Collection Scope and Coverage Policy

DARIAH 2010

Table of Contents

DARIAH
Digital Research Infrastructure
for the Arts and Humanities

Executive Summary

At the core DARIAH services will be the data and resources which researchers require in order to undertake research. Enabling access for researchers will require building relationships with organizations responsible for the objects of digital research in the arts and humanities, as such DARIAH will aim to work with repositories, collections, and institutions to improve discovery, access and use of the resources which are most important to researchers within the digital humanities.

This policy is primarily aimed at organizations that manage collections of digital resources with the potential to be deployed through the DARIAH infrastructure. It aims to define the types of interaction which DARIAH will foster with organizations which hold and provide access to digital data and resources. This includes a wide range of institutions, including universities, national data services, archives, museums, and libraries, as well as collections created by research projects. The policy sets out DARIAH's expectations for developing relationships and interactions with these organizations, and aims to facilitate the building and maintaining of relationships between the DARIAH partnership and data providers with regard to the exposure of digital resources and the services provided.

Introduction

DARIAH is an infrastructure to support digital research in the arts and humanities. Enhancing the possibilities available to researchers working with digital tools and resources and improving the outcomes of that research is DARIAH's *raison d'être* . All aspects of the infrastructure and the services it provides to the user including will work toward this goal. The collective intelligence will provide spaces and opportunities to develop knowledge about the use of digital tools and methods, the technical architecture will provide an e-infrastructure upon which research activities can be built and the data curation activities will be the basis of providing the researcher with the content they require.

At the core of all of these services will be the data and resources which DARIAH researchers require. DARIAH will provide researchers with access to the data and resources required to do advanced research. Although repository and collecting services are undertaken by some DARIAH partners, holding and managing content will not be a primary activity of DARIAH.[1]

---

[1] The types of services offered by these DARIAH organizations can be found on their websites and in the DARIAH policies 'Data Ingest, Management and Preservation' and 'Compliance as a Trusted Digital Repository'

The breadth of potential sources for arts and humanities research means that many of the digital resources which DARIAH will make available are not held by DARIAH partners. Enabling access and enhancing resources for researchers will require building relationships with those organizations that are responsible for resources which are the objects of digital research in the arts and humanities. DARIAH will work with repositories, collections, and institutions to improve discovery, access and use of the resources which are most important to researchers within the digital humanities.

This policy sets out the parameters for interaction between organizations with which DARIAH will work in order to allow users access to digital data and resources. This includes a wide range of institutions, including universities hosting institutional repositories, national data services, archives, museums, and libraries, as well as collections created by research projects. The policy explains how DARIAH will develop relationships and interactions with these organizations. It will set out the mutual responsibilities of the DARIAH partnership and the data provider with regard to the exposure of digital resources and the services provided.

Audience

This policy is primarily aimed at organizations that manage collections of digital resources with the potential to be deployed through the DARIAH infrastructure. This could include national and institutional research data repositories, arts and humanities research institutes, research libraries and memory institutions. It is also intended for the benefit of large-scale projects which involve a number of organizations that maintain and provide access to content. These varied institutions, projects and consortia will make up the providers of data and resources for DARIAH users and the primary organizations that will interact with DARIAH.[2]

Aims and Scope

This policy aims to ensure the DARIAH user community has access to the high quality resources and data necessary to perform advanced research tasks. The real long-term research value of digital data and resources resides in access and interoperability. The

---

[2] Because of the nature of humanities research there is often an overlap between the roles of 'researcher' and 'data provider'. DARIAH recognizes that this is a significant consideration and will aim to work with both researchers (mainly through project liaisons) and organizations.

researcher requires data to be available and to be linkable to other relevant materials. To promote access and interoperability of resources DARIAH will build formal relationships with organizations providing data. The policy sets out the areas and parameters of interaction between organizations that manage collections and the DARIAH consortium. It will guide this process by setting out which collections and organizations the policy is applicable to, explaining what DARIAH is looking for from these collaborations, and how DARIAH will expose the resources to its users. It also sets out the types of organizations which DARIAH will aim to work with, and the parameters of that interaction. It will set out the requirements which DARIAH will expect collections to meet, and the responsibilities which DARIAH undertakes with regard to those collections. In addition it provides guidance on DARIAH priorities with regard to collections, resources and large-scale projects. This will provide the basis for relationships between DARIAH and the data providers.

Collections and collaborations

The diversity and dispersion of resources for the study of the humanities is a significant challenge for DARIAH. The DARIAH infrastructure will support digital research across the arts and humanities, being only limited by the needs of its users. This requires an approach which is open to the range of types of resources and the ability to work with many different types of data providers. As such DARIAH will be open to resources held in digital repositories, but also to resources which exist outside of repositories. Museums, libraries and archives are just the most obvious examples of organizations whose data will be potentially available through DARIAH services, making it possible for users to use the resources available in combination with their own.

