

★ The availability of vast amounts of data via the Internet has fundamentally changed research methods in the arts and humanities, but its widely scattered nature means it is often difficult to access, says **Peter Doorn** coordinator of the DARIAH project, who expands on the benefits of establishing a digital research infrastructure

Virtual infrastructure for humanities research

Humanities and arts research has traditionally involved extensive trips to archives, libraries and museums to study sources and analyse material. The advent of the Internet has revolutionised study methods however, and an ever-increasing amount of

resources are now available online, a trend which Peter Doorn of the DARIAH (Digital Research Infrastructure for the Arts and Humanities) project is keen to build on further. "The overall aim of the DARIAH project is to bring together resources and tools for the

arts and humanities, in order to improve their availability to researchers," he says. "The use of computers and digitisation in various areas within the arts and humanities has dramatically

increased over recent times. This has been especially apparent since the late '80s/early '90s in various fields, such as linguistics, social and economic history, and textual studies," continues Peter. "A lot of material has been digitised over the years and made accessible, but this work has faced all kinds of difficulties. The main problem is that this material is widely scattered – all kinds of institutions put material on the web, but this doesn't mean that it is easy to find, easy to reach or easy to use."

Enriching knowledge

Established to enhance and support digitally enabled research across the arts and humanities, DARIAH is working to address these issues. An EU-funded initiative which brings together 14 partners from 10 countries, the project aims to explore and apply ICT-based methods to enable researchers to ask new questions about their field, as well as to link distributed digital source materials in a coherent and

accessible way, which has the potential to greatly enrich our understanding of historical events. "Our knowledge of slavery, for example, has been transformed, widened and deepened because of the availability of databases on the slave voyages. This work has been going on for the past 20-30 years or so – it is quite painstaking, and many countries have been trawling their archives to find information such as shipping registers and so on concerning these voyages," outlines Peter. Bringing these disparate sources of data together allows researchers to gain a more detailed picture of their subject, an approach which DARIAH plans to replicate across other areas of the arts and humanities. "We are building an infrastructure to provide

the basic facilities on which researchers can work. Physical institutions like libraries, archives and museums are, in a sense, arts laboratories," he says. "In the digital age we need a parallel to that, a virtual laboratory for humanities research. That is what we are creating."

“**We are building an infrastructure to provide the basic facilities on which researchers can work. Physical institutions like libraries, archives and museums are, in a sense, arts laboratories. In the digital age we need a parallel to that, a virtual laboratory for humanities research. That is what we are creating**”

This encompasses a number of subjects from across the wider arts and humanities field, including archaeology, textual studies and social history. The scope of the project is too broad to work on all these data types simultaneously, making the 2 year preparatory phase of DARIAH, which involves detailed work on several demonstrator projects, crucial to its further development. "At the moment we are working on two demonstrator projects – one in the field of archaeology and one in textual studies. We are also currently negotiating to get another project off the ground in the area of Holocaust studies, in which the material from all the major archives on the Holocaust will be brought together," Peter states. Integrating this kind of data within a single infrastructure is no easy task; not only are there large volumes of data, but its varied nature is also an important issue, particularly given the broad geographical scope of DARIAH. "No one country has the same type of data as another," Peter points out. "They all have archaeological data, but the type varies widely. Greece has a particularly rich cultural heritage, for example, and much of its archaeological data is

from different periods and in different styles than that from Western Europe. And if we are talking about textual sources then of course the languages differ, while historical documents pertain to the national history of the individual country. These collections will be kept separate, although users

will be able to perform facile searches across the entire DARIAH infrastructure."

Sharing data

The willingness to share this kind of data varies widely, not just from country to country, but also between disciplines. Some researchers have historically been wary of sharing their data; Peter believes attitudes are changing, however. "In the past archaeologists tended to keep their own findings to themselves, but now they see how the material can be enhanced if it is available to everyone from a central place. For example one archaeologist could use material from his colleagues, who are maybe working on the same period but in another area, which enriches his knowledge of his own area of study. There is also cross-disciplinary use, but this happens less frequently," he says. Establishing the mutually beneficial nature of this relationship, and ensuring potentially sensitive data does not get into the wrong hands, is central to encouraging more researchers to share their findings. "Archaeological data sets can get very technical. If you were to make detailed information about archaeological sites available to everybody then every hobbyist would know where to dig and where to go with their metal detectors," Peter points out. "So if we are to protect



At a glance

Full Project Title

Digital Research Infrastructure for the Arts and Humanities (DARIAH)

Project Objectives

DARIAH will contribute to the efficiency and effectiveness of arts and humanities research by:

- ensuring data can be found and accessed without any need for extensive travel;
- making innovative interpretation tools available to the broader research community
- standardising tools and datasets to allow for interoperability.

Project Partners

- Please see DARIAH website for details

Contact Details

Project Coordinator,

Peter Doorn

PO Box 93067

2509 AB The Hague

The Netherlands

T: 0031 7034 94450

E: info@dariah.eu

W: www.dariah.eu

Peter Doorn

Photograph
by Wiebe Kiestra



Project Coordinator

Peter Doorn studied human geography in Utrecht and completed his PhD at the same university. He taught computing for historians at Leiden university between 1985 and 1997. He was director of the Netherlands Historical Data Archive and head of department at the Netherlands Institute for Scientific Information Services (NIWI). He is currently Executive Director of DANS (Data Archiving and Networked Services). He has acquired and directed a considerable number of digitisation and humanities computing projects.



our archaeological heritage we must ensure that potential treasure hunters cannot get access to the information, we need some sort of protection there. In the Dutch case we restrict access to detailed archaeological data – you need to be an accredited archaeologist in order to get access.”

Allowing access to data also opens it up to further scrutiny, meaning that it can be checked and verified, and its

as open as possible, but they also take into account some of the very justified restrictions that sometimes need to be imposed,” explains Peter. As such, the maintenance of the database is an important issue. Based at DANS (Data Archiving and Networked Services) in the Netherlands himself, Peter says there are clear differences in the roles and functions of the individual DARIAH partners. “We in the

“ **In the past archaeologists tended to keep their own findings to themselves, but now they see how the material can be enhanced if it is available to everyone from a central place. For example one archaeologist could use material from his colleagues, who are maybe working on the same period but in another area** ”

credibility established. Although the risk of data being misused cannot be removed entirely, it can be minimised with the imposition of certain limitations; in the case of DARIAH the project must also take the legal frameworks of participating nations into account. “One of the DARIAH work packages is looking at legal issues. This is in terms of licences to collect data – that is to get it deposited by researchers – and also to use it. The licence models that we have made are

Netherlands have the explicit task of acting as a data archive and to preserve the data that we make available over the long-term. But not every country has that, so DARIAH itself, as an organisation, will not take responsibility for long-term preservation of information,” he says. “DARIAH will certainly maintain the systems that it develops, but the data on which it applies – that will be the task of the individual Member States themselves.” ★

