CREDS 2014 – Position Paper

Four ethical issues in online trust
Topics for a moderated workshop

Robin Wilton
Technical Outreach Director – Identity and Privacy
Internet Society

wilton@isoc.org

Document ref: CREDS-PP-2.0

Background: The Internet Society's Trust and Identity initiative.

Our experience is in the application of ethical principles to questions of online trust and privacy, and digital identity. Our approach is based on the principle of involving and responding to multiple stakeholder interests. In doing so, we have encountered consistent types of problem, each of which highlights a particular ethical facet of the trust/privacy issue.

This paper sets out four such problem types as the basis for:

- multi-stakeholder discussion
- formulation of a candidate problem statement
- derivation of general ethical principles.

Four ethical issues in online privacy/trust:

1. The “no surprises” principle
2. Ethical dilution
3. Multi-stakeholder issues
4. Multi-context issues

In engaging with the CREDS community, our goal is to use this position paper as the basis for a moderated multi-stakeholder workshop, with the aim of producing two pieces of output:

- A candidate problem statement for ethical guidance on data handling in the domain of cyber-security research;

- Statements of ethical principles which extend beyond the cyber-security research domain, and provide more general guidance for ethical data handling in the modern, mediated digital environment.

1 - The principle of "no surprises"

One practical "rule of thumb" for privacy is that the use of data should not come as a surprise to the data subject. This is reflected in the basic principle that the use of personal data should conform to a stated purpose. However, even the most comprehensive privacy laws do not guarantee this as an outcome, and as a result, there is frequently a gap between the expectations users have concerning personal data, and the ways in which it is actually used, shared and re-used.

It may be useful, here, to think of the difference between "legal" and "legitimate", and how it might relate to the principle of "no surprises". We can probably all think of cases where the use of personal data is, strictly speaking, legal, but where it would come as an unpleasant surprise to the data subject. Perhaps they signed up to a usage statement that was general enough to cover a multitude of sins; perhaps they didn't sign up, but were simply opted in by default; or perhaps there was an implicit consent step of which they were not made sufficiently aware. These all represent failures of legitimacy – but it is also worth noting that none of the issues above is new. The principles of notice, consent and purpose are well established in law, regulation and good practice, but by no means universally obeyed.

Regulations are generally framed to include exceptions, exemptions and interpretations that permit data collection and use beyond the norm for specific purposes or contexts (e.g. public safety, law enforcement, security, research). A major challenge is to ensure that such carve-outs remain consistent with what is just and fair, particularly since data use practices tend to evolve much faster than the related laws and regulatory measures.

The data protection exemptions for law enforcement and security have, of course, recently come in for a great deal of scrutiny. Privacy advocates and civil society representatives have, for some years, pressed for such exemptions to be qualified – and in some pieces of legislation that has been the case. For instance, European human rights law applies a test of "necessity and proportionality in a democratic society". The Council of Europe, a European human rights institution, has published a usefully detailed analysis of the implications of this phrase[1].

Citizens have been unpleasantly surprised to find that the "necessity and proportionality" clause has had little or no effect in constraining pervasive monitoring of the Internet by state actors, under the aegis of the law enforcement and national security exemptions described above. Again, this could be seen as a failure of legitimacy: what has been done may be legal under some interpretation of the wording, but nevertheless seems to lack fairness, and to have escaped the "adequate safeguards" also called for by human rights legislation.

Legal issues aside, in today's data-driven economic environment the regulatory focus is increasingly turning to "data use" as a pivotal point of influence for privacy protection. In part, this is a response to the difficulty of deleting data once it has been published, shared or stored online. For example: since it may be impossible to remove all trace of an image that was shared via social media, some are calling instead for the focus to be on control of its subsequent use - e.g. in the context of recruitment, research and so on. This is a topic we will examine in the following section.

[1] http://www.coe.int/t/dghl/cooperation/lisbonnetwork/themis/echr/paper2_en.asp The Margin of Appreciation – Council of Europe

2 - Ethical Dilution

A perennial problem in online privacy is the lack of a clear ethical equation to solve. For instance, it is easy to formulate principles such as: “If passive collection of personal data causes the individual harm, then it is ethically bad”, but the lack of a clearly quantifiable and demonstrable harm can make it hard to justify the conclusion - and even harder to justify doing anything to prevent or constrain passive collection. The case of passive collection (i.e. collecting data that is disclosed incidentally, such as CCTV images captured while the individual is walking down the street) raises particular questions because it gathers personal data without there being any related intent on the part of the individual.

