Pers Ubiquit Comput
DOI 10.1007/s00779-016-0939-3

ORIGINAL ARTICLE

Enabling the new economic actor: data protection, the digital economy, and the Databox

Andy Crabtree[1] · Tom Lodge[1] · James Colley[1] · Chris Greenhalgh[1] · Richard Mortier[2] · Hamed Haddadi[3]

Received: 12 February 2016 / Accepted: 26 July 2016

Abstract This paper offers a sociological perspective on data protection regulation and its relevance to design. From this perspective, proposed regulation in Europe and the USA seeks to create a new economic actor—the consumer as personal data trader—through new legal frameworks that shift the locus of agency and control in data processing towards the individual consumer or "data subject". The sociological perspective on proposed data regulation recognises the reflexive relationship between law and the social order, and the commensurate needs to balance the demand for compliance with the design of computational tools that enable this new economic actor. We present the Databox model as a means of providing data protection *and* allowing the individual to exploit personal data to become an active player in the emerging data economy.

Keywords Privacy · Personal data regulation · Sociology · Digital economy · Databox

1 Introduction

Slogans such as "Big Data" and the "Internet of Things" (IoT) herald a new economic market that is largely predicated on the trading of "personal data"—i.e. data that pertain to identifiable human beings. McKinsey global estimate is that Big Data could generate from $3 to $5 trillion in value every year [1], and Gartner forecast $1.9 trillion aggregate benefit from the sale and use of IoT technology by 2020 [2]. Personal data are rapidly becoming the "new currency" [3] in the digital economy, though not without comment. A steady drip of media stories detailing the misuse and abuse of personal data is complemented by large-scale leaks, all of which combine to create broad societal concern and engender what the world economic forum (WEF) describes as a "crisis in trust" [4], a crisis that motivates new data protection regulation in a bid to rebuild consumer confidence.

The authoritative view of data regulation [5] is that it is there to protect the individual from the misuse and abuse of data that pertains to them, whether the data are generated by the individual and used by other parties or it is generated by other parties and is about an identifiable individual. The view offered here is that new data protection regulations being put forward in the USA and adopted in Europe are *also* about enabling a new kind of economic actor: an actor who is an active player in, rather than a passive victim of, the digital economy in general and the emerging data economy in particular. From this point of view, proposed data protection regulations can be seen to promote the data economy by creating legal frameworks that shift the locus of agency and control in data processing towards the individual consumer or "data subject".

This alternative perspective on data protection regulation reflects a sociological orientation to the law. From this point of view, the law is not "simply" a system of rules devised to regulate action, a mechanism of social control as it were: the system is reflexively tied to the social order [6]. Seen from this perspective, the efforts of lawmakers to

✉ Andy Crabtree
Andy.Crabtree@nottingham.ac.uk

[1] School of Computer Science, University of Nottingham, Nottingham, UK

[2] Computer Laboratory, University of Cambridge, Cambridge, UK

[3] School of Electronic Engineering and Computer Science, Queen Mary University of London, London, UK

Published online: 02 August 2016

define new regulation are not restricted to defining data protection measures and compliance procedures. They can also be seen to be concerned with creating a *new social order*, one that in this case enables the widespread and even global trade in personal data and the individual's active participation in it.

Thus, the sociological perspective shifts the focus of technology development from developing support for data protection to also envisioning how this new economic actor might be enabled through design, i.e. through the building of computational infrastructures, services, applications, and devices or "tools" in the round that enable the individual to become a player in the data market: an active *data trader*. This is not saying that the individual citizen will start to sell his or her data. This may happen to some extent but the notion of data "trading" does not necessarily imply, and nor is it restricted to, financial exchange. Indeed, it is likely that financial trading will be the weakest form of *exchange* insofar as it provides low value returns [7] and that the value of the trade in personal data to the individual and digital economy will instead be primarily derived from the exchange of data to deliver *personalised services*.

This is not to dismiss a concern with compliance, clearly the law places binding requirements on design, and it is important that developers build technology with respect to them. It is, however, to recognise that focusing on compliance alone is not sufficient to ensure the manifold social and economic benefits that are tied (at least prospectively) to the trade in personal data [8]. Building in data protection needs to be balanced then with the building of tools that enable personal data to be exploited *by individuals*. Thus, in addition to elaborating a distinctive sociological view on proposed legislation, we also articulate the Databox model, which provides an "in principle" approach to enable the compliance, control, *and* utility that is required to foster broad participation in the data economy.

The Databox model marries together the principle of Individual Control, which is core to proposed legislation [9, 10], with the local control recommendation for IoT devices and applications proposed by the European Union's Article 29 (data protection) Working Party [11] and the Utility model for personal data proposed by the WEF [4]. We elaborate each of these principles in turn and how they provide for the Databox model, which enables individual control over the flow of data in the digital economy as per the overarching goal of proposed legislation. In enabling *direct control,* the Databox model makes personal data harvesting accountable to individuals, enabling both privacy protection and the utility that are needed to deliver projected social and economic benefits.

