DATA CRITIQUES

Request For Comments

The Challenges of Data Custody & A Testable Plan for Data Trust

Data Critiques,
June 2019

Community Collaboration Disclaimer: *We intentionally embrace an ethos that values the diversity of perspectives and experience of our community. This means that, while we focus our efforts on areas of shared critique, one or more of us may disagree with the stated community position on any given topic. We also allow ourselves, individually and as a group, to learn, debate, and change our minds as we progress.*

Table of Contents

Request For Comments

The Challenges of Data Custody & A Testable Plan for Data Trust

Table of Contents

Executive Summary
- The Data Trust
- Better Consent through Proxies
- Data Availability
- Data Portability
- Certification
- What Comments Are We Asking For?

Introduction

The Challenges of Data Custody
- Weaponized Data
- Several privacies, many freedoms
  - Value of data and its impact on privacy
- Usability Dilemmas
- Walled Gardens Limit Public Good
- Incomplete Datasets Lead to Biased AI

Data Trusts
- Formulating an approach
- Data Trust Explained
- Reducing Scope
  - Accepting Reality
  - Governance Tradeoffs
  - Technology Trade Offs & Considerations
    - Usability versus customization
    - Data loss, breach, or completeness
    - Homomorphic encryption is not a magic bullet
    - Claims, not facts
    - Redistribution of wealth
- Key Concepts
  - Custody, not Stewardship

- Meaningful consent
- Consent by design
- The Data Exposure Spectrum
- Consent Proxy
- The Mechanics of Consent
  - Separation of Data Collectors & Data Consumers
  - Consented Rules
    - Collection Rules
    - Access Rules
    - Usage Rules
    - Summarization Rules
    - Money Rules
    - Export and Deletion Rules
- Auditing is required

Making this real
- Use Cases
- Bootstrapping and The Sustainable Economic Model
  - Startup phase:
  - Operating costs:
  - First-mover costs
  - Conflicting incentives
- Incentives: why would anyone do this?
- Failsafes

Caveats

Appendix A
- Could consent proxies help us navigate privacy concerns?
  - Consent proxies
  - Requirements for effective consent proxies
  - Would consent proxies work in the wild?
  - To conclude

Appendix B
- Data Portability, Federation And Portable Consent
  - The separation of powers: protocol, platform and license
  - Portable Consent: An unavoidable complexity
  - Conclusion

Executive Summary

In this RFC we present our concept of a data trust as one way to overcome the main challenges of data custody and to provide **a way to make data available for the common good while safeguarding our various privacies**.

When it comes to the custody of data about us, we are faced with a plethora of failure modes. In a race to collect ever growing piles of data, corporations and governments have failed to make this data available to the common good and instead weaponized, exploited and enclosed it. Unfortunately, privacy policies coming into force today only address part of the problem. They focus on protecting a strict notion of individual privacy, without concern for the negative externalities to society at large of both over- and under-sharing data. Individuals, meanwhile, are faced with false choices - having to decide between opting into inappropriately permissive terms and conditions or leaving their networks of friends behind - and incomprehensible, obtrusive consent banners. Finally, enclosure of data mostly leaves us unable to challenge claims about us and add missing data to existing datasets, we are bound to create biased artificial intelligences.

The Data Trust

In our view, any meaningful solution to this set of challenges starts with a separation between those *holding* the data, and those collecting and/or using it. Enter a Data Trust: **A piece of infrastructure to separate companies that *collect* and/or *use*[1] (personal) data, from the *custody* of that data.** The custodian is a non-profit, acting under the direction of a trust. This infrastructure allows data subjects to control under what conditions data about them is used. It places a fiduciary responsibility on that custodian (the trustee) to act in the best interest of the data subjects, as well as those impacted by the sharing of data under the custodian's control.

In our design of a data trust we accept a number of real-world constraints and scope restrictions:

- We do not require any laws to be changed for our concept to be implemented.
- Implementing a democratic governance process within a data trust is out of scope; therefore, the trust will have a rule-making process that is permanent.
- We do not believe short-cuts (for instance in the form of homomorphic encryption or federated learning) are advanced enough to provide true privacy.
- We do not require human beings to have endless time and energy to spend on making decisions about their privacy. We call this the usability requirement.
- Any data trust needs to make explicit whether it prioritises prevention of loss, prevention of breach, or a guarantee of completeness. This is equivalent to the CAP theorem, for custody.

[1] See "Sara Ahmed: The uses of use"

- We treat data as a set of claims, not a set of facts. A data set can contain contradictory claims.

To navigate these constraints in light of our vision of a data trust, we rely on a few key components, as discussed below.

Better Consent through Proxies

To satisfy the Usability Problem, we propose to put consent proxies in charge of assessing the privacy risks of collecting and sharing data in specific contexts. Consent proxies would be trusted entities with domain specific knowledge charged with the creation of consent profiles that cover specific types of data in that domain. For instance, a trade union could function as a consent proxy for workers. Data subjects, in turn, can then adopt various consent profiles as their own. The use of a consent proxy makes it reasonably straightforward for data subjects to understand what they consent to and change their minds down the line.[2] As well as defining and updating consent profiles, the consent proxy is tasked with negotiating the "Terms of Service" (or EULA) agreements between the Data Trust and Data Subjects. These agreements cover what data can be collected, stored and shared. It falls on the data custodian to ensure that data users adhere to the rules deriving from these consent profiles, through access control and periodic audits.

Data Availability

Through a data trust, data that is currently enclosed by single platforms would be made more widely available. In principle, if company A can license specific data for a specific purpose and duration, then company B should be able to do so as well. In addition, data could also be made available to researchers or governments. Of course, the license attached to the data and specific access rights depend on the consent profiles adopted by the data subjects, negotiated in balance against general concerns regarding the impact of such data sharing on society at large.

Data Portability

In order to ensure that individuals can move freely between data trusts, it is not enough to ensure that the data within the trust remains portable - we need to ensure that the granted consent (and the details of the rules) are portable as well. We do so by making consent portable. We conceive of a **Consent Rules Engine** and associated grammar that provides a cascading hierarchy of consent statements. Of course, true portability also requires data subjects have other data trusts to which they can move their data. We believe we can achieve a federated and portable ecosystem of data trusts by separating Protocol, Platform and License.

[2] For more on consent proxies see Appendix A

Certification

In order to be able to trust the data trust, the data custodian should be required to obtain certification from an independent body. If it loses its certification, it should no longer be allowed to create data trusts - in any jurisdictions - and any existing data trusts should be dissolved (with data moved to other trusts). A healthy ecosystem of data trusts therefore also requires that no single data custodian ever holds a large enough share of the entire number of data trusts in a jurisdiction.

What Comments Are We Asking For?

We are looking for general feedback and comments on the ideas presented in this RFC. In addition, there are still a number of open questions we would like your input on, specifically:

- **Incentives**: Are the moral, technical and financial incentives strong enough for meaningful participation, without regulatory pressure?

- **Liability**: Is the liability currently placed on the company/entity holding data reduced by offloading data to a trust? Is it reduced enough?