Part of the problem is “dilution”. Individuals’ ethical judgements about the disclosure/collection of personal data are distorted by a number of practical factors. Even if an individual is actively and willingly disclosing data, he/she may be doing so on the basis of a flawed, incomplete or misleading set of assumptions.

- The harm resulting from disclosure/collection is often remote, in time and place, from the behaviour that gave rise to it. Examples include identity theft, or spam mail.

- Harm may be fragmented or dispersed, such that its impact on any given individual may not be enough to provoke the individual to react, while its over-all effect may still be one of significant harm. An example is the chilling effect of mass interception of communications. Even knowledge of the possibility monitoring can be enough to change behaviour, as was well understood by both Bentham [1] and Foucault [2] in their work on panoptical systems.

- Some forms of harm, such as damage to a person's reputation, can be difficult to quantify in ways that allow a clear remedial action to be defined.

- The data itself becomes fragmented/dispersed. Everything we do online is mediated through at least one other party, and the ecosystem of data monetization is complex and highly populated. This dilutes the accountability for data use and any resulting harm, and can make recourse impractical or even impossible.

- Sometimes the “dilution” is intentional: data may be aggregated, pseudonymised, anonymised and so on: again, the harm that might result from future use or re-identification of such data is often remote enough to make pre-emptive measures impractical. There are two risks, here: first, that supposedly anonymous data can be re-identified with increasing sophistication; the science of assessing *how* anonymous a given dataset is, and with what reliability over what timescale, is still in its infancy. Second, it is no longer necessary to re-identify data in order for that data to have a privacy impact. We are all, every day, affected by inferences drawn not from the data we have disclosed ourselves, but from the data diclosed by others. The science of anonymity may be in its infancy, but the science of profiling is mature and sophisticated.

At any given point in these distributed, mediated chains of use, it is hard to construct a clear ethical equation that leads to a simple, conclusive and actionable proposition. If there are effective remedies to this, they are most likely to be found in three areas:

1. More nuanced understanding of “harm”, including risk, potential harm, and forms of harm other than physical/financial.
2. Use, impact and outcome-based approaches to ethical data handling.
3. Accountability as a way of ascribing ethical considerations beyond the data collector, and a potential basis for designing more effective monitoring of data-handling.

3 - Multi-stakeholder issues

In its simplest form, one statement of the multi-stakeholder issue in online privacy/trust is that the online services of today involve too many competing and unequal interests to be easily resolved. Power relationships are asymmetric, and this leads to distorted outcomes; users accept a "bargain" that is far less oriented towards their interests than they may suspect.

Stakeholders in and among the individual, commercial, government and societal domains all have their reasons for participating online, different goals and success criteria, and different ethical calculations to make as a result. A Syrian anti-government activist, a farmer in Tanzania, or an advocate for women's rights in Pakistan, will have a very different perspective on the purpose and benefits of the Internet from that of an NSA analyst or a multi-national telco.

The topic of Internet governance is high on the international agenda for 2014, and raises multi-stakeholder issues at every turn. Governments, telecommunications bodies and corporations are all legitimate stakeholders, but is any of them a legitimate candidate to hold ultimate executive control over the governance regime? What should be the role of civil society? Is the Internet a tool for economic growth or social well-being, or are those two criteria inseparable? When is it better to regulate commercial activity rather than let the markets take their course?

All of these questions bear directly on online privacy, and have an inescapable ethical dimension.

In the privacy domain, the possibility of using privacy *outcomes*, as a metric, is generally disregarded in favour of attempts to regulate particular kinds of data (personal, sensitive, identifiable) or abstractions such as the relationship of "purpose of collection" to "actual use". As technology advances, these become increasingly ineffective as a means to provide effective privacy.

Sometimes, as indicated above, under Ethical Dilution the argument is made that no harm comes from collection, only from use. Leaving aside risks of stored data breaches, this neglects an important ethical dimension: is it fair, in principle to collect data that has the potential to have a privacy-impact?

There is also often a danger of trying to reach ethical conclusions through false opposition. A typical example would be argument such as: "This is a matter of drawing the balance between individual privacy and national security interests"), rather than either trying to optimise for both, or arrive at optimised relationships of interests/power.

