

## NetLab workshop on web archiving

Do you work with web material in your research? Do you need to archive web material to preserve your research object? Or do you wish to use material that has previously been online, but is now only available in a web archive?

Perhaps you are doing research into history or society, and are working with an event that has been intensively covered online? Maybe you are researching online fan culture? Or online works of art? Perhaps you are researching language use in Internet forums? Maybe you are studying social media, and are interested in finding out how to archive that type of data? Perhaps you want to save parts of a website, so that, for one reason or another, you can study it and document your observations?

### Purpose

If so, NetLab's research workshop on web archiving will be useful for you. The purpose of the workshop is to give you an introduction to working with archived web material in research, including:

- Insight into the challenges of working with archived web material as a research object
- A presentation of the various ways you can archive web material yourself – and hands-on experience of using various tools
- Knowledge of existing web archives, and of how these can be used in research

### About NetLab

NetLab is a national research infrastructure project for the study of web materials, under the auspices of Digital Humanities Lab Denmark (DIGHUMLAB). NetLab's task is to spread knowledge among researchers and students about the use of research infrastructure in connection with studies involving web materials.

The teachers at the workshop are NetLab employees:

- Niels Brügger, head of NetLab and professor with special assignments at Media Studies, Aarhus University
- Ulrich Have, IT architect, NetLab and Aarhus University
- Asger Harlung, research assistant, NetLab and Aarhus University
- Janne Nielsen, assistant professor, NetLab and Aarhus University



## Structure

The workshops are structured around several available modules so each workshop can be adapted to your needs and wishes in relation to your field. Each module has a main topic and an approximate duration (depending on the number of participants, and with the possibility of scaling). Module 1 will usually be included in all workshops, but apart from that there is great flexibility, and the modules can be combined in various ways.

### MODULE 1 Web Archiving

An introduction to web archiving and to archived websites as research material. The module addresses the following topics:

- What is web archiving?
- The main types of archiving, such as web crawling of entire websites, screen recording and screen captures
- Archiving strategies
- Characteristics of the archived material
- Ethical considerations in research
- Referencing archived web materials
- Examples of research projects that use archived web materials

*Duration:* 45-60 min.

*Preparation:* Read the relevant chapter in *Using web archives in research – an introduction* (Nielsen, 2016).

### MODULE 2 Existing Web Collections

An introduction to existing web collections. The focus is on the contents of the collections, the archiving strategies applied, access, search options, documentation, and possible use by researchers. The module deals with the following web archives:

- Netarkivet, the Danish web archive
- The Internet Archive
- Other national archives
- Other archives and collections

*Duration:* 45-60 min.

*Preparation:* Read the relevant chapter in *Using web archives in research – an introduction* (Nielsen, 2016)



**MODULE 3**  
Creating Your  
Own Collections

An introduction to tools that you can use to create your own web archives. The focus is partly on software that you can install on your own computer, and partly on professional on-demand archiving services. The module deals with software that can be used for:

- Archiving entire websites
- Screen recording
- Screenshots and/or saving multiple pages as a PDF, etc.
- Archiving social media (e.g. Facebook, Twitter, Instagram) using API

*Duration:* 45-60 min.

*Preparation:* Read the relevant chapter in *Using web archives in research – an introduction* (Nielsen, 2016).

**MODULE 4a**  
Mini-project

The benefit of the workshop is increased if there is a direct link to a specific research project. In this and the following two modules, work is done with the participants' own mini-projects. You can either work with an ongoing or a planned project involving web materials, or one which will benefit from web materials being added, or a pilot project created for the occasion. It may be that only some of the participants choose this module. The following topics are covered:

- Presentation of the participants' mini-projects
- A discussion of the potential feasibility of the projects

*Duration:* 30 mins.

*Preparation:* Prior to the workshop, prepare a short project (1/2 -1 page) on a research project that will use archived web materials, either in existing web archives or in your own archive. The projects can be created individually or in groups.



**MODULE 4b**  
Mini-project  
Follow-up

This module follows up on module 4a, and is therefore held an agreed number of weeks after module 4a, so that it can take place while work is ongoing on the participants' mini-projects. The module will cover:

- What went well/badly?
- What are the next steps?

The module takes the form of a video meeting.

*Duration:* 45-60 min.