- **Governance: Can we avoid voting?** At the moment our design relies on low barriers of entry for the creation of data trusts, certification of data trusts, as well as low switching costs for data subjects as the main ways to ensure data subjects are sufficiently represented by the data custodian, without the need for democratic processes to govern control over data. Is this enough to prevent evil data trust empires from emerging? What other models could we envision? For context, the representative interests are:
    - Trustees for subjects and impacted parties
    - Data Subjects
    - Impacted non-subjects

- **Failure Modes?**
    - Can we develop a “FAIL-SAFE” Model for this trust, where it can “degrade gracefully”?
    - What happens when faith on data trust is eroded?
    - How do we prevent one data trust to gain monopoly power? Are our ideas around portable data and consent sufficiently strong to mitigate this risk?
    - How do we ensure survival of data trust does not come at the expense of ethical data sharing?
    -
- **Consent over shared data:**
    - For simplicity, we have assumed a model of “joint account with sole survivorship”, which implies that shared data is wholly owned by each party. Is this enough? Or are new kinds of consent required to broker the relationship between data parties? If Kim messages Peter, can Kim then take a copy of those messages with them as they leave the trust? If the trust loses its certification, who has rights over those messages - if anyone?

    - Who controls data that is not attached to an individual/entity? Who moves it from trust to trust or selects consent profiles for it? E.g. data about car movements in cities. Data that is somewhat personal, but for which it would be too cumbersome to figure out which person it's about.

Introduction

The last decade has been characterized by the weaponization and exploitation of *personal data* for commercial interests. In addition to the ethical and socio-political impacts of this dramatic concentration of wealth and erosion of our very notions of (and opportunities for) privacy, the ongoing concentration of data into the hands of a few will undermine any possibility of collaborative use[3].

> "Building up huge stores of data inherently weaponizes them."
> https://twitter.com/lkanies/status/1116791694026928128

What would an alternative look like? This RFC describes a testable blueprint for the "minimum viable data trust". It has a set of interconnected legal and technical elements that are intended to allow each actor to behave only in their own natural self-interest.

What is a Data Trust? Put simply, it is a mechanism to separate companies that *collect* and/or *use*[4] personal data, from the *custody* of that data. Because the **custodian** is a non-profit, acting under the direction of a legal trust, it allows the subjects of the data to control under what conditions data about them is used, and places a fiduciary responsibility on that custodian (the trustees) to act in the best interest of the data subjects, as well as those impacted by the sharing of data under the custodian's control. We call these parties "**data collaterals**", whether they are individuals or other externalities such as the environment.[5]

There is a well-defined set of rights that people individually, and societies collectively, expect to have in relationship to data about them.[6] This has been, and continues to be, defined in regulation. But there are inherent challenges to ensuring these rights only through regulation, and there are key issues with the *usability* of many other proposed solutions.

At the moment most concepts of data trusts are still highly theoretical. Our design is no exception. However, with this RFC we hope to obtain sufficient input and feedback to create testable hypotheses and break our design into components that we can build and test in the wild. We very much welcome collaborations, criticisms and alternative visions!

---

[3] For example of for-good uses of data, see McKinsey 2018
[4] See "Sara Ahmed: The uses of use"
[5] See also "Friendly Fire" and "Splash Damage".
[6] See The Bill of Data Rights

The Challenges of Data Custody

A data custodian needs to balance the societal value of data against a myriad of privacy concerns. Here we discuss the key challenges to overcome.

Weaponized Data

There is a temptation for those holding valuable datasets to weaponize that data for either profit or surveillance (control). Designing systems that prevent valuable data from being abused is hard. Given the emergent shape of the utility curves of data, there will always be a future state where any for-profit entity will have a strong financial and fiduciary motivation to subvert any regulations that are limiting their ability to profit off of the data they hold. Therefore, data Custodians HAVE TO BE non-profit entities.

Several privacies, many freedoms

When we talk about privacy, we often refer to an individual's ability to stay secluded. In reality, there's a bit more to it than that. Helen Nissenbaum defines privacy as the "appropriate flow of personal information". Meaning that data about us can be shared, but it should be done appropriately and in accordance with the contextual norms we hold. For instance, you may expect to be overheard telling a story in a restaurant, but would not expect - or find it acceptable - for that story to then be shared on Twitter.

We hold that:

- Privacy is context specific: A piece of information shared in one context may be relatively meaningless, while in another context - or connected to another piece of data - it could reveal someone's identity.
- Privacy is not always an active choice, as information about us can be inferred against our will. Doing so can also affect our future privacy, as data about us today can be employed to make predictions about our behaviors tomorrow.
- Privacy concerns exist on multiple levels, ranging from the individual to the collective.[7]

We specifically identify four ethical dilemmas that the data trust should grapple with, perhaps with help from a regulator. Each of these ethical dilemmas emerges primarily from trying to balance the needs of the data subjects, with the impacts to data collaterals.

a. **The unborn child problem**
   When you share your DNA, you do not just share information about yourself, but also reveal information about your family members. What if Adam and Eve decided to share their DNA with 23AndMe. Their decisions would also reveal

[7] https://medium.com/@anoukruhaak/why-your-privacy-is-about-all-of-us-a1f7a9f0035

information about their as of yet unborn baby Robin, as well as any other future offspring. How do we balance Adam and Eve's interests and needs against those of Robin?

b. **Collective privacy problem**
   In late 2017, Strava, an app that enables runners and cyclists to track their exercise routes, accidentally revealed the location of secret US military bases around the world. It did so by aggregating all the routes generated by its users on a publicly available data visualisation map. The map did not reveal any information about any individually identifiable user. But, as some of those users happened to be US soldiers, their collective running patterns perfectly traced the formerly secret army bases. This is but one example of fully anonymised data revealing collective patterns that could be weaponized.

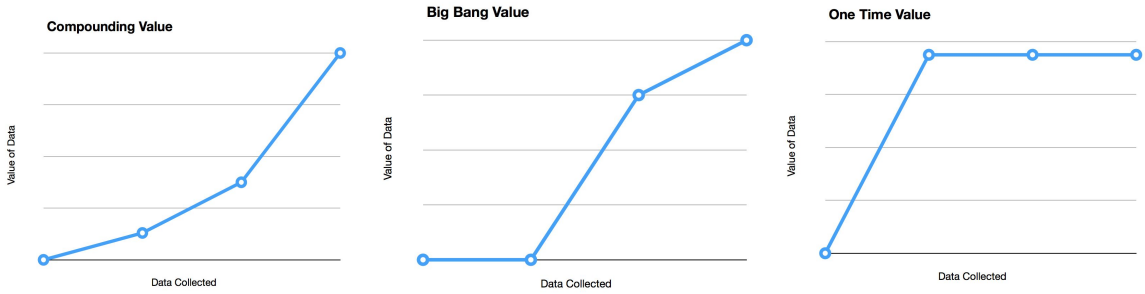
c. **The non-actor problem (aka opt-out problem)**
   Data sharing can also impact those who decide not to share their data. For instance, what if eighty percent of the population shares fitness tracker data with a health insurer in exchange for a discount on their premium? Suddenly we can make assumptions about the remaining twenty percent who have not chosen to do so. Are they less healthy perhaps? How will this inferred information impact their premium?

d. **The relational data problem**
   Closely related to the unborn child problem, the relational data problem refers to the situation where a piece of data is created by more than one subject. How do we govern this data? Do all the subjects have to consent to this data being used? Or just one (which would be similar to copyrights in the case of co-authorship)?

