ACCUMULATING ARCHIVES
DANS - Data Archiving and Networking Service 2005-2010

KNAW Evaluation Report
November, 2010
CONTENTS

Preface 3

1. INTRODUCTION 4

2. REVIEW OF THE INSTITUTE 5
   2.1 Mission, policy, and strategy 5
   2.2 Leadership 6
   2.3 Resources and funding policies 6
   2.4 Facilities 7
   2.5 Academic reputation in general 7
   2.6 Societal relevance 8
   2.7 Strengths and weaknesses 8

3 ASSESSMENT OF THE RESEARCH 10

4 RECOMMENDATIONS 13

APPENDIX
   1 Programme evaluation committee 16
   2 References / Received material 18
PREFACE

This is the first evaluation report written about DANS (Data Archiving and Networking Service). The institute was founded in 2005, is funded by both KNAW and NWO and its mission is to coordinate research databases on a national level. It deals with storing and disseminating data, and therefore differs from other KNAW institutes, in the sense that it is not primarily a research institute.

This posed a challenge for the visitation committee that was invited by the KNAW to conduct the first full evaluation of the institute. The committee consisted of several scientists and specialists in the field from Dutch as well as international universities and institutes.

DANS proved to be a kind and generous host during the two days that the committee visited DANS headquarters for the site visit. After two days of interviews and exchanges of information, the first findings were presented in oral form: this is the written version that was promised at that occasion.

On behalf of the committee, The Hague, February 2011,

H.J. van den Herik, chair
B. Mourits, secretary
1. INTRODUCTION

Once every six years, each KNAW institute is assessed in accordance with the *Standard Evaluation Protocol 2009-2015 for Research Assessment in the Netherlands by KNAW, VSNU, and NWO*. The specific aims of this system, as well as the aspects on the basis of which an evaluation takes place and the ratings that can be given, are described in this protocol and therefore not repeated here.

The review of DANS (Data Archiving and Networking Service) was the first full assessment of this institute that was founded in 2005. A midterm review had taken place in 2007. The current review was commissioned by the board of management of the KNAW. The committee was appointed in June 2010 and consisted of:

- Prof. Jaap van den Herik (chair) – Tilburg University
- Prof. Heinrich Best – Friedrich-Schiller University of Jena
- Prof. Frank van Harmelen – Free University Amsterdam
- Prof. Bert van der Knaap – Erasmus University Rotterdam
- Prof. Julian D. Richards – University of York
- Prof. Kevin Schürer – University of Leicester

Dr. Bertram Mourits was appointed secretary of the review committee.

The assessment dealt with the work performed in the institute, which entails providing permanent access to digital research data for the humanities and social sciences. DANS’ position in the KNAW is unique, for two reasons. First, it is the only institute that is funded by the KNAW and NWO. Second, it is the only institute where the emphasis is on service, not research. The Standard Evaluation Protocol therefore does not apply completely to DANS.

The period of assessment was 2005 until 2010. The review committee had received a detailed self-evaluation as well as several other documents – see the Appendix 2 for an overview. Information available in the self-evaluation is mostly not repeated in detail in this report.

The committee interviewed the directors and employees of DANS, the stakeholders (KNAW and NWO) and a group of users. The committee also attended several presentations about the work: national as well as international projects dealing with service as well as research aspects. These interviews took place during a site visit on November 22 and November 23, 2010. The program of the site visit is included in the Appendix (1).
2. REVIEW OF THE INSTITUTE

Because, as has been said, DANS is not a research institute like the other institutes functioning under the umbrella of the KNAW, not all aspects of the standard evaluation protocol applied equally well. In an attempt to make this report as useful as possible, we attempted to comply with the structure as dictated by the SEP.

2.1 Mission, policy, and strategy

DANS came into being as a result of an advisory report from December 2004, in which KNAW and NWO suggested to combine the work of several data archiving units (among which the Steinmetz Archive and the Netherlands Historical Data Archive of the former KNAW institute Netherlands Institute for Scientific Information Services, NIWI-KNAW, and the Scientific Statistical Agency of NWO). The explicitly formulated mission was for DANS to ‘improve access to research databases at the national level, and to guarantee long-term access to such databases, for researchers in the social and behavioral sciences and in the humanities’.