These multi-stakeholder issues increase in complexity when (like the Internet) one takes a global view without regard for national and cultural boundaries. Is it either desirable or achievable to aim for ethical guidelines that (a) can be applied globally, in a global Internet environment, but (b) respect differences between regional/national cultures, social aspirations and individual morals? Is this the 21st-century version of the debate over moral absolutism and moral relativism?

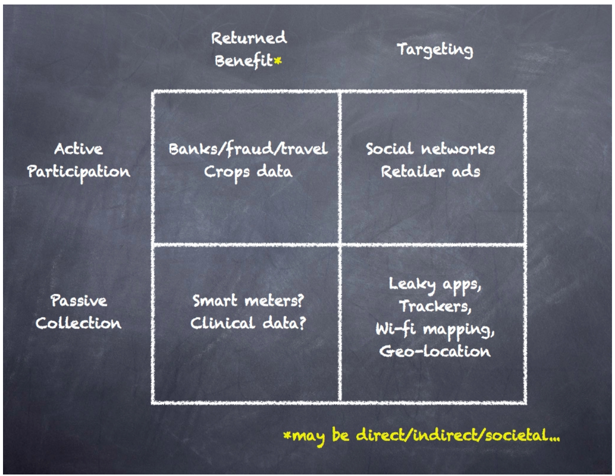
The Internet Society's experience is also that different stakeholder groups "speak different languages" when describing identity and privacy, and this presents an immediate obstacle to productive multi-stakeholder discussion about the ethical dimension of these topics. One way to overcome this obstacle is to develop a shared conceptual framework for the discussion; this approach has proved successful in the identity and privacy domain, and extending existing models to cover ethics would be a valuable exercise.

4 - Multi-context issues

As mentioned above, multi-stakeholder issues of privacy and trust become more complex when considered in the context of different nations, jurisdictions and cultures.

This is just one set of multi-context issues; there are also those that arise when data crosses contextual boundaries between industry sectors, application contexts, or domains of personal interaction. Helen Nissenbaum's concept of "contextual integrity" [4] remains entirely relevant. Contextual control over the collection and use of data remains a problem which is hard to solve either by exclusively non-technical means, or by exclusively technical ones. Contextual metadata, 'tagging', and various means of privacy preference expression are all promising technical avenues, but none has yet led to usable privacy-enhancing solutions. The general problem is that of finding the appropriate mixture of technical, policy, regulatory and procedural measures to achieve the best result (where "best", in this instance, relates to ethical outcomes. The continued absence of widely-adopted solutions in this area perpetuates an ethical problem, in that data subjects lack the ability to express and enforce their preferences with regard to personal data.

The multi-context issues also surface in different forms, if one analyses data disclosure according to a number of criteria: active disclosure versus passive; disclosure with a direct, returned benefit versus disclosure that results in asymmetric (or no) benefit, and so on. Figure 1, below, gives a simple example of this analytical approach.



	Returned Benefit*	Targeting
Active Participation	Banks/fraud/travel Crops data	Social networks Retailer ads
Passive Collection	Smart meters? Clinical data?	Leaky apps, Trackers, Wi-fi mapping, Geo-location

*may be direct/indirect/societal...

Figure 1: A simple consent/benefit model

The current landscape (of mobile devices, app-borne functionality, hyper-connectedness and "big data") is characterised by services which purport to offer one specific direct benefit to the end user and data subject, while (overtly or otherwise) seeking a different benefit for a different stakeholder (e.g. data capture for later mining, or direct revenue for online advertisers). There is frequently a "consent gap" between the consent granted in expectation of the direct, stated benefit of the service, and the implied consent to other uses of the resulting data.

Context changes over time – particularly with "big data" and its reuse. This is of direct relevance in, for example, clinical and healthcare data; there is a de facto assumption of legitimacy in the reuse of such data for diagnostic, research and national healthcare planning purposes, even when the data originated in a different context (generally, a hospital or general practice visit). The potential future benefits of such use are perceived to outweigh the short-term privacy interests of the individual. This position is strengthened if it is further claimed that the data in question has been rendered pseudonymous or anonymous.

However, the question of re-identification deserves serious consideration. The research published by Narayanan and Shmatikov [5] is one example of how assumptions of the robustness of pseudonymisation can be ill-founded – and yet regulators, data controllers and policy-makers continue to make decisions about data safety, based on unreliable assumptions about anonymisation and pseudonymisation. Similarly, research done by Prof. Sandy Pentland [6] illustrates how information which is, on the face of it, anonymous from the outset, can in fact be used to generate accurate predictions of the movements and behaviour of individuals.