Value of data and its impact on privacy

The types of privacy related to the sharing of specific types of data depends on how and when that data becomes valuable[8]. For some types of data the value will show diminishing marginal returns. For instance, when the aim is to verify whether someone is old enough to drink, the utility of the first piece of data (e.g. their birth date) will hold a lot of value, while any subsequent piece of data will be essentially worthless. In other cases the value curve stays flat for a long time and spikes as a certain threshold of data collection has been reached. Agricultural data is a good example. Data on the production levels of a single farm tell very little, while data about half of all the farms in a country would reveal enough information about supply trends to manipulate markets. This is especially true for data that is used to train AI models - until the threshold for training is reached, the data is essentially worthless; once that threshold is approached, the entire dataset becomes extremely valuable, but there is little additional value for new data.

[8] https://www.schneier.com/blog/archives/2019/06/data_surveillan.html, by Bruce Schneier, June 2019



Much of the data now being used to power AI models was collected before the specific algorithms powering those models were developed. Thus, **the shape and eventual value of these data value curves are emergent and subject to change.**

Usability Dilemmas

"If no one reads the terms and conditions, how can they continue to be the backbone of the internet?" asks the New York Times editorial board in an article titled 'How Silicon Valley puts the 'con' in consent'[9]. Despite data protection laws becoming commonplace, we have yet to find consent models that do not rely on individual users blindly clicking 'Agree' or opting into tracking cookies they hardly understand.

First of all, consent is meaningless when saying 'no' is not really an option. We may not want Facebook to collect data about us, but we equally do not want to lose our friends. And so we click yes. We will address this problem further below.

Secondly, and more relevant to the question of usability, **our ability to consent to what data we share about ourselves should never rely on us having endless time and expertise.**

In the words of Maciej Cegłowski: "*The infrastructure of mass surveillance is too complex, and the tech oligopoly too powerful, to make it meaningful to talk about individual consent. Even experts don't have a full picture of the surveillance economy, in part because its beneficiaries are so secretive, and in part because the whole system is in flux. Telling people that they own their data, and should decide what to do with it, is just another way of disempowering them.*"[10]

Think about it, if you actively had to decide who you want to collect what data about you and for what purpose, you would need multiple lifetimes. These decisions would take time to contemplate. They would require extensive research into how data about you might violate your privacy when combined with other data about you that you are, as of yet, unaware of. Add to

[9] https://www.nytimes.com/2019/02/02/opinion/internet-facebook-google-consent.html

[10] https://idlewords.com/2019/06/the_new_wilderness.htm accessed June 2019

that the fact that sharing data about you may also affect other people's privacy and you start to see how unusable our current approach to online consent really is. In the words of Helen Nissenbaum: "*Proposals to improve and fortify notice-and-consent, such as clearer privacy policies and fairer information practices, will not overcome a fundamental flaw in the model, namely, its assumption that individuals can understand all facts relevant to true choice at the moment of pair-wise contracting between individuals and data gatherers."*[11]

We hold that we need to make reasonably sure that we can all participate in the services we create. In a similar way that we ensure government forms can be understood by most citizens and do take a limited amount of time to fill out. This criterion immediately invalidates many of the consent proposals we are presented with today.

Walled Gardens Limit Public Good

The problems of data being held and controlled by just a few extends beyond the risk of weaponization. In addition, data that *could* be put to work for the public good, is not available to, for instance, researchers. Making it available while safeguarding the multiple dimensions of privacy is a hard problem that is as of yet unsolved (at least in the real world).

Incomplete Datasets Lead to Biased AI

A proliferation of automated decision-making is increasingly replacing decisions made by humans. Unfortunately, the algorithms we use are all-too-often trained on incomplete datasets, that leave out data on entire swaths of the population[12]. The end results can be disastrous: from facial recognition system misidentifying people of color and women, to people of color receiving longer prison sentences because algorithms wrongly predicted their recidivism rates to be higher. And, remarks Ben Green: *"The issue goes much deeper than today's prevalent critiques that the training data behind predictive policing algorithms is biased due to a history of over-enforcement in minority neighborhoods—our very definitions of crime are the product of racist and classist historical processes."*[13]

Those making decisions about when and how to share data about us need the appropriate tools and agency to prevent undiverse and incomplete datasets being used for these purposes. In addition, we should all have the opportunity to challenge the validity of data about us.

[11] https://www.amacad.org/publication/contextual-approach-privacy-online
[12] https://www.schneier.com/blog/archives/2019/06/data_surveillan.html by Bruce Schneier, June 2019
[13] https://arxiv.org/pdf/1811.03435.pdf by Ben Green

Data Trusts

Below we outline how we propose to address the challenges mentioned above, with the help of a data trust.

Formulating an approach

When considering a mechanism to address these challenges, we draw inspiration from other researchers and organisations working on data trust. The Open Data Institute[14] considered data trusts as a piece of infrastructure, created for a single purpose (e.g. making food waste data available to NGOs). In their view, the role of the data steward (the trustees) would be to balance the interests and concerns of the various stakeholders to the trusts, including data creators, users and subjects.

Silvia Delacroix is pioneering bottom-up data trusts[15]: a more grassroots version of a data trust where people collectively decide to collect data about or by them in a trust, and generate rules for how that data can and cannot be shared. This form of a trust approaches the idea of a data commons and aims to rebalance power between data subjects and platforms.

Our concept of a data trust is closely connected to these ideas, but refocusses the role of the data custodian (the trustees) to be specifically responsible for safeguarding the interests and concerns of the data subjects, while mitigating collateral damage. It comes closer to the data trust proposed by, among others, Alphabet's Sidewalk Labs as part of a smart city project in Toronto[16]. This data trust was intended to, in addition to negotiating the interests of the various stakeholders, also determine what data can be collected and who would be allowed access to it. That said, details on governance and control of Sidewalk Lab's 'civic data trust' remain vague. In this RFC we aim to take a closer look at the possible mechanics of a data trust governed by a data custodian.

[14] See the ODI Data Trusts summary report.
[15] See https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3265315
[16] https://business.financialpost.com/technology/make-public-libraries-custodians-of-smart-city-data-board-of-trade

Data Trust Explained

Let's start with a story about Luna. Luna, a late-millennial, never warmed to the quantified self hype. She never felt a particular desire to know how many steps she had taken on any given day, or whether her heart rate was behaving as prescribed. But at 33, her sedentary lifestyle started to take its toll. Time for a Fitter.

'A wearable fitness tracker you can trust', the box reads, 'To get started, download our app'. Luna sighs. Years of Facebook scandals and data breaches have eroded her capacity for blind enthusiasm. But while her head is generating doubts, her fingers are typing 'Fitter' into the App Store search bar. Two minutes later she's in, ready to hand over information about her heart rate and calorie intake to a group of virtual strangers.

But then a popup appears: 'Select your data trust', followed by a clickable list of data trusts to choose from. Not the usual terms and conditions she was expecting. She clicks on the info button, which explains to her the basic concept:

"In order to make accurate health predictions, we need to access data generated with your Fitter. To ensure we do not abuse your privacy and to help you share data with others, we have opted to not hold it ourselves. Instead it will be held by a non-profit data custodian, a data trust, with a fiduciary responsibility to safeguard your privacy as well as that of others. You can choose from a number of certified trusts."

