

**KB as Buzz Lightyear of Dutch Digital heritage:  
saving the Dutch internet from the digital black hole**

**Dr Kees Teszelszky**

Researcher web archive Koninklijke Bibliotheek

[www.kb.nl](http://www.kb.nl)

@keesone

## Post history period is coming

Buzz Lightyear: (insanely) Gone! It's all gone! All of it's gone! Bye-bye! Whoo! See ya!

**Pre-history:** objects, images

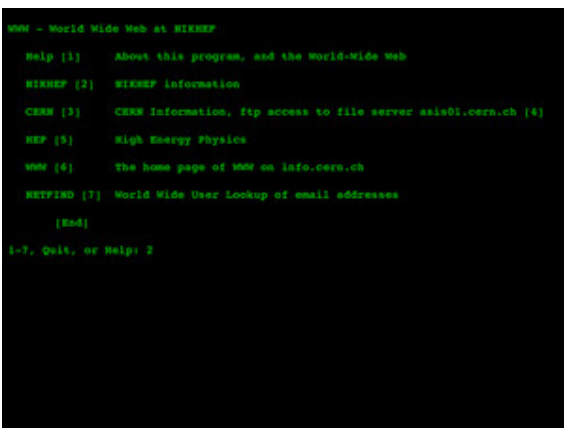
**History:** written sources

**Post-history:** digital or no sources at all

**Web archiving:** preserving digital heritage from the web

## Dutch as pioneers of the world wide web

.nl country code Top Level Domain: **1986**

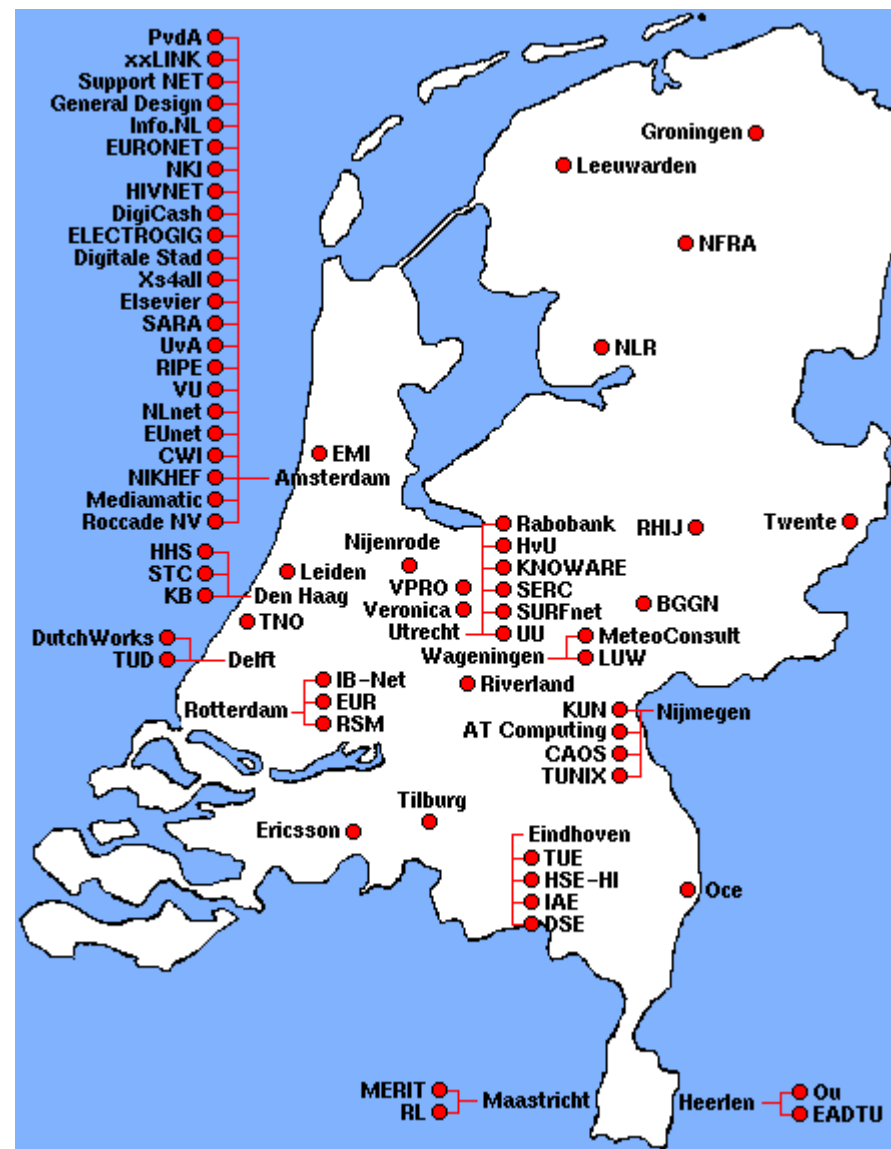


**Website of Nikhef, 1992,  
3th website in the world  
Beginning of www!**

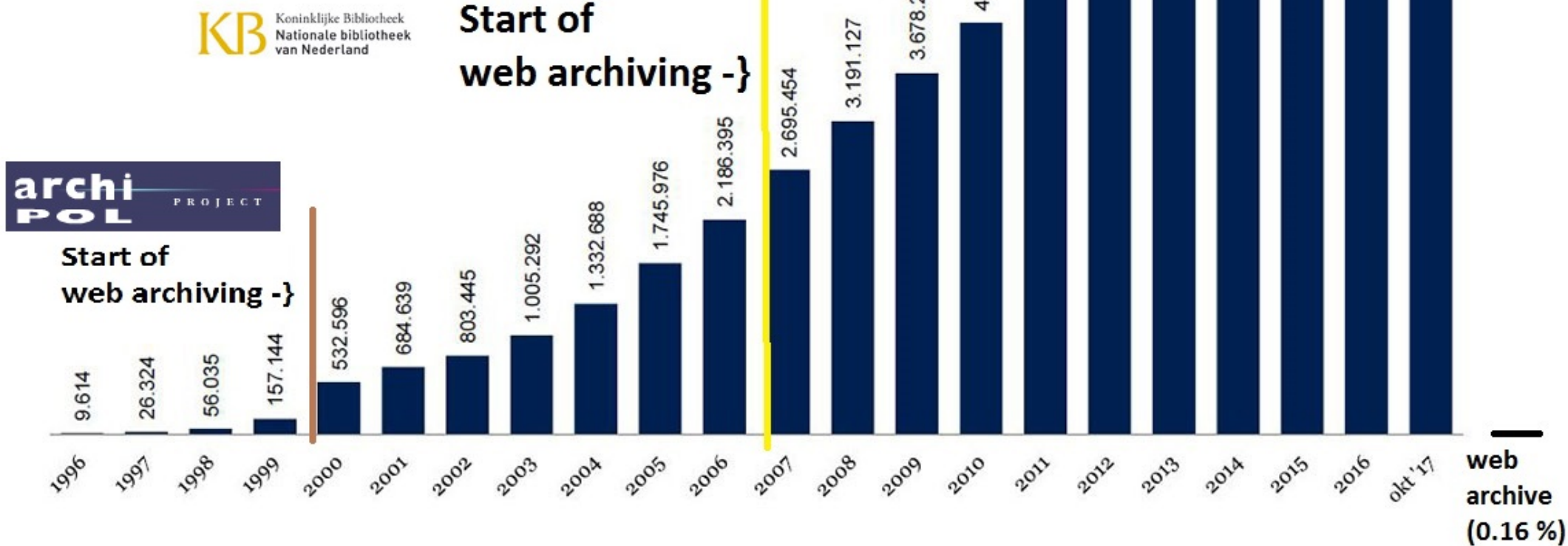
General characteristics:

1. early and **innovative**, fast-growing
2. local or regional, neither centralised, nor one center
3. less attention to heritage (similar to other institutions...)

**Dutch web, mid-1994**

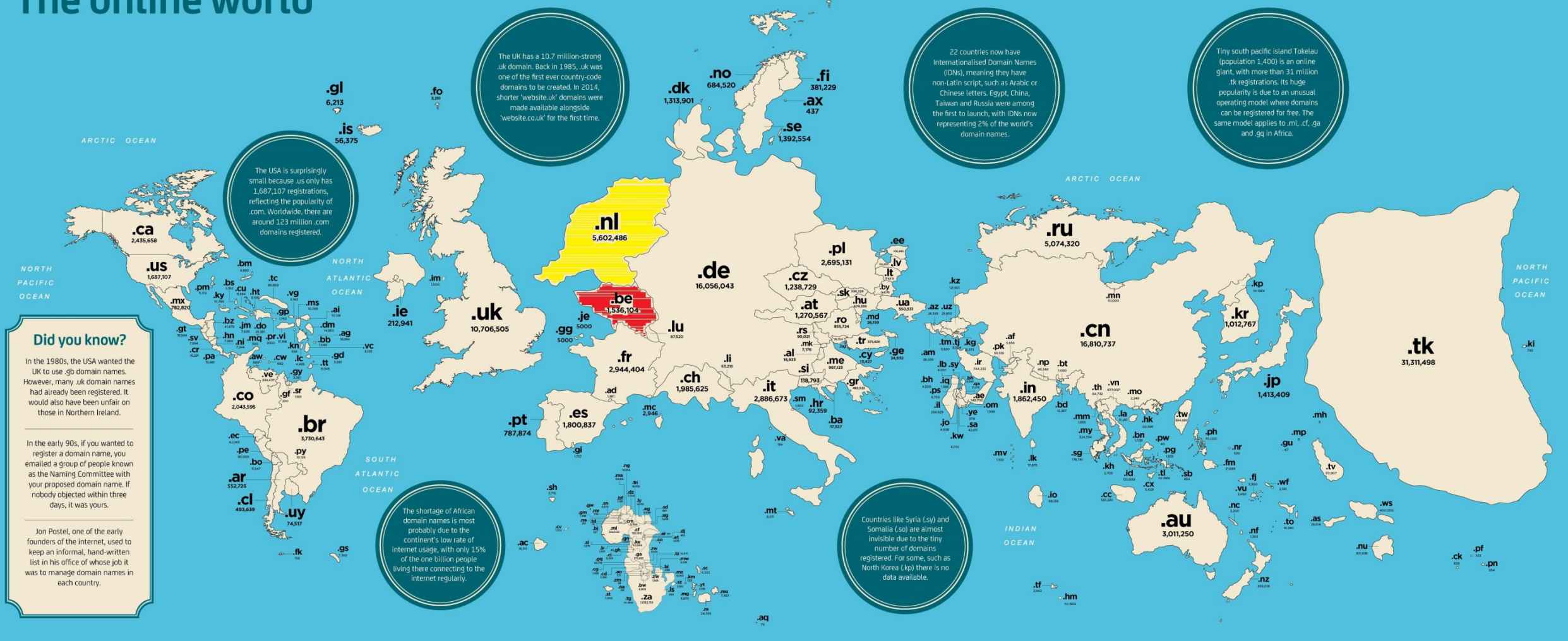


**.nl-domain names: 5.78 million**  
**national web domain: +/- 10 million sites**  
**KB-NL web archive: 13,000 sites**  
**Archipol (DNPP): 1,000 sites**



NOMINET

## The online world



The USA is surprisingly small because us only has 1,687,107 registrations, reflecting the popularity of .com. Worldwide, there are around 123 million .com domains registered.

The UK has a 10.7 million-strong .uk domain. Back in 1985, .uk was one of the first ever country-code domains to be created. In 2014, shorter 'website.uk' domains were made available alongside 'website.co.uk' for the first time.

