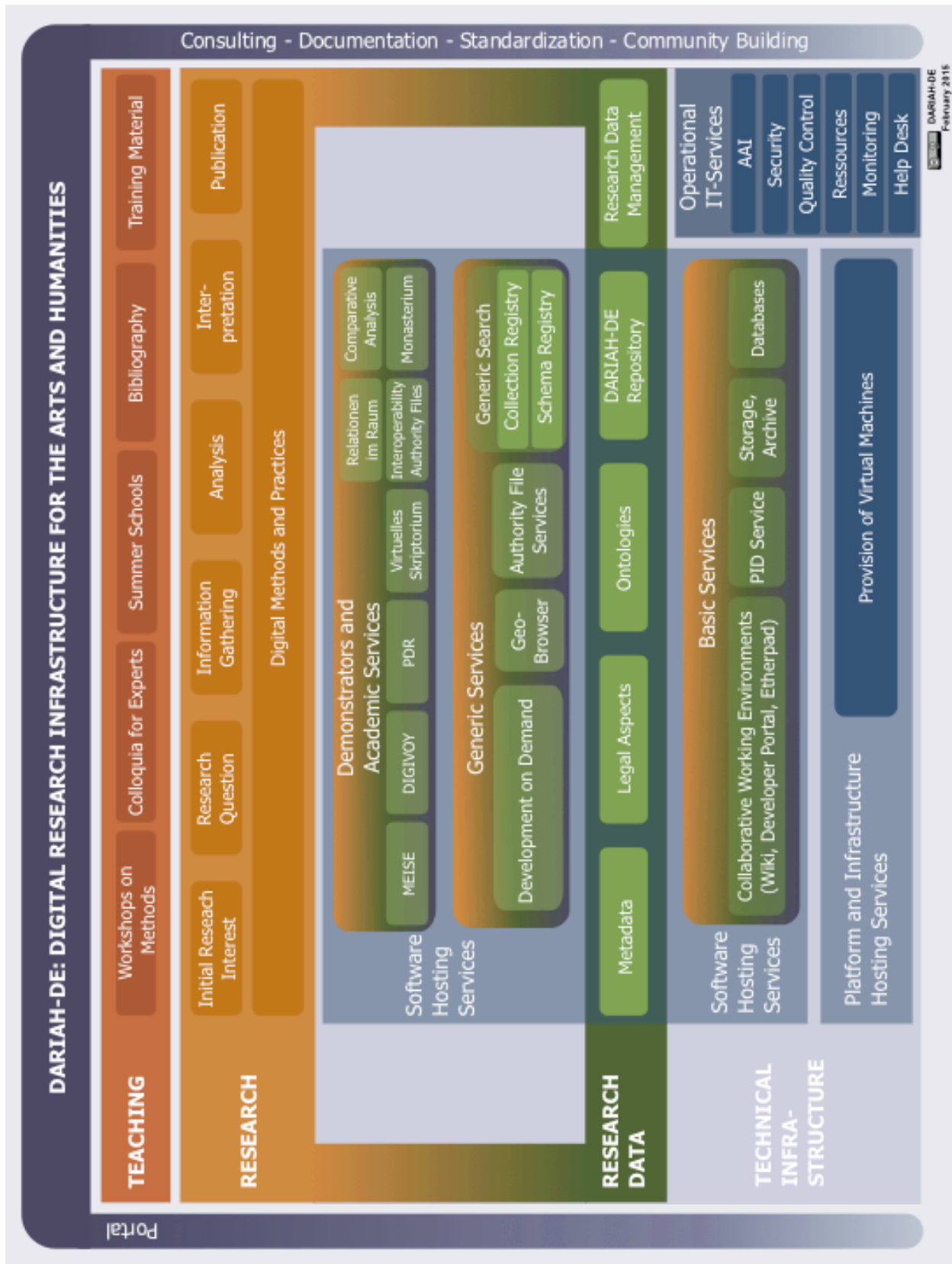
Overview of the DARIAH Infrastructure

Figure 1 – Overview of the DARIAH-DE Infrastructure

RiR – A DARIAH-DE user success story

Authors: Peter Gietz, Thomas Kollatz

The interdisciplinary research network "RiR – Relationen im Raum" (Relations in Space) focuses on spatial relations on Jewish cemeteries. RiR is developing a topography-visualizer. The project integrates data collections from two different disciplines, Judaistic epigraphy and history of architecture, into one repository and visualizes results of searches in these data on maps on a shared platform that allows access to and research on the material. A continuous workflow throughout the different project stages is created, from the drawing of the maps, creating descriptions of the headstones and encoding of the epigraphical texts to the spatially oriented visualization of all materials and with their interlinkages. The appropriate TEI markup and architectural schemata as well as the database technology and visualization of data structures along with the software that ties it all together hypertextually is already developed. This virtual research environment could in future be easily adapted to data material other than Jewish cemeteries, i.e. to any projects with "spatial" relations between objects and/or texts and images.

RiR uses DARIAH infrastructure offerings wherever possible and reasonable (see Figure 2).

Thus the RiR server – a virtual machine, which holds the data base and the retrieval interface – is provided by DARIAH-DE. Having the productive environment integrated into the DARIAH-DE

Figure 2 – RiR usage of the DARIAH Infrastructure

infrastructure gives RiR a number of advantages: The system administration of the operating system can be outsourced to DARIAH. After the end of the project, the service can still be operated. Of course RiR is aware that operating system updates can potentially lead to malfunction of software, because of incompatibilities with new library versions, so on long term new development work on the software will have to be done. To ensure the necessary maintenance and long term support of the DARIAH Infrastructure the service is currently evaluated through the DARIAH Service Life Cycle (see Quality Assurance below).
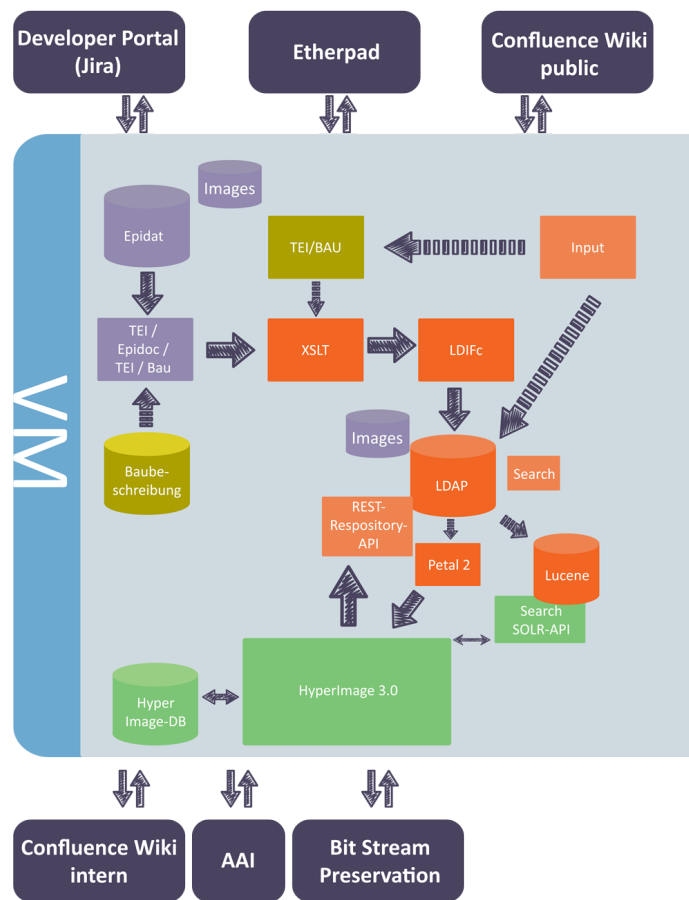
But DARIAH-DE services have proven to be very usefull throughout the RiR project life cycle. RiR is provided with two Confluence wikis, one for internal communication and project management, the other public for project presentation. Both have proven to be valuable tools for the project internal and external communication. Minutes and articles are invariably written and prepared in the DARIAH etherpad tool, providing collaborative editing so that minutes can be created by all participants during the meeting. Etherpad is also used by RiR for collaboratively working on texts (such as this one). Besides the increasing functionality a major point to use this instead of comparable commercial offerings, is the safeness of the data. They will not be used for marketing purposes and we know that they are stored on a secure hardware within a research computer center. The team of developers uses the DARIAH JIRA Issue tracker, as central communication means between the developers and the judaistic epigraphers and architecture historians.

Access to the DARIAH infrastructure is granted by the DARIAH AAI (**A**uthentication and **A**uthorization **I**nfrastructure). Thus all the tools mentioned above can be used with the same user account, which can be a dedicated DARIAH account, but also the federated account of the researcher's home organization. Moreover, RiR has perfect control on who is able to access RiR resources. By applying for memberships into one of the 4 RiR permission groups (rir-admins, rir-user, rir-contributors and rir-developers) we control, e.g., who is able to provide content to the internal RiR-Wiki, the Ticket system, etc. Currently 24 persons are member in rir-contributors and 14 in rir-developers.

More information on the project: https://dev2.dariah.eu/wiki/display/RIRPUB/RiR and https://de.dariah.eu/beteiligte-projekte

Sustainability concept

Author: Heiko Hütter (DAASI)

The goal of DARIAH-DE to provide a holistic research environment for the digital humanities raises the question of sustainability for this environment itself. The individual components of the infrastructure (data, services, tools, publications, courses, …) need to have a long term perspective of their own to be of interest for researchers. To meet this requirement DARIAH-DE will be iteratively converted from a research project with a strictly defined beginning and end, into an institutionalized organization. In this organization the different share- and stakeholders can collaboratively work together to provide and further develop the infrastructure. At the moment DARIAH-DE analyses the huge set of German legal forms to find the best one fulfilling the requirements. Amongst many others these are the most important ones:

1.) First and foremost the organization will enable and support the integration of the diverse digital humanities landscape, consisting of individual researchers, research projects, (research) institutions, service providers and sponsors, which all can be both users and contributors (see Figure 3 for an example). Furthermore the organization has to be able to integrate new members and allow members to withdraw at convenient periods of time.
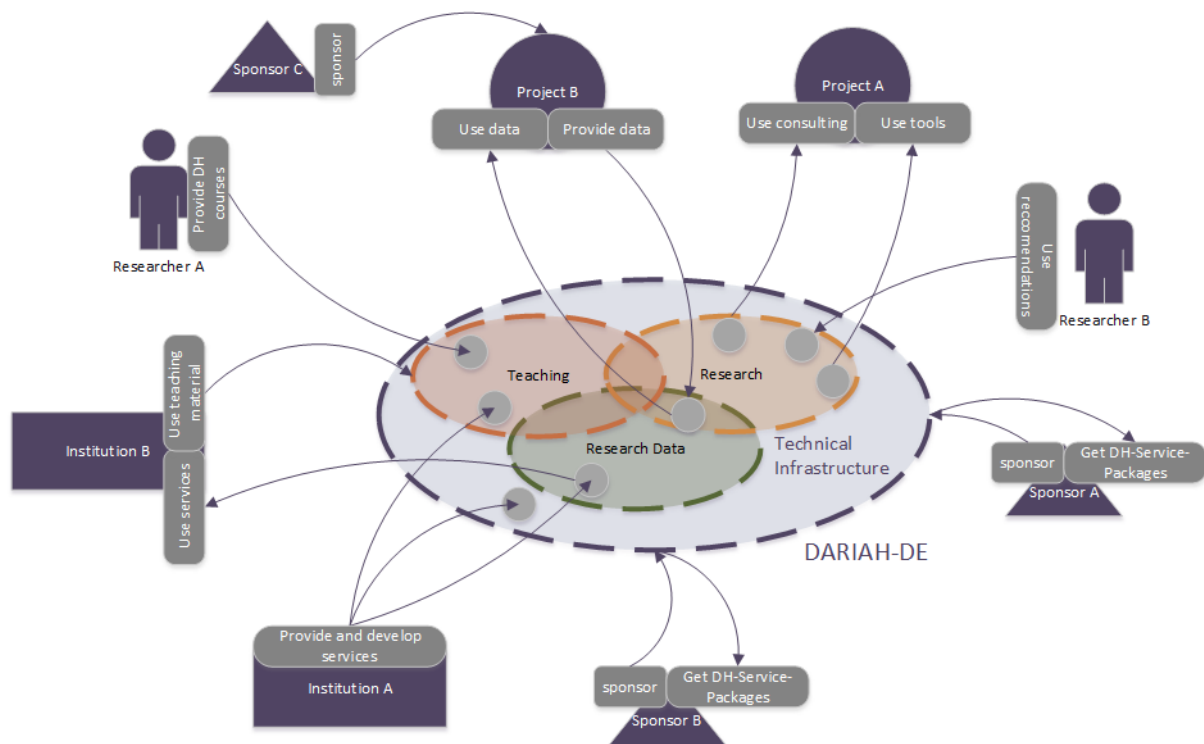

Figure 3 – Example Use Cases for the DARIAH-DE organization

2.) The organization has to encourage the collaboration of researchers amongst many different fields and with different skills and competencies. This is a requirement for the organizational structure, its processes and the provided infrastructure components. To

give an example for the latter, this means that the User Support has to be able to empathize both with computer scientists as well as historians.