Luna randomly selects a data trust from the list. The description tells her this data trust has a proven track record, specialises in health data and is certified by the data trust authority. 'Good enough for me' she thinks and accepts. Another window appears. This one still looks nothing like the terms and conditions Luna has grown accustomed to. She's asked to select a consent proxy. Would she like to use the default privacy profile created by the Ministry of Health? Or would she rather go for the one set up by the World Health Organisation (WHO)? She decides to trust the WHO with decisions about her privacy regarding Fitter data. A few more questions. Is there any type of data about herself she would never want to have collected? Any institutions or companies she would never want to receive this data? Any data she would like to share with a specific institution? She's further told that her consent expires in a month, unless she opts for auto-renewal.

Months later, Luna's situation changes, she is pregnant. As she is unsure of whether she wants to keep the baby she is keen to keep both her pregnancy and her medical records private. Aware that her fitness tracker data might give her away, she starts to look into different consent profiles. "Which organisation is best suited to handle these types of privacy questions?" she wonders. After a bit of a search she settles for the profile created by Planned Parenthood. The profile goes into effect immediately.

Luna' story sums up the basic premise of a data trust. Instead of a commercial entity holding data about its users, the data is sent directly to a data trust, where it is governed by the data custodian. The data subject gets to decide which data trust to put in charge. Each trustee on the data trust board (collectively called the data custodian) represent different categories of data subjects, as well as data collaterals. Their job is to look after the best interests of the data subjects and consider the larger impacts of data collecting and sharing on society.

They do so with help of the consent proxies who advocate on behalf of the data subjects, usually focussed on specific interests. They can be any organisations you would entrust with decisions about your data and privacy. They may be different for different categories of data and people. For instance, someone with a specific type of as of yet incurable cancer might elect to have data about their health (and maybe even financial transactions) governed by an NGO that advocates for patients with that disease, knowing that they will understand both the risks and opportunities that stem from collecting and sharing this data.

Companies, researchers, or policymakers access the data in a data trust by obtaining a license that describes their entity (e.g. commercial, academic etc), type of usage, level of access, as well as duration of usage. It falls on the trust to audit the data users and ensure they behave in accordance with their license. The data trust will only make data available under a license when doing so does not violate the consent rules set up by the data subjects and consent proxies, as well as their own ruling about the impact of sharing specific types of data on society at large.

The data trust itself is certified by an independent body. Should it lose its certification, it is no longer allowed to share any data in the data trust. At that or any other point in time, data that is tied to a specific data subject can be moved to a different data trust.

Reducing Scope

Our blueprint is **testable** by design. We aimed to create a concept that could be implemented tomorrow without changes to regulation or statutory frameworks, and that, in some basic form, could be applied in most jurisdictions. This implies that we have accepted a set of constraints, which we see as the *Limits of the Current World.*

Accepting Reality

Data Trusts exist within a regulatory regime at the national and local level; they are not intended to supercede or subvert it. Essentially, having your data in a data trust will not protect it from Uncle Sam.

Governance Tradeoffs

Implementing a democratic governance process within a data trust is out of scope; therefore, the trust will have a rule-making process that is permanent.[17] We mitigate this with strong portability guarantees, and an easy mechanism to set up new, compatible data trusts. (Low barriers to entry and low switching costs). In addition, trust certification means trusts can be disqualified as trusted custodians if they don't meet a fixed set of requirements.

Technology Trade Offs & Considerations

Usability versus customization

The consent interface can either be usable for most data subjects, or completely customizable for any single data subject, but not both. We have to rely on a consent proxy to strike the right balance of usability and protection of rights.[18]

Data loss, breach, or completeness

We acknowledge that there is a fundamental technology tradeoff between prevention of loss, prevention of breach and guarantees of completeness. Each data trust will have to make EXPLICIT decisions about what compromise they are accepting.

Homomorphic encryption is not a magic bullet

Today's differential privacy and homomorphic encryption capabilities are not strong enough to support ad-hoc queries without potential violation of privacy guarantees. Data Trust Custodians must take responsibility for carefully providing summarized data interfaces, which will likely be limited to median and histogram-type reports.

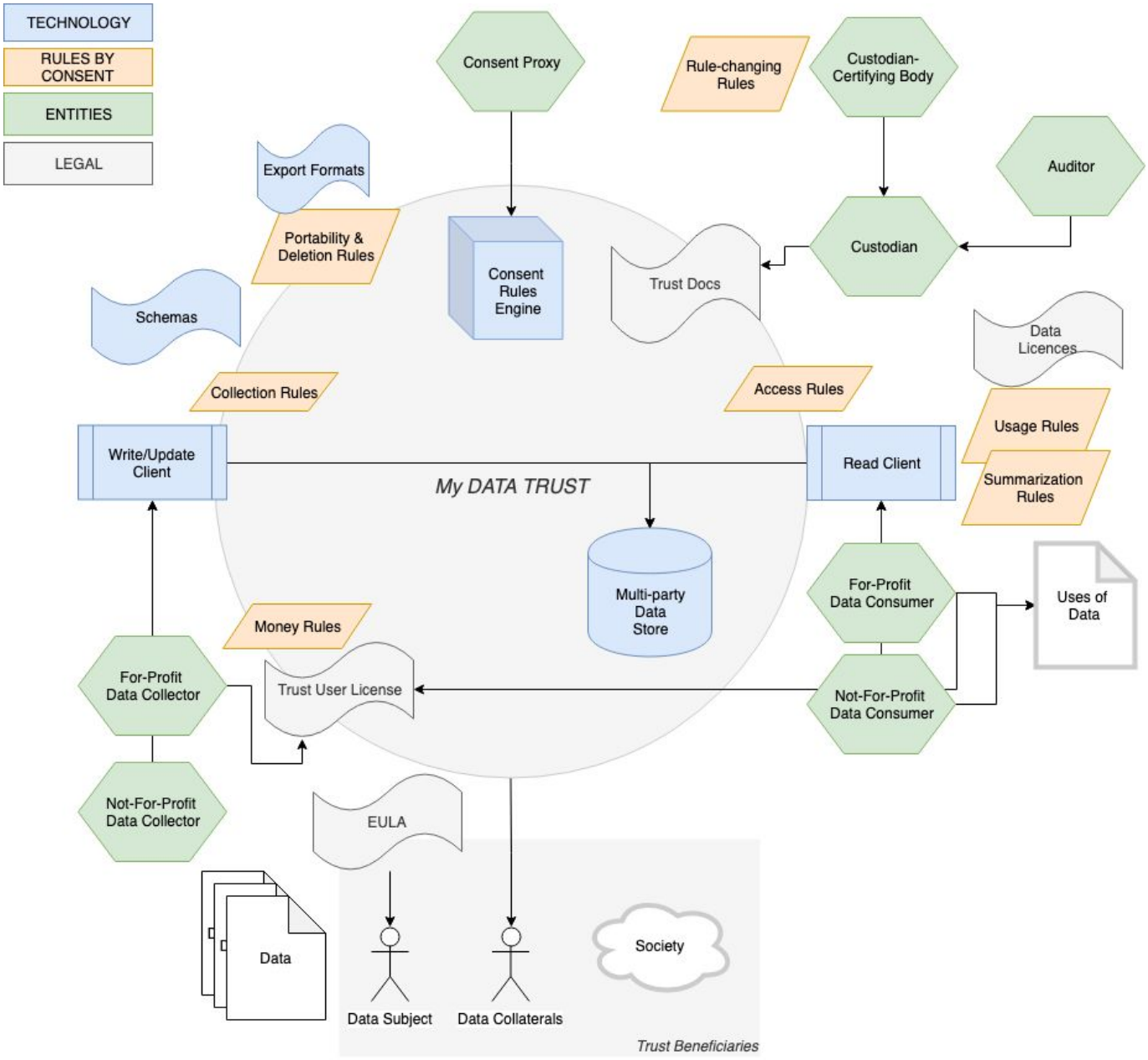
[17] See Ostrom's work on operational rights and constitutional rights.
[18] See the failures of Google+ Circles

Claims, not facts

Data is treated as a set of claims, not a set of facts. As a practical matter, this means that we should ALWAYS be able to store multiple (and potentially contradictory or competing) claims about the same “fact”. Emergent “truthiness” can be statistically produced if and when required, but the outlier claims have immense value.

