

# NATIONAL WEB ARCHIVES — THE LAND OF PROMISE FOR RESEARCHERS

---

NIELS BRÜGGER  
PROFESSOR, HEAD OF THE CENTRE FOR INTERNET STUDIES,  
AND OF NETLAB



# AGENDA

---

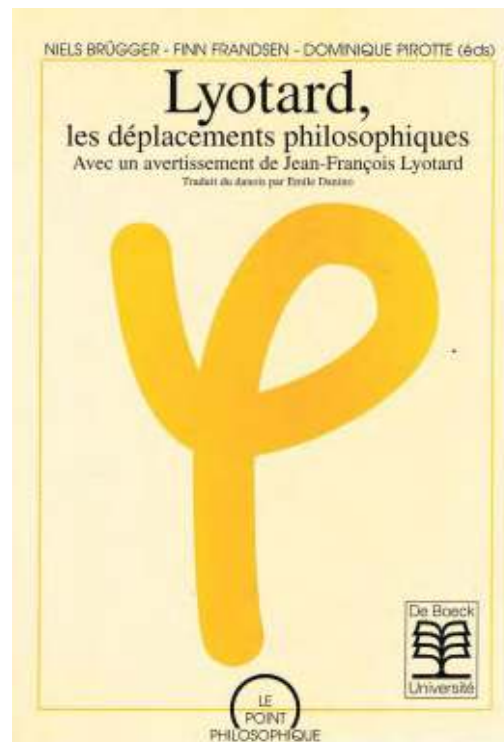
1. Why national web archives matter — a personal story
2. Mapping different web archiving forms
  - a) the US vs. Europe
  - b) mapping trends in Europe
3. Impact on history writing — the case of national webs
4. Belgium — the land of PROMISE

# WHY NATIONAL WEB ARCHIVES MATTER — A PERSONAL STORY

---

From French philosophy to  
Belgian beer — and web  
history.

*Lyotard, les déplacements  
philosophiques* (Eds. N.  
Brügger, F. Frandsen, D.  
Pirotte). Bruxelles: De  
Boeck-Wesmael, 1993.



# WHY NATIONAL WEB ARCHIVES MATTER — A PERSONAL STORY

De Objectieve  
Bierproevers — Les Taste-  
Bière objectifs



In 2002 became Zythos.

Survived in some regions  
— Essen, 'Objectieve  
Bierproevers van de  
Essense Regio'

› I would like to study the  
history of the website of  
'De Objectieve  
Bierproevers', 1993-2002

# WHY NATIONAL WEB ARCHIVES MATTER — A PERSONAL STORY

No Belgian web archive —  
try the Internet Archive.

Success — many copies  
found from 2000-01.

Consequence: if we want  
to study the history of the  
Belgian web we have to  
go to the US.



obp.be, 17 Oct 2000, the Internet Archive

## 2. MAPPING DIFFERENT WEB ARCHIVING FORMS

---

- › the US vs. Europe
- › mapping trends in Europe

## 2A) THE US VS. EUROPE

---

## 2A) THE US VS. EUROPE

---

- › legal deposit or not
- › independent, non-profit organisation like the Internet Archive ([archive.org](https://archive.org)), the biggest collection worldwide
- › one national actor, the Library of Congress
- › several other initiatives, at universities, museums, etc.
- › cumulative archiving vs. based on national borders

