



DARIAH Newsletter

Digital Research Infrastructure
for the Arts and Humanities

Contents

- Looking back on the preparatory phase of DARIAH
- Establishing a stable e-Infrastructure for the humanities: a community endeavour
- DARIAH-EU Coordination Office
- Göttingen State and University Library sets kick-off date for DARIAH-DE
- Two countries, one voice for the humanities: France and Germany
- TGE ADONIS launches ISIDORE research platform
- Digital humanities in Denmark
- Research news from Croatia
- Collective Intelligence Workshop report

This free newsletter is published quarterly. If you find it useful, please tell your colleagues and friends about it.

Looking back on the preparatory phase of DARIAH: What has changed over the past 2,5 years

By Peter Doorn
DARIAH Preparatory Phase Coordinator

If there is one constant over the past two and a half years of the preparatory phase of DARIAH, it is change. With hindsight, this is what we should have expected. When we started in the autumn of 2008, we did not yet know what the blueprint for DARIAH would resemble. I think that most of us had an idea of a central, European coordinating office, and in each member country a “national DARIAH” in a variety of forms.

I remember that the idea of Virtual Competency Centres (first six, later four) came up during a meeting of the “Paris

Group”¹ in London in the summer of 2009. Who mentioned it first I don’t remember, maybe it was really an example of “collective intelligence”, but if it was an individual, excuse me if my recollection is not clear who it was. What I do remember is that from the first moment everybody present found it to be a most promising idea, striking the right balance between centralisation and fragmentation.

Another change which occurred had to do with the scope of DARIAH. Should it have a data orientation, with a central archiving and access function or should it link up repositories? In the end, we chose a wider ambition of support for digital humanities research, in which e-Infrastructure, e-Research, education and scholarly content all have a place.

Of course, at the start the ideas about the technical setup were still rather theoretical. Over time, a reference architecture and conceptual model took

¹ The Paris Working Group was an *ad hoc* group formed by the Management Board during an earlier meeting in Paris... hence the name.

shape. I think what we need to keep in mind is that because of fast technological developments and rapidly changing user needs, we need to stay alert and ensure that the technological framework is sufficiently flexible and adaptable to new circumstances. For instance, the rising importance of linked data and semantic methodologies, or the increasing prevalence of mobile devices such as smart phones and tablets.

We also did not have a sense of what the legal and organisational framework for DARIAH should be. It soon became clear that the ERIC¹ (European Research Infrastructure Consortium) model was the best choice for legal structure, and the clear statutes DARIAH has produced based around the ERIC are a very good piece of work.

Remarkably, the rough total cost estimate for DARIAH is still about the same now as it was at the very beginning. This is one of the few things that has remained fairly stable over time. Of course, it is now based on real cost estimates and on discussions with funders. A new aspect, however, is the idea of having a separate budget for community-driven projects, and we can be glad that European Science Foundation was so positive at being involved.

In terms of personnel and internal organisation there were also considerable changes: several DARIAH partners changed because of national developments, several people assumed new positions in other institutes and even other countries. We can safely assume that this process will go on.

If I compare my experiences with DARIAH with those of my home institute DANS, I also feel that adaptability to

changing conditions is indeed very important. In six years time, DANS has gone through two strategic cycles in which we modified our mission and main goals since we started on the basis of a blueprint written in 2004. Because of the growth of the institute (from 10 to about 45 people in six years) we also needed constant revisions.

The world is changing. The humanities are changing. Our ideas about the world and the humanities are changing as well. DARIAH will grow and expand. Stagnation will be a sign of decline.

From this looking back I have a few recommendations for the future: be prepared for changes, and keep DARIAH flexible. Always keep in mind that an infrastructure supports things: it is not an aim in itself. It is there for humanities researchers to be used. DARIAH needs a service-oriented approach, and we need to give the user a central position.

Also keep in mind that there are many infrastructure initiatives, and it is likely that not all of them will survive independently in the long run; therefore, it is wise to link up with surrounding infrastructures: CLARIN, CESSDA, Europeana, OpenAire, Bamboo, etc.

Finally, thank you all very much for your contributions and for the collaboration. I have learned a lot from DARIAH, which was, as I expected from the start, one of the most complex projects in the humanities ever, at least for me. But I think we can say that the mission of the preparation phase was accomplished, that we have a dependable business plan plus the support from funders in enough countries to continue building DARIAH in the construction phase.