Redistribution of wealth

Some of the activists involved in data privacy issues are primarily concerned with the economic concentration that is occuring. They see data trusts or other mechanisms as a way to capture and redistribute the wealth generated by this data. While data trusts could be used as an ingredient in such a plan, we have not aimed to address the various issues around such a “data dividends” approach. Our discussion of economic models is limited to addressing how the data trust itself could be sustained.


TECHNOLOGY
RULES BY CONSENT
ENTITIES
LEGAL
Consent Proxy
Rule-changing Rules
Custodian-Certifying Body
Auditor
Export Formats
Portability & Deletion Rules
Consent Rules Engine
Trust Docs
Custodian
Schemas
Data Licences
Collection Rules
Access Rules
Usage Rules
Summarization Rules
Write/Update Client
My DATA TRUST
Read Client
Multi-party Data Store
For-Profit Data Consumer
Uses of Data
Money Rules
For-Profit Data Collector
Trust User License
Not-For-Profit Data Consumer
Not-For-Profit Data Collector
EULA
Data
Data Subject
Data Collaterals
Society
Trust Beneficiaries

Key Concepts

Our concept of a data trust relies on a number of concepts that require further explanation. Below we briefly mention the most notable ones, in our view. We very much encourage any input towards further development of these concepts.

Custody, not Stewardship

The trustees of a data trust are considered custodians of the data, with a fiduciary responsibility to safeguard the rights of both data subjects *and any other person, group of people, or organism* that could be affected by the sharing of data in the trust. We call the latter group the data collaterals.

One of the advantages of a trust over a contract is that while contract law, such as that which governs a EULA (End User License Agreement), requires that all the parties to the contract are known in advance, trust law does not. Under trust law, we can treat unknown or future beneficiaries (such as unborn children or the global environment) as parties to the trust.

Our concept of a data custodian, moreover, differs from the notion of a data steward, as put forth by among others the ODI, which is generally viewed as an entity that negotiates the interests and constraints of data subjects, data holders (collectors in our grammar) and data users.

Finally, the Trust Custodian may have a mandate to increase the number and diversity of data collectors and consumers participating in the trust, but they should never become a data collector or consumer. In the long run, this would lead to conflicts of interest.

Meaningful consent

Within the context of a data trust, we differentiate between individual consent and consent by proxy. Following from the problems with current notions of individual consent outlined above, we hold that individuals can only give consent when four criteria are satisfied:

1. Individuals can reasonably be expected to understand the context and content covered by their consent. Privacy specialist Helen Nissenbaum has argued that for consent about data to be meaningful, we have to relate it to real-world events. What expectation of privacy do we have in the real world and how does that relate to what we should reasonably be able to expect online? Consent extended by individuals will only cover the elements of data sharing that can be intuited, while the more technical aspects of consent will be handled by a consent proxy (see below).

2. Consent is given enthusiastically, meaning individuals can decide against sharing data about them without it impacting their lives in any significant ways. Much the same way

someone can only be expected to meaningfully say ‘yes’ to sexual intercourse, when saying no would not yield any significant negative consequences.

A data trust helps mitigate this problem by making it easier to effectively share and move data (and consent expressions) between different services - imagine, for instance, individuals being able to move more freely between different social media. This would help users escape the walled gardens.

3. Consent once given can be updated and retracted. All consent to the data trust should be time-bounded and retractable. In addition, when data subjects move their data from one data trust to another their consent statements move with the data. This relies on common standards for (portable) consent. See Appendix B on Data Portability and Federation.

4. Consent is only meaningful when the organisations we consent to can be trusted to uphold our decisions and act in good faith. We expect a data trust to become one of those organisations. This trust should be ensured through certification of the data trust, combined with frequent audits as well as low barriers for entry for new data trusts.

Consent by design

When we think of privacy as the appropriate flow of personal data, it no longer holds true that data minimization is the holy grail. For instance, it is not hard to imagine that someone with an incurable disease may want to collect every bit of data about themselves, if that could benefit the finding of a cure. We should allow this person to do so, especially when their interests do not violate anyone else’s, or those of society at large.

In a consent by design framework, data can only be *collected* if such collection is consented to; this requires that the EULA governing data collection is between the data subject and the trust, *not* the commercial entity. There are an increasing number of regulatory frameworks (the GDPR first among them) that make this a desirable compromise for the company involved, as it offloads the regulatory burden (and potentially the liability) for data governance to the trust.

The Data Exposure Spectrum

Governance of data collection and access is a set of decisions that trade-off individual freedoms against societal rights.

We have applied a grammar of limits on data exposure. This is more than an access model, and conceives of restrictions on initial *collection* of data as well as guarantees of simultaneous and future access.

Step 0: Unknown (and uncollected)

Step 1: Unspoken (Collection is as per subject's consent, existence is not acknowledged)
Step 2: Secret (existence is acknowledged, no access other than subject)
Step 3: Private (Access and use is as per subject's consent, and can be revoked)
Step 4: Public
Step 5: SEC Public (guaranteed synchronized public access, ala wire notice)
Step 6: Immutable (Strong guarantees for future access)
Step 7: Viral Freedom (Derived and related data will also be public)

We believe that a Data Trust must support Step 0-4 at a minimum. Extending this blueprint to have copyleft characteristics requires consent rules that apply specific licenses to the accessed data, such as Creative Commons Share-Alike.

Consent Proxy

To not force every individual to become data privacy experts and make a million decisions about the data they share or is collected about them, we propose to put consent proxies in charge of assessing the context specific risks and privacy norms of sharing data. This would allow data subjects to adopt the consent profile that best describe their norms, interests and beliefs about the world. Data subjects should be able to select multiple profiles to describe different sets of data, or the same sets of data in different contexts. In addition, they should be able to change profiles easily. The use of a consent proxy satisfies the Usability Problem, making it reasonably straightforward for data subjects to understand what they consent to and change their minds down the line.

An ideal consent proxy is an existing legal entity with a strong relationship with the data subjects (such as a worker's union, public library or medical provider). As well as defining and updating consent profiles, the consent proxy is tasked with negotiating the "Terms of Service" (or EULA) agreements between the Data Trust and Data Subjects. These agreements cover what data can be collected, stored and shared.

For more on consent proxies see Appendix A.