Exceptions: Portugal, and the Netherlands, not based on legal deposit



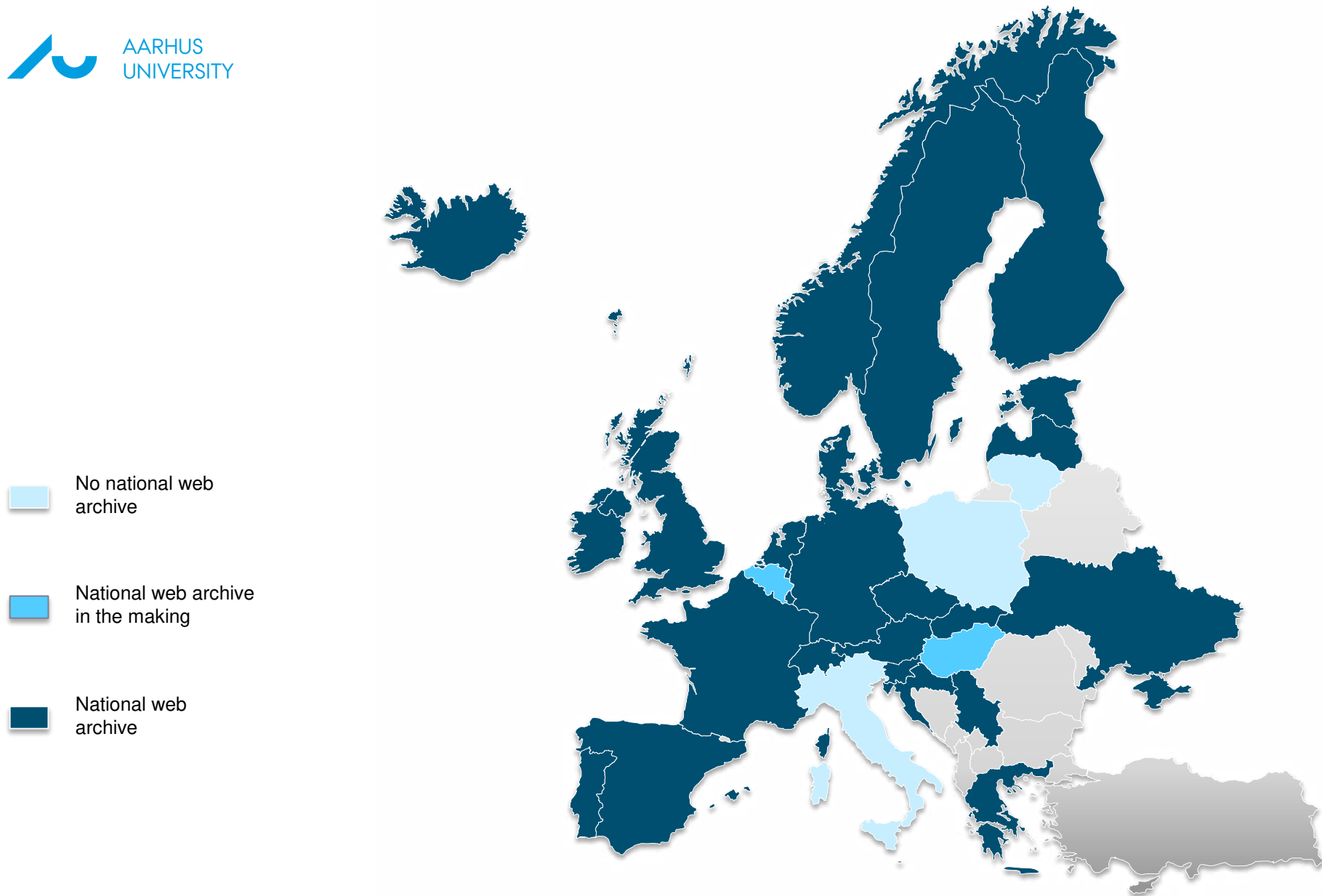
## 2B) MAPPING TRENDS IN EUROPE

---

The 'have' and the 'have-nots':

- › European countries without a national web archiving initiative, e.g. Poland and Italy
- › national web archives in the making — PROMISE in Belgium, a pilot project in Hungary
- › countries with a national web archive

The specific challenge of transnational web domains:  
.eu — pilot project run by the Portuguese Web Archive



## 2B) MAPPING TRENDS IN EUROPE

---

The 'have' and the 'have-too-many':

- › some European countries have more than one national web archive — examples:
  - › UK: the UK Web Archive, The UK Government Web Archive, The UK Parliament Web Archive
  - › France: BnF, and the Institut National de l'Audiovisuel

## 2B) MAPPING TRENDS IN EUROPE

---

The different archiving strategies:

- › intention to archive an entire national web domain
  - › based on list of web domains of the ccTLD only, e.g. UK since April 2013
  - › list of web domains of the ccTLD supplemented with manual tracking of material outside ccTLD such as .com, .net, etc., e.g. Denmark

## 2B) MAPPING TRENDS IN EUROPE

---

The different archiving strategies:

- › intention is to archive a selection of the national web domain — actors, time frame, topic, valuable cultural heritage, e.g. Dutch

## 2B) MAPPING TRENDS IN EUROPE

---

The different archiving strategies:

- › differences within the same web archive — combination of two or more strategies, e.g. the Danish case: entire national domain, a selection of rapidly changing websites, and websites in relation to selected events
- › differences over time, e.g. the UK Web Archive, 2005 selected websites of historical, social and cultural significance — from 2013, entire web domain

## 2B) MAPPING TRENDS IN EUROPE

---

The different access conditions — legal issues:

- › free online access for everyone, e.g. Portugal, the Icelandic Web Archive
- › for researchers only:
  - › onsite only, e.g. Netherlands
  - › onsite, but distributed, e.g. UK (national libraries), France (regional libraries)
  - › online, e.g. Denmark
- › no access, e.g. the Norwegian web archive, the National Library of Ireland

## 2B) MAPPING TRENDS IN EUROPE

---

The different access conditions — technical issues:

- › Wayback view, presents the web page in a manner close to how it looked when online — commonly used access form in most national web archives





