DANS: a new data initiative in the Netherlands

Introduction

DANS, the acronym for: 'Data Archiving and Networked Services', is the new Dutch organisation tasked with the preservation and permanent access to research data in the humanities and the social sciences. The Royal Netherlands Academy of Arts and Sciences (KNAW) and the Netherlands Organisation for Scientific Research (NWO) are jointly responsible for DANS. DANS aims to promote (interdisciplinary) collaboration with and among researchers. DANS is also a Dutch national partner in European and other international data organisations. DANS will assist foreign researchers in accessing Dutch data and will be an intermediary for Dutch researchers in retrieving data from abroad. DANS takes the form of a network with a centre which is responsible for the organisation of the data infrastructure. From October 10th DANS will announce its activities on the website www.dans.knaw.nl.

Reasons for merging three institutions

Three existing data archives merge in DANS: The social science data archive 'Steinmetzarchief' and the Netherlands Historical Data Archive (NHDA), that were part of the Netherlands Institute for Scientific Information Services (NIWI) from 1997 until 2005, are now part of DANS. In addition, the tasks of the Scientific Statistical Agency (Wetenschappelijk Statistisch Agentschap – WSA), have been taken over by DANS. WSA was part of the Dutch Research Council, and was the intermediary for the delivery of large data bases from organisations such as Statistics Netherlands, The Social and Cultural Planning agency and the Topographical service/Kadaster to researchers. From now on researchers can order and deposit their data sets for archiving by DANS.

There were three reasons for this merger[1]:

- Researchers in the field of humanities and social sciences need interoperationability of data-archives.
- The need for technologic innovation of archives and data infrastructure.
- To enhance efficiency by creating a 'One stop shop' for researchers that need data and institutions that create relevant data, by bundling split up activities of several organisations.

---

[1] Described in two documents: 'Behouden toekomst. Een advies met betrekking tot de toekomst van de diensten van het Nederlands Instituut voor Wetenschappelijke Informatie', Oktober 2003, dr. N.M.H. van Dijk and 'Networked Data Services. Towards a future data infrastructure for the social sciences in the Netherlands. An advisory report by the Social Sciences Council', Royal Netherlands Academy of Arts and Sciences, Amsterdam 2003.

Network structure

Archiving and secondary use of research data in the humanities and the social sciences has been very important for the development of the field in the last four decades. In many countries, also in the Netherlands, several organisations (Steinmetz, NHDA, WSA, university faculties) started to archive research data. Digitalisation of the data and the possibilities raised by the development of the internet has changed this process enormously. Unfortunately, the research data archiving activities have been split up between several institutions. Sometimes this was caused by institutional reasons, sometimes this was driven by the researchers themselves. Because of this research data was difficult to find, obtain and use. Besides these difficulties, for scientific use of research data, the data needs to have certain minimal quality level. If the source or method to obtain research data is not clear, the data can not be used for scientific purposes. The internet has now reached a level of usage and the speed capability that researchers rightfully expect different data sources to be combined in one national data-archiving network. This network must then be linked to international sources[2]. Having one organisation steering this process by making agreements with data producers, one can guarantee the quality of the research data and the continuality of the available data.

Structure: a single, core organisation that combines specialist knowledge of the humanities and social sciences fields with technological expertise in digital information and communication structures. (Figure 1).
Organising principle: a decentralised network; a strong core in a cluster of satellites, local data centers (LDC's).
The core will bear responsibility for organising and supporting the network, for the basic ICT infrastructure, and for the method and means of communication.
The decentralised 'hubs' will bear responsibility for the specific thematic or disciplinary expertise. The hubs will be prominent institutes and research networks with a leading role within the European context. The model is an open one and will be able to embrace new, promising fields that are as yet unable to play such a leading role in Europe.