The Mechanics of Consent

Portable Consent can be conceived of as a set of rules, each of which has a strong set of attributes that define what data this rule applies to, within what contexts, and for what duration. There are a specific set of enforcement points for these rules, which support the data freedoms discussed above (specifically the freedom to have uncollected data). We conceive of a **Consent Rules Engine** (based on the Open Policy Agent framework and the XACML attributes-based access-control model) and associated grammar that provides a cascading hierarchy of consent statements.

It is not enough to ensure that the data within the trust remains portable - we need to ensure that the granted consent (and the details of the rules) are portable as well. Similarly to how the internet was structured to achieve a federated and portable model, we believe we can achieve a federated and portable ecosystem of data trusts by separating Protocol, Platform and License.[19]

Just like permission bits in a unix filesystem, consent grants can be thought of as a set of bits applied to a set of groups. However, the groups are defined by attribute-based access, rather than role or membership. (E.g., all non-profits, or all b-corps, or the company named “CoverUs”, or all municipal governments within Estonia, etc.) Some example attributes would be:

- The data exposure spectrum level: E.g. uncollected (step 0) or private (step 3)
- The context that applies to this consent statement: E.g. this government can have access to my data as long as it is currently not a dictatorship. Or, my DNA data can be public as long as it does not indicate that I will soon die based on today’s medical science.
- The duration of consent, after which they expire. There is a maximum duration of consent that can be declared, based on the underlying data type. (See relevant informed consent regulations e.g. dentistry, etc).

```
{
 If: user_is: organization_type: B-Corp, AND
 usage_activity_is: scientific_research, AND
 user_is: certified_as: COMPETENT_SECURITY, THEN
 allow: 30 days access to deidentified_data, AND
 allow: aggregate queries with differential privacy level 5;
}
```
[20]

Separation of Data Collectors & Data Consumers

In today’s data architecture, we assume that the same entity that collects your data is the one that controls access and usage of it. Thankfully, we don’t have to make that assumption in the context of Data Custody. We have conceived of data collectors and data consumers as being separate licensees of the data trust; this allows us to easily support organizations who will only act in one or the other of these roles; to consider models for data trusts where some activities are for public benefit and without fees, where others are strictly commercial. At a technical level, we describe the “Read Client” and “Write and Modify Client” as being separate components; they also have different license keys and very different requirements for audit.

---

[19] For more on consent engine mechanics, see: https://docs.google.com/document/d/1ASPvCyTw8FIzVWeK2MJooIHLFbVVekb8QXOQHtZzXM8/edit#

[20] See grammar from OPA link here: https://www.openpolicyagent.org/docs/v0.10.7/how-do-i-write-policies/ https://github.com/open-policy-agent/contrib/blob/master/data_filter_example/example.rego

Consented Rules

Following the framework of the Bill of Data Rights[21], we have identified the following enforcement points:

1. Collection Rules
2. Access Rules
3. Usage Rules
4. Summarization Rules
5. Money Rules
6. Export and Deletion Rules

Collection Rules

What data can actually be collected? We hold that data, once recorded, can never be effectively deleted. Therefore, there are types of data we should never collect in the first place. This is especially true for data that can never be changed. For instance, we can change our address, but we cannot change our DNA. Once you surrender your DNA, it is known, never to be unknown again.

Access Rules

Access rules determine what entities can access the data under what conditions. It is important to treat access to the data store as a separate privilege from the underlying use of the data. Access Rules also limit those to whom we will admit the existence of particular data sets, whether or not they have the privilege to query or fetch data from that set.

"Metadata equals surveillance; it's that simple." **- *Bruce Schneier***

Usage Rules

Usage rules limit to what use data can be applied. They can also (similarly to open source) apply so-called "viral" rules to derived works and systems. Whereas access is relatively straightforward to limit and police, usage (especially derived use, including training AI models or the sale of summarized data products) will not be simple to monitor.

Summarization Rules

"Collections of personal data have emergent properties that aren't present in observations on a single individual."[22]

Data can almost never be adequately anonymized in a meaningfully large dataset, since correlation will typically undo any attempts to de-identify individual data subjects. A useful

[21] https://docs.google.com/document/d/125dzNMcHOT3ZvYT3jGHb_ZRKDW2DJy2Xmgf3pKdo2Qc/edit
[22] https://www.alexpghayes.com/blog/consent-in-the-presence-of-correlation/

workaround is to allow queries that only return “Summary” results - counts, averages, means, standard deviation, etc. However, a series of these queries can still “leak” identifying data - typically a “privacy budget” is used to estimate how many queries can be tolerated, and to inject “noise” into the summary results. (See current work on Differential Privacy). There is much more work to do in this area - today’s best tools of differential privacy, privacy budgets and the analysis of correlation-safe privacy measures are all likely insufficient.

Money Rules

In a perfect world, we would all enjoy economic and social freedom and therefore never see ourselves forced to sell our body parts, or our right to privacy. Unfortunately, our current reality does not reflect this utopia. In addition, as data about one person often also reflects behaviors of another, you selling data about you could have adverse consequences for all of us. Therefore, we hold that individuals should never be able to *directly* sell their right to privacy. However, we could imagine consent proxies making ethical decisions about what data can and cannot be sold, with the data subjects receiving a share of the pool of resources that are won through this selling of data.

Export and Deletion Rules

Archival guarantees are the ones that can’t be changed. To put it bluntly, donating a kidney is acknowledged to be an irreversible transaction. Similarly, we acknowledge that preserving the right to export as well as the right to delete is impossible - one right must be sacrificed for the other. The responsibility of the data trust is limited to guaranteeing the deletion of their COPIES of the data, and the revocation of any use consents.

(This is a difference of opinion from the “blockchain” communities, who believe that somehow, magically, using the word blockchain will allow them to maintain perfect remote control over all copies of all data, everywhere.)

Auditing is required

To ensure these rules are enforced, it falls on the data custodian to regularly audit the data consumers. We do not believe this responsibility can be (entirely) replaced by technology and instead will require compliance professionals to review the usage logs and access environments of the data consumer, as well as related business activity. It is especially important for auditors to understand the impact of data licensing on derived works, and to be alert for unconsented commercial exploitation.

Making this real

Use Cases

Data Trusts, as conceived in this RFC, are particularly-well suited for situations where there is a natural conflict between the data subject's privacy, and potential benefits for society at large. Some obvious and non-obvious examples of this would be:

- Personal health data, including fitness and activity monitoring
- Agricultural data, including crop performance and resource use
- Sentiment data, especially political but also on other key social issues

Bootstrapping and The Sustainable Economic Model

Startup phase:

We see two possible scenarios: either the data trust is initiated by a company wishing to offload user data onto a data trust, or by a government or NGO wishing to keep data about citizens out of the hands of corporations. In the latter case the startup costs would likely be financed by that government body, in the former case funding would come from the initiated company.

However, to avoid the initial funding establishing power dynamics between the data trust and the initiating company, we need to ensure the startup costs of data trust are spread out over any future beneficiary. One way to achieve this is by creating the data trust from a loan payment from the initiating company. Any other collector that would like to join the data trust later on would take on part of the loan. The loan would be paid off with income from the licensing fees from future data users (including the collectors themselves).

