



NetLab is a research infrastructure project
for the study of internet materials,
affiliated to [Center for Internet Studies,](#)
[Aarhus University](#)
and a part of the national
Danish Digital Humanities Lab

Proposal: A Research infrastructure for the Study of Archived Web materials. (RESAW)

Project

The project aims at developing a coordinated, but distributed workspace (e.g. for searching, analyzing, and visualizing) and to build the relevant skills for development and use of software supported methods in the study of internet materials across different national web archives and to build strategies for archiving relevant web materials which are not taken care of within existing institutional frameworks, such as materials on the domain .eu and other non-national domains (.net .info .biz .mobi etc.).

The project should be driven by research questions which will lead to results of relevance for the wider community of scholars and researchers within the relevant fields who are not specialized in studying digital materials by help of software supported methods.

Thus, the establishing of a sustainable digital research infrastructure must be based on a close interplay between web archiving institutions and the relevant research communities, and the research questions which scholars want to inquire.

Scope

Of all media the internet today holds the richest set of materials documenting contemporary social, cultural, and political life. These materials belong to the original cultural heritage of our own age on the level of national cultural heritage but even more on the level of European and global cultural heritage. Web archives are also of relevance as source for the study of the past even before the origin of the web. In the years to come such archives of digitized and born digital materials will be of still growing significance for research and scholarship as well as for the wider public as they become a major source for a growing variety of disciplines and fields of study.

The building and the study of internet and particularly web archives pose a number of new issues as compared to most formerly known digital sources and databases. Contrary to digitized cultural heritage materials web materials are born digital, the formats are not defined by scholars or researchers or limited due to a restricted set of editorial standards. The materials are also often modified, moved or removed very fast. Therefore these materials are today archived on a national basis in many countries. Today the international Internet Preservation Consortium (IIPC) counts more than 30 major and primarily national web archives as members and more than 20 being national European archives.

These archives cannot be complete and different strategies are employed for collecting, storing, accessing, and searching. There is a need for a thorough comparison, validation and possibly standardization of the archives, a need for developing adequate frameworks for the handling of research data obtained from searching and analyzing the web materials and for the development of appropriate software supported methods for the study of these materials.

Thus to access, validate and study these archives new methods and infrastructural facilities are needed both on a national and a European level. While the archives are now established in most EU countries, the infrastructural facilities have to build on close collaboration between archives and research communities.

Since a major part of the content of web materials is articulated in many different languages, and the various national archives are build due to different criteria for collection building, storage and access there is a huge need for more coordinated developments which would allow large scale international research collaborations on a wide range of issues as for instance: Coordinated harvesting and studies of particular European and/or global events (e.g. EU elections, all sorts of studies of public discourse on say, climate debates, EU-debates etc); the study and understanding of the growing importance of the internet in its actual state as well as the history of the internet (e.g. how the internet has developed throughout the last two decades within a European frame as carrier of identity). Projects could include comparative analyses of the use of the internet in identity building in different national and/or regional settings, and at different times; how national identities and the wide range of minority identities within Europe are connected to European identity formation processes in different countries? Is it possible to identify overarching historical trajectories in national and transnational opinion building as articulated on the internet? What are the typical and characteristic cultural patterns of network relations?

The project will serve 1) to favour comparative approaches to different archives and interoperability across web archives in different countries and with different sets of linguistic materials, 2) to validate the quality of the various archives due to the different combinations of methods used in building the archives, and 3) to build a platform and to develop software supported methods allowing large scale European comparative research in web materials. The project will be aimed at scholars within media studies, history, cultural studies, cultural memory and identity studies, political science, sociology, psychology all sorts of studies of public discourse.

The proposal is developed in collaboration with DARIAH (dariah.eu) and NetLab (netlab.dk), which is a part of the Danish research infrastructure DIGHUMLAB (dighumlab.dk). The collaborations

between the Danish internet archive netarkivet.dk (a joint venture between The Royal Library, Copenhagen and The State Library, Aarhus) and the research community at The Centre for Internet Research, Aarhus University, Denmark have developed gradually since 2000 initiated by a Pilot Project on archiving the municipal elections in Denmark in 2001. The collaboration today continues within the Danish DIGHUMLAB project which includes the establishing of the NetLab research infrastructure.

We are looking for possible partners based on close collaboration between archival institutions and scholarly communities.

Further issues to be addressed: Privacy protection and access for scholars, for students and for the general public across national jurisdictions.

Aarhus 27.9 2012

Niels Brügger
Head of The Centre
for Internet Research

Niels Ole Finnemann
Professor, dr.phil.

Netlab, Dighumlab Denmark

Aarhus University

Denmark