SURFnet Data Sharing Policy

Legal and ethical guidelines relating to data sharing for research purposes

Authors: Chloë Baartmans[1], Aimee van Wynsberghe[2], Roland van Rijswijk-Deij[1], Evelijn Jeunink[1], and Floortje Jorna[1]

Date: June 2016
Version: 1.0

[1]SURFnet bv
http://www.surf.nl/en/about-surf/subsidiaries/surfnet

[2]University of Twente
http://www.utwente.nl/en/

Contents

1 Introduction

As the Dutch National Research and Education Network (NREN) SURFnet is committed to advancing scientific research in the areas of computer networking and security. Part of this commitment involves the sharing of network traffic data and operational information about the services SURFnet operates with universities and research institutes, for scientific research. Providing access to this kind of data is essential for generating knowledge in network and security research and helps researchers in SURFnet's constituency reach the global top of academic research.

While recognizing the great potential of sharing data, SURFnet also recognises that methods for sharing data will vary and thus careful assessment is needed. In fact, there are multiple reasons for which assessment is necessary: the sensitivity of the data, size and complexity of datasets, and/or volume of the request anticipated. The fact that this data can encompass (very) sensitive information makes it, prior to sharing, susceptible to checks that safeguard the privacy, confidentiality and security of the data in an ethically responsible manner.

SURFnet believes its responsibility as steward of the data extends beyond legal implications and conditions, and wishes to explicitly take the ethical implications of research performed with the data it collects into consideration. When entering into agreement with SURFnet, it is thus important that sharing partners consider and share these implications.

For these reasons, SURFnet has established this policy. With it SURFnet strives to act as a role model, setting general conditions for sharing operational data with researchers. As such, the policy aims to uphold and enforce a high level of privacy protection, which tackles privacy issues as efficiently, effectively and early as possible. This allows SURFnet to continue to pursue and even broaden its possibilities for data sharing, while extending the legal and ethical responsibilities that come with it to its future sharing partners.

2 Policy

1. As operator of the SURFnet network and as operator of advanced ICT services for education and research, SURFnet has access to operational data on and from these services. For example, SURFnet monitors traffic across the network and derives flow information for this traffic as part of its routine operational practice. At the application level, SURFnet, for example, collects logging information for authentication requests through the SURFconext identity federation. The main purpose of these kinds of information is to **ensure the proper operation of the network or service** and to **protect customers and users of the network or service**. These kinds of information can also be useful for all types of research on networks and security that SURFnet believes will help it improve and safeguard its future operations. SURFnet will therefore endeavour to work with researchers who wish to make use of operational data for these purposes and uses the possibilities Dutch privacy law offers to share personal data in the context of research.
2. Any use of operational data for research must not hinder the normal operation or use of the network or service, nor influence (have an impact on) any users or organisations connected to the network. The sharing of data must also comply with contractual obligations of SURFnet and the law, in particular the *Wet Bescherming Persoonsgegevens (WBP) 2001*, and associated legislation. The conditions set out in this policy are thus applicable to the: collection, use, storage, curation, dissemination, sharing and destruction of the data.
3. In all cases SURFnet requires that data is processed in accordance with the data protection principles of the *Wet Bescherming Persoonsgegevens (WBP) 2001*, affording data subjects their legal rights and protections. Under no conditions will SURFnet share exceptional personal data as defined in Article 16 of the WBP 2001[1].
4. SURFnet can make the following information available for use in research:
   a) network flow data;
   b) captured traffic data;
   c) DNS data;
   d) logging data;
   e) aggregate data, statistics.

[1]This includes, for example, information about a person's religious beliefs, sexual orientation, ...