¹ <http://bit.ly/9H3XPD>

Establishing a stable e-Infrastructure for the humanities: a community endeavour

*By Laurent Romary
DARIAH Transition Phase Coordinator*

Let me first express my warmest thanks to Peter Doorn who has brought DARIAH through the preparatory phase to a stage where we can actually contemplate giving it a concrete embodiment as a European infrastructure for the humanities under the joint coordination of France and Germany. When a small group of colleagues¹ representing national endeavours met in January 2006 at the headquarters of the CNRS in Paris, we could only dream of joining our forces to offer better services and a more sustainable infrastructure to the scholars we served.

Over the subsequent years, we, together with our colleagues from across Europe who joined the initiative, managed to define a concept for what DARIAH could be. We supported this idea within the human and social sciences working group of ESFRI, and have achieved our goal of having DARIAH become officially supported by the European Union, not only through the funding of the preparatory phase, but also by accompanying us in our endeavour to put together concrete actions, such as the EHRI project, that would demonstrate our capacity as a digital infrastructure for the humanities.

Still, as Peter puts it, the world is changing, and we now need to see where DARIAH should act to become the efficient infrastructure we have all dreamt of. The main change I would like to stress here is indeed the move from a highly centralized consortium of partners to a network of competence centres, which are now to be coordinated both at a national and European level.



DARIAH Transition Phase Consortium – Berlin, February 2011.

¹ Sheila Anderson (then AHDS), Peter Doorn (DANS), Ralf Schimmer (Max Planck Society), and myself (then CNRS)

This move is accompanied by a change of governance for the DARIAH infrastructure, where countries are now taking the lead in providing the necessary forces that DARIAH will be made of, both from the point of view of the actual budget, but also the scientific and technical priorities that we will put on our agenda. The work in front of us is now to harmonize national contributions and to use the results of the preparatory phase as a reading guide to establish stable services for various communities in the humanities.



Laurent Romary

We have to deal here with a highly complex environment where national policies and the actual expectations of scholars will have to be brought together. We will definitely need to be very pragmatic in order to define concrete activities where our technical capacity, our joint expertise, and the actual understanding we have of the issues raised by digital humanists can be made concrete.

Clearly, quoting again Peter, we will have to articulate our work with the

other initiatives covering complementary domains such as publication repositories (OpenAire) or cultural heritage (Europeana). But above all, we need to remember that the technical and scholarly world around us will sometimes change more quickly than we, as a single infrastructure, will ever manage to anticipate. DARIAH must contribute to this change and be a facilitator for scholars themselves to define new methods to carry out their own research in the digital world.

DARIAH-EU Coordination Office (DCO)

As of March 1, 2011 the transition phase of DARIAH starts, leading into the construction phase of DARIAH by end of 2011.

From now on, all matters pertaining to the coordination and the management of DARIAH-EU should be addressed to the DARIAH-EU Coordination Office (DCO).

The DCO will be run by Germany and France. At present, the German office, located at the Göttingen Centre for Digital Humanities, is at your service. For any information and questions on DARIAH please contact:

- dariah-dco@dariah.eu

Christiane, Ele and Sally will be happy to offer support.

We also like to take this opportunity to thank Milena Piccoli for her services on behalf of DARIAH during the two-and-a-half-year preparatory phase. She performed the not always easy job of coordinating DARIAH with courage and success. Bringing together different cultures of communication is essential

for an infrastructure project on a European scale. Milena managed all critical situations brilliantly with the aid of her charm and a large portion of pragmatism.

Göttingen State and University Library sets kick-off date for construction phase of DARIAH-DE



The Göttingen State and University Library is pleased to announce a kick-off meeting on March 31st for all seventeen partners of the German DARIAH-DE consortium. The meeting, which will take place in Göttingen, marks the beginning of the construction phase for DARIAH-DE.