3.) The organization and especially the legal form has to enable most flexible funding and financing structures which can be, but needn't be, independent of the organizational steering.

The current plan is to have the basic infrastructure components be funded by the Federal Ministry of Education and Research and the ministries of some federal states. This funding should enable the organization to operate the basic infrastructure components and to run the organization itself.

All resource intensive components will be provided on demand and prices for the usage will be calculated on cost prices of the organization itself. Users can then either subscribe themselves for the services or include them into their research proposals. This indirect funding guaranties a user and demand driven infrastructure landscape.

For two reasons DARIAH-DE will offer *DH Service Packages.* First, this should make the selection of needed services for users easier and second, it allows DARIAH-DE to offer incentives for the basic infrastructure sponsors. *DH Service Packages* include sets of services at respectively different service levels. These sets are assembled based on the need of typical projects of different scales. Although these sets are currently in the concept phase, Figure 4 shows the planned structure. The *DH Service Packages* are only meant as a good starting point and can then be upgraded as needed and of course users will be able to subscribe to single services, too. The sponsors of DARIAH-DE will get a quota of *DH Service Packages,* which they can delegate to projects funded by them.

DH-Service Packages

	Bronze	Silber	Gold
HelpDesk Support for all Services	✓	✓	✓
Monitoring	✓	✓	✓
Etherpad	✓	✓	✓
Fachw. Beratung und Integration	4 Stunden (Remote)	8 Stunden (Remote)	16 Stunden (Remote)
Confluence inkl. AAI	1 Space, 20 NutzerInnen	2 Spaces, 60 NutzerInnen	3 Spaces, 150 NutzerInnen
VM(s)	1	1	2
Storage	1 TB	2 TB	3 TB
Repository	5 GB	50 GB	100 GB
Langzeitarchivierung		10 Jahre	20 Jahre
SVN oder GIT Space		5GB	50GB
PID Service		50.000 PIDs	1.000.000 PIDs

Figure 4 – DH Service Packages

Technical Infrastructure components

Information about the technical infrastructure components is listed in the following sub-sections in an alphabetical order to give an overview of the DARIAH-DE environment.

AnnotatorJS

Author: Harald Lordick (STI)

The topic of web annotation has promising potential and is therefore in lively discussion. A tool like Annotator.js can expect broad disciplinary, cross-application usage scenarios. The work package *Methods* of the DARIAH cluster *Annotations* has especially chosen Annotator.js for its generic and interoperable character, which allows for convenient adaptation, development and integration into the DARIAH-DE infrastructure. Annotator is characterized by a simple, intuitive user interface that allows the annotation of text and images in a huge space of various formats. In the scientific field, the infrastructural support, however, is indispensable. This includes, among other things, differentiated authentication and authorization abilities, sustainable data management of annotations as micro publications, precise, unambiguous and permanent referencing with backlinks to source formats in repositories (e.g. to XML nodes and IDs), the inclusion of semantic structures and normative data.

More Information at
https://dev2.dariah.eu/wiki/pages/viewpage.action?pageId=33689418
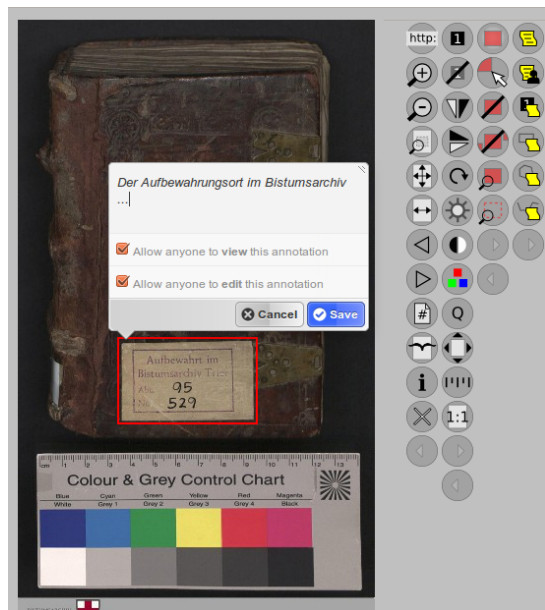
Authentication and Authorization Infrastructure (AAI)

Author: Peter Gietz (DAASI)

Due to the very nature of the research process and due to the scarcity of service- and storage resources, authentication and authorization infrastructure (AAI) is a necessity for every technical research infrastructure. Such an AAI also provides for the re-usage of user accounts of user management systems of the research institutions and it thus prevents that every service has to manage its own user base, which would lead to very many usernames and passwords that have to be memorized by the user.

As reported for the last TAB meeting, DARIAH-DE has set up a SAML based Authentication and Authorization Infrastructure, which basically consists of the following components:

1.) A central OpenLDAP based LDAP server that contains:

- Accounts for researchers that are not connected to a SAML federation or whose Identity Provider (IdP) do not release needed Attributes to DARIAH Service providers ("homeless")
- Similar information objects for researchers that can login with their home organization's account, needed for referencing and for storing additional data provided by the user, that were not released by their IdP
- Groups, the memberships of which provides privileges to specific DARIAH services or data. Such groups either represent projects that e.g. have their own space in the DARIAH wiki, or services. In fact there is a group per privilege role. Thus we, e.g. have 4 groups for the EHRI project: ehri-users with read privileges, ehri-contributors with write privileges in the EHRI wiki space, ehri-developers for access to the EHRI specific parts of the development portal and ehri-admins for the administration privileges.

   This Server is clustered having a provider instance deployed at GWDG and a consumer server deployed at RZG.

2.) A Shibboleth IdP (at RZG) that acts as authentication authority for the homeless accounts and as an attribute authority for all DARIAH users

3.) A number of Shibboleth SPs that are configured to do attribute aggregation by sending attribute queries to the IdP after a user has authenticated

4.) A web-based administration portal that allows for managing group memberships and homeless accounts

5.) A web-based self service interface for providing additional data by all users and for password reset for the homeless users

Although this infrastructure had been operated in production mode since many years a number of improvements could nevertheless be achieved.

1. The administration tool has been extended to distribute the account management. For this a hierarchical role system has been specified and implemented that allows for distributed delegation of management rights within the portal. The roles again are managed in a hierarchical way, so that the representative of the DARIAH Coordination Office can assign national representatives (NP) for each country participating in DARIAH-EU. These NPs can then create organization objects below the respective country subtree and assign the role of organization administrators (OA). These OAs now are given the rights to create DARIAH accounts (by observing a specified policy) and project groups. Now the membership in the project's admin group also gives the privilege to administrate all project groups. Thus each project can manage their own groups. With these new features the management burden can be distributed and, even more important, the person creating a new account has face 2 face acquaintance of the user.

2. The self-service interface, that only had a password reset functionality now additionally has implemented an account registration process, a password forgotten process, and a functionality to view all data stored (including a PDF printout), with the possibility to modify the data.
3. Due to the account generation policy, DARIAH was able to include the DARIAH IdP into the productive DFN-AAI federation.
4. All DARIAH SPs were also included into the DFN-AAI
5. As to the reluctance of campus IdPs to release personal data to the DARIAH SPs, one practical remedy was the self assertion of data via the self service interface. A more political strategic approach which DARIAH took in parallel was the support of the GÉANT Data Protection Code of Conduct (CoCo, cf. http://www.geant.net/uri/dataprotection-code-of-conduct). The complete DARIAH-EU board of directors signed the CoCo support letter that encourages IdPs to release data to CoCo supporting Service Providers, and all DARIAH-DE Service Providers comply by configuration to CoCo.

Another challenge for the DARIAH AAI was how to integrate non-web based services, especially web services. For the REST based DARIAH-DE storage API a solution based on SAML ECP Profile, could be implemented. In addition DARIAH-DE set up an OAuth2 based solution, where OAuth2 tokens are issued upon SAML based authentication.

In addition to the technical work, the DARIAH-DE AAI team liaised with a number of activities, namely:

- brought the AAI solutions into the DARIAH-EU VCC1
- organized the DARIAH/DASISH Access and Authentication Infrastructures Workshop (http://www.gesis.org/en/events/conferences/dasish-aai/).
- regulary actively participated in the FIM4R workshops (Federated Identity Management for Research, cf., e.g., https://indico.cern.ch/event/358127/) where DARIAH is regarded as one of the Pilot projects
- participated in the RDA Special Interest Group Federated Identity Management (https://rd-alliance.org/groups/federated-identity-management.html)
- actively participated in the AAI workshop organized by the European Commission Information Society and Media DG and TERENA (cf. https://www.terena.org/activities/AAI-Workshop/) in preparation of Horizon2020

Bit Preservation

Author: Rainer Stotzka (KIT)

The sustainable management of large amounts of research data is gaining importance for research projects all over the world. The DARIAH Bit Preservation, as a part of an archiving system for the arts and humanities, allows for a high performance, sustainable, and distributed storage of research data as the basis of virtual research environments. A great challenge in designing such a service is to provide a standardized, consistent yet easy-to-use API for accessing the data that remains stable even if backend technology changes over time. We developed a RESTful API for the DARIAH Bit Preservation which includes an administrative extension, and which is secured by an Authentication and

Authorization Infrastructure (AAI) based on SAML (see text above). The implementation of the API offers distributed access by usage of the HTTP protocol and is able to handle a high number of files. Data transfer rates of up to 45 MB/s were achieved for uploading large files in the local network. Recent developments show a growing number of services relying on the DARIAH Bit Preservation to be a safe and persistent storage backend and thereby illustrate the convenient integration in a modular infrastructure. Additionally as only close international cooperation will facilitate outcomes, a DARIAH-EU working group "Data Preservation" including members from various countries is being established to coordinate preservation efforts on a European level.
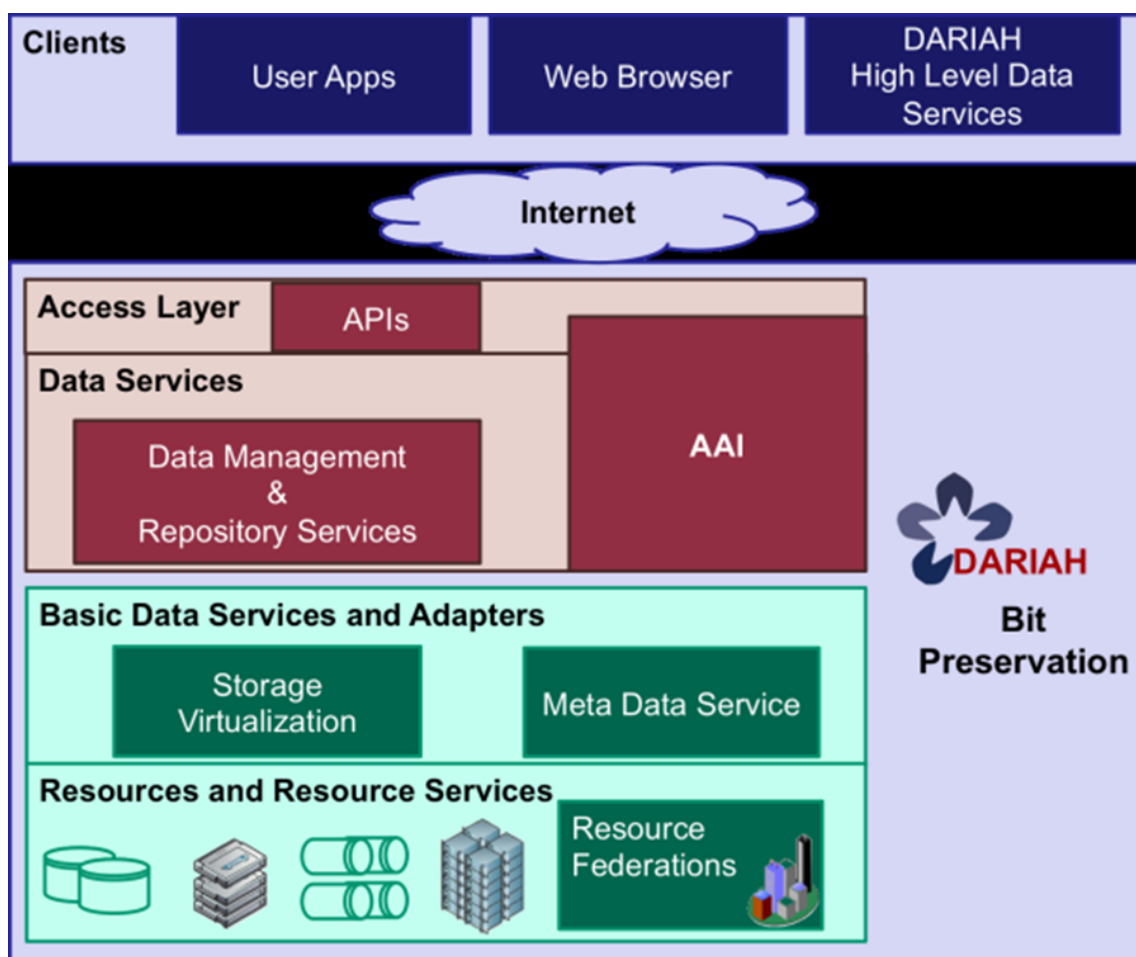
Figure 5 – Architectural overview of the DARIAH Bit Preservation service

