

---

# Making it Open and Keeping it Safe

## *e-Enabled Data-Sharing in Sweden*

Ann-Sofie Axelsson

*Department of Technology Management and Economics, Chalmers University of Technology,  
Göteborg, Sweden*

Ralph Schroeder

*Oxford Internet Institute, University of Oxford, Oxford, United Kingdom*

**abstract:** This article focuses on data-sharing – a central aspect of an ambitious e-Science programme recently embarked on in Sweden. Data-sharing has become a powerful and promising direction in e-Science in general, even though fraught with difficulties. Sweden has a unique position in relation to data-sharing: a world-unique set of social science and medical data collections, a well-established tradition of regulations concerning data protection, a widely used form of personal identification that allows integration of databases, and a population that generally trusts researchers and the Swedish state with personal data. The aim of this study was to find out how Swedish database owners/managers and database users – key actors in the Swedish e-Science enterprise – anticipate the way that databases will be built up, managed and used in the future, and how this will influence e-Science. For this purpose, these actors were interviewed and official documents on the topic were studied. It is concluded that openness and the integrity of personal data are particularly critical elements for the success of a range of future e-Science endeavours in Sweden and elsewhere.

**keywords:** data-sharing ♦ e-Infrastructure ♦ e-Science ♦ research ethics ♦ research policy ♦ Sweden

---

## Introduction

Data lie at the core of science. In all times, scientists of all disciplines have collected, analysed and drawn conclusions about the world around us and about ourselves on the basis of data. In the early days of science, however, the systematic collection and analysis of data was cumbersome for two practical reasons. These processes were not automated in any sense, and researchers and their assistants had to gather, enter and analyse the data manually. In the twenty-first century, science has become, in many ways, a highly automated process, and involves processes such as collection and analysis of data and publishing and dissemination of research results. New technologies, such as digital audio and video recorders, the mobile telephone and the Internet, have automated data-collection; statistical software programs and software for qualitative analysis have automated data analysis; and the Internet has made research communication and diffusion of research results fast and automated activities. For

the most part, however, these new technologies have been fairly small scale and used only locally.

In recent years, though, development has taken technologies to another level, with large-scale storage capacities, high bandwidth computer networks, the processing power of supercomputers and the capability to share tools and resources on a global scale. This change from small scale to large scale and local to global is having major implications for research in changing both the pace and the nature of the research process. Not only can research activities such as data-collection, data-analysis, data-visualization and publication of research results be carried out much more quickly, thus allowing more scientific studies, they can be carried out in new and innovative ways. Advanced information and communication technologies (ICTs) and grid technologies, to mention just a couple of examples, have made it possible for researchers to collect, store and analyse vast amounts of data, e.g. gene sequences (Gibbons et al., 2007); to analyse different formats of data simultaneously (e.g. video, audio and text) (Crabtree et al., 2006); to make data (and research results) available to other researchers (see, for example, [www.hapmap.org](http://www.hapmap.org)); and to make research findings transparent for experts, decision-makers and the general public (Heldal, 2007).

Data-collection about individuals is a particularly sensitive issue and constitutes a special case of scientific data. Thus the collection, storage, use and sharing of data about individuals have for some time been governed by laws and regulations to varying degrees in different countries. The laws and regulations are there in order to protect information that can be linked to individuals from being seen and used by others and by authorities and commercial organizations in ways that could be perceived by the individual as intrusive or damaging.

The integrity of personal data has been an issue in all modern societies, and especially since public authorities began to keep registers of their citizens. As the possibilities for collecting information about people have grown with ever more sophisticated technologies (for example with closed-circuit television cameras (CCTVs), transaction logs and computerized medical registers etc.), the debate concerning the integrity of personal data has intensified. However, as described later in this article, these debates seem to rise to new levels when authorities, researchers and in rare cases individuals have broken the law or violated prevailing norms.

As discussed in more detail later, in Sweden there are a number of well-established data collections and registers in addition to a system of national identifier numbers linking individuals to information – pieces of information that on the one hand constitute the very foundation of e-enabled and improved research, and, on the other, a potential risk of crimes or threats to integrity. Large databases, grid technology and e-Science tools add further to the complexity of the situation, since these amplify existing problems with the scale, scope and nature of data about individuals. There will be an increased amount of data gathered as well as new types of data, an increasing number of data collectors, and, not least problematic, an increasing number of users of these data and enhanced access to the data and tools for analysing them.

The aim of this study was to find out how key actors in the Swedish e-Science enterprise – mainly database owners/managers and database users – perceive (1) the potentials and risks associated with increased accessibility to research data, (2) the obstacles that may prevent increased accessibility to research data, and (3) how the unique Swedish conditions in terms of well-established data registers and a high level of trust among citizens towards authorities and researchers should best be utilized and maintained. In more general terms, we were interested in how these central actors anticipate how databases would be built up, managed and used in the future, and how this would influence e-Science.

We discuss these findings against the backdrop of some well-publicized historical incidents in Sweden involving the use and misuse of research data, incidents that can be drawn upon when increased data accessibility is being planned in the future.

Finally, we discuss whether other countries can derive lessons from Sweden, countries that are developing their national research infrastructures for data-related practices.