# De Digitale Bierproever



## HOP & Poperings Bierproeversgenootschap Presenteren

Menu	Welkom bij de Objectieve Bierproevers	
OBP vzw	Welkom op de website van OBP	Websitenieuws:
Werkgroepen		
Acties	OBP zijn de Objectieve Bierproevers, een vzw ter bescherming van de Belgische Biercultuur	<b>10/2000</b> Een nieuwe website voor OBP. Graag uw <a href="#">reactie</a>
Verenigingen	<a href="#">&gt;&gt;&gt;Meer Info</a>	<b>Den Biersurfer</b> Neem je Bierproever in de linkerhand, je muis in de linkerhand en surf achter bier op het internet.
OBP Biertrofee		Alle linken uit Den Bierproever voor u verzameld op 1 plaats.
24-Uur		<a href="#">&gt;&gt;&gt;Meer info</a>
Dossiers		
Den Bierproever	<b>Agenda</b>	<b>Biernieuws</b>
Den Biersurfer	<b>14-15 oktober 2000:</b> <a href="#">24 uur van het Belgische bier</a> Antwerpen-Meir	<ul style="list-style-type: none"> <li>• Interbrew naar de beurs</li> <li>• Riva betaalt milieuboete</li> <li>• Waarom daalt de prik van Guinness</li> <li>• Prijskartels Interbrew toegegeven</li> <li>• Bockor verhoogt capaciteit</li> <li>• ....</li> </ul>
Bieragenda	<b>26-27-28 oktober 2000:</b> <a href="#">13de Heikantse Bierfeesten</a> Familia-Heikant	<a href="#">&gt;&gt;&gt;Meer info</a>
	<b>28-29 oktober 2000:</b> <a href="#">Karakterbierfestival</a> Hotel Palace-Poperinge	
	<b>10-11-12 november 2000</b> <a href="#">Weekend der Belgische Bieren</a> De Schakel-Hasselt	
	<b>9-10 december 2000</b> <a href="#">6de Kerstbierfestival</a> Parochiecentrum-Essen	<b>Standpunten</b> Onze vereniging heeft tot doel het organiseren en stimuleren van projecten die het kenbaar maken van de Belgische bieren en hun smaken bevorderen, het aanklagen van onregelmatigheden ten overstaan van de consument, en het behouden van de Belgische biercultuur.
	<a href="#">&gt;&gt;&gt;Meer info</a>	
	<b>Lidmaatschap</b> Voor slechts 500 Bfr (= 12.39 €) wordt je een	

### Niet te missen



### Interessant

[Gastenboek](#)

[Contacteer ons](#)

Zoek op de website

## 2B) MAPPING TRENDS IN EUROPE

---

The different access conditions — technical issues:

- › Wayback view, presents the web page in a manner close to how it looked when online — commonly used access form in most national web archives
- › open API, e.g. The Portuguese Web Archive
- › prepared datasets, e.g. “JISC UK Web Domain Dataset (1996–2013)” in the UK
- › extraction, taking material out of the web archive and handing over to researchers, e.g. Denmark

## 2B) MAPPING TRENDS IN EUROPE

---

In summary, mapping trends in Europe:

- › 'have', the 'have-nots', the 'have-too-many'
- › different archiving strategies
- › different access conditions

How to make transnational European historical studies based on web archives?

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

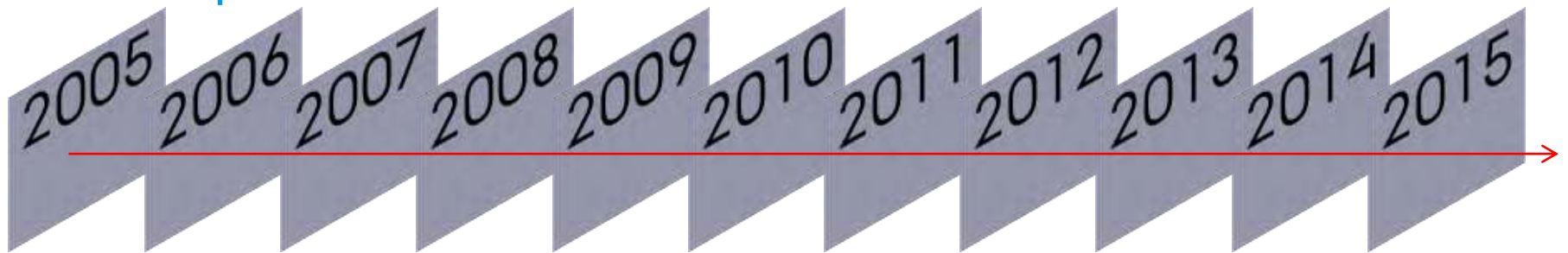
- › not unusual that national collections are not made in the same ways accros borders
- › never the less these differences between different national web archives have a great impact on future studies — in particular cross-national studies

Case: national web domains — the historical development of the Danish web, compared to other national web domains.

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

Aim of "Probing a Nation's Web Domain":

- › analyse the historical development of the entire Danish web domain — using a variety of 'probes'
- › develop Big Data methods on the archived web
- › establish methods to create a corpus
- › a blueprint for web archive studies at a smaller scale



# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

- › Data: material from the national Danish web archive Netarkivet
- › Hardware: the DelC National Cultural Heritage Cluster at the Royal Danish Library, computes 70TB
- › The project team: Janne Nielsen (Aarhus University, NetLab), Ulrich Have (Aarhus University, NetLab), Ditte Laursen (Royal Library), Per Møldrup-Dalum (Royal Library), Asger Askov Blekinge (Royal Library), and me
- › Project start: 2014!

