5 Personal Data and Individual Access Control

A key ethical dilemma regarding personal information is data asymmetry. Our personal information fundamentally informs the systems driving modern society but our data is more of an asset to others than it is to us. The artificial intelligence and autonomous systems (AI/AS) driving the algorithmic economy have widespread access to our data, yet we remain isolated from gains we could obtain from the insights derived from our lives.

To address this asymmetry there is a fundamental need for people to define, access, and manage their personal data as curators of their unique identity. New parameters must also be created regarding what information is gathered about individuals at the point of data collection. Future informed consent should be predicated on limited and specific exchange of data versus long-term sacrifice of informational assets.

There are a number of encouraging signs that this model of asymmetry is beginning to shift around the world. For instance, legislation like The General Data Protection Regulation (GDPR)[liv] is designed to strengthen citizens' fundamental rights in the digital age and facilitate business simplifying rules for companies by unifying regulation within the EU. Enabling individuals to curate their identity and managing the ethical implications of data use will become a market differentiator for organizations. While some may choose minimum compliance to legislation like the GDPR, forward-thinking organizations will shift their data strategy to enable methods of harnessing customer intention versus only invisibly tracking their attention.

We realize the first version of The IEEE Global Initiative's insights reflect largely Western views regarding personal data where prioritizing an individual may seem to overshadow the use of information as a communal resource. This issue is complex, as identity and personal information may pertain to single individuals, groups, or large societal data sets.

5 Personal Data and Individual Access Control

However, for any of these scenarios it is our candidate recommendation that policy should be created that:

- Allows every global citizen/individual access to tools allowing them control over a minimum common denominator of attributes that define his/her identity.
- Allows the possibility for citizens/individuals to access, manage, and control how their data is shared.
- Provides easily understandable ways for citizens/individuals to choose how or whether to share their data with other individuals, businesses, or for the common good as they choose.
- Provides for future educational programs training all citizens/individuals regarding the management of their personal data and identity, just as many countries provide training in personal finances and basic legal understanding.

We realize there are no perfect solutions, and that any digital tool can be hacked. But we need to enable a future where people control their sense of self. Augmented and virtual reality [lv] will soon provide lenses through which we perceive the world. Virtual worlds and social networks already blend our online identity with our physical sense of self. Autonomous and intelligent systems will apply virtual identities that impact the physical world.

Our goal is to champion the tools and evolved practices that could eradicate data asymmetry today to foster a positive image for our future.

5 Personal Data and Individual Access Control

Section 1 – Personal Data Definitions

The following definitions, resources, and candidate recommendations are provided to realign the systematic tracking, distribution, and storing of personal data to overtly include individuals and their predetermined preferences in the process.

Issue:

How can an individual define and organize his/her personal data in the algorithmic era?

Background

Personal data needs to embrace an individual's definition and clarification of his/her identity, mirroring unique preferences and values.

Candidate Recommendation

Where available, individuals should identify trusted identity verification resources to validate, prove, and broadcast their identity.

Further Resources

The following are two examples of identity programs along these lines:

- **eIDAS**
  Work is underway to explore extending the U.K. Verify Program to commercial applications and not just government. This aligns to the implementation of the eIDAS scheme throughout the European Union, known as Regulation (EU) N°910/2014.

  Adopted by the co-legislators in July 2014, the eIDAS scheme is a milestone to provide a predictable regulatory environment that enables secure and seamless electronic interactions between businesses, citizens, and public authorities. It ensures that people and businesses can use their own national electronic identification schemes (eIDs) to access public services in other EU countries where eIDs are available.

  The aim is to create a European internal market for eTS—namely electronic signatures, electronic seals, time stamp, electronic delivery service, and website authentication—by ensuring that they will work across borders and have the same legal status as traditional paper-based processes.

  With eIDAS, the EU has provided the foundations and a predictable legal framework for people, companies, and public administrations to safely access services and do transactions online and across borders in just "one click." Rolling out eIDAS means higher security and more convenience for any online activity such as submitting tax declarations, enrolling in a foreign university, remotely opening a bank account, setting

5 Personal Data and Individual Access Control

up a business in another Member State, or authenticating for internet payments.

