http://networkedsystemsethics.net/

Networked Systems Ethics - Guidelines

These guidelines aim to underpin a **meaningful cross-disciplinary conversation** between gatekeepers of ethics standards and researchers about the ethical and social impact of technical Internet research projects. The iterative reflexivity methodology guides stakeholders to identify and minimize risks and other burdens, which must be mitigated to the largest extent possible by adjusting the design of the project before data collection takes place. The **aim** is thus to improve the ethical considerations of individual projects, but also to streamline the proceedings of ethical discussions in Internet research generally.

The **primary audience** for these guidelines are technical researchers (e.g. computer science, network engineering, as well as social science) and gatekeepers of ethics standards at institutions, academic journals, conferences, and funding agencies. It is certainly possible to use these guidelines beyond in academic research in civil society, product development, or otherwise, but these are not the primary audience. Some sections point the reader to other groups – such as the data subjects, lawyers, local peers, etc. – who can also use (parts of) the guidelines to help assess the impact of a project from their expertise or point of view.

Do not be intimidated by the length of this document; not everything is relevant for each project. Relevant questions will be decided on a case by case basis and should become apparent when using the document. The ethical reflexivity process exists in four parts. First, the context and impact of the project is established. Second, the ethical tensions and design causes are determined. Third, feasible alternatives to the scope, the technological, and methodological design are evaluated through an iterative process. Finally, the necessity of informed consent is assessed. The guidelines are designed to be a linear process, with points of iteration included. Not everything will be relevant for each project.

The guidelines have been created through a lengthy multidisciplinary and collaborative effort. Computer scientists, network engineers, lawyers, philosophers, and social scientists contributed valuable input in specifically organized workshops and meetings. We are thankful to all who contributed in this process! The framework has been inspired by discussions with participants, as well as the Association of Internet Researchers Ethics Guidelines, the Menlo Report, general emerging technology ethics methodology, Constructive Technology Assessment, Value-Sensitive Design (pdf), and Ethical IT Innovation.

We suggest to include these guidelines in a Call for Papers or institutional review processes, whereby the ethical tensions of a project can be resolved collaboratively through conversation **before** submission, rather than when the project has already been completed and risks rejection on ethics grounds.

This work was edited by Ben Zevenbergen at the Oxford Internet Institute as part of an Open Technology Fund Fellowship, with the help, input, support, and guidance of many, many amazing people. To edit or suggest changes, please first create an account and email bendert.zevenbergen [at] oii.ox.ac.uk with the changes you suggest. Citation suggestion: *Zevenbergen, B., et al. (2016) "Networked Systems Ethics". http://www.networkedsystemsethics.net.*

Summary questions (TL;DR)

These fundamental research ethics questions lead you to more elaborate explanations in these guidelines, as well as additional questions to consider. Ideally, the engineer and ethics boards should follow the Iterative Reflexivity Methodology, but these initial questions are a good start to identify where more attention is needed. These questions may also serve as a starting point for ethics committees of a department, journal, or conference, for their internal considerations:

A. Context: How would you describe the context within which data is collected, information flows are created (or affected), or phenomena are measured?
   - Aims: What are the aim and benefits of the project?
   - Benefits: Why are the benefits good for stakeholders?
   - Purpose limitation: Can the scope of data collection be limited whilst still achieving the project aim?
   - Politics and Power: Are particular stakeholders empowered or disempowered as a result of this project?
   - Risk of Harm: Could the collection of the data in this study be reasonably expected to cause tangible harm to any person's well-being?
   - Law: Which bodies of law are likely to be applicable to the operation of the project?
   - Values: Which values will the project conceivable impact?
   - Burdens: Who carries the burden of harms or impacted values, and how?

B. Technology Ethics: Can the harms and impacted values be traced to parts of the technological design of the project?
   - Function Creep: Does the project potentially set a precedent for unethical methodologies that could be misused by others in the future?
   - Data Governance: Using current techniques, can the data used in this study reveal private or confidential information about individuals? If so, discuss measures taken to keep the data protected from inappropriate disclosure or misuse.
   - Data Retention: When will the collected data be deleted?

C. Evaluate Alternatives: Have you considered measures to mitigate the identified risk of harm or impacted values?
   - Tech Alternatives: Can alternative technologies be employed or devised to mitigate some issues?
   - Scope: Can you limit the scope of the project (geography, knowledge generated, etc.)?
   - Methodology: Have others used alternative methodologies to achieve similar ends?

D. Justify Design Choices: state clearly the design choices related to ethical aspects as well as remaining ethical tensions
   - Informed Consent? Do you need to rely on informed consent from participants and stakeholders?

Reasons for Ethics Guidelines about Networked Systems

The Internet is a highly complex and pervasive information environment that mediates an increasing amount of human activities in areas such as social life, commerce, government, finance, and culture. To understand and make sense of the Internet architecture, engineers and researchers collect data, measure information controls, and manipulate environments to carry out experiments. While such measurements could be considered a largely technical exercise many years ago, these activities now increasingly affect Internet user's lives and thus require ethical scrutiny.

Scientific and technological advancements transform society in a variety of ways. They are designed by scientists and engineers who make decisions about how their technologies work. In order to contribute to the advancement of social good, engineers should consider the social and ethical impact of their work. The socially engaged engineer brings a normative vision for how their disciplines' work should or will impact society.

A discrepancy exists, though, between human subject research – where there are relatively strict and widely-understood ethical traditions – and Internet engineering where the consideration of these issues is relatively new. Existing best practices from other fields don't easily translate to a the digital environment where projects can have a global reach, making consequences of actions difficult to comprehend and assess. Further, the logic of digital information on networked systems differs substantially from the logic of analogue information, of which the collection and distribution are far more limited in time and space.

It may be difficult or even impossible to determine ahead of time whether a particular technological advancement will be used for good or bad. This is no reason to ignore the value of thinking about the obvious uses of a particular technology, who could use them for which end, which values they promote, and which they suppress. The aim of these guidelines is to inform actors in the scientific process, stakeholders, as well as engineers and developers alike about the ethical impact of this new technology and how to bridge social requirements and technical implementations.

One of the tasks of technology ethics is to analyse the impact of a project, and guide decision-making processes to reduce harms and maximise benefits. This document presents a set of guidelines for engineers and researchers that inform an ethical assessment of research projects. The aim is not to prescribe a right way of dealing with the human element in Internet research, but to offer questions that will guide the researcher through the thought process. The answers to questions create a base for conversation with ethics boards, experts from other disciplines, or stakeholders in target communities.

Internet Research Ethics

Several formal frameworks for scientific ethics exist to protect human subjects from potential risks of harm in a variety of disciplines. For example, the US Belmont Report concerns scientific and medical research involving human subjects, and established "respect for persons," "beneficence" and "justice" as the guiding principles of research ethics. The Common Rule is the federal regulation that tasks Institutional Review Boards (IRBs) with reviewing research to ensure it meets those principles. The Menlo Report builds on the Belmont Report and translates the scientific ethics research principles into the computer science and network engineering domain. The Association of Internet Researchers (AoIR) also maintains a useful set of guidelines for Internet researchers broadly.

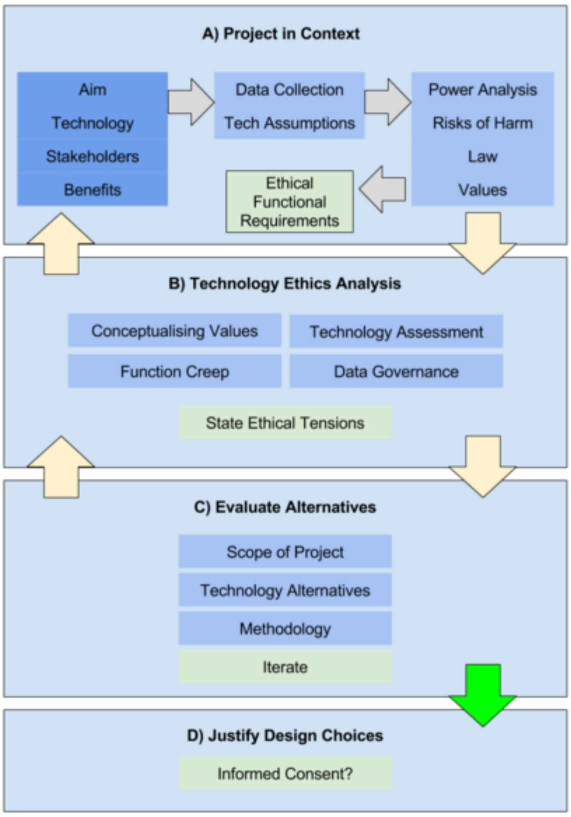
Despite these efforts, ethical dilemmas continue to arise in technical Internet research projects, whereby actors in the scientific process base their decisions on conflicting reasoning. The mechanism for ethics in Internet research projects appears to be somewhat broken and guidance is needed about how to move forward. These guidelines do not aim to fill the gap of Internet research ethics broadly, but aim to assist in creating a research design based on integrity and knowledge of the social impact of projects.

