

Humanities eInfrastructure initiatives in Denmark

Hanne Fersøe, Bente Maegaard, Bolette S. Pedersen

University of Copenhagen, Njalsgade 140, 2300 København S, Denmark

hannef@hum.ku.dk, bmaegaard@hum.ku.dk, bspedersen@hum.ku.dk

ABSTRACT

This paper presents an overview of the policies and decisions which have resulted in the current research infrastructure landscape in Denmark within the humanities, and it describes individual initiatives and how they fit into the landscape.

The landscape was fragmented, and it still is to a degree, but one key factor has played a determining role in shaping it into what it is today: the European strategic processes and policies following from the development of the European Research Area, and the ESFRI (European Strategic Forum for Research Infrastructure) process, which has been adopted by Danish policy and decision makers. The strategies and priorities based on the European ESFRI process has resulted in funding of Danish research infrastructure (RI) initiatives within the humanities in the last 5-6 years, such that we can now define a landscape with national, Nordic and European collaboration and networking, extended into the social sciences community as well.

The list of initiatives is comprehensive: CLARIN, META-NORD, META-NET, DARIAH, DASISH, CLARA, ASTIN, DIGHUMLAB, DIGDAG, LARM.

KEYWORDS: Denmark, eInfrastructure, Humanities, CLARIN, META-NORD, DASISH, DIGHUMLAB.

1 Introduction

In the following, we first present a brief overview of priorities and processes leading to funding of Danish research infrastructure (RI) initiatives within the humanities in the last 5-6 years, including the international collaboration and networking leading to or resulting from these initiatives. Following this overview we describe the individual initiatives with regard to funding, national and international collaboration, objectives, results and future plans. The conclusion sums up the status and the future development.

2 European Priorities

This section describes the ESFRI process which is central in establishing a common European approach to RI, and which recommended the national strategic funding schemes for RI currently in process in the member states.

2.1 The ESFRI Process

It is well known, traditionally, that RIs were conceived as single sited physical installations and large scale facilities such as for instance telescopes, electron lasers or particle accelerators, and that these installations were extremely costly to build and maintain. They were therefore often jointly funded by several countries, and scientists from the relevant disciplines would subsequently apply for and be granted access to use the RI facility for a period of time or scientists would be employed as permanent staff at the facilities. In the humanities and social sciences disciplines with different research traditions and methods, national libraries and archives constituted, and still generally constitute, the primary type of research infrastructures. They are, by nature, single sited, but with the emergence of more advanced IT technology they have developed and offer on line access to their databases, thus also defining themselves as virtual facilities (European Commission, 2007).

Steps towards a coordinated European approach on RI were taken in 2001 in connection with the development of the ERA, The European Research Area, leading to the creation in 2002 of ESFRI (European Commission, March 2013).

The first ESFRI Roadmap (European Strategy Forum on Research Infrastructures, 2006) was published in 2006 with updates in 2008 and 2010. The roadmap lists RI initiatives, which are identified as new large scale RIs to strengthen the European Research Area. Five initiatives within the Social Sciences and Humanities are included CESSDA, Council of European Social Science data Archives (CESSDA web site, April 2013), CLARIN, Common Language Resources and Technology Infrastructure (CLARIN web site, April 2013), DARIAH, Digital Research Infrastructure for the Arts and Humanities (DARIAH web site, April 2013), ESS, The European Social Survey (ESS web site, April 2013), SHARE, Survey of Age, Health and Retirement in Europe, (SHARE web site, April 2013). These applied, successfully, for FP7 funding to prepare their implementation phases, and are today ERICs (European Research Infrastructure Consortium) or in the process of becoming ERICs or equivalent structures.

Partner institutions from these five ESFRI RIs, including University of Copenhagen, Department of Language Technology (UCPH-CST), are today collaborating in the so called cluster project DASISH, Data Service Infrastructure for the Social Sciences and Humanities (DASISH web site, March, 2013), where they work on joint activities related to data access, data sharing, data quality, and data archiving to provide common solutions to common problems.

2.2 The META-NET Initiative

There is a long European tradition through the framework programmes for scientific collaboration across Europe in the areas of computational linguistics, language technology, and resource and tool development. This has resulted in the creation of META-NET (META-NET website, March 2013), a research infrastructure initiative for the Humanities, which focusses specifically on language technology. The European Commission's 7th framework programme and the ICT Policy support programme gave support to the 4 projects T4ME, CESAR, METANET4U, and META-NORD, and to the umbrella project META-NET. UCPH-CST was a partner in META-NORD.

