

A NATIONAL WEB TREND INDEX

Niels Brügger, Aarhus University, DK
Ditte Laursen, Royal Danish Library
20190606 IIPC, Zagreb, Croatia



**DET KGL.
BIBLIOTEK**
Royal Danish Library



AARHUS UNIVERSITY

Web trend index

- Web (not on the internet or computers in general)
- Trend (regular time intervals)
- Index (metrics)
 - could have two main components
 - a *standard index* with statistical information about size of the web domain, number and size of websites, file types, software types, response codes, number of words, etc.
 - an *extended index* with link graphs, linguistic analyses, analyses of where specific topics are mentioned, analyses of the use of Named Entities (locations, names, etc.), etc.



Why a national Web Trend Index?

—researchers and society at large need reliable and systematic information about the development of the web

—to be used in studies of the web's development, and in studies of 'x', where the web is a part of the study



**DET KGL.
BIBLIOTEK**

Royal Danish Library



AARHUS UNIVERSITY

Why a national Web Trend Index?

- focus is on 'the available web', that is the web that was available for users to use
- different from other existing statistical annual information where focus is usually on user statistics, and existing statistics have a very limited scope (e.g. number of web domains included is often quite small)



Method

- based on the collections of a national web archive, potentially supplemented with material from the Internet Archive
- must be created in a systematic, transparent and reproducible way to enable comparisons over time
- must be made on the basis of a reconstruction of what was once online (cf. Brügger et al. 2018, on corpus creation)
- corpus algorithm may have to change year by year as the online web and the web collection changes



Organisation

- joint venture between archiving institution and research institution(s)
 - web archives have the needed knowledge about how to unlock the collection (curatorial, IT-developers...)
 - researchers have knowledge about how to establish an annual corpus that can be used for studies, and they can help ensure the relevance of the measure points of the index
- must be based on a close and sustainable collaboration to ensure continuous adjustments and improvement



Organisation

the web archive can gain:

- collections made 'available' in a way that is not violating copyright and privacy legislation
- potentially get in contact with more users
- get research based insights in their collections

participating researchers can gain:

- access to reliable and aggregated statistics
- ensure that the Web Trend Index is based on research methods
- can bring new knowledge to their domain knowledge

