Personal Data, Privacy and the Internet of Things: The Shifting Locus of Agency and Control

Andy Crabtree
School of Computer Science
The University of Nottingham
Email: andy.crabtree@nottingham.ac.uk

Richard Mortier
Computer Lab
University of Cambridge
Email: richard.mortier@cl.cam.ac.uk

Abstract. This paper examines the shifting the locus of agency and control in the processing of personal data, occasioned by the emerging digital ecosystem and the Internet of Things in particular. It reviews legal, policy, and industry initiatives that seek to respond to broad societal concerns over privacy, and initiatives within design that can be seen to respond to the core problem of empowering users and giving them active control over the processing of personal data. It is, in many respects, a clarion call for HCI to respond to a grand challenge in contemporary life. It suggests that current efforts to address the problem have been piecemeal and that a coherent interdisciplinary approach is required to affect the shift. It proposes that the nascent field of Human Data Interaction provides the necessary design framework needed to drive the research community forwards and put users in control in practice.

Keywords. Privacy, personal data, data protection, Internet of Things, agency and control, human data interaction.

1. Introduction

Privacy is currently a topic of widespread societal interest and debate as digital technologies generate and trade in personal data on an unprecedented scale. The calls of industry and government alike proclaim the social and economic benefits to be had from personal data and run alongside a steady flow of scare stories detailing its misuse and abuse. Current industry efforts to quell anxiety offer security as the panacea to public concerns. Encryption in turn becomes a matter of concern to those charged with state security. It also glosses, hides or at least renders opaque a key threat to consumer or user privacy: the ability to 'listen in' and stop devices 'saying' too much about us. As Winstein [37] puts it,

> "Manufacturers are shipping devices as sealed-off products that will speak, encrypted, only with the manufacturer's servers over the Internet. Encryption is a great way to protect against eavesdropping from bad guys. But when it stops the devices' actual owners from listening in to make sure the device isn't tattling on them, the effect is anti-consumer."

The Internet of Things (IoT) further complicates the situation, reshaping the nature of data collection from an 'active' feature of human-computer interaction to a 'passive' one in which devices seamlessly communicate personal data to one another across computer networks. Insofar as encryption is seen as the panacea to privacy concerns this gives rise to 'walled gardens' in which personal data is distributed to the cloud before it is made available to end-users. Encryption offers the consumer *no guarantee* that privacy will be protected then, as consumer data is still open to the kinds of industry abuses that we are all familiar with. Open IoT platforms, such as Samsung's ARTIK, do not circumvent the problem either; they are only open to developers.

This is not an objection to the IoT; clearly security is an important part of the privacy equation, though it is equally clear that *more* is required. There is need in particular to put the end-user into the flow of personal data; to make the parties about whom personal data is generated into active rather than passive participants in its distribution and use. The need to support personal data management is reflected in a broad range of legal, policy and industry initiatives. This paper explores a range of these initiatives to tease out formative perspectives on the challenges posed by new digital ecosystems to privacy in the 21st century.

Here issues of trust, accountability, and user empowerment are paramount. They speak not only to the obligations of 'data controllers' – the parties who are responsible for processing personal data and ensuring compliance with regulation and law - but seek to shift the locus of agency and control towards the consumer in an effort to transform the user from a passive 'data subject' into an active participant in the processing of personal data. That is, into someone who *can* exercise control and manage their data and privacy, and thus become an active player or participant in (rather than a passive victim of) the emerging data economy.

Enabling this transformation is where HCI can play a key role. Thus, our exploration of current legal, policy and industry concerns is complemented in turning to consider current design efforts that seek to address the contemporary privacy challenge. Here we explore design efforts to enable user control and the tracking of personal data. We also consider the need to move beyond current piecemeal efforts and develop a coherent design framework that is explicitly focused on personal data management: one which moves beyond a generic focus on human-computer interaction to a specific focus on *human-data interac-*

tion. This focus recognizes that data has become an object in its own right [7]. Our aim here is to outline key design challenges for human-data interaction, each of which requires significant HCI research, to enable the transformation to take place and *put users in control*.

2. Legislating Privacy

We first explore legislative initiatives in Europe, the US and Japan. These of course are not the only areas to implement data protection measures [10], but they are leading the way on the challenges to privacy posed by the emerging digital ecosystem. It might be asked why exploring these perspectives on privacy is relevant to HCI? The simple answer is that they shape the social context in which computers, and their designers, operate and that their systems, applications, services and devices *must* be accountable to if they are ever to make it into everyday use. The perspectives at work here are powerful; literally, they can and do exert influence on systems development.

It is not possible (and probably not desirable) to cover the twists and turns of legal thought in each case, but we can draw out what is distinctive about the respective *orientations to privacy* at work here and issues of common concern. Common agreement exists as to the potential social and economic benefits to be had from personal data, for example, and also to the risks occasioned by the emerging digital ecosystem and the need to build *trust*. Furthermore, there is common agreement as to the absence of clear 'ground rules' for handling personal data in the emerging digital ecosystem [34], that personal data is becoming a 'gray area' [33], that current laws and regulations create 'uncertainty' [6] and ultimately a barrier to social and economic benefit. All are in agreement that new frameworks are required to address privacy in a massively networked world, though all but the EU proposal have yet to be ratified.

2.1. Europe

The European proposal extends the existing overarching legal framework [12], motivated in no small part by "what happened under Nazi rule" [20], which enshrines the individual's "right to the protection of personal data". Personal data, simply put, is data that relates to an *identifiable* "natural person" (for a more detailed definition see [8]). The aim of encoding the rights of natural persons in law is not simply to protect EU citizens from the *unaccountable processing* of personal data. It is also to "put individuals in control of their own data", and to provide "legal and practical certainty" for those would make use of it as to what this entails [6]. Key to ensuring the individual's right to control the processing of their personal data are the issues of consent and accountability; it is worth noting that 'processing' generally refers to the collection, retention, use, disclosure and/or disposal of personal data.