22 countries now have Internationalised Domain Names (IDNs), meaning they have non-Latin script, such as Arabic or Chinese letters. Egypt, China, Taiwan and Russia were among the first to launch, with IDNs now representing 2% of the world's domain names.

Tiny south pacific island Tokelau (population 1,400) is an online giant, with more than 31 million .tk registrations. Its huge popularity is due to an unusual operating model where domains can be registered for free. The same model applies to .ml, .cf, .ga and .gq in Africa.

**Did you know?**

In the 1980s, the USA wanted the UK to use .gb domain names. However, many .uk domain names had already been registered. It would also have been unfair on those in Northern Ireland.

In the early 90s, if you wanted to register a domain name, you emailed a group of people known as the Naming Committee with your proposed domain name. If nobody objected within three days, it was yours.

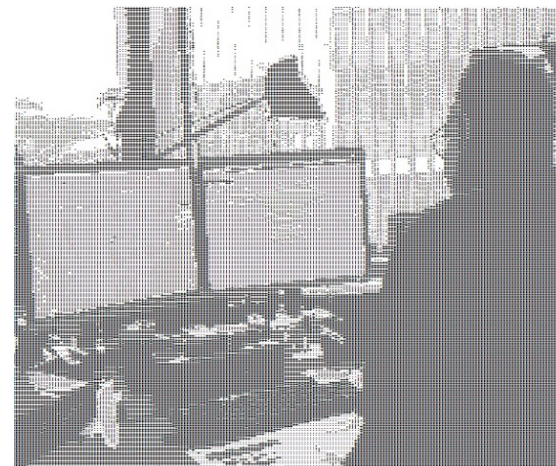
Jon Postel, one of the early founders of the Internet, used to keep an informal, hand-written list in his office of whose job it was to manage domain names in each country.

The shortage of African domain names is most probably due to the continent's low rate of internet usage, with only 15% of the one billion people living there connecting to the internet regularly.

Countries like Syria (.sy) and Somalia (.so) are almost invisible due to the tiny number of domains registered. For some, such as North Korea (.kp) there is no data available.

## Aim of KB-NL web archiving

To select, preserve and make accessible a representative set of Dutch websites of the Dutch national domain (only organisation in the world with this aim)



## Web archiving @ KB in numbers

- 13.000 websites in total webcollection since 2007
- 31 Terabyte (second biggest digital collection)
- Annual growth: +/- 1.000 sites
- 300 million (hyper)links
- Only selective harvests: no legal deposit



## **Innovation @KB**

### **Webarchiving projects:**

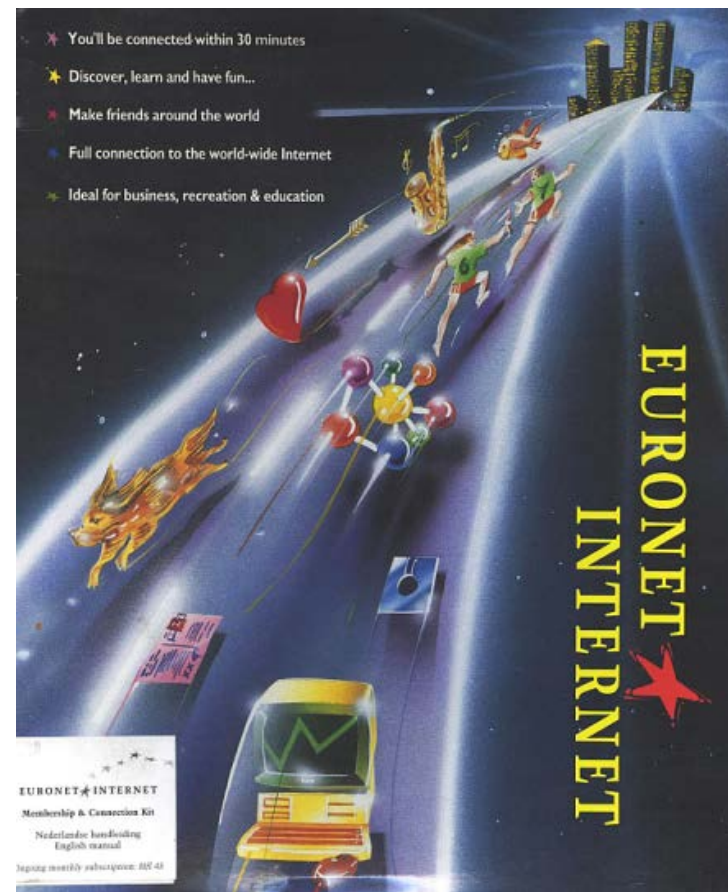
- 1. Web archaeology**
- 2. Preserving online news**
- 3. Web sources of the Post-truth period**

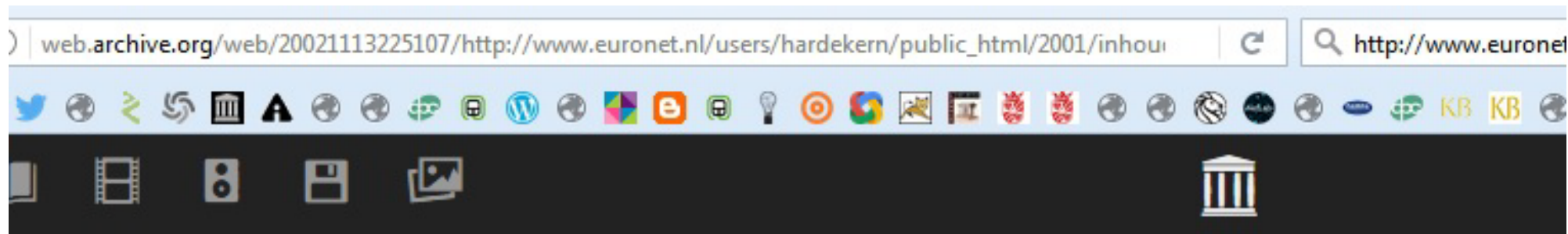


Special webcollection Dutch **web archaeology**:  
find the pearls before 2007, esp. 90's



**Casus:**  
**Euronet**  
**provider**  
**user sites**





INTERNET ARCHIVE  
**WayBackMachine**

Hard to crawl due to  
bad construction of sites  
(Not in Internet Archive!)