Operating costs:

Operating costs should be covered by licensing fees, or in the case of a public data trust, ongoing government financing. It's important to make the distinction between data trusts that serve a specific public benefit and those that, in addition, also have a commercial function. The commercial viability of a data trust should never be a requirement for its existence, as data trust could also be imagined to serve a pure social benefit that could never be monetized. This is especially true when the users of a data trust are low-income organisations, like universities or journalists.

Part of the operating costs is a liability insurance, which would allow for the data trust to be held accountable in case of a data or consent breach.

First-mover costs

The first data trusts will have to incur costs for research and legal drafts that could be partially copied by subsequent data trusts. This is similar to the effect that the open source FAST[23] and SAFE[24] documents have had on the early startup fundraising processes.

Conflicting incentives

One potential problem with this business model is that it relies on income from data users. This could create a perverse incentive, where the data trust - even though it is a non-profit - would see itself forced to navigate between income generating activities to guarantee its survival and the privacy concerns of the data subjects it has a fiduciary responsibility to protect. At the moment, we envision the certification of the data trust could help ensure that its fiduciary responsibilities are always prioritized. However, we are open to other models (e.g. risk or income pooling between data trusts).

Incentives: why would anyone do this?

- Offloading liability: with data protection laws becoming commonplace, holding large amounts of personal data increases the liability of the company holding it. Especially for smaller companies, becoming data protection law compliant is a relatively cumbersome endeavour. Switching to a data trust model could be a way to offload (some of) that liability and transfers costs of compliance onto the trust.

- Data pooling could give early-stage startups a competitive edge. By offloading data to a trust, startups are no longer able to attract investors by promising returns on the valuable data they will collect in the future. However, we could imagine a scenario where startup accelerators offer an opportunity for their startups to pool all data in a trust that they can access first. Thereby giving them a competitive advantage over other startups. One way this would work is if licensing for usage by a commercial entity required that entity to also be a contributor to the data. Possibly combined with a time-delay on access to the data for different types of licensees.

- Protecting data about citizens: when a corporation is hired by a government to collect data about citizens (as is often the case in smart city projects) the government has an opportunity to mandate that all data collected is held in a data trust.

- Governments promoting public data: grants stemming from institutions with a mandate to serve the public interest have an opportunity to require that all data collected through the grant-funded projects is stored in a data trust and, through the data trust, be made

[23] https://fi.co/fast
[24] https://www.ycombinator.com/documents/

available for research and policy - for as long as doing so does not violate the consent agreements the data is governed by.

Failsafes

In much the same way that the data custodian represents “adult supervision” of the data collectors and data users, the certification body also represents “adult supervision” of the custodian. If the trust itself “becomes evil” for some definition of evil, it can lose its’ certification which *(presumably) would automatically trigger the end of consent for many/most of the data subjects. Consent proxies could either work within their collective bargaining role to bring the trust back into compliance, or they could migrate the data elsewhere.

Loss over breach: in a scenario where the protection from breach is more critical than protection from loss, the trust can employ strong encryption of all subject data. Given the collapse of the data trust, this data would become essentially lost.

Caveats

No plan such as this will satisfy all critics. We have attempted to strike a balance between *personal* freedoms, and *societal* freedoms. (Consider, for example, the use of “my” DNA data for medical research, the potential (mis)use of that data for pricing or limiting my access to insurance, and the downstream/sidestream impacts on my children or my siblings and other relatives.) When such tradeoffs are unclear (and they almost always are), our rule is simple: the choice that can be reversed in the future is better than an irreversible one.[25]

[25] https://en.wikipedia.org/wiki/Precautionary_principle

Appendix A

Could consent proxies help us navigate privacy concerns?

"If no one reads the terms and conditions, how can they continue to be the backbone of the internet?" asks the New York Times editorial board in an article titled 'How Silicon Valley puts the 'con' in consent'[26]. Despite data protection laws becoming commonplace, we have yet to find consent models beyond those that rely on individual users blindly clicking 'Agree' or opting into cookies they hardly understand. Online consent is severely broken and the wreckage extends beyond the impossible-to-navigate consent windows. Helen Nissenbaum argues that the model of notice-and-consent is inherently flawed as we can never fully understand the repercussions of data about us being used in different contexts: *"Proposals to improve and fortify notice-and-consent, such as clearer privacy policies and fairer information practices, will not overcome a fundamental flaw in the model, namely, its assumption that individuals can understand all facts relevant to true choice at the moment of pair-wise contracting between individuals and data gatherers."*[27]

One of the underlying problems, Nissenbaum observes, is that consent is not meaningful when we cannot trust the system in place to protect our rights. We can, for example, meaningfully consent to have a surgeon operate on our kidneys; not because we have a perfect understanding of what such an operation might entail, but because we trust the medical system works and protects us. The only decisions left for us are not the specific way we want to surgeon to cut our kidney, but whether we have any problems with blood transfusions or resuscitation. Decisions that depend on our norms and beliefs; things we understand.

"*Choosing is not mere picking but requires that the subject understands that to which he or she is consenting".* Especially when considering raw data, like the number of keystrokes, or seconds spend staring at a screen, context is missing. The data lacks sufficient meaning to allow us to understand possible privacy implications of sharing such data. Only once we think through specific use cases and combinations of data do we start to understand how sharing it could help or hinder us. Such an analysis, however, requires extensive time and domain knowledge.

But we should not have to be experts to make decisions about how data about us is collected and shared. Likewise, we should not have to trust Facebook to make those decisions for us. One solution is to play safe, and severely limit the data that is collected about us. This is the approach taken by the General Data Protection Regulation (GDPR) that came into force in

[26] https://www.nytimes.com/2019/02/02/opinion/internet-facebook-google-consent.html

[27] https://www.amacad.org/publication/contextual-approach-privacy-online

Europe last year. Whilst an improvement, this defensive attitude towards data sharing does not help us in circumstances where we might *want* to share more data about ourselves: where doing so is in both our personal and collective interests.

Consent proxies

Enter consent proxies: organisations to make consent decisions on your behalf. These could be any organisation you trust to make decisions for you on specific types of data about you. For instance, someone infected with HIV might have specific privacy preferences, ranging from a desire to keep this information private, to wishes to help research towards better treatments. HIV advocacy bodies would potentially be well-suited to understand the context necessary to navigate these concerns, which are both complex and interlinked. In a similar vein, trade unions could be well-positioned to create consent profiles that would specifically deal with data collection and sharing relating to future employment, or wage calculations.

Consent proxies, therefore, are organisations that hold expertise in specific domains (e.g. health, mobility, human rights) and use that expertise to draft consent profiles regarding data collection and usage that individuals can adopt as their own. These profiles reflect the ethical considerations and risk assessments performed by the consent proxies, as well as the values and norms the proxy stands for. The trust placed in consent proxies is founded in their proven capacity for understanding the specific norms, values, needs and expectations of a specific demographic, given a specific context.

Requirements for effective consent proxies

In order for a consent proxy to effectively address the problems laid out above, it needs to fulfill a number of roles and requirements:

1. Consent profiles should be usable
The user should be presented, by the proxy they are entrusting with their data decisions, with a menu of possible consent profiles. The consent profiles should provide an easy overview of the main norms, functions and goals that underlies them, as well as a clear overview of the types of data the consent profile cover. It thereby replaces the spider web of choices users are currently faced with (for instance in the form of Facebook's privacy settings). Underneath this easy-to-navigate profile sits a fine-grained set of rules that govern the relationship between the platform users and those collecting, using and sharing data about them.