## Background and previous research

### *e-Science in Sweden and elsewhere*

Research carried out with the help of advanced computing and grid technologies is usually denoted as cyber-infrastructure, e-Science or e-Research. Whatever term is used depends on one's understanding of the concept, which differs between nations and disciplines. Jankowski (2007: 552) points out that all the concepts share the idea that technology, when combined with science, creates added value. In this article the term 'e-Science' is used, since this happens to be the established term in Sweden (see, e.g., <http://www.vr.se/mainmenu/pressandnews/newsarchive/news/howcansciencebeimprovedbyscience.5.315d803911100fdce718000382.html>).

As already mentioned, the use of computers for research purposes is nothing new. However, in recent years, technologies have developed that have transformed computer-supported research into e-Science. In making the most of these new technologies in supporting research, there have been a number of e-Science programmes initiated in various national and supra-national (EU) settings, especially in recent years (starting around 2000). The earliest initiative, which can nowadays be seen as a forerunner, is the United States' 'cyber-infrastructure' initiative, which is developing e-Science as yet another 'big science' challenge (Galison and Hevly, 1992). The UK e-Science initiative has a different structure, one in which a series of specific research programmes has been initiated. And, in Europe, there are various programmes for developing 'e-Infrastructures' (for an overview of the different initiatives, see Schroeder and Fry, 2007). As might be expected, the specific agendas of the e-Science initiatives reflect the heritage of their respective systems of innovation (Schroeder and Fry, 2007; Schroeder, 2008). Nevertheless, these initiatives can all be regarded as attempts to create infrastructures for e-Research, with all the historical parallels to the creation of other societal infrastructures that are implied in this (Jackson et al., 2007; Schroeder, 2007). For the purpose of this article, we define e-Science or e-Research as the use of networked computing for sharing distributed digital data and tools in scientific or academic research.

The Swedish e-Science initiative, which is the focus of this article, takes a UK-like approach, since it confronts the e-Science challenge from several angles closely related to previous socio-technical innovations, namely a national grid infrastructure, a university computer network and a well-established but under-used set of databases and registers. Briefly, the Swedish e-Science initiative can be described as follows: In 2005, the Swedish Research Council (Vetenskapsrådet, 2009) established the Committee for Research Infrastructures (Kommittén för forskningens infrastrukturer, KFI) with the remit to support the building and use of infrastructure for Swedish research. A year later, in 2006, a new committee was established by the Swedish Research Council, i.e. the Database Infrastructure Committee (DISC). At the time there were two already existing parts to the research infrastructure: the Swedish National Infrastructure for Computing (SNIC) and the Swedish University Computer Network (SUNET). Together with these two already existing parts, and according to the Swedish Research Council, it was assumed DISC would 'comprise the trinity which constitute[d] the base in the work of the Swedish Research Council with an e-science infrastructure for Swedish Research' (Vetenskapsrådet, 2009). With the establishing of DISC there is now a strong emphasis on databases and data-sharing within the Swedish e-Science enterprise, since DISC's general mission is to 'create an advanced coordination of existing and new quality-assured research databases and provide this national resource to Swedish and international research' (DISC, 2009). It can be mentioned in passing that data in this case fall mainly within the province of the social and medical sciences; the former mainly the province of the US cyber-infrastructure for social science

(Berman and Brady, 2005) and in the UK under the National Centre for e-Social Science (NCeSS, <http://www.ncess.org/>), the latter mainly under the NIH (National Institutes of Health) in the US and in the UK various e-Science and Medical Research Council initiatives.

Just recently, in December 2008, the Swedish Research Council announced that it would finance six database research projects and one postgraduate school, and inject a sum of 135 million Swedish kronor during a five-year period within the 'Swedish Initiative for Micro-data in the Social and Medical Sciences' (SIMSAM). The aim of SIMSAM is to fund a small number of database research projects per year, and that these projects, led by junior researchers, would be joined together in a network with joint workshops, conferences and a postgraduate school. The vision for this part of the enterprise, the Swedish Research Council says, is to create a new generation of researchers with a knowledge of Swedish databases and registers, knowledge of database methodology, and that has start-up funding for new and interdisciplinary database research (Vetenskapsrådet, 2009). The research will focus on public health and social issues. In the autumn of 2008, the Swedish Research Council established a national service via DISC, the Swedish National Data Service (Svensk Nationell Datatjänst, SND), which will be responsible for maintaining existing databases, primarily within the social and medical sciences, and for supporting the creation of new ones. In short, Sweden is in the midst of a major overhaul or upgrading of its data infrastructure for research and of pushing it in the direction of e-Science. This seems a good time to review its past and potential data-related practices.