5. SURFnet will, in every case, require a binding commitment from the sharing partner's organisation that shared data are **subject to technical and organisational measures** to ensure protection of the data against unauthorised access or modification, loss or misuse of the data. In particular the data must only be used for the particular research for which it was requested and disclosed, and must not be used for any other purpose.
6. The conditions to which the research is subject depend on the privacy risk level in which the requested dataset is categorised. This risk level is determined by SURFnet employees based on the classification that accompanies this policy[2], for each individual request. This decision is based on a number of factors, for example: type of data, amount of data, sensitivity of data, complexity of the dataset, the anticipated data volume, the specific research purpose, retention terms, etc. There are three risk levels: **low risk**, **medium risk** and **high risk**. Each data request is labelled according to the potential risk to privacy, unless otherwise agreed between SURFnet and the researcher requesting data. Each request will receive a binding agreement based on the risk level.
    a) **Low risk** – Data that are anonymous (do not contain original IP addresses, network numbers, original log-in data or any other information that can associate it with a particular individual or institution connected to the SURFnet network). Appropriate methods of anonymisation may be discussed with the individual researcher. SURFnet will in all cases, in collaboration with the researcher, determine whether the research in question can be performed with merely/only anonymised data. Where data sharing concerns low risk (anonymous) data, SURFnet's principle aim is collaboration with data sharing partners through means of a binding agreement between SURFnet and the researcher, or the researcher's organisation/institution in question.
    b) **Medium risk** – Personal data that are pseudonymised or functionally separated from any personally identifiable data[3] where research can be undertaken on pseudonymised or functionally separated data (absolute anonymisation cannot be guaranteed). Appropriate methods of pseudonymisation or functional separation may be discussed with the individual researcher. Where data sharing concerns medium risk data, SURFnet will require (in addition to a binding agreement with the researcher or researcher's institution) a binding undertaking from the researcher's organisation that no attempt will be made to strip away the pseudonymity and no data or statistics will be published that would allow others to do so. SURFnet reserves the right to request information from sharing partners concerning the outcome of research related to such data and any consequences that may arise from sharing.
    c) **High risk** – Data that are neither anonymous nor pseudonymous and thus personally identifiable data. In this case the researcher must explicitly demonstrate which specific (parts of the) data cannot be anonymised and why; the aim being that any remaining part will be anonymised or pseudonymised in the normal way. Where data sharing concerns high risk data, SURFnet will

[2] See Appendix A for details on the classification.
[3] Data that cannot be attributed to a specific data subject without use of additional information, to which the individual researcher has no access.

require a **mandatory ethical review**[4] of the request to assist in approval or denial of a binding agreement with the researcher or researcher's institution. In addition, SURFnet reserves the right to only allow in-house access (the researcher is invited to access the data on SURFnet's premises, and data may not leave said premises) to researchers that are granted access to high risk data from the network. SURFnet reserves the right to request information from sharing partners concerning the outcome of research related to such data and any consequences that may arise from sharing.

7. The risk level of a data sharing request is the sole responsibility of SURFnet. Final approval of the risk level is performed by SURFnet's **privacy officer**.
8. Researchers requesting data from SURFnet will in all cases be asked to provide a short **assessment of the ethical implication(s)**[5] of their research. This will aid SURFnet in its assessment of the request. Where data sharing concerns medium risk data, SURFnet reserves the right to require, and allows the researcher the right to request – in exceptional cases – an **ethical review** in addition to an ethical implication assessment. Where data sharing concerns high risk data, an ethical review is mandatory in addition to the ethical implication assessment.
9. Unless otherwise agreed between SURFnet and the researcher, medium and high risk data will either **be destroyed or returned to SURFnet** in its capacity as steward of data from/about the SURFnet network. If the data are returned to SURFnet, the data will in principle also be destroyed. SURFnet realises that academic morals increasingly require researchers to retain source data used for specific research to aid reproducibility and safeguard scientific integrity. Therefore, in exceptional cases, SURFnet will strive to provide an acceptable manner to curate specific data, under sole control of SURFnet.
10. Unless otherwise agreed upon between SURFnet and the researcher, the researcher shall under no circumstances **disclose the data to any third party**, nor **use the data for any purpose other than the research purpose defined**. Again, SURFnet reserves the right to request information from the sharing partner in the event that SURFnet believes the data to be used for unintended purposes.
11. Unless otherwise agreed upon between SURFnet and the researcher, the researcher's institution is **responsible for the internal management** of the low or medium risk data immediately after transfer of the data, and will have to agree non-disclosure with any individual(s) who will gain access to the data on their behalf and disclosure shall only be granted within limits of the research purpose.
12. Where joint research involves countries outside the European Economic Area, further conditions relating to **cross-border transfers** need to be considered. If a request originates from a country outside the EEA, and its level of data protection is lower than the minimum level of protection required by Dutch law, SURFnet will deny the research request.
13. Unless otherwise agreed upon between SURFnet and the researcher, **SURFnet reserves the right to publish all research**, which has made use of SURFnet network data, in its public repository. The purpose of this public repository is to

