Next Generation e-Researchers: 
Doctoral Students in Social Sciences and Humanities in Sweden and their Attitudes towards Open Access and Open Repositories

Ann-Sofie Axelsson\textsuperscript{1}, Carina Carlhed\textsuperscript{2}
\textsuperscript{1} Chalmers University of Technology, Sweden
\textsuperscript{2} Mälardalen University, Sweden and Swedish National Data Service
Email address of corresponding author: annaxe@chalmers.se

Abstract. In Swedish research politics there are, at the moment, several actions taken to enhance existing and develop new research infrastructures (Axelsson and Schroeder 2007). As one example, the Swedish National Data Service (SND, \texttt{http://www.snd.gu.se}), an operative key actor is currently working on coordinating existing data resources and increasing deposition of research data to the national data service and re-use of those data. SND has, as a part of this endeavour, recently conducted two survey studies, targeted at professors (N=549) and doctoral students within social sciences and humanities departments (N=1147) at Swedish universities and university colleges, in order to obtain knowledge regarding existing use and re-use of digital research data and archiving practices as well as obstacles to increased digital research data sharing. This paper focuses on the doctoral students’ data and the results are compared with the results from the parallel study of the professors and from a recent survey targeted at professors in various social sciences and humanities disciplines at Finnish universities (Kuula and Borg, 2008).

The results from the current study show that doctoral students in general expressed a great uncertainty about questions of amounts of reusable digital data and effective interventions to enhance accessibility to digital research data. They identifies research ethical aspects as important barriers to sharing digital research data, while professors emphasize lack of resources for researchers to document and make their data accessible for others as the most important obstacle. Concerning interventions to enhancing re-use of digital data, the majority of the doctoral students and the professors thought it should be effective to get more information about accessible research data in data archives or databases. Nearly 100 \% in both groups reported that also more of training in research methods, digital research databases and information about accessible e-tools would be effective interventions. The most effective interventions for enhancing accessibility to digital data were reported in terms of that research grants should include funds for preparing the data for sharing and archiving and that archiving data for the use by the scientific community is acknowledged to be scientific merit. Surprisingly, when it comes to the degree of urgency in sharing their own data, the professors seem to be a bit more eager to share data than the doctoral students.
Introduction

This current paper deals with only a few aspects of this complex issue, partly due to the limited space available, but mostly due to the fact that every single aspect of the larger question concerning research data availability needs to be examined and discussed very carefully and the pieces put together to be able to provide the bigger picture.

This paper is based on two recently conducted survey studies, targeted at professors and doctoral students within humanities and social sciences at Swedish universities and university colleges, with the broader aim of investigating existing practices and attitudes when it comes to availability and re-use of research data. The questions dealt with in this paper are the following: How do doctoral students within social sciences and humanities departments at Swedish universities and university colleges use, archive and re-use research data, and what obstacles to increased digital research data sharing do they perceive? This group of interest, doctoral students, is compared with the senior group of professors, targeted with the same questionnaire, as a comparison.

The two survey studies were carried out by the Swedish National Data Service (SND, http://www.snd.gu.se), an operative key actor in the Swedish enterprise towards an enhancement of existing and development of new research infrastructures (Axelsson and Schroeder 2007). The SND’s mission, from the Swedish Research Council (VR, http://www.vr.se/), is to collect research data from researchers within the humanities and social sciences, and provide access to these data to the Swedish research community. The reason for the specific interest in the senior researchers, the professors, and in the junior, or future, generation of researchers, the doctoral students, is that these to stakeholder groups could or should have both experience from and interest in research infrastructures in general and data repositories in particular.

The reason for a specific focus in this paper on doctoral students is that it is particularly important to understand how this young generation of researchers carries out research today and how they may want to conduct research in the future. Also, the considerations they may have concerning depositing own data and re-using other researchers’ data are important to be aware of and crucial to address when creating and maintaining a digital data repository, or else the repository will not be able to convince its potential users, and the whole idea of sharing will fail. If a national data archive shall succeed, the least it should be able to do is to convince its two most important stakeholders, the researchers that deposit data, and the researchers that use those data. One important step, therefore, towards a data repository and a data sharing service that are widely trusted and used, is to get to know the users and make their knowledge, experiences and doubts parts of the system.

Previous research

Recent national and international research data sharing initiatives have also called upon research on the varies issues related to data sharing within and between disciplines in order to tackle practical problems and theoretical issues that arise when research data, the hard currency of research, are made available to researchers on a broad scale.

