Cultural Heritage, Digital Humanities & E-Science

Peter Doorn
DANS, The Netherlands

Plan-E Plenary, Rome 15-16 April 2019
Contents

- From Humanities Computing to Digital Humanities
- Digital Cultural Heritage: Examples
- Curating and Archiving Digital Heritage
- Linking up the containers of heritage data
- Research Infrastructures for Humanities
- Future challenges of e-scholarship
- Questions for the Breakout
The Humanities

The Humanities are academic disciplines that study aspects of human society and culture.

The humanities use a wide range of methods that are primarily critical, sometimes speculative, and have a significant historical element — and some use empirical approaches like in the natural sciences.

The humanities have no central discipline. They include:

- Ancient and Modern Languages
- Literature
- Philosophy
- History
- Human geography (social science?)
- Law
- Politicology (social sciences?)
- Religion
- Arts (visual, performing, art history)
**Debates & Claims:**

- “L'historien de demain sera programmeur ou il ne sera plus” (Le Roy Ladurie, 1967)
- Quantification and the application of computers will bring new insights in human behaviour and socio-cultural processes
- Methods from the “hard sciences” will revolutionize the humanities
- Qualitative research has little to gain from computing

**Timeline:**

- **1950s**: Bible studies
- **1960s**: Historical databases
- **1970s**: Content analysis
- **1980s**: Archaeological GIS
- **1990s**: Speech processing
- **2000s**: Collaboratories
- **2010s**: Linked data
- **2020s**: Driven by data
Data archives in the social sciences and humanities

- Data collection and data processing ➔ awareness of the value of preserving data for re-use:
  - for validating the results of earlier research
  - for comparative analysis
  - answering new research questions with existing data (secondary analysis)

- Emergence of data archives:

<table>
<thead>
<tr>
<th>social science data archives</th>
<th>text archives for linguistics and literary studies</th>
<th>historical data archives</th>
<th>archaeology data archives</th>
</tr>
</thead>
<tbody>
<tr>
<td>ICPSR, ZA, UKDA Steinmetz</td>
<td>Oxford Text Archive</td>
<td>NHDA, HDS, IPUMS</td>
<td>ADS, EDNA</td>
</tr>
</tbody>
</table>

The example of Data Archiving and Networked Services in The Netherlands

- Created in 2005, combining the forces of:
  - Steinmetz Archive for the social sciences (1964)
  - Netherlands Historical Data Archive (1989)

Mission:
- Promote and provide permanent access to digital research information

https://dans.knaw.nl

Driven by data
Humanities Computing Examples: Bible studies

https://shebanq.ancient-data.org

Dutch Bible: only 917,264 words (21,554 different)

https://www.statenvertaling.net
Historical Population Censuses

- Census digitization projects since 1996
- Collaboration with Statistics Netherlands
- 80,000 pages of tables turned into numbers
- Images of the original source books
- Up to 60,000 visitors per year

www.volkstellingen.nl
Oral History (qualitative social science)

- Thematic collection from several projects (especially about war and conflict)
- 1822 interviews, many Audio/Video files, partly transcripted or summarized
- Value of reuse of qualitative data demonstrated in enhanced publication on “Veteran Tapes”

Driven by data
Digital Archaeology

- The biggest (but youngest) collection in DANS (40,000 out of 46,000 humanities datasets)
- In collaboration with Cultural Heritage Agency of the Netherlands
- Obligatory deposit of data
- *European Convention on the Protection of the Archaeological Heritage* ("Malta Convention")
- Great variety in data formats
Language and speech analysis

Annotations on video and audio

Language variation

Archives of multi-media and multi-modal language resources

Lexical databases

Speech analysis

Driven by data
Thousands of data silos in the humanities

- Historical databases
- Archaeological GIS
- Linguistic corpora
- Literary text bases

Driven by data
Collaborating to preserve digital heritage (and research data)

Netherlands Coalition For Digital Preservation:
- National Library
- Sound & Vision
- National Archives
- DANS
- Heritage (Museums)

Dutch Digital Heritage Network:
- National Library
- Sound and Vision
- Cultural Heritage Agency
- Humanities Cluster of the Academy (KNAW)
- National Archives

Driven by data
Three pillars:

- Visibility
- Usability
- Sustainability

- Increase public value of collections: users are the focus
- Cross-sector collaboration & facilities
- Joint standards to link collections
- Building on existent hubs

Driven by data
Examples of Connecting data silos: Historical shipping

Bringing together shipping records from projects over the decades: South Chinese Sea Trade (1681-1792); Dutch-Asiatic Shipping (1602-1795); Climate of the World Oceans (weather observations from ships’ logs, 1750-1854), Atlantic Connections, Trans-Atlantic Slave Trade, etc.
Journal entry, 26 September 1758

Month: September
Year: 1758
Date: 26th
Day: Tuesday

Ship's name: Noordbeveland
Weather on board

Peculiarities

Wind
Shipping Routes 1750-1850
(Spain, Netherlands, England, Argentina)