[4] For more information, see Appendix B 'the ethical review process explained'.
[5] For more information, see Appendix C 'Writing an assessment on ethical implications of research'.

provide insight in, and transparency regarding, past uses of SURFnet network data. Publication will not infringe upon the privacy of individuals.

14. In all cases SURFnet requires that **researchers provide public acknowledgment** of the use of data provided by SURFnet in their research publication (including web pages, papers published by a third party, and publicly available presentations). Please use the following generic acknowledgement text (and edit it as appropriate):

    - (Part of) the data used for this work was provided by SURFnet, the National Research and Education Network in the Netherlands. For more information see `http://surf.nl/datasharing`

Appendix A Data Risk Classification

Each data sharing request is subject to a risk classification based on the expected impact on the privacy of users and/or connected institutions on the SURFnet network. The table below specifies the risk categories and describes when data fall under that specific risk category. Note that this table is not exhaustive and that the decision is based on a number of factors besides type of data, for example the amount of data and the complexity of the dataset. If the risk category for a data sharing request cannot be determined accurately based on the classification below, then the request will be treated as a 'high risk' request until such time that the classification is updated to cover that request.

Risk category	Data description
Low risk	Aggregate data or statistics. The data does not identify individual IP addresses, individual users (e.g. by login name) or individual connected institutions or network segments of the SURFnet network.
Medium risk	Data with anonymised IP addresses, either fully anonymous or with prefix-preserving anonymisation (this is a form of anonymisation that preserves the IP allocation structure of a network and allows researchers to identify network segments). DNS data captured at authoritative name servers operated by SURFnet. Logging data in which IP addresses have been anonymised or on which prefix-preserving anonymisation has been performed, and/or in which individual user login names have been anonymised.
High risk	Data containing original IP addresses on the SURFnet network. DNS data captured at recursive caching name servers ('resolvers') operated by SURFnet. Logging data in which original IP addresses or (partial) original user login names are present (where partial can, e.g., be the *realm* in which a user resides).

Table 1: Risk classification

N.B.: When a data sharing request is composed of data from multiple sources, then the entire request will be treated according to the highest data risk classification for the individual sources that comprise the request. If it is likely that combining shared data – either by combining sources inside the request or combining the data supplied by SURFnet with data from external sources – will lead to a higher risk, then SURFnet will treat the entire request according to this higher risk category. Finally, SURFnet is solely responsible for determining the risk category and may decide at its discretion to increase the risk category of a request. The power of final decision on the risk category rests with SURFnet's privacy officer.

Appendix B Ethics Review

An ethical review will take place:

- when the request for data is labeled high risk;
- when a cross border transfer is being considered;
- or when there is uncertainty about a medium risk request.

There may be other instances in which an ethical review will take place (e.g. when the question of re-labeling standard research arises) and it is up to the discretion of the SURFnet employee who has received the data sharing request or another employee at SURFnet to place a formal request for an ethics review.

The ethical review process is not meant as a legal review, this will be done by the legal team and not the ethics review panel.

The ethics review process will be a meeting in which a small number of individuals (herein referred to as the Board) will gather to discuss the ethical issues related to a data sharing request. The purpose of the meeting will be to address the concerns raised by the data risk classification and to understand if, and how, the concerns may be mitigated. The overall goal is to understand whether or not the data should be shared with the researcher and under what conditions.

The board will consist of:

- one SURFnet employee that received the data sharing request (*voting member*);
- one independent SURFnet employee (*voting member*);
- one non-SURFnet member with subject matter expertise (*voting member*);
- one ethics adviser for SURFnet (*non-voting member*).