### ***Research data and data-sharing***

The importance of data in all scientific research, independently of disciplines, has been clearly pointed out by Borgman (2007), who has overviewed data-sharing practices as they relate to e-Research. Borgman points out, on the one hand, that there are many similarities between scientific disciplines when it comes to data, e.g. that most disciplines generate large amounts of data, that researchers in all fields use their own and others' data, tools to handle the data, and have certain rules and regulations concerning data-handling and data-sharing. On the other hand, Borgman emphasizes that there are also several differences between disciplines. For example, data differ in terms of their very nature or shape (in meteorology, data can be figures indicating temperature; in sociology they can be voice or video recordings of interviews; and in archaeology they can be bones) and, therefore, in terms of degree of sensitivity. As Borgman (2005: 21) points out, natural science data have a low degree of sensitivity, since they describe natural phenomena, while data within the social sciences and medicine have a high degree of sensitivity since they describe humans. There are other differences between disciplines too in relation to data and data-sharing, says Borgman. For example, data have different economic value depending on their re-use (chemistry data have high value and social survey data low value), and there are disciplinary differences, says Borgman (2007), when it comes to pressure from funders and scientific journal editors to make data available (where there is high pressure within medicine and low pressure within the humanities).

Data-sharing means a larger amount of data available to the individual researcher, which, from a scientific point of view, is a positive thing. As Vickers (2006) points out, most research activities, for example replication of analyses, development of statistical methods, meta-analyses and teaching, can benefit from data-sharing. From the research funding agencies' points of view, data-sharing can be seen as a way of getting better value for the money spent. In the end, research is about knowledge-creation, which at some point should reach society and benefit its citizens by providing better medicines and treatments, better schools, better political systems, and the like. Awareness of the benefits of increased data-sharing has provoked individual researchers (Vickers, 2006), editors of scientific journals (Gardner et al., 2003) and public authorities (NIH, 2009) into raising their voices in defence of open access to research data.

However, although e-Science initiatives actively promote data-sharing activities, most researchers seem to support the idea, and an increasing number of funding agencies and scientific journals are beginning to make this a requirement – the large-scale sharing of research data is still to a large extent a vision. From a technical point of view, storing large amounts of data, and having a large number of distributed users of that data, does not present a problem. From social, ethical and legal points of view, however, there are many problems still to be resolved.

First of all, there are general problems that apply to all scientific disciplines regardless of the kind of data being shared, i.e. problems that have to do with how research is conducted and the crucial role of data in that process. To most researchers, external research funding is indispensable in research; the results and scientific publications arising from them, in turn, are the preconditions for future funding and future research. In the light of this, one can readily see that sharing data with other researchers is almost like giving away money, merit and future possibilities of the same, which may be seen as a problem by many researchers. To some extent, however, this could be compensated in an arrangement by which the data-users acknowledge the data-collectors. Another general problem is in ensuring that the quality of the data is maintained even after having been shared; that is, after having left the initial data-collector (Carlson and Anderson, 2007).

A further issue may involve solutions such as thorough and controlled documentation, well-established standards, well-suited training and reliable services for researchers. This relates to another general issue, which is how research data can be made available for re-use in the first place; this involves digitizing non-digital materials (Carlson and Anderson, 2007), maintaining data over long periods of time and using the data for purposes other than the original within and outside the borders of a nation (Borgman, 2007). Then there are discipline-specific issues; that is, issues relating only to research dealing with data about humans pertains to how the integrity of personal data should be protected when made available to a large and unknown number of users. This is partly a technical question of how data can be securely stored in and retrieved from databases, but has also to do with the morals of researchers, what they do with the data once in their possession.

### ***Data collections and registers in Sweden***

Sweden has long built up a unique collection of databases and registers (Welin, 1990; Jonsson and Landegren, 2001). There are a number of official registers dating back to the eighteenth century (Demographic Database, 2009) and current registers administered by the official authority Statistics Sweden (Statistics Sweden, 2009). These data collections and registers are also to some extent used in research. The uniqueness of Swedish data collections and registers and related national conditions can be described as a combination of several important characteristics (Jonsson and Landegren, 2001).

Many of the registers have been compiled over a long period of time. The first census registers, established in the eighteenth century on the basis of parish statistics, can support genetic studies by providing information regarding the relatedness of sampled individuals.

In Sweden, there are many registers of official data available to researchers via Statistics Sweden, the Swedish central government authority for official statistics. Over a long period of time, Statistics Sweden has been carrying out population-based surveys (of the entire Swedish population), which guarantee a high degree of reliability of the survey data.

The Swedish healthcare system is well-functioning and is today highly advanced, which ensures that sampled individuals have been correctly diagnosed and characterized in medical registers.

In sparsely populated parts of Sweden, there is an increased likelihood that two individuals who have a disease also share the genetic risk factors, which is an advantage in genetic studies.

Since 1947, Sweden has had a system by which every newborn person, or immigrant, is given a national identification number which is used by all public administrations and in most commercial contexts – including economic transactions – to identify the individual (see Otjacques et al., 2006 for a European comparison). This identifier can be used in research for cross-referencing between registers.

Finally, a number of data collections and registers have been built up within individual research projects, and these are available to other researchers through contacting the managers of the individual data collections. To mention a few prominent examples, there is the Swedish Twin Registry, described in detail by Lichtenstein et al. (2002), the Stockholm Birth Cohort Study, described in Stenberg and Vågerö (2006) and the Demographic Database, described at the database's website ([http://www.ddb.umu.se/presentation/index\\_eng.html](http://www.ddb.umu.se/presentation/index_eng.html)).