- **IDNYC – New York Residents ID Program** IDNYC is a free identification card for all New York City residents. It is a government-issued photo identification card fulfilling the requirement for New York residents to permanently carry an acceptable form of ID. Eligibility extends to the most vulnerable communities; including the homeless, youth, the elderly, undocumented immigrants, the formerly incarcerated, and others who may have difficulty obtaining other government-issued ID.

  More importantly, IDNYC has implemented leading privacy practices and policies in order to further protect the vulnerable groups the program serves. These privacy enhancing processes include strict limits on the amount of time physical application documents are held before destroying them and who can access the enrollment information and the ID database, including other government and security agencies.

  Pursuant to NYC Administrative Code Section 3-115(e)(4), information collected about applicants for the IDNYC card shall be treated as confidential and may only be disclosed if authorized in writing by the individual to whom such information pertains, or if such individual is a minor or is otherwise not legally competent, by such individual's parent or legal guardian.

  The card features a photograph, name, date of birth, signature, eye color, height, and unique ID number. Residents can choose whether or not to include gender (including self-declared), emergency contact information, organ donor status, preferred language, and option to display Veteran status.

  Data privacy provides options for those that are survivors of domestic violence, or have legitimate security concerns, regarding address disclosure.

Issue:

What is the definition and scope of personally identifiable information?

Background

Personally identifiable information (PII) is defined as any data that can be reasonably linked to an individual based on their unique physical, digital, or virtual identity. As further clarification, the EU definition of personal data set forth in the Data Protection Directive 95/46/EC[lvi] defines personal data as "any information relating to an identified or identifiable natural person." The Chairwoman of the United States Federal Trade Commission has also suggested that PII should be defined broadly. The new GDPR legislation also provides

5 Personal Data and Individual Access Control

definitions for genetic and biometric data[lvii] that will become even more relevant as more devices in the Internet of Things track these unique physical identifiers.

Candidate Recommendation

PII should be considered the sovereign asset of the individual to be legally protected and prioritized universally in global, local and digital implementations. In the U.S., for instance, PII protection is often related to the right of the people to be secure in their persons, houses, papers, and effects, pursuant to the fourth amendment to the Constitution (e.g., the Supreme Court's ruling in US v. Jones from 2012, 565 U.S.).[lviii] In the EU, PII protection is commonly framed in terms of informational self-determination and defense of human dignity. In both cases, (See generally United States v. Jones, 565 U.S. 400 (2012)) the aim should be to tackle key ethical dilemmas of data asymmetry by prioritizing PII protection universally in global, local, and digital implementations.

Further Resources

- Different laws and regulations around the globe define the scope of personally identifiable information differently. The use of data analytics to derive new inferences and insights into both personal data and technical metadata raises new questions about what types of information should properly be considered personal data. This is further complicated by machine learning and autonomous systems that access and process data faster than ever before.
- The U.S. Federal Trade Commission (FTC) has taken the position in its 2009 staff report on online behavioral advertising and in its more recent 2012 Privacy Report that data is "personally identifiable," and thus warrant privacy protections, where it can be reasonably linked to a particular person, computer, or device. As a result, in many circumstances, persistent identifiers such as device identifiers, MAC addresses, static IP addresses, or cookies are considered personally identifiable under U.S. federal law. More recently, the European Court of Justice (ECJ) Advocate General has also proposed that IP addresses are personal data protected by European Union law. U.S. Federal Communications Commission (FCC) officials approved broad new privacy rules on October 27, 2016, that prevent companies like AT&T and Comcast from collecting and giving out digital information about individuals—such as the websites they visited and the apps they used— in a move that creates landmark protections for internet users. The new rules require broadband providers to obtain permission from subscribers to gather and give out data on their web browsing, app use, location, and financial information. Currently, broadband providers can track users unless those individuals tell them to stop.
- For additional discussion of how to think about what constitutes personal data, we recommend the U.K. Information Commissioner's Office paper, *Determining What Is Personal Data*, which provides guidance on how to decide whether data falls within the definition of personal data in non-obvious circumstances.