In principle, we can expect any anonymisation/pseudonymisation mechanism to be subject to several kinds of pressure over time:

- improvements in the techniques of re-identification
- ability to identify or re-identify sparser and sparser datasets
- likelihood of datasets becoming richer over time (e.g. through combination)
- compromise of anonymisation/pseudonymisation technique

In many respects, these pressures are analogous to those faced by cryptographic algorithms. The risks associated with re-identification increase, as data becomes richer and more persistent; but the strength of anonymisation/pseudonymisation mechanisms is not subjected to the same kind of scrutiny as cryptographic mechanisms, and the discipline of quantifying their robustness (e.g. the "work factor" required to render data identifiable) is neither mature nor widely practised.

Comparison of three ethical approaches

Defining an ethical strategy for personal data processing, whether in the cyber-security research domain or in general, leads down many of the same pathways as any other ethical enquiry. Is any of the 'standard' ethical approaches (consequentialist, rule-based and justice-based) particularly well suited to the task?

Consequential

Ostensibly, a consequentialist view is a good fit for ethical data handling: it would lead us to base our ethical assessment on the outcomes of particular data handling practices. However, in practice, this approach suffers most from the “dilution” effects described above. Harm, risk and accountability are all viable criteria in principle, but the distributed and asymmetric nature of the personal data ecosystem make them hard to enforce.

The consequentialist approach can also be subject to manipulation, as recent experience with UK healthcare data policy illustrates. A case is made for the consent-less aggregation of patient data, on the basis of predictions about future clinical utility; however, competing interests and dis-benefits are played down in the policy debate (for instance, the risk to the individual arising from inappropriate disclosure of health data, the possibility that the hoped-for clinical benefits do not materialise, and the reality that commercial third parties stand to profit from the policy regardless of both these factors).

There are observable parallels between this example and the kinds of policy argument sometimes advanced in favour of cyber-security related work. Consent-less collection and use may be justified (for instance, under the 'national security' exemptions described earlier) on the basis that it contributes to the greater good, improving security for all, even if this is at the cost of some personal privacy. Benjamin Franklin was conscious of the civil liberties risks of this approach, observing (in various wordings) that “those who sacrifice essential liberty for a little temporary security deserve neither”.

That said, the consequentialist approach should not be written off: data-handling policies based on the 'traditional' approach - defining lists of what counts as 'personally identifiable information' (PII) and then prescribing what may or may not be done with that – can only take us so far, and cope relatively poorly with the modern ecosystem of attribute monetisation. An approach based on risk-assessment and outcomes can help address the shortcomings of a data-driven approach – perhaps helping us to redefine as “privacy-impacting information”.

Rule-based

Arguably, one could characterise the current state of affairs as a rule-based approach (founded, as noted above, on the idea of categorising data as personal or not); if the information falls into the category of “personal” there are constraints on what may be done with it, and if it does not, its use is unconstrained. The behaviour of data controllers can be influenced not just by enforcement and penalties, but also by the notion of compliance; evidence of efforts at compliance may reduce the data controller's liability in case of a failure, or may reduce cost and risk in other areas.

However, some of the practical problems with a rule-based approach can be seen if we look at experience with Safe Harbour agreements. These are, on the face of it, a pragmatic rule-based approach to dealing with the issue of cross-border data transfers and compliance with data protection laws. Consentless cross-border data transfers are acceptable under data protection law, provided the data in question benefit from equivalent safeguards in the destination country to those they would have enjoyed in the country of origin.

What we find in practice is that this approach can fall short of delivering the ideal outcome for a variety of reasons, of which the following are examples:

- Differences in the approach to remediation. In the EU, the route of recourse would be via the national data protection authority, in the expectation that they would rule on questions of principle and impose a statutory penalty; in the US, it is more likely that recourse would have to be sought via the courts. This may not be a practical route for a remote plaintiff, and may lead to arguments about 'provable harm' which (as indicated in the section on Ethical Dilution) can be problematic. Only comparatively recently have some US class actions suits resulted in awards of damages not directly related to 'provable harm', in cases of failure on the part of the data controller).

- Invalid assumptions about “equivalent safeguards”; for instance, Canadian public sector organisations have long-standing concerns that, under the provisions of the USA PATRIOT Act, citizens' data under Canadian jurisdiction might be accessed by the US Government because it was out-sourced to US 'cloud' service providers.[7]
- Lack of clarity in the enforcement of Safe Harbour certification. The certification mechanism is essentially self-policing – in other words, if a data controller, whether through ignorance, carelessness or malice, fails to abide by the Safe Harbour rules, it's possible that this could go undetected, or that even in the case of a data breach, the opportunities for effective redress are limited.