---

[2] The importance for international access was clearly stated by the OECD in the January 2004 "*Declaration on access to Research Data from Public Funding*."
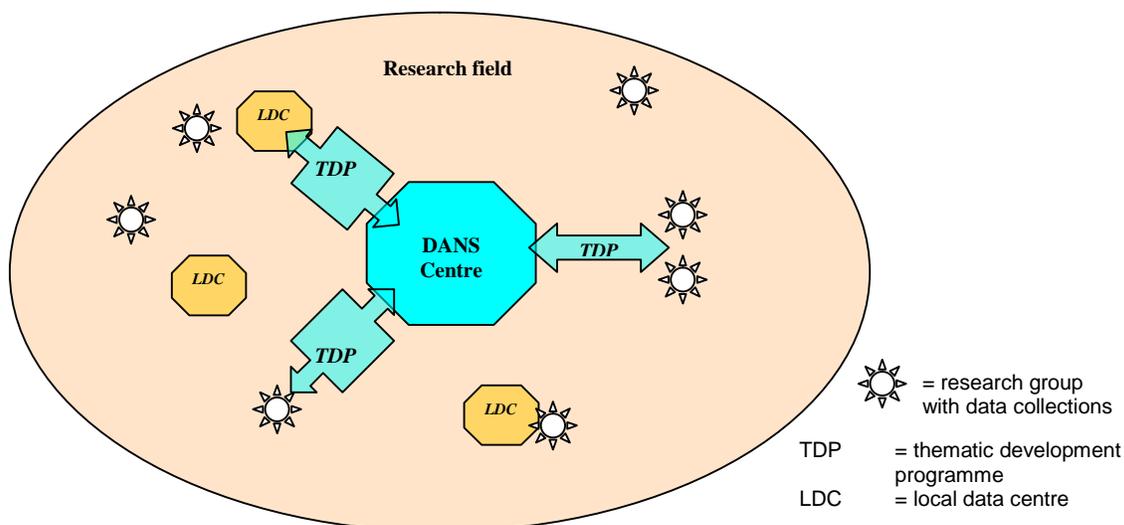
Figure 1: DANS core linking TDP's and LDC's