5 Personal Data and Individual Access Control

Issue:

What is the definition of control regarding personal data?

Background

Most individuals believe controlling their personal data only happens on the sites or social networks to which they belong. While taking the time to update your privacy settings on a social network is important, the logic of controlling your personal data is more holistic and universal in nature. Instead of individuals having to conform to hundreds of organization's terms and conditions or policies, in a world where people control their own personal data, those organizations would conform to an individual's predetermined requirements.

Candidate Recommendation

Personal data should be managed starting from the point of the user versus outside actors having access to data outside of a user's awareness or control.

Further Resources

- For an introduction into these issues, we recommend the Project VRM website. VRM stands for vendor relationship management, a concept created by Doc Searls and outlined with great specificity in his book, *The Intention Economy: When Customers Take Charge*. In marketing terms, customer relationship management (CRM) describes the tools utilized to track, message, and influence individuals that companies want to attract. The current Internet economy is built on this CRM model.
- Providing individuals with tools like a personal data cloud as described in the Fast Company article, "Personal.com Creates an Online Vault to Manage All Your Data," can empower users to understand how their data is an asset as well as how much data they produce. Tools like these vaults or clouds also let individuals organize their data around various uses (medical, social, banking) to potentially create an individual version of their own terms and conditions. For an example of this, we recommend reviewing Meeco.me and their Signal feature.
- For more specifics on this topic, we recommend reading Introduction to the Personal Data Ecosystem created by The Personal Data Ecosystem Consortium (PDEC).
- Hasselbalch, Gry, and Pernille Tranberg. *Data Ethics. The New Competitive Advantage*. Copenhagen: Publishare, 2016.

Section 2 – Personal Data Access and Consent

If you cannot access your personal data, you cannot benefit from its insights. Also, you will not be able to correct erroneous facts to provide the most relevant information regarding your life to the actors you trust. Multipage agreements written to protect organizations must also quickly and genuinely inform users of their choices for trusted consent in the algorithmic era.

Issue:

How can we redefine data access to honor the individual?

Background

Much of the contention associated with the concept of "privacy" actually relates to access and consent. The challenges are often around transparency and providing an explicit understanding of the consequences of agreeing to the use of our personal data, complicated by the data handling processes behind true "consent." Privacy rights are often not respected in the design and business model of services using said data.

Candidate Recommendation

Practical and implementable procedures need to be available in order for designers and developers to use "Privacy-by-Design"/Privacy-by-Default methodologies (referring to the practice or business philosophy of privacy embedded in the development of a service).

In order to realize benefits such as decision enablement and personalization for an individual, open standards and interoperability are vital to ensure individuals and society have the freedom to move across ecosystems and are not trapped by walled gardens. In order to safeguard this freedom, for example, Article 20 of the EU regulation on data protection (Right to Data Portability) sets up the right to receive PII that individuals have provided to a data controller, in a structured, commonly used and machine-readable format and have the right to transmit those data to other controllers without hindrance from the controller to which the personal data have been provided. [lix]

Paradigms like "differential privacy" may also allow for designers and developers to bake privacy into the design and development of services.[lx] Differential privacy shifts the focus from "your data" to finding general usage

5 Personal Data and Individual Access Control

patterns across larger data sets. Differential privacy is not about anonymization of data, as that can be easily de-anonymized through intelligent cross-referencing. Instead differential privacy uses hashing, sub-sampling, and noise-injection techniques to obfuscate personal information about individuals. However, while differential privacy may provide a methodology for better usage of private or public data, it should be implemented in complement to tools and methodologies empowering individuals to manage and control their data.

As a tool for any organization regarding these issues, a good starting point is to apply the who, what, why, and when test to the collection and storage of personal information:

1. Who requires access and for what duration—is it a person, system, regulatory body, legal requirement "or" input to an algorithm?
2. What is the purpose for the access—is it read, use and discard or collect, use and store?
3. Why is the data required—is it to fulfill compliance, lower risk, because it is monetized, or in order to provide a better service/experience?
4. When will it be collected, for how long will it be kept, when will it be discarded, updated, re-authenticated—how does duration impact the quality and life of the data?