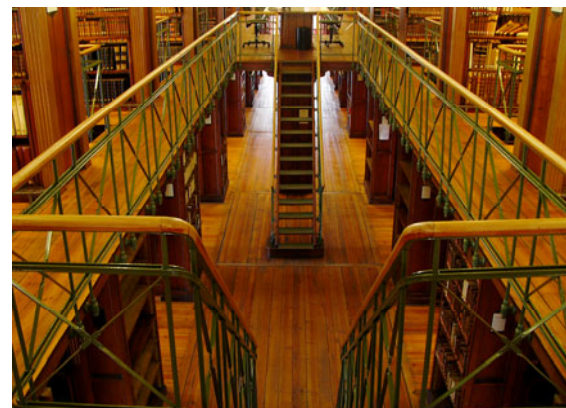
As joint project leader (together with the Max-Planck-Society Digital Library) of the Virtual Competency Centre for "e-Infrastructure" in the DARIAH-EU consortium, the Göttingen SUB and its partners are responsible for providing the technology for establishing DARIAH as a trusted intermediary for developing tools, standards and services which can be preserved, shared and integrated with the arts and humanities community.

Additional contributions by the German partners to DARIAH-EU include the development of digital humanities curricula, demonstrator projects and scholarly content.

The members of DARIAH-DE are drawn from a broad range of university institutes, centres and seminars as well

as industry with proven competence in computer science, technology and digital humanities. These include large, national institutions such as the Berlin-Brandenburg Academy of Sciences and Humanities or the Jülich Supercomputing Centre as well as much smaller groups such as the Salomon Ludwig Institute for German-Jewish History or the newly founded Göttingen Centre for Digital Humanities.

As a result, we are particularly looking forward to the unique opportunities for innovative collaboration between the humanities and cutting-edge technology which this consortium will offer its partners.



Göttingen State and University Library

Two countries, one voice for the humanities.

French and German research ministries meet to join efforts on establishing DARIAH

As the preparatory phase of DARIAH reaches its end, and the main organisational aspects, but above all the scholarly perspective, for establishing a research infrastructure in Europe for the humanities are being finalized, it was a major milestone for the project to meet

with representatives of the French and German Ministries of Research in Bonn on January 11th, 2011.

In a very friendly atmosphere, and in presence of acting parties from both countries (TGE Adonis and The Göttingen State and University Library), the ministries confirmed their willingness to jointly coordinate DARIAH, whose seat will be situated in France. Following a presentation of the main objectives and plans for the establishment of DARIAH as a European Research Infrastructure Consortium (ERIC), the two countries compared their own strategies in establishing a national roadmap for the humanities which in turn should contribute to providing expertise and services at the European level.



French and German research ministry representatives meet with DARIAH. From left to right: Jacques Dubusc (MESR), Dietrich Nelle (BMBF), Jean-Luc Pinol (DARIAH, TGE Adonis), Helge Kahler (BMBF), Heike Neuroth (DARIAH, UGOE), Laurent Romary (DARIAH, INRIA), Norbert Lossau (UGOE), Hans Nerlich (PT-DLR).

The importance of involving scholarly communities as well as of ensuring the re-use of technologies developed at a national level was given particular focus during discussions. The desire is that the experience of setting such roadmaps which reflect the national strategies in support of e-humanities become a seminal activity for DARIAH partners at

large. The meeting ended with the expression of the willingness to such bilateral exchanges be renewed in the context of a new meeting in the summer of 2011.

TGE ADONIS launches ISIDORE social science and humanities research platform



The DARIAH partner TGE ADONIS, a department of the French Centre National de la Recherche Scientifique (CNRS), announces the launch of the beta version of the ISIDORE research platform.

ISIDORE allows researchers to search across a wide range of data from the social sciences and humanities. ISIDORE harvests records, metadata and full text from databases, scientific and news websites that have chosen to use international standards for interoperability. Once harvested, this information is turned into an RDF structure and enriched with cross references (vocabulary lists, thesauri, references) produced either by the scientific community or by major higher education and research institutions. The information thereby harvested becomes points of entry to the full text, which is also indexed when possible.

ISIDORE is available via a dedicated website,¹ but it can also be integrated with thematic, disciplinary or academic

¹ <http://www.rechercheisidore.fr>

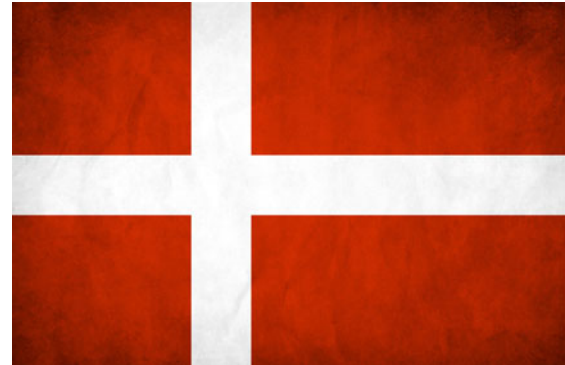
portals. At the moment, ISIDORE is also an RDF triple store for data providers. ISIDORE is therefore not a conventional search engine; it is a flexible research platform that aims to federate access to digital research data in the social sciences and to provide unified access for teachers, researchers, and PhD students.

