Accountable Internet of Things?

Outline of the IoT Databox Model

Andy Crabtree, Tom Lodge, James Colley, Chris Greenghalgh
School of Computer Science
University of Nottingham, UK
{first.last name}@nottingham.ac.uk

Richard Mortier
Cambridge Computer Lab
University of Cambridge
United Kingdom
richard.mortier@cl.cam.ac.uk

***Abstract*—This paper outlines the IoT Databox model as a means of making the Internet of Things (IoT) accountable to individuals. Accountability is a key to building consumer trust and mandated in data protection legislation. We briefly outline the 'external' data subject accountability requirement specified in actual legislation in Europe and proposed legislation in the US, and how meeting requirement this turns on surfacing the invisible actions and interactions of connected devices and the social arrangements in which they are embedded. The IoT Databox model is proposed as an in principle means of enabling accountability and providing individuals with the mechanisms needed to build trust in the IoT.**

Keywords—Internet of Things (IoT); trust; accountability; IoT Databox

I. Introduction

While the Internet of Things (IoT) holds great social and economic promise, it is accompanied by deep-seated concerns that drive new approaches to data protection in Europe and the US. The EU has adopted new General Data Protection Regulation (GDPR), which will come into force in 2018 [1], and the US is reconsidering the proposed Consumer Privacy Bill of Rights [2,3,4]. Key to both is the notion of *accountability*, which refers to the demonstrable implementation of enforceable privacy policies and procedures. Any organisation using IoT technologies to process personal data in these economic zones will have to comply with the accountability requirement.

Two fundamental categories of accountability are to be found in actual and proposed legislation: 'internal' and 'external'. Internal accountability refers to the policies and procedures that a data processing entity puts in place to *demonstrate to itself* that its processing operations comply with the requirements of data protection regulation (e.g., privacy impact assessments). More relevant to this paper is the notion of 'external' accountability, which refers to the policies and procedures a data processing entity puts in place to *demonstrate to others* that its operations are legally compliant. Those 'others' fall into two categories: regulatory authorities, and the 'data subjects' or individuals whose data is processed. The individual accountability requirement concerns us here.

In the following section we briefly outline the individual accountability requirement as given in actual and proposed legislation and how it has been translated into 'best practice recommendations' for IoT developers by data protection agencies. These recommendations seek to enable individual control over the flow of personal data through the design of computational mechanisms that make data collection transparent, enable consent, and permit fine-grained data flow management, data portability and access. Satisfying the accountability requirement requires that we surface and articulate hidden aspects of the IoT: not only machine-to-machine or M2M actions and interactions but also, and importantly, the social arrangements connected devices are embedded in, for it is not only the data collected by Internet-enabled 'things' that must be made accountable but also *what* is done with that data and by *whom*.

We outline the IoT Databox model as a means of making device actions and interactions and the social or cooperative arrangements they are embedded in transparent to enable accountability. The IoT Databox is an edge device intended to be situated within the home. It collates data from IoT devices, either directly or via APIs, and makes them available to 'apps' that enable data processing and actuation. Data processing takes place on-the-box. This has a range of potential benefits including resilience (actuation does not need to rely on continuous connectivity), low latency (data does not have to be moved to and from remote data centres), efficiency (centralised data processing costs are significantly reduced), and data minimisation (only the results of processing queries are distributed). Making the IoT accountable may, then, have manifold advantages.

II. The Accountability Requirement

The accountability requirement plays a key role in the processing of personal data. 'Personal data' are any data that relate to an identifiable person, including data generated by connected devices. 'Processing' includes collecting, using, retaining, disclosing and/or disposing of personal data. In addition to specifying that personal data processing be conducted under the auspices of the 'data minimisation' principle (Article 5), EU legislation or GDPR [1] specifies that data processing should also be lawful. The lawfulness of data processing turns in significant part upon satisfying the external data subject accountability requirement, particularly a number of matters to do with 'consent' as laid out in Articles 6 to 22.