Confluence Wiki

Author: Stefan Schmunk (SUB)

DARIAH-DE uses the Confluence Wiki system from Atlassian as a central research management system for internal project management, more specifically for tasks such as the creation of research-specific content and the collaborative preparation of reports and publications within each project. Additionally, the Confluence Wiki is made available to and prepared for external DH research projects on request. The production of texts is a central element in contributing to knowledge in the arts and humanities and is therefore of significant importance as a fundamental instrument for the Digital Humanities. The Wiki system not only allows research projects to manage their content

but also to display it interactively to a wider audience via different access rights, which differentiate between read/write access and read only access.
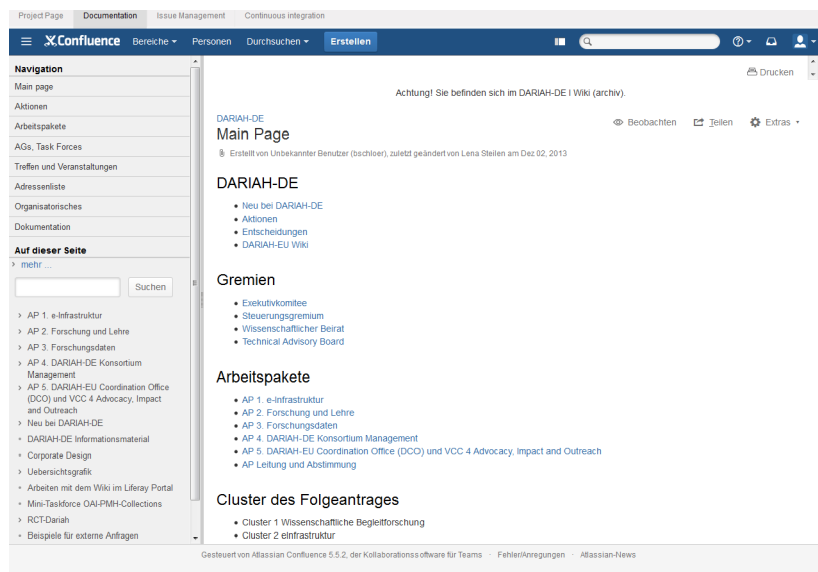
Figure 3: Main page of the DARIAH-DE Confluence space

From a technical and administrative perspective, Confluence also offers an easy handle (including JIRA integration) via the provision of spaces in the installed DARIAH-DE instance and is already shibbolized. Currently the DARIAH-DE Confluence Wiki is already being used by more than 60 external DH projects with a total of more than 800 users are using this collaborative instrument. The following screenshot shows a visualization of typical use patterns.

Figure 4: Typical use pattern of the DARIAH-DE space

DARIAH-DE Portal

Author: Nadja Gruppe (SUB), Stefan Schmunk (SUB)

The DARIAH-DE portal serves as the central access point to the infrastructure of DARIAH-DE and is visited by an average of 65 users a day. The users access scientific and technical services and data. Also research and education in the Digital Humanities are supported with various materials, which can be downloaded or requested. Besides, the portal provides information about recent developments in the fields of Digital Humanities 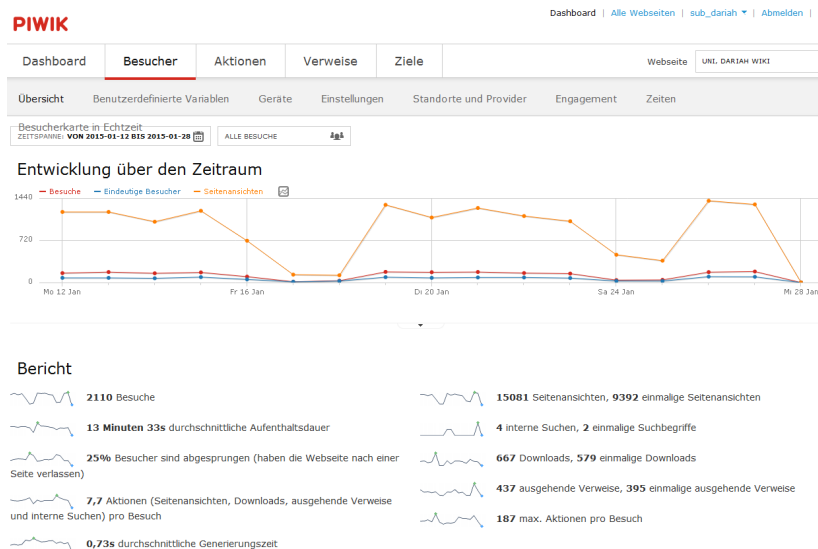by introducing current research projects, weblogs, a collection of re-usable documents for teaching and online-courses.[1] In cooperation with the Dirt Directory, DARIAH-DE is developing a SKOS oriented taxonomy called TaDiRAH.[2]

The portal is implemented with Liferay, an open-source solution that displays data, applications and information within a consistent user interface through a conventional web browser. Liferay is based on Java and offers in addition to a server-oriented content management system (CMS) the opportunity to enhance the portal through the integration of Liferay Portlets or self-developed portlets.[3] This integration e.g. has already been successful implemented with the Zotero-based DARIAH bibliography and other portlets. In the meantime, several other research projects use Liferay instances, so the development can be coordinated with them.

This modular approach brings the benefit of being able to gradually integrate various applications of the DH community into the portal:

1. cursory integration via linking,
2. embedding of the service through iFrame and
3. deep integration through a self-developed portlet.

For the creation of portlets, developers not only can access the technical infrastructure for application programming (developer portal), but also the guidance and support of the experienced DARIAH colleagues.

The cross-linking of services, data and applications is an important issue. Through integration of data to methods, tools, services, projects and research questions the community-based initiatives gain a new level of semantics, which can be of use not only for DH newcomers but also for the experienced DH researchers.

---

[1] https://de.dariah.eu/schulungsmaterial-sammlung

[2] https://github.com/dhtaxonomy/TaDiRAH

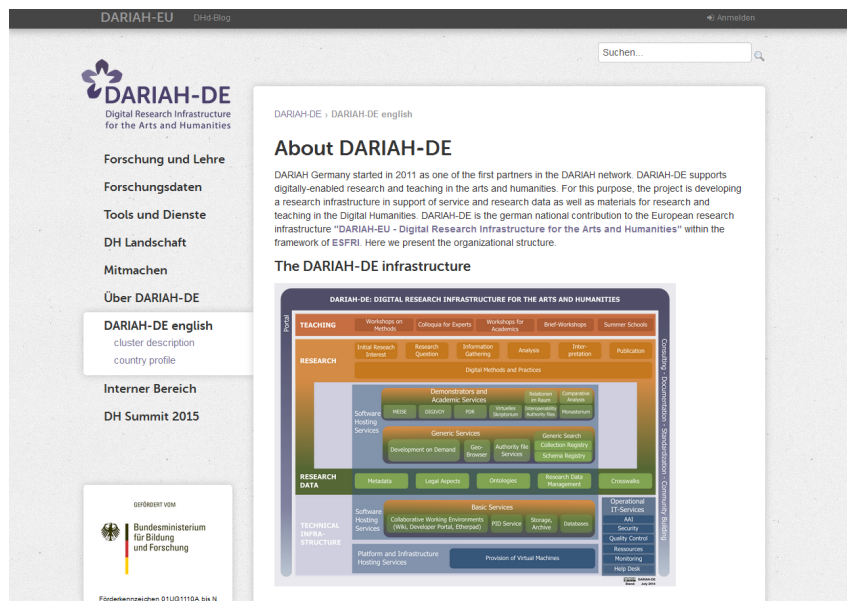[3] Vgl. http://www.liferay.com/de resp. http://de.wikipedia.org/wiki/Liferay_Portal
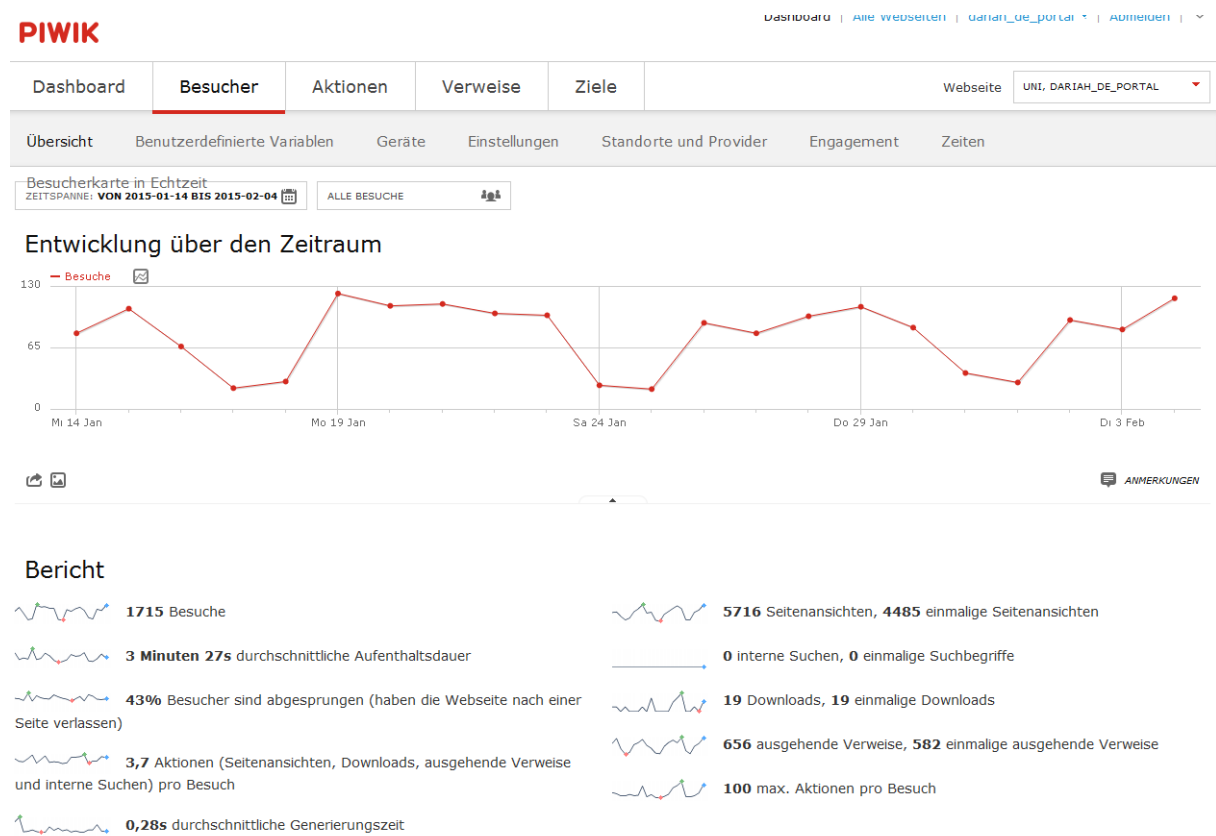
Figure 6 – The current DARIAH-DE portal

Figure 7 – Statistic of website visits, January 2015

In the last two years in particular, the focus was on the editorial research-driven expansion in addition to the technical development of the portal. Thus, the existing teaching materials will gradually be made available as Open Educational Resources for

the communities[4]. In the upcoming month the integration of the DARIAH-DE repository for research data and the implementation of bilingualism are on the agenda.

Databases

Author: Benedikt von St. Vieth (JSC), Jedrzej Rybicki (JSC), Ulrich Schwardmann (GWDG)

Motivation

The efficient storage and sustainable deployment of information is of essential importance:

- the growing amount of data implies challenges on the extraction and accessibility of data.
- for the participation of third parties a simple and useful access is necessary.
- Relational and Non-SQL databases are established instruments to describe relationships between data, to display them and to allow a most easy and performant access.

Problem

At least the following points have to be considered for the operation of a database system (DBS, Datenbanksystem):

- Hardware resources (operation, maintenance).
- Operating systems (installation, maintenance).
- Database system (installation, maintenance, security, backup).