The above-mentioned characteristics of the Swedish circumstances in relation to data collections and data registers are said to be part of the underlying rationale for the current Swedish e-Science endeavour. Several internal and official documents by the Swedish Research Council – which is the driving force behind the Swedish e-Science enterprise – have expressed the view that the existing national data collections and registers are unique resources for research (Vetenskapsrådet, 2005: 5). What they also say, however, is that they are highly under-used, and mainly for two reasons. One is that many researchers are unaware of what collections and registers exist, since, as mentioned already, they are stored at and available only via particular institutions (Vetenskapsrådet, 2005: 27). The other reason is that it is considered complicated and costly for researchers to access research data from registers and data collections (Vetenskapsrådet, 2005: 27).

A generally accepted description of data collections, according to the Swedish Research Council (Vetenskapsrådet, 2005: 9), distinguishes among registers at official authorities (Statistics Sweden, 2009), at universities and colleges and at hospitals and healthcare institutions. Swedish official statistics are collected, stored and made available by 25 different official authorities, each responsible for its own specific area (e.g. higher education, health and welfare, agriculture). These official authorities all belong to a network, 'Official Statistics of Sweden' ([http://www.scb.se/templates/Listning1\\_\\_\\_139369.asp](http://www.scb.se/templates/Listning1___139369.asp)) headed by Statistics Sweden (SCB). 'Official statistics', according to the website of 'Official Statistics Sweden', are defined as 'statistics for public information, planning and research purposes in specified areas'. In addition, Swedish universities and colleges, as pointed out by the Swedish Research Council in their report (Vetenskapsrådet, 2005: 13), are managers of several research data collections. Over a long period of time they have built up a number of extensive and world unique social science and medical longitudinal databases.

### ***Collecting data in Sweden – public attitudes and media coverage***

A very important precondition for database research is public trust in the data collectors. If the public distrusts researchers or other data collectors (like the state) and denies them information about themselves, little database research could be carried out. However, the public has much to gain from research based on information about themselves. Better medical care, a better educational system and better solutions of social and economic problems are some of the societal benefits of database research. In Sweden, it seems that individuals have long trusted the data authorities with information about themselves. As described above, Sweden has a world-unique set of data registers about its population and the widely spread national identification number has been in use for 50 years. In addition, a survey of the public's attitudes to the collection of individual data shows that people are positively disposed to giving information about themselves to the Swedish authorities. In a recent report from the Ministry of Justice (Justitiedepartementet, 2007) concerning integrity issues in Swedish society, 39 per cent (N = 1,000) stated that they thought the authorities collected a reasonable amount

of information about its population, while 12 per cent said they collected too little information. Only 10 per cent of respondents believed that the authorities collected too much information. In relation to the national identification number, a majority of the respondents, 66 per cent, thought that it was used to a reasonable extent today and only 1 per cent considered that it should be done away with.

However, despite the relatively strong support for continuous data collection and few concerns about integrity issues in relation to collection, storage and sharing of data in Sweden, there have been a few incidents in recent years that have produced headlines in the Swedish newspapers, heated discussions in the media, and even demonstrations by concerned individuals. One incident (described in detail in Welin, 1990) took place in the 1980s, when the largest newspaper, *Dagens Nyheter* (D, 1986), revealed the details of a large social science study of 15,000 individuals in Stockholm, all born in 1953. The study had been ongoing since 1966 and researchers, the newspaper revealed, had collected qualitative as well as quantitative data about the individuals, partly from population-based registers and without the individuals' knowledge. The Metropolit study, as it was known, prompted a huge debate and even demonstrations by the public reacting against the way that researchers could make use of information about individuals unknowingly and for their own undisclosed purposes with financial support from the state. Without going into detail about the study and the reactions to it here, this incident, as pointed out by, for example, Hermerén (1986), initiated the first open discussion of research ethics in Sweden and a strong and negative reaction against the 'secret' activities of the researchers.

Another more recent case in Sweden, which caused not only uproar in the media but also became a legal issue, was the so-called 'Gillberg case'. Briefly, this was about a professor in child psychiatry, Christopher Gillberg, who was ordered to hand out research material to another researcher for inspection, but instead destroyed the material in order to protect the study participants' integrity (children diagnosed with Attention-Deficit Hyperactivity Disorder and their parents). For his action, Professor Gillberg was conditionally sentenced and fined by the court of appeal in February 2006.

Yet another, and even more recent, debate in Sweden had to do with the new FRA law ('Försvarets radioanstalt' or the National Defence Radio Establishment), which allows signals intelligence to be carried out by the National Defence Radio Establishment on commission by the Swedish authorities on data traffic (Internet and telephone) over borders to and from Sweden. The FRA law was proposed in the spring of 2007 and approved by the Swedish parliament in the summer of 2008. It, too, was heavily criticized and widely discussed by politicians, the public and academics in political forums, in blogs and in the Swedish media before and after the decision. The strong criticism finally led to a number of changes to the law, the main one being stronger protection of the integrity of the individual. The law finally came into force on 1 January 2009.

