

PhD seminar

Web Archiving and Archived Web — a new Research Method, a new Object of Study?

Aarhus University, Denmark, 11-12 June 2014

Organised by the Danish Digital Humanities Lab/NetLab & Aarhus University, the PhD programme 'ICT, Media, Communication and Journalism'

For a number of years the web has played a pivotal role in our communicative infrastructure as well as in cultural and political life. But the web has not only been important in society at large, it has also played a still more dominant role for the social sciences and the humanities, as a research tool for gathering and analysing empirical material, as well as as an object of study.

Web archiving has been used since the mid 1990ies. On the one hand, contemporary studies of the web has used web archiving as a means of preserving and documenting the object of study, and, on the other hand, large web archives have been established with a view to preserving the cultural heritage, either on a national or a transnational scale. Thus, the archiving of the web has been used as an element within the research methods and as a cultural heritage initiative.

This PhD seminar focus on web archiving and web archives with a view to investigating the nexus between web archiving and web archives as a new research method and as a new object of study. The aim of the seminar is double. On the one hand it is to introduce web archiving as a research method to be used by scholars studying contemporary political, social, and cultural phenomena within the humanities and the social sciences, and, on the other hand, the aim is to introduce to the methodological and theoretical issues related to the use of existing (trans)national web archives, in the main in relation to historical studies involving the web.

The course format

There will be four lectures, as well as paper presentations from the PhD scholars. Each PhD scholar will hand in a short paper (5-6 pages, max. 15.000 type units) and a short video-presentation (max. 5 minutes). With a view to promoting discussion rather than presentation PhD scholars will only briefly present the main ideas of their videos and papers during the seminar (1-2 tracks depending on the number of participants), and each PhD scholar must prepare feedback to one other paper/video. The papers and videos will be distributed in a closed Facebook group for all participants before the seminar.

Costs

Participation as well as coffee and lunch are free of charge. Participants must pay for dinner, travel and accomodation themselves.

Dates and venue

The PhD-seminar will take place 11-12 June 2014 at Aarhus University in Denmark.

Number of participants

The number of participants is limited to 20.

ECTS-points

4 ECTS.

Application deadline and seminar enrolment

Deadline for submission of application is Monday 24 March 2014.

The seminar application consists of a one-page abstract. The abstract can be either a project description or an expression of interest in the field. Notification about enrolment will be sent out no later than Friday 28 March 2014.

Participants who are enroled must submit the video and the paper no later than Thursday 15 May 2014. For questions about course content please contact Niels Brügger, <mailto:nb@imv.au.dk>.

Submission is open from 1 March at <http://webarc.netlab.dk>.

Seminar readings

Seminar readings will be made available one month before the course, with required reading and suggested literature for the seminar. PhD students are expected to have read the literature before the beginning of the seminar.

The lectures and the lecturers

Virtual Digs: Excavating, Preserving, and Archiving the Web

Meghan Dougherty, Assistant Professor, Digital Communication, Loyola University Chicago

The Web contains a wealth of information about institutions, governments, marketplaces, and societies. This information has potential to make up a rich digital historical record that is built by content producers, conserved and preserved stewards of cultural heritage, and invaluable as evidence of the cultural past for researchers in the social science and humanities. But this historical record is often overwritten with new status updates and revised sites. The Web is dotted with only occasional elements of permanence, which is still rather unreliable. We trust in the infrastructures built around Web content that aid in information seeking as the arbiters of what is current, true, vetted, and so worth remembering. And so every second that ticks by, crucial information in that unfolding historical record is lost.

There are a variety of approaches to Web archiving common in national libraries, but these often adapt traditional archival theories and practices to digital cultural landscapes. These efforts treat websites as documents that can provide evidence of the past. But this document-centric approach does not necessarily provide scholars of Web history with the archival records they need to conduct retrospective study of Web culture.

This lecture focuses on the research needs of humanities and social science scholars, the nature of observation and objects of study online, and how to ethically collect evidence for the purpose of study and for a robust and replicable scholarly knowledge base.