Iterative Reflexivity Methodology

These guidelines are developed for a range of technologies and methodologies, which are in a state of rapid innovation. Due to the worldwide or unexpected scalability of these projects, the values affected can vary widely. Therefore, these guidelines suggest to consult experts from other disciplines for a social assessment of the impact of the chosen technologies on the jurisdictional, political, and cultural context of stakeholders. Technical researchers may need to consult appropriate lawyers, ethicists, social scientists, peers in target regions, or local experts at different moments throughout this iterative process.

The four sections of these guidelines provide fundamental considerations for a research design. The questions in each section should initially be answered by the researcher or engineer, but later serve as a base for conversation between actors in the scientific process. While not every single question needs to be answered for every project, they should provoke technical researchers to think about their technology choices, the impact of newly created information flows, or the political implications of the knowledge generated in a potentially little-understood social and political context of data subjects. The answers will allow less non-technical actors – who may not possess the same level of knowledge and understanding as the researcher or engineer – to engage in an iterative conversation that mitigates ethical issues during the the project design.

The suggested iterative approach is shown in this flowchart



In **Section A**, technical researchers set the scene of the project by explaining the aim, benefit, methodology, and technology of a project, as well as listing the risks of harm and affected values. This allows a third party who will scrutinise the project to have a good understanding of the facts and context of the project. Through cross-disciplinary discussion of these facts, the actors can set the **ethical functional design requirements**. **Section B** encourages the researchers or engineers to pinpoint the causes of specific harms in the research design and to analyse them. **Section C** then enables actors to creatively assess and reflect on technological and organisational alternatives to the initial project design. Alternatives should be assessed through the process outlined in Section B, which in turn affects the facts in Section A - the scope of the project.

Once actors agree on the final project design, the remaining ethical tensions should be stated clearly. While informed consent may not be necessary when the risks have been minimized, actors should still consider how risks of harm could be conveyed to potential data subjects in **Section D**. This approach will allow all actors and stakeholders to understand which steps have been taken to mitigate ethical concerns to the furthest extent, while still maintaining the purpose of the project.

Incentives for Actors

Applying the Iterative Reflexivity Methodology outlined in these guidelines will lead to several benefits for researchers, ethics boards, conference program committees, editorial boards of academic journals, and the larger research community. The input from various **stakeholders** and experts will incentivise researchers to create ethically sound research projects that will have a lower probability of rejection on ethical grounds in publication venues. Further, **researchers** will be able to demonstrate how risks and harms have been meaningfully mitigated to the furthest extent possible, which can help mitigate personal or institutional liability claims if harm is caused to stakeholders. **Programme committees** of conferences and **editorial boards** of journals will be able to meaningfully articulate their ethical requirements and engage with researchers to ensure that valuable publications can be accepted. The larger networked systems research community will benefit from having a clear framework for ethical analysis and articulating accepted solutions to ethical dilemmas more clearly.

The guidelines in this document have been developed to protect the interests of both researchers and data subjects. They are based on existing examples of best practice, wide consultation with researchers, and a number of interdisciplinary workshops around the world. Their use will contribute to public trust in Internet research, which is essential for the future development of the field. They will also help researchers and engineers demonstrate they have taken reasonable and appropriate steps to ensure that ethical dilemmas have been mitigated to the furthest extent possible.

How to use these Guidelines

It has been proven to be very difficult - if not impossible - to quantify the risks of harm when vague terms such as privacy, trust, and autonomy are violated or affected. Therefore, these guidelines use a more qualitative approach, whereby factors are considered on a scale of high, medium, and low. This applies mainly to the probability and magnitude of harm and the scope of mitigation techniques. High risks should be counteracted with strong mitigation techniques, whereas low risks may require less stringent measures. This scale merely functions as a guide for cross-disciplinary conversations about the project and should not be considered as a robust method of assessment.

The answers to the questions in this document should be discussed with the gatekeepers of ethics standards. Researchers should refer to this website, or send along a pdf version of the version they used. This will clarify the approach taken by researchers and allow a meaningful conversation based on the Iterative Reflexivity Methodology.

Terminology

Project

The work under scrutiny is referred to as a ‘project’. The project design should be considered as a plan of the various components of the project and how they will operate to achieve a desired result. The focus of the guidelines are technical Internet research projects, but this document could also be used for public, commercial, or activist projects.

Stakeholders and Actors

Several types of stakeholders exist in an Internet research project. Below we classify them briefly:

Actors in the scientific process Actors in the scientific process are the researchers, their consortiums, their departments or institutions, relevant ethical review committees or institutional review boards, conference programme committees, editorial boards of journals, etc.

Direct Stakeholders Individuals, parties, or organisation who interact directly with the project, for example whose data is collected, whose systems are measured, or whose devices are used to carry out the project.

Indirect Stakeholders Indirect stakeholders refer to all other parties who are affected by the operation of the project.

These guidelines refer to direct and indirect stakeholder simply as 'stakeholders'.

Representatives In some research setting it is not feasible nor expected that a research team contact every data subject with an informed consent sheet. However, it is important that a suitable representative is identified and contacted. These could be NGOs that represent Internet users rights in a particular country, or about a particular issue.

When discussing a project design with a representative or stakeholder, it is important to keep in mind their political (or other) interests. The the researchers work, they may be able to gain a competitive advantage due to a shift in power. The recommendations given by representatives or stakeholders should be analysed with the section on Power Shifts and Political Analysis in mind. An advice from a single actor should not be considered sufficient and triangulated with the views of others actors.

Section A: Project in Context

The first step of the iterative reflexivity process encourages the researcher or engineer to create an understanding of the technology, its operation, and social impact based on relevant contextual factors. It is important to scrutinize the social, political, and economic context within which a technology operates, as much as the technology itself (for background reading, see pdf) This section unveils the effect of a project design, such as the benefits and risks of harms, as well as power shifts, and political implications. The social context(s) in which the project will perceivably operate is the **unit of analysis**.

Further, this section creates an awareness of the engineering assumptions embedded in the technology design, as well as the resulting effects of the technology that underpin the analyses in further sections of these guidelines. It is important to keep in mind that technologies are hardly ever neutral: design choices are a human process and thus typically contain assumptions and political elements. Even the most neutral technologies will still cause a shift of power or affect entrenched values to some extent in society, which can be unknown to the technological researcher initially (see the Module on Sociotechnical Systems and Embedded Values).

Aim, Benefits, Technology & Stakeholders

The aim, purpose, and perceived benefits of a project need to be clearly stated as a crucial first step of an ethical analysis. The answers to this section will give reviewers an understanding of the importance of the project and thus give weight to technical and organisational decisions taken in further sections of these guidelines. When drafting the answer to these questions, it is important to think through the lens of the following actors who may be affected by the project:

- Personal benefit of the **researcher**, their team, their department, their consortium, and their institution(s). These include scholarly benefits, as well as responses to other incentives, such as career advancements.
- Benefits for **stakeholders** and data subjects (ie. the persons directly or indirectly involved in the research, whose data is collected or whose living environment potentially affected).
- The benefits to broader **society** due to the knowledge generated, as well as the direct and indirect effects of the project. “Society” is a broad term which includes (but is not limited to) the particular research field, politics and the economy where the researcher is based, politics and the economy where the research is carried out, and society at large.

1. In short, what is the main aim of the project and what will an answer to the research question uncover?
2. How will this research contribute to the state of the art in understanding Internet phenomena?
3. Who are the actors and stakeholders involved directly or indirectly in this project?
4. What are the relevant existing information flows?
5. Explain how the project (a) creates new information flows, (b) measures existing information controls, and/or (c) creates new databases that previously did not exist in the context within which the project operates (a UML sequence diagram is highly recommended).
6. How will the research benefit society and specific stakeholders?
    - What knowledge are you basing your answer on?
    - Can you verify your answer?
    - Did stakeholders request the experiment or knowledge?
7. If the project is (partly) carried out in a different country, how do the local stakeholders perceive the benefits of the project?
8. Will the collected data or inferred knowledge be directly relevant to and applicable in some specific government, business or academic processes?
9. How can the knowledge generated support future research?

Scope of Data Collection

The principles of *purpose limitation* and *data minimization* suggests that the amount and types of data collected should be relevant and not excessive for the project purpose. The data categories should support the aim of the project, such as answering a stated research question. This is not only a legal obligation in many countries, but also minimises the risks of liability for researchers. Reducing the size of the database also simplifies considerations for data management and project governance (see Data Governance).