The research platform ISIDORE differs from classical research engines in several characteristics:

- Harvesting specific scientific structured data compatible with open access and four connectors (OAI-PMH, RSS and Atom, RDFa);
- Indexing metadata and full texts in French using the AFS engine from Antidot.
- Standardizing metadata and enriching data by using thesauri and taxonomies provided by teams of scholars and monitored by the Mondeca ITM software.
- Offering API developed by Sword and Sparql endpoint access (W3C Standard) making it possible to embed ISIDORE in other software.

ISIDORE is based on the principles and methods of the web of data, a new extension of the web supported by W3C. ISIDORE enables data providers to re-expose their metadata enriched in accordance with web principles and particularly with open access to public data. In this way, ISIDORE contributes to disseminating French scientific data in the social sciences and humanities and offers the community a way to continuously enrich these data. In December 2010, for the opening of the beta version of ISIDORE, 905 168 documentary units from 745 sources organized in 35 collections were accessible and searchable. ISIDORE development is ongoing.

Denmark seeks to establish major centre for digital humanities research



Major educational and cultural heritage institutions in Denmark have come together to jointly propose the establishment of Digitalt Humaniora Laboratorium (DigHumLab), a laboratory of digital humanities research, development and support which will serve the entire nation. The laboratory will be designed to organise, coordinate and develop national research infrastructure in the humanities subject area, and will work closely with DARIAH through its construction and operational phases. It is expected that the Danish Agency for Science, Technology and Innovation will fund DigHumLab for 5 years, beginning in 2011, at approximately DKK 6,000,000 (€800,000) per year.

Denmark is poised to act as co-chair, with Ireland, of one of DARIAH's Virtual Competency Centres, and DigHumLab will form the organisational backbone of this leadership. As a national entity with an international perspective, DigHumLab will engage with local and EU-wide projects with the goal of enhancing research opportunities and improving data quality and access. DigHumLab will eventually act as the single Danish point of entry to all digitised resources relevant to the

humanities in both Danish and European research infrastructures. Researchers and students will use DigHumLab services to keep abreast of and contribute to national and international developments in the creation and storage, use, copyright and ethical issues related to humanities digital data of any kind. The laboratory will develop new research tools and methods, and inspire new research methods and issues.

DigHumLab will provide researchers and students the ability to search, analyse, organize, link, compare, share and combine resources with analysis tools which are under continuous development, according to researchers' needs. The types of resources that will be included DigHumLab will include text, audio, images, physical objects (in 3D rendering), and combinations thereof.

Partners committed to Digitalt Humaniora Laboratorium include:

- Faculty of Humanities, University of Copenhagen
- Faculty of Humanities, Aalborg University
- Faculty of Humanities, Aarhus University,
- Faculty of Humanities, University of Southern Denmark
- Roskilde University
- Royal School of Library and Information Science
- Kolding Design School
- National Library
- Royal Library
- National Museum
- Society for Danish Language and Literature
- Danish Language Council
- School of Conservation

Croatia signs DARIAH MoU, develops new national research data strategy



On December 23rd 2010, the Croatian Ministry of Science, Education and Sports signed the DARIAH Memorandum of Understanding agreement, declaring Croatia's commitment, in principle, to contribute towards the construction of DARIAH in the years ahead, leading ultimately towards a distributed European research infrastructure for the arts and humanities.

Croatia joins other MoU signatories, including Germany, France, the Netherlands, Greece, Ireland, Switzerland, Italy and Slovenia.

DARIAH's Croatian partner is the Centre for Informatics and Computing of the Ruder Boskovic Institute (RBI), one of the biggest and the most prolific institutes in Croatia. In collaboration with FFZG (Faculty of Philosophy, University of Zagreb), RBI recently organised a round table meeting titled *The storage system for digital data - the development of national strategies* to which the most eminent institutions, departments and universities in Croatia were invited. The aim of the round table was for all institutions to have an opportunity to exchange views and experiences, and to discuss the

establishment of a joint national strategy for the development of digital content storage in full compatibility with EU initiatives.

