

What are your information needs? Three user studies about research information in the Netherlands, with an emphasis on the NARCIS portal

Arjan Hogenaar; Marga van Meel; Elly Dijk

Royal Netherlands Academy of Arts and Sciences, Research Information,
P.O. Box 19121, 1000 GC Amsterdam, The Netherlands

Arjan.Hogenaar@bureau.knaw.nl

Abstract

The NARCIS portal (www.narcis.info) provides access to science information (information about research, researchers and research institutions) and scientific information (full-text) publications and datasets. The portal is very popular, with 1.2 million users annually. NARCIS is also an important supplier of information to international services such as Google/Google Scholar, WorldWideScience.org and DRIVER. In 2009, the KNAW conducted a three-part user survey, with two online surveys and a series of semi-structured interviews. The aim was to learn more about the people who use the portal, why they use it and their ideas and wishes for improvements to the portal. Another purpose of the survey was to identify changes that could be made to improve the match between the services provided by NARCIS and the needs of existing and potential users. Surveys showed that more than half the users of NARCIS are from universities, research institutions or universities of applied science. Most searches conducted on NARCIS are for dissertations. The existence of a single gateway to different types of information is regarded as very useful. The most frequently mentioned improvement in the service would be to provide access to information from other countries as well. Respondents also mentioned the provision of *tools* for performing complex analyses of the information available via NARCIS as a worthwhile option for enhancing the service. The interviews revealed, among other things, the need for the presentation of information in context and that senior officials are often confronted with information overload. The user survey has led to a series of proposals for modifications or improvements in

the service; some of them may be implemented immediately, while others will require consultation at national or international level.

Key words: user survey; questionnaires; portal; evaluation of integrated services

1. Introduction

The most important task of the Research Information (KNAW-OI) department of the Royal Netherlands Academy of Arts and Sciences (KNAW) is to help national and international users to find information about research, researchers (and their expertise), research institutions and the results of research (publications and datasets) in the Netherlands.

The Dutch Research Database (NOD)¹ is a service provided by KNAW-OI and forms the basis for its role as the national focal point for research information. Before 2005, the KNAW was involved in the development of DAREnet (network for Dutch Academic REpositories) [1], at the time the central portal for access to publications in the repositories of research institutions. Since then, KNAW-OI has been developing the national focal point for research information and research results at European level. The final result is NARCIS (National Academic Research and Collaborations Information System) [2].

NARCIS now plays a central role in searching all research-related information in the Netherlands and serves as the national showcase for researchers working in the Netherlands. Via NARCIS users have access to both the information from the Current Research Information Systems (CRISs) and the information from the Open Access Repositories (OARs).

A problem is that in the Netherlands the (OARs) and the CRISs generally fall under different organisational units of the universities: the libraries or the research administration departments. The datasets, if they are kept at all, are stored in the DANS² system. With all these different systems, it is inevitable that variant versions of the names of authors and researchers are in circulation. To cross-reference the different information types, unique Digital Author Identifiers (DAIs) are used. Every author or researcher is assigned a personal DAI, which creates the Academic Information Domain [3], the domain where all information relating to research is collected. Thanks to the

¹ www.researchinformation.nl

² Data Archiving and Networked Services (www.dans.knaw.nl)

What are your information needs?

DAI, a personal page can be compiled in NARCIS for every researcher, containing a complete overview of his or her research, publications and datasets in context, as illustrated by the example for Professor W.H.J. Meeus.³

NARCIS already offers users many useful functionalities such as RSS feeds, the Zotero⁴ reference tool and personal pages for researchers. The portal is visited 1,200,000 times a year by researchers, policy makers, people in the media and members of the general public. The proportion of Open Access publications available is rising steadily, as Peter Suber has observed [4]. Thanks to NARCIS, these Open Access publications can be traced quickly and easily.

It is easy to discover how often NARCIS is used from the log data. In keeping with the department's tradition of conducting regular surveys, KNAW-OI conducted a user survey in 2009 to identify who the users are, where they come from and what they use NARCIS for.