[http://www.euronet.nl/users/hardekern/public\\_html](http://www.euronet.nl/users/hardekern/public_html)

Latest

Show All

## Digital incunables of the Dutch web

# Finding unique sites

### Journalistieke en andere vertrekpunten

#### Journalistieke adressen

- [ASNE](#) American Society of Newspaper Editors
- [Awesome Lists](#) Met mooie doorverwijzingen
- [Beat Page](#) Lijst met journalistieke links van reporter.org
- [Committee to Protect Journalists \(CPJ\)](#) Laatste nieuws over problemen rond persvrijheid in de wereld
- [Eurojournalism](#) Europese adressen voor de journalist
- [European Institute for the Media](#) Persberichten, agenda, publicaties
- [European Union Newsroom](#) Nieuwsoverzicht, agenda, persberichten
- [Finding Data on the Internet](#) A Journalist's Guide
- [Hotzones](#) Informatie en tips voor journalisten die naar gevaarlijke gebieden reizen
- [IFJ](#) International Federation of Journalists
- [IFEX](#) Schendingen van persvrijheid
- [Index Persverwijzingen](#) van collega Ren de Vree van het AD *Aanbevolen!*
- [Internet Exercises for Journalists](#) Het Net als Uitdaging...
- [Internet Newsroom](#) Begin hier de jacht op informatie
- [Investigative Reporters](#)
- [Journalism College Florida](#) Met doorverwijzingen plus instrumenten om op het Internet informatie te vinden
- [Journalism Ethics](#) Een index met theorie en praktijk
- [Journalism List](#) Pagina met fraaie doorverwijzingen
- [Journalism SourceFinder](#) Met be- en onbekende adressen
- [Journalist's Toolbox](#) Veel mooie doorverwijzingen
- [Megasources](#) Mega, dus zeer veel links

- [National Press Club in Washington](#) Persberichten, adressen
- [Newseum](#) Museum over het nieuws
- [NewsPlace](#) News, resources
- [Nipkow](#) Van de schijven
- [Online Journalist](#) Bronnen, links, een Webtour
- [People's Communication Charter](#) Handvest van de Amsterdamse hoogleraar Hamelink
- [Press News Site](#) Current events, press room, tools
- [Pulitzer](#) De Amerikaanse prijs voor journalisten
- [Reference Guide](#) Gopher-overzicht met veel naslagwerken
- [Reporters Network](#) Met Journalism Resources
- [Research on the Net](#) Met behulp van Wired's editor guide
- [Scoop](#) Internet Resource Guide voor m.n. journalisten
- [Smoking Gun](#) Site met materiaal over schandalen
- [WebOvision](#) Journalism, dus met nuttige links
- [World Wide Events](#) Agenda van aanstaande evenementen

#### Andere vertrekpunten

- [WWW Virtual Library](#) From A to Z, hier komt geen eind aan, van Aboriginal Studies tot Zoos
- [Yahoo](#) Met heel veel mooie doorverwijzingen en zoekmogelijkheden (zie ook 'Zoekinstrumenten')
- [Yanoff's List](#) Zeer volledige List van een veteraan op het Internet *AANBEVOLEN!*

	<a href="#">De visie van</a>	<a href="#">Meer informa</a>	<a href="#">Opinie-</a>	<a href="#">D66 Interact</a>	<a href="#">Wat staat er</a>	<a href="#">Verwijz</a>
--	------------------------------	------------------------------	-------------------------	------------------------------	------------------------------	-------------------------

### [Chatten met Kamerleden!](#)

Een partijcongres is altijd een goede gelegenheid om eens een babbeltje te maken met die mensen die midden in het politieke centrum staan. Heeft u gerichte vragen of wilt u gewoon eens een babbeltje maken, dan bent u zaterdag en zondag van harte welkom op het [IRC-kanaal van D66](#). U heeft hiervoor wel een apart programma nodig. Hoe u dit kunt downloaden, installeren en hoe dit verder werkt (het is NIET ingewikkeld!) dat leest u [hier](#). De volgende mensen kunt u bereiken op de volgende tijden:

Zaterdag:  
10.30 [Bert Bakker](#)  
12.30 [Roger van Boxtel](#)  
14.30 [Nicky van 't Riet](#)  
16.30 [Pieter ter Veer](#)

### [Verkiezingscongres via Internet](#)

Zaterdag 14 en Zondag 15 Februari zal in Noordwijk het D66 congres plaatsvinden. Ook dit congres zal weer via internet te volgen zijn. Met behulp van teksten, audio en foto's zal iedereen in staat zijn om het congres zo goed mogelijk te volgen. In [Emmen](#) zette D66 de trend om congressen live te verslaan en deze keer belooft het nog weer beter te worden. Maar wie het congres echt optimaal wil meemaken kan [hier](#) een omschrijving van het congres vinden en [hier](#) een routebeschrijving...

### D66 vindt sport

### Virtuele wandelgang

Op het komende D66-congres lopen uiteraard ook alle D66-prominenten rond. De beste plaats dus om even een kamerlid of een minister aan te schieten. Wie niet naar het congres in Noordwijk kan komen kan toch van de gelegenheid gebruik maken. [Mail](#) u vraag naar [webmaster@d66.nl](mailto:webmaster@d66.nl) en vermeld aan wie u de vraag wilt stellen. Het D66 Nieuwe Media Team loopt voortdurend door de wandelgangen, dus als de gelegenheid zich voordoet zullen wij u vraag aan de desbetreffende persoon voorleggen. Alle antwoorden op uw vragen vindt u dan dit weekend terug op de site - in [Realaudio](#) of [Realvideo](#).