The meeting's agreements regarding future Croatian strategies and activities include recognition of the importance of:

- Promoting the vision of society based on knowledge.
- Synchronisation of terminology with FP7 applications.
- Inclusion of knowledge about the content in the archive.
- Overcoming the separation between science and culture.
- Including representatives of public authorities in national strategy creation.

RBI also recognized future partners in carrying out this strategy:

- ICLL (The Institute of Croatian Language and Linguistics)
- FFZG (Faculty of Philosophy, University of Zagreb)
- HAZU (Croatian Academy of Sciences and Arts)
- HIT (Croatian Institute of technology)
- HDA (Croatian State Archives)
- HRZ (Croatian Conservation Institute)
- NSK (National and University library of Zagreb)

With regards to DARIAH project, RBI's Centre for Informatics and Computing will be involved in creating digital archives of 3D objects and developing intelligent methods for analysing, reconstructing and searching archived objects.

RBI has recognized experts in the field of 3D visualization and distributed computing as well as instrumentation and expertise in 3D object acquisition, analysis and reconstruction. As part of several production grid infrastructure environments (EGI, SEEGRID, CRO-NGI), RBI also holds considerable computing power resources and storage capacity. The 3D archive will be a digital data collection of archaeological objects, sculptures and structures and will be stored in the grid.

The Centre for Informatics and Computing at RBI is also developing novel methods for 2D and 3D image data feature extraction as well as image mining methods which would be used for advanced 3D object multispectral (IR and visible) analysis and for improving registration and search of the objects in the archive.

Collective Intelligence Workshop report

A two-day workshop at King's College, London in February explored the idea of "Collective Intelligence" in relation to DARIAH and the Digital Humanities. Two dozen participants, representing numerous countries, organisations, domains and backgrounds were in attendance, including DARIAH partners from London, Oxford and Dublin. The workshop kicked off with the presentation of position papers from Jan-Cristophe Meister (participating remotely), Andrew Prescott and Susan Schreibman.

Jan-Cristophe Meister (Hamburg University) outlined the plans of the Association for Literary and Linguistic Computing (ALLC) to re-launch its

website with three major functions, namely provide a moderated Digital Humanities information platform for the association's members and affiliates, that will offer a "one-stop" overview on current DH activities, funding opportunities and services, with links to more detailed external repositories.

As a precondition to the wider sharing of such data, Meister emphasised the need "to define a data curation protocol stipulating standards for the moderation and validation of DH information by information gatherers and providers", warning that without such a protocol there would be too much variation in the shared information, making it obsolete and "creating 'white noise' that will frustrate information seekers."

Meister therefore proposed "the definition of a DH atlas or a DH taxonomy enabling us to systematize DH information".

Andrew Prescott (University of Glasgow) proposed that we need a new generation of tools that will work with publishers and other content providers, and enable new perspectives on data and humanities research questions.

Susan Schreibman (Digital Humanities Observatory, Ireland) outlined the detailed and extensive work done in the DRAPier¹ project (Digital Research and Projects in Ireland) to scope and describe digital humanities work and act as a collaboration space to share expertise. Susan proposed greater use of Web 2.0 technologies in future initiatives in this area.

There was also a presentation of the arts-humanities.net² portal, and a discussion of the lessons to be learned

from its six years of existence. The possibility of preferring to follow a design path more oriented towards 'apps', 'gadgets' or 'toolkits' was considered.

Group discussions focused on how to move forward in creating more interoperable metadata. Do we already have adequate standards and procedures for sharing information? Do we need the carrot or the stick to encourage data creators to follow them? Do we need to link communities and expect the metadata to follow, or vice versa?

Some concrete suggestions emerged for potential ways forward to capture, disseminate and use the potential knowledge that is embedded in our current and past activities. An aggregation of information about events was strongly promoted, and the idea of a service for mining the collected knowledge of past discussions on relevant email lists and forums was mooted.

There are plenty of organisations and initiatives producing useful information, that there is a general willingness to share, but due to various factors there is a certain inertia tending to block efforts to do so. Measures to overcome this inertia and to make it easier to exploit our collective intelligence should be a key guiding principle of our next steps.



University of London

¹ <http://dho.ie/drapier>

² <http://arts-humanities.net>