Ethics in Data Sharing

a best practice for NRENs

Roland van Rijswijk-Deij
SURFnet bv
roland.vanrijswijk@surfnet.nl

Keywords

ethics, privacy, data sharing, policy, best practice, NREN

1. INTRODUCTION

NRENs have a unique position in the world of networking; we operate large and often complex networks with ultra-high bandwidth and cutting edge technology. But we are also – generally – publicly funded. This makes us attractive to academic researchers who perform research on networks and network security, since we have access to large amounts of operational data for our networks, but no commercial goals that inhibit us, in principle, from sharing this data for research purposes. This data is highly valuable to researchers as they often lack ground truth for models that they develop or because they have the need to perform intricate measurements on operational high-speed networks.

SURFnet, like many NRENs, regularly receives requests from researchers to share all kinds of operational data ranging from network topologies to flow data and all the way to packet captures for specific services. Our current practice is that we only share data with *trusted researchers* (i.e. people we know) under strict conditions set out in a Non-Disclosure Agreement (NDA), that stipulates:

- what data is shared;
- for what (research) purpose the data may be used;
- who may access the data;
- the period for which data is shared;
- what conditions apply for publication (e.g. anonymisation requirements, review by SURFnet, ...);
- when the data must be destroyed.

There are several problems with our current practice. First, we only share data with trusted researchers. This hampers scientific research as we have no process to share data with legitimate researchers outside our circle of trust. Second, we require researchers to destroy the data we provide after a certain, often short period of time. This is bad for reproducibility of research and goes against what is rapidly becoming a core academic value, the long-term curation of research data. Finally, depending on the kind of data that is shared there may be privacy and wider ethical concerns. For instance, flow data is highly privacy sensitive and reveals a lot about the individual users of our network. These concerns are now dealt with on an ad hoc basis, where we err on the side of caution; a more well-defined process where these concerns are addressed more methodically is, in our opinion, warranted.

2. DAGSTUHL SEMINAR

In early 2014 a so-called Dagstuhl Seminar[1] was organised by an international team of academics from Computer Science, Ethics and Law. The goal of this week-long seminar was to discuss the ethical implications of computer science research with a particular focus on network and network security research. The seminar was attended by twenty representatives of academia and industry. Like the organising team, representatives present at the seminar represented three disciplines: computer science, ethics and law. The starting point for the seminar were the ACM and IEEE Ethics Codes of Conduct and the so-called Menlo Report [1, 2]. The goal of the seminar was to come to a joint statement or model on how to deal with ethical issues in computer science research as the participants recognised that there is a need for practical guidance, especially where the sharing of, for example, network data for research purposes is concerned.

2.1 Dagstuhl Ethics Model

The principal outcome of the seminar was a novel model that strives to examine ethical values in all stages of research. The research stages are defined as:

- Research definition
- Research design

[1]For more information see `http://www.dagstuhl.de/programm/dagstuhl-seminare/`

- Data Collection
- Data Storage
- Analysis
- Verifiability
- Dissemination
- Curation

In every stage of the research, all actors involved in the research should make explicit, given their role (e.g. researcher, network operator, ...) and their context (e.g. working for a commercial enterprise, a university, non-profit, ...), what their ethical values are with respect to the research at hand and the kind of data that is to be exchanged or analysed. Figure 1 shows this schematically.