Thematic Development Programmes

In addition to continuing existing services, DANS has also been mandated to expand and improve the current data research infrastructure. New initiatives for data services and data archives are being developed in fields where such facilities have been lacking until recently. The new initiatives are organised as 'Thematic Development Programmes' (TDPs), which are set up in collaboration with research groups. Archaeologists and researchers in literary and linguistic computing are already preparing a TDP in their respective fields. In the area of the social sciences, DANS is working with Statistic Netherlands on creating access to a conglomerate of registration data bases and sample surveys, which are known as the 'Social Statistical Database', from which the Census 2001 was extracted. DANS has taken on the challenge to create three TDPs within one year. New infrastructure for data archiving and 'Open Access' is being created. Work on the renewal of the infrastructure for digital data-archiving is in progress. For the acquisition of data sets, collaboration with faculties, research schools and academic institutes will be invigorated. In the near future researchers will be able to place their research data in DANS compliant data repositories. This means that they can always rely on a safe backup of their own data, as well as share their data with other researchers. DANS aims to make access to research data as open as possible, in accordance with the 'Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities' of 22 October 2003.
This declaration requires researchers to make their scientific materials (publications, data, etc.) freely accessible to everyone through repositories on the Internet. This declaration has been signed by a large number of leading scientific organisations all over the world.
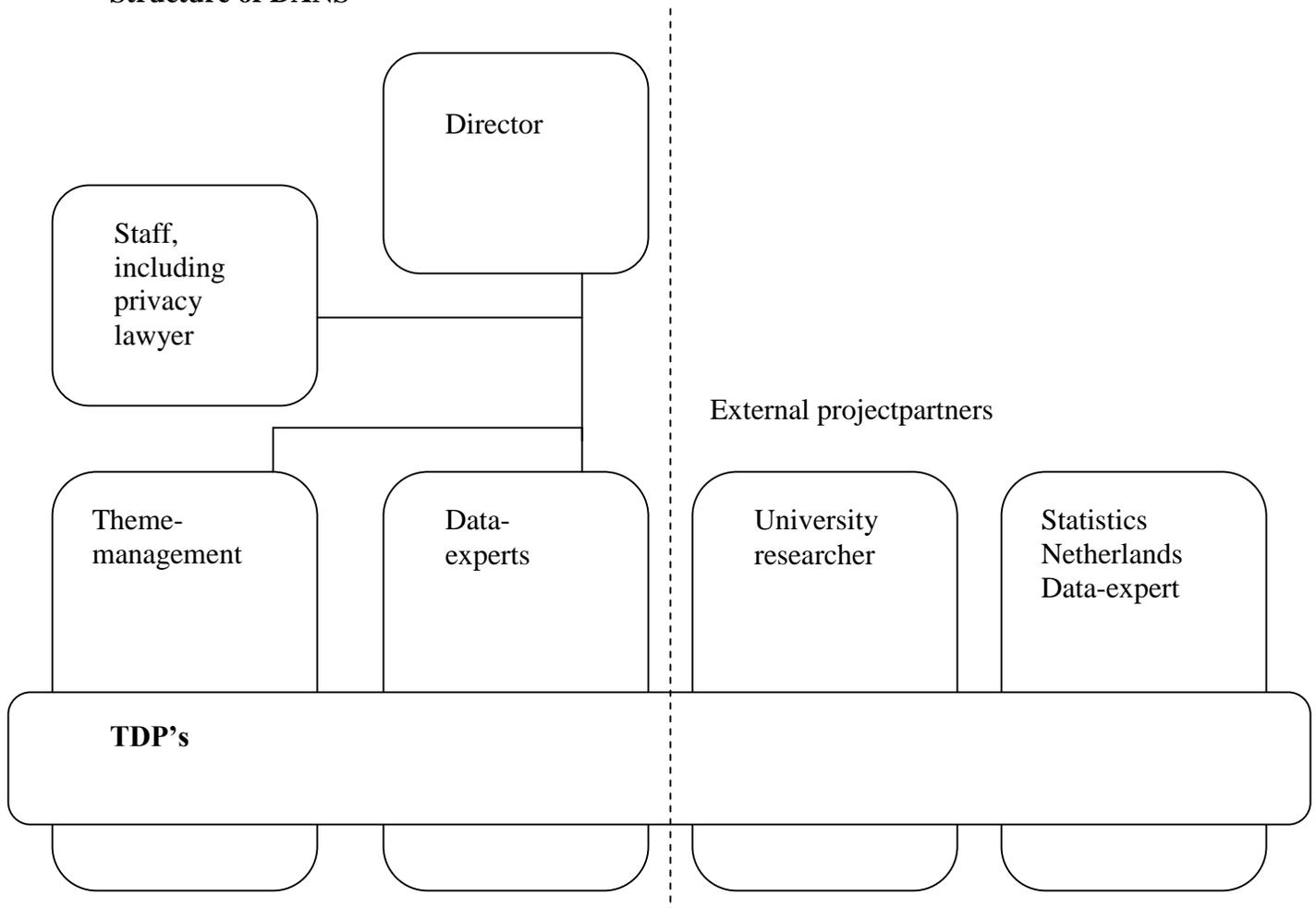
Structure of DANS

Figure 2 structure of DANS

The structure of the DANS organisation will be a matrix organisation. (figure 2). Thememanagers will run TDP's as projects with dataexperts and staff support in projectteams containing several university researchers.

Paying for Data?