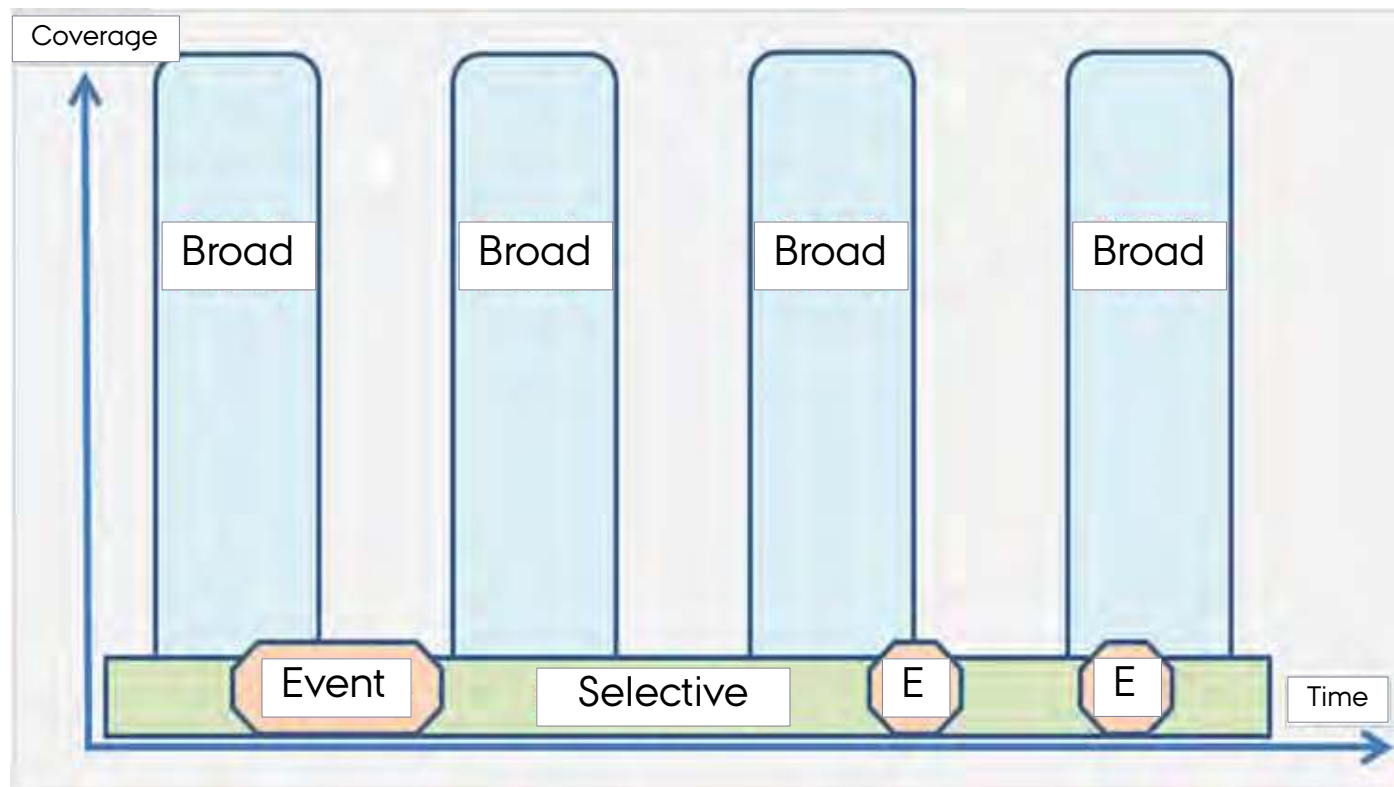
# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

Research questions:

- › What is the size of the Danish web?
- › How much written text and how many images are there on the Danish web?
