

Consultation on possible topics for future activities for integrating and opening existing national research infrastructures

Title	
Title of the proposal -open reply-(compulsory)	A Research infrastructure for the Study of Archived Web materials (RESAW)
Contact person	
Family name -open reply-(compulsory)	Finnemann
First name(s) -open reply-(compulsory)	Niels Ole
Organisation -open reply-(compulsory)	
Aarhus University, DIGHUMLAB/NetLab DENMARK	
Position in the organisation -open reply-(compulsory)	Professor, Head of NetLab
Postal address -open reply-(compulsory)	
Department of Aesthetics and Communication, Helsingforsgade 14, DK 8200 Aarhus N	
Country -open reply-(compulsory)	Denmark
e-mail address -open reply-(compulsory)	finnemann@imv.au.dk
Is your proposal representing your own personal view or are you responding on behalf of your organisation as a whole? -single choice reply-(compulsory)	Organisation view
Description of the research infrastructures covered and the trans-national access and /or services provided	
<p>Indicate the type of research infrastructures to be covered by the proposed topic, and list the research infrastructures in Member States, Associated Countries and Third Countries, that would provide transnational access and/or services to researchers, with brief descriptions of the state-of-the-art equipment and services offered to users that make them rare or unique in Europe. Outline the specific areas of research and scientific communities normally served by the infrastructures, as well as new areas opening to users, if any. Indicate what would be the overall access modalities necessary to be developed. Text of maximum 4000 characters including spaces.</p> <p>-open reply-(compulsory)</p>	
<p>The proposed research infrastructure, RESAW, aims to bridge the gaps between the fast growing number of web archiving institutions, who holds a still more significant part of 21th century cultural heritage, and a fast growing range of research communities and individual researchers for whom these archived materials are of increasing importance. It is proposed that such a research infrastructure should be build around general (not least) national Internet archives holding collections of materials primarily published on the web, in collaboration with an international community of scholars and researchers representing the widest possible range of disciplines and methodological approaches to the study of archived internet materials. On a more detailed level the purpose is to 1) foster the collaboration and networking activities between the archival institutions and research communities on an European and global scale, 2) develop a coordinated, but distributed workspace for searching, analyzing, and presenting, 3) build the relevant skills for development and use of</p>	

software supported methods in the study of internet materials across different national web archives, 4) give access to related research data storages, and 5) 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 also to be acknowledged by a 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 web research infrastructure should be based on a close interplay between web archiving institutions, the relevant research communities, and the research questions which scholars want to inquire. The relevant research communities come from a wide range of fields (see 5 and 6b below) The great majority of relevant archiving institutions are members of the “International Internet Preservation Consortium” (IIPC) who today counts more than thirty major and primarily national web archives as members. Twenty of these are national European archives. RESAW will serve to 1) favour comparative approaches to different archives and interoperability across web archives in different countries, which use different principles for collection, and with holdings in different languages, 2) validate the quality of the various archives due to the different principles and combinations of methods used in building the archives, 3) establish a shared portal for accessing and documenting the web archives, 4) develop software supported methods allowing large scale European comparative research in web materials, 5) initiate investigations as to how other internet activities (e.g. email, apps to smart phones and tablets, facebook data etc.) can be integrated into general internet & web archives, and 6) facilitate online access to the archives and research data, and in so far online access is restricted due to national legislation to facilitate in situ access for researchers. Furthermore the project should address the long term perspective of coordinating web archives with other sorts of digital collections, including digitized cultural heritage collections whether available online or not. The project is primarily, but not solely, aimed at researchers and scholars within the humanities and social sciences but is obviously also of high relevance for a wider public in modern societies.

Scientific domains served by the research infrastructures

Select the scientific domain(s) served by the research infrastructures -multiple choices reply-(compulsory)	Social Sciences and Humanities - Mathematics and ICT
Indicate the main scientific domain served -multiple choices reply-(compulsory)	Social Sciences and Humanities

Key potential partners

Indicate a list of key potential partners. Text of maximum 3000 characters including spaces, with 1 line per potential partner (participant organisation name, country and contact person)

-open reply-(**compulsory**)