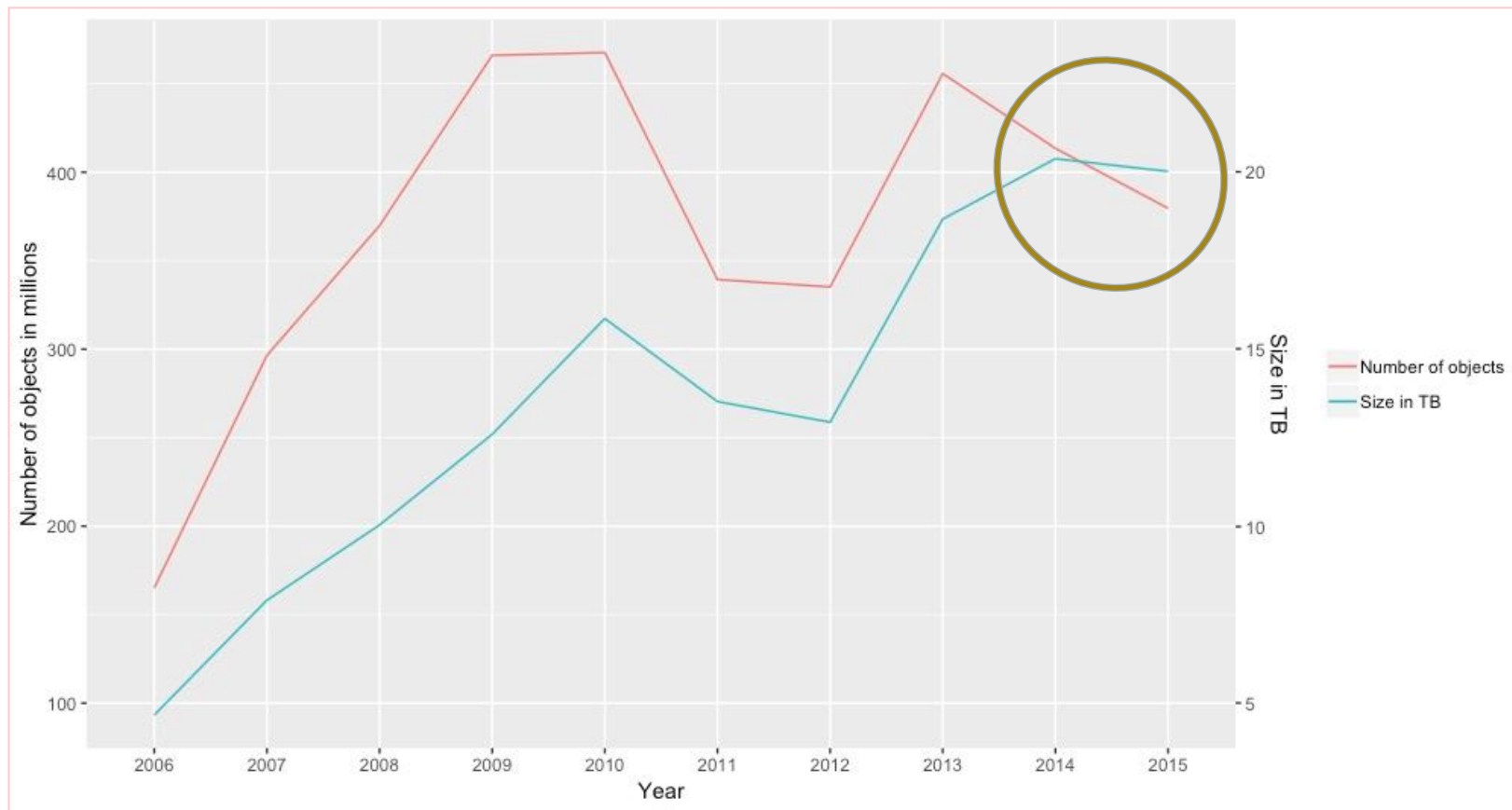


Organisation – the Danish case

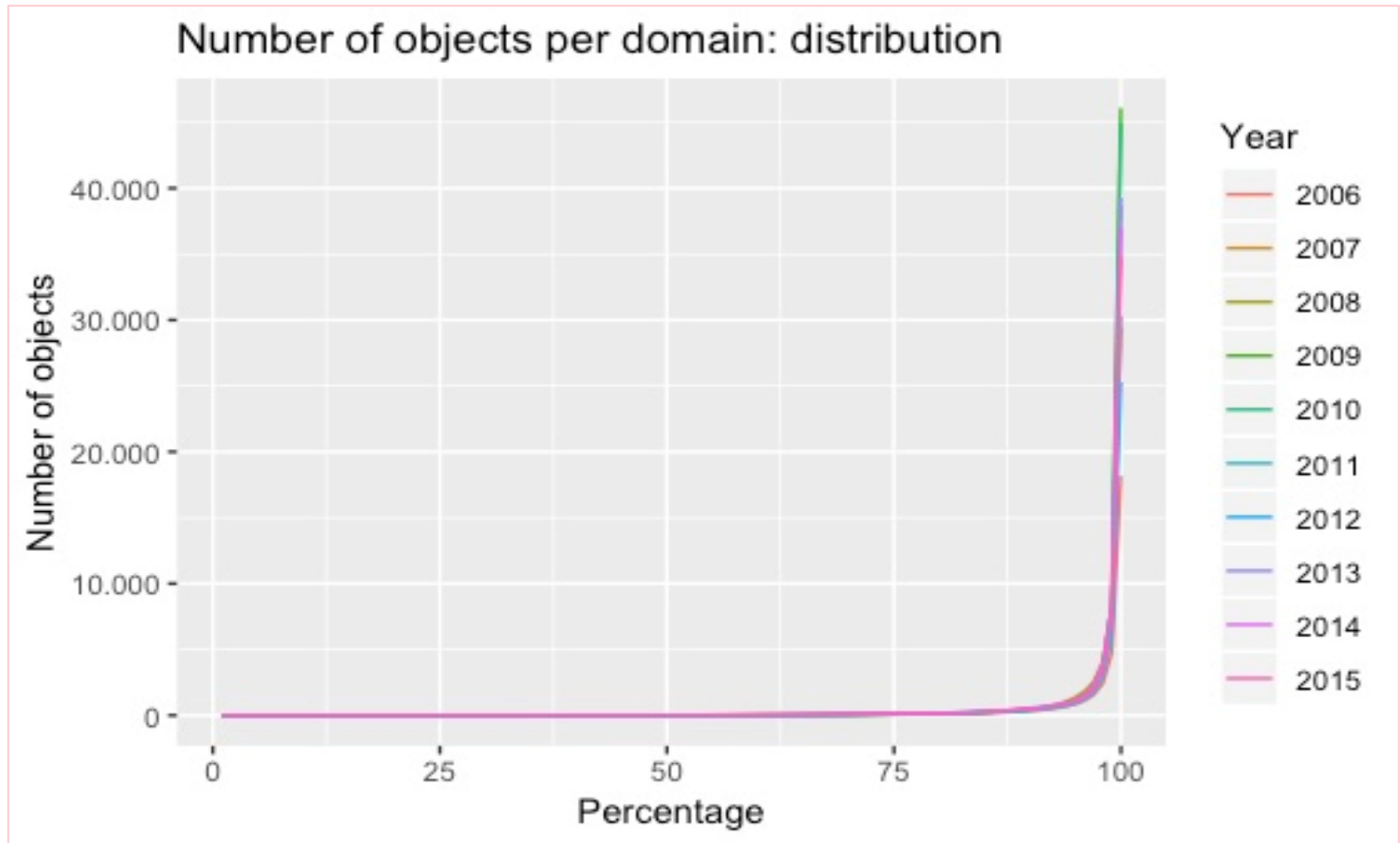
- the Danish case: drafted a 'Note about a national Web Trend Index' with information about the idea, planning, organisation, and funding
- pilot phase has been approved by the Royal Library, approval from Aarhus University is still pending
- Corpus algorithm has been created (as part of a more general research project)
- metrics to be decided