- › Which are the most popular social media on the Danish web?

Strategies:  
 > Broad/bulk  
 > Selective  
 > Special collection/event



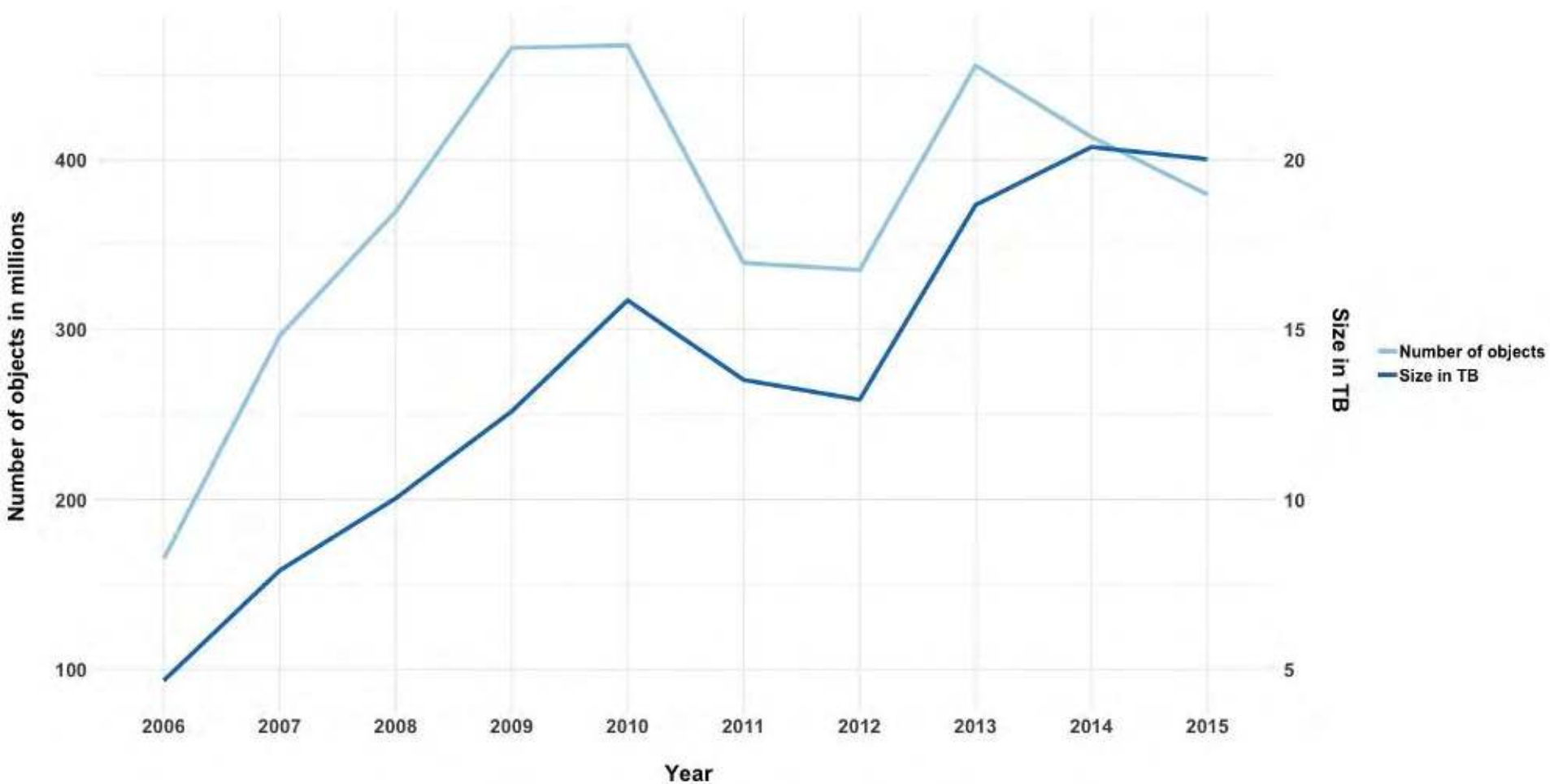
*From <http://netarkivet.dk/om-netarkivet>*