2.1.1. Consent

Consent is the primary mechanism of personal data protection in the EU proposal – "such data should not be processed, unless the data subject gives his explicit consent" [6]. What constitutes consent is further defined by the new EU proposal: it should be given freely by a method appropriate to data collection. The method should clearly inform the data subject of the purpose(s) of processing and all processing activities, and should enable the data subject to provide a specific indication of their wishes through clear affirmative action, written or otherwise. The data controller has the burden of proving that the data subject has given their consent to processing operations. It is also of note, especially where the digital is concerned, that consent does not provide a valid legal ground for the processing of personal data where there is a clear "imbalance" between the data subject and the data controller. That is, in situations where the individual is not able to refuse or withdraw consent without detriment and so has no genuine free choice.

2.1.2. Accountability

Consent is closely coupled to the "accountability principle" [1, 2] or the "fair, lawful and transparent" processing of personal data. This means that parties who would make use of an individual's personal data must have legitimate grounds for data processing, which do not breach the fundamental rights of the data subject. Personal data should not be obtained in a deceitful or misleading way or processed in ways that are otherwise unlawful. It also means that in the process of consent the need for collecting and processing personal data, and the consequences of not providing it, should be clearly specified along with the data retention period. Data processors must also specify if the data will be disclosed to another recipient. The data must also be relevant to the purposes for which it is processed, be limited to the minimum necessary for the purposes for which it is processed, and be processed in ways that data subjects might reasonably expect.

Data processors must also ensure that processing does not have a detrimental effect on individuals unless this is justified (e.g., by legal requirements). Thus, the accountability principle seeks to ensure that "responsibility is exercised" [2] in the processing of personal data. This also includes informing data subjects of their "rights" as part of the consent process including the right to lodge a complaint; the right of access, rectification or erasure; the right not to be subject to profiling by means of automated processing, unless explicitly consented to; and the right to be forgotten where the retention of personal data is not in compliance with EU data protection regulation.

2.1.3. Compliance

Compliance with the proposed data protection regulation is to be provided through a European Data Protection Board made up of Commission representatives and the heads of independent supervisory authorities in Member States. The regulation applies to the processing of personal data by any data controller whose "establishment" operates within the EU, if the processing "takes place within the Union or not" [6]. Furthermore, the regulation applies to data processing outside of the EU when processing is "related to the offering of goods or services" to EU citizens or to "the monitoring of the behaviour of such data subjects". The regulation does not apply to data rendered anonymous in such a way that "the data subject is no longer identifiable".

2.2. America

The American approach is, as we might expect, somewhat different. Here regulation seen as something that may stifle innovation. It is also important to appreciate that the starting point is different. Data protection in the US is "sectorial" [20] covering specific areas of data protection (e.g., health care) rather than personal data protection *in general* as in the EU. The Administration therefore seeks to foster the uptake of new Fair Information Practice Principles (FPPP) rather than place overarching legal requirements on the processing of personal data. These new principles are intended "to provide consumers who want to understand and control how personal data flows in the digital economy with better tools to do so. [34]"

2.2.1. New Fair Information Practice Principles

The Consumer Privacy Bill of Rights [34] outlines 7 principles of data processing. (1) Individual Control revolves around consent, which is seen as the "critical" initial act in "sharing" personal data. This principle requires that consumers be offered "clear and simple choices", which allow them to make meaningful decisions about data collection and processing. Furthermore, that consumers are informed of data disclosure to other recipients, including the constraints (if any) on its subsequent use by them, and that similarly "clear and simple" means are provided that allow consumers to withdraw or limit consent.

(2) Transparency is key to Individual Control and is concerned to make the nature of data processing accountable to consumers. It requires that clear and concise descriptions of the data to be collected, why it is needed, how it will be used, and when it will be deleted or de-identified be provided to consumers. (3) Respect for Context is concerned to ensure that the use and disclosure of personal data is limited to purposes that are consistent with the relationship the data processor has with consumers and the context in which consumers originally provided the data. If it isn't then this this should be "prominent and easily actionable by consumers" at the time of data collection.

(4) Security is concerned with putting in place reasonable safeguards to control risks, notably data loss; unauthorized use, modification or destruction of data, and improper disclosure (the EU is similarly concerned with security, see Article 30 of the proposal). (5) Access and Accuracy is concerned with other data risks, particularly that data processing is accurate and that consumers are provided with appropriate means and opportunity to access data, to correct inaccuracies, and to limit its use or request its deletion. (6) Focused Collection is concerned with data minimization and that "only as much personal data as needed to accomplish specified purposes" is collected and processed. (7) Accountability is concerned with putting a framework in place that ensures compliance with these principles.

2.2.2. Compliance

Compliance is not a matter of encoding these principles in law, but of fostering the widespread and voluntary uptake of "enforceable codes of conduct" that embody the principles [34]. Even though adopted on a voluntary basis, such codes of conduct become enforceable by the Federal Trade Commission (FTC) under federal acts and state statutes prohibiting *unfair* or *deceptive* acts or practices. Enforcement will apply to US regulated entities even if personal data leaves the US and is processed by overseas subcontractors. Thus the new FPPP package not only provides "a clear set of ground rules to apply in the commercial arena", but in doing so is explicitly intended to increase interoperability between domestic and international privacy frameworks.

2.3. Japan

As with EU data protection law, the Japanese proposal extends current regulation, specifically the 2003 Act on the Protection of Personal Information or APPI. Explicit consent is *not* generally required by the APPI insofar as the purposes of data collection are clearly specified and it is processed in ways that data subjects might reasonably expect. Explicit consent is only required if the data is subsequently used for purposes other than those originally specified or if it is transferred to another recipient, in which case the data subject must be notified and offered an "opt-out". The regulation excludes small businesses whose databases identify less than 5,000 individuals.

As in the EU and US, a range of responsibilities and rights attach to personal data processing including rights of access, complaint, correction and deletion if the data is inaccurate or used for purposes other than originally specified. The APPI only applies to companies that carry out business in Japan. It makes no specific provision for trans-border dataflow beyond providing an "opt-out" notice as described above. Compliance is governed by the government ministry having jurisdiction over the area of activity a business operates within. Ministries do not have the power to take direct enforcement action, they can only issue enforcement notices. The APPI has been deemed "not sufficient" to meet international standards [14].