What can be said about these three incidents – each causing, at different times and in different contexts, massive debate and uproar among the public as well as within the scientific community and the political sphere – is that they touched on several of the issues at the very core of e-enabled data-sharing. In the Metropolit and Gillberg cases, the researchers were accused by the public, by other researchers and by the media, of not being open about their actions. However, the two cases differ. In the Metropolit case, the lack of openness was due to the fact that the researchers were unaware of the concerns that the public, and especially the study participants, might have had about having information about themselves stored in computers and shared among and studied by researchers. Moreover, Swedish legislation did not prohibit their way of using the data or of conducting research. Nevertheless, owing to the public reaction, further studies on the material were prohibited without the consent of the study participants. In the Gillberg case, on the other hand, the lack of openness was a way for the

researcher to protect the integrity of the study participants. In this case, however, the researcher was judged to have acted incorrectly in law. Nevertheless, further research on the data and perhaps new knowledge regarding the disease under study was rendered impossible. The most recent case, the FRA law, also has a clear bearing on e-enabled data-sharing in Sweden. One reason is that the law, and even more so the debates preceding the law, made Swedes in general think, worry about and debate the fact that even the (Swedish) state might want to collect and use information against its own citizens. This awareness may affect attitudes among Swedes in general in a negative way towards attempts to increase accessibility to research data, among other things putting people off participating in studies where sensitive data are collected.

## **Method**

In order to gain insight into the possibilities and constraints of increased data-sharing in Sweden, interviews were conducted with 16 informants, 15 of whom belonged to two or three of the stakeholder categories: database managers, database researchers or database funders (the informant who can be said not to belong to any of these categories was the coordinator of a bio bank programme at a Swedish university; this programme arranges and supports national activities such as research, development and education within the bioengineering field). All participants live and work in Sweden. During the interviews, the informants were introduced to the concept of e-Science both in our initial request to participate in the interview and at the beginning of their interview. e-Science was broadly defined as science that would be carried out through distributed global collaborations enabled by the Internet. A central feature of such collaborative scientific enterprises is that access will be required to large data collections, large-scale computing resources and high performance visualization.

The interview questions were specifically designed to elicit the informants' perspectives on the Swedish e-Science data-sharing enterprise and, particularly, to influence the way in which the databases built up, managed and used would influence e-Science in the future. We were interested in the implications of increased accessibility to data in research databases and the possibilities of handling data via e-Science tools among different stakeholders. Among other things, for example, one can foresee new and increased research possibilities, new types of legal issues to be dealt with and new integrity matters.

All interview questions were open-ended, and follow-up questions were asked to ensure that informants' views were captured. Because all interviews, apart from one, were conducted in Swedish, quotes from the informants in the sections below have been translated from Swedish. Every effort has been made to ensure that the voice of the informant resonates in the translation.

## **Results and discussion**

It is impossible in the limited space available to present a full discussion of the issues raised, or to present the range of views expressed by our informants. In what follows, we nevertheless highlight key points and provide some illustrative responses.

In some respects, the problems of data-sharing are not unique to e-enabled data, but rather relate to research collaboration in general. Intellectual Property (IP) issues, security issues (especially relating to sensitive data), data-formatting (data curation, metadata creation), the need for training and support for database researchers in their use of databases, and the like, all occur whenever data are shared. Nevertheless, as one interviewee stated: 'Qualitatively there will not be any differences, but quantitatively' (Database funder and database researcher). Put differently, the more users, or the more distributed users and unknown users, the more problems can be expected.

Our interviewees were well aware of the unique possibilities in Sweden:

Longitudinal population-based research has been asked for in international research where limited studies have been conducted with measures only at one point in time /... / We have had unique conditions here in Sweden to conduct this kind of research. Among other things we have a well-functioning national registration from which we can select a certain sub-group and we can follow these individuals wherever they go and move. Our American colleagues would go mad if they even tried to get the same kind of data input to work. (Database manager and database researcher)

The same researcher also put this point in terms of 'competitive advantage':

In Sweden we have a niche in this, to conduct longitudinal population based research, which is a competitive advantage in comparison to other countries, based on the fact that we can collect data of this kind. (Database manager and database researcher)

At the same time, they were well aware that this advantage could be lost in the new environment:

I can see a great danger in that we are about to create legal systems such that the incentives to carry out research and collect data about people, which used to be easy in this country, will be so difficult that people rather give up /... / I think we have imported the worst parts of the American system, which is where we assume that misuse is what is going to take place /... / This is in a way to hinder what could be the most prominent part of this country's research, namely longitudinal research. (Database manager and database researcher)

On one side, this is the question of preserving the special relationship that obtains in Swedish society as a whole:

Openness, openness, openness ... Information [to study participants] about a particular study but also information and education in general about how important knowledge is in our society /... / In other societies this is not the case, people do not think they need to participate in relationship to the state and to society. ... (Database funder and database researcher)

On the other side, it is not a question simply of consensus, but also of ensuring that the views of the minority are taken into account:

... then we have those 1–1.5% of people that would not accept this [giving out sensitive information about themselves to research]. If we do not show consideration for them, then the whole system will break down. (Database funder and database researcher)

Questions about data protection were also considered, not just in the context of researchers and research participants, but in terms of how trust might be broken outside this relationship:

I am not at all afraid that people who primarily have a scientific interest in this would cross the line of what is appropriate. The individuals [in these databases] are of no interest to us researchers. We are looking for patterns, we have no interest in whether there is someone in this whom we know that has some specific genetical structure. (Database manager and database researcher)

However, several of our interviewees mentioned the risk of journalists or other members of the public who might obtain sensitive information and misuse it, especially about famous people, and thus threaten trust and the possibilities for collecting data in the future.