Previous research of relevance to the questions asked in this study have been carried out within disciplines such as for example library and information science, social informatics, and sociology, each discipline dealing with specific parts of the larger question of how to make researchers collaborate distributed, on a global scale, using technology. Some research has mainly focused on technological issues (e.g. Crabtree et al. 2006), and other on the social
(e.g. Sonnenwald 2006), legal (e.g. Burk 2007) or ethical (e.g. Dutton, Carusi, and Peltu 2006) issues that are involved.

Previous research of most interest for this current study relates to the huge amount of research data collected today, the so called ‘data deluge’ (Hey and Trefethen 2003), one of the most central prerequisites for the national research infrastructures and for e-Research initiatives taken in recent years. The many problems related to the collection, storage and re-use of research data have been studied and described by for example Carlson and Anderson (2007) and Borgman (2007). Both mention for example the problem with assuring a high quality on research data that have been disconnected from the researchers who collected them (Carlson and Anderson 2007).

The different views on and attitudes towards data and data sharing within different disciplines, something of relevance for this study, has been described by Borgman (2005), who emphasizes, among other things, that data have different values within different disciplines, and also to other stakeholders, where for example medical research data and social sciences data have a higher value than research data within humanities disciplines.

Studies on attitudes among professionals towards knowledge and information sharing between organizations, which is in many senses closely related to data sharing, has been studied during recent years (e.g. Axelsson, Spante and Sonnenwald, 2009; Sonnenwald 2007) showing that there are a number of reasons for people wanting to share information and knowledge and for not wanting to. One somewhat surprising finding in those studies is that many professionals have not only professional but personal or social reasons for wanting to collaborate and share knowledge and also for not wanting to do the same. These findings are immensely important to consider when developing a service of the kind discussed in this paper, which, more or less, replaces, or at least complements, the personal and social information sharing between professionals.

Open Access, Open Repositories and e-Science in Sweden

The Swedish research infrastructure initiative, which has been described in more detail elsewhere (Axelsson and Schroeder forthcoming) takes on the challenge of developing the infrastructure on the basis of an existing national grid infrastructure, a well-functioning university computer network, and a well-established but under-used set of databases and registers. In order to increase the use of already existing research data and to expand the research data base with new research data, a technical and social infrastructure consisting of a national data service – an open repository for open access to data – was initiated by the Swedish Research Council in 2005 (The Swedish Research Council 2007). The two surveys reported on in this paper, is an important part of this work. However, the results from these two studies needs to be contrasted with and understood against Axelsson and Schroeder’s study of the Swedish e-research enterprise (forthcoming), since the main findings in this study have implications for how to interpret and make use of the results from the current one. First, the small size of the country and the dense and personal research networks suggests that a data service may have difficulties establishing itself as a third party, a broker of research data. Second, the well-established trust between the Swedish population and data researchers needs to be taken care of. Third, the Swedish legislation regarding data handling, seems to be a crucial issue to deal with, maybe the most important one. If the Swedish law does not allow for data sharing in the sense as it is planned, which is the case at the moment, it matters very little if researchers are positive towards the idea.
Method

The two surveys, one directed to Swedish professors in humanities and social sciences and the other directed to Swedish doctoral students in the same domains of disciplines, contained 80 items covering the researchers affiliations, domain of discipline, gender, age, familiarities with research policies and ventures, and use, re-use, archiving practices of digital research data. Furthermore, there were questions about possible reasons for not using digital data, interventions and barriers to enhanced re-use and accessibility to data, possible agents in overcoming barriers and willingness to engage in promote alternations in this area and to share their digital research data. The surveys were carried out through email-questionnaires and with lists of respondents based on retrievals from databases at the universities’ offices for IT or Personnel Administration. In some of these lists it was easy to recognize respondents’ disciplines; others were sorted by thematic or interdisciplinary departments and/or no information about discipline were accessible. Departments, which were within Science and Technology, Educational Sciences, Social medicine were included, but only departments which described themselves as inter-disciplinary on their websites. However, most of the departments were within Humanities and Social Sciences. Because the population was broad and had somewhat non-distinct boundaries, we asked respondents to reply to us, if they did not use perspectives of social science or humanities in their research. In that case they were cancelled from the survey. Initially, there were 1589 professors from 35 universities/university colleges, and after the cancelling procedure of non-social science or non-humanities researchers (by the definition above) there were 1436 professors. The response rate was 38 %, with 549 responses. The same procedure was carried out with the population of doctoral students. However, the lists from the universities which formed the respondent list had minor inaccuracy problems, due to some “natural” conditions, namely doctoral students becoming doctors. This affected the update status on information in the university personnel information systems, which had in some cases inaccurate information about the doctoral students. In addition, doctoral students at university colleges could also appear at a list from another university, hence with double mail addresses. A check up was made before the distribution of the email questionnaire in order to avoid obvious doubles, however in some cases the email addresses were abbreviated and impossible to relate to the names of the doctoral students. Like the professor survey, the population were broadly defined, which called for a similar procedure for cancelling, by respondents reply stating their non-social sciences or non-humanities affiliation. Initially, the doctoral student population included 4697 respondents and after the cancelling procedure (mentioned above), 4065 remained. The response rate was 28 %, with 1147 responses. When comparing how the professors’ response rate patterns related to the distribution among a selection of the universities which received the largest proportion of questionnaires, we can conclude that the response rate from the larger respondent groups’ universities alternated between 22 to 44 %.