2. Variety of consent
To reflect the fact that human beings hold many identities that are subject to change, we should be able to a) choose from a wide range of consent proxies and b) elect different consent proxies to govern different types of data, as well as different types of concerns. For instance, we might want to have health data about us governed by both our GP and the Diabetes Foundation. As such, different profiles could end up governing the same piece of data. In those cases it makes sense for the most stringent privacy preferences to become the default, unless otherwise

stipulated. In order to automatically negotiate the differences between various profiles, their rules should be available in a computer readable format. The result of these various compatible data proxies and profiles, working in concert is a set of API rules dictating what data can and cannot be collected, stored and used by data users, and under what conditions.

3. Low barriers of entry and exit
As our values and circumstances are subject to change, we should be able to switch between consent proxies easily. This, in addition, allows us to vote with our feet.

Would consent proxies work in the wild?

At this point, the careful reader will wonder why any platform or service would agree to have their activities curtailed by consent profiles, when at the moment they have free rein. Fair question. For the moment, consent profiles are likely most relevant to services that rely on their users trusting them with data about them. It could, for instance, be a solution for companies that allow users to make data about them available to third parties. What if citizens would like to make location data about them available to local governments to improve flow of traffic? Or perhaps salaried workers would like their salary data available for research into gender disparities? A consent proxy could be a great way to help them better share data about themselves, while ensuring the way this data is used corresponds with their values and privacy expectations.

But to make consent proxies work for the average internet user, the creation of profiles alone will not be enough. Consent proxies would be toothless without ways to enforce the profiles they create. Their power relative to, for instance, a social media platform depends on their ability to steer individuals away from platforms that do not comply with their consent profiles. At the moment that is unlikely to be the case. We may not want to hand over data about us to Facebook, but we are equally unwilling to give up our social network. This problem will not be solved by a consent proxy, but will instead require true data portability - the ability to take data about us from one platform to the next.

A number of solutions to level the playing field have already been proposed, ranging from decentralized web technologies (e.g. Solid, Holochain, MaidSafe) to data trusts. While still in their infancy, these technologies and infrastructures promise to shift the power from the creators of closed platforms, back to the consumers of those platforms. Once the playing field is sufficiently leveled, a consent proxy - or consortium of consent proxies - would hold enough collective bargaining power to alter the behavior of data collectors and users. Equally, consent proxies could meaningfully advise against using a platform that fails to adopt any of its profiles.

To conclude

Consent proxies will not be a magic bullet, but rather part of a range of infrastructural solutions that together pave the way for better data sharing while safeguarding individual and collective privacy. Of course, there are a number of remaining questions to work out. Who would cover the

costs of setting up the consent proxies and creating the consent profiles? How do we ensure that the organisations we trust to become these proxies have enough understanding of data (in addition to their understanding of their specific domain)?

One of the great failures in the current data and privacy debate is the users who are routinely set up to fail; expected to fend for themselves and make wise data choices with too little information, even if they have the understanding. Much like we do not need to ask whether the water we drink is clean, everytime we take a sip, we should not have to evaluate a wide range of privacy concerns everytime we log into an app. Trusted consent proxies, with area-relevant expertise, would be one way to relieve this burden.

Appendix B

Data Portability, Federation And Portable Consent

Online consent is broken. The enclosure of user data by social media platforms and services have resulted in power imbalances that undermine any meaningful notion of consent. Our ability to freely choose how and when we share our data breaks down when the 'choice' is between surrendering data about ourselves or social exclusion. A different reality is possible: one where we can leave a social media platform, and take our social graph and data with us; one in which we can vote with our feet, and thereby change the power dynamic between us and the platforms we rely on for our everyday communication.

The internet itself was conceived of as an “open” platform - where open, in this case, means that the legal entities, the technical solutions used, and the future classes of content, can all be evolved independent of each other. This combines **portability** and **federation**. Portability refers to the idea that any single website or network-connected device can be moved between existing entities. Federation means that new websites and devices can be connected without requiring the participation of existing entities. Taken together, the combination of portability and federation can provide a robust set of protections for individual and societal freedoms.

Under the General Data Protection Regulation that came into effect in Europe last year, individuals gained a right to data portability. We can now decide to leave Facebook and take our personal data with us to another service. In theory, we are no longer beholden to one social media platform or banking app and are free to break out of the walled gardens that have held us hostage. In reality, we rarely exercise our right to data portability. For starters, it's hard to actually obtain the data about us. Additionally, data derived from one service is often not directly usable for another.

What mechanisms and infrastructure need to be in place to realize true portability and federation?

The separation of powers: protocol, platform and license

We can divide the problem of providing portability into three segments: **Protocol**, **Platform**, and **License**[28]. We argue that data portability and federation requires each of these segments to be created independent of the others, such that no single segment can ever prescribe rules to the remaining two.

[28] This is analogous to http, httpd, and GPL in the internet space.

Protocol describes the relationship between data and the set of explicit consents that have been granted on that data. Take the git protocol as an example, which describes the history of changes to a data set and the authors of those changes.

Platform is envisioned as a server environment that supports real-time API-driven access to data sets, mediated by a rules engine that conforms to the Protocol. This includes a set of read and write clients, as well as standalone policy engines that enforce pre-collection permissions and post-read usage and summarization consent. All of the components of the platform are designed to allow robust audit capabilities, and a tamper-proof activity log.

To return to the git example, this open protocol works with various platforms, like GitHub and GitLab. If users of one of these platforms stop trusting in its governance, they can easily take their code base to a different platform. In the extreme case, they can run a platform themselves.

License provides a permanently-attached set of limits on usage and derived work, and functions exactly the same as copy-left or permissive licensing in the open source world. Note that extended permutations of the copy-left principles (including licenses that explicitly prohibit commercial activity) are likely to emerge in this space. Additionally, licenses and other legal frameworks are the only portion of a data trust structure that remains ATTACHED to the data after it is exported; thus, they are critical to protecting the integrity of the consent(s) during migrations.

Portable Consent: An unavoidable complexity

When transferring data between services, or organisations, we need to be able to ensure that the privacy statements attached to the data are upheld. In other words, if data is moved from one social media to another, a user should be able to trust that the privacy settings agreed to on the first platform, equally apply on the second. This means that the mechanisms for managing consent need to exist at the protocol and license layer, decoupled from the platform.

In order for protocols to provide meaningful portability guarantees, there must be at least one platform available that implements this protocol (such as GIT for the git protocol, httpd for the HTTP protocol) and that platform must be a) operated by at least two different entities, and b) be relatively simple for an additional entity to operate. Any data trust operator could swap out the underlying platform without breaking portability, provided the same underlying protocol was supported.

Moreover, the entity that defines and evolves the protocol should ideally be separate from any of the entities that operate these platforms - much in the same way the W3C is separate from any specific browser.

Conclusion

Each of the three aspects of portability represents a check or balance on the others. Licenses and protocols can both be abused to tightly couple data to a specific platform; similarly, protocols can be structured in a way to limit the potential licenses with which they could be used [29]. Users and their consent proxies should remain free to select the license of their choice, unconstrained by platform or protocol limitations.

[29] See https://en.wikipedia.org/wiki/Tent_(protocol) and http://xanadu.com/xuTco.html