The rule-based approach is what we end up with in practice; it suffers where there are regional differences in the attitude towards regulation as a valid control, and depends critically on effective enforcement if it is to be viable. However, the reality is that in the distributed, cross-border environment of modern data processing, effective enforcement is both challenging and rare.

'Justice'-based

Just as in other ethical debates, the shortcomings of consequential and rule-based approaches lead us towards a justice-based option. Here, we come back to some of the principles set out in the opening section of this paper: ideas of “no surprises” and legitimacy. In particular, it is worth unpacking the notion of legitimacy to look at its foundations, so as to assess whether it can take us further than the consequential or rule-based alternatives. Again, I want to distinguish between “legal” and “legitimate” - since legality is, unfortunately, no guarantee of legitimacy.

I suggest that legitimacy implies fairness; that even if something is legal, if it is manifestly unfair, then it is probably also unethical. Intuitively, if a data controller acts unfairly, they have the option of another course of action which would be more fair (but might, of course, benefit them less than the unfair alternative).

Legitimacy also implies lack of deceit; if the data subject is kept in the dark about what is done with personal data, it is hard to see how notions of consent and respect can be upheld. A criticism often levelled at social networks and their bedfellows, the targeted advertisers, is that users find their behaviour 'creepy'. I would argue that this is because users are deliberately under-informed about the collection and re-use of data about them, such that when the re-use becomes apparent, users are unpleasantly surprised by it. It is, of course, in the interests of social networks to lull the user into a feeling of confidential interaction with their peers. Users reveal more about themselves if they don't feel there is a third party listening in on the conversation.

As long as unfair behaviour is risk-free, it will always be tempting – especially if there is a commercial pay-off. Ethical data-handling therefore needs the support of a system in which bad behaviour can be detected and remedied.

With these factors in mind, what might a justice-based system of data-handling look like? Here is a shortlist of characteristics one could look for:

- A fair balance of the rights and interests of stakeholders;
- Transparency concerning motivations, risks and benefits;
- Accountability for behaviour and outcomes;
- Effective redress in case of failure;
- Above all, a focus on “should we do this?”, rather than “can we do this?”.

Notably, these are all non-technical criteria. That's not to say technology won't have a place in the solution, but if the principles aren't in place, all the tools in the world are unlikely to produce a good ethical outcome.

Proposal

We propose a facilitated, multi-stakeholder discussion on this set of topics:

- The principle of “no surprises”
- Ethical dilution
- Multi-stakeholder issues
- Cross-context issues

The discussion would examine the effectiveness of current data protection principles (collection, usage, retention, sharing) and proposals for user-requested deletion, and attempt two outputs:

- A candidate problem statement for ethical guidance on data handling in the domain of cyber-security research;

- Statements of ethical principles which extend beyond the cyber-security research domain, and provide more general guidance for ethical data handling in the modern, mediated digital environment.

We would also like to use this as the first step in the creation of a sustainable community of interest to continue work in this area, on problem definition and solution analysis. We welcome comments, suggestions, volunteers and follow-up contact requests.

Christine Runnegar
Robin Wilton

Internet Society
April 2014

References

[1] Council of Europe – The Margin of Appreciation:
http://www.coe.int/t/dghl/cooperation/lisbonnetwork/themis/echr/paper2_en.asp

[2] Jeremy Bentham – The Panopticon as an institutional architectural tool:
https://en.wikipedia.org/wiki/Panopticon

[3] Michel Foucault - “Discipline and Punish”
http://plato.stanford.edu/entries/foucault/

[4] Helen Nissenbaum - “Privacy as Contextual Integrity”
https://crypto.stanford.edu/portia/papers/RevnissenbaumDTP31.pdf

[5] Narayanan and Shmatikov – Robust De-anonymization of Sparse Datasets
http://www.cs.utexas.edu/~shmat/shmat_oak08netflix.pdf

[6] Prof. Sandy Pentland – Mobile data and behavioural predictions
http://hd.media.mit.edu/tech-reports/TR-680.pdf

[7] US Presses Canada To Relax Privacy Laws
http://www.huffingtonpost.ca/2014/04/03/canada-us-privacy-laws_n_5083034.html