**Stuur email aan Els Borst**

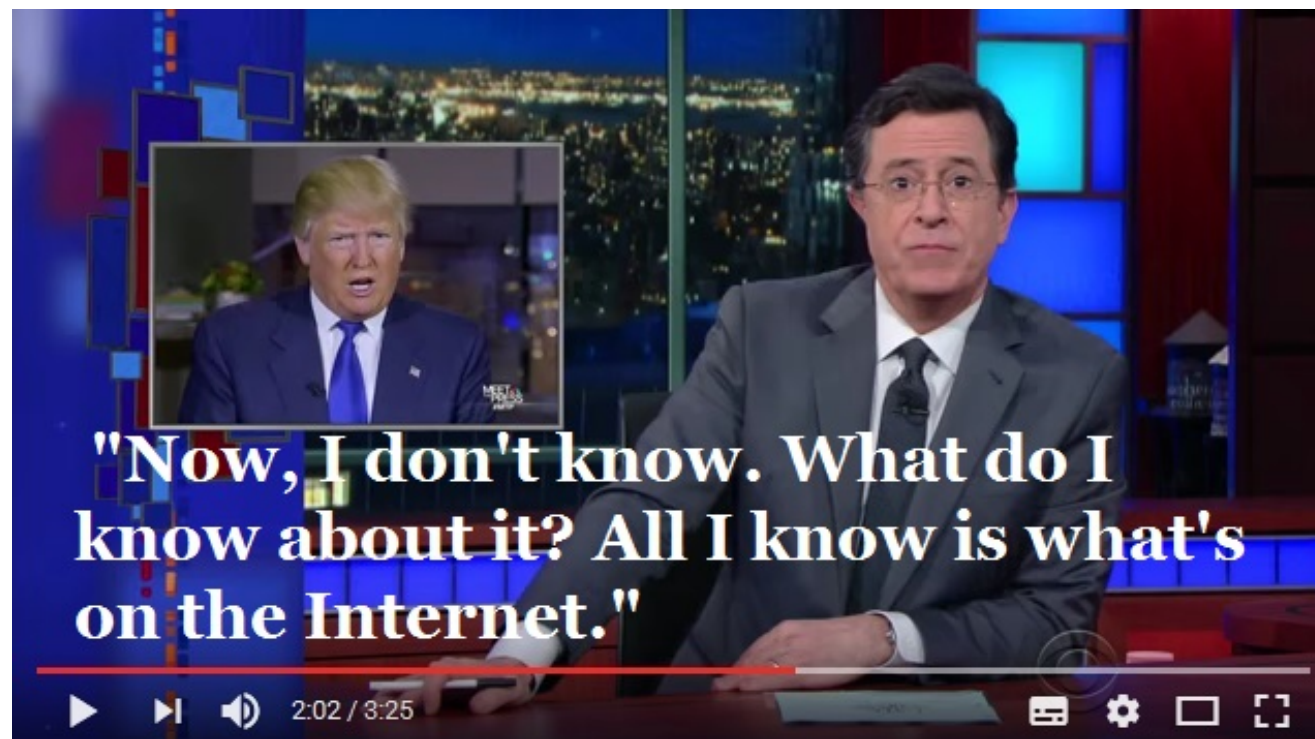
## Archiving actual sites from the web: mirror of Post-truth period

“The KB aims to preserve a relevant selection of Dutch websites.”

(Content strategy 2017-2020)

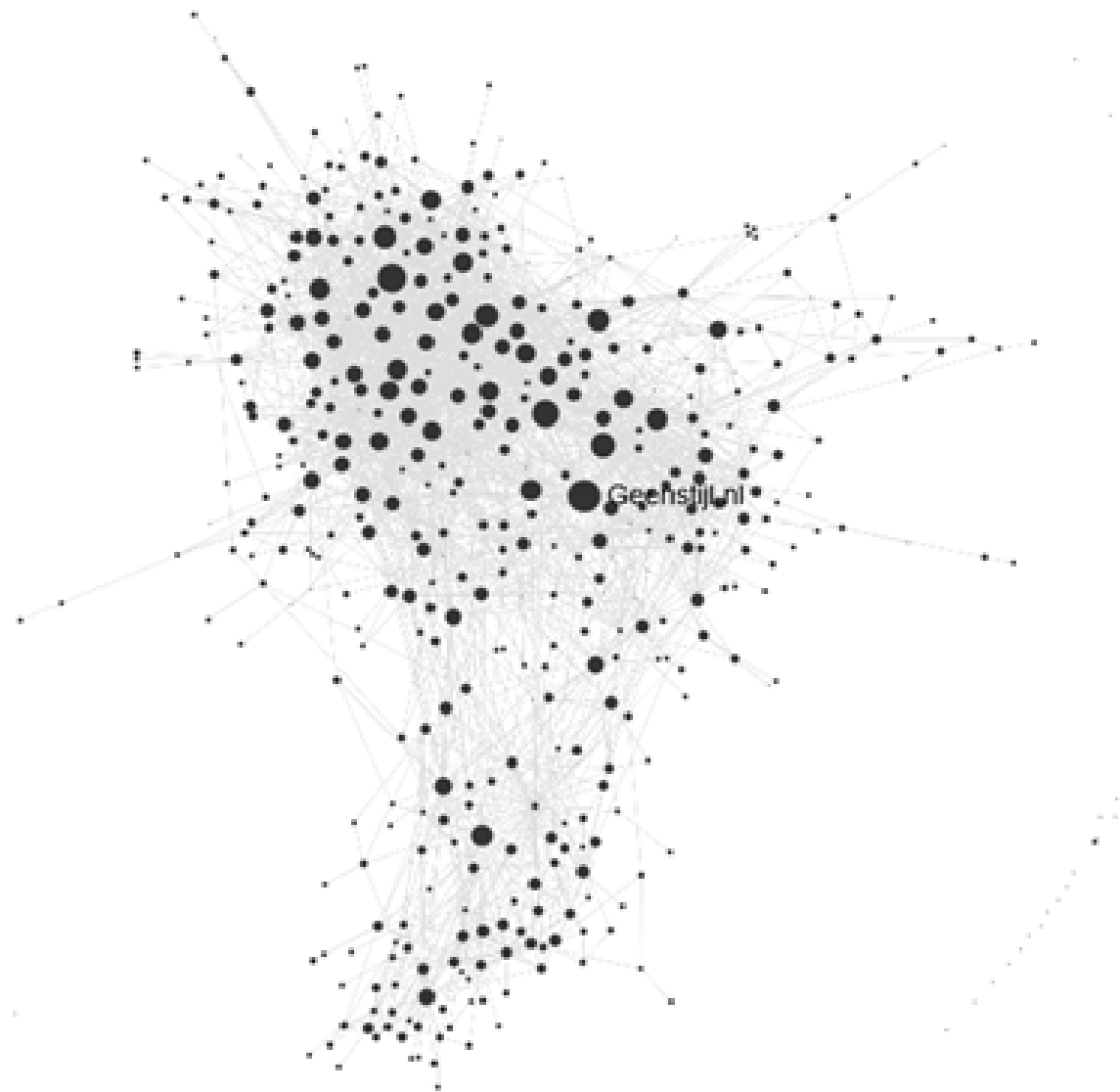
Must select now, no time for  
“**historic distance**” towards  
source material