2.3.1. Revision of the Current Law

A new "policy outline" was issued by the Cabinet of Japan's IT Strategic Headquarters in 2014 [33] motivated by the growing concerns of Japanese consumers and the need to revitalize industry. The new bill outlines "a system that makes it possible to utilize the personal data possessed by businesses" while at the same time responding to consumer concerns about personal data "being used for purposes other than the person's intent." As in Europe and America the new bill seeks to balance economic opportunity with consumer control over the processing of personal data.

Consent is again key, though perhaps not as expected insofar as it is seen as a major "barrier to the utilization of personal data". Consent will not be required as per the current law for reasonable use that meets expectations. Nor, it is proposed, will be it required for changes of purpose or disclosure to other recipients "under certain regulations", which mandate the processing of personal data "into data with the reduced identifiability of individuals" [33]. The proposed framework will make it possible to provide personal data to third parties and use it for purposes other than those specified without obtaining consent insofar as it is effectively *anonymised* then, otherwise the regulations of current law apply.

The framework also extends definitions of what constitutes personal data in a bid to further reduce "ambiguity" and "interpretation" to include biological and sensitive data that "may cause social discrimination". The proposed framework will "prohibit such data from being handled … without consent of the persons" [33]. It will also abolish the small business exemption and regulate trans-border transfer of personal data. To transfer data to a foreign company, contracts will need to be drafted that comply with the new data protection framework, unless the destination country is certified as having equivalent data protection standards as Japan's. The extension of data protection regulation to trans-border data transfers is intended to provide for an "internationally harmonized, reliable system" of data protection and privacy.

2.3.2. Compliance

Compliance with this harmonized system is to be provided through the creation of an independent data protection commission referred to as the "third-party authority". Much as the American approach is to establish voluntary codes of conduct enforced by the FTC, compliance in Japan trades upon "private sector-led self-regulations" enforced by the third-part authority. The third-party authority will be responsible for certifying self-regulations and supervising both government and industry frameworks for data distribution beyond borders. It will also have powers of complaint handling and compliance monitoring, be able to perform onsite inspections, issue warnings and cancel certification. It will, in short, have teeth.

2.4. Summary

It is clear to see that legislation is unique in Europe, America and Japan. The EU emphasizes compliance with legally binding requirements, Japan and the US voluntary codes of conduct enforced by the legal agencies. Nevertheless, and despite these differences, these legal initiatives are punctuated by common concerns, particular with consent and accountability. These matters are addressed in different ways in different jurisdictions – in Europe consent is always required, whereas the grounds upon which it is required vary in the US and Japan depending on the context of data processing and the nature of the data involved – though there is a clear push to make data processing accountable and transparent especially where data is used for purposes other than those than might reasonably be expected and where data crosses borders. 'Harmonising' data processing is a common goal and includes not only specifying the need for consent and accountability but also increasingly circumscribes what constitutes personal data and bringing it under the control of regulation. These measures are intended to build consumer trust in the digital economy and drive data processors to provide consumers with the tools they need to exercise control over the processing of their personal data. Of course, a key challenge is to figure out how to put the principles enshrined in legislative frameworks into practice, and it is towards such concerns that we turn next.

3. Policy

Given the head start the US and EU has on the personal data challenge it is perhaps unsurprisingly that they are leading the way on working out how to put legislative frameworks into practice and *shape the design* of devices and services that make use of personal data. The Internet of Things (IoT) is a matter of particular concern in this respect. The IoT is defined by the US Federal Trade Commission (FTC) as devices or sensors other than personal computers, smartphones, or tablets that connect, communicate or transmit information with or between each other through the Internet [35]; and by the EU's statutory Article 29 Working Party on Data Protection (WP29) as an infrastructure in which billions of sensors embedded in common, everyday devices seamlessly record, process, store and transfer data and interact with other devices or systems using networking capabilities [2]. Both bodies are tasked with turning legislative principles into policies that provide practical measures for IoT development.

3.1. US FTC

The FTC suggests that three and a half billion sensors are already in the marketplace and predict that data traffic will exceed fifteen exabytes per month by 2018. Much of the US-based traffic is currently unregulated and where regulation does apply (e.g., the Fair Credit Reporting Act) then it generally excludes companies that do their own in-house analytics (e.g., an insurance company that offers consumers a wearable fitness tracker to lower health insurance premiums). The current state of play is thus seen to pose a range of privacy risks that extend beyond security, including [35]:

- The "massive volume of granular data" revealing "sensitive behaviour patterns" either directly or through inference.

- The risk that either authorized or unauthorized parties could “eavesdrop remotely, intruding into an otherwise private space.”
- The risk of consent being undermined by devices that “have no screen or other interface to communicate with the consumer” or, even if a device has screens, that data may be collected “at times when the consumer may not be able to read a notice (for example, while driving).”

Furthermore, these risks combine to “undermine the consumer confidence necessary for the technologies to meet their full potential and may result in less widespread adoption.”

The FTC is wary of legislation and “choking off innovation”. It does, however, recognize the need for industry to balance personal data harvesting by adopting a responsible use framework, which shifts the emphasis from whether or not consent was “properly obtained” to holding “data collectors and users accountable for how they manage the data and the harms it causes”. It is accompanied by a series of practical measures that seek to engender “best practices” in IoT development [35]:

- *Security by design*. The recommendation here is that security is built into devices at the outset, rather than as an afterthought and that “defense-in-depth” approach is adopted to secure information passed over consumers’ home networks.
- *Data minimization*. This recommends that reasonable limits are imposed on the collection and retention of consumer data. As the FTC report puts, “Thieves cannot steal data that has been deleted after serving its purpose; nor can thieves steal data that was not collected in the first place.”
- *Notice and choice*. This recommends various measures to ensure informed consumer choice “remains practicable in the IoT.” The measures include providing “management portals or dashboards” to enable consumers to (re)configure privacy settings for particular devices along with general settings that can be applied by default; “out of band” communications allowing consumers to configure devices so that they receive important information through emails or texts; and “icons” for quickly and easily turning device connections on or off.