To answer these questions, the researcher should create a database template that explains what the various data fields represent.

1. Is it necessary to collect new data, or do existing datasets contain the necessary information? (If so, see the Module on Easy Data Sources)
2. Can the same data be collected in a confined test setting, or must the data be collected on operational networks?
3. If new data will be collected, are the identified datafields relevant and not excessive in relation to the research purposes (i.e. strictly necessary)?
4. To what extent will data in the database identify individuals directly, or indirectly through inference? (**Consider consulting a re-identification expert**)
5. Is it feasible to take into consideration auxiliary datasets which can identify individuals when combined with the data collected for this project?
6. If the raw data is disclosed, does the database reveal any sensitive information about the data subjects? (**Consider local political conditions, cultural specificities. Consider consulting local peers or regional experts for a review of your assumptions**)

Technology Assumptions

Technology is very rarely ethically or politically neutral. It is not uncommon among engineers to claim otherwise and analogies with other technologies are sought to support this position (e.g. knives, because they "can be used to cut vegetables or kill people"). While there is some truth to the idea that the ethics of a technology depends in part on the ends for which it is used, Internet engineers have the power to make decisions and to limit the harmful consequences of the new information flows, databases, or knowledge that are generated by them.

A designer typically has several options during the design phase of a project and makes choices about how the finalised artefact should work. In fact, the inspiration or necessity of a new project or technology usually stems from the idea that something is missing, or something should be changed. Design choices or the choice to use a particular technology component will have an effect on the material world, whereby established powers shift or are amplified, values are impacted, or existing (social) problems are solved, whereby unexpected consequences may create entirely new problems.

Some questions are important to transcend a potentially siloed tunnel vision, or a researcher's overestimation of their ability to predict the social impact of the particular technologies chosen for a project. It is important to take an objective view when answering these questions, instead of making self-serving arguments. For this section, the project and its technological components are the **unit of analysis**:

1. Describe succinctly what the technical functionality is that has been developed or deployed for this project.
2. Fundamentally, what will be the impact of the project's direct and indirect benefits on stakeholders? Why are these benefits good?
3. If you put yourself in their shoes and context, would the the data subjects, actors, or stakeholders perceivably have objections to the (1) technological choices, (2) the potential harms, or (3) the new knowledge created?
    - Are their interest being served or undermined?
    - Is conceivable that these stakeholders could argue that the engineer is using their technological superiority to impose their own ethics on a particular context?
4. Is it likely that the project sets a standard or precedent, to be followed by others with less benign intentions?

Power Shifts and Political Analysis

Technologies embody political and ethical choices by the designers. Designers of a project have a position of significant power to create new databases, information flows, and knowledge that did not exist before. However, the impact of the design choices will differ in the context within which the technology is used or operated, because social, political, cultural, and economic forces in the target location or community will influence how the technology and information may be put to (unexpected) use. Those who can potentially access the project's technologies or information can therefore gain power relative to others within the context.

Sources of power differ per country - or regions within countries - based on political factors such as its institutions, the power of the executive, the independence of its judiciary, the extent to which civil liberties are respected and protected (e.g. freedom of speech, privacy, the freedom of assembly), and several others. The governance of information is an important part of maintaining and distributing power. In addition, revealed information and databases may cause different actions in various contexts.

A project that introduces new information flows, measures existing information controls, or creates new databases has the potential to disrupt existing (political) relationships, which in some cases may be fragile. Socio-political contexts of actors and stakeholders are prone to sudden or more gradual shifts that cannot be predicted precisely. For example, a war, revolution or a coup in country is difficult to predict years in advance, but equally important to consider are gradual shifts in the attitudes of governments towards data collection, privacy and data protection. Governments are already showing tendencies to use available information to make decisions about citizens in several new areas that were unheard of just a few years ago.

The potential consequences for stakeholders in a target location or community should not be neglected, and the legal and social context of participants cannot always be assumed to be comparable to those of their peers in the Western world. Stakeholders can be exposed to risks of harm when the political and social contexts are poorly understood, especially when gaining informed consent is problematic. Similarly, passive measurement techniques that do not directly involve end users can implicate entirely uninvolved users of studied networks, who may face risks that are poorly understood without knowledge of the particular political context.

The researcher should therefore conduct an assessment of the informational and political economy of the target location's context in order to meaningfully evaluate the potential benefits and risks of harm. It is recommended that researchers consult experts or peers of the particular country. While it will be difficult to gain a full picture of the political effects of a project from a cross-disciplinary conversation, **researchers can create a database with a small sample of the data that will be collected and present this during a discussion with a country expert, in order to demonstrate the type of data that will be collected.**

The aim of this section is twofold: First, we establish the roles of stakeholders and actors in the contexts within which the technologies will operate. Second, we assess to which their respective positions of power, especially informational power, shift due to the introduction of a new technology. The answers to these questions should establish the social, cultural, and economic framework from which risks of harm can be identified. **The researcher should consider addressing these questions with the help of a local peer or target country expert.** Further, consider consulting reports such as the Freedom House's "Freedom on the Net score", or the Economist's "Democracy Index." Where possible, the researchers should refer to risks of harm as either low, medium, or high.

Roles

1. Who is likely to control or have access to the project's technologies, databases, or information flows? What are their roles in the target society?
2. Who is likely to be empowered or disempowered by the technology? And how?
3. Are other (indirect) stakeholders disempowered in any way as a result of the project? If so, how?
4. Whose interests are promoted by the project? Whose interests are compromised?
5. Will the project magnify the existing powers that were already in place, or does the project create new powers in the context?

Impact of benefits

1. How do the benefits identified in Section A materialise for the identified actors and stakeholders as a direct or indirect result of the project?
2. Are the benefits and potential empowerment a result of the new information flows, the data generated, or the knowledge inferred?

Political impact

1. Will the new information flows measure or alter an existing politically sensitive control or management of information?
2. Does the knowledge generated by the project challenge or solidify the power or reputation of an authority?
3. What is the conceivable political impact of the information that will be inferred from the new information flows and collected data? Will the behaviour or actions of particular actors or stakeholders be uncovered, and would this be considered to be politically sensitive in the country of experimentation?
4. Will the project upset or influence power relations between actors and stakeholders in the context, or in other areas (such as national or international politics)?

Individual stakeholder impact

1. Does the created database reveal anything that may be considered (politically) sensitive in the country of operation?
2. If the new database or information flow is scrutinised by a powerful party in the country, what could conceivably be the impact on the rights and freedoms of individuals?
3. If plausible deniability would be an acceptable legal defence in the country of the researcher, to what extent can a stakeholder in the experiment country realistically rely on this defence?
4. What impact may this research have on uninformed stakeholders?

Extreme contexts: Do No Harm Principle Some contexts are extreme to the extent that potential harm posed to individuals by data insecurity and personal identification can likely include arrest, torture, death and longstanding discrimination. Instead of harm mitigation, as outlined in these guidelines, a principle of do no harm is more appropriate, whereby data should be de-identified to the extent that it realistically cannot be re-identified, regardless of the data utility. If this is not feasible, the data should not be collected at all.

Risks of Harm

Researchers must identify the likely risks and harms to participants, as well as their probability and likelihood of occurrence. Before risks and harms can be identified, however, they must first be defined. Due to the complex, dynamic, and innovative nature of the Internet, it is difficult to define the harms concretely, or to set standards that stand – even a short – test of time. What may be considered harmless today, may become a much larger threat in future when (currently) protected databases are fused, computing power increases to enable new inferences, or new data collection methods enable new ways of (re-)identification. However, for an ethical analysis it is crucial to be aware of the risks of harm that result from the implementation of a project.

A comprehensive and reliable quantitative method to assess risks does not yet exist. Therefore, the researcher must assess the balance of risks and benefits based on reasonableness, which can best be achieved through the iterative reflexivity approach. This test should be a cautious one, taking into account future technical developments, auxiliary data and unexpected changes in a political landscape, amongst several other factors that could affect the use of sensitive information contained in mobile connectivity datasets.

Risks need to be assessed in light of any likely adversary who could be motivated to use the new information flows, databases, or knowledge to their advantage, and the broader context in which resulting information could be used (see the Module on Adversaries). For example, the researcher must consider what the collected data, if re-identified, tells the adversary about the data subject. Some information may be fairly benign, whereas other contexts could be sensitive when interpreted by a specific adversary.

The overall risk level of the research project should be adjusted based on the expected capacity, motivation, skill, time, and available auxiliary information the adversaries are likely to possess. In addition to this consideration, the researcher should give due deliberation to possible future adversaries, to the extent possible. These parameters are necessary to determine suitable mitigation strategies. The researcher may reuse risk assessment profiles from previous, comparable research designs.