Two central awareness activities were accomplished within this initiative: the Language White Paper series on language technology in Europe (META-NET white papers, March 2013) as well as the Strategic Research Agenda (SRA) (META-NET SRA, March 2013). These publications aimed at attracting the attention of and informing politicians and policy makers in their decisions regarding language policy and language technology, especially with regard to the upcoming European funding opportunity Horizon 2020.

The project funding ended 31st January 2013, but the META-SHARE repository (META-SHARE Repository, March 2013), which is an infrastructure facility offering resources, technologies and services is in place and is being maintained. The META projects had partners from a large number of European countries, and many of the partners are also members of the National CLARIN consortium of their country.

2.3 Other supporting initiatives

Among recent European initiatives which have objectives that support the RI agenda, we can mention the Marie Curie programme, which supports the training of 19 research fellows in the area of language resources and their application through the Initial Training Network, CLARA (CLARA web site, March 2013). UCPH-CST employs two of these fellows. Denmark also participated in the FlaReNet (FlaReNet web site, March 2013) project, which developed a common European vision for language resources and technologies with the support from the eContentPlus programme. There were also other projects supported through FP7 which contribute similar types of results, e.g. the MEDAR (MEDAR website, March 2013) project, coordinated by UCPH-CST, which built networking mechanisms within Arabic Speech and Language Technologies, updated the Arabic BLARK, and created Arabic language resources as well as a prototype MT system for English-Arabic.

3 Nordic Collaboration priorities

Formalised collaboration between the Nordic countries takes place through the Nordic Council, one of the oldest and most comprehensive intergovernmental collaborations in the world (Nordic Council website, March, 2013).

Various programmes offer instruments to support collaborative research between the Nordic countries in the humanities area, the overall objective always being to support mutual understanding of Nordic languages and culture. There have been no specific research infrastructure initiatives in the language and culture areas, but the language technology communities in the Nordic and Baltic countries have collaborated in a number of other types of initiatives over the years, so a tradition for knowledge exchange exists, which has been useful in connection with the formation of European RI consortia such as CLARIN, META-NET and DASISH. Examples of such initiatives are e.g. the Nordic Language Councils collaboration in the working group ASTIN (ASTIN website, March 2013), which as one of its goals has “att verka för ... tillgång till nödvändig infrastruktur i form av grundläggande språk- och teknikresurser för de nordiska språken”, and NEALT the Northern European Association for Language Technology (NEALT web site, March 2013).

4 National Danish Priorities

Already in 2005, a strategy paper, surveying existing Danish infrastructures and needs, and proposing a strategy for future RIs had been published (Forsknings- og Innovationsstyrelsen 2011). Based on this strategy paper calls for RI proposals were published, and two humanities proposals were awarded grants: the Danish CLARIN project proposed by UCPH-CST, which referred itself to the ongoing ESFRI RI initiatives in Europe, and LARM proposed by UCPH-Department of Arts and Cultural Studies.

In parallel, Denmark decided to join the ESFRI roadmap process and became one of the EU member states who agreed to develop a national roadmap. Working groups with members from 6 major research areas were established, and their survey work resulted in publication of the Danish National Roadmap by the Danish Agency for Science, Technology and Innovation (DASTI) in 2011. The roadmap recommended prioritized initiatives, following from the ESFRI process, within 6 scientific fields, humanities and social sciences being one of the fields.

The Digital Humanities Laboratory, DIGHUMLAB, comprising a research infrastructure for the humanities, was awarded a five year grant to the four proposing Danish universities, University of Copenhagen being one of them. Other interested parties such as e.g. the Royal Library and the State and University Library have a standing invitation to join the consortium, at their own cost.

5 Research infrastructure activities in Denmark

This section describes briefly each of the international infrastructure initiatives with Danish involvement within the last six years, and two national projects.

5.1 CLARIN

The CLARIN preparatory project, CLARIN PP, started in January 2008 with the aim of preparing the full scale implementation of the CLARIN RI. The initiative met a huge need in Europe for a unified and concerted approach to a previously fragmented scientific field, and there was and is massive back up from the communities in all the European countries and from CLARIN members worldwide to providing access to resources, tools, and services to all language based sciences, particularly the humanities.