Both points led to successes; in addition to the social science and historical archives, DANS now provides a vital digital archive for Archaeology in the Netherlands: it is one of only three groups working in this sector so far in the world. Currently, DANS’ strategy is to remove barriers for access to data along two lines: (1) researchers may deposit their data with DANS, and (2) users are encouraged to mine DANS databases, which are not limited to academic generated research material but also include government data and other materials, both publicly and privately funded.

A significant part of DANS’ activities aims at innovation and is discipline independent. We mention four of them: (1) work on a general Data Seal of Approval (guidelines and best practices), (2) the creation of persistent identifiers, (3) the storage and technical archiving, and (4) legal matters. Other work is discipline-dependent (formulating subject-specific meta data, subject-specific standards).

DANS’ ambition – formulated in a number of strategic goals – is to increase its scope in two ways. First, by entering into the field of natural, technical and life sciences. DANS wants to become the ‘enabler’ for one national data organization for research (with links to SURF, NCF, NCDD). The second way is to initiate more research projects by itself (e.g., in the fields of digital data archiving and high performance computing for data processing).

Besides these strategic goals, concrete goals of the organization are: increase the number
of files, users, research areas, participation in other projects, and co-operation with other institutes.

2.2 Leadership

DANS has a devoted, knowledgeable and dedicated staff. Most of the people working at DANS have a humanities background (with an emphasis on computing and/or information management) or an IT background.

Since its inception, DANS has grown considerably. More storage of data, more users, more research projects: DANS has built successfully on the tradition of the institutes on which it was founded. The adjustments that have been made in the last couple of years have been executed quite successfully: the organization functions well and organically.

DANS is an innovative center with a strong national and international reputation - it has shown leadership in the development of the Data Seal of Approval. Peter Doorn, its director, shows leadership and is ambitious in his attempts to broaden the scope of DANS. The institute now works mainly for humanities and social sciences, with a strong emphasis on social and political studies, and less focus on economics. In several strategic goals the ambition becomes apparent. Four strategic goals (out of a total number of 12) were discussed with the committee.

1. DANS provides data services to at least 10 new research areas, including at least 6 in the sciences.
2. Data Seal of Approval used by at least 100 repositories at home and abroad.
3. Introduction of data reviews: in 2015 50% of the data sets that have been downloaded more than 10x have been reviewed over 5x.
4. Formulation and implementation of an R & D program on the cutting edge of e-Science, e-Humanities and e-Social Science; in 2015, at least 5 R & D projects.

The committee was critical about several of these goals (see chapters 3 and 4 for the assessment and recommendations) but they do show that DANS is ambitious and has a wide focus.

2.3 Resources and funding policies

The core funding of DANS is the responsibility of both KNAW and NWO, but there are many projects that are also partially funded by outside sources. With 40 people working at DANS, and
more than a third of its work financed on a project basis, the outside funding plays an important part.

Next to the established resources by KNAW and NWO, DANS has developed the following additional funding policies. DANS aims, successfully, to capture a high proportion of contract work. Furthermore, DANS is a partner in several European organizations. The size and depth of this involvement is impressive. For example, DANS has taken on the coordination of a project with 20 partners and a 3.7 million Euro budget. Then, it has a partnership in three important European programs.

Moreover, DANS has taken the initiative in several projects that aim to improve the data infrastructure in the Netherlands and Europe; sometimes as leader of the project, sometimes in its role as a service institute. Funding for these projects is arranged in several ways, sometimes through KNAW and/or NWO, sometimes through European funds. There are certain risks though, in the dependence on outside funding. We mention two of them (1) does it yield the kind of data that are most needed? (2) Is it a good strategy to play a minor role in many projects? Especially the second point is crucial. Yet, it should be noted that the external funds make up part of DANS’ institutional success.