3.2. EU WP29

European policy considerations are similarly marked by a concern with the risks posed by the IoT. At the heart of these lies the “seamless” nature of data distribution in the IoT, which “may result in situations where the user can lose all control on the dissemination of his/her data”. A range of risks in addition to security are also concomitant to this [2]:

- Much of the data generated by IoT devices is not, at the current moment, “adequately reviewable” by the data subject. Not only does this reduce individual control it creates “information asymmetry” – i.e., the user does not know just what data is actually being disseminated.
- The potential combination of multiple IoT data sources raises the risk of the “intrusive bringing out” of “an individual’s habits, behaviours and preferences”, along with the possibility to “infer information with a totally different meaning” to that for which it was originally collected by a device.
- The impact of devices that are “not noticeable” is seen as a key risk to consent, along with the

ability to discern “unique fingerprints … which IoT stakeholders will be able to attribute to specific individuals”, which undermines the “anonymous use of services”.

WP29 has similar concerns to the FTC with the granularity of data processing and problems of consent raised by the IoT, complemented by a concern with asymmetrical differentials in data distribution. WP29 furthermore recognizes the inadequacies of “classical mechanisms” for providing consent and managing data flows and thus proposes a series of practical measures be taken by industry to mitigate the risks occasioned by the IoT [2]:

- *Privacy by design*. The recommendation here is that in addition to security by design, privacy be built into devices at the outset “by default” and that a “privacy impact assessment framework” [2] be developed and applied for particular IoT applications to ensure this.
- *Data minimization*. This recommends that application developers apply the data minimization principle. This includes limiting “the amount of data leaving devices by transforming raw data into aggregated data directly on the device” or “at the nearest point of data collection” as possible.
- *Transparency and control*. Similar to the FTC, this recommends various measures to ensure that consumers can exercise an informed choice. These include providing “granular choices” when granting access to data sources, not only in terms of the data itself but also the time and frequency at which it is captured. Furthermore, tools should be provided that enable consumers to “read, edit and modify the data” and that this should be possible “without [first] having to transmit the data to the device manufacturer.” These measures thus seek to provide “local control” over “processing entities”.

3.3 Summary

Again the policy recommendations and practical measures recommended in the US and EU are unique: the emphasis on local control in Europe is notably different to the US, for example. Nonetheless, both parties emphasize the need to build in security and privacy measures at the outset of design; both emphasize the need to build in data minimization, and both emphasize the need to build in mechanisms that not only inform consumers but also enable active *and ongoing* control over data processing. These measures seek to move beyond existing consent models and build accountability into the actual *use* of IoT devices and services. Before considering how these issues might be seen to play out in design it is first worth considering how industry is responding to the personal data challenge.

4. Industry

While law and policy makers recognize that the economic push from industry to open up and valorize personal data needs to be responsive to societal concerns, industry itself is not blind to the underlying concerns of consumers. The World Economic Forum (WEF), perhaps best known for its annual Davos conference, is a key player in shaping global industry agendas. The WEF has released several reports over recent years, drawing off a wide range of research, on personal data and the challenges that confront industry [38, 39, 40]. Top of the list is what the WEF calls “a crisis in trust” [39]. The crisis is occasioned by the “sophistication of ubiquitous computing”, the “sea change” in data processing created by the IoT, and the “borderless flow of data” [40]. The situation is exacerbated by the inadequacies of traditional data-protection approaches, which assume that data is “actively collected from the individual with some level of direct awareness”; an assumption which starts to breakdown “as billions of sensor passively collect data (without individuals being aware of it) and as computer analytics generate and synthesize more ‘bits about bits’” [40].

A new approach is required that addresses foundational causes of “global anxiety” in today’s data-driven world. Key here is the need to “empower individuals”. This includes, as we might expect, the need to provide consumers with increased control over their personal data and its use such that people have “a say in how data about them is used by organizations” and can “more effectively engage in understanding (and managing) the intended impact of data usage.” At the same time, the WEF also recognizes that the value of personal data is all too often *concealed* from individuals and that empowerment also turns upon promoting “fair value exchange”. Bringing this about and enabling individual empowerment turns upon redistributing power, providing meaningful transparency and strengthening accountability [40]:

- *Redistributing power*. The WEF is of the view that current data protection approaches reflect “an asymmetry in power that broadly favours institutions (both public and private).” Consequently debate is “framed by incumbent interests from a highly concentrated set of powerful actors”, who “have greater resources to design the legal terms of data transactions to advance their interests.”
- *Meaningful transparency*. Key to redistributing power is meaningful transparency. This involves opening up the data flow from the current “one-way street” that reduces individuals to “being spectators” on how their data is used. Meaningful transparency involves making the ecosystem and not just the front end transparent, exposing “business-to-business processes of data handling” and the “ways that data flows out the back door” to “facilitate data flows which empower individuals in meaningful transactions and experiences that are consistent with their expectations.”
- *Strengthening accountability*. Compliance with regulation is seen as deeply problematic in practice: “accounting for the complex realities of today’s data flows in a precise and granular manner remains a grand challenge”. While regulatory frameworks are seen as a key to addressing the challenge, “tools are needed” as well to “simplify, automate and facilitate trustworthy data flows.”

Empowering individuals is not only a matter of strengthening regulation then, but also of developing technology to enable this. The WEF identifies 4 “layers” where development is required – infrastructure, data management, and user interaction [40]:

- *Infrastructure*. Developments here should seek to “assure the availability, confidentiality, security and integrity of the data, both while in transit and at rest” and enable developers to meet the requirements of the subsequent layers.
- *Data management*. Developments here should seek to provide “a dynamic level of individual

control over data uses" and regulate the "flow and use of personal data based on specified permissions and policies". The aim here is to foster an "architecture that links permissions and provenance with the data to provide the means for upholding the enforcement and auditing of the agreed upon policies." In turn it is envisioned that such developments will create a new "social stack" that strengthens the confidence individuals have in global data flows.

- *User interaction*. Developments here should seek to provide the "elements that enable individuals to have a meaningful interaction with service providers" as per the transparency and fair value exchange requirements and specify "the permissions and policies associated with the use of their personal data."
- *Accountable algorithms*. A final technological consideration focuses "on a new nexus of control and influence: the algorithm." Algorithms reflect "the intentions and values of the individuals and institutions which design and deploy them" and "augment existing power asymmetries". The aim of development here is to ensure that algorithms are "lawful, fair and can be explained intelligibly" to individuals.