One of the organisations to merge in DANS was the Scientific Statistical Agency (WSA) of the Dutch Research Council. This organisation used contracts to acquire data from large (government) institutions. Sometimes these organisations were payed to get scientific access to their data. The philosophy of WSA was that paying for data was better in order to get access, then getting no access at all. Researchers also had to pay WSA to get WSA data, but costs were kept low for university researchers. The amount WSA payed to get access to data, was always higher that the amount yielded by researchers

paying for the data. Sometimes however; data contracts needed only a small investment of WSA and semi-commercial researchers paid most of the contract. This meant non-profit researchers also got access to this data for a minimal fee. Not every commercial researcher could get access to the data. Researchers had to work at institutions that were complaint to the criteria set by the independent Central Commission for Statistics (CCS) of Statistics Netherlands. The most important criteria being the research institute must make all results public. The data researchers could order at WSA were modified to meet Dutch privacy laws.

Contracts

WSA always made contracts 'three party arrangements' (figure 3). WSA being the intermediary, the researcher being the user and the institution that gathered data being the provider. For judicial reasons the user was defined as the faculty where the researcher works. A faculty that ordered the data bought the right to have all their researchers use the data.
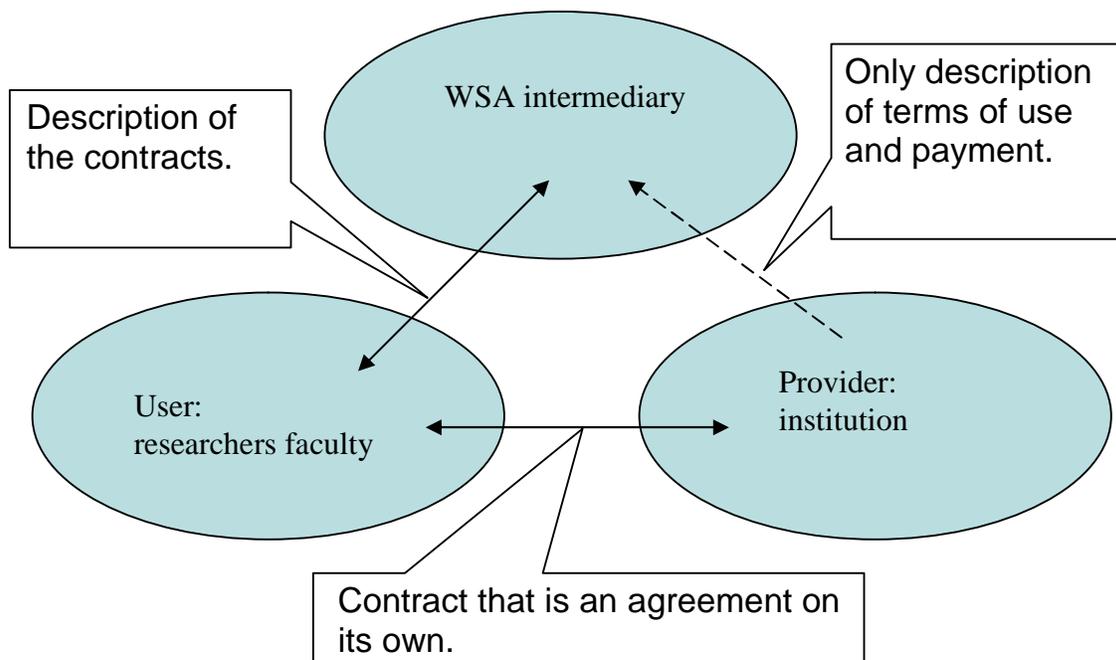
WSA intermediary

Description of the contracts.

Only description of terms of use and payment.

User: researchers faculty

Provider: institution

Contract that is an agreement on its own.

Figure 3 Three party contracts

Some contracts only brought temporally access to data. In those contracts there is a special clause for using the data for archiving purposes. DANS will take over the WSA

contracts although DANS will try to make the financial burden for researchers as low as possible.

Conclusion

The humanities and social science data infrastructure in the Netherlands has been greatly improved by the creation of DANS. Researchers now have a 'one stop shop' for ordering data. With thematic development programs, driven by researchers, DANS will have a better focus on the needs of researchers in the humanities and social science field.