1. Who is likely to be interested to use the information created by this project to their advantage and for which reasons? (See the Module on Adversaries for a general classification of adversaries and assess their motivation and subsequent level of risk).
2. What capacity (in terms of time, skill, computing power, etc.) do any identified adversaries likely have to re-identify datasets?
3. What type of attack could cause harm? What is it's likelihood of occurrence?
4. What are the roles, relationships and power structures of the stakeholders (data subject, likely adversary, and other beneficiaries)? What is the political context? Does this change the sensitivity of the revealed activities?
5. What activities would the information flows reveal about individuals, if identified?
6. To what extent would the use of the information harm individual data subjects, or specific stakeholder groups? Could the type of information be considered a higher risk (e.g. financial and medical information, even if indirect), or rather a lower risk (only IP-address)?
    1. Will the magnitude or likelihood of these harm likely increase over a longer period of time than the life of the project (consider political volatility, or the development of data subjects? See also Data Retention and a blog on "Speculating about technology in ethics".
7. Which known auxiliary information could the adversary use to re-identify data subjects, and would the available information increase the potential harm? How sensitive is the known auxiliary data that can be combined with the research dataset to re-identify data subjects? Is it reasonable to assume that more linkable auxiliary information exists?
8. Are there any meaningful statutory privacy protections in the jurisdiction of the data subject that offer extra protection for the data subject?
9. To what extent can the researcher predict reputational harm to stakeholders, such as organizations or government agencies? Can the researcher justify this harm?

The Law

Applicable laws should generally be considered as legitimate and must thus be respected by all operations of the project. Laws can and do differ widely between jurisdictions, so it is not sufficient to only abide by the local laws of a particular research team. However, enforced laws may not always be codified into law or could be subject to interpretation by political officials. It may thus be difficult to assess which laws will be applicable should a legal challenge arise (for example, hosting measurement nodes in a foreign network could be interpreted to constitute an act of espionage on critical national infrastructure). Further, due to the fast pace of innovation of information technologies, existing laws may not - or inadequately - cover the actions carried out by a project, which may leave further room for (unpredictable) local interpretation.

A global survey of applicable law and policies for an Internet project would be a near impossible task for a legal researcher, let alone a team of computer scientists. Enumerating all possible (albeit remote) legal risks to actors and stakeholders is similarly infeasible. However, to gain some understanding of the legal risks, some issues could be discussed with a counsel or local lawyer.

1. Which bodies of law are likely to be applicable to the technical operation of the project? (**Consider consulting a legal counsel**)
2. Are fundamental rights of actors or stakeholders impacted? (**Consider consulting a legal counsel, peers in target context, or digital rights NGOs**)
3. Will the project violate identified laws?

4. If the project will violate laws, can the project team justify this action? (**Consider consulting an ethicist**)
5. To what extent do countries within which the project will operate guarantee a fair trail (pdf)?
6. How have other project teams identified applicable laws in the given context?

Ethical Functional Requirements

The responses to questions in this section will have unveiled the benefits, affected values, political power shifts, and risks of harm. This part reflects on the answers of the questions and encourages the researcher to establish a list of affected values and harms, which will function as a set of functional requirements for the project in Section B. The aim is to create an understanding of the impact of the project - both positive and negative - on a high level. For each harm or affected value, state the probability and magnitude of risk to each on a scale of low, medium, or high. This exercise may be most useful for an ethics board to summarise the understanding created in the previous parts. For an overview of some values, see the Module on Values and Harms.

1. Can the researcher list the identified potential harms as a direct result of the project for each identified stakeholder group?
2. Can the researcher list the identified affected values as a direct result of the project for each identified stakeholder group?
3. Which harms or values are affected by power shifts caused by the project?

Section B: Technology Ethics Analysis

The previous section uncovered and collected relevant facts of the project and the environment in which it will operate. These facts form the **units of analysis** for the technology ethics analysis of this section. The next section encourages researchers to consider alternatives to ethical dilemmas that will be uncovered by technology in this section.

The ethical analysis of this section forms the central part of the iterative reflexivity approach. First, the affected values and harms are assessed in more detail. Second, the technological causes are pinpointed and linked to specific design choices. Third, options for project and data governance are explored. Finally, remaining ethical dilemmas are stated, which informs the next section.

Conceptualizing Values and Harms

Conceptual investigations are philosophically informed and thoughtful considerations of the issues by which stakeholders may be impacted as a result of the project design. Values and harms must be broken into their relevant constituent parts to understand where the pain lies. While these guidelines cannot realistically demonstrate the parameters of every existing value, we offer a guide to think about values in the Module on Values and Harms. Still, it will be a difficult tasks for technical research teams to adequately assess values and harms comprehensively. **Therefore, we recommend researchers discuss the identified values with experts from the humanities to identify the relevant constituent parts**. For an overview of how values can be defined and identified during the design process, see the paper "Value sensitive design and information systems" (pdf) by Friedman, B., Kahn Jr, P. H., Borning, A., and Huldtgren, A.

1. Which definitions or explanations will be used to assess a value? Can you cite sources?
2. Can the researcher pinpoint the constituent part of the value that will be affected by the project?
3. Is the risk of harm high, medium, or low?
4. If materialised, is the impact of the harm to different stakeholders high, medium, or low?
5. Which stakeholder carries the burden? And who carries the burden of proof?

Technology Assessment

Specific properties of technologies can affect values or cause harm. Some properties may cause a particular harm intrinsically due to their nature, such as privacy violations caused by online trackers. Extrinsic harms occur when the technology is put to a particular use, or a harm becomes amplified. For example, switching on the camera of a laptop without the user's knowledge is a potential extrinsic privacy harm of having a camera above a laptop's screen. For this section, the technological cause for identified harms and values that are affected will be sought.

1. To what extent does scientific knowledge exist about the intrinsic harms of a technology? To what extent are these harms materialising?
2. Can you pinpoint the technological causes of harms? Would these harms hypothetically not occur if this component would not be used?

Unethical Methods and Precedent Setting

Internet researchers and engineers occasionally apply methodologies that may be considered unethical, possibly unlawful. For example, some well-intentioned projects intentionally harm information systems, participate in fraud, circumvent a robots.txt file, imitate humans, intrude into systems, infiltrate botnets, collect WiFi signals from mobile devices, exploit zero-day vulnerabilities without giving due notice, etc. While these practices may be legitimate in some cases, they do warrant strong scrutiny.

Even if the research aims are beneficial for a wide group of stakeholders, research methods need to be published and thus will set a standard or precedent and initiate function creep and unintended consequences. For example, publishing the methodology of how to conduct surveillance on mobile phones raises the likelihood of others learning and applying these methods, possibly for malevolent purposes ranging from more commercial applications to widespread and unheralded surveillance. Questionable methods for benign goals can be misused for destructive uses. It is therefore important that researchers engage actively with the fact that their methods may be misused, and find ways to mitigate risks and harms.

It is ultimately the responsibility of individual researchers – in dialogue with ethics boards and other stakeholders in a specific project - to agree on the limitations based on a thorough understanding of the project. It is important to consider the precedent and standard setting effects of condoning ethically improper research when weighing risks and benefits.

1. Do ethical standards exist elsewhere for the proposed methodology or technology? (See for example the standards set in Resonsible Disclosure of security vulnerabilities).
2. What are the researcher's incentives to develop and deploy potentially unethical methodologies?
    1. To what extent is the researcher showcasing a new and sophisticated technical possibility, and to what extent are you using the technology to create knowledge of Internet phenomena?
3. To what extent could stakeholders conceivably perceive the researcher to be acting as a vigilate?
4. Would the methodologies, actions, and resulting knowledge be more suitable in another context than research?
    1. For example, would the project be more suitable for an activist group, law enforcement, or intelligence agencies?
    2. If not, can the researcher articulate the reason why the project is relevant for science and research?
5. Which actors will likely be interested to use the methodology for malevolent purposes, and how?
    1. Is it possible to contain the potential malevolent future uses by design?

Data Governance

The data management processes can also influence whether the risks of harm are high, medium, or low. For example, the chosen dissemination method of the research data and generated knowledge affects to what extent sensitive information is voluntarily revealed by the research team. It is therefore important to take into account the method of dissemination into an ethical assessment, and to let it guide possible methods of de-identification or anonymization.

Dissemination

Several methods of dissemination exist:

1. Open Data - No restrictions on dissemination - Higher risk;
2. Restricted data sharing - Legally enforceable restrictions - Medium/higher risk;
3. Managed access - limiting access to data - Lower risk;
4. Interactive methods - dissemination of statistical information about dataset - Lower risk.