The ethics review will be moderated by an ethicist – the ethics adviser for SURFnet – who is responsible for:

- initiating and structuring the discussion as it relates to the ethical issues raised in the initial assessment;
- raising ethical concerns not already mentioned in the initial assessment;
- moderating the discussion so as not to get too off topic;
- ensuring each party acts respectfully and appropriately to one another.

While not mandatory, it is recommended that the researcher, if possible, is present at the meeting, to respond to questions from the board that can help clarify the data sharing request.

At the end of the meeting, each member will express their vote: approve sharing request, disapprove sharing request, or abstain from voting. Only the ethics adviser, as moderator, will not have voting rights. A majority vote will determine whether or not the sharing request is accepted.

In case no final decision can be made, a follow-up meeting will be scheduled with an additional SURFnet employee and/or an additional non-SURFnet employee. This is to be decided based on the nature of the request and the nature of the disagreement in ethical acceptability (e.g. it may be necessary to have a legal scholar present or an expert in specific cybersecurity issues related to the sharing request etc.).

At the end of each meeting the ethics adviser will write a report of the discussion as well as the outcome of the meeting (i.e. the votes). This report will be sent to all members to agree upon the contents.

The report will be stored for future reference and is freely available upon request.

If anyone present at the meeting wishes to express their concern or disapproval with the decision (and the reasons for said disapproval) they are free to do so in the (email) discussion concerning the report from the ethics adviser. Expressing concerns or disapproval in email form does not open the doors for the other members to provide feedback or comments on the disapproval or reasons for disapproval; it is for the records only.

In the instance that the moderator does not agree with the decision made by the three voting members of the Board and wishes to express this they are free to do so in written form for the records.

The records of the decisions made by the Board will be kept and used in future decision making of cases with similar variables. The decisions may also be used to update the sharing policies and mission of SURFnet.

Appendix C Writing an Assessment on the Ethical Implications of Research

C.1 What is it?

The Ethical Implication Addition (hereafter referred to as the EIA) is a short portion of the initial assessment in which the researcher requesting data from SURFnet must identify potential ethical issues related to the research they plan on conducting with the data they have requested from SURFnet. When ethical issues have been identified the researcher must also indicate why and how there is no other way in which the research can be conducted to mitigate said concern. Alternatively, the researcher must indicate what steps have been taken to mitigate or minimise the ethical issue raised. This portion of the assessment ensures that researchers have given due consideration to the impact their work will have on the data set in question as well as the impact said research might have on the field of cybersecurity, data analytics and so on.

C.2 Scope of the EIA

Researchers are requested to consider the potential ethical implications of their research in terms of both research ethics (i.e. how the data is collected, stored, analyzed, handled, disseminated, and so on) as well as the larger societal impact (i.e. the societal issues that the research aims to address). Researchers must show that they have considered both of the categories of issues listed above (research impact vs societal impact) and must include a statement concerning each.

C.3 Sections to be included

1. Potential or possible ethical considerations concerning the collection, analysis, use, dissemination or sharing of the data (i.e. issues in research ethics).
2. Ways in which the concerns raised in (1) will be mitigated or minimised.

3. Potential or possible ethical considerations concerning the larger issue the research aims to address.
4. How the proposed research aims to contribute to the larger issue raised in (3).
5. Whether or not ethics approval from another ethics committee will be sought (i.e. if the researcher works at an institute where an ethics committee or Institutional Review Board (IRB) exists they may also be obliged to send the request to that committee or IRB).

C.4 Length

The EIA need not be more than half a page but must be at least a full paragraph.

C.5 Further reading

For more information on the role of ethics in computer science research, and for more information about writing ethics paragraphs, we suggest reading the Menlo Report [1] and the companion to the Menlo Report [2]. Both documents can be found on the website of CAIDA:

`http://www.caida.org/publications/papers/2012/menlo_report_actual_formatted/`

References

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

[2] D. Dittrich, M. Bailey, and E. Kenneally. Applying Ethical Principles to Information and Communication Technology Research: A Companion to the Menlo Report. Technical report, U.S. Department of Homeland Security, Oct 2013.