2 The sociological perspective on legislation

The sociological perspective on the law might be viewed as new and provocative by the design community, but it is really very old and uncontroversial, reaching back to the beginnings of sociology in the nineteenth century and to Emile Durkheim in particular [12]. In many respects, the sociological perspective reminds us, as [13] puts it, of something that we all take so much for granted that we tend to *forget* it. *Ergo* the sociological perspective reflects what anyone knows about the relationship between law and society, and what anyone knows is that the law is an integral part of the *social order*, not simply in the sense that it is key to maintaining order but that it reflects in its writing, rewriting and use the order that is to be maintained. Thus, in sociological terms, the law "functions" (in contestable ways) to define and shape social order [6], which in the developed world at least is essentially *capitalist* in nature.[1]

It might be argued that this somewhat obvious but often forgotten "functional" view of the law is outdated and speaks only to the discredited theories of Structural Functionalism. Dispensing with Durkheim and Structural Functionalism more generally does not do away with the idea that the law has a sociological function; however; it only dispenses with particular *explanations* of that function. Marx, for example, saw the law as a key part of the "superstructure" of society functioning alongside other superstructural elements (e.g. politics, religion and the media) to mask the "contradictions" that capitalism depends upon for its existence [15]. Marx's explanation of the law is itself contestable and indeed contested by sociologists of different theoretical hue [16]. What is not contested, whatever theoretical perspective it is viewed from, is the fundamental observation that the law performs a sociological function which is essential to the production of social order. Sociological explanations can be dispensed with then. What anyone can see cannot. And what anyone can see is that the law is not only occupied with maintaining social order, but is clearly implicated in *re-ordering* it too.

Capitalism's historical evolution provides a ready example. It is not only different in different countries but that difference is provided for through a historically situated sequence of laws that have shaped and reshaped capitalism's unique "local" order. In the UK, for example, capitalism can be seen to have emerged locally over centuries through a succession of legal statutes regulating

[1] We use the term 'what anyone knows' in accordance with Bittner's caveat 'any normally competent, wide awake adult' [14].

labour, and not always positively.[2] Thus, the decline of feudal social order in the early part of the fourteenth century was marked by statutes such as the Ordinance of Labourers 1349 and the Statute of Labourers 1351, which sought to prohibit increases in wages and the free movement of workers (not that they were particularly effective). The same laws were still being reformed 200 years later, as reflected in the Statute of Artificers 1562, and it would be another century until the feudal social order was finally dispatched by the Tenures Abolition Act 1660. Such examples demonstrate the reflexive relationship between law and the social order, revealing its role not only in maintaining order, but also in remaking it, and creating it anew.

Thus, the old feudal order was replaced by "a new division of labour", which underpinned the wealth of nations [18]. With it, a new economic actor—one long in the making—was born. An actor whose labour was premised on a contractual relationship rather than his relationship to the feudal estate. In turn, the law came to encode this new actor and the new social order in regulation. The Employers and Workmen Act 1875 dissolved the Master and Servant Act 1823, which made breach of contract by a worker into a criminal matter. The Truck Act 1887 abolished payment in goods rather than money. The Trade Boards Act 1909 introduced minimum wage criteria, and the Representation of the People Act 1918 and the Equal Franchise Act 1928 eventually enfranchised the economic actor (male and female) in Smith's "new" social order. Thus, it continues, with an ongoing series of historically situated and locally unique laws not only regulating the social order but also, at the same time, reflexively shaping and reshaping it. This reflexive relationship between the law and social order is consequential for technology development.

The consequence turns upon setting questions concerning the meaning of the law to one side and asking instead what is its sociological function? When viewed from this perspective, the debate about what the law requires of design with respect to privacy and the processing of personal data shifts from a matter of understanding data protection measures and compliance procedures to understanding the *social arrangements* the law seeks to bring about through such measures and procedures. This, to reiterate, is not to set a concern with data protection and compliance aside. It is to ask *what kind of social order does the law seek to create?* It is this foundational matter that we take for granted and all too often forget when considering matters of law. Nevertheless, it is a matter that concerns us here and is one that we seek to address in considering proposed data protection regulation in Europe and the USA and its relevance to technology design.

3 Data protection legislation sociologically construed

Data protection regulation generally focuses on the obligations of the "data controller"—i.e. the party who determines the purposes for which and the manner in which personal data are processed—and regulates the act of "data processing", which may be carried out by another party on the controller's behalf (including machines). It also specifies the rights of "data subjects"—i.e. living individuals to whom personal data relate. There is much about the obligations of data controllers and processors in proposed European and American regulation. However, in both cases, it is clear that regulation is not "simply" concerned with specifying data protection measures. The *economy* looms large in both sets of proposals.[3]

In draft European legislation [9], the need to revise data protection regulation is firmly *premised* on economic considerations. The explanatory memorandum prefacing the proposal outlines the concerns that motivate the introduction of the new data protection framework. Thus, it is explained that "heavy criticism", "particularly by economic stakeholders", motivates the need to "adapt" the existing framework due to "fragmentation" in the ways in which data protection is currently implemented across the Union, and the need for "increased legal certainty" and "harmonisation of rules" across international borders given the "rapid development of new technologies". These concerns "constitute an obstacle to the pursuit of economic activities" and "distort competition".

> This is why it is time to build a stronger and more coherent data protection framework in the EU, backed by strong enforcement that will allow the digital economy to develop across the internal market, put individuals in control of their own data and reinforce legal and practical certainty for economic operators and public authorities.

The economic imperative is similarly marked in draft the US legislation. The proposed Consumer Privacy Bill of Rights [10] seeks to extend the reach of the Federal Trade Commission's Fair Information Practice Principles

[2] This of course is not to deny the influence of other social factors on the rise of capitalism, including the development of machines, financial markets, and the protestant work ethic [17], it is merely to point out that the legal system played a formative role too.

[3] We are aware that new regulation is also being proposed in Japan [19]. Here too the emphasis is on enabling the "utilisation" of personal data in order to "revitalise the economy".