Again, the specifically Swedish nature of this issue was recognized:

In Sweden we have more problems with secrecy compared to other countries, because our surveys are complete while in other countries the quality of the sources are too low, they have only selections. (Database manager)

At the same time, in the event of collaboration with international colleagues, the informal relations of trust would need to be extended in studies requiring shared data. As one researcher using longitudinal data told us:

The study was conducted in collaboration with American colleagues and no single individual had been capable of carrying out this on his own /... / The collaboration was not the least initiated or administered from above but totally based on trust, on personal contacts, on the fact that people liked each other ... (Database manager and database researcher)

Trust, in short, would need to extend into an e-Science environment:

The Swedish legislation is very rational, both for researchers and individuals. It takes into account the needs of the society as well as the needs of the individual for protection ... The question of trust is the key! (Database funder and database researcher)

In view of the new e-Science initiatives, the funding council and lawmakers would therefore need to tackle problems of data re-use:

Swedish legislation does not allow re-use of data /... / which is something the Committee for Research Infrastructures [Swedish KFI] really should try to do something about /... / The whole idea with Swedish National Data Service [Svensk Nationell Datatjänst, SND] builds on this, but the legislation does not allow it. (Database manager)

Apart from issues of trust and privacy in society, our interviewees anticipated a number of specific issues on the creation of shared databases that would need to be addressed. One was the relevant expertise:

To have someone from the outside to advise us when managing our database would be very valuable, in order to configure it for future e-Science uses that we do not have knowledge about today. (Database manager and database researcher)

Perhaps Sweden has come relatively late to e-Science, or perhaps this researcher believes that other e-Science efforts are more advanced than is warranted. Other issues have been recognized elsewhere, such as the short funding cycles, which may mean that databases do not come into widespread use, or focus on new data rather than re-using it:

People move in and out of studies. What is it that ensures the continuity of a database? A lot of existing databases are of high quality, but have fallen into oblivion. That is a waste of resources. But the research funding system tends to reward collecting data above using already collected data. (Database manager and database researcher)

Developing shared data sets and repositories is not necessarily to do with e-Science, but with being able to have access to linked resources without needing to be a specialist in particular tools:

For me [as a database researcher] the most important thing is to be able to put forward my research questions and link different databases. I should not have to know how databases are constructed or anything about e-Science to use the tools. That is the real challenge. (Database funder and database researcher)

e-Science is not about technology, but about making use of it. Sharing data, however, does require researchers willing to make their data available. Several of our interviewees mentioned that the predecessor to the Swedish National Data service (SND), the Social Science Data Service, had not been very successful in this. One way that is becoming increasingly popular is the requirement of a deposit of data for publicly funded projects, but this has not yet been decided in Sweden:

At what point should data be released to other researchers? We have not yet come to a conclusion on that point yet . . . However, there is the ethical aspect of it and requirement from the main national research funding agency in Sweden, that researchers should make their data available to others so they can check whether the research is sound, but there is no requirement to make raw data available to others to carry out research on that data. NIH [National Institutes of Health, USA] requires that researchers should share data and that is a good model, however not yet developed to perfection. (Database funder and database researcher)

## Conclusion and outlook for the future

The core of the Swedish e-Science enterprise is, with good reason, data registers, database research and data-sharing. The unique Swedish circumstances have been identified as providing a competitive advantage in international research, both by researchers and by funding bodies, but there is still a long way to go before these uniquely favourable conditions can be used to their full potential. E-enabling of Swedish data registers is only partly a technical issue, and this is understood by the various stakeholders in the Swedish e-Science enterprise. The stakeholders that we interviewed in the current study, for example, seem to worry little about storage capacity or about which middleware to use to make data available, and much more about how the exceptional Swedish conditions that have been built up over a long period of time can be preserved and prevented from being misused and thus ruined for research. In other words, as Sweden tries to build up an infrastructure for e-Research around its unique data collections, stakeholders are aware that the key will be to maintain the unusual conditions of trust that have obtained in Sweden.

Several incidents in relation to research on Swedish data registers and data-sharing indicate that the public's trust, in the state and in researchers, is fragile and could readily be destroyed, and that it would take years to re-construct that confidence. Significantly, the two main overarching issues that the interviewees in this study stressed were openness and safety. Openness with what researchers are up to, to the public and within the scientific community, and the safety of – or protection of – individuals who contribute with information about themselves to the data registers. Like all science, e-Science depends on the data to analyse; the difference is that in e-Science the numbers of data points and users are much larger. However, if Swedish database researchers, database managers and database funders take the demand for openness and integrity protection seriously, and as a joint challenge already now, in the process of scaling up the quantity of – and accessibility to – data with e-Science, then there will be few obstacles for Swedish database research in the future. However, if the Swedish e-Science enterprise pushes ahead too quickly and starts making data gathered from the public accessible to wider groups of researchers too soon, then Swedish e-Science may encounter a number of challenges. More specifically, without ensuring transparency towards the public and researchers regarding how data will be used, and by whom, and without protecting individual integrity, it may be that individuals will be reluctant to provide researchers with information about themselves, or that researchers will hesitate to deposit their research data in accessible repositories or use other researchers' data. Therefore Sweden needs to tread the path to data-sharing carefully. It will be especially important to inform the public about the benefits of making research data accessible, about the role of data in research in general and about the legal rights of individuals in relation to data collections. It will also be very important to inform and educate researchers regarding the benefits of sharing research data, regarding their rights to their own data after deposition, and regarding the effective legal framework concerning the protection of individuals. In other words, the uniquely high level of trust that has so far governed the relation between researchers, the public and a social order in which personal data are thought to be safely and transparently managed – this fragile trust must now be