At all three levels critical situations can appear, that endanger the security of data. Be it a blackout, malicious intents of third parties or errors in the software.

Solution

Database systems are administered and maintained by datacenters (Rechenzentrum) and allocated for usage:

- The demanding party receives host/ port and user account/ password for direct access to the database.
- The maintenance of the server resides with the datacenter.
- The backup of the database is ensured.

Leading to

- A Clearly structured responsibilities.
- No administrative overhead for the customers.

The offer includes at first PostgreSQL (JSC) and MySQL (GWDG), but can be extended continually and is dependent on the datacenter.

Data Registries and Generic Search Framework

Authors: Tobias Gradl (MInf-BA), Christof Plutte (BBAW)

---

[4] https://www.oercommons.org/browse/provider/dariah-de

Data Registries and Generic Search Framework – Integrating Heterogeneous Research Data Collections

Access to research data is one of the most important aspects in digital research but often very difficult on a larger scale due to the diversity of digital data sources and the heterogeneity of the information they contain. Due to the high level of diversity and heterogeneity of research data in the Humanities, data integration plays a very important role in the data infrastructures.

Traditional approaches to data integration are typically based on the harmonization of heterogeneous data with respect to the constraints of a globally integrative data structure, such as a global schema or ontology. With the *Data Registries* and the *Generic Search Framework,* DARIAH-DE supports an alternative in order to cope with the limitations of the harmonization-based approach in the specific context of the Digital Humanities: The data federation is established by interactions between *Collection Registry*, *Schema* and *Crosswalk Registry* and utilized e.g. by the *Generic Search* – facilitating a research-oriented and case-based, integrative approach.

The architecture and the data schemata it utilizes can be registered and flexibly associated – resulting in the ability to create broad and comprehensive as well as discipline-specific views of research data.

The collection is registered and described with information required to access the data

A scholar wants to analyze and compare a collection of research data, for which no sufficient query mechanisms are available

1 Describe collection — Collection Registry

Reuse

The data schema is selected or – if not created – modeled.

2 Define schema — Schema Registry

3 Create mappings — Crosswalk Registry

4 Generic Search — Search

Other scholars can utilize the description of the collection, the schema and its mappings for similar requests

Schemata receive semantic associations

The generic search indexes data of the collection.

Figure 8 – Exemplary user interaction for facilitating access to digital research data

The DARIAH-DE federation infrastructure aims to address integration problems by building a comprehensive framework of registries and generic services: The *Collection Registry* serves as an online registry to hold and publish descriptions of research data collections and their machine-readable access points (e.g. OAI-PMH). Metadata schemas used within the collections are registered and semantically enriched in the *Schema- and Crosswalk Registry*, which also allows scholars to define disciplinary- or collection-specific mappings and transformation rules to other schemas (e.g. DC, MODS). The registered collection can then be harvested by the *Generic Search* via the registered

access points using the schema and crosswalk information to map data and to facilitate a faceted, federated search that is dynamically tailored to relevant collections and data structures – all from a single user-oriented portal.
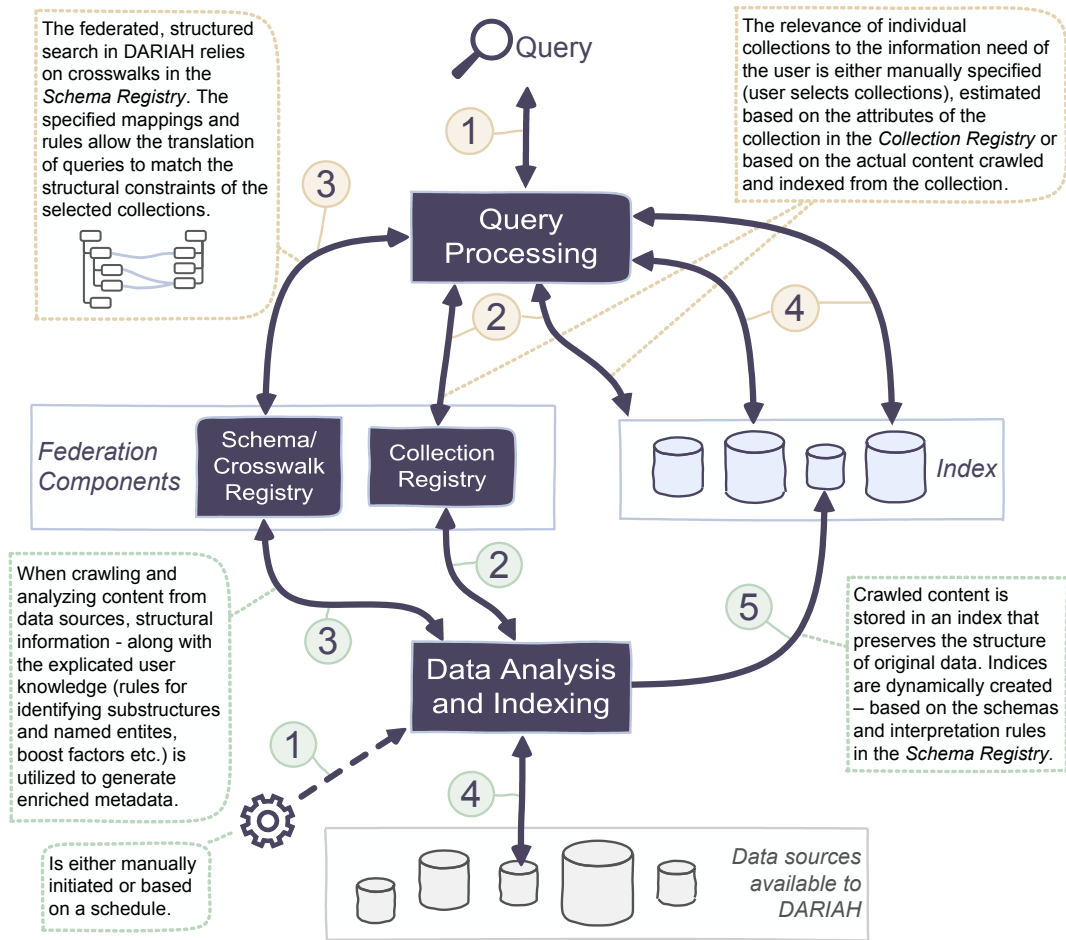

The federated, structured search in DARIAH relies on crosswalks in the *Schema Registry*. The specified mappings and rules allow the translation of queries to match the structural constraints of the selected collections.

The relevance of individual collections to the information need of the user is either manually specified (user selects collections), estimated based on the attributes of the collection in the *Collection Registry* or based on the actual content crawled and indexed from the collection.

Query

Query Processing

Federation Components

Schema/ Crosswalk Registry

Collection Registry

Index

When crawling and analyzing content from data sources, structural information - along with the explicated user knowledge (rules for identifying substructures and named entites, boost factors etc.) is utilized to generate enriched metadata.

Crawled content is stored in an index that preserves the structure of original data. Indices are dynamically created – based on the schemas and interpretation rules in the *Schema Registry*.

Data Analysis and Indexing

Is either manually initiated or based on a schedule.

Data sources available to DARIAH

Figure 9 – The DARIAH federation infrastructure

The user interface of the *DARIAH-DE Generic Search* provides capabilities for a quick full-text search as well as extended options allowing the specification of facets on both content- and collection-level.

Aside from the typical content search, users can choose to retrieve collections relevant to their information need and navigate into the results.
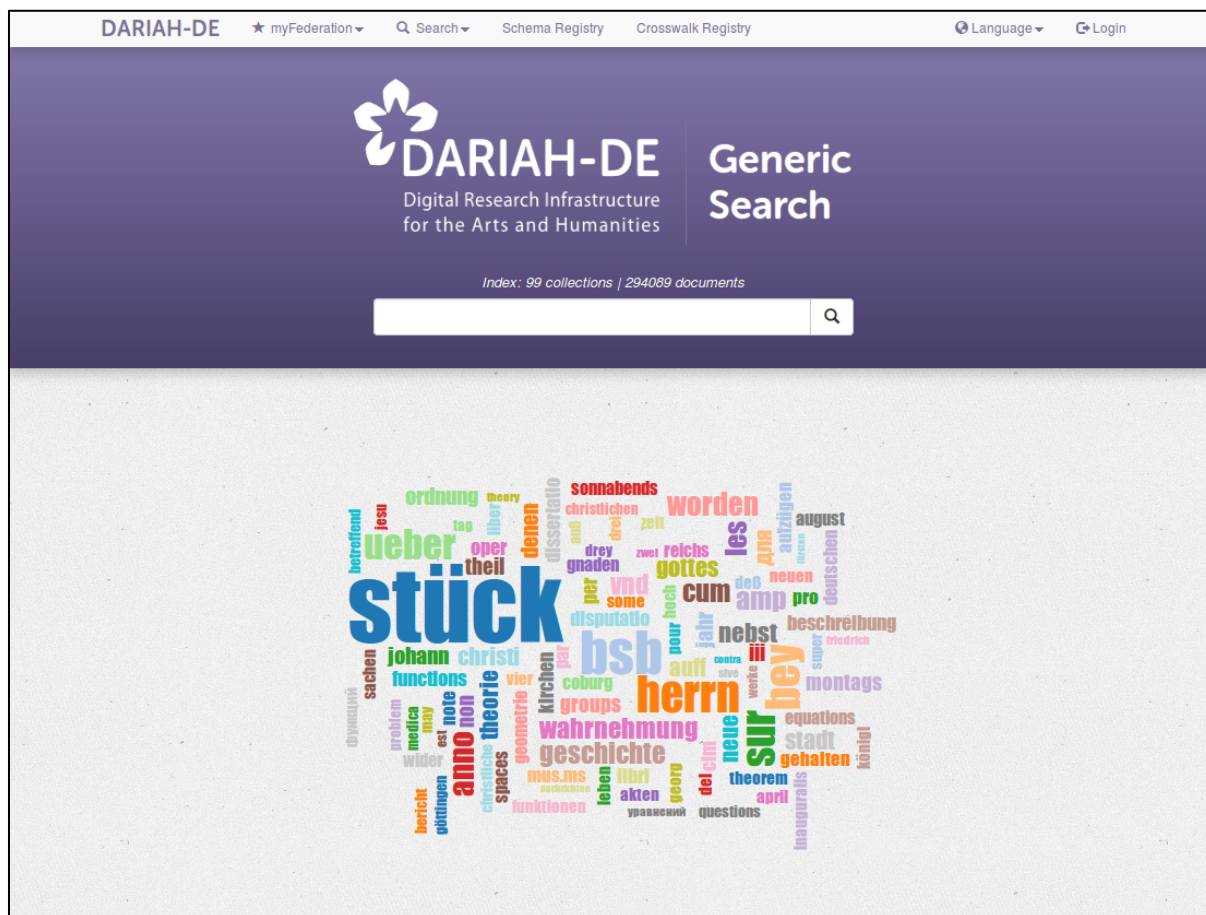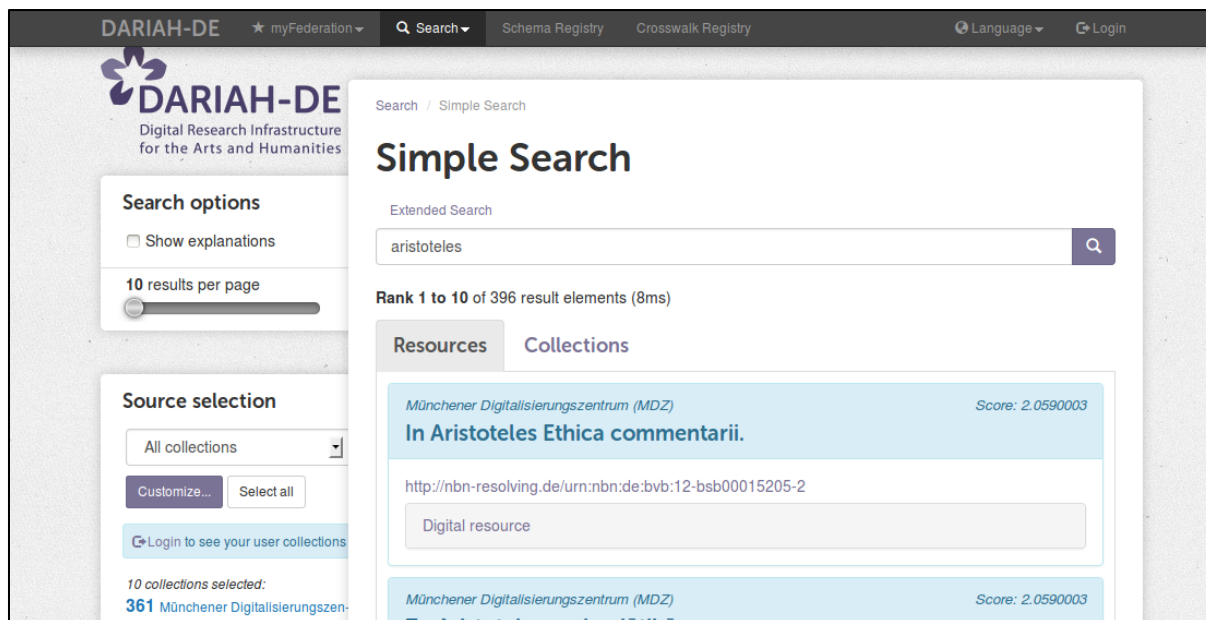
Figure 10 – Start page DARIAH Generic Search

**Figure 11 – Result presentation in the DARIAH Generic Search**

Geo-Browser

Authors: Stefan Schmunk (SUB) and Stefan Funk (SUB)

The Geo-Browser is a tool to visualize, analyse and present spatial data in a temporal context. DARIAH-DE is reusing the Open Source Tool Europeana 4D and their services and interfaces to develop a new user-friendly web-accessible service, which is integrated in the research infrastructure. With the Geo-Browser temporal and geographical relations can be visualized on the basis of „historical" maps of different ages.