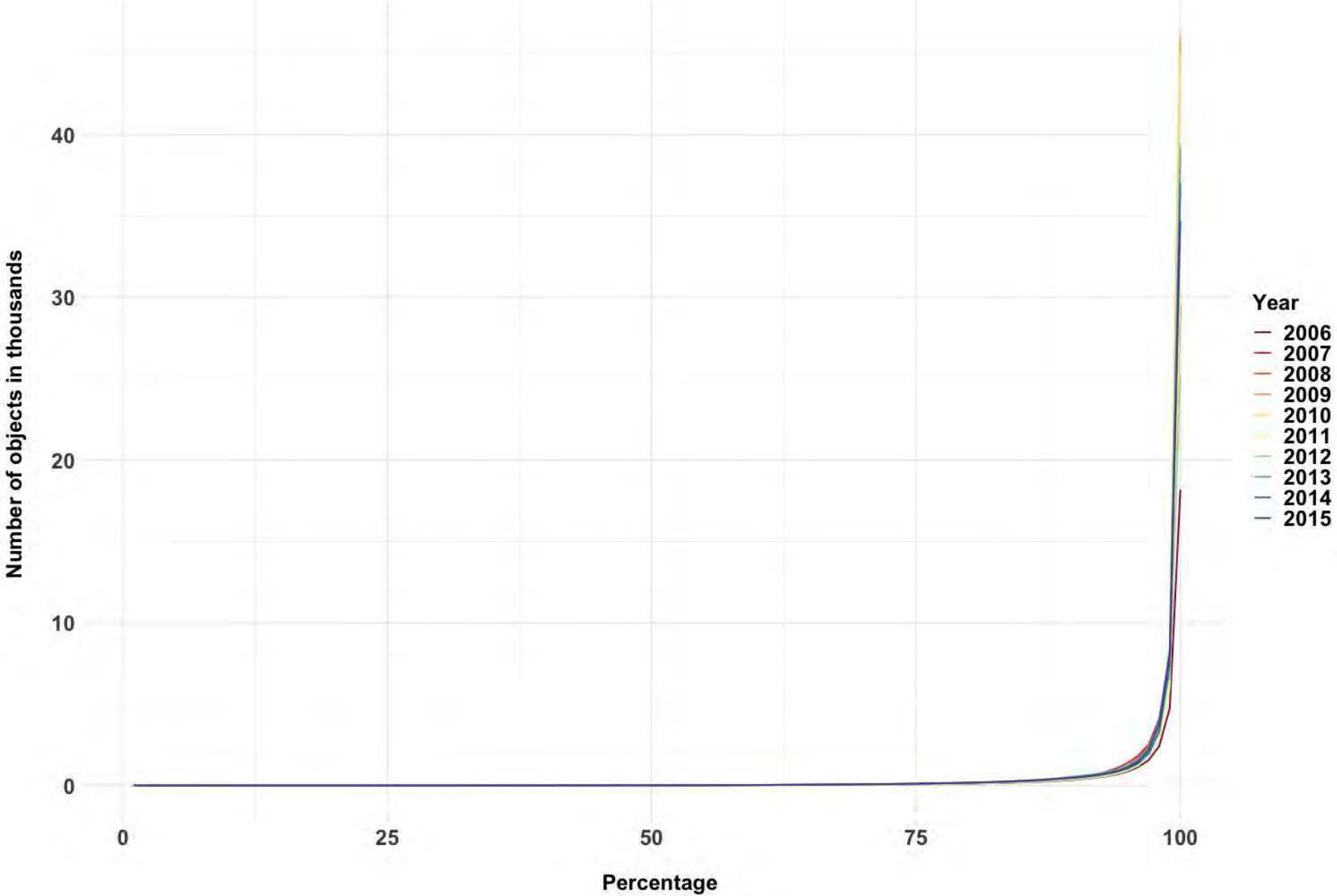


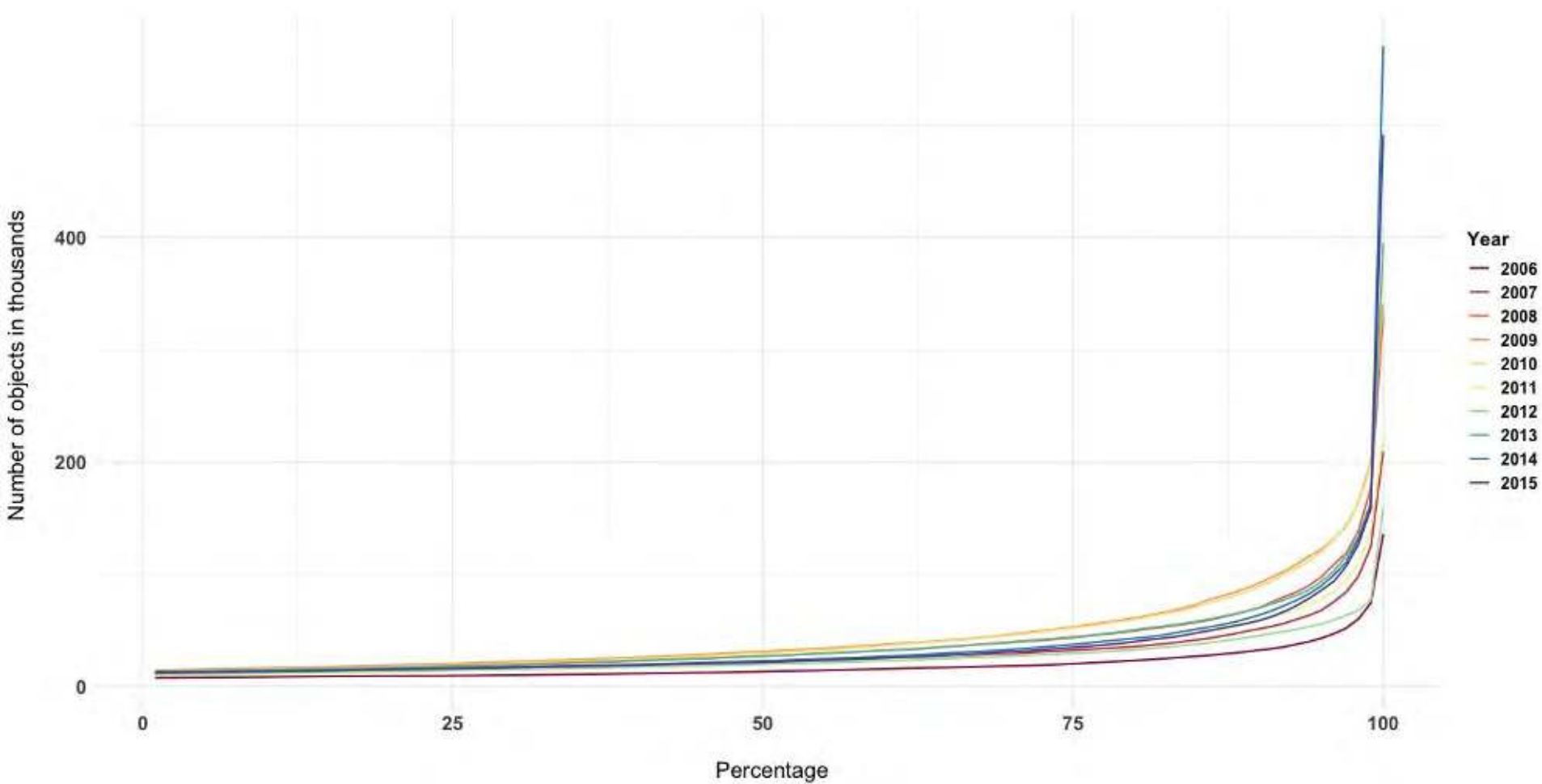
# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

RQ: What is the size of the Danish web?