The WEF recognizes that significant development is needed for this approach to be "technologically feasible", and that these issues cannot be solved independently by either industry or government but will require a "multistakeholder approach to gain traction".

4.1 Summary

The WEF demonstrates a level of awareness and sophistication, both socially and technically, that rivals that of law and policy makers. Like them, industry emphasizes the need to engender broad societal trust in the emerging digital ecosystem and to enable increased consumer control over the actual processing of personal data, proposing a raft of social and technical measures to enable this. This is not to say that there is in anything like unanimous agreement about the WEF recommendations in industry, only that the crisis in trust and need to empower consumers is broadly recognised through its efforts.

5. The Shifting Locus of Agency and Control

The emerging digital ecosystem, and IoT in particular, is clearly a matter of broad social and economic concern, at the heart of which lies the perceived inadequacies of traditional data protection approaches. Law-makers, policy-makers and industry alike recognize the need to develop and implement new approaches to privacy that shift the locus of agency and control *towards the consumer*. In traditional data protection frameworks, the locus of agency and control centres on the 'data controller' – i.e., the party responsible for processing personal data. The various initiatives reviewed above seek to strengthen the responsibilities of data controllers, while at the same introducing measures that enable the consumer to play a much more active part not only in the collection of personal data but also in its use. The steps proposed by the various stakeholders seek to extend the consumer's role in personal data harvesting then, from an 'up front' matter of consent (with various rights attached) to an *ongoing* and *accountable* matter that runs throughout the processing of their data, including the 'back door' as the WEF puts it.

The shift is arguably occasioned by the ways in which the emerging digital ecosystem *reconfigures* human environments, which undermines the control that consumers have over their privacy. As Marmour [26] puts it,

> "The right to privacy … is there to protect our interest in having a reasonable measure of control over ways in which we present ourselves to others. The protection of this interest requires the securing of a reasonably predictable environment about the flow of information … So what would count as a violation of a right to privacy? The answer is that your right to privacy is violated when somebody manipulates … the relevant environment in ways that significantly diminish your ability to control what aspects of yourself you reveal to others … One typical case is this: you assume, and have good reason to assume, that by doing X you reveal Y to A; that is how things normally work. So you can choose, on the basis of this assumption, whether to X or not. Now somebody would clearly violate your right if he were to manipulate the relevant environment … making it the case that by doing X you actually reveal Y not only to A but also to B et al., or that you actually reveal not just Y but also W to A (and/or to B et al.), which means that you no longer have the right kind of control over what aspects of yourself you reveal to others; your choice is undermined ..."

The IoT creates fundamental change in the environments people live and operate, and the flow of information from them. The change is often glossed as a shift from 'active' to 'passive' data production, where just what is being glossed is that *the individual is being transformed* by the IoT from the user of a computer into what is effectively a core part of the machine itself, a machine that is designed to be parasitic on human action. Radical reconfiguration of the environment is accompanied then by an equally radical reconfiguration of 'the user' (though 'the used' might be more apposite at the current moment in time) in hooking 'things' up to the Internet. The problem this gives rise to, and which underpins the various initiatives outlined above, is how to mitigate this 'Orwellian' situation [19] by providing individuals with a *reasonable measure of control* over the machine to enable them to determine with a *reasonable degree of predictability* what it reveals of them to others.

5.1 HCI and the Personal Data Challenge

HCI is, of course, no stranger to the topic of privacy. Its interest reaches back to the l970's [18] and reviewing the field would be a significant job of work in itself [again see 18 for an overview]. Even if we restrict ourselves to contemporary developments, the literature is considerable and diverse in character (see, for example, [17]). There has been a wealth of work over recent years in the design community focused on understanding the personal data challenge and the IoT [e.g., 16, 21, 23, 32, 36, 41] and developing tools that enable direct end-user control over their personal data and its distribution. Examples of the latter include personal data vaults [30], personal data servers [3], and personal data stores [11]. Such research underpins the rapid commercialization of personal data management services. The WEF reports, for example, that "from January 2013 to January 2014 more than one new personal data service was launched per week" [40]. The issue of control is not a settled matter, however. As Haddadi et al. [15] put it,

> "Unfortunately, all these approaches provide both a logical and a physical single point of control over our personal data: typically, they entail lodging information in the cloud where the service is running. This naturally leads to a host of trust issues for users, who find themselves not just having to trust the service directly but also the infrastructure providers involved, other parties such as local law enforcement in the specific jurisdiction, and the possibility of collusion between these cloud services to build ever more detailed models of individuals."

Personal data management services may offer some level of control then, but not necessarily to a sufficient degree to engender the trust that is essential, a pony underscored by recent research investigating users' attitudes to such solutions [22]. Further research foundational research is required.

The development of personal data management services is complemented by efforts to enable users to track data usage. This includes recasting consent in terms of an ongoing social process rather than an up-front point of severance [25], and developing policy infrastructures that make personal data usage transparent and traceable. Examples include sticky policies [31], avenance tags [4], and consent transaction receipts [24]. Such developments have yet to come to market however, being little more than a "hypothesized" [24] "promise" [31] at this point in time, and one with its own inherent problems. The prospect of "cheating" is a real problem to be reckoned with:

> " … even with a cryptographic binding, after the personal data has been decrypted, the binding is broken in the sense that the users' data is then fully available to the authorized party and subsequent actions could be taken that contravene the policy … [Furthermore] if encryption is applied only to text files that adhere to a predefined structure, it can be relatively easy to corrupt policies; thus, a skilled hacker could tamper with the file and make the policy illegible. Watermarking schemes and obfuscation techniques also can provide content protection, but they do not ensure policy enforcement or offer protection for the data after access." [31]

Enforcement or compliance is key to the success of policy-based mechanisms. Thus in addition to further developing policy mechanisms to support traceability, there is need to develop a socio-technical infrastructure to support compliance. [4] posit an 'ecosystem' that consists of identity providers (acting as proxies for users to manage complexity), policy interfaces (for managing the complex relationships between stakeholders), and auditing parties (having the ability to detect and punish violations). Again these are not settled matters. Like control mechanisms they require foundational research, as does the challenge of making algorithms accountable, which is largely a moot point at the current moment but requires urgent attention by the design community.