Source: CLIWOC project, KNMI
http://www.knmi.nl/cliwoc/

Dutch ships in South Chinese Sea, 1681-1792 (decades)

South China Sea trade, 1681-1792
urn:nbn:nl:ui:13-hts-h7z

Source: CLIWOC project, KNMI
http://www.knmi.nl/cliwoc/

Driven by data

Every dot an entry in a ship’s log

Fort Zeelandia, Formosa
Wind speed and direction, and temperature per month, 1750-1850

Maps courtesy of CLIWOC project, KNMI
http://www.knmi.nl/cliwoc/
Trans-Atlantic Slave Trade

Embarked: 12.5 million
Disembarked: 10.7 million
Died on board: 1.8 million

Dutch share: 550 thousand (4.5%)

About 36,000

https://www.slavevoyages.org
Clio-Infra: historical data on worldwide economic growth & inequality

• Data collection from thousands of sources from all over the world by hundreds of specialists
• Solving massive problems of data interpretation, cleaning, linking, harmonization, comparison...

From source to database: example on age data about Ceylon, 1770

Reconstructing Global Inequality

Solving the problem of “age heaping”

<table>
<thead>
<tr>
<th>Poentje hamij</th>
<th>Ittepane</th>
<th>dorp1</th>
<th>wed</th>
<th>v</th>
<th>65</th>
</tr>
</thead>
<tbody>
<tr>
<td>Toepaijge Basti</td>
<td>Aloetkoer korle</td>
<td>distr1</td>
<td>geh</td>
<td>m</td>
<td>nv</td>
</tr>
<tr>
<td>Christina</td>
<td>Aloetkoer korle</td>
<td>distr1</td>
<td>geh</td>
<td>v</td>
<td>nv</td>
</tr>
<tr>
<td>Masan</td>
<td>Ittepane</td>
<td>dorp1</td>
<td>doet</td>
<td>na</td>
<td>geh</td>
</tr>
<tr>
<td>Tosienda</td>
<td>Ittepane</td>
<td>dorp1</td>
<td>geh</td>
<td>v</td>
<td>30</td>
</tr>
<tr>
<td>Adea</td>
<td>Ittepane</td>
<td>dorp1</td>
<td>nv</td>
<td>m</td>
<td>30</td>
</tr>
<tr>
<td>Lokoe natjere</td>
<td>Aloetkoer korle</td>
<td>distr1</td>
<td>nv</td>
<td>v</td>
<td>nv</td>
</tr>
</tbody>
</table>

Driven by data
Long-term trends and comparisons after reconstruction

GDP per capita, 1270-2000 ($ of 1990)

- Indonesië
- Nederland

Average years of education per capita, 1880 (reconstruction) and 2000

Driven by data

Jan Luiten van Zanden
3. Digital Collaboratory for Cultural Dendrochronology

Data collections of ‘old wood’ for The Netherlands

- Private sector in The Netherlands (6000 BC-present):
  - > 2000 research projects
  - > 20,000 measurement series of 13,000 trees (60% dated)
- Private sector and universities in Germany:
  - Archaeology: e.g. Dorestad
  - Cultural heritage: many objects from The Netherlands and Flanders
  - Architectural history: North and East NL, Amsterdam

Driven by data
European Research Infrastructures for Humanities and Heritage

DARIAH-ERIC
Digital Research Infrastructure for the Arts and Humanities

Explore 58,331,022 artworks, artefacts, books, films and music from European museums, galleries, libraries and archives

EHRI
EUROPEAN HOLOCAUST RESEARCH INFRASTRUCTURE

PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

CLARIN
Common Language Resources and Technology Infrastructure

Driven by data
Opening up the silos: Linked Open Data and Semantic Web Technologies

[Image of a TED Talk by Tim Berners-Lee]

[Diagram showing connections between research data, community content, and electronic publications]

Driven by data
The CEDAR project - linked open census data

- Population, Housing and Labour censuses
- 507 Excel files
- 2,288 tables
- 33,283 annotated cells

Harmonize semantics and table structure and express as RDF
Impact on the humanities and social sciences: new models of scholarship

- e-Humanities modelled after e-Science
- Impact of computing science:
  - new questions are being posed
  - old questioned can be answered in new and better ways
  - increase of scale and division of work
  - new tools and methods are being developed
  - the new scholar is rather an information scientist than a programmer
- New challenges of scholarship for the coming decades
  - to preserve and open up the resource containers and to link up data of all kinds
  - to recognize high level patterns in huge volumes of structured and unstructured data (big humanities data)
  - to maximize information content while minimizing noise in fuzzy data
  - to combine different perspectives on the same data (re-use) by combining studies of language, media, historic interpretations and socio-cultural impacts
Thank you for your attention
www.dans.knaw.nl

Finally ... an integrated collaboration solution :-)