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
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

Start of web archiving -

web archive (0.16 %)
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
Webarchiving projects:
1. Web archaeology
2. Preserving online news
3. Web sources of the Post-truth period
How to improve the selection and harvest?

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

**Casus:**
Euronet provider user sites
Hard to crawl due to bad construction of sites
(Not in Internet Archive!)
How to improve the selection and harvest?

Digital incunables of the Dutch web

Finding unique sites
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
Linkanalyse

http://members.ziggo.nl/henryv/index.htm (oud) http://www.openbaring.org/ (nieuw)
Content analysis on basis of links and web nodes
Trend: web in 4D

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)
Content analysis
Questions?