2.4 Facilities
Infrastructure is obviously a crucial factor for an institute, which is predominantly dedicated to data. Storage, maintenance, and online services are among the most important tasks. The facilities to perform these tasks are ample available and maintenance is an essential part of the policy. The committee sees no problems here: the IT-component in the Institute is strong and obviously, most of the material is available online through several websites and the “EASY”-archiving system.

Still, compatibility with archives in other institutes could be improved (the CBS, research institutes at universities in particular) and more attention should be paid to the data demands and requirements of graduate and postdoctoral students.

2.5 Academic reputation in general
The academic reputation in general is difficult to assess in case of a service institute. The academic value of DANS’ work is decided by the research projects that make use of the data, a factor in which DANS’ staff cannot be decisive since it largely does not initiate these projects.
itself.

The two important tasks for DANS are to warrant (1) the quality of the data and (2) the reliability of references. The second task is reflected in DANS’ commitment to the creation of persistent identifiers.

As for the quality of the material: before a data set can enter the DANS-archives it must conform to DANS’ standard norms: archives will be judged and, if necessary, upgraded to these norms. The norms are no longer DANS-specific, since the inception of the Data Seal of Approval. The DSA has set a standard and now functions independently of DANS, which is definitely a success (see above).

2.6 Societal relevance
The societal relevance for DANS lies foremost in providing research material in several scientific fields. However, for the future it may be expected that the scope of its reach might become wider. Especially the research material in the fields of humanities and social sciences are relevant to a wider audience. DANS has initiated several projects that make this information accessible to such a broad audience. Two examples will illustrate this statement. “Wat Veteranen Vertellen” (What Veterans Tell Us) is one of these, in which Dutch the stories of war veterans are the ground material for oral history; “Getuigenverhalen” (Stories of Witnesses) is a second example. Also it is clear that websites of municipal archives are visited often by a larger public than just researchers.

In contrast, DANS has been less successful in raising visibility among (potential) users. The institute recognizes this problem. The midterm review already recommended DANS to increase the ‘brand awareness’, but this is, in 2010, still one of the areas which needs work (see the Self Evaluation, p. 31-32).

2.7 Strengths and weaknesses
The SWOT-analysis in the Self Evaluation is mostly realistic and convincing. It is remarkable though, that EASY is not mentioned here at all, while it is one of the most important products of DANS’ work. In an attempt to broaden its scope, in research as well as subject matter, DANS runs the risk of losing sight of its basis activity.

On the positive side we mention that among the most notable strengths there is the key role DANS has in research infrastructures in social sciences and humanities, nationally and
internationally. To be specific: (1) the expertise in the fields of data, digital durability, direct access, legal matters; (2) its strong orientation towards innovation, and (3) its strong network, including research funders, large (institutional) data collectors, foreign and international organisations.

One weakness of DANS – in its own estimate – is the fact that it does not have an own research program. This means there is a ‘dissimilar supply of data for different disciplines’. Supply of data is research driven, and the institute wishes to make the research program more coherent, and therefore the influx of data more focused.

However, when DANS communicates as one of its weaknesses, that ‘the role of DANS in projects with researchers [is] more focused on carrying out the work than on cooperating’, it seems that the importance, as well as the quality of its own contributions is being underestimated.

Obviously, new opportunities are sought in data acquisition in the health sciences, life sciences, and natural sciences, but this could be risky as well, when DANS loses focus on its main activities. Other opportunities – more cooperation, international as well as interdisciplinary but within the same fields DANS is active now – seem more promising.

Among the threats, the institute itself sees (1) costs of storage and (2) high expectations among researchers with regard to DANS’ capabilities as software developers.