Because our method of selection was somewhat unstable, we found it necessary to investigate our precision further. The Swedish National Agency for Higher Education produces statistics about the universities and university colleges.1 By comparing statistics of professors and doctoral students and their affiliation to university and disciplinary domain from 2008 and our response patterns gives a view of how our survey succeeded in target the population. It seems that the population of professors (constructed from statistics, i.e. number of professors in different domains of disciplines and university), is well covered by our group

---

1 http://www.hsv.se/statistik/statistikomhogskolan/personal.4.6df71dcd1157e4305158001770.html.

In this paper, comparisons have consistently been made between statistics from The Swedish National Agency for Higher Education for year 2008 and the background information about the participants in our surveys.
of professors, who have participated in the survey. In concordance with this one can argue that our response frequencies are higher in reality, when comparing it with the statistic population, which we constructed for comparison reasons. For some universities however, the response rate was lower in this comparison. It signals distortion in our precision about the doctoral students. In conclusion, our generalization opportunities are limited due to these aspects that have been discussed above. It seems that the ground for conclusion about the group of professors is more stable than the group of doctoral students. Nevertheless, a large number of professors (N=549) and doctoral students (N=1147) have participated in our studies, which implies considerable opportunities to valid conclusions.

In the professors’ group, there were a majority of men who had answered the questionnaire, 73 % compared to 27 % women. This reflects however the demographics of the larger population, whereas 23 % of the professors in social sciences, humanities (and law) are women. In the group of doctoral students the conditions were opposite, 61 % of the doctoral students in our survey were women. In comparing with the statistics from The Swedish National Agency for Higher Education, the larger population consisted of 56 % women. In both cases we can conclude that women were slightly a bit more inclined to participate in our surveys than men. Considering age, with our survey we seem to engage a larger part (25 %) of the younger doctoral students (younger than 29 years old), than expected (16 %). The same counts for the group of professors, but there were only a minor difference. Two percents more of professors participated that were younger than 50 years old (18 %), compared to statistics from The Swedish National Agency for Higher Education (16 %). According to domains of disciplines, it seems that our groups of professors and doctoral students reflect the structure of the larger population. Based on the discussion above, our conclusion is that the results from our surveys could be treated as fairly valid, in spite of the relatively low response rate. The amounts of responses from professors and doctoral in different domains of disciplines, age and gender is corresponding to the official statistics that has been described and discussed.

Results

The Swedish Research Council has, in a current venture made a long-term strategic plan - a roadmap [The Swedish Research Council’s Guide to Infrastructure (2007)]. In the questionnaire we asked the researchers about their knowledge about this venture and their opinions about it. There were 11 % of the professors, which were familiar with the venture and the guide and only 1 % of the doctoral students. Half of the professors group did know about the venture but not its details and 40 % did not have knowledge about it at all. This was also true for the majority of the doctoral students (82 %). Professors were more inclined to express positive opinions about the venture and the doctoral students followed the same pattern. The knowledge about [OECD Guidelines on Open Access to Research Data from Public Funding (2007)] was generally low, 75 % of the researchers (both groups) did not know about it at all. Surprisingly, 61 % of the professors were not aware of its existence. Breaking down results to different domains of disciplines, it seems that professors within Social Sciences are the most informed about the OECD guidelines, and the group which were at least informed was the doctoral students within Law. Considering the situation of being a doctoral student, we are not surprised of the large amount of them not having knowledge about the guidelines and/or the research venture mentioned above.

Archiving practices and re-use of digital research data

The primary condition of archiving and reusing digital research data is that data are collected and compiled in some way. 73 % of the professors stated that digital empirical data are used
in research and 16% stated that the use of digital empirical data is unusual or are never used. The major part that did not use empirical digital data was the professors within Humanities (56%). Among the doctoral students, they declared that digital empirical data was used (42%), but they expressed a great uncertainty about these questions generally.