Issue:

How can we redefine consent regarding personal data so it honors the individual?

Background

Technology leaders give innovation teams and engineers too little or no direction on what human values should be considered, protected and designed for regarding personal data. When implemented correctly, solutions providing transparency and choice for the individual can be designed within the increasing regulatory environment (as is the case currently with the GDPR in the EU) to allow for minimal viable collection for maximum viable access. However, it should be noted that fundamental issues regarding the processing of data need to be addressed before exchanges happen so individuals aren't consenting to commercial or scientific usage of their information that is unclear without methods for recourse or control. A final issue to consider along these lines is how to design for portability when the convergence of digital, mobile, and Internet of Things results in the perpetual creation of data.

Candidate Recommendations

In order to realize benefits such as decision enablement and personalization for an individual, open standards and interoperability are vital to

5 Personal Data and Individual Access Control

ensure individuals and society have the freedom to move across ecosystems. Explicit consent provided by individuals in the exchange of their data via methodologies previously described in this document can inform future requirements for data to be stored, shared downstream, anonymized, or identified. By developing a decision matrix between individuals and external actors about their information, personal data can be used to process high-volume anonymized data for general insights, through to low-volume identified data used for tailored experiences.

The needs of society, communities, and research will factor into this decision matrix and introduce the need to consider security, roles, and rights management. For example, a doctor may need medical data to be identified in order to treat a patient. However a researcher may require it simply for statistical analysis and therefore does not require the data to be identifiable. Additionally mechanisms for dynamic consent as use-cases change, or data moves from the original collection context to a change of context are critical design features. This is particularly important to explicitly surface if the primary reason for data collection masks the secondary use post-collection. A European context along these lines will also require for the "right-to-be-forgotten" as a core design capability.

Further Resources

- European Commission, Factsheet on The Right to Be Forgotten Ruling.

Issue:

Data that appears trivial to share can be used to make inferences that an individual would not wish to share.

Background

How can individuals be sufficiently informed to give genuine consent?

Candidate Recommendation

While it is hoped AI/AS that parse and analyze data could also help individuals understand granular level consent in real-time, it is imperative to also put more focus on the point of data collection to minimize long-term risk.

Further Resources

As analysis becomes more autonomous, not even the analysts will necessarily know what conclusions are being drawn and used in the process. This means that informed consent could become too complex for companies to ask for or consumers to give. This is why we need to move focus away from the consent of the user to the point of data collection. Too much data is collected for no immediate purpose. There needs to be limits and exact purposes for the collection of personal data. Use limitations are also important and may be more feasible than collection limitations. Organizations should commit not to use data to make sensitive

5 Personal Data and Individual Access Control

inferences or to make important eligibility determinations.

- For an example along these lines: Felbo, B., P. Sundsøy, A. Pentland, S. Lehmann, and Y. de Montjoye. "Using Deep Learning to Predict Demographics from Mobile Phone Metadata." Cornell University Library, arXiv: 1511.06660, February 13, 2016.

Issue:

How can data handlers ensure the consequences (positive and negative) of accessing and collecting data are explicit to an individual in order for truly informed consent to be given?

Background

It is common for a consumer to consent to the sharing of discrete, apparently meaningless data points like credit card transaction data, answers to test questions, or how many steps they walk. However, once aggregated these data and their associated insights may lead to complex and sensitive conclusions being drawn about individuals that consumers would not have consented to sharing. A clear issue, as computational power increases with time and algorithms improve, is that information that was thought private can be linked to individuals at a later stage in time. Furthermore, as data is stored in terms of summaries rather than as raw observations, and may be key to training algorithms, keeping track of data usage and potential risks to privacy may be increasingly complex.

Candidate Recommendations

To guard against these types of complexities we need to make consent both conditional and dynamic. Safeguards are required to surface the downstream impact of data that appears to be trivial that can be later used to make inferences that an individual would not wish to share. Likewise, resources and legislation should be afforded to an individual so they can retract or "kill" their data if they feel it is being used in ways they do not understand or desire.