The strong (and almost exclusive) reliance on internally developed software can turn into a risk factor. There are good reasons why most large professional organizations rely on external software vendors. More use should be made of externally available software either commercially or in the open source world. Selecting software that complies with open standards is sufficient warrant against the fear of commercial vendor lock-in. Examples of this are map-overlay software and dBase converters. The Streaming Media Server project was a good example of a mix of buying where possible, developing where necessary. It is advisable to seek out competitive offers with regard to storage possibilities; and explore other possibilities (possibly open source software) when looking for new software.
3. **ASSESSMENT OF THE RESEARCH**

The assessment of the research is the part of the Standard Evaluation Protocol that is the most difficult to apply to DANS. As mentioned, the institute provides the facilities for research and mostly does not initiate research itself. Here the committee faced with two problems: (1) how to assess the role of DANS with respect to research, and (2) what performances elsewhere in the world could be used as comparison. DANS’ reason of being is possibly most closely comparable to that of research libraries, for which the SEP provides no standards to measure quality.

This problem was also addressed in the Self Evaluation by a thorough review of the strategic goals, which were in the strategy policy document and the interim evaluation. In this report we follow a more general approach, describing as much as possible the way in which DANS’ activities support research, and assessing those activities.

The committee addressed this problem along a variety of directions. One of them was meeting with users who were invited to talk about their research. In this way, the committee hoped to obtain an impression of the way in which DANS contributed to scientific research. It proved to be a useful addition to the program.

The integration of archaeology was hugely appreciated, DANS’ role in the preservation of cultural heritage is big. Enriched publications and new types of publication are becoming possible. The EASY archiving system works well, particularly for those who have discovered the possibilities.

In other words, DANS’ value for research became apparent in these conversations. It was also clear that more cooperation between the Dutch data-archives would improve the conditions for research: most often mentioned was the Central Bureau for Statistics but also material from the Royal Library (KB), National Archive, Sound and Vision (Beeld en Geluid) and Cultural Heritage (RCE) should be more integrated within the existing infrastructure.

So, even though the direct academic impact of DANS is limited, the focus on research is very important indirectly. Hence, we see the following four functions of the institute:

1. collecting data
2. providing data
3. facilitating research
4. initiating research
The first two functions are the proverbial bread and butter of DANS, the other two are important but should be seen in relation and service to the main functions of the institute, that is mostly an ‘enabler’ for research.

Furthermore, DANS’ explicit wish to devote more time and facilities to research, seems an underestimation of the amount of research performed at the institute. The committee was impressed by the variety and quality of the various projects that were presented. Contrary to some of the comments in their own report, these projects demonstrate that DANS is actively participating in research. DANS’ research activities are also confirmed by the reputation for innovation that DANS has among its sister organizations, and by the international visibility of some of DANS’ projects and members of staff.

From an international perspective DANS is hugely respected, without a shadow of a doubt, technologically, intellectually, and strategically. In conclusion, the assessments yield the following results with regard to DANS position for the academic community, society in general and education.

Institute
The quality of the institute is very good. It functions well, has adapted to new circumstances. On the field of productivity: this is good. There is no question about the relevance of the institute: good as well; it has an important function, which it handles with responsibility. One aspect deserves attention: communication with users, especially students, is satisfactory. Vitality and feasibility are very good – the feasibility might even be close to excellent, thanks to international initiatives and the strong reputation it has in the field.

Research
DANS’ function as a research-assistant institute is clear. It helps in delivering research and international programs of high quality. Here DANS’ contribution should be considered very good, with respect to quality as well as productivity. Societal relevance is good; and so is the vitality of the institute. Looking at the future, the feasibility of DANS is satisfactory. Research ambitions should be handled with care – DANS’ core functions might have to be protected more carefully to warrant its feasibility in the future.
### Assessment of Institute Research

<table>
<thead>
<tr>
<th></th>
<th>Institute</th>
<th>Research</th>
</tr>
</thead>
<tbody>
<tr>
<td>quality</td>
<td>very good</td>
<td>very good</td>
</tr>
<tr>
<td>productivity</td>
<td>good</td>
<td>very good</td>
</tr>
<tr>
<td>relevance</td>
<td>good / satisfactory</td>
<td>good</td>
</tr>
<tr>
<td>vitality and feasibility</td>
<td>very good / excellent</td>
<td>good / satisfactory</td>
</tr>
</tbody>
</table>

In other words, although there is obviously room for improvement, for a relatively young institute, with ambitious leadership and a devoted and knowledgeable staff, the committee is optimistic about possibilities for the future.
4. RECOMMENDATIONS

The midterm review resulted in a number of recommendations, which the management of DANS tried to follow; their attempts are the core of the self evaluation. Our recommendations first deal with the current state of affairs: ambition should not get into the way of doing best what DANS was set out to do. Our recommendations are as follows.