extended into a new environment. Obviously, other countries do not share the same setting of trust, and would need to build on their relation between the public, researchers and the way in which personal data are governed – though some of the lessons may be transferable.

## Acknowledgements

We thank the informants and the Swedish Foundation for International Cooperation in Research and Higher Education (STINT) for making this study possible. Schroeder's work on the article was supported by ESRC grant RES-149-25-1022 for the Oxford e-Social Science (OeSS) project: Ethical, Legal and Institutional Dynamics of Grid-enabled e-Science.

## References

- Berman, F. and Brady, H. (2005) Final Report: NSF SBE-CISE Workshop on Cyberinfrastructure and the Social Sciences. Available at: <http://www.sdsc.edu/about/director/pubs/SBE/reports/SBE-CISE-FINAL.pdf> (accessed April 2009).
- Borgman, C. L. (2005) 'Disciplinary Differences in e-Research: An Information Perspective'. Keynote talk at the First International Conference on e-Social Science, Manchester, UK. National Center for e-Social Science. Available at: [http://www.ncess.ac.uk/events/conference/2005/papers/presentations/ncess2005\\_borgman.pdf](http://www.ncess.ac.uk/events/conference/2005/papers/presentations/ncess2005_borgman.pdf) (accessed April 2009).
- Borgman, C. L. (2007) *Scholarship in the Digital Age: Information, Infrastructure, and the Internet*. Cambridge, MA: MIT Press.
- Carlson, S. and Anderson, B. (2007) 'What are Data? The Many Kinds of Data and Their Implications for Data Re-Use', *Journal of Computer-Mediated Communication* 12, article 15. Available at: <http://jcmc.indiana.edu/vol12/issue2/carlson.html> (accessed April 2009).
- Crabtree, A., French, A., Greenhalgh, C., Benford, S., Chevherst, K., Fitton, D., Rouncefield, M. and Graham, C. (2006) 'Developing Digital Records: Early Experiences of Record and Replay', *Computer Supported Cooperative Work: Journal of Collaborative Computing* 15: 281–319.
- Dagens Nyheter* (1986) '15 000 svenskar i hemligt dataregister' [15 000 Swedes in Secret Computer Register], (10 February), Stockholm.
- Demographic Database (2009) Available at: [http://www.ddb.umu.se/index\\_eng.html](http://www.ddb.umu.se/index_eng.html) (accessed April 2009).
- DISC (2009) Available at: <http://www.disc.vr.se/> (accessed April 2009).
- Galison, P. and Hevly, B. (eds) (1992) *Big Science: The Growth of Large-Scale Research*. Stanford, CA: Stanford University Press.
- Gardner, D., Toga, A. W., Ascoli, G. A., Beatty, J., Brinkley, J. F., Dale, A. M., Fox, P. T., Gardner, E. P., George, J. S., Goddard, N., Harris, K. M., Herskovits, E. H., Hines, M., Jacobs, G. A., Jacobs, R. E., Jones, E. G., Kennedy, D. N., Kimberg, D. Y., Mazziotta, J. C., Miller, P., Mori, S., Mountain, D. C., Reiss, A. L., Rosen, G. D., Rottenberg, D. A., Shepherd, G. M., Smalheiser, N. R., Smith, K. P., Strachan, T., Van Essen, D. C., Williams, R. W. and Wong, S. T. C. (2003) 'Towards Effective and Rewarding Data Sharing', *Neuroinformatics* 1: 289–95.
- Gibbons, S. M. C., Kaye, J., Smart, A., Heeney, C. and Parker, M. (2007) 'Governing Genetic Databases: Challenges Facing Research Regulation and Practice', *Journal of Law and Society* 34: 163–89.
- Heldal, I. (2007) 'Supporting Participation in Planning New Roads by Using Virtual Reality Systems', *Virtual Reality* 11: 145–59.
- Hermerén, G. (1986) *Kunskapens pris: forskningsetiska problem och principer i humaniora och samhällsvetenskap* [The Price of Knowledge: Problems in Research Ethics and Principles within Humanities and Social Sciences]. Stockholm: HSFR.
- Jackson, S. J., Edwards, P. N., Bowker, G. C. and Knobel, C. P. (2007) 'Understanding Infrastructure: History, Heuristics, and Cyberinfrastructure Policy', *First Monday* 12(6). Available at: <http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1904/1786> (accessed April 2009).