Data can be imported in various formats, eg. KML, KMZ or CSV. The most important aspect is at all, that a special tool called „Datasheet-Editor" was developed, so scholars from different disciplines can generate in an easy way their own data samples. Especially the semi-automatic identifying of geo-references, using the Getty Thesaurus of Geographic Names (TGN) and/or Open Street Maps (OSM), allows an easy generation of datasets.

At the moment more then 50 different projects are using the Geo-Browser, coming from all field of the Arts and Humanities but also from e.g. Bio-Diversity and other disciplines of the Natural Sciences. Until March 2015, a new version of the Geo-Browser will be released. Mainly new functionalities in the Data Sheet Editor are implemented, in addition to the transition to a new code base - Java Scipt. In the new version it will be possible to directly reference the data using interfaces to TGN and to OSM. In addition, the KML format has been enhanced so more detailed information of the datasets can be seen in the Geo-Browser. All the developments are based on user requirements. In the next planned release in fall 2015 a deeper integration in the AAI and the DARIAH-DE Repository is forced.

More information at http://geobrowser.de.dariah.eu
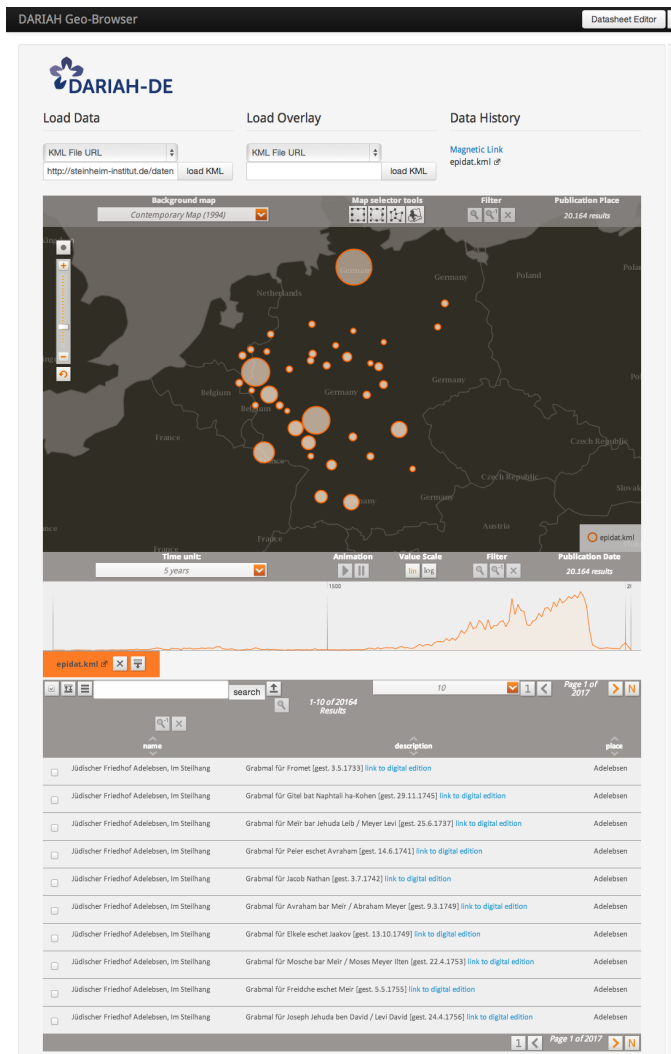
Developer Portal

Author: Bastien Saquet (MPDL)

The DARIAH developer portal has been developed during the DARIAH development phase and has been immediately used as a central component to achieve the other tasks of the DARIAH project. Its goal is to support software development within the DARIAH community. This is done by:

- Making software development tools available.
- Helping collaboration in distributed teams and between the different actors in software development (developers, scientists, librarians, etc.).
- Improving software quality with guidelines and services of the DARIAH infrastructure.

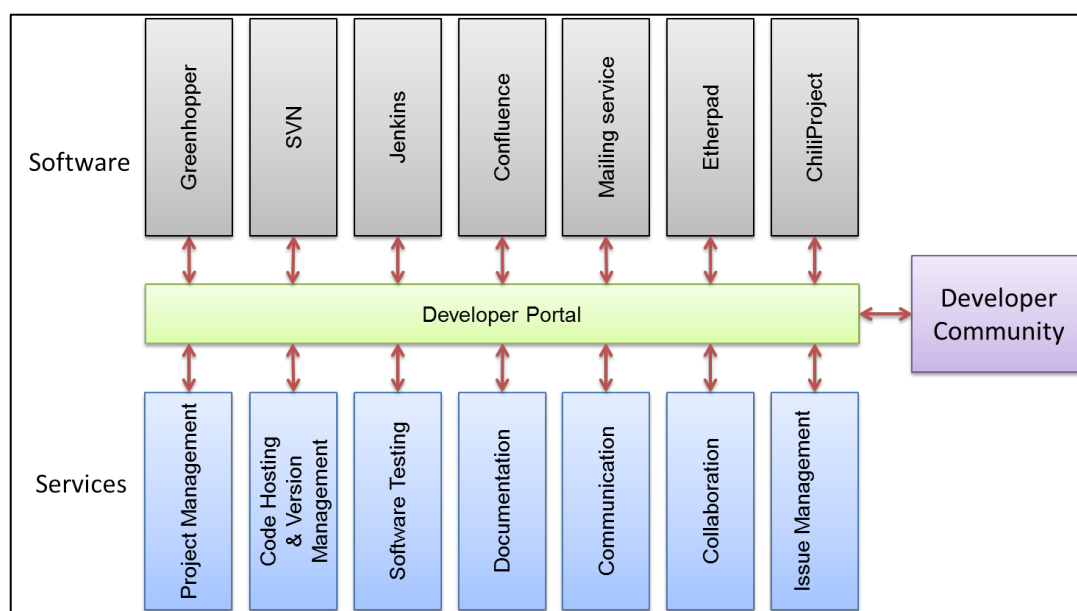The list of the supported tools and services can be found in the figure below.

Figure 12 – Developer Portal: Tools and Services

All this tools and services are available for productive usage to all DARIAH partners and associated projects like EHRI,[5] CENDARI,[6] for a total usage[7] of:

- More than 30.000 pages views per month
- 20 software teams using the portal to develop and maintain their source code
- More than 2.300 users registered

50 projects are using the developer portal.

---

[5] EHRI: European Holocaust Research Infrastructure: http://www.ehri-project.eu/

[6] CENDARI: Collaborative European Archive Infrastructure: http://www.cendari.eu/

[7] Statistics for December 2014, based on the tools Confluence, Jira/Chili, Jenkins and SVN of the developer portal.

Hosting Environment

Authors: Jedrzej Rybicki (JSC), Tibor Kálmán (GWDG)

Introduction

Digital methods become an important item in the everyday arsenal of the researchers from the arts and humanities. Hence, the provision of a means to access compute resources is an important goal of the DARIAH-DE project. Resources offered within the DARIAH-DE e-Infrastructure can be primarily used for either hosting services for the humanities or performing computation intensive tasks. In this regard DARIAH-DE offers state-of-the-art solutions resulting from the long-term experience of the computer and data centers participating in the project:

- Juelich Supercomputing Center (JSC)
- Rechenzentrum Garching (RZG)
- Gesellschaft für wissenschaftliche Datenverarbeitung mbh Göttingen (GWDG)
- Karlsruhe Institute of Technology (KIT)

The DARIAH-DE hosting offer aims to strike a balance between the maximal flexibility for advanced users and seamless access to typical scholarly services for less experienced users. Therefore, different kinds of hosting services are offered. Maximal flexibility is covered by the VM hosting option. In this case, users get access to a virtual machine running on the resources provided by the data centers and can install and configure the services inside as they prefer to, or use the VM for data processing. For less advanced users demanding an instance of existing, well-established services, the option of assisted hosting is more suitable. In this scenario a request for an existing, popular service is made, while the installation and initial configuration is done by the experts at a given computing center. The requester is then provided with an URL to an up-and-running instance. The following list provides an overview of the common features of these two options:

- access to virtual machines with required operating system
- for testing (on-demand: use and throw away) or production purposes
- high-end resources (compute, storage, network)
- basic system configuration provided (firewall, domain, security updates)
- computer center services (monitoring, backup if needed)
- integration with other DARIAH services (distributed monitoring, AAI, bitstream preservation)

There is a growing repository of well-known DARIAH-DE services and service containers, which the computing centers are able to pre-configure automatically. In this case the user only needs to deploy the service. Such applications are called embedded applications (as opposed to stand-alone applications) throughout the rest of this document.

It should be stressed at this point that the DARIAH-DE Hosting Environment is an object of research itself. Computer scientists are seeking the best ways of providing resources and assisting the users. It is still in the pilot phase and thus is undergoing a number of

changes; nevertheless the provided services are of high quality and can be used for both testing purposes and production-ready hosting.

Requirements

Before a service for humanities can be transferred into the DARIAH-DE Hosting Environment, it has to be justified that the service is important for DARIAH-DE and its communities. Depending on the type of service (short time, testing purposes or long term usage), this clarification has to be done by different boards. Furthermore, every hosted service must fulfill some general and site-specific technical requirements. The users of the DARIAH-DE hosting infrastructure and the services hosted within must accept the DARIAH-DE Terms of Use (see page 40).

General technical requirements

Each service request should be justified and sustainability must be proven by the requester. This is something that only the requester can do and that the DARIAH-DE board should evaluate, e.g., by proving that the application is in fact of interest for DARIAH-DE communities, providing the expected number of users, and load and the usage/maintenance time horizon. Furthermore:

- for each service a short description of its functionality is to be provided, including the requested version, a link to the development history (releases) and a future roadmap
- the services should be open-sourced and have a permissive license (Apache, BSD, etc), and the data stored in the machines (i.e. content managed by the services) must also have a permissive license.
- a set of start/stop scripts with documentation must be available, and a VM should survive a restart without any problems.
- the name and email address of a person responsible for the service is to be provided, and this person should be able to answer questions regarding configuration, etc.
- documentation of the service should be available e.g., on the internal DARIAH-DE wiki

An exact specification of the runtime environment must be provided:

- for stand-alone applications, only open-source operating systems will be supported (Debian, Ubuntu LTS, SLES)
- an estimation of resources needed by the application is needed (CPU, RAM, HDD size). Those estimations will be taken into account but there is no guarantee that all the resources are provided from the day one; instead, dynamic reallocation of resources to account for increasing load will be applied
- estimated load should be provided (the expected number of users)

A specification of the runtime environment must include requested dependencies:

- for the service dependencies, the same criteria hold with regard to open-source, documentation and licenses,

- for system libraries dependencies: current, supported, and secure versions available in generic open-source operating systems must be used
- for embedded applications, current versions of application servers (Tomcat, Apache) will be supported
- for services running in application servers or service containers, necessary changes in the general config files should be provided (if applicable)

Security requirements

All services should fulfill basic security requirements (e.g. passwords must not be stored nor exchanged in clear text, etc). After a vulnerability is detected in an application, updates have to be provided and applied. Depending on the type of vulnerability, it is possible that the service will be disabled until an appropriate fix is provided. In case of the VM, the hosting person responsible for the machine should take care of operating system patches and security upgrades when they have to be applied.

Network requirements

For stand-alone services, exact specification of port (ranges) needs to be provided (necessary to configure firewall rules properly). Embedded services hosted within Apache or Tomcat will get the port as configured by the computing center during the server installation. The network traffic of the applications will be monitored. Typically, only passive applications (responding to external requests and not initiating own network connections) will be supported.