Selection for future research  
must be done **now**



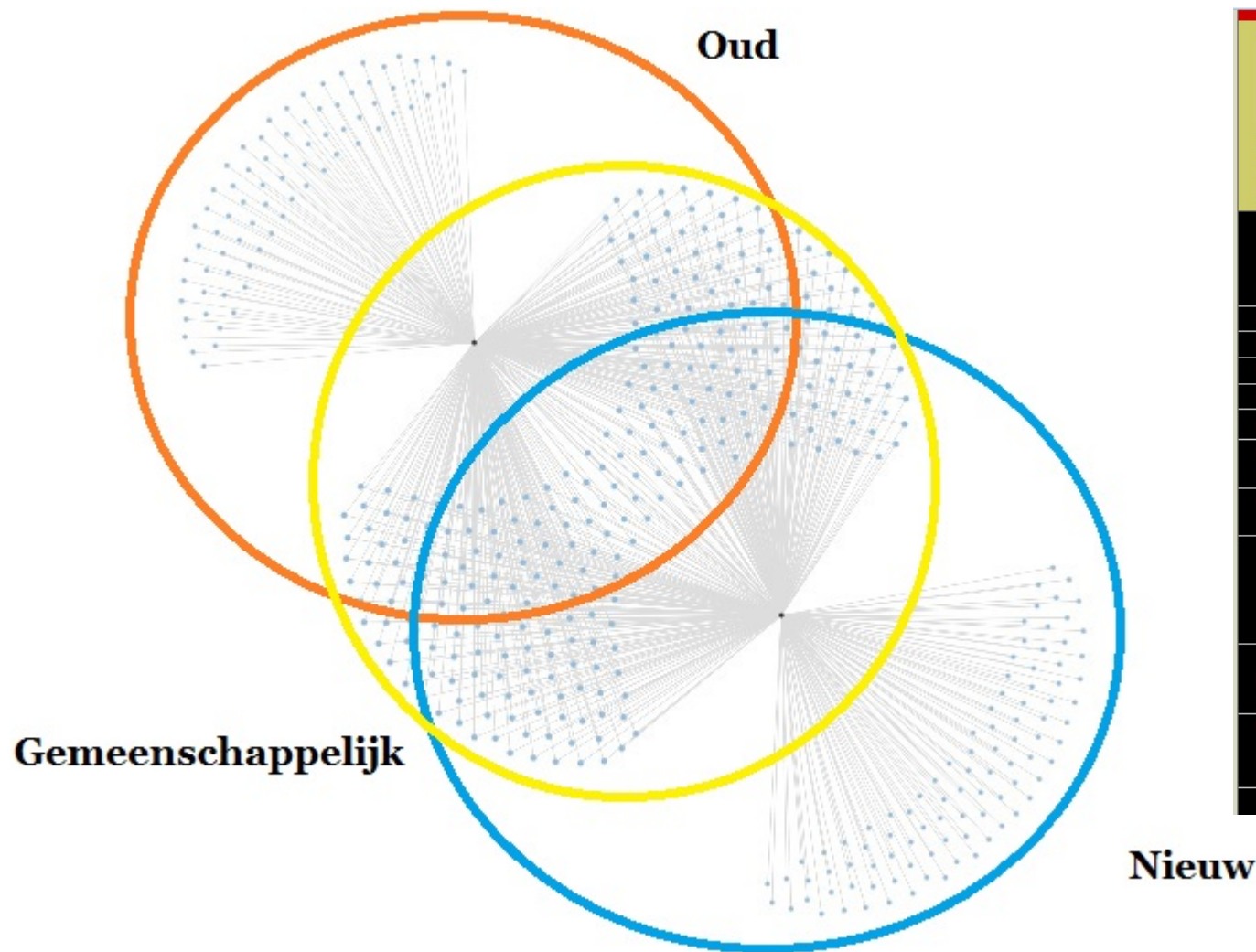
## Network analysis

- 500 relevant sites based on PhD thesis and academic research
- Blog Geenstijl.nl most cited website
- First time: mapping of web sphere through links



<http://hyphe.medialab.sciences-po.fr/demo/#/project/maatschappijkrit/network>





**EINDTIJD INFORMATIE WEB**

Microsoft Internet Explorer, Netscape Now!, Adobe Reader, real, WinGrave New Version Download, Get Adobe Reader

**EIW is een christelijke website met betrekking tot bijbelse eindtijdpredicties en de vervulling ervan**

**Belangrijk:** EIW is niet verantwoordelijk voor de inhoud van de reclamebanners en links vanaf deze banners die op de pagina's van deze site zouden kunnen verschijnen. Site best bekeken met een resolutie van 1024\*768, min. 16 bits kleuren. Als u hier nieuw bent leest u dan a.u.b. [dit](#) eerst.