(FIPPs). While FIPPs is not enforceable, it does form the basis of laws regulating the use of personal data in specific sectors (e.g. health, education, finance). The proposed bill "carries FIPPs forwards" and seeks to apply it through self-regulation enforced by the FTC Act (Sect. 5) prohibiting "unfair or deceptive acts or practices" to "the interactive and highly interconnected environment in which we live and work today". Although adopting a different approach to data protection, the concerns that motivate the proposed bill are similar to those in Europe. Thus, the proposed bill of rights seeks to address the problems occasioned by a fragmented "sectorial" environment, provide "greater legal certainty" to companies, and "create interoperability between privacy regimes" in order to "promote innovation" and "drive the digital economy".

Evidently, the *purpose* of proposed legislation is not "simply" to lay down and spell out data protection measures and compliance procedures. It does this of course, but to a *social* rather than a legal end: to engender individual or consumer *trust*. Furthermore, as the following extracts make clear, the purpose of proposed regulation is not to engender trust per se, but to engender trust *in the digital economy*; an economy that increasingly relies upon the trade in personal data.

> Preserving trust in the Internet economy protects and **enhances substantial economic activity**. Online retail sales in the United States total $145 billion annually. New uses of personal data in location services, protected by appropriate privacy and security safeguards, could create important business opportunities. Moreover, the United States is a world leader in exporting cloud computing, location-based services, and other innovative services. To preserve these economic benefits, consumers must continue to trust networked technologies. **Strengthening consumer data privacy protections will help to achieve this goal**. [10, our emphasis]

> The scale of data sharing and collecting has increased dramatically … Building trust in the online environment is **key to economic development**. Lack of trust makes consumers hesitate to buy online and adopt new services. This risks slowing down the development of innovative uses of new technologies. **Personal data protection therefore plays a central role in the Digital Agenda for Europe, and more generally in the Europe 2020 Strategy**. [9, our emphasis]

Clearly, new data protection regulation is motivated by economic concerns, but what of the new economic actor? Where is the individual or consumer as data trader and linchpin of the data economy? Proposed EU regulation states that it seeks to "put individuals in control of their own data" through the implementation of "appropriate technical" (as well as organisational) "measures" that apply at the time of "the design of [data] processing" and at the time of "the processing itself." These measures should provide for informed consent "at the time of [data] collection or within a reasonable period" and informed choice through the implementation of "certification mechanisms and data protection seals" that allow individuals to "quickly assess the level of data protection" offered by digital products and services. Furthermore, individuals should be able to "obtain a copy of the data concerning them" and "transmit those data" from one automated application into another one to "further strengthen the control over their own data" [9].

The US Consumer Privacy Bill of Rights similarly seeks to provide "consumers who want to understand and control how personal data flow in the digital economy with better tools to do so." The proposed bill goes a step further than the EU proposal, however, and seeks to enshrine the principle of Individual Control in regulation:

> Consumers have a right to exercise control over what personal data companies collect from them and how they use it.

The Individual Control principle is the first of seven key "rights" laid out in the draft bill and has two key aspects to it: one, "providing consumers with easily used and accessible mechanisms" with which to exercise control and two, "consumer responsibility", which recognises that the use of personal data turn upon the individual's decision to share data with others. Indeed, the draft bill views control "over the initial act of sharing" as "critical." This turns upon consumers having the tools and mechanisms to hand to make informed decisions and exercise control. The draft bill suggests that "innovative technology can help to expand the range of user control" and cites examples such as "detailed privacy settings", "do not track", and "opt out" mechanisms. However, it also goes so far as to say that while such mechanisms "show promise" they "require further development" [10].

Now, it might be argued that this is a thin legal basis on which to ground the claim that proposed regulation seeks to enable a new economic actor. However, we are not making a legal argument but a sociological one. From this perspective, the need to enable *individual control over the flow of personal data* in the digital economy is plain to see in both EU and the US proposals, and it is on this basis that we say proposed legislation seeks to *shift the locus of agency and control* in data processing towards the individual. Furthermore, as it is also plain to see, the measures proposed to affect this shift are not purely legal in nature—

not "simply" a matter of specifying data protection measures and compliance procedures—but reach out to "technical measures", "tools", "easily used and accessible mechanisms", and "innovative technologies" to *enable* the actor's participation *in* the digital economy.

The underlying need to enable the new economic actor—the individual as data trader—through technology development can be further apprehended when we turn to those parties tasked with transforming legislation (actual and potential) into best practice guidance; in this case, the Article 29 Working Party (WP29), established under the 1995 Data Protection Directive, and the Federal Trade Commission (FTC) were tasked with enforcing data protection in the US. Both parties have issued guidance with regard to the Internet of Things (IoT), which is set to be a primary engine of personal data production and distribution, over the last 2 years. Both parties offer a broad range of recommendations for best practice to industry. Of particular note here are those recommendations that speak to the principle of Individual Control.

The FTC proposes a number of practical measures to put the individual in control of personal data generated by IoT devices [20]. These include "general privacy menus" enabling the application of user-defined privacy levels (e.g. low, medium, high) across all their IoT devices by default. The use of icons on IoT devices to "quickly convey important settings and attributes, such as when a device is connected to the Internet" and to enable individuals to quickly "toggle the connection on or off." The use of "out of band" communications to convey important privacy and security settings via other channels, e.g. via email or SMS and the use of management portals or "dashboards" that enable individuals to configure IoT devices and accompanying privacy settings.

> Properly implemented, such 'dashboard' approaches can allow consumers clear ways to determine what information they agree to share.

WP29 also proposes a number of practical measures "in order to facilitate the application of EU legal requirements to the IoT" [11]. These include providing individuals with "granular choices" over data collection, including the "time and frequency at which data are captured", and scheduling options to "quickly disable" data capture. Individuals should also be "in a position to administrate" IoT devices "irrespective of the existence of any contractual relationship" and "easily export their data" from IoT devices "in a structured and commonly used format." Furthermore, settings should be provided that enable individuals to distinguish between different people using shared devices "so that they cannot learn about each other's activities." Most of these recommendations complement the dashboard approach towards putting the principle of Individual Control into practice, insofar as they are to do with providing and enabling individuals to specify *privacy settings*.