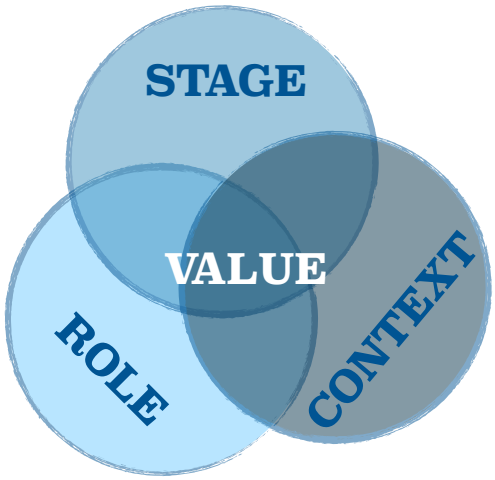

Figure 1: Dagstuhl Ethics Model

Take, for example, a research project in which the principal investigator wishes to use flow data from an NREN and asks this NREN for the data. The investigator wants access to all flow data for a certain link and would like full access (i.e. not anonymised) to the data. Ethical considerations for the researcher could for instance be *effectiveness, non-malificence* and *serendipity*. An NREN staff member dealing with this request, on the other hand would consider *privacy* and *reputation*. The goal of the model developed at Dagstuhl is that these values are expressed explicitly, inviting discussion between all actors about their motivations and ethical concerns. This is radically different from what is common practice where this discussion either stays within the silo of the own organisation or some ethical values are implicitly (and often incorrectly) considered to be universally held by all actors.

A position paper on this model was presented at the CREDS workshop co-hosted with the IEEE Symposium on Security & Privacy in May 2014 [3].

3. BRINGING IT INTO PRACTICE

There are two shortcomings in current discussion on ethics in computer science research. First, it often stops at discussion and is seldom brought into practice other than in the form of an ethics paragraph in a paper. Second, in cases where data is acquired by researchers from a third party (such as an NREN) there is often no clearly defined process that considers the ethics from both sides of a data sharing relationship.

As a publicly funded entity, SURFnet feels that it has an obligation to facilitate network and security research. Indeed, we believe that we should share *more data*, *more often* but under clear conditions that respect *the privacy of our users and institutions*, *ethics* and *the law*. In order to take our data sharing practice to a new level of professionalism, we have started a project to design a new policy for dealing with data sharing for research purposes. The project team was formed after the Dagstuhl seminar and comprises of participants from this seminar. Again, multiple disciplines are represented in the team (computer science, ethics and law) as well as multiple roles (researcher, ethicist, NREN employee, legal counsel). During several project meetings over the course of 2014, we have revisited some of our discussions at Dagstuhl as well as added specific legal expertise that covers EU Privacy Law as any data sharing between SURFnet and researchers will be subject to that law. Our end goal is to draft a new data sharing policy which will be ready by the end of 2014. We will then pilot this policy for any data sharing requests we get over the course of the next year and evaluate our experiences. We intend to publish the results of our evaluation as a best practice for other NRENs or network operators that deal with similar requests from researchers. We already have a number of data sharing requests lined up that will be subjected to the new policy.

4. PRESENTATION AND BOF

We believe that we are not alone in dealing with the dilemmas around data sharing requests, nor do we believe that our approach is a panacea that solves all these dilemmas. Nevertheless, we believe we have a novel approach that may have worth to fellow European NRENs and we would like to have an active discussion with international colleagues to further improve our approach. As such, we would like to present our research during a TNC 2015 session (in a 25 minute slot) and we are also open to the idea of organising a birds-of-a-feather (BoF) session if there is sufficient interest in this.

5. ACKNOWLEDGEMENTS

The author would like to thank all participants to Dagstuhl seminar 14052 on Ethics in Data Sharing, held from January 26th to 31st 2014[2].

Part of this work has been supported by the EU-FP7 FLAMINGO Network of Excellence Project (318488) and by the GigaPort3 programme funded by the Dutch Economic Structure Enhancing Fund (FES).

6. REFERENCES

[1] David Dittrich and Erin Kenneally. The Menlo Report: Ethical Principles Guiding Information and Communication Technology Research. Technical report, U.S. Department of Homeland Security, August 2012.

[2] Michael Bailey, David Dittrich, Erin Kenneally, and Doug Maughan. The Menlo Report. *IEEE Security & Privacy Magazine*, 10(2):71–75, March 2012.

[3] Sven Dietrich, Jeroen van der Ham, Aiko Pras, Roland van Rijswijk-Deij, Darren Shou, Anna Sperotto, Aimee van Wynsberghe, and Lenore Zuck. Ethics in Data Sharing: developing a model for best practice. In *2nd Cyber-security Research Ethics Dialogs & Strategy (CREDS II), co-located with the 35th IEEE Symposium on Security and Privacy (IEEE S&P)*, San Jose, CA, USA, 2014. IEEE Comput. Soc.

[2] `http://www.dagstuhl.de/en/program/calendar/semhp/?semnr=14052`