Meghan Dougherty is an Assistant Professor of Digital Communication at Loyola University Chicago's School of Communication, and studies the preservation of Web cultural heritage, collaboration tools for research, and Web archiving as an emerging cyberinfrastructure for e-research. She has a Ph.D. in Communication from the University of Washington, Seattle.

A Data Driven Approach to Web Archive Research

Anat Ben-David, post-doctoral researcher with the WebART project, University of Amsterdam

In recent years, the single URL approach to accessing Web archives is being gradually replaced by a big data approach. Several web archives and research initiatives have already adapted faster and more efficient clustered storage of web archives to develop advanced search interfaces and analytical tools. While this transition potentially increases the value of web archives for research, it also proposes new theoretical and methodological challenges that need to be

addressed.

How would Web archive research methods change once Web archives become searchable on a large scale? This presentation attempts to answer this question by introducing data driven methods for critically studying archived Web data. Such methods open up a variety of analytical practices, such as cross-collection analysis, the re-assemblage of existing collections around a theme or an event, the study of archival artefacts, and the (re)scaling of the unit of analysis from a web element or single page, to a collection or the full archive. The presentation includes examples using tools and methods developed by the WebART project at the University of Amsterdam, in collaboration with the National Library of the Netherlands.

Anat Ben-David is a post-doctoral researcher with the WebART project, University of Amsterdam. The project develops Web Archive Retrieval Tools to critically examine existing Web archives as an object of study of scientific research and in relation to preservation of future cultural heritage. She holds a PhD from the Science, Technology and Society program, Bar-Ilan University. Anat is affiliated with the Digital Methods Initiative, University of Amsterdam, which develops methods and tools for Web studies.

Archiving web material for future research?

Ditte Laursen, senior researcher and curator at the State Media Archive, State Library in Denmark

The Danish Netarchive has captured Danish websites since 2005 and is one of the largest webarchives in the world. In this presentation, I will talk about our curational experiences with archiving the Danish internet over the years, with a particular focus on the challenges we face in archiving new kinds of electronic network content, like apps, social media sites, interactive games etc. I will also talk about a combined webarchiving/research project in which we experimented with different kinds of tools to capture the synchronous interplay between real-time internet and live television in cross-media productions like Voice and X factor. Finally, it is discussed how webarchiving in national web archives affect the possibilities for writing historical studies of web material in the future, including how specific methods of web archiving provide both limited and privileged access to the original data, and accordingly, frame what research questions can be answered.

Ditte Laursen, PhD, is senior researcher and curator at the State Media Archive, State Library in Denmark. Participating partner in RESAW (Research infrastructure for the Study of Archived Web materials), IIPC (International Internet Preservation Consortium), and DigHumLab (Digital Humanities Lab Denmark). Post.doc. 2009-2011 at DREAM – Danish Research Centre of Advanced Media Materials with a project on digital technologies in museums. Research interests are digital archiving, technologically mediated communication and social interaction in, around and across digital media.

Probing a nation's web sphere

Niels Brügger, Associate Professor, Aarhus University, Head of the Centre for Internet Studies

With a view to discussing and illustrating one of the possible roles of the web within historiography this presentation will focus on an aspect of the web which media, communication, and internet historians (and historians in general) does usually not consider either an object of study or a historical source, namely a nation's web sphere: What does 'a nation's web' actually look like? And how to study a nation's web sphere and its developments over time?

In this presentation some of the methodological challenges that this new type of historical study and historical source material puts on the research agenda are discussed, and an analytical design is outlined. The presentation will use Denmark and the material in the national Danish web archive Netarkivet a case. However, all the debated methodological challenges are generic and will therefore in the main apply to any country.

Niels Brügger is Associate Professor, Aarhus University, and Head of the Centre for Internet Studies. His primary research interests are web archiving, web historiography, and Digital Humanities. He has edited the books *Web History* (Peter Lang, 2010) and *Histories of Public Service Broadcasters on the Web* (co-ed. with M. Burns, Peter Lang 2012), and he has published monographs and a number of edited books as well as articles about web history and web archiving.

The seminar is organized by Niels Brügger, the Danish Digital Humanities Lab/NetLab & Aarhus University, the PhD programme 'ICT, Media, Communication and Journalism'.

netlab.dk