An open data format disclosure means that datasets need to be de-identified as much as possible, thereby losing much utility, we suggest some other types of disclosure that the researcher may want to consider (see the Module on Dissemination).

1. Will the research dataset be shared with specified individuals (lower risk), a wider research consortium (medium risk), or be released publicly in open data format (higher risk), possibly via a data repository?
2. Will the disclosed data be limited to fulfil certain specified tasks (e.g. developing anti-spam lists, lower risk)?
    - If the answer to the previous question is positive: Will the further use be left to the receiving party to decide (higher risk), or will the researcher discuss the needed functionality with the recipient and tailor the data for this use (medium risk)?
3. If the researcher chooses a data sharing approach, which legal restrictions will be included in the data-sharing agreement?
4. How will the researcher enforce compliance by the receiving party?

5. If an interactive method is chosen, does the researcher disseminate only general statistics about the data (lower risk), truncated data (medium risk) or is more detailed information including identifiers shared (higher risk)?

De-identification

The researcher may apply multiple de-identification techniques and methods of dissemination to a single dataset (or a sample thereof), depending on a case by case basis on the assessed risk level of the recipient, amount of control exercised over the dataset and the sensitivity of the dataset.

The researcher should ensure that high risks of harm in the data fields of collected data types, the probability and magnitude of identified affected values and the risk-factor of the type of foreseen dissemination method are counterbalanced by the robustness of the de-identification technique used.

1. Is the chosen threshold of de-identification technique proportionate to the:

   - Sensitivity of the data?
   - Foreseen disclosure method?
   - Capacity of the identified adversary?
   - Has the researcher consulted a re-identification expert to discuss whether the foreseen data collection categories can lead to re-identification of individuals in the dataset? (Lower/medium risk)
   - Has the researcher successfully carried out experiments to test re-identifiability? (Lower risk)

Managing unforeseen risks

Systems and research designs will never be as robust as intended. To mitigate unforeseen risks, the researcher must be prepared and manage the unknown in the best way possible. When a dataset is disclosed unexpectedly, it is part of the ethical process to alert data subjects, so they too can take precautions. The researcher should give consideration on whom a burden falls when unforeseen risks materialize.

1. Is there a containment policy and what does it oblige the researcher to do?
   - How will other stakeholders be burdened?
2. Will the researcher contact the data subjects and/or the relevant privacy regulator directly about a breach?
   - To what extent does this depend on the seriousness of the disclosure or the sensitivity of the data?
3. Has a contact person been designated in the case of a data breach?
4. How will harmed data subjects or stakeholders be compensated?

Security

Sensitive data must be protected through good information security practice, such as physical and personnel security measures. As a structural requirement researchers should require from their employers that robust infrastructure is provided where data can be stored securely. Datasets containing personal information should be stored along with some sort of metadata that describes the dataset and its intended use, where feasible.

Systems should implement strong access controls, to ensure that only those who are authorized to do so access personal information. Access to the datasets should be logged, and logs regularly audited. Further audits should be carried out on topics such as data management, configuration control, intrusion detection, and incident response.

1. Is the collected data stored securely?
2. Are sensitive datasets encrypted?
3. Is the provided infrastructure upon which the project operates robust enough for the type of data that will be collected?

Data Retention Period

Projects should consider a length of time for which data is stored. For example, once the purpose for which data is completed (e.g. generating knowledge), it could be argued that data can be deleted. Of course, academic integrity prescribes that data should be kept for falsification, or further research. However, not all data will be useful for secondary purposes or longitudinal research, and some data may actually be harmful in the longer term. Researchers and ethics boards should consider to what extent it would be more beneficial to delete some data after a period of time, or upon completion of the project.

1. Is it necessary to keep the collected data upon completion of the project?
   1. If so, why?
   2. Can the research justify whether it is useful to store the data for longer than the life of the project?
2. How can the researcher ensure that the deletion of the data is thoroughly managed?

State Residual Ethical Tensions

An ethical tension arises when the researcher or engineer has several conflicting ethical incentives to address an issue, but doing both actions is not possible in the context of the project. For example, an a project can uncover useful data for future policy-making about a particular global network phenomenon, but collecting this data would necessitate to collect data of people's web browsing without their consent, which would constitute a grave violation of the unsuspecting data subject's privacy.

The identified harms and affected values can typically not all be mitigated in a project design. In the first iteration of these guidelines, several harms and values will have been identified, and their technological causes pinpointed. The technologies are chosen based on a perceived benefit, which may even be key to the project's aim. However, by now, the researcher and other parties are aware of the issues caused by the technological choices.

In the current iterative cycle of analysis, the research team should explicitly state the ethical tensions caused by the technological and methodological choices in the project. We suggest the researcher states the tensions in a short and concise form, so that a reviewer can understand the issue. For example, a famous dilemma in ethics is the trolley problem, which can be summarised as follows:

"A trolley is running out of control down a track. In its path are five people who have been tied to the track by a mad philosopher. Fortunately, you could flip a switch, which will lead the trolley down a different track to safety. Unfortunately, there is a single person tied to that track. Should you flip the switch or do nothing?"

1. Given the stated benefits in Aims, Benefits, Technology, and Stakeholders, the impacted values of the project and risks of harm in Conceptualizing Values and Harms, and the technological causes of these negative consequences in Technology Assessment, can the researcher state the residual ethical tensions?

Many ethical theories exist to address ethical dilemmas (See the Module on Ethical Lenses). However, these theories will likely give conflicting opinions of the right course of action. Therefore, the theories in the Module on Ethical Lenses should only be used as a lens through which to view and assess the ethical dilemmas, rather than a concrete answer to the dilemmas.

1. How would the researcher solve the ethical dilemmas using a common sense approach or their own intuition? Does the benefit of the research outweigh the potential harms, given the context in which the data is collected and research results are likely to be used?
2. What does the lens of consequentialism reveal about the right way to act?
3. What does deontology reveal about the right way to act?
4. Are any other ethical theories relevant to assess the right course of action?

Section C: Evaluate Alternatives

The initial project idea and design is not set in stone, but should be open to scrutiny by all actors in the scientific process, as well as stakeholders. Several ways exist to design a project that mitigates risks, but still achieves the original aim. This section balances the integrity of the project with the the overall scope and design choices that are available to a researcher. This section requires creativity from the researcher, and possibly a brainstorming session with experts from different fields, as well as stakeholders. John Rawls' thought experiment of the Original Position is a useful starting point for a deliberation, whereby participants take an objective view of the social and political that will be impacted by the project.

First, the scope of the project will be assessed. The research question and the field of deployment of operation will be assessed critically, and alternatives considered. Second, the assessment of technologies will be reviewed and other technical options explored. Third, the project's methodology is scrutinised, and the project governance process adapted.

Scope of Project

Internet experimentation projects can be scaled to a worldwide level. Engineers are typically incentivised to deploy a project as widely as possible to maximise the reach. It is sometimes also just easier to let a project operate without limitations and to see later which data is collected, rather than limiting its scope artificially.

However, the knowledge gained in Sections A and B will have exposed some problems in specific political and cultural contexts. Risk levels can vary widely based on target countries or for particular target groups. Therefore, trying to mitigate the risks and shifts in values in all areas will result in a race to appease the lowest common denominator (or: reduce the utility of the project to appease the context with the highest risk factors).

1. Can the researcher limit the scope of the research question and project aim to avoid some risks or affected values?
2. Can the researcher limit the scope of stakeholders, by excluding particular groups or countries? If so, would the data collected still be a representative sample to answer the research question?
3. Are any risks averted if the researcher limits the project duration to a shorter amount of operation time?

Technology Alternatives

Engineers can apply a wide range of information technologies or create their own. For each technological option, several alternatives exist that will yield similar results. It is also possible to combine existing and/or new code to devise sophisticated solutions. The researcher should **consult peer researchers or external experts** to list some alternative technology options for the research design, or discuss the feasibility of developing new code.

1. Which suitable technological alternatives exist for the technologies that cause a specific harm?
2. Can the researcher create new code, or adapt a technology to mitigate harms of affected values?
3. Do any practical cases or theoretical arguments exist regarding negative consequences of the alternate technological components used in the project?

Methodology and Project Governance

Similar to the technological alternatives, the methodology and the governance of the project can be reconsidered. Many alternative methodologies exist - some particularly sophisticated - that may require less data to be collected. Improvements to the governance of the data may also mitigate identified risks of harm or affecting values. The researcher should assess the options similarly to the questions in Technology Alternatives.

Iteration

The answers to this section will have uncovered some potential alternatives to the initial project design. At this point, the alternatives can be assessed through the steps outlined in Section A and Section B. The answers to the initial project design should be kept, and the new answers based on design alternatives added and identified as a new consideration. The outcomes can then be compared to assess the most suitable design strategy.

