What do Digital Humanists want from a National Library?

dr. Steven Claeyssens | @sclaeyssens | [dr. Martijn Kleppe | @martijnkleppe]
Part 1

How we assist Digital Humanities scholars
Data
Vastknellende aan en in der daad
tegelijk nog nimmer
terleven de schrijvers overtroffen
België aan de ware
maar ook het goed
koest, en om de k"u"den. Inderdaad
over alle landen ge
Gok hier la ander vooral uit den Boel
achtwaardigste en
vooroordelen, in
Franschen zeiven.
Nooitens waardige internationale kon
opofferingen doen,
onderen zouden in veronderstelling komen
moeten onderwerp
aangewezen zijn. Z
schuilplaats te zijn
waarnemde ijdelheid
in een zeer nauw
Wij bevelen het be
publiek en aan ons
eens, even goed al
worden. Zij verwielt
tegenovergestelde
rede als het liefst
INTERNATIONAAL

de Edinburgh Revi
nummer van 1852
opnamen, als uitz
d van dat, in een vro
meedegedacht:

De toehoors
letterd
land, dat een lezen
drukwerk van Engel
toegezeggen van Fr
België door de rep
bekende andere landen
toch dat Belgie th
 nadrukken en nab
Data

= result of
  more than 200 years of collecting
  over 30 years of digitisation
+ almost 10 years of collecting born-digital publications

= machine readable, mostly textual
= structured or semi-structured
= legally as open as possible
Datasets of the KB, National Library of the Netherlands

The collections of the KB National Library of the Netherlands are being digitised on a large scale. The KB publishes books, periodicals, newspapers and other textual heritage freely accessible on the Web.

A significant amount is also made available in bulk for research purposes. Datasets containing digital texts, images and meta-data, can be accessed through www.kb.nl/datasets.

Characteristics

• extensive Dutch text corpora
• optical character recognition with vector coordinates
• machine-readable access
• documents in PDF and/or JPEG
• full text in XML
• metadata in Dublin Core and MPEG-21 DIDL
• access via SRU and OAI-PMH

Research projects

HETEN, BLAIR-CLARK, PolitieMedia, PubMedia, CATON, SEAL, YFMT, Radio Pieterkie Verhoeff, WEBSP, and more...

Datasets

• Medieval Theiran Manuscripts
• 1,300 images from 20 medieval manuscripts (manuscripts.snr.kb.nl)
• Historical Newspapers
• 124,000 newspaper pages from the Netherlands and its former colonies, 1666-1945 (kranten.kb.nl)
• Early Dutch Books Online
• 4,000 books from the Early Modern region, 1450-1750 (www.earlydutchbooks.nl)
• States-General Digital
• 220,000 Dutch parliamentary papers, 1535-1848 (www.statengenerealdigitaal.nl)

Periodicals
• 13 million pages from Dutch periodicals, 1888-1945 (periodicals.kb.nl)

AMI indexes

• 13 million indexes of radio bulletins, 1920-1945 (AMI-indexen.nl)

And more...

Contact

admin@kb.nl
www.kb.nl/dataservices

www.kb.nl
Lab
Lab

= online **and** offline sandbox
  1 co-ordinator
  1 curator
  2 programmers

= experimental data sets

= tools and software

= workshops

www.kb.nl/lab
Programs
Programs

1. Fellowship Program
   = 4 months, 1 fte | invited | successful academic

2. Researcher-in-Residence Program
   = 6 months, 0.5 fte | open call | early career

Call is open! – blog.kbresearch.nl
Part 2

Our experiences with assisting Digital Humanities scholars
Additional analysis
Findings

Particular interest in the newspaper corpus

Very diverse research questions and use cases

Various ways of data delivery wanted

Most computer scientists have data-driven research questions

Most humanities scholars are driven by research questions
Findings

Apart from a lot of good points of critique, also:

Conflicting wishes about the corpus

Poor knowledge about the corpus

Poor understanding of the difference between the data and the tool
Abstracts received for Researcher-in-residence 2016

JANUARY 7, 2016 / LOTTE WILMS

Below you will find all abstracts that were submitted and unfortunately not accepted for the 2016 run of the Researcher-in-residence programme. The abstracts are in alphabetical order. The accepted projects and their abstracts can be found [here](#).

We want to thank all researchers for their interesting proposals, wish them all the best for 2016 and hope to see them again in a following year!

[blog.kbresearch.nl](http://blog.kbresearch.nl)
Findings

Again: very diverse research questions and use cases

Proliferation of tools and software

Interest in several corpora
Part 3

Lessons
The hard ‘news’

There are no typical Digital Humanities scholars.

Their needs change over time, change with skills.

There are hardly any standard tools or techniques.
In the short term

We need to be more transparent about the data (and the tools).

An API is not enough. Multiple ways to deliver the data are needed.

We need to be more involved in research projects.
In the medium term

We need to (collectively) develop a digital source criticism.

We need to (collectively) develop a tool criticism.
In the long term
Any questions?