5.2 From HCI to HDI: Shifting the Locus of Agency and Control

Regulators agree that there is strong need to build the user into the flow of personal data, but continue to debate the pros and cons of privacy protection frameworks. At the same time there is active and ongoing interest within design in building mechanisms that engage users in the flow of personal data. The challenge ahead is to tackle the problem head-on and develop a coherent approach towards building trust into the emerging digital data ecosystem. Current design initiatives add value but are essentially piecemeal in nature. A more holistic approach is required, which focuses expressly on the problems occasioned by *personal data*. This is reflected in the emergence of a distinctive strand of research called Human-Data Interaction or HDI [28].

The term HDI has various uses in HCI (see [7] for an overview). The version that is of interest here chimes with the kinds of design initiatives touched upon above, which seek to enable users to manage their personal data and track related data flows. A cornerstone of DARPA's Brandeis program [9], HDI seeks to develop a coherent platform that incorporates the infrastructures, data management facilities, and interfaces required to enable user control over personal data processing. An early technical attempt to put HDI into practice and build an ecosystem of 'data sources' and 'data processors' enabling user management of 'my data' is provided by the Dataware platform [27].

At the heart of HDI lies three core principles: *legibility, agency* and *negotiability* [29]. Legibility is premised on the recognition that interactions with data flows and data processes are often opaque. It is concerned with making data and analytic algorithms both transparent and comprehensible to users. Visualization is a starting point for enabling users to reason about their data and what it reveals of them, though the challenges here are considerable. Data is often poorly understood by users and the analytic techniques used to process it and draw inferences incomprehensible. Nevertheless, legibility is an essential prerequisite of users being able to exercise agency over personal data processing. Agency of course refers to the ability of users to interact with data flows and processes. This not only includes the ability to opt-in or opt-out of data collection and processing, but the broader ability to engage with its collection, storage and use, and to understand and modify data and the inferences drawn from it. Agency also implicates negotiability, enabling the ongoing engagement of users so that they can withdraw from data processes either completely or in part, and can derive value from data harvesting for themselves.

The principles of HDI underscore the need to develop a user-centric platform for personal data processing in the 21st century. While in its infancy, it is increasingly clear that HDI poses a broad range of challenges that are only now beginning to be elucidated [e.g., 7]. Many arise from engineering decisions taken early in the life of the Internet, where many features were eschewed in favor of building something that worked [5]. Thus, application (and higher) layer data flows are not a thing with which the Internet is concerned. The focus was and is on moving data packets between network interfaces, and supporting delivery of those packets to the correct application. It is hard to envision completely redesigning the entire basis of the Internet at this late stage. However, a number of discrete challenges are key to putting HDI's principles into practice.

- *Accountability*. The potential efficacy of HDI fundamentally turns upon opening the Internet up as it were and making it accountable to users. What we mean by this is that at the network layer, the Internet only really supports accounting to the extent required for settlement between ISPs (e.g., counting the number of bytes exchanged over particular network interfaces) to enable usage-based billing. With the kinds of intimate data the IoT is envisioned to make available, this low-level "bits and bytes" accounting will be completely inadequate. It will be necessary to surface what data devices generate, how that data is recorded and processed, by whom, where it flows to, etc. This metadata must be made visible to users to enable legibility, agency and negotiability without infringing users' privacy.
- *Personal infrastructures*. The advent and growth of the IoT coupled with the present lack of facility for easily managing ensembles of network-connected devices increases the likelihood that we will suffer harm by leaking intimate information. There is need to complement the opening up of the Internet with the development of personal infrastructures that enable users to manage the flow of data. One possible approach might be to provide smarter home hubs that support a range of interfaces and control points developed for specific purposes. Another is to support users in building their own infrastructure to a far greater extent than is possible today. Instead of relying on others (e.g., ISPs) to provide, configure and manage infrastructure to support users, we might seek to make it straightforward for users to create their own infrastructure services, configuring and managing facilities such as firewalling, virtual private networks, DNS and other

services.

- *Resilience*. Resilience is a key ingredient in the mix between the Internet, personal infrastructures, and IoT applications in critical domains, such as health and well-being or smart-device energy management. In short, we might ask what happens to such applications when the Internet goes down (e.g., when the router dies or there is a problem at the local exchange, etc.)? There is a critical need to build resilience into IoT infrastructures if applications in critical domains are to be relied upon. One possible solution is to build IoT infrastructure into the local physical environment – e.g., into the fabric of the home – to provide the necessary fallback. This might be complemented by formal modeling techniques to enable the 'in house' management of complex networked systems of 'dumb' devices, which in turn raises the challenge of how users are to understand such techniques and interact with them to ensure quality of service and the ongoing protection of privacy in the face of contingency.
- *Identity*. As Peter Steiner put it in a cartoon in The New Yorker (1993), "On the Internet, nobody knows you're a dog". Identity touches all aspects of HDI and requires that meaningful statements can be made about *just who* has access to a user's data. The Internet, being concerned with moving packets between network interfaces, provides no inherent support for higher-level expressions of identity. Application layer means of supporting identity do exist, - e.g., TLS client certificates and PGP public keys - but they are very complex to manage. Specific challenges here include how to ensure the availability of the necessary 'secrets' (keys, certificates) on all devices that may be used to access relevant data; how to support the management of data corresponding to multiple identities held by a user; and how to handle the revocation of access.
- *Dynamics*. Devices generating data change context as they are shared between individuals, and individuals change context as they move around in space and time. Applications and services will come and go as well. Enabling users to be aware of and to manage the dynamics of ongoing data processing – who or what has access to which data, for which purposes, etc. – is a critical challenge to the sustained harvesting of personal data. That ongoing data harvesting will be dynamic, and potentially implicate multiple parties (users and data consumers), also raises the challenge of understanding the dialogues that are needed to sustain it; particularly the 'work' these dialogues need to support and how they should be framed, implemented and maintained.
- *Collaboration*. Systems developed to support personal data management typically focus on the individual. But personal data rarely concerns a single person. It is far more common for sources of personal data to conflate information about multiple individuals, who may have different views as to how personal it is. For example, smart metering data gives a household's energy consumption in aggregate, and different household members may want that data to be shared with data consumers at different levels of granularity. Supporting the collaborative management and use of personal data is another critical ingredient in the mix, all of which trades on making the data and data processing legible and putting the mechanisms in place that enable users to exercise agency and negotiability locally amongst their own cohorts as well as globally.