The data portability requirement is unique, however, as is the *local control* recommendation:

> To **enforce transparency and user control**, device manufacturers should provide tools to locally read, edit and modify the data **before they are transferred to any data controller**.
>
> Device manufacturers should enable **local** controlling and processing entities allowing users to have a clear picture of data collected by their devices and facilitating **local** storage and processing **without having to transmit the data to the device manufacturer**. [11, our emphasis]

The local control recommendation is radical. It undermines the current approach to privacy being widely adopted by industry—i.e. encryption—which puts personal data online for cloud processing *before* making it available to the individual. As Winstein [21] 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 into make sure the device isn't tattling on them, the effect is anti-consumer.

Security is of course needed, but the current model as described by Winstein does not satisfy the Individual Control principle and neither is it sufficient to satisfy individual privacy requirements, as encryption does not stop device manufacturers from exploiting an individual's personal data. The local control recommendation provides an alternative pathway, one that allows designers to strike a balance between privacy protection and the individual control needed to enable the new economic actor.

4 Striking the balance

The sociological perspective on legislation makes it perspicuous that the principal *function* of proposed regulation is to engender consumer trust in the digital economy. This raises the issue of balancing the design of tools that enable data protection with the building of tools that enable the individual's participation in the digital economy. The need to strike this balance is underscored by the World Economic Forum (WEF), which emphasises a "lack of empowerment" as a key issue "undermining trust" in the digital economy [4]. The WEF recognises that current data

protection approaches reflect "an asymmetry in power that broadly favours institutions (both public and private)", which enables them to "orient notice and consent agreements to advance their interests." The best practice guidelines outline above may go some way to redress the imbalance.

However, another key issue wrapped up in this asymmetrical relationship, and one that is often overlooked, concerns "individuals are being able to use their own data for their own purposes", which is an area where the "power dynamics" (or differentials) really bite. As Lanier [22] puts it,

> The dominant principle of the new economy … has been to conceal the value of information.

Following Lanier, the WEF argues that individuals not only need to be able to "assert more control" over data processing, but also be able to *benefit* from the ways in which personal data "are leveraged and value distributed". Thus, the WEF proposes an "alternative model" that enables personal data to "be used as a utility" by the individual, rather than it being something that is simply handed over to others albeit with appropriate notice and consent agreements in place.

The WEF goes on to suggest that this alternative *utility model* might be enabled through the development of Personal Data Management Services (PDMS). It notes that "there is growing momentum in the area" and that "more than one new personal data service was launched per week" between January 2013 and January 2014 [4]. Despite growing commercial interest, public uptake of PDMS, such as MyDex or OpenPDS, has been problematic. A recent report suggests that this might be due to "perceptions of privacy and security risks" that consumers attach to storing their personal data on cloud-based services [23]. The situation is compounded by the fact that personal data are distributed across a great many silos (e.g. Facebook, Google, Twitter, etc.), with no standard data formats, no standard APIs for access, and no easy way of obtaining a *holistic* overview. Furthermore, as [24] point out, most personal data do not belong to a single individual but are *social* in nature (e.g. communications data), and PDMS solutions have yet to address this foundational matter.

Current PDMS approaches do not strike the balance then between data protection and control, let alone enable personal data to be used as a utility for individual benefit. An alternative approach is provided by the local control recommendation—i.e. developing local PDMS rather than cloud-based ones. One such example is provided by the Databox model [25]. At the centre of this model sits a physical device located in the individual's home, which is under the direct control of the individual. The device allows the individual to collect a distributed array of physical (e.g. sensors) and digital (e.g. internet or social media) "data sources". Data sources may, then, connect directly or indirectly to the device (e.g. via an embedded VPN server and/or SOCKS proxy) to enable the individual to control access to their personal data.

The device or "Databox" provides a *gateway* to an individual's, or collection of individuals (e.g. a family's) data sources. The Databox leverages a "containerised" approach to data processing, enabling data to be held in stores that can be written to by data sources but are isolated from reading by data processors until appropriate permissions are presented. For additional security, these data stores may be implemented as "unikernels", i.e. application-specific virtual machines that eschew use of a general purpose operating system with the attack surface and management problems it entails, for a library operating system approach where only the specific system-level code required by the data store is linked into the resulting unikernel.[4]

This approach enables raw data to be retained by the individual and supports both local processing of data "requests" and local hosting of computation, which includes running algorithms on the box to deliver local services. Selected raw data, at a chosen granularity, can be released to specific data processors should the individual wish to do so, though processing should still be limited to only those operations that have been explicitly permitted by the individual. In each case, data are encrypted and tagged [26] in a bid to prevent data processors using the data for any but the specified purposes. Data transactions may also be lodged with a trusted third party or a distributed ledger to enhance accountability. The Databox is embedded in an interactional systems model that enables both compliance with proposed data protection legislation and the utility that is needed to drive active participation in the digital economy and achieve the overall goal of proposed legislation.

4.1 The Databox model

The Databox model assumes that a number of distinct actors, of which there may be many of each, are directly implicated in data processing:

1. The individual or "data subject".
2. The "data controller" or party who wants to consume an individual's data for some (lawful) purpose.
3. A "data processor" or party who carries out data processing on the controller's behalf, which we assume will be a machine.[5]

[4] http://unikernel.org.

[5] The model also assumes that data subjects may consume one another's data for 'domestic purposes' as provided for by existing and proposed regulation, which *exempts* data processing done for such purposes from the requirements of that regulation.