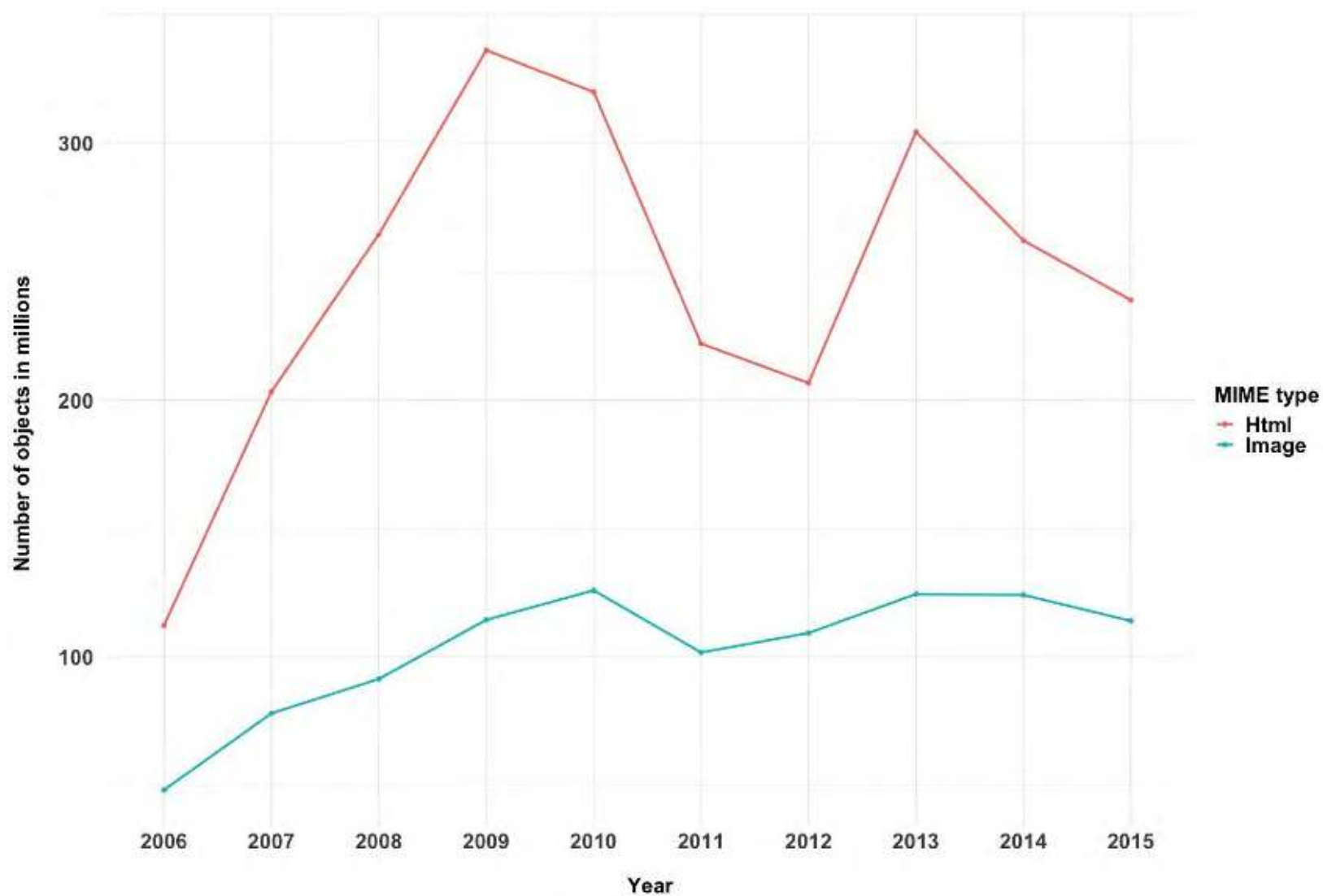




# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

RQ: How much written text and how many images are there on the Danish web?

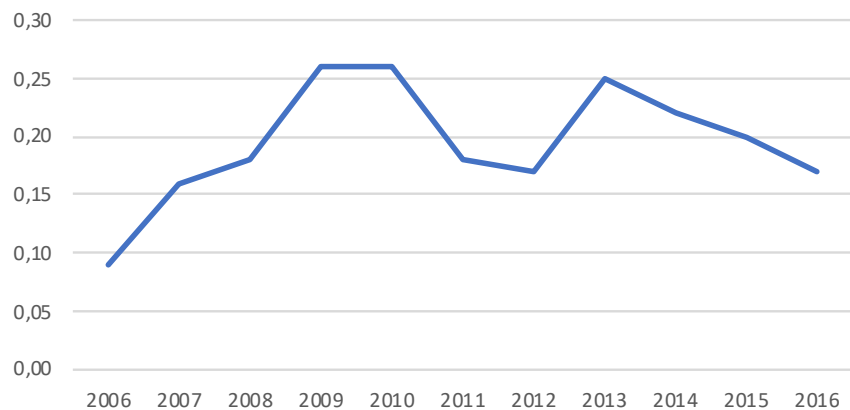


# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

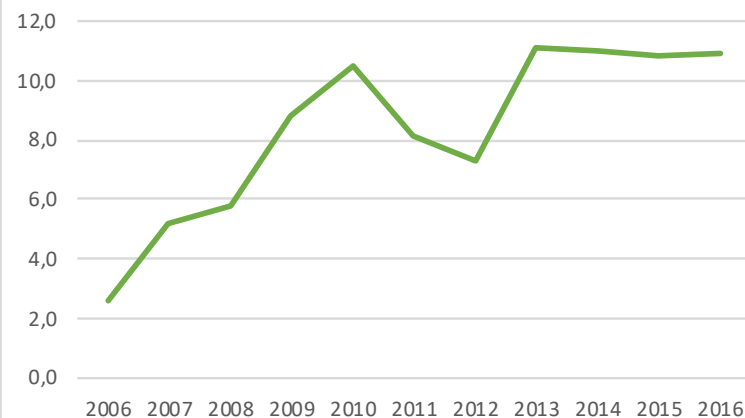
---

RQ: Which are the most popular social media on the Danish web?