Through this process, the researcher should strive to reduce the residual ethical tensions stated in section on State Residual Ethical Tensions. A creative combination of all parts of this section (scope, technologies, methodology, and data governance) will likely lead to a more ethical (and better) project design. However, if the tensions are not mitigated and risks of harm remain, the project may have some inherent ethical flaws and the researcher should consider not pursuing this study at all.

1. If the researcher has adapted the scope of the project, then the project should be analysed from Section A to reassess the basic facts of the project.
2. If the scope of the project is maintained, but technological alternatives have been identified, the researcher should return to the sections on Technology Assumptions and Ethical Functional Requirements, as well as Section B.
3. If the methodology and project governance have been adapted, the researcher should consider whether the impacts are relevant for Section A and Section B, but focus on answering questions in Data Governance and State Residual Ethical Tensions.

Section D: Ethical justification and Informed consent

By this stage the researcher has identified and assessed the benefits and risks of the research design and made informed decisions about the operation of the project as a whole, as well as the collection, processing and dissemination of data. Through the iterative approach the ethical tensions will have been minimized, though some risks will likely remain. These risks should be communicated to stakeholders and a relevant ethics board.

The Module on Informed Consent outlines important considerations for whether the project requires informed consent from research participants. While it is not always feasible or desirable to gain informed consent, the procedure offers an opportunity to be transparent about the technologies and data collection.

If a waiver for informed consent is sought:

1. Can the researcher demonstrate that the project poses minimal risk to users and data subjects?
2. Can the researcher demonstrate how they applied the iterative reflexivity methodology of these guidelines to mitigate risks to the furthest extent possible?
3. Can the researcher demonstrate why it is impossible or impractical to carry out the project and address the research questions properly if prior consent of participants is required?
4. Can a data subject still object to the data collection? How will their wishes be respected?

If informed consent is sought:

1. Have stakeholders whose data will be collected **meaningfully** consented to their involvement with the research project?
2. Could a lay person understand the scope and risks of harm, as well as the residual ethical dilemmas?
3. Has the researcher informed the data subject about possible and foreseen further uses?
4. Will the researcher give data subjects insight into what data is collected, what secondary uses the data is used for, and share the research results with the data subject?
5. Could a lay person understand how long the data will be stored and how it will be disseminated?
6. If the purpose of the collected data changes, can the data subject be informed and can she retract her consent?

Once a data subject has given her informed consent or the relevant ethics oversight committee has waived the neessity of informed consent for this project, the researcher is in principle free to start the project.

Supporting modules

Sociotechnical systems and Embedded Values

Scholars widely refer to Internet technologies as sociotechnical systems, where humans and networked technologies interact to mediate an increasing amount of activities. Internet technologies allow technically proficient persons to collect, analyse, and further disseminate vast amounts of data about people's behaviour, preferences, activities, and other parts of their lives. It is also possible to create new information flows or measure the controls on existing information flows, thereby potentially altering the environment of data subjects.

Sociotechnical systems methodologies generally analyse and inform the interplay of human, social, organisational, and technical factors in the design of technological systems. It is acknowledged that sociotechnical design approaches are more likely to lead to acceptance by end users. A crucial factor is thus understanding 'sociotechnical constituency', which is comprised of the underlying social, political, economic, and other forces shaping the development of technology.

Common characteristics in the definitions of sociotechnical systems are that :

1. they consist of interdependent parts;
2. these parts comprise of social and technical subsystems;
3. they pursue goals in the external environment;
4. design choices have been made during their development; and
5. the performance relies on the optimisation of the subsystem’s technical and social cooperation.

Technologies are typically not politically or value neutral, so their operation will likely cause a shift of power of some sort in the society in which it operates. The values of their designers are typically embedded in a technology. It is important to be aware of the values that a researcher imposes on (unsuspecting) users, data subject, or context. For example, below is a list of values that are inherent in blockchain technologies, with a short statement about some potential consequences on each:

- **Decentralisation** - can lead to scale-free power laws, which typically leads to inequality.
- **Disintermediation** - can lead to a lack of accountability, eroding the usefulness of consumer rights laws, for example.
- **Transparency** - is usually a positive value, but radical transparency erodes values such as privacy, or disallows trade secrets.
- **Trustless system** - many people would prefer to trust a bank with their savings, rather than volatile cryptocurrencies.
- **Computationally mediated contexts** - while a few may find this concept interesting, it makes most people feel uncomfortable to be controlled by algorithms and systems.
- **Security** - the security provided by Bitcoin, for example, uses vast amounts of energy, which may be a cost that is politically too high.
- **Interoperability** - proprietary systems give some comfort to people who are not technically proficient enough to adjust their IT devices and software to suit their needs.
- **Self-organisation** - while and interesting idea, would these organisations reflect the libertarian techno-anarchist values embedded into blockchain technology?
- **Independent identity reputation systems** - will this lead to people carrying a reputation number for the rest of their lives, being held accountable for actions in the past?
- **Loss of control** - what does a loss of control mean in the long term? How are future externalities dealt with?

Easy Data Sources

Data is available in abundance on the Internet. Data can be measured, or existing databases, some of which may have been hacked or gathered unlawfully. It in enticing to collect the available data and make create analyses or projects around them. However, you may be violating rights or ethical principles by doing so. When thinking about easy data sources, keep in mind that just because you can, doesn't mean you should.

Privacy laws, data protection frameworks, and ethical principles likely apply to data that are generated about identifiable humans, their behaviour, their devices, etc. These people likely did not consented to be part of your research. Identifiable means identifiable by anyone, including those with more computing capacity or resources than your average research team (although it is difficult to say where to draw the line in such a hypothetical assessment).

The oft repeated statement "I don't see how currently anyone could have an expectation of privacy any more" once data has been published, does not hold. Likely, these people had an expectation of privacy at the time of communicating their data, or interacting with the website whose databases had been leaked. They did not intend for their data to be used for academic research, or any other processing outside of the context of the service they trusted. The researcher would be changing the context of this information and creating new unintended audiences by publishing about it and making the database available. The researcher should take into account the potential harm of using such information in new contexts and new audiences.

Further, it could be argued that by using this data unlawfully or unethically collected data, you're (implicitly) condoning these act. Stronger, still: If you profit from using information from this breach (publications and other career enhancements), you could entice others to also work with dubious data sources, therefore potentially incentivising (and even justifying) these acts. The researcher would be setting a precedent about working with unlawfully or unethically collected data that may be difficult to reverse.

Values and Harms

The Module on Sociotechnical Systems and Embedded Values explains how values can be impacted and people can experience harm as a consequence of technical research projects on the Internet. It is beyond the scope of these guidelines to list all possible values that could be affected by Internet research, or to provide an encyclopedia of human values. Instead, below is a list of typically affected values. Further, we provide a brief analysis of two values (privacy and trust), a violation of which can cause harm to people.

Typical Values: Batya Friedman has identified the following values in her papers on Value Sensitive Design:

- Human welfare,
- Ownership and property,
- Privacy,
- Freedom from bias,
- Universal usability,
- Trust,
- Autonomy,
- Informed consent,
- Accountability,
- Courtesy,
- Identity,
- Calmness,
- Environmental sustainability.

Privacy

Privacy is a difficult term to define. Daniel Solove has developed an overview in his paper "A Taxonomy of Privacy", which builds on previous works such as William Prosser's "Privacy" and Alan F. Westin's book "Privacy and Freedom". Many types of privacy have been identified, but for the scope of these guidelines, we will only focus on informational privacy (also known as data privacy) and the related principle of informational self-determination.

Informational/data privacy and self-determination

Informational privacy concerns the individual's right to exercise control over the disclosure or processing of her personal information. In this sense, Westin defines privacy as "the claim of individuals, groups, or institutions to determine for themselves when, how and to what extent information about them is communicated to others". Personal information is generally defined as any information about or otherwise relating to an identified or identifiable individual. Whilst there has been a debate about whether truly innocuous data is covered by this definition, it should be stressed that the type of data used in network measurement research will, if identifiable, fall within this definition. Information will still be identifiable even if this is only possible when the data in question is matched to information stored in another (or even several other) auxiliary databases.

Solove identifies four harmful activities with regards to privacy, all of which are directly relevant for informational privacy:

- Information collection (surveillance and interrogation);
- Information processing (identification, aggregation, storing, second uses, exclusion);
- Information dissemination (disclosure, breach of confidentiality, exposure, increased, accessibility, distortion, blackmail and appropriation);
- Invasion (intrusion or interference into one's life).