2. Methodology

2.1 Analysis of IP addresses

The simplest way of discovering who is using the NARCIS portal is to check the users' IP addresses.⁵ We identified the IP addresses of the 400 most frequent users (in terms of the number of NARCIS views) in January 2010 (through AW-stats⁶).

Those IP addresses were then linked to the names of institutions using IP locators Topwebhosts⁷, Geobytes⁸ and ip2locations.⁹ In Table 1 those institutions are broken down into the following categories: University, Research institution, University of Applied Sciences, Government, Not-for-profit sector, Hospital, Business, Media and Provider.

Table 1: Share of users in each category

Category	Share of NARCIS use
----------	---------------------

³ <http://www.narcis.info/person/RecordID/PRS1237369>

⁴ <http://www.zotero.org>

⁵ For reasons of privacy, no attempt has been made to connect IP addresses to individual users.

⁶ <http://www.ubiquityhosting.com/web-hosting/service/awstats>

⁷ <http://www.topwebhosts.org/>

⁸ <http://www.geobytes.com/ipLocator.htm>

⁹ <http://www.cqcounter.com/whois/domain/ip2locations.com.html>

What are your information needs?

University	36%
Research institution	4%
University of Applied Sciences	8%
Government	6%
Not for profit	2%
Hospital	1%
Business	11%
Media	1%
Provider	32%

The last category, 'provider', is a very special one: many users access NARCIS via a provider. This means that the KNAW can see which providers were used, but naturally cannot identify the individual or organisation that visited NARCIS via those providers.

2.2 Online surveys

In addition to this analysis of users on the basis of IP addresses, we also conducted two online surveys to discover more about the work environment and the professions of the users. The surveys were also designed to learn more about the types of information the users were searching for, how they rated the different functionalities in NARCIS and what developments they would like to see in the system. The two online surveys were compiled by using SurveyMonkey.¹⁰

The first survey was held in June 2009. It could be completed only via the NARCIS website (www.narcis.info), so that only actual portal users were aware of the survey's existence. In view of NARCIS's international character, the survey was presented in both Dutch and English. The participants were asked whether they would also be willing to take part in a follow-up survey. The survey was deliberately kept short and confined to just six questions.

The Dutch-speaking respondents who had said they were willing to participate in a follow-up study were asked to take part in another study, again compiled in SurveyMonkey, in December 2009.

As already mentioned, the two surveys were completed only by actual users of NARCIS. They left two questions unanswered:

¹⁰ www.surveymonkey.com

What are your information needs?

- To what extent does the target group for whom NARCIS may be relevant actually use the service? Can any conclusions be drawn about non-users?
- To what extent could NARCIS be useful for non-users, given their information behaviour?

2.3 Interviews

To answer these questions, semi-structured interviews were held with 17 individuals in the final quarter of 2009. It was not known in advance whether or not they used NARCIS, but it was known that they all held senior positions in which they handled a lot of information.

The point of departure for the interviews was to identify the information ecology [5] of the interviewees, in other words what technologies they use to search for and process information.

These 17 individuals represented NARCIS's various target groups: nine researchers [humanities (3), natural sciences (3) and social sciences (3)], four policy makers, two information specialists and two journalists. This method corresponded with that used in a previous survey of needs in 2002 [6], although that earlier study related exclusively to the Dutch Research Database (NOD).

A script was written in advance to ensure that at least the following subjects were discussed with the interviewees:

- What sources of information do they use?
- How do they search for information?
- What problems do they encounter in searching for information?

Although the interviews related to the respondents' general information needs and information behaviour – in other words, their information landscape – the interviewers focused mainly on NARCIS's potential role in it.

Each interview lasted one hour. The interviews were all recorded and a report summary of each interview was produced according to a fixed format. The reports were all approved by the interviewees.

3. Results

3.1 Surveys

What are your information needs?