4. An intermediary, which enables data controllers to discover data providers and data subjects to discover data consumers.

This mediated model goes beyond "walled in" data transfers between, for example, IoT device manufacturers and individuals to enable the broader use and reuse of personal data and open up the data market.

The Databox model puts in place a set of interactional arrangements and supporting system architecture that enables data subjects and data controllers to exploit an individual's data for mutual benefit and, at the same time, enables demonstrable compliance with proposed legislation, particularly the "external data subject accountabilities" it requires: i.e. transparency and consent, granular choice, data portability, and access. Interaction between the parties to personal data processing (i.e. the actors) is provided in the following ways.

The data subject first configures data sources. This entails associating physical (local) and online (remote) data sources with the Databox. Data sources may then be assigned to ownership (e.g. collective, shared by specific individuals, or a single individual) and be annotated (e.g. fridge smart plug, kitchen humidity sensor, etc.). Individual accounts may also be created to enable individuals to manage the data sources they own (including shared data sources). The data subject may then register with an intermediary "discovery service". This entails establishing a secure association with the service, e.g. setting up an account and lodging an authenticated public encryption key. The data subject may then post metadata about the data sources they own and wish to make available to data consumers.

Before a data controller can access a data subject's data sources, they must also register with the discovery service and create an account. This also entails establishing a secure association with the service, as well as declaring the legitimate purposes for which the controller seeks to process personal data, the kinds of data sources it wishes to exploit, and registering any data processor APIs. The latter enable individual access to processed data and allow data subjects to inspect data uses, retention, sharing, etc. The data controller can also post containerised "apps" on the discovery service, which can be downloaded by data subjects and enable data processing or the local hosting of computation on the Databox.[6]

The discovery service reviews a data controller's application, rates it based on the information provided (e.g. no processor API might result in a poor rating) and issues a revocable machine-readable token that will allow the data processors acting on the controller's behalf to search the data source registry for the required data sources. The discovery service also enables individuals to post reviews about data consumers and rate them. Reviews and ratings are lodged with the controller's account. Ratings are displayed alongside apps on the Databox, from where reviews can also be accessed, and the data subject can actively search the service via the box to find reviews and ratings for other data controllers should they wish.

Interaction between data subjects and data controllers is mediated by a "multi-layered notice" [27] providing a service level agreement or SLA (Fig. 1) that identifies the controller, the purposes of processing, and the other mandatory information that is required to be provided to the data subject prior to data processing by existing and proposed legislation. The SLA also defines the benefits of data processing, and the risks that attach to particular categories of data (e.g. that occupancy can be inferred from $C0_2$ data). The data subject may use data visualisation apps to preview the data that are requested by the controller and also exercise granular choice over data collection via the SLA, configuring which data sources may be used and setting data sampling frequencies. This may reduce the service options that are available to the individual, which is dynamically reflected in the SLA.

SLA's are attached (like terms and conditions) to apps. An app cannot be used without an SLA being in place and data cannot be transferred to a controller's processors without an SLA being completed. SLAs are machine configurable, though it is assumed that they will initially be drafted by human actors (i.e. the controller's representatives). Once an SLA is accepted, the Databox either enables local computation (e.g. allows an algorithm contained in an app to access data sources and deliver a local service) or runs a data processing request on the box and returns the results to the controller's processors. As noted above, raw data streams from specific data sources may on occasion, as the data subject sees fit, also be made available to the controller's processors.[7]

Data subject interactions are provided for through the Databox Catalogue. In addition to the interactions outlined above (data source configuration, discovery service registration, metadata publication, app discovery, ratings and reviews, and SLA configuration), the Catalogue enables data processing auditing. Auditing enables the data subject to inspect all data processing operations, historical and live

[6] Data subjects can discover apps, and with them data consumers, from the Databox. Apps may also be provided by other parties. They are automatically made available to the data subject based on the data sources an individual owns and are used locally for various purposes including data processing, analytics, and visualisation.

[7] The discovery service provides a 'domestic purposes' SLA, which enables individuals to make specific data sources available to one another. Individuals may also use apps to share the results of data processing with one another should they wish retain control over data sources.