Site-specific requirements

There are differences between the sites offering hosting facilities, so that some services can be offered only by particular computing centers. In particular, the assisted hosting option is only offered by the Juelich Supercomputing Center. Although all involved parties try to keep most of the technical details hidden from the users and make the process of gaining access to resources as opaque as possible, there are a few differences in their offerings which should be accounted for.

Juelich Supercomputing Center (JSC)

The operating system can be selected from Debian, Ubuntu LTS, and SuSE Linux Enterprise Server. Due to some internal regulations and external obligations towards the DFN, Juelich Supercomputing Center can only host services offered for "limited" group of scientific users (project partners). A simple way to guarantee that is to offer services secured by DARIAH AAI or filtered DFN AAI. All services accessible to the public must undergo a security audit before respective ports are opened in the firewall. For computation jobs and short-living instances, experimental access to the OpenStack-based private cloud can be provided.

Rechenzentrum Garching (RZG)[8]

The RZG VM hosting cluster is based on a XEN virtualization environment running in a blade center. Entitled admins may request VMs for deploying their specific services. The preferred operating system is SuSE Linux (SLES). VMs from different projects are protected against network isolation and IP spoofing with Iptables rules on the host. The purpose of the hosting environment is to provide VMs for test environments as well as for production services. Due to security and administration reasons, administration of the VMs requires the application for an RZG account. For safe data storage, VMs can connect using data management software to the HPSS system.

Gesellschaft für wissenschaftliche Datenverarbeitung mbh Göttingen (GWDG)

The DARIAH-DE Hosting Environment at the GWDG contains general hosting services offered to research infrastructures and special hosting services offered to DARIAH-DE. The general hosting portfolio of the GWDG has been redesigned and the general hosting services has been broadened by a new cloud environment of the GWDG. The general services include templated virtual machines with current Debian, Ubuntu LTS and SuSE Linux Enterprise Server (SLES) operating systems, monitoring of virtual machines, backup of virtual machines, and setting up HPC clusters on demand. Access is provided to GWDG users and project partners.

For the DARIAH-DE project special hosting services are also provided. These services include both managed hosting of virtual machines on DARIAH hardware, as well as virtual machine images prepared not by the GWDG (external images). The services are currently provided within the DARIAH-DE project, but might also be extended for later phases.

Karlsruhe Institute of Technology (KIT)

KIT does not offer compute resources at this time. They might be, however, incorporated in the later phases of the project.

Process of requesting hosting and compute resources

The procedure of requesting hosting and compute resources has been elaborated within a collaboration of several DARIAH-DE working groups. The procedure is technically supported by the DARIAH-DE Helpdesk System. The requester specifies its request by filling a web form (reference) and submitting such a request. Request from the web form are posted on an internal DARIAH-DE mailing list. Representatives of all computer centers (hosting officers) should be included in this list. The monitoring of the request processing is done via traceable tickets. One of the hosting officers is selected to be a hosting manager. He is responsible for creating tickets for each incoming request. The request should also be attributed to a hosting officer. This attribution should be done in such a way that either the best suited computer center (e.g. all requests for assisted

---

[8] German designations appear as long as they are common in the project context and make the attribution for the reader more easy.

hosting will be attributed to an officer from Juelich) or the center with most available resources will be selected.

Before the preparations for hosting are started, it is proven by the DARIAH-DE Board whether the requested service is needed by DARIAH-DE. Once the hosting manager creates a ticket, the ticket should be assigned to the DARIAH-DE board manager and the computing centers wait until the decision of the Board. A requesting user should be able to view and comment on the request ticket. For requests for VMs, a SSH public key should be provided to the responsible center.

Hosting in DARIAH

|	DARIAH Provider	DARIAH-DE	Developer/ Community
Infrastructure	🟢	🔴	🔴
(template) VM	🟢	🟡	🔴
(basic) Service	🟡	🟢	🟢 🟡
Service	🔴	🟡	🟢

🟢 = responsible   🟡 = may be responsible   🔴 = not responsible

Figure 13 – Responsibilities at different DARIAH infrastructure levels

MEISE – MEI Score Editor

Author: Nikolaos Beer (DT/PB)

The MEI Score Editor (MEISE) is an Eclipse-based client software for viewing and editing MEI-encoded music documents in CMN (Common Music Notation), facilitating proof reading as well as modifying and correcting MEI files. MEISE is designed to satisfy special requirements of research in musicology, currently unsupported by existing music notation software tools. MEISE should be used especially for working with MEI-data as a processing and editing tool, e.g. for creating variants, generating, completing and modifying MEI-files. As a key merit, MEISE gains of its homomorphic internal data structure to the MEI schema tree, allowing for native support for theoretically all MEI encoding principles. This allows for some unique features arising from the field of scholarly music edition, such as the encoding and visualization of variants of different musical sources in one digital MEI representation.

More information at https://de.dariah.eu/mei-score-editor
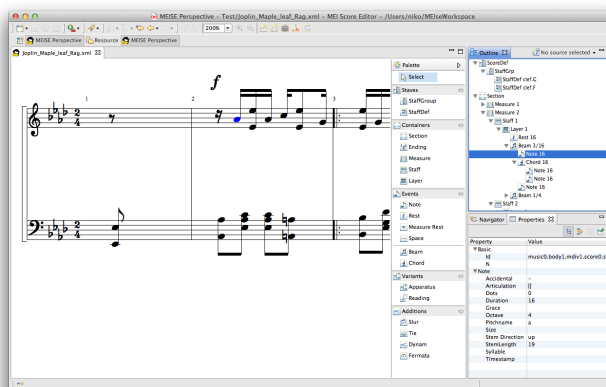
Monitoring

Author: Benedikt von St. Vieth (JSC)

The DARIAH-DE infrastructure and service monitoring is a system based on well-known building blocks and it integrates the existing monitoring systems of the different participating computing centres.

Icinga is used to check the availability of hosts and services and to collect the information of the distributed monitoring systems. NagVis is used to visualize the data collected by Icinga to create a view that shows service dependencies and to create an infrastructure overview.
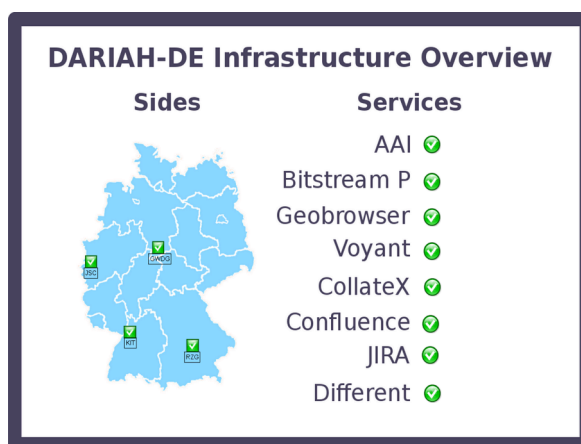
Figure 14 – Infrastructure overview

To monitor the availability of all pieces of the infrastructure, we use different methods:

- Direct checks: Especially DARIAH demonstrators are available for everybody, therefore we are able to start function checks at the central monitoring server that use http requests or other protocols to make sure the service is running and working properly. These function checks can use existing Icinga Probes but some of them are proprietary scripts/programs that service developers provide.
- NSCA (Nagios Service Check Acceptor): For computing centres that already have running monitoring systems the central monitoring server accepts their results. At the central monitoring server, a daemon is running that accepts the incoming results and redirect it to Nagios.
- NRPE (Nagios Remote Plugin Executor): Whenever a computing centre does not want to monitor parts of the DARIAH-DE infrastructure we are able to use a installed NRPE at the target machine to execute remote checks.

Because there is no central registry of hosts/services, we use a combination of git and Jenkins to maintain the configuration of the central monitoring server. Known administrators are allowed to download, modify, and commit the changes to a repository, afterwards this configuration is checked for validity and then copied to the central monitoring.

PID-Service

Author: Tibor Kálmán (GWDG), Daniel Kurzawe (SUB)

To ensure the sustainable long-term preservation of digital objects not just the data content is important, but also a sustainable references to the data. Therefore, persistent identifiers (PIDs) are used to build a stable layer between the reference to an object (URI, etc.) and the digital object itself. In DARIAH-DE, PIDs are used in different contexts, from data citation to referencing digital objects in long term archiving systems. One of the most common use cases is the identification of digital objects on the preservation layers in several data centres. The PID Service is also used for referencing objects in the DARIAH-DE and TextGrid Repository for humanities data. Furthermore PIDs are used to linking files and its metadata together to complex digital objects.

The services required for the PID resolution and the PID management are not built up by DARIAH-DE, but DARIAH-DE relies on the PID-Service of the European Persistent Identifier Consortium (EPIC)[9]. The PID activities within DARIAH-DE are coordinated with EPIC. Therefore, DARIAH-DE hosts a PID test environment[10].

DARIAH-DE also faces some issues related to current PIDs. Since the research data in the arts and humanities domain is highly heterogeneous, the PID usage scenarios and, respectively, the types of the utilized PIDs are also diverse. Interoperability on the level of the PIDs would help further the collaboration of the various A+H disciplines. For example, the possibility of converting an EPIC PID to another type of identifiers (especially DOI PIDs) is an emerging issue, as well as referencing subsets of digital objects and giving more possibilities for the granularity.

---

[9] http://www.pidconsortium.eu/

[10] http://dariah-vm07.gwdg.de/

In DARIAH-DE, PID related issues are handled in the e-Infrastructure cluster. DARIAH-DE also aims to contribute to the DARIAH ERIC with its existing know-how and services. Therefore, DARIAH-DE co-initiated the establishment of a new DARIAH-EU Working Group on PID Policies, where the needs of data producers, the technical constraints of PID service providers, and the usage scenarios of researchers are going to be discussed.

Quality assurance

Author: Rainer Stotzka (KIT)

Quality assurance in DARIAH is concerned with the quality of services offered to scholars in the academic disciplines. All basic services provided by the computing centers are monitored to guarantee continuous access. Higher level software services produced by the partners or by external groups are already closely guided within the development process.

In DARIAH-DE a "Service Life Cycle" state diagram has been developed in which the development steps as well as the transition conditions to the next steps are clearly defined. A team of scientific and technological mentors escorts the service and helps with advice to assure reasonable exploitation of existing DARIAH services and to check the usability and value of the service for the humanities scholars.
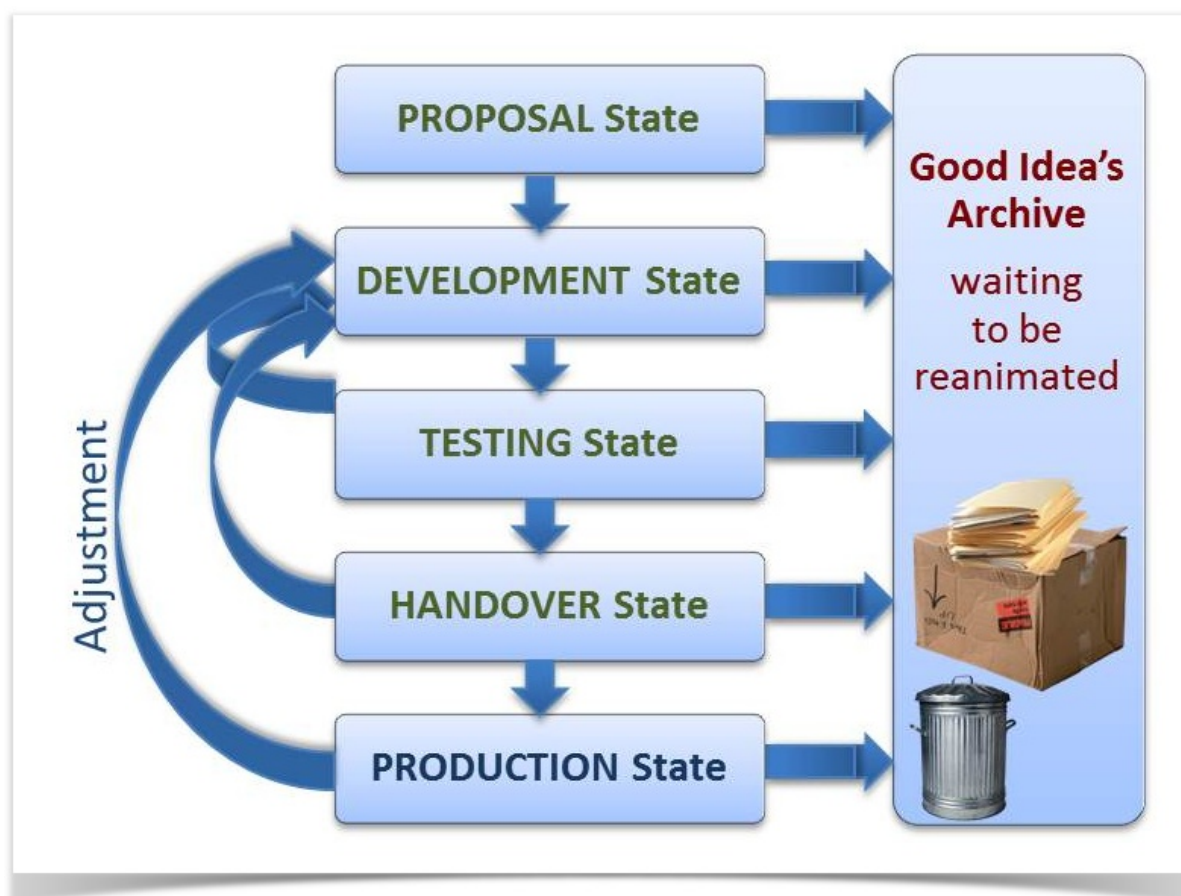

Figure 15 – The DARIAH Service Life Cycle

31

In order to ensure the serviceability of the technical basic services a criteria catalogue has been developed which deals with questions regarding documentation, security aspects, support, licence policies, verification, and migration of a service.