There were 434 respondents in the first survey, of whom 268 completed the full questionnaire; 61 respondents completed it in English.

Of the respondents, 59% were from universities, universities of applied science or research institutions. Another 15% were from the business community and 15% were from the not-for-profit or public and semi-public sector (see Figure 1).

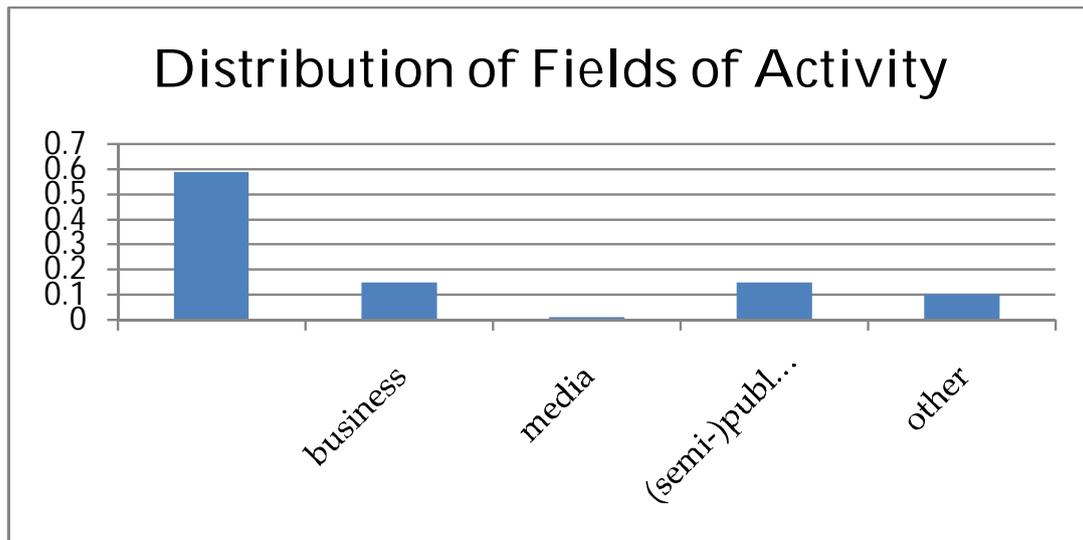


Figure 1: Distribution of fields of activity

As mentioned, it was already possible to gain an impression of the fields of activity of the users of NARCIS by identifying the names of the institutions corresponding with the IP addresses of the 400 most frequent users in January 2010. That analysis showed that at least 48% of the users were from universities, universities of applied science or research institutions, a figure that corresponds closely with findings from the survey, especially bearing in mind that quite a number of those who visit NARCIS via a provider have a position at one of those research institutions.

Thirty-six percent of the respondents described themselves as researchers and 21% as information specialists. Few described themselves as members of the other professions (e.g., policy assistant, journalist). Many respondents answered the question concerning their profession in their own words. Analysis of the information they provided suggests that almost half of those surveyed can be described as academics.

During a six-month period, 21% of the respondents use NARCIS more than 10 times, 18% use NARCIS between four and 10 times and 60% use it between one and three times. Relatively speaking, information specialists use

What are your information needs?

NARCIS most frequently (58% use NARCIS four times or more in a six-month period).

NARCIS users search mainly for dissertations, other types of publication and information about researchers. The number of searches for datasets is remarkably small (7%).

Being able to download full-text publications was mentioned as the most important feature of NARCIS by 78% of the respondents, while 60% regard the links to additional information (for example, from the description of a person to his/her publications) as important. Other important features are being able to search simultaneously in different information types as well as the presentation of an individual's entry in combination with all the relevant information about him or her.

Asked to say what they felt the most interesting development would be for NARCIS, 57% of the respondents mentioned the presentation of similar information from other countries. Other frequently mentioned suggestions for upgrading NARCIS were to make improvements in its functionality (for example, the possibility to browse) and to offer tools for complex analyses.

