

An integrated approach for data and tools in the humanities

Searching beyond the borders of topic and time

The Dutch projects CLARIN-NL and DARIAH-NL have merged and evolved into a new infrastructure for data and tools in the humanities: CLARIAH.

Patricia Alkhoven

Libraries, museums and archives conserve and manage large collections of data (texts, images, structured data, audiovisual materials) which are gradually being made available in digital form for re-use. Until now there has not been an integrated approach to dealing with digital data and tools for the humanities; existing datasets are not interconnected, and researchers and students experience a lack of training in applying digital methods to analyse large datasets. CLARIAH aims to deal with these issues and find integrating solutions.

The CLARIAH infrastructure will give researchers access to large collections of digital data and innovative, user-friendly applications for the processing of these data. Both data and tools will be managed in a sustainable way, i.e. they will be easily accessible to humanity scholars such as linguists, historians and media scientists.

CLARIAH is the natural continuation and extension of two national projects: CLARIN-NL (linguistics, texts) and DARIAH-NL (socio-economic history, structured data), complemented with a third pillar: Media Studies (audio-visual data). CLARIAH combines the European CLARIN and DARIAH infrastructures in the Netherlands.

Beyond topic and time

Until recently, addressing questions about culture and identity depended on experts' ability to identify potentially relevant pieces of information in archives, libraries and museums. Because such research was extremely time consuming, it was hardly possible to look at all the data or test alternative analyses.

In the CLARIAH infrastructure digital data will be made accessible to a large number of researchers, who will be able to process them using digital tools. They will be well equipped to verify data selections and other researchers' interpretations. CLARIAH will open up their world by enabling them to search beyond the boundaries of topic and time, to tackle the vast universe of data and study the development of culture, cultural identities and ideas.

On 1 October 2013, the CLARIAH proposal for a 'National Roadmap for Large-Scale Research Facilities' was submitted to the Netherlands Organisation for Scientific Research (NWO). A decision about the project proposal is expected by June 2014.

If funding is awarded, the project will be launched on 1 January 2015. The five demo projects described here show the current and future possibilities of CLARIAH.

clariah.nl

TROVe Transmedia Observaory

Different kinds of media, such as newspapers, TV broadcasts, blogs and other online social media, are getting more and more entangled. It therefore makes sense to use all news sources available if one is interested in studying the diffusion of news items across media channels and over time. TROVe has built a search engine that allows scholars from the humanities to analyse multi-media content over time. Specific roles of the various media, their interrelationships and their mutual influences can be examined in one interface. TROVe not only allows the identification of key players, but also analyses the roles they play in discursive constructions. Eventually it will be able to show how the various media and actors (people, organisations) interact, in other words: how they influence public debate.

clariah.nl/en/trove/summary

Johan Oomen

OHT Oral History Today

Oral History collections are a rich basis for various types of scholarly research. However, scholars need tools to explore these collections in a way that is suitable for answering research questions. In the first phase of research, scholars need to select relevant collections. After that, they will want to analyse the selected materials thoroughly. The tool developed in OHT aids the scholar during the selection phase by means of keyword search, content visualisation using word clouds, and various filter options based on information available in the metadata, such as time periods and location. For the analysis phase, the OHT tool provides features like a personal folder to store interesting segments selected using a virtual cutter. The segments can be annotated and shared with colleagues. The tool also recommends related interviews or interview segments in the collection. oht.staging.dispectu.com/#
Roeland Ordelman

HLZ (HSN Links Zeeland)

Microdata on 'changes in life courses' provide a unique insight in how societies change over time. The Historical Sample of the Netherlands (HSN) collects this kind of data for the Netherlands and is largely based on municipal registers. However, some registers are incomplete, do not provide information on stillbirths and sometimes provide conflicting information. LINKS (Linking System for Historical Family Reconstruction) is based on civil certificates and aims at the reconstruction of all nineteenth and early twentieth century families in the Netherlands.

HLZ demonstrates the value of combining differently structured datasets by extending and enhancing the HSN database with data from the LINKS database for the province of Zeeland.

iisg.nl/hsn/index.html

Kees Mandemakers

CLIO-DAP

The purpose of CLIO-DAP is to improve communication in research by asking leading scientific journals in the field of economic and social history to make available the research data associated with articles they publish. The software should facilitate the process where data are deposited in reliable digital data archives; ideally, these data collections are also published and reviewed, and subsequently linked to the corresponding publications.

CLIO-DAP has built a demonstrator service for CLARIAH with enhanced publications to showcase existing output from socio-economic history journals with a Data Availability Policy (DAP) and to illustrate the workflows needed. It thus contributes to convincing stakeholders in other journals to adopt similar policies. cliodap.dans.knaw.nl

Leen Breurs

Nederlab

Nederlab's goal is to enable scholars in the humanities to find answers to new, longitudinal research questions. For this purpose Nederlab aims at setting up a user-friendly tool-enriched web interface, allowing researchers to simultaneously search, analyse, and enrich the digital historical texts made available by scientific libraries and institutes, at the text and metadata levels. The CLARIAH demonstrator is a working Nederlab prototype.

It consists of a website offering a Virtual Research Environment for diachronic research, with a transparent and efficient user interface that is designed to meet the specific needs of end users. In this respect, Nederlab differs from existing user interfaces such as those currently offered by CLARIN, which mainly focus on the supply of tools and data.

www.nederlab.nl

Nicoline van der Sijs

-