Informational self-determination is a very important concept to mitigate these harms for several reasons, including (according to Westin):

- Personal autonomy– the development of the personality and prevention of manipulation by others;
- Emotional release – the ability to escape everyday tensions;
- Self-evaluation – understanding events and experience from the individual's perspective;
- Protected communication – sharing information with trusted individuals and setting interpersonal boundaries.

Privacy harms

The need to protect privacy can best be explained from the perspective of possible harms and risks. Protection from privacy harms is a legal right for data subjects in many countries. Privacy breaches can be first-order harms (e.g. identity theft, blackmail), or a second-order harm, which increases the risk of other first-order harms (e.g. disclosing data collected by surveillance). Finally, privacy breaches can also lead to less immediately obvious harms, such as the loss of individual autonomy.

Decisions based on computer algorithms

Decisions affecting people's lives are increasingly based on inferences generated by aggregated information and automated processes, which are linked to natural persons. Modern societies have increasingly placed trust in algorithms to make these decisions. However, the linked data is often incomplete and thus only represents a facet of people's lives. Further, when aggregated incorrectly, misguided decisions can be made, which can have significant impact on people's lives.

Identity theft

It has become an easy and common criminal practice to use another person's identity to commit crimes, accessing resources or obtaining credit in another person's name. The victim of identity theft will often be left with a tainted digital identity, whereby decisions will subsequently be made about her based on information which classes her as a criminal or debtor. It can take a long time to fix such problems, and the victim may have trouble finding employment or mortgages during this time. Careless dissemination of mobile data containing personal information may give criminals more information about a specific person, to possibly make identity theft easier, or more comprehensive.

Blackmail

Blackmail is a crime in many countries, where the criminal threatens to publish certain information about the victim if certain demands are not met. Extensive mobile connectivity datasets can potentially show some incriminating information or expose behaviour, which may be used against a victim. The blackmailer may find further unexpected information about his victim on close inspection of a research dataset, while looking for other information.

Use of data by governments

Governments in different countries, but also over time, use data for many different purposes. It cannot be guaranteed that their motives will always be in line with the (good) intentions for which the data was originally collected. Data collected from mobile phones can be highly sensitive personal information. In her book Privacy in Context, Helen Nissenbaum states: "[…] information is a more effective tool in the hands of the strong than in those of the weak."

The revelations about the NSA and other intelligence agencies have shown that it is not only authoritarian governments that can misuse data gathering powers, but also democratically elected governments. Further, surveillance by public (e.g. government intelligence agencies) or private organizations (e.g. mobile connectivity researchers), can affect people's behaviour and sense of freedom.

Privacy in context

Privacy proponents do not simply want to restrict the flow of information, as many uses of data can be beneficial for data subjects or for the wider public good directly. However, there is a societal interest to ensure information flows appropriately, in order to prevent harms from being inflicted from an uncontrolled flow of information. To assess the appropriateness of the flow of information, contextual considerations such as the capabilities of an adversary or the political environment of the data subject must be taken into account.

Trust

The concept of trust is also complex. However, Friedman, Kahn, and Howe propose (pdf) a more simple and usable analysis of the concept, which is still based on a philosophically informed conceptualisation. It is summarised below:

- The authors propose that people trust when they are vulnerable to harm from others, and
  - yet believe that those others would not harm them even though they could.
- In turn, trust depends on people's ability to make three types of assessments:
  - One is about the harms they might incur.
  - The second is about the goodwill others possess toward them that would keep those others from doing them harm.
  - The third involves whether or not harms that do occur lie outside the parameters of the trust relationship

Adversaries

A wide variety of adversaries exist in the Internet environment. It is therefore most useful to conceptualise adversaries in terms of their motivation, capacity, and interest, rather than the type of person or organization. These classifications have been taken from work by El Emam et al. in the context of medical ethics:

- **"Prosecutor risk" - High risk**
  - Wants to re-identify a specific data subject;
  - Possesses auxiliary information which can be combined to reveal certain information about data subjects;
  - Has legal powers to compel the production of stored information.

- **“Journalistic risk” - Medium to high risk**

    - Searches a specific target in the dataset;
    - Possesses auxiliary information which can be combined to reveal certain information about a data subject.

- **“Marketer risk” - Lower to Medium risk (depending on how many individuals can be re-identified)**

    - Adversary tries to identify as many people as possible;
    - The more people are identifiable, the higher the risk.

Dissemination

Not all datasets will be suitable to be published in an open data format - for example, when the sensitivity and granularity of the data is high. In such cases, the risk of re-identification will be too high to publish in an uncontrollable open data format. This section recommends methods that do not make a dataset available freely and without restrictions.

Restricted data sharing

The researcher only disseminates research datasets to persons or organisations on request, refusing dissemination when the level of risk is considered too high. The researcher should discuss the expected types of recipients and the corresponding risk level with colleagues and/or a legal expert. Generally, the following risk level can be assigned, although the interest of the recipient and their general information security and privacy standards need to be taken into account:

- An individual researcher from same organisation - Lower risk;
- Sharing with a research consortium - Medium risk;
- Sharing with a commercial entity or government - Higher risk.

The risk level can be lowered if the researcher attaches certain requirements or limitations to the use and further dissemination of the dataset. The researcher should inform data subjects of the restrictions on secondary dissemination and also enforce the requirements and limitations when the third party does not act in accordance with what has been promised to the data subject. Not enforcing such commitments would be even worse for public trust in network research than not making promises to the data subject at all.

Managed access

Instead of disseminating the dataset to specified third parties, the researcher can provide managed access to the dataset. Third parties can query the dataset and conduct statistical (or other) analysis. Such an approach allows the researcher to ascertain exactly who accesses the datasets, while maintaining control over its dissemination. The risk level of a managed access system can be considered to be lower.

Interactive methods

The most well-known interactive method of publishing research data is called Differential Privacy, developed by Cynthia Dwork and explained in her paper on the topic (pdf). It is a particularly robust method, which only gives statistical answers to queries about an underlying dataset. To protect privacy even further, a certain amount of noise is added to the disclosed statistical data.

In principle, differential privacy offers a lower risk for privacy, but there are certain limitations to this approach that need to be understood. For example, the uncertainty related by the addition of noise to the data can be exhausted, which means the dissemination must then stop.

Hybrid

The researcher may consider splitting a database that contains personal information that is likely to be re-identified. For example, direct identifiers can be stored in a managed access system, whereas the other non-identifying data fields are published freely in a repository. The researcher can then attach a certificate to the dataset, giving a contact address that informs the third party how she can request access to the full dataset. Such approaches limit the risk of re-identification while maximizing utility.

De-identification

De-identification techniques are useful tools to make it more difficult for an adversary to identify individuals in a dataset. Full de-identification is very difficult to achieve, however, and "anonymised" datasets have often been re-identified. For example, a former governor of a US state was identified by combining a public "anonymised" healthcare dataset with auxiliary data (pdf).

More sophisticated methods exist, such as using *fingerprinting* techniques, where inferences about individuals can be made based on seemingly non-identifying data types. Especially in multi-dimensional datasets, such as often created with mobile Internet measurements, for example, it has proven to be possible to uniquely identify a large part of the dataset. Human mobility traces, for example, have been found to be highly unique: only four location data points over the course of one day were necessary to identify 95% of individuals uniquely. The settings, configurations and combination of plug-ins of certain applications - such as a web browser - can also be used to distinguish individuals and identify them with adequate auxiliary data.

The premise of these guidelines is the generally accepted understanding that full de-identification is not possible without significant loss of utility. Exemptions for "anonymised" datasets in existing privacy laws are therefore not a suitable ethical standard. To guide the researcher in the choice of suitable de-identification technique, we describe some methods below and attribute a level of robustness to them (higher, medium or lower robustness). None of these techniques are perfectly resistant against determined attackers with access to sufficient auxiliary data sources, computing power and determination. This does not mean, however, that only the strongest de-identification technique is suitable for network research. The choice of technique should be guided by an assessment of the risk.

Perturbation

One of the simplest approaches to de-identifying a datasets is to add 'noise' to genuine values. For numeric quantities this can simply be the addition of random figures according to an appropriate probability distribution. For categorized data this can result in attributes being re-assigned in various ways. Rick L. Wilson and Peter A. Rosen discuss the use of perturbation and its impact on the ability for knowledge discovery in their paper "Protecting Data through 'Perturbation' Techniques: The Impact on Knowledge Discovery in Databases" (pdf).

Truncation

In numerical data, truncation limits the number of significant digits stored, thereby making such values less accurate. Truncation can also be applied to IP addresses, postcodes or other key-identifiers.

- Lower robustness;
- Simple and useful when dealing with fields that contain sensitive data;
- Accuracy is decreased.