The follow-up survey was held among a sub-population of the respondents in the first survey, but with a similar composition. The purpose of this survey was to find out how the respondents rated the functionalities and content offered by NARCIS. For 95% of the respondents, having a single gateway to different types of information was felt to be useful or very useful.

The respondents were impressed with the option of searching on (full-text) publications and on current research. However, also this group of respondents – who search in NARCIS more frequently than the wider group of participants in the first survey – does not perform many searches on datasets.

Although the search options are highly rated (79% of respondents were satisfied or very satisfied), the respondents were not always aware of all the search functionalities in NARCIS (combining search terms; tailored RSS feeds). The respondents were most impressed by the large number of Open Access publications available through NARCIS and the overview of experts. They were also pleased with the response time.

3.2 Interviews

The interviews gave an impression of how the interviewees are using information and of what could be the potential importance of NARCIS for

What are your information needs?

them. The demand for the information in NARCIS differed from one interviewee to another, so it is impossible to draw any general conclusions from the interviews. Nevertheless, a certain trend could be discerned in the interviews. The various information types to be found in NARCIS are briefly described below.

Information about individuals, organisations and current research

Information about individuals and organisations is occasionally important for researchers and non-researchers alike. Researchers use this information as background material to help them assess the value of a particular individual or organisation's publications. Non-researchers are often searching for experts in a particular field in order to gather more information.

The interviewees often have their own network and do not need to consult a database, unless they want to explore a new area or need information about less well-known individuals.

Information about current research is important to gain an early impression of work being done new fields of research.

Information sources

The interviewees use a variety of channels to gather information. The sources mentioned include those available via the Digital Library of the respondent's own university, preprints, search engines (Google, Google Scholar), personal contacts and participation at conferences and workshops in the Netherlands and elsewhere. Blogs and Twitter were also mentioned as a source of very up-to-date and opinion-forming information.

Dissertations and datasets are especially important for researchers; dissertations are a particularly valuable source of information outside the natural science sector. The same applies for other types of publication of a monographic nature. While many dissertations are nowadays available in electronic form, this is unfortunately not true of monographs in general. Nevertheless, it is shown that there is a growing demand for digital monographs [7].

In the Science-Technology-Medicine (STM) sector, the interviewees were more interested in journal articles (which also are the main elements of the dissertations in this sector).

Policy makers seek inspiration from the results of research to formulate and roll out new policy, while journalists report on that research. At most, dissertations and other scientific publications are useful to them as background information.

What are your information needs?

Most researchers and non-researchers subscribe to services alerting them to new information. One disadvantage of this method that was mentioned was that it causes information overload.

A noteworthy finding was the importance the interviewees attached to personal networks, including online networks. Some have created their own networks and they often also establish special interest groups on networks such as LinkedIn.¹¹ Scientific information is quickly disseminated in these networks. Trust is important in this context, which is why the digital networks are built on existing networks in real life.

Datasets

Datasets are mainly important for researchers. These datasets may consist of statistics but may in a broader sense also include, for instance, audio and video recordings. There is a certain tension between, on the one hand, the desire to write publications based on one's own raw material first, and sharing and re-analysing this material on the other.

At most, non-researchers need pre-packaged statistical information.

Context

Many researchers refer to the importance of the context of the information they find. This relates to a functionality such as links (for example from the raw data to the related publications), on the one hand, and the presentation of background information (about the author, the organisation or the research programme) relating to the information that has been found, on the other. In this way, the user can assess the merits of a particular source.

Problems in searching for and selecting information

Examples mentioned by the researchers include:

- *Quality*: it is not always easy to distinguish between information of a high quality and less valuable information.
- *Accessibility*: publications are not always available under Open Access.
- *Coverage*: a lot of material that is relevant for research and education is not available.

¹¹ <http://www.linkedin.com>

What are your information needs?

- *Context*: search engines like Google provide no information about the context.
- *Information overload*: search and alert systems are not intelligent enough, which results in a surplus of information or in irrelevant information.
- *Persistence*: researchers and documents are difficult to trace permanently on the Internet. Assigning Digital Author Identifiers and persistent identifiers to documents could solve this problem.