Further Resources

For examples along these lines:

- Duhigg, C. "How Companies Learn Your Secrets." *The New York Times Magazine*, Feb. 19, 2012.
- Meyer, R. "When You Fall in Love, This Is What Facebook Sees." *The Atlantic*, Feb. 15, 2014.
- Cormode, G. "The Confounding Problem of Private Data Release." *18th International Conference on Database Theory* (2015): 1–12. DOI: 10.4230/LIPIcs.ICDT.2015.1.

Section 3 – Personal Data Management

For individuals to achieve and retain a parity regarding their personal information in the algorithmic age, it will be necessary to extend an Identity Assurance paradigm to include a proactive algorithmic tool that acts as their agent or guardian in the digital, and "real" world ("real" meaning a physical or public space where the user is not aware of being under surveillance by facial recognition, biometric, or other tools that could track, store, and utilize their data without pre-established consent or permission).

Issue:

Could a person have a personalized AI or algorithmic guardian?

Background

The creation of a personalized AI would provide a massive opportunity for innovation in AI and corporate communities. Some might view an individual's desire to control and manage their data as hindering innovation since higher choices may conflict with well-intentioned efforts to amass vast data sets for public good. However, this view inherently assumes all individuals in a certain context would want their data utilized for a certain project, even if it was for the "public good."

The sophistication of data-sharing methodologies have evolved so these scenarios can evolve from an "either/or" relationship ("we get all of your data for this project or you provide nothing and hinder this work") to a "yes and" one—by allowing individuals to set their preferences for sharing and storing their data they are more likely to trust the organizations conducting research and provide more access to their data.

It should also be noted that providing these types of platforms and paradigms is of value to organizations at large because contrary to rhetoric saying, "privacy is dead," individuals and governments around the world have become more focused on the control of privacy and personal data in the past few years. In the United States, according to a May 20, 2015 report, "93% of adults say that being in control of *who* can get information about them is important; 74% feel this is 'very important,' while 19% say it is 'somewhat important'" and, "90% say that controlling what information is collected about them is important—65% think it is 'very important' and 25% say it is 'somewhat important' (Madden and Rainie)."[lxi]

5 Personal Data and Individual Access Control

Candidate Recommendation

Algorithmic guardian platforms should be developed for individuals to curate and share their personal data. Such guardians could provide personal information control to users by helping them track what they have agreed to share and what that means to them while also scanning each user's environment to set personal privacy settings accordingly. The guardian could serve as an educator and negotiator on behalf of its user by suggesting how requested data could be combined with other data that has already been provided, inform the user if data is being used in a way that was not authorized, or make recommendations to the user based on a personal profile. As a negotiator, the guardian could negotiate conditions for sharing data and could include payment to the user as a term, or even retract consent for the use of data previously authorized for a breach of conditions.

Nonetheless, the dominant paradigm for personal data models needs to shift to being person-based and away from system and service-based models not under the control of the individual/human. Personal data cannot be controlled or understood when fragmented and controlled by a myriad of entities in legal jurisdictions across the world. The object model for personal data should be associated with that person, and under the control of that person utilizing a personalized AI or algorithmic guardian. *Specifically*:

- For purposes of privacy, a person must be able to set up any number of agents/guardians or profiles within one agent with different levels or types of personal data associated.
- During the handshake/negotiation between the personal agent and the system or service, if the required data set contains elements the personal agent will not provide, the service may be unavailable. If the recommended data set will not be provided, the service may be degraded.
- Default profiles, to protect naive or uninformed users, should provide little or no personal information without explicit action by the personal agent's owner.

Further Resources

- We wish to acknowledge Jarno M. Koponen's articles on Algorithmic Angels that provided inspiration for portions of these ideas.
- Companies are already providing solutions for early or partial versions of algorithmic guardians. Anonyome Labs recently announced their SudoApp that leverages strong anonymity and avatar identities to allow users to call, message, email, shop, and pay—safely, securely, and privately.
- Tools allowing an individual to create a form of an algorithmic guardian are often labeled as PIMS, or personal information management services. Nesta in the United Kingdom was one of the funders of early research about PIMS conducted by CtrlShift.