5.3. Summary

HDI is a nascent field. Nevertheless, it creates significant challenges for the HCI community. Many of these may appear to be challenges for other areas of computing research at first glance, but HCI cuts through each of the matters we have outlined above. Making the Internet accountable to users; making personal infrastructures configurable and manageable by users; making formal modeling techniques, identity mechanisms and the dynamics of data processing intelligible to users; and making the collaborative management of personal data possible are all matters that require significant input from the HCI community to ensure that the HDI platform meets user needs. HCI is critical to making the shift happen and ensuring that users actually are in control of their data and can ultimately act with a reasonable degree of predictability as to what it reveals of them to others.

6. Conclusion

The annals of HCI are peppered with works that address grand challenges of the times, whether it be the problem of human-machine communication, the turn to the social, the shift away from the workplace, the challenges of ubiquitous computing in the home, post-colonial computing, etc. Shifting the locus of agency and control to enable users to effectively manage *privacy* in the emerging digital ecosystem is a grand challenge for contemporary computing in the round and HCI in particular insofar as 'the user' is its core business. This paper has sought to demonstrate that the shift is taking place. That the momentum is real and pressing. That it is driven by law-makers, policy-makers and industry leaders in response to broad societal concern and the recognition that the IoT drives radical transformation of the social fabric, which cannot be controlled by traditional data protection approaches. The proposals are radical, recasting the nature of consent and accountability from matters that occur up front to matters that permeate data processing. New privacy frameworks go beyond regulation to mandate the *building-in* of mechanisms within the emerging digital ecosystem that meet data protection needs in the 21st century and enable users to actively exercise control over the processing of personal data. However, the problem, simply put, is how to bring this about?

The current emphasis in design is on security but this is no panacea to the problem, it merely creates walled gardens with the potential for ongoing abuses of personal data to take place. Alternative approaches also seek to enable users to manage and track personal data, but these are not settled matters and require further foundational research, as does the critical challenge of making algorithms accountable to users. Furthermore, the effort to provide control is piecemeal at the current moment in time. There is then a pressing need to adopt a coherent interdisciplinary approach to the problem. Human Data Interaction provides a cogent design framework for driving a sustained approach to foundational research forwards. It is not a solution in itself but a problem that needs to be *worked out* to shape the infrastructures, data architectures, and interfaces required to actually affect the shift and put the user in control. The principles of the approach and early examples have been considered elsewhere [15, 27, 29]. Here we have sought to outline key challenges that need to be addressed, including making data flows accountable to users, enabling users to configure personal infrastructures, providing for resilience, identity, and the dynamics of data processing, and supporting the fundamentally collaborative nature of personal data management and use. Each of these challenges requires significant input from the HCI community. We would urge members to act now and "innovate along new trajectories" [13] to drive forwards the much-needed revolution in privacy occasioned by the emerging digital ecosystem and Internet of Things.

Acknowledgements

The research on which this article is based was funded by the EPSRC grant EP/M001636/1.

References

[1] Article 29 Data Protection Working Party (2010) *Opinion 3/2010 on the Principle of Accountability*, 00062/10/EN WP 173. http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp173_en.pdf

[2] Article 29 Data Protection Working Party (2014) *Opinion 8/2014 on Recent Developments on the Internet of Things*, 14/EN WP233. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2016/wp238_en.pdf

[3] Allard, T., Anciaux, N., Bouganim, L., Guo, Y., Le Folgoc, L., Nguyen, B., Pucheral, P., Ray, I., and Yin, S. (2010) "Secure personal data vaults: a vision paper", *Proceedings of 6th International Conference on Very Large Data Bases*, pp. 25-35, Singapore, VLDB Endowment. http://www.vldb.org/pvldb/vldb2010/pvldb_vol3/R02.pdf

[4] Birrell, E. and Schneider, F. (2014) "Fine-grained user privacy from avenance tags", Cornell University Technical Report. https://ecommons.cornell.edu/handle/1813/36285

[5] Clark, D. (1988) "The design philosophy of the DARPA internet protocols", *Proceedings of SIGCOM*, vol. 18 (4), pp.106-114. http://dx.doi.org/10.1145/52324.52336

[6] COM (2012) 11 Final - Proposal for a Regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation). http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:en:PDF

[7] Crabtree, A. and Mortier, R. (2015) "Human data interaction: historical lessons from social studies and CSCW", *Proceedings of ECSCW*, pp. 1-20, Oslo, Springer. http://dx.doi.org/10.1007/978-3-319-20499-4_1

[8] Cumbley, R. and Church, P. (2008) *EU – What is Personal Data?* http://www.linklaters.com/Insights/Publication1403Newsletter/PublicationIssue20081001/Pages/PublicationIssueItem3513.aspx

[9] DARPA Brandeis Program (2015) "DARPA to purse 'revolutionary' privacy tools", *Communications of the ACM*. http://cacm.acm.org/news/184344-darpa-to-pursue-revolutionary-privacy-tools/fulltext

[10] Data Protection: Country Q&A Tool, *Practical Law*. http://uk.practicallaw.com/2-502-1510

[11] de Montjoye, Y., Wang, S. and Pentland, A. (2012) "On the trusted use of large-scale personal data", *Bulletin of the IEEE Technical Committee on Data Engineering*, vol. 35 (4), pp. 5-8. http://sites.computer.org/debull/A12dec/p5.pdf

[12] Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML

[13] Greenberg, S. and Buxton, B. (2008) "Usability considered harmful (some of the time)", *Proceedings of CHI*, pp. 111-120, Florence, ACM. http://dx.doi.org/10.1145/1357054.1357074

[14] Greenleaf, G. (2010) *Comparative Study on Different Approaches to New Privacy Challenges in Particular in the Light of Technological Developments: B.5 Japan*. http://ec.europa.eu/justice/policies/privacy/docs/studies/new_privacy_challenges/final_report_country_report_B5_japan.pdf

[15] Haddadi, H., Howard, H., Chaudhry, A., Crowcroft, J., Madhavapeddy, A. and Mortier, R. (2015) "Personal data: thinking inside the box", *Computing Research Repository*, http://arxiv.org/abs/1501.04737