The non-researchers report the following problems:

- *Absence of very concise abstracts* of scientific publications
- *No free access* to some texts
- *Difficulty in finding experts* (who are needed to assess the content of news items)

Suggestions for improving the NARCIS service were made in both the surveys and the interviews. Some of the suggestions may be put into effect immediately, but some call for national or international consultation. Some of the most imaginative suggestions were:

Intelligent search and alert systems; text mining; internationalisation; permanent storage of new information types such as blogs; access to enriched publications [8, 9].

4. Discussion

The surveys have shown that a significant number of NARCIS users come from universities, research institutions or universities of applied science. They are the portal's principal target group. At the same time, it became apparent that NARCIS users are often unaware of the possibilities of the portal. For example, they are not all aware of the possibility of combining terms in a search command and do not all take advantage of the benefits of the customised RSS feeds. It is very important to display these options more clearly in NARCIS.

The interviews revealed that the interviewees first consult Google (Scholar) when searching for information. Only the biomedical specialists among them also use PubMed.

However, these interviewees are also aware of the limitations of the giant Google, the most prominent being information overload and the uncertainty about the quality of the information that is found.

What are your information needs?

The information overload can be eased by introducing the option of personalising the presentation of information in the NARCIS system by giving the greatest prominence to the information types that are most relevant to him or her.

Although search engines, and particularly Google, are popular, the interviewees did say that they would like information to be presented by subject. To present information by subject (for example, on the topic of historical sciences), a service depends on the metadata that is supplied.

Past experiments by the KNAW with tools for automatic categorisation suggest that it does not lead to acceptable results in a multidisciplinary database. Thematic presentation might be possible with Web 2.0 facilities (along the lines of Flickr), with users applying their own tags to information objects.

The NARCIS information is already highly accessible in Google. Google often shows users information from NARCIS without their realising it. Google can therefore be regarded as a supplementary source of access to NARCIS. The benefit for the user of searching directly in NARCIS is the availability of additional functionalities that Google does not offer. A public relations effort is needed to inform users of these extra options in NARCIS.

This user survey was confined to the reaction of human users. However, a service like NARCIS is also for non-human users. For example, NARCIS provides crucial information to services that operate at European and global level (DRIVER ¹² ; Scientific Commons ¹³ ; Google (Scholar); WorldWideScience.org¹⁴). It is clear, for example, that the co-ordination between NARCIS and Google works well from the fact that the website www.narcis.info has the high page ranking of '8'¹⁵, a scale that is awarded only to one percent of the websites displayed by Google.

According to the interviews, NARCIS is not adequately promoted or publicised. None of the senior officials interviewed use the service, which is not surprising since until 2010 there had never been a publicity campaign for NARCIS. The launch of a new version of NARCIS in March 2010 is now being used to bring NARCIS to the attention of a wider public.

The campaign will stress the key role that NARCIS plays in the Dutch national information landscape. By using Digital Author Identifiers (DAIs) and showing relationships between types of information, NARCIS is the leading site for searching for and finding scientific information in context.

¹² <http://search.driver.research-infrastructures.eu/>

¹³ <http://www.scientificcommons.org/>

¹⁴ <http://www.Worldwidescience.org/>

¹⁵ <http://www.thegooglepagerank.com>

What are your information needs?

However, broader applications of identifiers are possible, particularly identifiers for persons, who may act as a researcher, as an author or even as the subject of a study. A number of interviewees suggested enriching the identifiers with a definition of the various roles as a way of improving the system.

The NARCIS concept is unique in the Netherlands, and even in Europe. There are no other services that provide a combination of scientific information (publications and datasets) and science information (information about researchers, research, research institutions).

The integrated supply of so many types of information automatically creates a desire for more complex text-mining tools, which can display clusters of researchers or publications. Naturally, that implies that users must be able to visualise the results of these analyses.