- Evaluate continuously the way in which data are acquired. A large amount of data sets are acquired through research projects and it is not always clear how relevant these materials will turn out to be for later research: try to prevent significant parts of the DANS archive to become largely unused.
- A related recommendation: show more initiative in actively acquiring data sets. Archaeology was a good example of how new data were integrated and used. This experience should be turned into a strategy of productively working with new data.
- As mentioned before: the users. This is currently one of DANS’ more vulnerable points: the relationship with users and what they need. DANS’ major goal should be to increase the foothold in the social science and humanities and focus on students: providing lecture material for lecturers, providing projects or exercises to be used in courses, etc. More direct and serious communications with users could improve the efficiency of the institute in several ways.
- We recommend to capitalize more on the high reputation of Dutch social sciences in the international social science community. The absolute key is the richness and quality of the data they hold. They have to develop and maintain a collection that researchers want to use and are essential for research purposes. The social sciences are an area where DANS might visibly ‘add value’ to research, which is indispensable for survival. DANS needs to make itself absolutely core and indispensable to at least one key research group.
- The issue of ‘brand awareness’ does not receive sufficient attention from DANS employees. DANS should carry out an active policy to increase brand awareness, also among students. PR-activities in research schools and universities seem straightforward, but are probably an effective way to increase the use of existing files, ultimately the reason of being for DANS. Reaching students is for DANS a relatively cost-effective way of keeping and increasing their market in the future.
- Try to establish more effective cooperation with other data centers. At universities –
some of which have their own data centers, with the Royal Library, National Archive and also – especially – the CBS. Because of their strong foothold in the social science, strengthening the contacts with CBS is crucially important to offer deeper access to data which is not to be neglected.

- We suggest DANS make more use of externally available software either commercially or in the open source world. Selecting software that complies with open standards is sufficient warrant against the fear of commercial vendor lock-in. Examples of this are map-overlay software and dBase converters. The Streaming Media Server project was a good example of a mix of buying where possible, developing where necessary. It is advisable to seek out competitive offers with regard to storage possibilities; and explore other possibilities (possibly open source software) when looking for new software.

Then there are recommendations specifically related to the strategic goals in the vision memorandum.

- With regard to the wish to initiate research projects the committee wants to stress that substantial increases in these existing research activities should only be considered under the condition that they do not distract attention or funding from any DANS primary activity, namely the acquisition, storage, dissemination and support of archived research data.

- The committee recommends to be careful with regard to increasing the number of fields in which DANS is active. Stressing this recommendation somewhat: the institute should aim to be a core element to a few communities instead of being peripheral to many. It is not a matter of modesty, it could turn out to be a matter of survival. DANS wants to be 1) a repository of the most important/best/popular datasets, 2) an exhaustive archive and 3) a research institute. It will be impossible to be all three at once.

- Broadening scope within the social science and humanities should be a key effort. Further developments can start from there. If DANS wants to move in the direction of beta studies, they should think of social health and medicine.

- DANS wishes to initiate research itself, and we recommend to proceed most carefully with this ambition. To be a research institute, instead of a service institute, demands a different organization structure. The strength currently lies in the combination, whereby the applied aspects are stronger than the theoretical qualities. It will take more than good intentions to change this and DANS runs the risk of neglecting its core activities by aiming
at too many different activities on a thin basis. The risk of uncontrollable growth should be avoided.

- If this route is pursued, we suggest writing a business plan for a modest expansion, not in the “hardcore beta”-field, but in a neighboring field, such as social health or economics. Do not underestimate the danger of cultural differences (they are the most prominent risk-factor in commercial mergers and acquisitions and there is no reason to assume that science is different).

- Furthermore, it is a question whether beta sciences need a centralized approach, and since DANS has no roots in these communities, it might be more difficult and not productive.