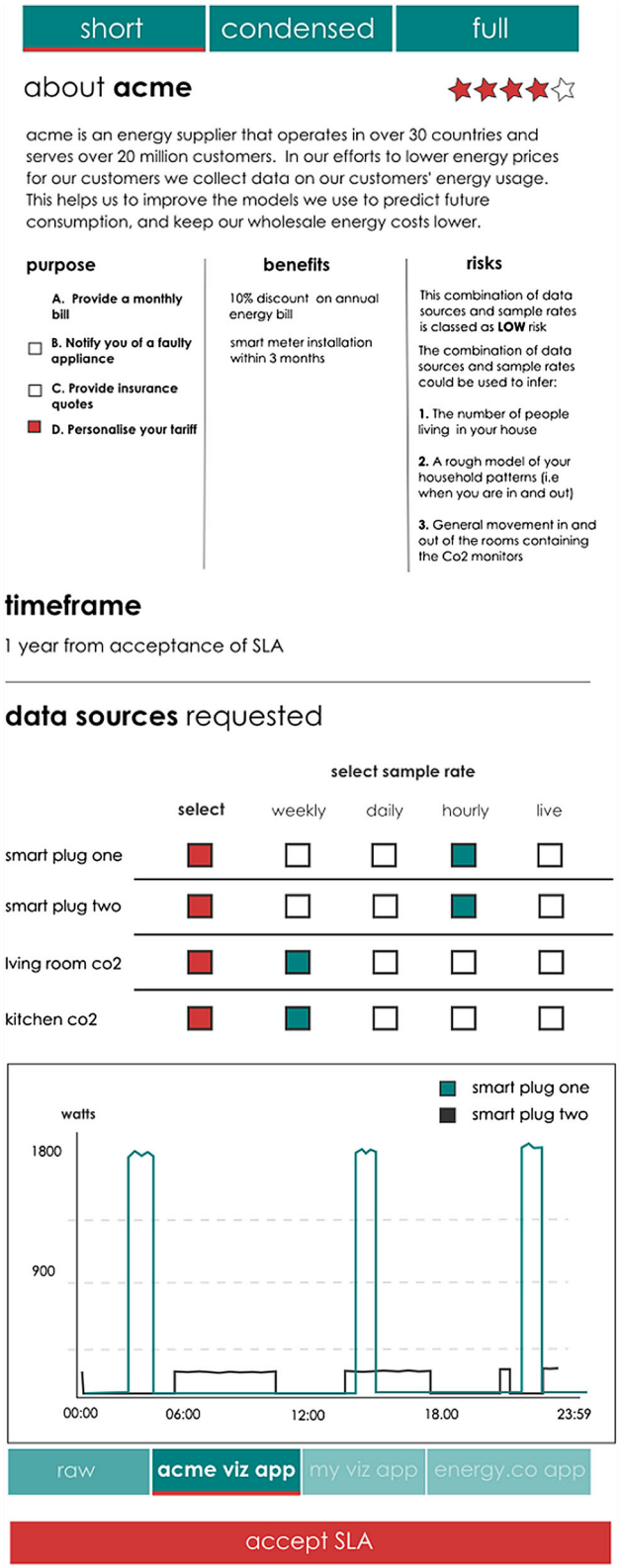

Fig. 1 Databox SLA

that have been permitted on the box and the SLAs that attach to them. The audit log may be lodged and updated dynamically with a trusted third party or distributed ledger (as may the processor's). The Catalogue also provides a messaging service that enables data subjects and data consumers to communicate with one another (e.g. to identify faulty data sources, such as a sensor, or faulty appliances that may, for example, need replacing).

Architecturally the Databox model consists of three key components: the Databox, a controller's processors, and the discovery service (Fig. 2). The Databox is a small form factor computer consisting of a web server and webapp containing the catalogue UI, which supports user interaction with apps and data. Apps, like data stores, run within isolated containers (e.g. using Docker) and interact with APIs provided by the Databox to perform a task. For example, apps may use the Databox's datastore API to query data sources for processing or the comms API to send data to external machines. The comms API is responsible for recording transactions which are encrypted and signed/countersigned by the Databox and recipient of data and stored in the transaction log. Accounts, raw data and indexes, and metadata are also stored on the box. Restrictions on the use of the APIs are determined by an app's SLA. Apps are installed/removed/updated using the Databox's app manager API.

The discovery service is a cloud-based service, which is interacted with using standard internet protocols (principally HTTPS). It consists of a web server and webapp containing the discovery UI providing for human interaction and a query API providing for programmatic (machine-based) interaction. It contains a key and security association manager or key server, which is utilised by Databoxes and a data controller's processors for signing data transactions. The discovery API allows data subjects to upload data source metadata (via the catalogue UI) and is stored as Databox metadata, which is made available to humans via the discovery UI or machines via the discovery API. The discovery service manages an app repository of all apps, which are uploaded via the app API and indexed by the metadata they operate on. It provides tools to help data controllers publish apps, one of which will be a set of skeleton SLA templates that can be specialised according to the particular aims of an app. And it manages accounts for data subjects and consumers, along with rating/reviews.

The discovery service provides most of the resources; a data controller requires to exploit the Databox model. However, the controller will need to put in place sufficient resources to support their own operation. While the specific components needed to process data will vary from case to case, all controllers will need to deposit an app in the app repository to support their operations and we anticipate that they will want to keep a record of data transactions and thus require a transaction log to meet their accountability requirements to supervisory authorities. They will also need to store their private keys. Minimally, perhaps