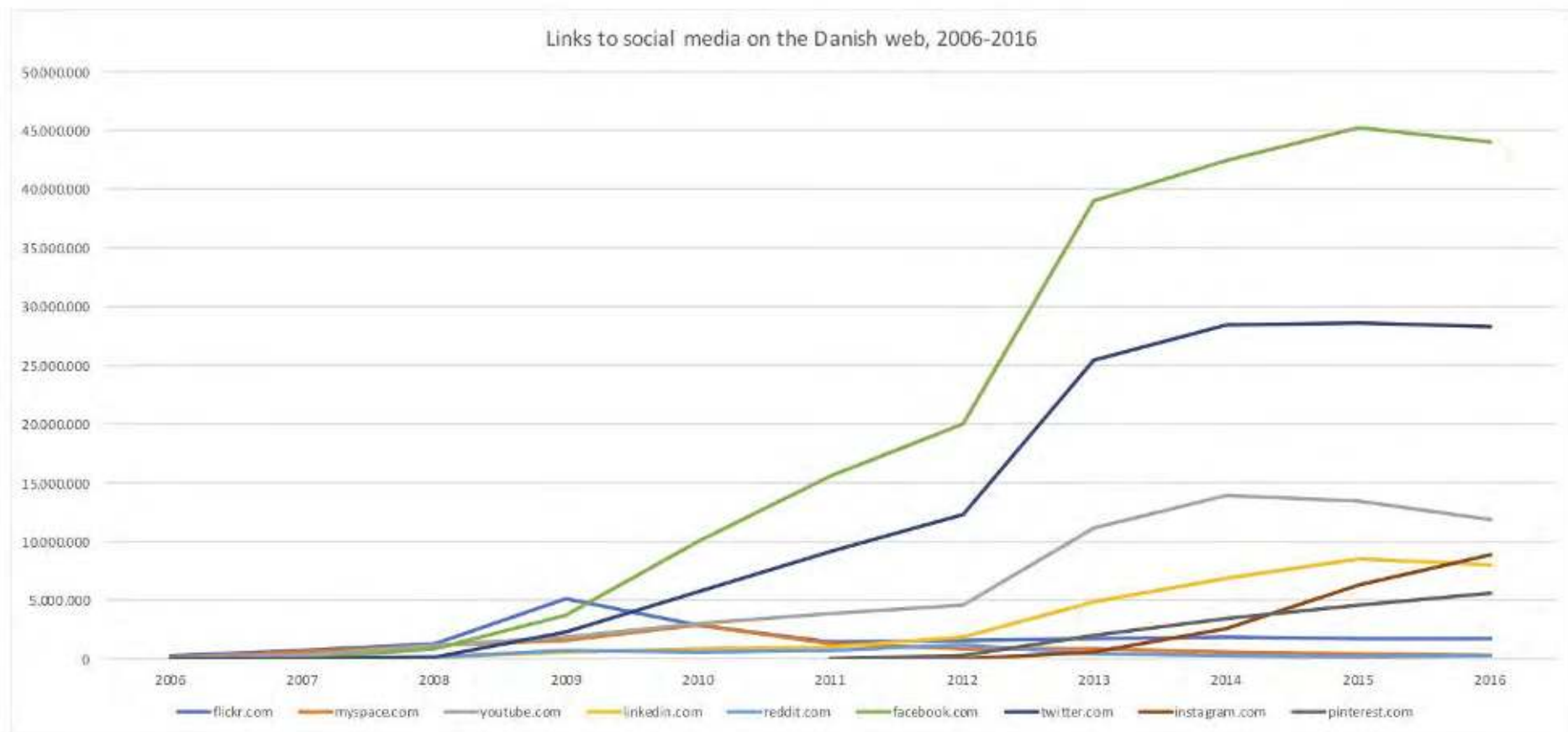
Number of webpages (billions)



Total number of links (billions)







# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

Challenges of remaking this study in another European nation state, like Belgium?

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

The 'have', the 'have-nots', the 'have-too-many':

- › if a country does not have a national web archive  
the study has to be based on e.g. the Internet Archive
- › dependent on their collection strategy and coverage

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

The 'have', the 'have-nots', the 'have-too-many':

- › if a country has more than one national web archive the study may have to combine the content from different web archives, potentially archived with different software, different strategies, etc. — not trivial

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

Different archiving strategies:

- › if an entire national web domain has been archived we may have the needed data
- › if only a selection of the national web domain has been archived we will miss a lot, and comparisons will be difficult — we may have to supplement with material from the Internet Archive or another national web archive
- › and if entire domain and selection were archived we also have to combine

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

Different access conditions:

- › one challenge is to get access to the web archives' holdings, may or may not be possible if coming from another country
- › maybe only onsite access, must go there
- › may not provide the data in a form ready for analysis — very likely
- › data cannot be extracted — also very likely

# 3. IMPACT ON HISTORY WRITING — THE CASE OF NATIONAL WEBS

---

Mitigating these challenges?

- › researchers need to voice their demands
- › and to do this in a coordinated, transnational manner
- › usually national web archives are more than happy to help, but it's a matter of resources

## 4. BELGIUM — THE LAND OF PROMISE

---

- › have already lost the first 25 years of Belgian web
- › get a national web archive established, the sooner the better
- › take stock of all existing national web archive initiatives — take the best and make the best
- › enable extraction, taking material out of the web archive and handing over to researchers
- › ensure enough resources to outreach to researchers — be proactive, get out in the research environments
- › for any researcher of contemporary history national web archives are the land of PROMISE



# NATIONAL WEB ARCHIVES — THE LAND OF PROMISE FOR RESEARCHERS

---

NIELS BRÜGGER  
PROFESSOR, HEAD OF THE CENTRE FOR INTERNET STUDIES,  
AND OF NETLAB

