

---

# What can we know about a website? Studying web archive content by mixing methods and data types

Niels Brügger\*<sup>1</sup>, Hans-Peter Degn , Line Hassall Thomsen , Maja Sonne Damkjær ,  
Christiane Særkjær , and Katharina Sølling Dahlman

<sup>1</sup>Aarhus University [Aarhus] – Nordre Ringgade 1 DK-8000 Aarhus C, Denmark

## Abstract

This paper aims at answering the question "What can we know about a website?" by discussing some of the methodological issues related to mixing methods and data types - including the archived web -, and by providing empirical results from a research project where one website was analysed from different angles.

In 2008 the Danish website *videnskab.dk* was founded, and in the following years it became a leading journalistic website that disseminates research related knowledge to the wider public in the form of articles, interviews, and the like. *videnskab.dk* is the Danish equivalent of *scientificamerican.com* (US), *futura-sciences.com* (France), and *scinexx.de* (Germany).

In 2022 the funder of the website wanted to have the website evaluated and the authors of this paper were commissioned to perform this evaluation which we decided to do by performing a multi-faceted analysis based on a variety of digital and non-digital methods and sources, some of which were historical material from the national Danish web archive Netarkivet as well as from the Internet Archive.

## The analytical design: Mixing digital and non-digital methods and data types

The project includes analyses of the website itself, of the staff producing the website, of different types of readers, and of the website's media context on the web to which the website relates through hyperlinks, and as to the latter a historical dimension was added. The following data types and methods were used:

- qualitative analysis of content and forms of expression on *videnskab.dk*'s website (genre, design, functionality, sender/receiver characteristics, and journalistic communication), on *videnskab.dk*'s social media, and in the website's Newsletter (including different versions, and how they are referred and adopted in other media),
- observation/field study and interviews with management and staff at *videnskab.dk*,
- interviews with researchers who have provided content to *videnskab.dk*,
- analysis of the use(rs) of *videnskab.dk*, based on panel data from Gemius, a research and technology company providing data about internet use, and traffic from the website itself,
- interviews with editors from some of the media that often publish articles based on content from *videnskab.dk*,

---

\*Speaker

- survey and interviews with teachers and pupils (elementary and high school),
- analysis of mentions of *videnskab.dk* in Infomedia (a database of all Danish media),
- network analysis of hyperlinks extracted from Netarkivet, from the period 2009-2022,
- network analysis of outgoing hyperlinks collected from a variety of international websites of similar type as *videnskab.dk*, that is journalistic websites that disseminate research.

In terms of methods the paper discusses:

- the challenges related to performing historical hyperlink network analyses of the archived web (e.g. how to establish the dataset, what to extract, what to do with temporal inconsistencies?), as well as the advantages (e.g. being able to include hyperlinks pointing at *videnskab.dk* and thereby make a genuine network analysis),
- how best to have the different methods interact (should they just be used in a compiling way, like pearls on a string, or should they inform each other, having (preliminary) results from different analyses inform one another thus providing information that can open new avenues for all elements of the analysis).

Regarding the latter point focus is on how the historical network analyses of the archived web can inform the other elements of the project about the role of *videnskab.dk* on the web, including who are the most central actor types who link to the website, and how this has changed over time.

### **Results**

The analyses have been performed and the results are being compiled into a report to the funder of *videnskab.dk* to be handed in December 2022. Therefore, results cannot be presented in this abstract, but will be available at the RESAW conference i June 2023.

**Keywords:** website mixed method