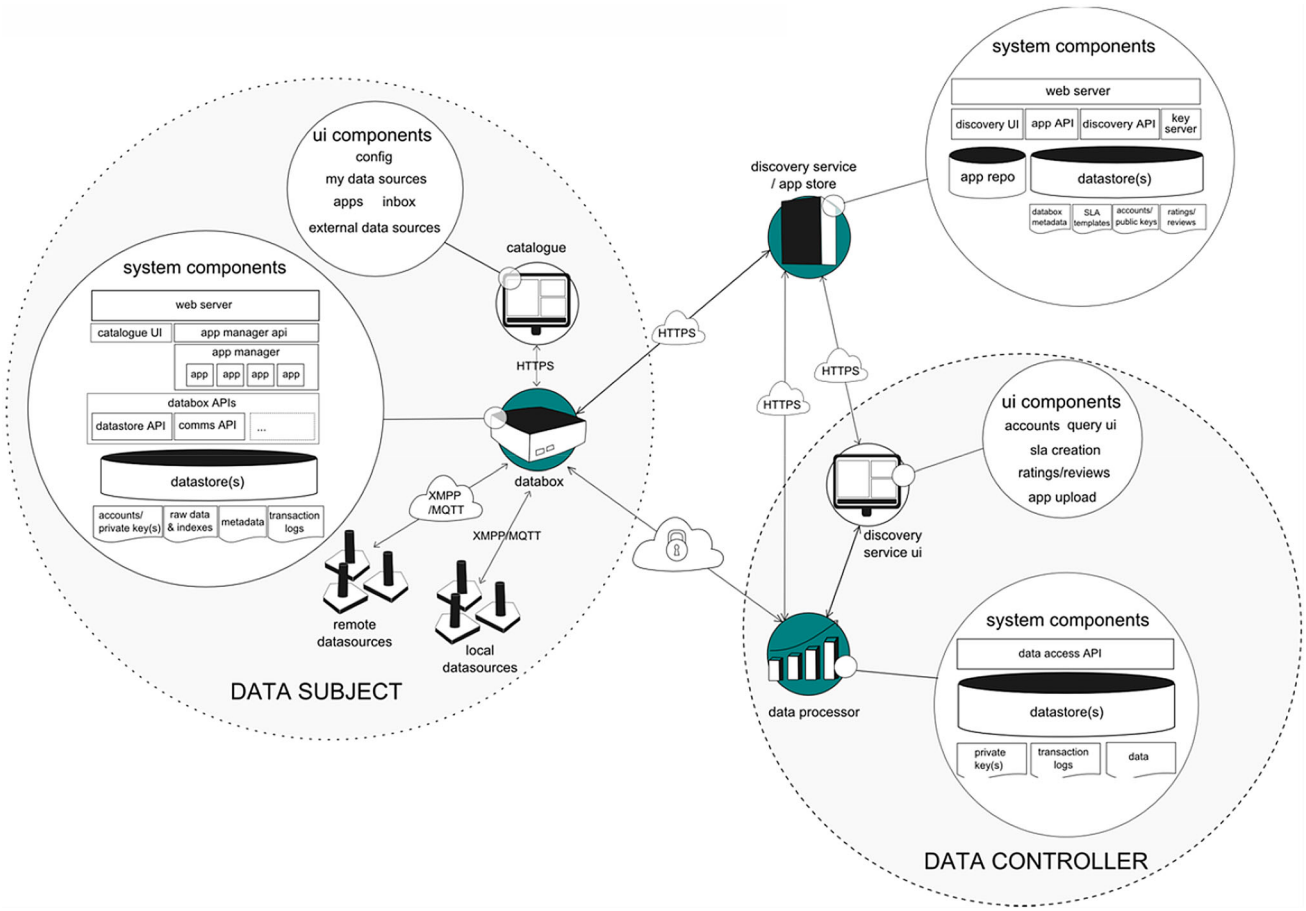

Fig. 2 Databox model

optimally, a controller's app might perform local processing and/or data visualisation but entail no data export. An SLA will still be required, but no further components are needed here. Where an app exports data, however, the data controller is responsible for providing a secure data endpoint and an encrypted connection for data transfer. Controllers are also encouraged to provide a data access API. This is not mandatory, but it is desirable as it enables a controller to meet the data subject accountability requirements of existing and proposed legislation and thus allows individuals to gain further assurances on how their data are used.

4.2 The Databox: enabling protection and utility

The Databox model provides an "in principle" approach towards meeting the sociological function of proposed data protection legislation. It combines the core principle of *individual control* with the *local control* recommendation to deliver a *utility* model that enables the data subject to throttle and drive the flow of data in the digital economy *and* to exploit data for personal benefit as she/he sees fit.

It enables privacy protection in supporting the external data subject accountabilities required by proposed legislation:

- Transparency and consent, and granular choice, both of which are provided for through the Databox SLA.
- Data portability, which is provided for through the collation of local and remote data sources via the Databox.
- Access, which is provided for through data processor APIs.

While the latter cannot be enforced by the Databox model, legislation requires that access be provided by data controllers to data subjects. The Databox model provides a coherent ecology for data controllers to demonstrate compliance with the accountability requirements of data protection legislation, and *demonstration* is key:

> Accountability refers to a company's capacity to demonstrate the implementation of enforceable policies and procedures relating to privacy (whether adopted voluntarily or as a result of legal obligations). [10]

Utility is enabled through the app-based approach that enables data controller's and other parties to create a *familiar environment* enabling the delivery of digital services predicated on the use of personal data. Just as apps are now, then so it will be for the "user" to decide which ones they wish to make use of to meet their needs and which not. But more than that, the Databox model enables the data subject to make a fundamental choice: to use accountable personal data services "on the box" or to use "unboxed" services and accept the increased privacy risks that accompany them. The Databox model is currently under development. As Naughton [28] puts it,

> Getting from here to a service that is usable by normal human beings will, no doubt, be a long and winding road.

However, insofar as it enables privacy protection *and* the utility that is essential to the digital economy, it would appear to be a road worth exploring.

5 Conclusion

The core proposition of this paper is that when viewed from a sociological perspective the law and proposed data protection regulation in the US and Europe, in particular, functions to create new social arrangements that enable a new kind of economic actor: the individual as data trader. While there is much about data protection and compliance in proposed legislation, the economy looms large and clearly motivates its introduction. Proposed legislation is not "simply" about putting protection measures and compliance procedures in place then. It is also, at the same time, about shifting the locus of agency and control to foster trust in and enhance the digital economy.

The shift is made perspicuous in the emphasis placed on the principle of *individual control* in proposed legislation. It is apparent too that the measures proposed in draft legislation to affect this shift are not purely legal in nature. Enabling the new economic actor is not only a concern for members of the legal profession then, but for technology developers as well, who legislators anticipate will drive innovation and provide the tools and resources that will actually enable the actor to exercise control over the flow of personal data in the digital economy.

The need to enable the new economic actor through design is underscored by the best practice guidance offered by the FTC and WP29, which emphasise the building-in of mechanisms to support the "external data subject accountabilities" of proposed legislation: transparency and consent, granular choice, data portability, and access. However, the most radical suggestion is encapsulated in the *local control* recommendation, which seeks to allow individuals to locally control data processing entities and view, read, modify, and edit data *before* they are transferred to a data controller.