Randomization & permutation

This approach refers to reordering the values of a column without losing the accurate values in the dataset.

- Medium robustness;
- Useful when dealing with fields that contain sensitive data;

- Useful for maintaining statistical utility and accuracy, such as aggregate counts; averages and distribution of data;
- Individual accuracy is lost.

Quantization

Similar to truncation, quantization constrains continuous values into elements of a defined set. This could, for example, take the form of grouping values such as exact height into a number of height ranges.

- Medium robustness;
- Care must be taken that all groups contain sufficient individual entries;
- Accuracy per individual record is lost.

Pseudonymization

This method replaces directly identifying fields of a dataset with non-identifying values. A common example is to replace identifiable IP addresses with (linkable) prefix-preserving pseudonyms or hashes of data. This is typically aimed at maintaining the links between a group of records, whilst removing the ability to easily identify the record identifier itself.

- Lower robustness;
- Preserves all key and secondary-attributes, so risk of fingerprinting or linking with auxiliary data remains.

K-Anonymity K-Anonymity is widely used in network research, which ensures that any record in a database must be identical to some number of other rows, forming a group of size k that is indistinguishable from each other. This approach may take the form, for example, of grouping subjects' locations into sufficiently large areas such that no set of locations is unique to any individual.

Latanya Sweeney describes this method as follows: *"A release provides k-anonymity protection if the information for each person contained in the release cannot be distinguished from at least k-1 individuals whose information also appears in the release"* (pdf)

- Medium/higher robustness, depending on suitable threshold for "k";
- A quantifiable probability that individuals could be re-identified exists. The probability that a data subject can be identified is 1/k, where k is the size of granularity chosen;
- It may still be possible to infer sensitive information about a person, even if direct identification is impossible. Further, the attributes shared by an entire group, such as a particular disease or condition, could be sensitive;
- Knowledge of the specific k-anonymisation algorithm could be sufficient to re-identify a dataset;
- Datasets with k-anonymity applied have been re-identified. See, for example, the AOL search data case, where k-anonymity was shown to be useless for some individuals in high-dimensional datasets and the information revealed was very damaging;
- An appropriate threshold for "k" depends on the context in which data is collected and disseminated. The researcher must consider this on a case-by-case basis.

Various extensions to k-anonymity have been proposed to mitigate weaknesses against various forms of attack. A full discussion of these is not appropriate here.

Differential Privacy

The concept of differential privacy (developed by Cynthia Dwork) is an interactive privacy method for statistical databases. Differential privacy does not guarantee that a privacy breach will not occur, but it guarantees that the privacy breach will not occur due to the data in the database. Breaches that can happen if data is in the database could have happened even if the data weren't in the database. This accommodates any and all possible auxiliary information available now or in the future. However, differential privacy has some limitations, so the usefulness of the concept must be well researched before it can be applied to specific network research.

Ethical lenses

Ethics is typically separated into three strands of analysis: consequentialism, deontology, and virtue ethics. More theories exist, but for the purpose of these guidelines it is sufficient to focus on these three. These lenses should be used mainly to assess the ethical dilemmas that are stated through the use of these guidelines. However, they may also be used in other steps of the process, for example to assess alternative technology choices or the scope of the research question.

Consequentialism

Consequentialism (or utilitarianism) is the view that the ethical rightness or wrongness of an action is determined entirely by the states of affairs it causes. In short, the ends justify the means. Consequentialism should help people choose the action that will bring the most good to the party the actor deems most important.

Deontology

Deontological theories stand in opposition to consequentialism. While most deontologists do take consequences into account when evaluating the ethics of an action, they do not consider consequences to be the only element that should be taken into account. More importantly, a fixed set of duties, rules, and policies define actions as ethical. Break these rules and you have behaved unethically.

Virtue Ethics

Unconcerned with duties or consequence, focusing instead on the subjective, idiosyncratic, and seemingly non-rational impulses that influence people in the absence of clear rules and consequences. Virtue ethics takes an alternative approach to determining the ethical acceptability of an action. Whereas consequentialism and deontology examine the quality of an action, virtue ethics is concerned with the character of the actor — it prescribes “how we should be rather than what we should do” (Darwall, 2003). Virtue ethics can provide guidance for ethical action by connecting the acceptability of actions to the character of the actor (Oakley, 2001). Right actions are those that would be chosen by a ‘virtuous’ agent.

Informed Consent (and exemptions)

Informed consent is recognized as a central and generally applicable principle and legal basis in scientific research when information is collected directly from the data subject. It is the process of obtaining a legally relevant approval from data subjects after they have been given the chance to understand and consider the use of their data for the research project. This demonstrates that participation is voluntary and that data subjects receive a comprehensible description of the research, including the risks they face and the benefits for research and society as a whole. When feasible, researchers should obtain informed consent from research participants, data subject, owners of information systems, and other affected stakeholders. It is important researchers do not consider stakeholders as means for a project, but as ends in themselves who deserve to be treated with full respect for autonomy.

Limitations and Challenges

Informed consent is not a litmus test for research and the concept has several limitations. It can be considered meaningless when users are presented with lengthy and complicated text to read that may be incomprehensible to them. Even if the texts is read in its entirety, it is uncertain whether the user can meaningfully assess the potential future risks of harm in the Internet’s dynamic information environment. It can be argued that in some contexts the concept of informed consent may impose unrealistic obligations on researchers and individuals, which then serve merely as a liability disclaimer, rather than allowing the user to act as an autonomous person. An informed consent notice to users may also significantly reduce the ability to collect data, which may be harmless or risk-free.

Exemption from Informed Consent

When the risk of harm is minimal, informed consent is generally not a requirement for scientific research. Further, the Association of Internet Researchers ethics guidelines state *"...providing notice may be particularly challenging given the scale and scope of many operational ICTR environments. It may be impracticable, it may not be technically feasible to identify subjects, or it may interfere with scientific integrity of the results."* **These guidelines aim to help minimize the risk of harm to the fullest extent possible in order to give researchers and other actors in the scientific process the means to discuss whether indeed a waiver of informed consent is justifiable**. A justification to waive an informed consent procedure should be strong, and agreed upon by relevant parties. The researcher should not make this decision in isolation. Typically, a debriefing procedure will be required to the extent that it's feasible.

Intermediaries

In projects where it is unfeasible to identify and contact every potential human involved, researchers may suggest to gain consent from intermediaries who act as representatives of data subjects. Examples include Internet service providers, platform or service owners, or representative bodies at a university. Researchers will need to explain why the departure from individual informed consent procedure is justifiable from the perspective of individual autonomy within the context of the project.

Requirements

If the actors in the scientific process decide that informed consent will be a necessary precondition to collecting data, the consent must be freely given, be specific, and an informed indication of wishes of the data subject. Data subjects should not be coerced into accepting the notice.

- **Free**: consent must be given without any pressure from the researcher;
- **Specific**: consent is for a specified purpose;
- **Informed**: The data subject must be informed how her data is processed for the specific purpose;
- **Indication of wishes**: the data subject must have indicated her wish to give consent by some action, for example ticking a checkbox. Silence, or implied consent, should not be equated with informed consent.

Key elements of notice

A thorough assessment of the information given to data subjects prior to consent should be part of the iterative reflexivity process. The average user is not aware of the risks of complex Internet research or how the data collection and processing systems work. The notice should therefore be written in layman's terms, but at the same time not oversimplify the risks involved.

Key elements of an informed consent notice:

- What data will be collected for the research purpose;
- How this data will be processed, de-identified;
- Whether measurements will be user-initiated or run in the background;
- Whether data will be published, and to what extent a person may be identifiable ;
- That the data will be used for research purposes and how these will benefit society and/or certain stakeholders;
- When the database is split, whether personal information will be stored securely
- Explicitly state if sensitive data will be collected, for example when IP addresses and geo-locations are collected;
- Length of time data will be stored;

- Explain that data will not be fully anonymous, but explain the measures that are taken to ensure the risk of identifiability is minimised;
- Highlight that identity may still be revealed, even after de-identification;
- Give data subjects an indication of risks they may need to consider, stating that one cannot anticipate all the secondary uses in the future, especially if the dataset is disseminated in an open data format.
- Researchers should inform subjects that they may not immediately benefit from the research directly, although society may benefit in the future.

Emphasis should be added to areas where the risk is identified as being higher, for example by adding bold text. When possible and feasible, the researcher should meet the data subject in person (or telecommunicate) to discuss the points above in person. This is especially important when datasets generated by the research may reveal very sensitive data or when the (political) context is particularly dangerous to a person. It has been suggested that a 30-second video explaining how the project works could already go a long way to explain the details to users and data subject who are not prepared to read text documents.

--END--

This page was last modified on 10 July 2017, at 10:05.