**ATTENTIE!** Vanwege de enorme hoeveelheid verkeer op deze site slaat de lokatie op Wanadoo (Orange) steeds vaker uit bandbreedte. Bookmark in ieder geval dit url: <http://www.openbaring.org> om de bandbreedte te spreiden.

Omdat u geïnformeerd dient te worden in deze laatste dagen (uren)

To our English speaking visitors: This is a non-commercial christian (Dutch) website containing a collection of links and articles that relate to Bible Endtime Prophecies. For an English version click [here](#) or on the Union Jack on top of page.

Vindt u ook dat de media soms eenzijdige en anti-Israëlische berichten publiceren? Onjuiste of onvolledige berichtgeving richt grote schade aan. Sommige verslaggevers lijken voortdurend Israël aan te vallen en verhullen hun sympathie voor terroristen nauwelijks. - U kunt er iets aan doen! - [Meer](#)

DE 13 SATANISCHE BLOEDLIJNEN -- De oorzaak van veel ellende en kwaad op aarde -- Robin de Ruiter en co-auteur Fritz Springmeier  
VERBIJEN UITGAVE: 'De 13 Satanische

**Tekenen v/d eindtijd**  
Wanneer gij dit alles ziet geschieden, beeft uw boezem omhoog want uw verlossing genaa kt  
LUCAS 21:28

**Aardbevingen:**

**Biometrische Identificatie**

**BIO-CHIP**

<http://members.ziggo.nl/henryv/index.htm> (oud) <http://www.openbaring.org/> (nieuw)