Archival Institutions: State and University Library, Aarhus Denmark Tonny Skovgård Jensen, Div. Head, Bibliothèque Nationale de France France Clément Oury, Head of Digital Legal Deposit Digital Repository of Ireland Ireland Sandra Collins, Director Koninklijke Bibliotheek The Netherlands Hildelies Balk, Head of European Projects National Heritage Institute, Bucureşti Rumania Dan Matei, Director The British Library UK Helen Hockx-Yu, Head of Web Archiving The National Library of Wales Wales Lorna Hughes, Chair in Digital Coll. Research Institutions, Communities and Researchers: DARIAH EU Sally Chambers, Secretary General WEBDATANET, EUCOST action 1004 - Pablo de Pedraza García, Chair of Action The Internet Memory Foundation France/the Netherlands Julian Masanès, Exec. Director Ghent University Belgium Pieter Verdegem, Ass. Professor et al State and Univ. Library, Aarhus Denmark Ditte Laursen, Senior researcher Aarhus Univ, NetLab - N.O. Finnemann, Professor Aarhus University, CFI/Netlab - N. Brügger, Head of CFI. Arcada University of Applied Sciences Finland Matteo Stocchetti, Adjunct Professor Inst. for Comm. Sciences (CNRS) France Valérie Schafer, Researcher et al Intl. College of Territorial Sciences - Marta Severo, Maître de conf.- en science politique Esaa, the Art school of Avignon - Lionel Broye, Researcher ISAM-IAE, Univ. de Lorraine, Nancy. - Fabienne Greffet, Maître de conf. en science politique Université Cergy-Pontoise - Valérie Aubourg, Ass. Prof. Berlin University of the Arts Germany Stephan Humer, Research Director Athena Research Centre, Digital Curation Greece Costis Dallas, Professor et al National Univ. of Ireland, Galway Ireland Stefan Decker, Professor et al. UCLAM, Cassino Italy Francesco Bolici, Ass. Professor Univ. del Piem. Orientale - Maurizio Lana, Researcher Univ. of Bologna, (CRR-MM) - Francesca Tomasi, Project Leader et al. Vilnius Univ. Fac. of Comm. Lithuania Ingrida Vosyliūtė, Project Manager Erasmus University, Rotterdam The Netherlands Payal Arora, Ass. Prof. Univ. of Amsterdam, Digital Methods - Richard Rogers, Professor Bergen University. Inf. and Media Studies Norway Hallvard Moe, Ass. Professor Univ. de Deusto, The iScience group Spain Ulf-Dietrich Reips, Professor Pablo de Olavide Univ. Seville - Amelia Almorza, Researcher, Lancaster Univ, Dept. of Sociology UK Adam Fish, Lecturer, Swansea University - David Berry, Senior Lecturer The Oxford Internet Institute - Eric Meyer, DPhil Programme Director The Australian Nat. Univ. (VOSON). Australia Dr Robert Ackland, Ass. Prof., La Trobe University - John Tebbutt, Senior Lecturer Univ. of Sydney, Media and Comm. - Gerard Goggin, Professor Univ. of Ontario, Inst.of Technology Canada Isabel Pedersen, Research Chair

Scope and activities

Describe the overall objectives of the activity. Describe the benefit that the proposal would bring about in terms of integrated provision of infrastructure related services. When appropriate, describe how the network would integrate with the relevant e-Infrastructures. Text of maximum 2000 characters including spaces.

-open reply-(**compulsory**)

The building and the study of 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 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 in many countries, most often on a national basis. 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 and presentations of these materials. 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 a range of research projects performed by different research communities each with their own set of research questions. The project will:

- Map the character of known general web archives due to criteria for and methods of selection.
- Test the quality and relevance of the selections due to a variety of criteria (incompleteness due to selections methods, technical limitations, economical limitations, fit with a wide range of research questions) as part of a variety of research projects.
- Add to the standardization of selection criteria and methods.
- Add to standardized sources of documentation and timeline based interface to documentation and collections.
- Develop methods and competences in software supported analysis of archived web mater

Indicate the Networking Activities that could be foreseen to foster a culture of co-operation between the research infrastructures and scientific communities. Indicate the Joint Research Activities that could be foreseen to improve, in quality and/or quantity, the services provided by the infrastructures. Text of maximum 4000 characters including spaces.

-open reply-(**compulsory**)

The overall goal for the network is to improve the quality of archives and ongoing development of methods in the study of these materials. A range of activities will be needed. A major issue for the project will be to incorporate the infrastructure developments into a range of collaborative research projects. This development include the above mentioned portal, standardized, but still modifiable interfaces and documentation formats allowing ongoing and research driven annotations, search, analysis and presentation. It also includes the initiation and coordination of research projects aiming at the widest possible range of approaches, methodologies and research questions posed. As the internet itself facilitate transnational activities there is also an increasing need of transnational research be it of transnational events (EU elections, Olympics, European song contest) and of comparative studies of political, social and cultural variations in the understanding of all sorts of issues on the European and global agenda. While the major collections are already established they are neither standardized nor optimized to research purposes. The web materials are heterogeneous, the methods are constantly renewed and the research communities comprise a wide range of different approaches. Due to these complexities the overall project will have to combine a sustainable facility for combined networking online and in situ both between archival institutions and research communities and in between different research groups. Such a facility would furthermore allow for ongoing matching, recruitment and inclusion of new research groups from any field and/or country. The proposed networked infrastructure will have to develop the cooperation between existing archives and facilitate emerging interdisciplinary cooperation between researchers from a growing range of fields and disciplines concerned with the study of archived web materials, as these materials are still more often the sole or most significant possible source materials. The proposal is supported by researchers in many different fields, ranging from information science, big data-analysis, semantic web-projects, network and web sphere analyses to a variety of – often combined – quantitative and qualitative approaches within sociology, psychology, political science, history, classical humanities, cultural memory and identity studies, and media and cultural studies of web genres, style and meaning of websites, and of individual messages. The various approaches contribute with their own conceptualisations of the main units of analysis, methodological framework and type of research questions. The proposal is furthermore supported by major European web archiving institutions and research communities, including also important non-European communities, which indicate the both central and broad scope of the proposal. To bridge the gap between the general purpose archive with an inherent long term perspective, also serving posteriority, and the particularity of any specific research project, the networked infrastructure should also facilitate the development of specialized IT competences which are solidly anchored