This section of the policy will provide a detailed explication of the types of organizations and collections with which the DARIAH infrastructure with undertake to collaborate. This includes a broad range of types of resources held by several different types of organizations. Research within the arts and humanities is as varied as the possibility of human endeavour. The three main kinds of data collections and categories (with examples of kinds of collections which might be included) that DARIAH will support the use of are:

- **Data produced by researchers** - primarily project based, this could be either open or only available to a selected group of users. Data of this type could be specialist digitized material produced to answer a specific research question, intermediate outcomes, and conclusions or publications.

- **Collections and material within the Higher Education sector** – This is research data and collections of digitized resources held in repositories and elsewhere. Increasingly institutional repositories within universities and research institutes are the caretakers of this type of material. This category also covers national data services which curate and provide access to research data and (e.g. Data Archiving and Networked Services in the Netherlands and the Swedish National Data Service). In addition there are many relevant collections which are not held by traditional repositories, which have the potential to be resources for humanists. There is potentially overlap between this category and the previous one, as data produced by researchers is often curated by institutional or national data repositories.

- **Cultural heritage collections** – It will be important to expose digital data and resources which are held by major Cultural Heritage institutions because these make up a large part of the potential research archive for researchers working within these disciplines. Much of the material of interest to humanists and arts scholars is held in a diverse range of research and memory institutions, including research libraries, museums and archives. Large amounts of digitization work are going on within these institutions internationally and DARIAH will aim to work with these organizations to provide access to these collections to academic researchers.

This is not an exhaustive list and should not be seen as prescriptive of all areas that the consortium will work with. If data and resources are of relevance to DARIAH researchers are held by other types of institutions the consortium will aim to provide the means to use these where feasible.

Enhancing Resources

It is the aim of the services DARIAH provides to allow researchers to rethink the possibilities of humanities research. The focus of DARIAH will be on enhancing data and resource driven transformative research which uses digital tools and methods. It is possible that such transformation could occur within fields across the arts and humanities, and any collection which provides resources and data relevant to arts and humanities research could be a potential DARIAH collection. In practice DARIAH will focus on the areas in which transformative research is taking place using digital tools and methods, or in which the consortium has identified a strong potential for discipline changing research.

The huge range of potential organizations from which DARIAH users may want to access data, means priorities will be need to be set, including disciplinary, geographical and chronological scope. These priorities will be updated regularly in line with the development of new collections and areas of digital engagement, but should never preclude DARIAH working with providers of resources which are of use to its community at any point, even if they are outside the identified priorities. Remaining aware of the needs of the researchers and user community will be a primary responsibility of the consortium and DARIAH will maintain a close watch on developments within the fields its users are engaged in.

Focus

From the beginning of the construction phase DARIAH has identified disciplinary and methodological areas on which it will focus its initial efforts. These areas have been identified as major fields of activity which will most benefit from the support of the infrastructural services provided by DARIAH. Based upon current areas of activity in the digital arts and humanities the following parameters have been identified. DARIAH aims to work in all of the disciplines of the arts and humanities in which researchers are utilizing these collections to do digital research. Specific fields and communities of practice with established levels of activity (that is, cooperating at the European level) include: history, textual and manuscript studies, classics and archaeology.

Disciplines

DARIAH will also aim to further identify disciplines and communities emerging on the European stage and seek to collaborate with them to add value to their activities. Examples of fields in this category include philosophy and practice-based research in performance and visual arts. It will be a crucial aspect of DARIAH's activities that it supports digital research and fosters best practice in the use of digital tools and methods in nascent fields which may not have the infrastructural support of some of the more well established digital humanities communities. Other disciplines will be identified and added to these lists as the DARIAH infrastructure matures and builds relationships with collections.

Methods

In addition to disciplinary foci, DARIAH aims to support interdisciplinary research and foster work in the digital humanities which has the potential for crossing disciplinary and methodological boundaries. Methodological communities of practice tend to cross disciplinary boundaries, so DARIAH will engage with researchers working with methodologies in the areas most likely to produce transformative results. Examples of these areas include geo-temporal methods, 3D visualization, and text mining. Currently the DARIAH partnership is best placed to support work in the digital humanities in the following

areas: data sharing and re-use, text analysis, interface development, and searching and linking of data.

As with the disciplinary focus, here too, DARIAH will aim to work with researchers to assist in the development of their digital humanities methodologies. It is a key aim of the infrastructure to build capacity in the knowledge and use of methodologies for digital data driven research.