According to the professors, the digital data are often kept at the researcher after analysis and reporting, without any actions to documentation (46%) but it also occurs from time to time (according to 15%). Archiving practices where digital data always are kept and documented in a catalogue/database at the university is quite rare (11%). Almost half of the professors’ group stated that these practices were unusual or never occurred. The same tendency showed concerning facilitating availability of digital research data at a data archive, that is, unusual practices. However, it seems that the research data are not regularly destroyed after analysis and reporting, at least 49% stated that destruction is uncommon and only 3% reported that it was common. The re-use of digital data are relatively common, 59% of the professors stated that data are re-used in PhD works or other research projects and only 3% reported that it never happened. The use of re-used digital data in teaching is also quite common according to 59% of the professors. Reusing all kinds of empirical data is most common in situations when researchers use the data themselves and approximately one third stated that they passes data on to other researchers, who are studying similar kinds of areas. 5% of the professors reported that this never occurred. About the amount of the digital data that are reusable, professors are more optimistic in general than the doctoral students, that seemed very uncertain and had difficulties to express opinions of estimates. There were small differences between both groups and the domains of disciplines in these issues. Important reasons for not reusing digital data are mentioned by the professors as uncertainty about the quality of data (62%), ethical aspects (57%), technical issues (53%) and juridical issues (49%). However the professors’ group is divided in opinions and the other part do not think that these factors are crucial (38% - 50%). According to those who think ethical aspects are crucial, we found that these professors were mainly from Social Sciences. That is true also for doctoral students in the same domain of discipline. The importance of juridical aspects is represented by the doctoral students in Law, but not the Law professors. Both professors and doctoral students in Humanities deviated in general from the others in these issues, i.e. the technical issues were considered important. They also report other reasons for not reusing digital data, such as not using empirical or/digital data at all, lack of knowledge and routines, decontextualized data have weak relevance for others etc.

Concerning interventions to enhancing re-use of digital data, 95% of the doctoral students and 93% of the professors thought it should be effective to get more information about accessible research data in data archives or databases. Nearly 100% in both groups reported that also more of training in research methods, digital research databases and information about accessible e-tools would be effective interventions (89-95%). It seems that professors and doctoral students in humanities are most positive towards more of education interventions and researchers in Social Sciences are the least positive, but all groups are generally positive to the interventions proposed. In addition, the most effective interventions for enhancing accessibility to digital data were reported in terms of that research grants should include funds for preparing the data for sharing and archiving (88% of the doctoral students and 83% of the professors) and that archiving data for the use by the scientific community is acknowledged to be scientific merit (87% of the doctoral student and 83% of the professors).

**Obstacles to sharing digital data**

Our seven suggested obstacles to sharing digital data have been ordered in precedence by the respondents. The professors regard deficiency of resources for researchers to document and arrange their data to reusable conditions, as the most difficult obstacle to sharing digital data. They also ranked lack of other resources like guidelines and directions for documentation as
important issue. Another obstacle highly ranked by the professors was doubtfulness about a correct use of their data, i.e. risks of mistakes and misuses. An additional impediment was the fact that their respondents were not informed that their contributions should be used in the research society in general, only for a particular study. Juridical obstacles and loss of one’s own advantage of competition in keeping data to oneself were not considered as crucial. The least important obstacle, according to the professors, was ethical aspects such as threats to confidentiality and delicate information. The doctoral students however, thought that ethical aspects mentioned above were the most difficult obstacles of all. After that, they considered the information to the respondents and the use of their contributions to the research society in general was an important issue. Deficiency of resources for researchers to document and arrange their data to reusable conditions, were also ranked as important, followed by juridical aspects. The least important obstacles according to the doctoral students were lack of other resources like guidelines and directions for documentation, loss of one’s own advantage of competition in keeping data to oneself and doubtfulness about a correct use of their data, i.e. risks of mistakes and misuses of data. The response pattern did not change depending on the researchers’ use of digital data or not. On the other hand, researchers in Social Sciences and women were more concerned with research ethical aspects and threats to confidentiality etc. while researchers in Humanities and men tend to emphasize lack of resources to document and arrange their data to reusable conditions. According to age, older researchers tend to emphasize lack of resources and juridical issues. Younger researchers pointed out ethical aspects, threats to confidentiality and doubts of incorrect use of their data. 65 % of the researchers meant that these obstacles prevent them from sharing data to The Swedish National Data Service. There were minor differences according to age, where the older researchers were more optimistic and the researchers in Humanities as well. There were very small differences between professors and doctoral students. When we asked if the researchers would consider to engaging in promoting alterations in these areas, the doctoral students tend to embrace issues of research ethics and changing values in accessibility and practices in sharing data, while professors were inclined to issues of jurisdiction. Surprisingly, when it comes to the degree of urgency in sharing their own data, the professors are a bit more eager to share data (30 %) than the doctoral students (24 %). In total, there were 53 % of the researchers that thought it was urgent to share data (55 % of the professors and 52 % of the doctoral students), but only 26 % in the total group reported that they intended to share their data. A large proportion of the total group expressed doubtfulness in sharing data (40 %). Researchers in Law were the least keen on doing it and thought it was not so urgent. Researchers in Humanities however, were those who distinguished themselves as potential “ sharers”. According to gender and age, there were the men and the older researchers who expressed more willingness to share than others.