Furthermore evaluation criteria for transferring and integrating higher level software services into the DARIAH hosting environment have been defined. They cover general technical, security, network, and monitoring requirements.

Regarding the provision of services by DeISU another criteria catalogue has been developed. E.g. any service-provider must provide service-level-agreements, usage limitations, provision costs, operation costs, dependencies, support costs, etc. for each service to be supplied.

Associated and external research projects are supported by a team of computer scientists and humanities and cultural scholars in the integration of tools and services in the DARIAH-DE research infrastructure.

Security

Author: Johannes Reetz (RZG)

The term security refers to confidentiality, integrity and availability of data, systems and services provided by or related to DARIAH-DE. Therefore, also critical functions, such as access controls and backup procedures, of sites participating in DARIAH-DE can be affected. Security is about how data, services and systems provided by DARIAH-DE are protected against risks according to best information security practices in accordance with national and EU laws. Security related assumptions of the DARIAH-DE service providers, users and other stakeholders are and will have to be agreed on common efforts to secure services and data related to DARIAH-DE during the course of the construction phase.

The computing centres in DARIAH-DE are engaged in national and international collaborations and e-infrastructures, and are operated in compliance to the best practices and guidelines for information security from sources such as the Trust Framework for Security Collaboration among Infrastructures (SCI), the security policy of EUDAT, policies of the European Grid Initiative or PRACE. Each of these centres has a security officer assigned who is responsible for incident coordination, risk assessments, continuous operational security and for developing and adapting security controls and guidelines at his site.

As one security-related task, the contact addresses of these site security officers which are of the form dariah-security@<site domain name> have been published on the DARIAH-DE wiki, together with further contact information which allow to direct any security notification to the computing centres in case of security incidents or discovered vulnerabilities.

The site security officers are responsible for ensuring secure systems configurations, vulnerability handling, incident handling, anomaly detection, monitoring and logging, providing security information within and between sites. This comprises also a secure authentication of machines and services using ssl-certificates from widely accepted Certification Authorities (DARIAH-DE uses ssl certificates of the DFN PKI).

As another security-related task of the computing centres, the common terms of use (ToU) have been specified during the reporting period. The ToU applies to all users of services provided by DARIAH-DE, and which makes them understand their security and privacy related rights and obligations.

Spaces of Jewish History (Orte jüdischer Geschichte)

Author: Harald Lordick (STI)

The App, especially tailored for mobile devices, opens up access to location-based, free data collections for German – Jewish history and groups them suitable based on geolocation data (GPS or user selection). First, only developed and used as a private tool, the web app now benefits from the hosting in DARIAH-DE and from the Geoservice offered as a REST API (TGN), which can easily be integrated in any application and in this case provides a core functionality of the application. Additionally the DARIAH-DE expertise on licensing turned out to be very helpful.

„As coding humanist I know the hardware, server operating system and security issues in best hands and I'm excited about the synergies of generic components. Thus I can focus more on the concept and the integration of data"

More Information at https://de.dariah.eu/web/guest/web-app-orte-juedischer-geschichte
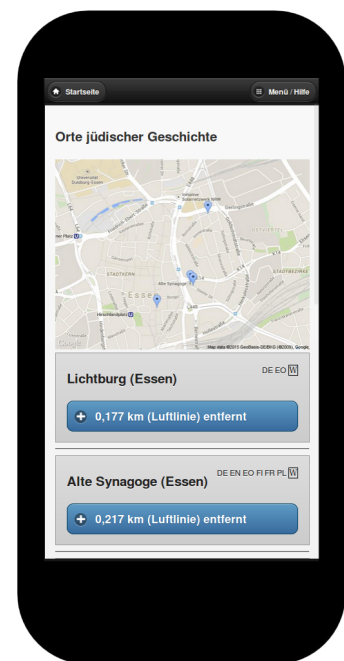
Storage Architecture

Authors: Rainer Stotzka (KIT), Danah Tonne (KIT), Johannes Reetz (RZG)

Scope

The internal DARIAH Storage Infrastructure establishes an *open, federated, distributed and dynamic* infrastructure services to ensure data preservation and interoperability across DARIAH. The DARIAH Storage Service provides a "low-barrier" entrance for new scientific disciplines and infrastructure providers to build-up scientific data networks:

- Support the creation and enhancement of data infrastructures for the management of research data
- Establish common (standardized) storage and archive interfaces to guarantee technological sustainability for high level services

Integration

The resulting DARIAH Storage Service and it's API are used as an interface between DARIAH High Level Services, e.g. the DARIAH Repository acting as the DARIAH Archive, and world-wide distributed data centers (see Figure 13).
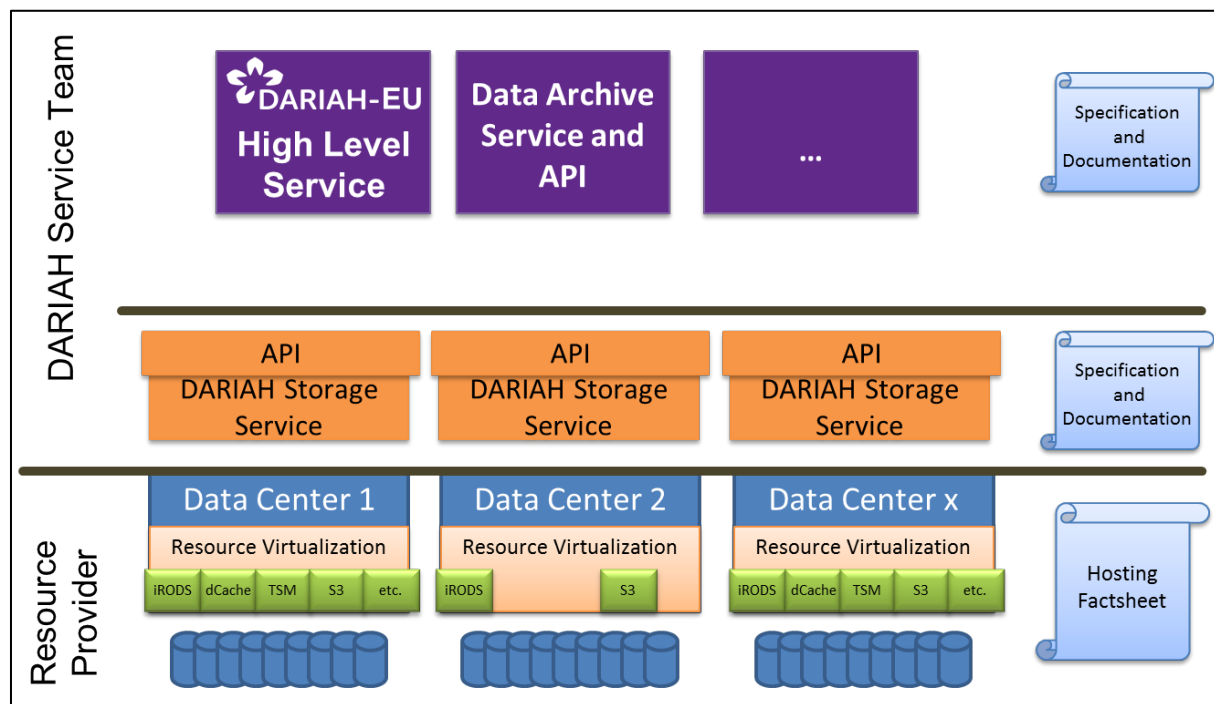
Figure 16 – Integration

At the moment several projects (February 2015: 12 projects) coming from different disciplines and different fundings use the DARIAH-DE storage service, e.g. the BMBF-project 3D-Joins, the BMBF-project Relationen im Raum, the academy-project Textdatenbank und Wörterbuch des klassischen Maya[11] etc.

Data Life Cycle

By reviewing similarities and differences of several research data cycles, like

• DDI Structural Reform Group. "DDI Version 3.0 Conceptual Model." (DDI Alliance. 2004. Accessed on 11 August 2008)[12], adapted by Daniel Kurzawe and Rainer Stotzka, 13. Feb. 2013 (see last TAB report)

• Boonstra, Breure und Doorn (2004),[13] page 22.

• Unsworth (2000),[14] a reference model with all essential functionalities can be designed.

The gaps between the different approaches had been reconciled in an approximation process. The following gaps have been identified:

• the need of a defintion of the term "research data"

• definition of the filetypes / objectclasses, which should be considered in an implementation of the Research Data LifeCycle

---

[11]http://www.iae.uni-bonn.de/forschung/forschungsprojekte/laufende-projekte/idiom-dictionary-of-classic-mayan

[12] see http://www.icpsr.umich.edu/DDI/committee-info/Concept-Model-WD.pdf

[13] see http://www.dans.knaw.nl/sites/default/files/file/publicaties/Past-present.pdf

[14] see http://people.brandeis.edu/~unsworth/Kings.5-00/primitives.html

- definition of the features behind the concepts of "curation" and "content preservation"
- definition of the term "annotation"
- a first LifeCycle definition

Since October 2014 the draft of a reference model of a Research Data LifeCycle for Digital Humanities is available. An overview of the order of activities within a researchproject has been developed:
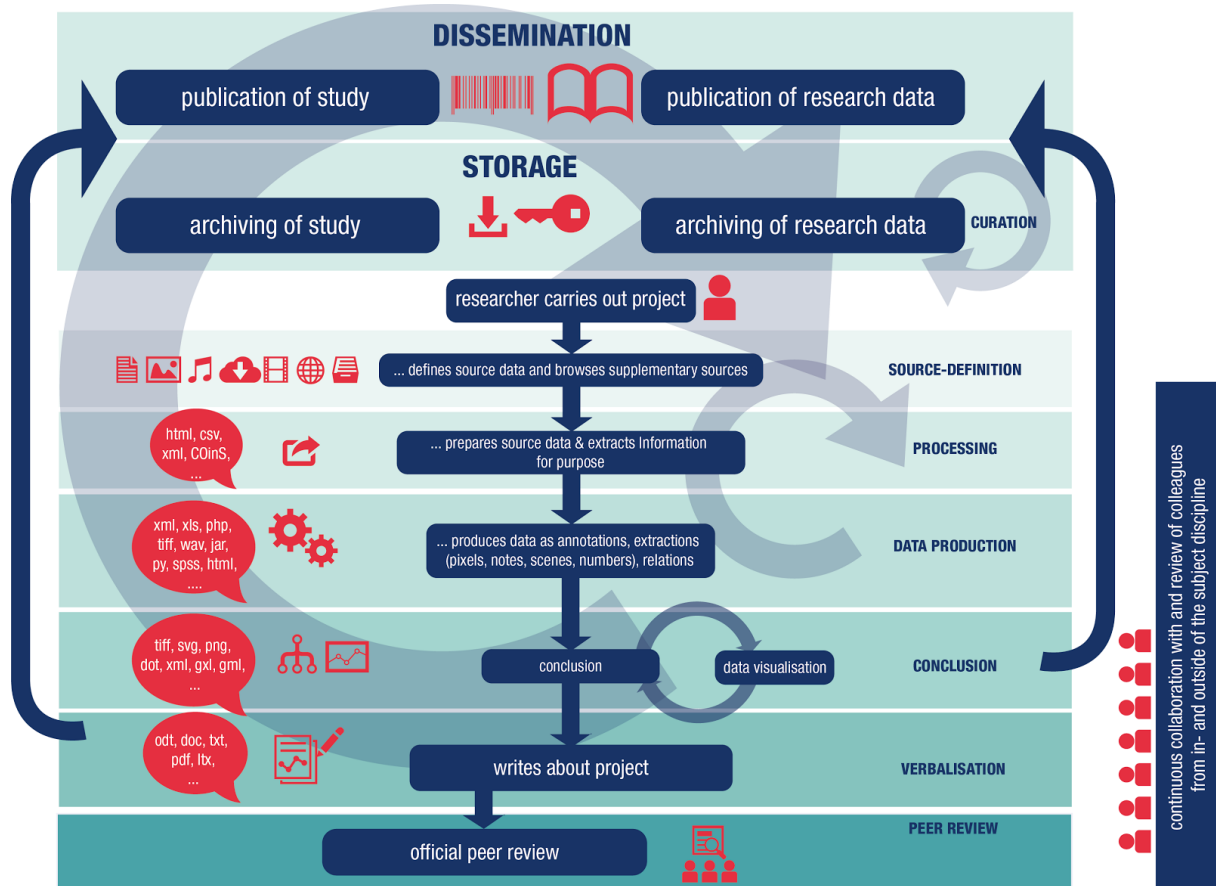

