



# NetLab Forum meeting

27 February 2018, 13.00 – 15.15, room 295 in the Nygaard-building

13.00 - 13.45	“If These Crawls Could Talk: Studying and Documenting Web Archives Provenance” (abstract below)	Emily Maemura, PhD Candidate, University of Toronto, Faculty of Information, guest researcher at NetLab
13.45 - 14.15	“The Study of Archived Web: User perspectives” (abstract below)	Maria Costea (author), presented by Niels Brügger, NetLab
14.15 - 14.30	“Experiences from NetLab Online Courses: Web Archive Readiness”	Asger Harlung, NetLab
14.30 - 14.45	“Status and experiences from a DeIC Cultural Heritage Cluster pilot project”	Janne Nielsen, NetLab
14.45 - 15.15	Current projects – what are we working on?	Everyone (voluntary)

---

## “If These Crawls Could Talk: Studying and Documenting Web Archives Provenance” / Emily Maemura

This talk will present preliminary research on the challenges to systematically document provenance information and capture the curatorial choices that shape materials in a web archives collection. I summarize work developed and presented over the past year in collaboration with Ian Milligan from the University of Waterloo as part of his fellowship with the University of Toronto’s Digital Curation Institute, web archivist and librarian Nich Worby from University of Toronto Libraries, and my doctoral supervisor Christoph Becker.

Through an analysis of select cases of web archives creation and use, we have identified a set of key factors that influence what material is selected and collected from the web, then cleaned and refined for the purposes of research. These elements are summarized in a preliminary framework that can provide curators with guidance on which choices to document, as well as provide researchers with a set of questions to ask that will help characterize and understand their data. The examples from these cases also highlight the interactions between individual decisions, and how the wider organizational, social, technical and legal contexts impact the choices made for a collection. This work will serve as my starting point to explore the activities, methods, and systems used by Netlab and the Royal Library as I study the work with Netarkivet during my time as a guest researcher in Aarhus.

---

## “The Study of Archived Web: User perspectives” / Maria Costea (author), presented by Niels Brügger

In the Autumn of 2017 Maria Costea who studies Sustainable Heritage Management was an intern at NetLab where she examined the researcher use of web archives. The final report from the internship is just finished, and it will be published 27 February on [netlab.dk](http://netlab.dk). Maria’s findings will be presented at the NetLab Forum meeting on the same date. The following paragraph is a paste of the abstract from the report:

**Purpose:** For many years, web archiving communities have dedicated much of their time and effort to developing the necessary archiving technology and procedures. With the better establishment of archiving infrastructure, their focus has now increasingly turned towards enhancing web archive use and usability. In support of these efforts, this paper’s aim is to provide an empirical perspective of how researchers currently engage with web archives, their needs in connection to this source, as well as to identify some of the reasons for their non-use.

**Methodology:** Data was collected using a mixed method approach, which consisted of an online survey in two Danish universities, and a series of semi-structured interviews.

**Findings:** Findings suggest that currently there is limited awareness in the humanities and social sciences research communities of the possibilities of using web archives as a scholarly source. At the same time, the

results describe researchers' various use methods, some of the challenges they are facing when using the archives, as well as their suggestions for potential improvements.

**Research limitations:** Limitations address sample size and sampling method. Further research is necessary in order to obtain more robust and representative results.

**Originality/value:** This study aims to contribute to the presently limited research on the use of web archives in general, and their scholarly use in particular.

---

### **Will you contribute to research?**

Emily Maemura would like to conduct interviews on how Netarkivet is used for research, including the different kinds of tools, methods and data used. Interviews will be of approximately one hour, and may be individual or group interviews. Please contact Emily at [e.maemura@mail.utoronto.ca](mailto:e.maemura@mail.utoronto.ca) if you are interested in contributing.