Discussion

The knowledge about OECD Guidelines on Open Access to Research Data from Public Funding was generally low, but professors within Social Sciences were the most informed and the least informed, was the doctoral students within Law. In comparison with the Finnish survey, which was carried out by The Finnish Social Science Data Archive in 2006 (Kuula & Borg 2008), where 81 % of the professors did not know about the OECD recommendations, compared to 61 % of the Swedish professors. 81 % of the doctoral students were not aware of its existence. Considering archiving practices, use and re-use of digital research data, 16 % of the Swedish professors stated that the use of digital empirical data is unusual or are never used. 18 % of the Finnish professors reported a similar amount of digital data non-use (ibid.) When comparing between the countries what happens to digital data after analysis and reporting, it seems that it was more common for Finnish professors to keep digital data
without any further actions to documentation (56 %) compared to Swedish professors (46 %). Data are destroyed to a larger extent in Finland (20 %) than in Sweden (3 %). However, the saved data is re-used by the researchers themselves in a greater extent in Finland (94 %) than in Sweden (54 %). The opinions of amounts of reusable digital data differs also, 50 % of the Swedish professors stated that more than half the amount of produced digital data is reusable, compared to 21 % of the Finnish professors. In analyzing responses to important reasons for not reusing digital data it appears that the Swedish researchers emphasize ethical, juridical, technical aspects and issues quality of data as more problematic than the Finnish researchers. Concerning interventions to enhancing accessibility of digital data, the researchers believe that the most effective interventions would be that archiving data for the use by the scientific community is acknowledged to be scientific merit and that research grants should include funds for preparing the data for sharing and archiving, like the stick and the carrot with both force and enticement. The last mentioned intervention were also one of the Finnish professors high ranked intervention (80 %), but their top priority of effective interventions was establishment of guidelines and principles by the Finnish universities together (84 %). The problems with assuring a high quality on research data is acknowledged by Carlson and Anderson (2007) and Borgman (2007) and argue that the quality assurance processes of research data have been disconnected from the researchers who collected them. This is in line with what the professors also points out when they state their opinions of obstacles to sharing digital data, the Swedish professors regarded deficiency of resources for researchers to document and arrange their data to reusable conditions, as the most difficult obstacle to sharing digital data together with lacking guidelines to documentation, while the Finnish professors reported that it was the situation when the respondents were not informed that their contributions should be used in the research society generally. They share this concern with the Swedish doctoral students. As we mentioned in the Result section we found that the degree of urgency in sharing their own data, the professors seems to be more eager to share data than the doctoral students. A large proportion of the total group was also expressing doubts in sharing data. Researchers in Humanities however, were those who distinguished themselves as “sharers”. The Finnish questionnaire did not have a pushing question like we had, but the Finnish professors were asked of their attitude to open access to digital research data collected in their own research and 76 % of them expressed positive attitudes. One might conclude that professors in Social Sciences and Humanities in Sweden and the Finnish professors differs a lot in opinions about digital research data. However, two years have passed with increasing focus on Open Access issues in research policies in these countries. It would be interesting to see if the Finnish professors have changed their mind since 2006. About our own results, it is always interesting when the research is surprising. We were surprised that the professors were the ones who seemed to be more positive and humble towards sharing and promoting accessibility to digital research data, than the doctoral students. But on the other hand, being a doctoral student means a lot in keeping on one own’s track, concentrating on the Ph D work and have little time to orientate among ventures, research policies and university practices. In conclusion and in spite of many prejudices about “conservative” professors, it seems that one have to acknowledge their positive orientation about e-science and put forward these survey results of barriers and opinions to be able to supporting and realizing sharing of digital research data in the future.

Acknowledgments

The authors wish to thank the doctoral students and professors who kindly answered the questionnaires and the Swedish National Data Service for carrying out these surveys and permission to use data.
References


Högskoleverket. (Swedish National Agency for Higher Education). http://www.hsv.se/statistik/statistikomhogskolan/personal.4.6df71dcd1157e43051580001770.html