Figure 17 –Preliminary result of AG Research Data LifeCycle in DARIAH-DE
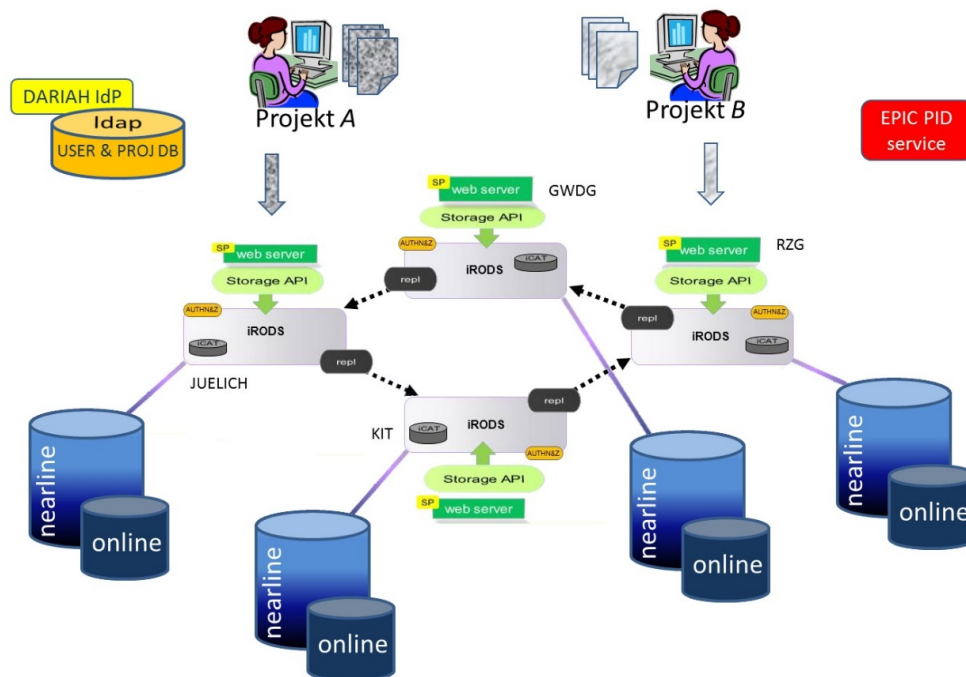
35

Storage Federation Architecture

Figure 18 – Storage Federation Architecture

The DARIAH Storage Federation consists of four federated iRODS zones provided by the computing centre partners. Within the federation a hard wired replication mechanism ensures safety of injected data.

Currently the federation is undergoing major enhancements:

- ability to pass metadata from one partner to another, which allows access failover and/or the possibility to access archived data at any partner
- flexible replication options
- easy integration of a new or retirement of an old federation partner
- introduction of accounting
- improved and enhanced access rights management

Terms of Use

Author: Ulrich Schwardmann (GWDG)

DARIAH has the claim to open its services to other projects and communities of the digital humanities. This implies that it is necessary to integrate groups of users, which go far beyond the institutions of the DARIAH consortium itself. Therefore it is necessary to have a reliable legal relation between the users of DARIAH services and the service providers with their partially different existing terms of use, which fulfills a sufficient set of their requirements.

The service centers agreed therefore to a common Terms of Use, that has to be signed by each user of the DARIAH services. Before the DARIAH has become a legal entity (ERIC) we agreed in DARIAH to set up only one contract between the users and only one of the

provider partners, that included a hint, that this has to be fulfilled also at use of any of the services provided by the consortia partners of DARIAH, to avoid the overhead for users to sign a contract with each service provider. Since the DARIAH ERIC is in place the Terms of Use contract is now agreed directly between DARIAH and the users.

The ToU attribute that contains all DARIAH-ToU-Versions (by filename), accepted by the user, is integrated into the next release of the WebGUI of the AAI and will be productive presumably already in Febraury 2015.

Until then this contract is agreed by the users at an IBM-Lotus system of one of the partners, which provides a database of the valid contracts to the service providers. Together with the introduction of the new WebGUI all ToU data from the Lotus System together with all according and available ToU data from TextGrid will be migrated to the ToU attributes in the user records of the new system.[15]

User Support

A survey from DARIAH-DE Cluster 3 lasting from July to October 2014 revealed that user support is of central importance for the infrastructure users. To meet this requirement and to support the special needs of DARIAH-DE as an Organization a Helpdesk System (OTRS) was selected and integrated in the DARIAH-DE infrastructure. It is currently in testing state in which it is evaluated. It is going to production state within the upcoming two months.

All kinds of requests are centrally gathered by the Helpdesk System – questions about services, teaching material, courses, standard recommendations, cooperation inquiries, bug reports, etc. As input channels exist web-forms, tools (for example the Textgrid Lab), E-Mails and the web-frontend, which additionally can be used by users to view their conversation history with help desk system.

The incoming requests are then assigned to predefined topics (queues) for which experts of DARIAH-DE are responsible for. Most requests get automatically assigned to their respective topics (queues) by the system. The assignment is done based on the incoming channel (recipient email address, web-form, …), text patterns ("Virtual Machine" is in the subject of the message) and other parameters. Requests which cannot be assigned automatically – for example, because they cover not anticipated topics – are answered directly by a small DARIAH-DE support-team or assigned to an expert manually.

Figure 19 – Ticket processing in the DARIAH Helpdesk System

Additionally, internal processes of the DARIAH-DE Organization can be modeled, controlled and are archived for later reference. Needed workflows and processes are

---

[15] Please also note the appendix which contains the current version of the ToU of August 2014.

already recorded and get iteratively implemented after the system is in production state. The Helpdesk System is already fully integrated in the DARIAH-DE AAI and therefore all requests can be mapped to DARIAH users, which is important for the internal process mapping.

Conclusion

After the first project phase, which was marked by initiating and developing the infrastructure itself, DARIAH-DE has produced a good basis for the upcoming efforts and is increasingly perceived as an established DH research environment. Today more than 2.300 researchers from DARIAH and over 50 affiliated projects use DARIAH infrastructure components and services. Even more are already profiting from the materials, tools and services, which don't need a DARIAH account to use. 13 more research projects, which plan to use the DARIAH infrastructure, are currently in the assessment phase.

The different infrastructure components are incrementally professionalized and the number of services considered in production state increases steadily. Additionally DARIAH works on its internal processes and environment to support the shift to an institutionalized organization. All these efforts are based on user studies and surveys among the DARIAH users and DH scholars.

The number of users, which has rapidly grown in recent months, the diverse requests, and the enormous need for advising and support services from various DH projects – from thematic and methodological in addition to disciplinary perspectives – demonstrates that digital research infrastructures for the humanities can only be effectively established in the context of close interaction and connections between the areas of instruction, research, research data, and basis infrastructures, and that they can only be successful when the entire spectrum of arts and humanities disciplines are involved. The DARIAH-DE consortium and its partners look forward to supporting these activities.

Appendixes

Terms of Use

Declaration of Consent to the Terms of Use

in the Context of Dariah-DE

The use of the software, the resources and the infrastructure of all providers cooperating in the context of the project DARIAH-DE takes place at the user's own risk and responsibility. The realization of economic profit from the use of these resources is prohibited.

The laws of the Federal Republic of Germany apply to the use of this infrastructure, software, and associated resources. You commit yourself, that you will not use this infrastructure to distribute content of an insulting, vulgar, obscene, pornographic or libelous nature, or content glorifying violence, in any form, particularly also not by abandonment of possibilities to others to use this infrastructure. In addition, the conscious or intentional introduction or spreading of viruses or other malicious or harmful programs is strictly prohibited. You concede to the operators of this infrastructure, software and associated resources the right to relocate and/or remove content at their discretion. In particular, user accounts which have been created, including stored data, may be deleted after the termination date of a project, unless arrangements have been agreed for a continuation of service. You are not permitted to use your access to resources and services to obtain personal data in the sense of the Federal Data Protection Act (Bundesdatenschutzgesetz) or to store or process personal data obtained by other means. You are not permitted to copy copyrighted data or programs from one computer system to another. The use of illegally obtained copies of copyrighted data or programs on the resources of the providers cooperating in the context of the project DARIAH-DE is strictly prohibited.

Breach, violation or infringement of these rules will result in immediate and permanent account closure. The operators of the infrastructure and associated resources reserve the right to take legal action and pass on network traffic data or similar data to the law enforcement agencies.

You hereby consent to the automated storage, processing and transfer of the personal data you provide at the time of registration and during operation to the institutions or other entities, whose resources are being used, as far as this is necessary to ensure proper operation of the resources. This information will be used for no other purpose.

The providers of resources cooperating in the context of the project DARIAH-DE only assume responsibility for damage caused by hacker attacks or computer viruses if such damage resulted from intent or gross negligence on their own part.

By registering, you agree to these Terms of Use. If you allow other persons to use ressources under these Terms of Use, you will also take the responsiblity that these other persons observe the Terms of Use you agreed to.

Figure 20 – DARIAH-DE Terms of Use

Abbreviations

AAI	Authentication and Authorization Infrastructure
BBAW	Berlin-Brandenburgische Akademie der Wissenschaften
BMBF	Bundesministerium für Bildung und Forschung (Federal Ministry of Education and Research)
CMS	Content Management System
DAASI	DAASI International
DAI	Deutsches Archäologisches Institut
DARIAH	Digital Research Infrastructure for the Arts and Humanities
DARIAH-DE	Digital Research Infrastructure for the Arts and Humanities - Deutschland
DARIAH-EU	Digital Research Infrastructure for the Arts and Humanities – European Union
DBS	Database System
DeISU	DARIAH-DE eHumanities Infrastructure Service Unit
DFN	Deutsches Forschungsnetz (German for "German Research Network")
DH	Digital Humanities
DHd	Verband Digitale Geisteswissenschaften im deutschsprachigen Raum
DT/PB	Universität Detmold/Paderborn – Musikwissenschaftliches Seminar
GCDH	Göttingen Centre for Digital Humanities
GPS	Global Positioning System
GWDG	Gesellschaft für wissenschaftliche Datenverarbeitung mbH Göttingen
HAB	Herzog August Bibliothek Wolfenbüttel
HKI	Universität zu Köln – Historisch-Kulturwissenschaftliche Informationsverarbeitung
IdP	Identity Provider
IEG	Leibniz-Institut für Europäische Geschichte
JSC	Jülich Supercomputing Centre
KIT	Karlsruher Institut für Technologie
LDAP	Lightweight Directory Access Protocol
MInf-BA	Otto-Friedrich-Universität Bamberg – Fakultät für Wirtschaftsinformatik und Angewandte Informatik

MPIWG	Max-Planck-Institut für Wissenschaftsgeschichte
MPDL	Max Planck Digital Library
NGO	Non-governmental Organization
NP	National Representative
NRPE	Nagios Remote Plugin Executor
NSCA	Nagios Service Check Acceptor
OA	Organization Admin
OAI-PMH	Open Archives Initiative Protocol for Metadata Harvesting
OKF	Open Knowledge Foundation e.V. Germany
OSM	Open Street Maps
PID	Persistent Identifier
PKI	Public-Key-Infrastructure
RiR	Relationen im Raum (German for Relations in Space)
RZG	Rechenzentrum Garching der Max-Planck-Gesellschaft
SCI	Trust Framework for Security Collaboration among Infrastructures
SKOS	Simple Knowledge Organization System
SP	Service Provider
STI	Salomon Ludwig Steinheim Institut für deutsch-jüdische Studien
SUB	Niedersächsische Staats- und Universitätsbibliothek Göttingen
TAB	CLARIN-D / DARIAH-DE Technical Advisory Board
TaDiRAH	Taxonomy of Digital Research Activities in the Humanities
TGN	Getty Thesaurus of Geographic Names
ToU	Terms of Use
TUD	Technische Universität Darmstadt – Interdisziplinäre Arbeitsgruppe Digital humanities (Germanistische Computerphilologie / Philosophie / Ubiquitous Knowledge Processing)
URI	Uniform Resource Identifier
UWÜ	Universität Würzburg – Institut für deutsche Philologie
VCC	Virtual Competency Center