[16] Hong, J. and Landay, J. (2004) “An architecture for privacy-sensitive ubiquitous computing”, *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*, pp. 177-189, Boston [MA], ACM. http://dx.doi.org/10.1145/990064.990087

[17] Hong, J. and Langheinrich, M. (2014) “Privacy challenges in pervasive computing”, *Computing Now*. http://www.computer.org/web/computingnow/archive/june2014

[18] Iachello, G. and Hong, J. (2007) “End-user privacy in human-computer interaction”, *Foundations and Trends in Human-Computer Interaction*, vol. 1 (1), pp. 1-137. http://dx.doi.org/10.1561/1100000004

[19] Kroes, N. (2011) “As the IoT matures into the connected society”, *High-level Internet of Things Conference*. http://ec.europa.eu/information_society/newsroom/cf/document.cfm?doc_id=827

[20] Kugler, L. (2015) “Online privacy: regional differences”, *Communications of the ACM*, vol. 58 (2), pp. 18-20. http://dx.doi.org/10.1145/2693474

[21] Langheinrich, M. (2002) “A privacy awareness system for ubiquitous computing environments”, *Proceedings of UbiComp*, pp.237-245, Gothenburg, Springer. http://dx.doi.org/10.1007/3-540-45809-3_19

[22] Larsen, R., Brochot, G., Lewis, D., Eisma, F. and Brunini, J. (2015) *Personal Data Stores*. https://ec.europa.eu/digital-agenda/en/news/study-personal-data-stores-conducted-cambridge-university-judge-business-school

[23] Lederer, S., Hong, J., Dey, A. and Landay, J. (2004) “Personal privacy through understanding and action: five pitfalls for designers”, *Personal and Ubiquitous Computing*, vol. 8 (6), pp. 440-454. http://dx.doi.org/10.1007/s00779-004-0304-9

[24] Lizar, M. and Hodder, M. (2014) “Usable consents: tracking and managing use of personal data with consent transaction receipts”, *Proceedings of UbiComp (Adjunct)*, pp. 647-652, Seattle, ACM. http://dx.doi.org/10.1145/2638728.2641681

[25] Luger, E. and Rodden, T. (2013) “An informed view on consent for ubicomp”, *Proceedings of UbiComp*, pp. 529-538, Zurich, ACM. http://dx.doi.org/10.1145/2493432.2493446

[26] Marmour, A. (2015) “What is the right to privacy?”, *Philosophy and Public Affairs*, vol. 43 (1), pp. 3-26. http://dx.doi.org/10.1111/papa.12040

[27] McAuley, D., Mortier, R. and Goulding, J. (2011) “The dataware manifesto”, *Proceedings of the 3rd International Conference on Communication Systems and Networks*, pp. 1-6, Bangalore, IEEE. http://dx.doi.org/10.1109/COMSNETS.2011.5716491

[28] MIT Technology Review (2015) *The Emerging Science of Human-Data Interaction*. http://www.technologyreview.com/view/533901/the-emerging-science-of-human-data-interaction/

[29] Mortier, R., Haddadi, H., Henderson, T., McAuley, D. and Crowcroft, J. (2014) “Human-data interaction: the human face of the data-driven society”, *Social Science Research Network*. http://dx.doi.org/10.2139/ssrn.2508051

[30] Mun, M., Hao, S., Mishra, N., Shilton, K., Burke, J., Estrin, D., Hansen, M. and Govindan, R. (2010) “Personal data vaults: a locus of control for personal data”, *Proceedings of CoNEXT*, Article No. 17, Philadelphia, ACM. http://dx.doi.org/10.1145/1921168.1921191

[31] Pearson, S. and Casassa, M. (2011) “Sticky policies: an approach for managing privacy across multiple parties”, *IEEE Computer*, vol. 44 (9), pp. 60-68. http://dx.doi.org/10.1109/MC.2011.225

[32] Sicari, S., Rizzardi, A., Grieco, L. and Coen-Porisini, A. (2015) “Security, privacy and trust in Internet of Things: the road ahead”, *Computer Networks*, vol. 76, pp. 146-164. http://dx.doi.org/10.1016/j.comnet.2014.11.008

[33] Strategic Headquarters for the Promotion of an Advanced Information and Telecommunications Network Society (2014) *Policy Outline of the Institutional Revision for Utilization of Personal Data*. http://japan.kantei.go.jp/policy/it/20140715_2.pdf

[34] US Consumer Privacy Bill of Rights (2012) *Consumer Data Privacy in a Networked World: A*

Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy. https://www.whitehouse.gov/sites/default/files/privacy-final.pdf

[35] US Federal Trade Commission (2015) *Internet of Things: Privacy and Security in a Connected World*. https://www.ftc.gov/system/files/documents/reports/federal-trade-commission-staff-report-november-2013-workshop-entitled-internet-things-privacy/150127iotrpt.pdf

[36] Weber, R. (2010) "Internet of Things - new security and privacy challenges", *Computer Law & Security Review*, vol. 26 (1), pp. 23-30. http://dx.doi.org/10.1016/j.clsr.2009.11.008

[37] Winstein, K. (2015) "Introducing the right to eavesdrop on your things", *The Agenda Magazine*, July edition. http://www.politico.com/agenda/story/2015/06/internet-of-things-privacy-concerns-000107

[38] World Economic Forum (2013) *Unlocking the Value of Personal Data: From Collection to Usage*. http://www3.weforum.org/docs/WEF_IT_UnlockingValuePersonalData_CollectionUsage_Report_2013.pdf

[39] World Economic Forum (2014) *Rethinking Personal Data: Trust and Context in User-Centred Data Ecosystems*. http://www3.weforum.org/docs/WEF_RethinkingPersonalData_TrustandContext_Report_2014.pdf

[40] World Economic Forum (2014) *Rethinking Personal Data: A New Lens for Strengthening Trust*. http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf

[41] Ziegeldorf, J., Morchon, O. and Wehrle, K. (2014) "Privacy in the Internet of Things: threats and challenges", *Security and Communication Networks*, vol. 7 (12), pp. 2728-2742. http://dx.doi.org/10.1002/sec.795