5.1.1 European CLARIN Preparatory Phase

Denmark's role in the CLARIN PP was to lead the work on governance in a future European organization. During the project period the Commission worked out and completed the ERIC Regulation, which sets out the legal framework for a totally new type of European organization with countries and intergovernmental organizations as members. The ERIC Regulation constituted a lever, which enabled the CLARIN consortium to thoroughly prepare the statutes and the application for the future CLARIN ERIC. Other preparatory tasks concerned for instance establishing a technical infrastructure, securing interoperability, and proposing a legal framework for the future use of data resources, tools and services.

5.1.2 Danish CLARIN Preparatory Phase

The first Danish CLARIN project ran in parallel with the European PP project, but not as a preparatory project. The Danish CLARIN consortium had to implement a technical infrastructure and populate it with content. This work is described in detail in (Fersøe et al., 2009).

The project has produced a data repository with a web portal accessible through the user authentication and authorization system WAYF (Where Are You From) (WAYF website, April 2013). The portal contains:

- 46,000 written text files with 143,000 annotation files for them
- 80 videos, 37 audios with 45 annotation files for them
- 3 lexica
- 5 tools

The portal can be accessed at www.clarin.dk.

5.1.3 CLARIN ERIC

In February 2012 CLARIN was granted ERIC status following a successful application process with the Netherlands as the hosting country and 9 founding members. Currently several countries are in the process of applying for membership and one country has obtained observer status.

Denmark's role in the ERIC is twofold: as a member we are obliged to deliver the contribution described in the CLARIN Agreement, which is made between the ERIC and each member (see section 5.2.1), and in addition, following from our governance work in the PP project, the CLARIN ERIC Vice Executive Director is Danish from University of Copenhagen.

The ERIC is established with a very lean administration, and all boards, standing committees and working groups have been established.

5.2 DIGHUMLAB

The DIGHUMLAB initiative is carried out in collaboration between Aarhus University, University of Copenhagen, University of Southern Denmark, and Aalborg University. The DIGHUMLAB project plan is structured into a project management and administration layer and three research themes. The Danish CLARIN and DARIAH fees and in-kind contributions are funded through the research themes 1 and 2.

5.2.1 DIGHUMLAB, Theme 1

Theme 1, Language based Materials and Tools, is anchored at UCPH-CST. The rights obtained by a country through membership, e.g. full access to CLARIN and all its services to the entire Danish research community, are described in the CLARIN ERIC statutes, while the obligations, e.g. in-kind contributions, are described in the CLARIN Agreement between CLARIN ERIC and Denmark.

In addition to paying the annual fee and contributing to the overall goals stated in the statutes, there are country specific in-kind contributions. All members must create a national consortium, provide a national co-ordinator, and provide a data and service center which gives user access in conformance with the access policies defined in the statutes and with an approved user authentication and authorization system; they must also contribute to international knowledge sharing. In addition, they must provide a work programme for the national consortium comprising improvements and enhancements to existing data resources, tools and services, creation of new resources, tools and services for CLARIN, promotion of standards, and a national knowledge sharing infrastructure. All these activities are listed and defined in the agreement, and they are also found in the DIGHUMLAB theme 1 work plan.

The portal developed in the Danish CLARIN preparatory phase constitute the point of departure for the in-kind contributions. Currently, a major activity is preparing for the national centre assessment. Another on-going activity is a revision of the existing metadata scheme. A third activity is to update and enhance the user interface to the web portal, which is the access gate to the Danish CLARIN resources, tools and services.

5.2.2 DIGHUMLAB, Themes 2 and 3

Theme 2 is divided into the sub themes NetLab, and Tools for Sound and Image Media. The sub theme NetLab was a joint initiative of the State University Library and the Royal Library in 2005 aimed at archiving the Danish internet. The NetLab focus in DIGHUMLAB is on the development of adequate tools for research.

The sub theme Tools for Sound and Image Media focusses on extracting metadata, developing tools, and on research maturation in relation to radio, TV, film and related media. Collections of sound and image media represent not only a significant cultural heritage resource for the 20th century in particular, but also a rich research source for the humanities and social sciences. The sub theme is carried out in close collaboration

with the LARM research infrastructure project, and serves as a Danish in-kind contribution to DARIAH.

Theme 3 will establish a Danish national research infrastructure with a focus on the experimental environments within the humanities. This will be achieved through i.a. the establishment of contemporary new laboratory facilities with a special focus on interaction between humans and technology. Supporting the development of relevant investigation protocols and development of and compliance with scientific ethical standards will therefore be a major focus area. The material thus produced will additionally be made available through CLARIN.