Chronology

There are no chronological limits to the collections DARIAH will work with. In order to accommodate the range of disciplines with a potential chronological focus, including literature, archaeology, history, and classics, it is necessary that DARIAH be open to collections with broad chronological range. This will require liaising with organization and projects which have expertise in specific issues to do with primary source material from a range of eras.

Capacity Building

One of the key facets of DARIAH's activities is building capacity for research with digital content. An important aspect of this is fostering international partnerships to develop access to and use of that content. Partnerships between the providers of the DARIAH infrastructure and data providers are one of the important ways in which DARIAH aims accomplish this. DARIAH is most interested in collaborating as partners with large scale consortia of archives and data providers which have a thematic coherence. Through these types of collaborations DARIAH will help to build capacity for research internationally and within communities of research practice.

A key means for doing this will be to support and foster the creation of critical masses of digitized material. In many cases this process will require DARIAH to facilitate the bringing together of material from a number of different providers so that researchers can discover, access and perform research task on these materials. Building thematic collections by encouraging large scale projects which bring together international partners from academic and memory institutions. These large scale projects to build interoperability between collections with a thematic relationship to each other are crucial for creating the necessary volume of material to enable high level research tasks to take place.

Interactions and Responsibilities

DARIAH will engage with the cutting-edge of digital research in the arts and humanities. This means interacting with diverse and dispersed data and allowing researchers to bring resources together in innovative ways to answer new questions. The digital arts and humanities researchers of the coming decades will want access to and meaningful deep linking of data from a variety of sources. DARIAH must enhance this type of digital research by interacting with researchers through the centres and projects in which they participate.

In order to make these collaborations possible DARIAH will aim to foster the building of strong and lasting relationships among a triangular set of stakeholders: data providers in the academic and cultural heritage sectors, researchers and research groups with a digital humanities focus, and members of the DARIAH consortium. In building these relationships DARIAH will enhance the value of these collections for researchers and educators. All collections to which DARIAH will provide access must have a clearly defined scope. This scope may be broad or narrow, and should fall – in a broad sense – within the disciplines which are part of the focus of DARIAH's efforts.

Defining Collection Scope

Having a well defined scope is necessary for a collection to interact with DARIAH because without a clear policy defining the types of data and the disciplinary or thematic coverage of the collection, DARIAH will not have a basis for working with the collection. It is crucial that any data provider who will provide access to collections through the DARIAH infrastructural services has a clear policy on the scope and coverage of their collection. Collection scope policies should include the following information:

- **The collecting remit of the organization**: providing information about the range of types of data and the resources which a user could expect to find and which the organization is responsible for.
- **User community information**: DARIAH will need to have a clear idea about who the users are of collections and resources supplied by the organization
- **Future collecting plans**: giving particulars of how collection will be developed and what resources, if any, the organization anticipates will be added to the collection.

In the case of new collections the data provider DARIAH members will aim to liaise with the organization from the outset to support the use of best practice in all areas. This will benefit both parties by ensuring that the organization develops in such a way as to encourage the use of its resources. This early contact will enhance the discovery and interoperability of the data

which is to be shared through DARIAH, through the promotion of standards in data creation, metadata, and preservation methodologies.

Metadata

With the vast increase of the landscape of metadata standards, and the continuing changes in this area, DARIAH does not specify which types of metadata are acceptable. DARIAH will aim to make it possible to utilize and link to metadata in any accepted standard format. However, collections to be exposed through DARIAH should aim to provide metadata in a single accepted standard. This will make it possible for the DARIAH technical team to develop the necessary services to allow the data within the collection to be interoperable with other collections.

Catalogues and Registries

DARIAH will undertake a number of important activities to promote the use of the collections and resources with which it has a relationship. These will be aimed at the goals of fostering wider use of the materials, promoting best practice and creation, curation and use of the data, and building a community of users.

One of the primary means for exposing and fostering the use of data will be the publication of a catalogue of the collections and resources which are available for use through the services and interfaces it provides. This catalogue will be maintained by members of the DARIAH consortium to enable researchers to understand the scope of collections which are exposed through DARIAH. This catalogue will utilize accepted metadata standards to provide users with information about the type of resource, the content of the collection, and the restrictions on the use of the resource. Registries will allow machine readability of the information while an online knowledge base will make the information available to DARIAH users and others across the digital humanities and the arts and humanities more widely.

Once the information about the materials which are available though DARIAH has been exposed DARIAH will provide gateways to those data and resources which will allow combination and recombination of the data which is available. In doing so the DARIAH consortium will build a 'collective intelligence' to provide researchers with information on best practices, tools, methods and potential areas of collaboration.

In addition to providing information about available collections DARIAH will foster the development of collections by providing trusted data curation for owners and managers of digital assets. These activities will foster interoperability and preservation from the earliest stages of the research lifecycle by providing organization and technical means for the ingest, management and preservation of the digital assets.