- The committee recommends that DANS remains a service organization that also performs some research, and should not aim at becoming a research organization that also performs some services.

New initiatives should preferably be funded by independent sources, thus making sure that these would not hurt the basics of collecting and providing data.

In conclusion, DANS is a very good service institute with a high international profile in its professional environment. Many of their data and data storage are being used internationally as well as nationally. We would encourage the Dutch potential customers to make more use of the facilities. All in all, DANS is better than its size would lead one to expect and is successfully ‘punching above its weight’. DANS is providing very good value for money to NWO and KNAW as a successful enabler of social science and humanities research, and as an internationally leading scientific data-archiving institute.
APPENDIX

1. PROGRAMME EVALUATION COMMITTEE DANS

Monday November 22

9:30 Welcome and introduction of the evaluation committee (plenary) Aims of the evaluation by prof. Jaap van den Herik, chairman of the Evaluation Committee, and Peter Doorn, director of DANS (plenary)
10:00 General overview DANS the first 5 years by Peter Doorn
10:30 DANS Business by Henk Harmsen, vice-director
11:00 Coffee break: poster session: the communication strategy by Martijn de Groot, communication advisor, and Brenda Sorensen, communication officer
11:30 The Archive by Laurents Sesink, coordinator Archive Group
12:00 Infrastructure by Dirk Roorda, coordinator Infrastructure Group
12:30 ICT Services by Rutger Kramer, coordinator Software Development Group
13:00 Lunch – Only members of the evaluation committee present
14:00 A selection of past and present DANS projects: Introduction to presentations by Henk Harmsen
14:15 Presentations by the respective project leaders (in parallel sessions)
   - The MIXED approach to digital preservation by Jan van Mansum, Vesa Akerman and Dirk Roorda
   - Data acquisition by Hella Hollander, Jetske van der Schaaf and Laurents Sesink
   - European Research Infrastructures: DARIAH / CESSDA / CLARIN by Milena Piccoli, Heiko Tjalsma and Marjan Grootveld
   - Presenting and access to social history data: HASH / NLGIS / Corporate Collective actions by Maarten Hoogerwerf and Rutger Kramer
   - EASY (Electronic Archiving System) / Dendrochronology (technical) by Paul Boon and Jan van Mansum
   - Enriched publications by Paula Witkamp and Marion Wittenberg
15:00 Tea break
15:15 Presentations continued 16:00 SWOT analysis by Peter Doorn
17:00 Reception (plenary)
18:00 End of the first day of the site visit to DANS

Tuesday November 23

9:30 A selection of present and future DANS projects: Introduction to presentations by Henk Harmsen
9:45 Presentations by the respective project leaders (in parallel sessions)
   - Multimedia and qualitative data by René van Horik and Ishan Sital
   - Alfalab by Maarten Hoogerwerf and Leen Breure
   - Data Seal of Approval (DSA) by Henk Harmsen and Henk Koning
   - Persistent Identifiers (PIf) by Laurents Sesink and Dirk Roorda
10:30 Coffee break: Poster session/exhibition: Small Data Projects by René van Horik, Lisa de Leeuw and Heiko Tjalsma
10:45 Presentations continued
11:30 The DANS Vision and New Strategy by Peter Doorn
12:30 Lunch
13:30 Meeting with DANS Steering Committee and the chairman of the Scientific Advisory Board
14:30 Meeting with users of DANS
15:30 Meeting of Evaluation Committee: towards the draft report
16:30 Presentation of the first preliminary conclusions to the personnel of the Institute by the Evaluation Committee (plenary)
17:00 End of the site visit
2. RECEIVED MATERIAL / REFERENCES

- The first five years of Data Archiving and Networked Services: Self-assessment DANS 2005-2010 (2010)
- DANS: design for a national organisation for a research data infrastructure in the Humanities and the Social and Behavioural Sciences (2004)
- Mid-Term Assessment of DANS by the Scientific Advisory Council (2008)
- [collection of pdf's from the presentation during the site visit]