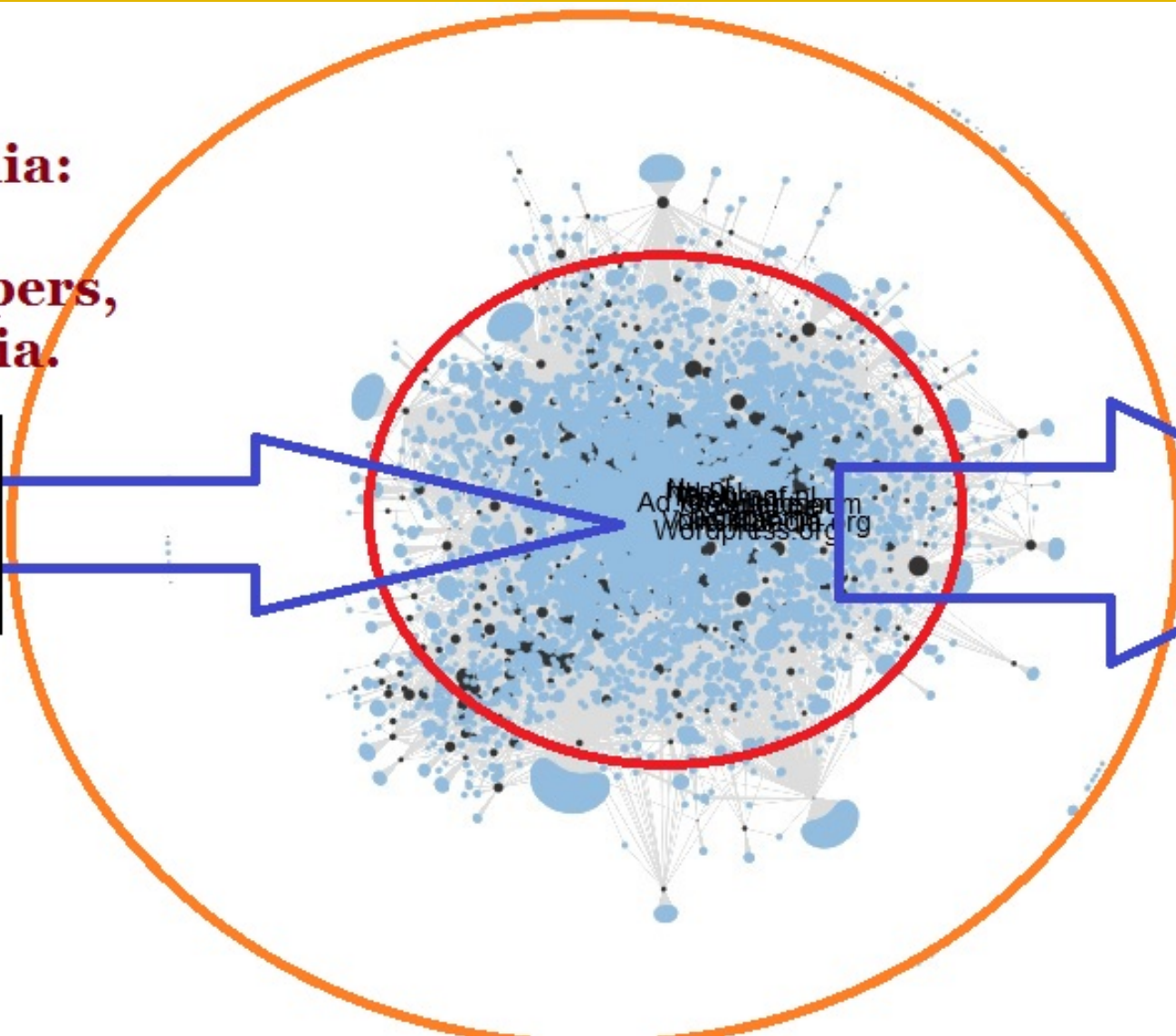




**Old media:  
Books,  
newspapers,  
wikipedia.**

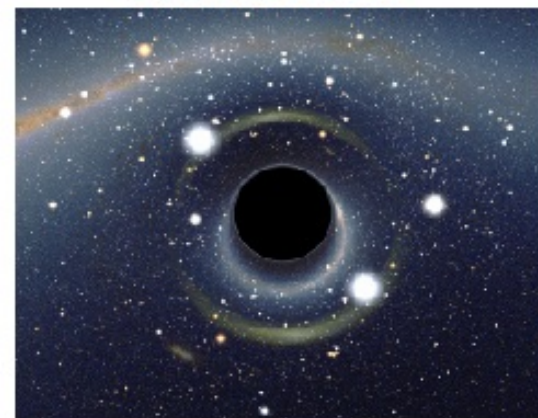


"Sun"



**New media:  
Social media  
(Youtube,  
Facebook)  
Apps  
(snapchat)**

**Digital black  
hole: cannot be  
archived.**

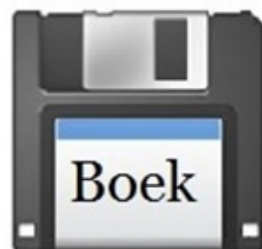


## Conclusion

Not only focus on **medium**, but to **content** and **trends**.

What can be of source value for the future?

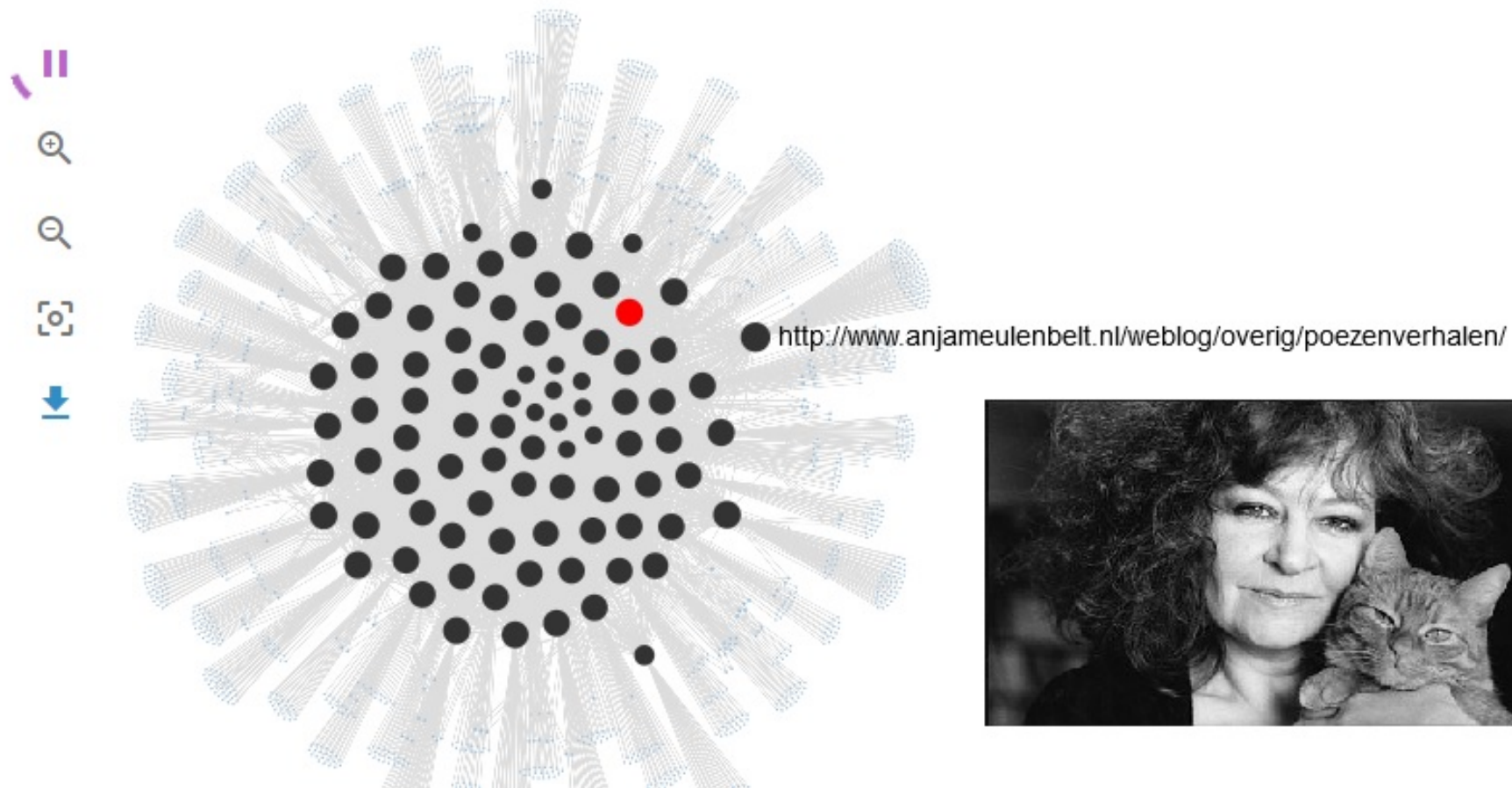
.



## Innovative possibilities for future research:

- **Building a bridge** between Humanities and Digital Humanities by presenting ready to use source data (data sets with URL lists, link lists)

• Pages network: [Anjameulenbelt.nl](http://www.anjameulenbelt.nl)



## Content analysis



**KB**  
**WEBARCHIVING**

**Questions?**