- Jankowski, N. W. (2007) 'Exploring e-Science: An Introduction', *Journal of Computer-Mediated Communication* 12, article 10. Available at: <http://jcmc.indiana.edu/vol12/issue2/jankowski.html> (accessed April 2009).
- Jonsson, L. and Landegren, U. (2001) 'Storing and Using Biobanks for Research', in M. G. Hansson (ed.) *The Use of Human Biobanks. Ethical, Social, Economical and Legal Aspects*, pp. 1–8. Report I, Uppsala University.
- Justitiedepartementet (2007) 'Skyddet för den personliga integriteten. Kartläggning och analys' [The Protection of Personal Integrity. Mapping and Analysis]. Statens offentliga utredningar, SOU 2007:22 [The Swedish Government Official Reports, SOU 2007: 22].
- Lichtenstein, P., deFaire, U., Floderus, B., Svartengren, M., Svedberg, P. and Pedersen, N. L. (2002) 'The Swedish Twin Registry: A Unique Resource for Clinical, Epidemiological and Genetic Studies', *Journal of Internal Medicine* 252: 184–205.
- National Institutes of Health (2009) Available at: [www.nih.gov](http://www.nih.gov)
- Otjacques, B., Hitzelberger, P. and Feltz, F. (2006) 'Identity Management and Data Sharing in the European Union'. Proceedings of the 39th Hawaii International Conference on System Sciences, 4–7 January, 4, IEEE Computer Society, Washington, DC, USA, p. 70.1.
- Schroeder, R. (2007) 'e-Research Infrastructures and Open Science: Towards a New System of Knowledge Production?', *Prometheus* 25(1): 1–17.
- Schroeder, R. (2008) 'e-Sciences as Research Technologies: Reconfiguring Disciplines, Globalizing Knowledge', *Social Science Information* 47(2): 131–57.
- Schroeder, R. and Fry, J. (2007) 'Social Science Approaches to e-Science: Framing an Agenda', *Journal of Computer-Mediated Communication* 12, article 11. Available at: <http://jcmc.indiana.edu/vol12/issue2/schroeder.html> (accessed April 2009).
- Statistics Sweden (2009) Available at: [www.scb.se/eng](http://www.scb.se/eng) (accessed April 2009).
- Stenberg, S.-Å. and Vågerö, D. (2006) 'Cohort Profile: The Stockholm Birth Cohort of 1953', *International Journal of Epidemiology* 35: 546–8.
- Svensk Nationell Datatjänst (SND) (2009) Available at: <http://www.ssd.gu.se/?lang=en> (accessed April 2009).
- Swedish Twin Registry (2009) Available at: <http://ki.se/ki/jsp/polopoly.jsp?d=9610&l=sv> (accessed April 2009).
- Vetenskapsrådet (2005) 'Strategi och infrastruktur för världsledande forskning på svenska register' [Strategy and Infrastructure for World Leading Research on Swedish Registers]. Internal report based on an investigation concerning the possibilities of improving the conditions for research on Swedish register data.
- Vetenskapsrådet (2009) Available at: [www.vr.se/huvudmeny/forskarservice/forskningsinfrastruktur/evetenskap/discdatabaseinfrastrukturcommittee.4.aad30e310abcb9735780001557.html](http://www.vr.se/huvudmeny/forskarservice/forskningsinfrastruktur/evetenskap/discdatabaseinfrastrukturcommittee.4.aad30e310abcb9735780001557.html) (accessed 18 March 2009).
- Vickers, A. J. (2006) 'Whose Data Set Is It Anyway? Sharing Raw Data from Randomized Trials', *Trials* 7(15). Available at: <http://www.trialsjournal.com/content/7/1/15> (accessed April 2009).
- Welin, S. (1990) 'The Computerized Scientist – Moral and Political Issues', in A. Elzinga, J. Nolin, R. Pranger and S. Sunesson (eds) *In Science We Trust? Moral and Political Issues of Science and Society*, pp. 155–76. Lund: Lund University Press.

**Biographical Note:** Ann-Sofie Axelsson is a lecturer in the Department of Technology Management and Economics at the Chalmers University of Technology. She has published extensively on the social aspects of how people interact in virtual environments, including the book (co-edited with Ralph Schroeder) entitled *Avatars at Work and Play* (Springer, 2006) and a forthcoming co-edited book (in Swedish) about technological literacy, *På spaning efter teknisk bildning* [In Search of Technological Literacy] (Liber, 2009).

**Address:** Department of Technology Management and Economics, Chalmers University of Technology, SE-412 96, Göteborg, Sweden. [email: [annaxe@chalmers.se](mailto:annaxe@chalmers.se)]

**Biographical Note:** Ralph Schroeder is Senior Research Fellow at the Oxford Internet Institute of Oxford University. His publications include *Rethinking Science, Technology and Social Change* (Stanford University Press, 2007) and *Max Weber and the Sociology of Culture* (Sage, 1992). Before coming to Oxford, he was Professor in the School of Technology Management and Economics at the Chalmers University of Technology.

**Address:** Oxford Internet Institute, University of Oxford, 1 St Giles, Oxford OX1 3JS, UK. [ralph.schroeder@oii.ox.ac.uk]