*Preparation:* Experience with working on a mini-project.

**MODULE 4c**  
Mini-project  
Reporting In

In this module, participants report on their mini-projects. This module takes place an agreed number of weeks after module 4b, so that it can be held after the participants have worked on their mini-projects. In the module:

- The mini-projects are presented and discussed.
- Experience is shared.
- The work with the archived web materials is evaluated.

*Duration:* 45-60 min.

*Preparation:* Completed mini-project.



## Examples of module combinations

### Entire package

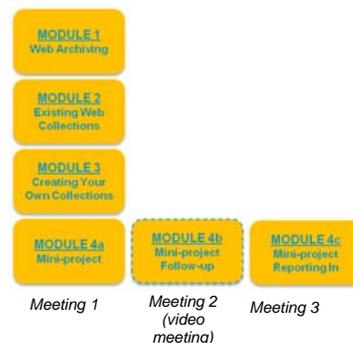
Introduction to existing web archives, the establishment of a private archive, and the implementation of your own mini-project, based either on a specific use of web archives or a private archive, or a combination of these.

*Duration, meeting 1:* 165-210 min.

*Duration, meeting 2 (video):* 45-60 min.

*Duration, meeting 3:* 45-60 min.

The meetings can be several weeks apart depending on the wishes of the participants.



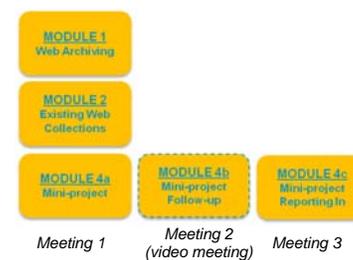
### Existing collections and own project

Focus is on the use of existing web archives and the implementation of your own mini-project, based on a specific use of web archives.

*Duration, meeting 1:* 120-150 min.

*Duration, meeting 2 (video):* 45-60 min.

*Duration, meeting 3:* 45-60 min.



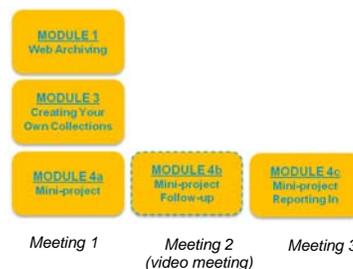
### Own archive and own project

The focus is on the establishment of your own web archive and the implementation of a mini-project, based on a specific use of this archive.

*Duration, meeting 1:* 120-150 min.

*Duration, meeting 2 (video):* 45-60 min.

*Duration, meeting 3:* 45-60 min.



### Existing collections

Focus is on the use of existing web archives.

*Duration:* 90-120 min.



### Establishment of own web archive

Focus is on the establishment of your own web archive.

*Duration:* 90-120 min.



### Existing collections and own archive

Focus is on the use of existing web archives and the establishment of your own web archive.

*Duration:* 135-180 min.



## Course materials

*Using web archives in research – an introduction* (Nielsen, 2016) is used as the textbook for the NetLab workshops. It is written in connection with the workshops and makes available knowledge about the use of web archives for research (especially in a Danish context and adapted to the needs of Danish researchers, but can also be used by others) The book is available free of charge as a PDF in advance of the workshop, and it follows a similar structure as the modules. Links to the teaching materials and the software used on the course are available at the website <http://www.netlab.dk/courses/workshops/>.

## Practical information

We seek to ensure the relevance of the workshop for scholars, so we wish to put together a workshop that focuses on the elements that are most relevant to you in your field. The modules can also be spread over several meetings. A workshop requires a minimum of four participants, and will be held locally at your premises.

In order to participate in the modules, which utilises the Danish national web archive Netarkivet (module 2 and possibly module 4), you must be a researcher, and you must have applied for and have been granted access to the archive. This is straightforward, and we can show you how to do it.

A workshop for researchers can later be supplemented by a workshop for your students, so that you can apply the methods in your teaching. Workshops for students are held separately, as these will not necessarily involve the same modules.

If you wish, we will be happy to come and provide a brief introduction to the ideas behind the workshop, for example at a meeting with your research group or research programme.

For further information, please contact Niels Brügger ([nb@cc.au.dk](mailto:nb@cc.au.dk)) or Asger Harlung ([asger@cc.au.dk](mailto:asger@cc.au.dk)).