5.3 DARIAH

The DARIAH PP ran in parallel with the CLARIN PP with 14 partners from 10 countries. The objective of the preparatory phase was to set up physical, strategic and human elements of the RI. The ERIC application process has started with France as the host country and 12 countries signing the Memorandum of Understanding, Denmark is one of them.

UCPH-Department of Scandinavian Research was a partner in the DARIAH PP project leading the work package on dissemination. After the completion of the PP project, there was an unfunded gap where the transition to ERIC status was prepared. DARIAH now operates through its European-wide network of Virtual Competency Centres (VCC). Each VCC is cross-disciplinary, multi-institutional and international and centred on a specific area of expertise. Denmark, University of Aarhus, together with Ireland, coordinates the Virtual Competence Center on Research and Education as a part of DIGHUMLAB, Theme 2, see section 5.2.

5.4 META-NORD

UCPH-CST participated in the Nordic and Baltic branch of the META-NET initiative, META-NORD, during the period 2011-2013. Apart from enhancing and upgrading a considerable number of Danish language technology resources and tools to agreed standards, the main, national focus of this participation consisted in a series of awareness actions concerning Danish language technology. The actions were realized as a white paper on the status of the Danish language in the digital age (Pedersen et al., 2012), a comprehensive press campaign in the Danish media, and as a national language technology workshop for researchers, industry and decision makers. By the end of the project, a Danish META-SHARE node was established, and 56 Danish language technology tools and resources were made accessible via this platform. These include analysis tools and resources provided by UCPH-CST, such as text corpora, a WordNet (DanNet), a computational lexicon for Danish (STO), and an annotated audio-visual corpus (the NOMCO Corpus). Furthermore, tools and resources are provided by University of Southern Denmark/GrammarSoft (constraint grammars, machine translation modules, Danish FrameNet among others) and from Copenhagen Business School (CBS) (treebanks) and The Society for Danish Language and Literature (the CLARIN Corpus). Most of the resources are monolingual and focus on the Danish written language, some, however, include cross-lingual links or alignments.

5.5 DASISH

DASISH is a cluster project between the five ESFRI RIs in the social sciences and humanities area, it has 19 partners, including 6 partners from the Nordic countries, and it is supported by a grant from FP7. The project is managed by the CESSDA president at Swedish National Data Service, University of Gothenburg; the Finnish and one of the Norwegian partners are also from CESSDA, while the second Norwegian partner, the Danish and the Estonian partners are from CLARIN.

The project started in January 2012 and runs for three years. The goal of DASISH is to identify areas of synergy in the infrastructure development of all five RI communities and to work on concrete joint activities in order to propose common solutions to common problems. The idea is to avoid double developments, and that the RIs should mutually benefit from advanced developments done by the others. Examples of such joint activities are: a) understanding the different architectural solutions in RI construction, b) data and metadata quality issues for instance in data collection in European-wide surveys and in the data management and curation methods used in text archives, c) development of a joint shared data access and enrichment framework, comprising for instance AAI, PIDs, joint metadata, workflow implementations, joint annotation framework, d) joint legal and ethical activities, for instance identification of issues and constraints, challenges imposed by new data types, e) development of training and education modules, and f) dissemination of activities and results.

UCPH-CST is leading the work on dissemination and is also involved in the joint activities on PIDs, workflow implementations and joint annotation framework.

5.6 DIGDAG

DIGDAG (DIGDAG web site, March 2013), a digital map, is a cross-institutional research project supported by the Ministry of Science Technology and Innovation under the national RI funding scheme. The name DIGDAG is an abbreviation of *Digitalt atlas over Danmarks historisk-administrative geografi*. The project started in 2009 with the purpose of establishing a historical-geographic database over the administrative subdivision of Denmark and thereby a) creating an infrastructure for research in Denmark's administration history from year 1600 to current time, b) developing a solid search engine to be used by archives, collections and libraries, and c) supporting research in administration history.

The project participants include a large part of the most important Danish cultural and research institutions: The State Archives, University of Copenhagen, University of Southern Denmark, Kort og Matrikelstyrelsen (an agency for maps which is now restructured), the National Museum, The Royal Library, and Kulturarvsstyrelsen (now restructured into the Danish Agency for Culture).