5. Conclusions

The user survey in 2009 proved very useful. The surveys gave an impression of the backgrounds of the NARCIS users, of the NARCIS functionalities they appreciated and of possible improvements. The interviews provided an understanding of the information needs of persons who use information a lot but are not yet familiar with NARCIS.

The surveys reveal that more than half of the NARCIS users come from universities, scientific institutions or universities of applied science. Most searches in NARCIS are for dissertations.

The most valuable functionalities are the option of downloading publications, the links from individuals to publications and the ability to search simultaneously for different information types. The existence of a single gateway to different types of information is regarded as very useful.

The possible upgrade that was most frequently mentioned is access to information from other countries. Providing tools to perform complex analyses on the material accessed via NARCIS was also mentioned as a worthwhile option. The conclusion to be drawn from these responses is that there is room for a service like NARCIS alongside a 'one size fits all' search engine like Google. NARCIS can already largely meet the wishes of the interviewees in its current form. With just a few minor modifications – such as the introduction of the possibility of browsing through the information – NARCIS will also be far better equipped to their needs.

What are your information needs?

The value of NARCIS can be further increased by intensifying the publicity about the portal and by continuing to improve the service.

Acknowledgements

KNAW-OI is very grateful to Repke de Vries and Henk Voorbij for their assistance in preparing and carrying out the user survey and processing the results.

Notes and References

- [1] WAAIJERS, L. The DARE Chronicle: Open Access to Research Results and Teaching Materials in the Netherlands. *Ariadne* 53, October 2007. Available at: <http://www.ariadne.ac.uk/issue53/waijers/> (March 2010).
- [2] DIJK, E; BAARS, CHR.; HOGENAAR, A.; MEEL, M. VAN. NARCIS: The Gateway to Dutch Scientific Information. Paper presented at ELPUB Conference 2006, Bansko, Bulgaria. Available at: http://elpub.scix.net/cgi-bin/works/Show?233_elpub2006 (March 2010).
- [3] BAARS, CHR; DIJK, E; HOGENAAR,A; Meel, M. VAN. Creating an Academic Information Domain: a Dutch example. Paper presented at EuroCRIS 2008, Maribor, Slovenia, 2008. pp. 77-78. Available at: http://www.eurocris.org/fileadmin/Upload/Events/Conferences/CRIS2008/Papers/cris2008_Baars_Dijk.pdf (March 2010).
- [4] SUBER, P. Open Access in 2009. SPARC Open Access Newsletter, issue #141, 2010. Available at: <http://www.earlham.edu/~peters/fos/newsletter/01-02-10.htm#2009> (March 2010).
- [5] NARDI, B.O.; O'DAY,V.L. Information Ecology : Using Technology with Heart. *First Monday*, 4(5), May 1999. Available at: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/672/582> (March 2010).
- [6] KOOPMANS, N.I. 'What is your question? The need for research information from the perspective of different user groups' in W. Adamczak and A. Nase (eds). *Gaining Insight from Research Information: Proceedings of the 6th International Conference on Current Research Information Systems, University of Kassel, August 29-31, 2002* (Kassel) pp.

What are your information needs?

- 183-192, 2002. Available at: <http://www.upress.uni-kassel.de/online/frei/978-3-933146-84-7.volltext.frei.pdf>
- [7] ADEMA, J; RUTTEN, P. Digital Monographs in the Humanities and Social Sciences: Report on User Needs. OAPEN report 2010. Available at: <http://www.oapen.org/images/D315%20User%20Needs%20Report.pdf>.
- [8] Enriched publications consist of a traditional publication, linked to information that is used in the writing of the publication (a dataset, for example) and to related information that has appeared since the publication (citations; publications based on it; commentary).
- [9] HOGENAAR, A. Enhancing Scientific Communication through Aggregated Publications Environments. *Ariadne* 61, October 2009. Available at: <http://www.ariadne.ac.uk/issue61/hogenaar/> (March 2010).