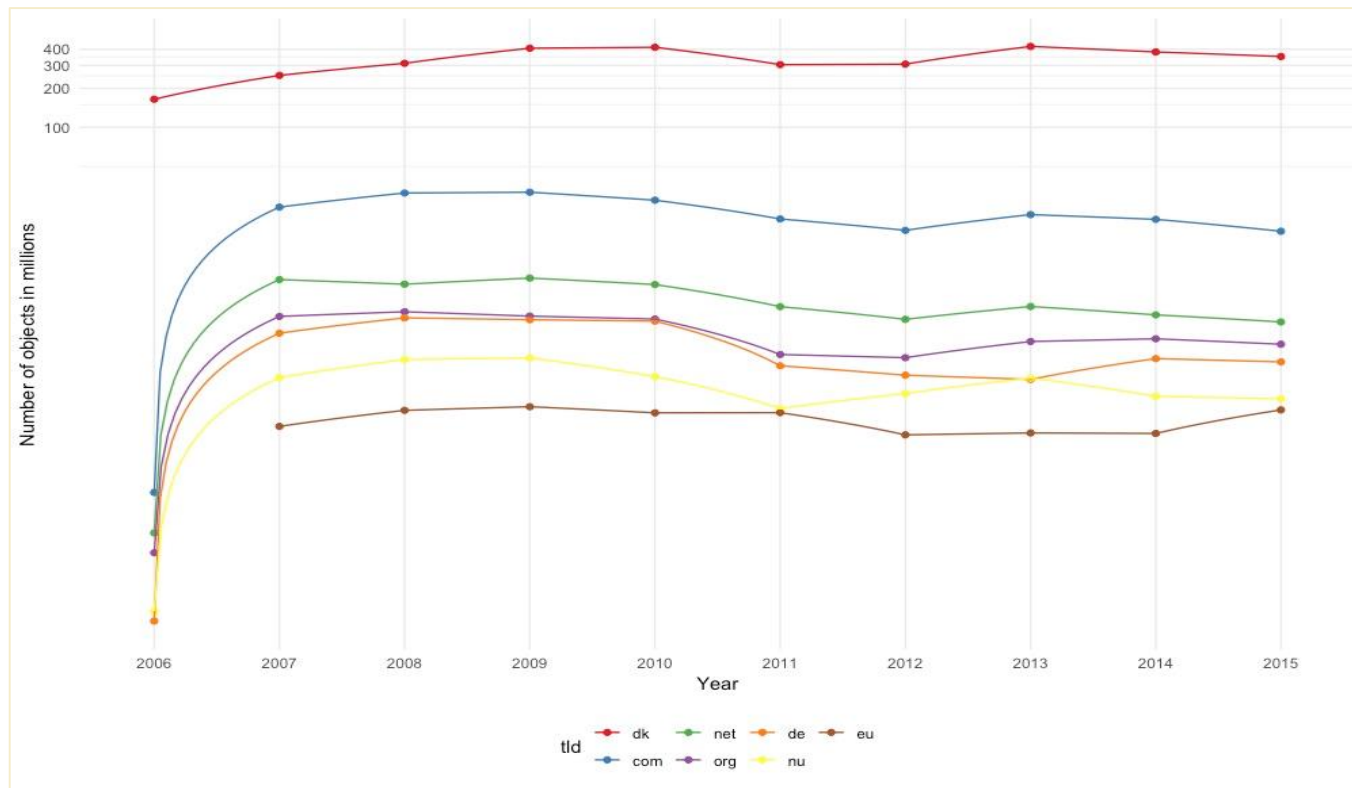
What is the size of the Danish web?



What is the size of Danish domains?



What is the size of the web Danica outside the Danish ccTLD, and which TLDs does the material live on?



Why national?

- basically because the most comprehensive collections are national
- the national Web Trend Index can serve as the template for cross-national comparisons — can shed new light on participating countries and on transnational trends
- challenges for making it international: the heterogeneity of the different web archive collections on which the index will be based



Read more

- Brügger, N. (2017). Probing a nation's web domain: A new approach to web history and a new kind of historical source. In G. Goggin & M. McLelland (Eds.), *The Routledge companion to global internet histories* (pp. 61–73). New York, NY: Routledge.
- Niels Brügger (2019). A national Web Trend Index. In N. Brügger & D. Laursen (Eds.), *The historical web and digital humanities: The case of national web domains* (pp. 178–187). Abingdon: Routledge.
- Brügger, N., & Laursen, D. (2018). Historical studies of national web domains. In N. Brügger & I. Milligan (Eds.), *The SAGE handbook of web history* (pp. 413–427). London: Sage.
- Brügger, N., Laursen, D., & Nielsen, J. (2018). Establishing a corpus of the archived web: The case of the Danish web from 2005 to 2015. In N. Brügger & D. Laursen (Eds.), *The historical web and digital humanities: The case of national web domains* (pp. 124–142). Abingdon: Routledge.



A NATIONAL WEB TREND INDEX

Niels Brügger, Aarhus University, DK
Ditte Laursen, Royal Danish Library
20190606 IIPC, Zagreb, Croatia



**DET KGL.
BIBLIOTEK**
Royal Danish Library



AARHUS UNIVERSITY