5.7 LARM

LARM¹ Audio Research Archive (LARM website, March 2013) is an interdisciplinary project aiming at producing a digital infrastructure to facilitate researchers' access to the

¹ LARM is the Danish word for noise

Danish radiophonic cultural heritage. The LARM project is a collaboration between a number of research and cultural institutions: The University of Copenhagen, Roskilde University, The University of Southern Denmark, Aalborg University, Aarhus University, The Royal School of Library and Information Science, The Danish Broadcasting Corporation, The State and University Library, Danish e-Infrastructure Cooperation, Kolding School of Design, and The Museum of Media. The project is made possible by a grant from The National RI funding scheme.

The infrastructure is a digital archive with the appropriate research tools which will give access to thousands of hours of national and local radio broadcasts from 1925 and onwards. The platform will enable researchers and university students to stream sound to their own computers directly from the digital archive, which at the conclusion of the project will contain more than one million hours of sound.

User driven innovation is a key element in LARM. The infrastructure and its interface are based on user needs and are developed in close collaboration between technicians, cultural researchers and designers, and the research projects deliver feedback to the development of digital audio search and audio description tools in both effective and innovative ways.

6 Conclusion and Outlook

In conclusion, the status of Danish eInfrastructure within the humanities is that many initiatives have been started in the last 5-6 years, they address various language based data types and media, including some tools and services for the research. The initiatives are anchored in European and national strategies and processes, and they provide excellent opportunities for networking and knowledge sharing across countries and communities and among institutions.

In the future there are a number of important aspects that need to be addressed. Continued funding is an issue; the universities are in the process of preparing their strategies concerning embedding into the university activities. Such an embedding will support the idea that research infrastructure is a general tool for research and consequently be the best possible support for further take up. On the other hand external funding will undoubtedly be necessary as the creation and maintenance of RI is a heavy task, just like the maintenance of a library.

References

- ASTIN website (March 2013). <http://nordisksprogkoordination.org/links/astin/>
- CESSDA web site (April 2013). <http://www.cessda.org/>
- CLARA web site (March 2013). <http://clara.b.uib.no/>
- CLARIN web site (April 2013). <http://www.clarin.eu/>
- DARIAH web site (April 2013). <http://www.dariah.eu/>
- DASISH web site (March, 2013). <http://dasish.eu/>
- DIGDAG web site (March 2013). www.digdag.dk
- DIGHUMLAB website (March 2013) <http://dighumlab.dk/>
- ESS web site (April 2013). <http://www.europeansocialsurvey.org/>
- European Commission, European Science Foundation (2007). Trends in European Research Infrastructures. Analysis of Data from the 2006/07 survey.
- European Commission (March, 2013)
http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-background
- European Strategy Forum on Research Infrastructures (2006). European Roadmap for Research Infrastructures.
- FlaReNet web site (March 2013). <http://www.flarenet.eu/>
- Fersøe, H. and Maegaard, B. (2009). CLARIN in Denmark – European and Nordic Perspectives. In Proceedings of the 17th Nordic Conference of Computational Linguistics - NoDaLiDa 2009, Workshop: Nordic Perspectives on the CLARIN Infrastructure of Language Resources, *NEALT Proceedings series no.5*.
- Forsknings- og Innovationsstyrelsen, Ministeriet for Videnskab, Teknologi og Udvikling (2011). Dansk roadmap for forskningsinfrastruktur.
- LARM website (March 2013). <http://www.larm-archive.org/>
- MEDAR website (March 2013). <http://www.medar.info/>
- META-NET website (March 2013). <http://www.meta-net.eu/>
- META-NET white papers (March 2013). <http://www.meta-net.eu/whitepapers/overview>
- META-NET SRA (March 2013). <http://www.meta-net.eu/sra-en>
- META-SHARE Repository (March 2013). http://www.meta-net.eu/meta-share/index_html
- NEALT web site (March 2013). <http://omilia.uio.no/nealt/>
- Nordic Council website (March, 2013). <http://www.norden.org/da/om-samarbejdet>
- Pedersen, B.S, J. Wedekind, S. Kirchmeier-Andersen, S. Nimb, J.E. Rasmussen, L.B. Larsen, S. Bøhm-Andersen, H. Erdman Thomsen, P. J. Henrichsen, J. O. Kjærsum, P. Revsbech, S.Hoffensetz-Andresen, B. Maegaard (2012). The Danish Language in the Digital Age - Det danske sprog i den digitale tidsalder. META-NET White Paper Series, Springer Verlag.

SHARE web site (April 2013). <http://www.share-project.org/>

WAYF website (April 2013). <http://www.wayf.dk/>