The need to put the individual in control is further underscored by the WEF, which identifies the asymmetry in power between individuals and organisations as a key driver of the public crisis in trust in the digital economy. The WEF proposes the adoption of a *utility* model that not only enables individuals to control the flow of personal data in the digital economy, but to derive *personal value* from it as well. Data protection, on this view, is not sufficient in itself then. Enabling the new economic actor, though not necessarily a financially motivated actor, is also required if the emerging data economy is to thrive and deliver anticipated benefits.

In response to these issues, we have presented an "in principle" approach towards enabling the protection *and* utility that are needed to enable the new economic actor. Thus, the Databox model combines the principle of individual control with the local control recommendation to provide a utility model that enables the individual to control the flow of personal data and, at the same time, embeds the flow of data within a sociotechnical ecology that enables data consumers to demonstrate compliance with the accountability requirements of proposed legislation.

The Databox model has the "in principle" potential to meet the needs of data subjects and data controllers, providing *both* with the tools they need to exploit personal data and comply with the requirements of data protection regulation. In doing so, it has the potential to meet the sociological function of legislation and thus bring about the new social arrangements that are sought by proposed legislation, building trust into the personal data ecosystem and enabling the individual to be an active participant in, rather than a passive victim of, the digital economy. The Databox model is currently under development. Future work will report on the implementation and in-the-wild deployment of the Databox to further explore the model's real world, real time viability.

Acknowledgments The authors acknowledge the support of the EPSRC, Grants EP/M001636/1, EP/M02315X/1, EP/N028260/1, and EU FP7 Grant 611001.

References

1. Manyika J, Chui M, Groves P, Farrell D, van Kuiken S, Almasi Doshi E (2013) Open data: unlocking innovation and

performance with liquid information. McKinsey and Company, New York
2. Middleton P, Kjeldsen P, Tully J (2013) Forecast: the internet of things worldwide. Gartner, Stamford
3. Gates C, Matthews P (2014) Data is the new currency. In: Proc. of new security paradigms workshop. ACM, Victoria, pp 105–116
4. World Economic Forum (2014) Rethinking personal data: a new lens for strengthening trust. http://www3.weforum.org/docs/WEF_RethinkingPersonalData_ANewLens_Report_2014.pdf. Accessed 12 Feb 2016
5. Gov.UK (2015) Data protection. http://www.gov.uk/data-protection/the-data-protection-act. Accessed 12 Feb 2016
6. Ferrari V (2007) Functions of law. In: Clark D (ed) Encyclopedia of law and society: American and global perspectives. Sage, New York, pp 611–617
7. van Rijmenam M (2015) From data ownership to data usage: how consumers will monetise their personal data. https://datafloq.com/read/data-ownership-data-usage-consumers-monetize-data/68. Accessed 12 Feb 16
8. BCG (2012) The value of our digital identity. Liberty global
9. EU (2012) General data protection regulation. http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2012:0011:FIN:en:PDF. Accessed 12 Feb 2016
10. US (2012) Consumer data privacy in a networked world. https://www.whitehouse.gov/sites/default/files/privacy-final.pdf. Accessed 12 Feb 2016
11. WP29 (2014) Opinion 8/2014 on recent developments on the internet of things. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp223_en.pdf. Accessed 12 Feb 2016
12. Durkheim E (1893) The division of labour in society. Presses Universitaires de France, Paris
13. Barnes JA (1966) Durkheim's division of labour in society. Man 1(2):158–175
14. Bittner E (1973) Objectivity and realism in sociology. In: Psathas G (ed) Phenomenological sociology. Wiley, Hoboken, pp 109–125
15. Marx K (1859) A contribution to the critique of political economy. Progress Publishers, Moscow
16. Bourdieu P (1987) The force of law: toward a sociology of the juridical field. Hastings Law J 38:814–843
17. Weber M (1905) The protestant ethic and the spirit of capitalism. Unwin Hyman, Crows Nest
18. Smith A (1776) An inquiry into the wealth of nations. Methuen & Co, London
19. Japan (2014) Policy outline of the institutional revision for utilization of personal data. http://japan.kantei.go.jp/policy/it/20140715_2.pdf. Accessed 12 Feb 2016
20. FTC (2013) 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. Accessed 12 Feb 2016
21. Winstein K (2015) Introducing the right to eavesdrop on your things. The agenda magazine, July edn. http://www.politico.com/agenda/story/2015/06/internet-of-things-privacy-concerns-000107. Accessed 12 Feb 2016
22. Lanier J (2013) Who owns the future?. Simon and Schuster, New York
23. Larsen R, Brochot G, Lewis D, Eisma F, Brunini J (2015) Personal data stores. https://ec.europa.eu/digital-agenda/en/news/study-personal-data-stores-conducted-cambridge-university-judge-business-school. Accessed 12 Feb 2016
24. Crabtree A, Mortier R (2015) Human data interaction: historical lessons from social studies and CSCW. In: Proc. of ECSCW. Springer, Oslo, pp 1–20
25. Chaudry A, Crowcroft J, Howard H, Madhavapeddy A, Mortier R, Haddadi H, McAuley D (2015) Personal data: thinking inside the box. In: Proc. of critical alternatives. ACM, Aarhus, pp 29–32
26. Pearson S, Casassa M (2011) Sticky policies: an approach for managing privacy across multiple parties. IEEE Comput 44(9):60–68
27. WP29 (2004) Opinion 10/2004 on more harmonised information provision. http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2004/wp100_en.pdf. Accessed 12 Feb 2016
28. Naughton J (2015) Fight back against internet giants' stranglehold on personal data starts here. The guardian, February 1, http://www.theguardian.com/technology/2015/feb/01/control-personal-data-databox-end-user-agreement. Accessed 12 Feb 2016