within the conceptual and methodological framework of any particular research community. Finally the network should aim to support and train scholars who are not yet skilled in the use of web archives and digital methods. The proposal is supported by DARIAH and the network would relate to DARIAH and other infrastructures under the DASISH umbrella.

Need for European integration

Explain why this proposed topic would require a European (rather than a national or local) approach. Describe how resources provided by EU would be mobilised. Indicate how account is taken of other national or international activities, and any resources that would complement an EU contribution. Text of maximum 3000 characters including spaces.

-open reply-(**compulsory**)

Since a major part of the content of web materials is linguistic, articulated in many different languages, and the various national archives are based on 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? These questions do not constitute an exhaustive list. The internet favours international collaboration. So does internet archives and the study of these archives. The bringing together in a dynamic and ever evolving network of archiving institutions and research communities will contribute to intensified international collaborations and – due to the shared object of study – to intensified cross fertilisation on the level of methods, theories and insights in contemporary society, ranging from insights in new communicational patterns, new business models, political processes and opinion building to new forms of self expression and identity building on both individual, national and transitional scales. With a majority of existing general web archiving institutions located within EU the RESAW infrastructure will be a dominant agency within the humanities and social sciences not only within EU but with the option to have a global reach. Even if web archiving today in Europe is primarily a national issue a full exploitation of these archives is only possible if they are increasingly incorporated into a general framework allowing the materials to be studied as close as possible to their original appearance in a transnational medium. And even if the materials are articulated in a manifold of different languages the software and the facilities which are needed to archive, search, analyze and visualize these materials can be shared, thus both fostering international collaboration between archival institutions and researchers, and reducing the developmental costs.

Expected impact

Describe the expected impact of the proposed activities on the scientific communities, on the functioning of the research infrastructures, and on the development of the European Research Area (including balanced territorial development). Highlight the contribution to socio-economic impacts, including for promoting innovation and developing appropriate skills in Europe. Text of maximum 3000 characters including spaces.

-open reply-(**compulsory**)

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. Internet materials, however, are rapidly disappearing except for those parts which are archived. Thus the study of archived web materials is increasingly the study of some of the most significant sources to contemporary and near past history. Without the study of these archives society will increasingly lose sight not only of the history of the internet but of society at large. Socio economic impacts in a variety of fields may be added, but are not easily predictable. However the study of web archived materials will no doubt foster the development of new and more appropriate forms of IT competences more fundamentally anchored in the various disciplines and fields. 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 proposal aims at the establishing of a new infrastructure and a new type of infrastructure with no known antecedents. It takes as its point of departure the growth of web archiving institutions in the last 15 years aiming at bridging between 1) these predominantly national institutions among themselves, 2) the archiving institutions and the growing number of research communities oriented towards the study of archived web

materials, and 3) between these research communities among themselves. While the materials in archives often represent linguistic and assumingly also cultural differences (including differences in the selection criteria) between the various countries, the research communities on their side represent a variety of epistemological and methodological approaches which undoubtedly will favour far reaching cross fertilizations. The infrastructure finally will also have to relate to different sets of legal criteria for accessing the collections, and thus include both restricted in situ access and full scale online access. While existing archives are funded on a national basis there is a need for funding of development costs in establishing and unfolding of the proposal into a full project aiming at the development of a sustainable international platform. An international infrastructure is the sole way to solve a range of limitations for the use of existing national archives as well as the most fruitful way to incorporate new research groups as they emerge in more fields, disciplines and countries.

Projects previously funded under FP7 and FP6

Only for those proposed topics that correspond to the follow-up of FP7 or FP6 funded Integrating Activities, please provide specific additional information on: the project(s) previously or currently funded and the level of funding; the main results and expected achievements of the funded project(s); the progress foreseen in the activities proposed beyond FP7. Text of maximum 4000 characters including spaces.

-open reply-(optional)

The proposal aims at the establishing of a new infrastructure and a new type of infrastructure with no known antecedents. This version replaces submission with IPM reference number: 503445402321429212 The only difference is the format of the confirmed potential partners, which should be 1 partner per line. Seems rather difficult to reach in the formular.