These include (but are not limited to) specifying the purposes of data processing, the period for which data will be stored, whether or not the data will be transferred to an international organisation or third country, and the level of protection afforded in such circumstances by reference to an adequacy decision by the European Commission. Data reuse does not require consent if reuse is compatible with the original purposes

978-1-5386-2723-5/17/$31.00 ©2017 IEEE

for which data was gathered [5]. Data subjects must be informed of their rights, including the right of access and the right to lodge a complaint (Article 15) the right to rectification (Article 16), the right to be forgotten and to erasure (Article 17), the right to data portability (Article 20), and the right *not* to be subject to measures which produce legal effects based solely on automated processing, including profiling (Article 22). Where automated decision-making, including profiling, is applied the logic, significance and envisaged consequences of data processing must be conveyed to the individual (Article 13). The required information must be provided in an intelligible form, using clear and plain language (Article 12), and at the time when data is obtained (Article 13).

Consent is a legal requirement of all personal data processing operations in Europe, except for specified exemptions including those done for personal or household purposes or purposes. Data processing occurring in such contexts is exempt from proposed regulation. However, the exemption does not apply "to controllers or processors which provide the means for processing personal data for such personal or household activities" [1, paragraph 18]. 'Controllers' are parties who commission data processing. 'Processors' are parties who act on the controller's behalf, and may include computational machines. Failure to comply with the external data subject accountability requirement may result in a fine of up to 4% of an organisation's annual worldwide turnover.

While the US proposal is voluntary, and such punitive sanctions are therefore absent, the emphasis on external data subject accountability is just as pronounced and enshrined in 6 of the 7 principles that underpin the proposed Consumer Privacy Bill of Rights Act of 2015 [2]; the other principle focuses on security (Section 105). The first of these principles, *Transparency* (section 101), stipulates that individuals should be provided with clear descriptions of the personal data to be collected and processed, data retention policies, access mechanisms, and any other entities the data will be disclosed to. The *Individual Control* principle (Section 102), stipulates that individuals should be provided with reasonable means to control personal data processing, including means to withdraw or limit consent that are as easily used as methods for granting consent in the first place.

The *Respect for Context* principle (Section 103) stipulates that data should be processed in a manner that is reasonable "in light of context" and otherwise implement "heightened" transparency and individual control. The *Focused Collection and Responsible Use* principle (Section 104) stipulates that an entity may only collect, retain, and use personal data in a manner that is reasonable "in light of context". The *Access and Accuracy* principle (Section 106) stipulates that individuals must be provided with reasonable access to, or an accurate representation of, personal data under the control of the processing entity, and means to dispute and resolve accuracy or completeness to avoid "adverse" consequences. The *Accountability* principle (Section 107) stipulates that data processing entities should build appropriate consideration for privacy and data protections into the design of its systems and practices.

While actual and proposed legislation is clearly different in the EU and US, the emphasis on external accountability to the individual is pronounced in both. This is unsurprising insofar as both are built on shared Fair Information Practice principles [6]. These principles encourage processing data for specific purposes, and for purposes that are contextually consistent or compatible with those for which data was originally collected, and mandate the provision of information about data processing to individuals in ways that clearly describe the purposes of data processing, data retention policies, data transfer policies, data access policies, etc.

Accountability thus 'frames' the processing of personal data, i.e., it does not specify *how* accountability is to be achieved, only *what* it must consist of and amount to, a matter succinctly summed up in the 2012 draft of proposed US bill: "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) [4]." Making data processing accountable to the individual whose data is processed is key to the demonstration. Furthermore, it is a demonstration that must be provided by any party using the IoT to process personal data in Europe, and it may soon be a requirement in the US.

III. Supporting the Accountability Requirement

The primary purpose of the accountability requirement is to demonstrably put the individual "in control of their own data" [1] and to enable them "to control how personal data flows in the digital economy" [4]. Historically, the accountability requirement has been treated as a non-functional requirement; a matter of providing information to enable decision-making (so called 'notice and choice'). However, 'best practice recommendations' from data protection agencies in Europe and the US - the Article 29 Data Protection Working Party or WP29 and the Federal Trade Commission or FTC respectively – make it clear that the accountability requirement is now a functional requirement to be implemented through design.

Thus, the FTC [7] proposes a number of practical measures to put the individual in control of personal data generated by IoT devices. These include "general privacy menus" enabling the application of user-defined privacy levels across all of 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 users to quickly "toggle the connection on or off"; the use of "out of band communications" to relay important privacy and security settings to the user via other channels, e.g., via email or SMS, and the use of management portals or "dashboards" that enable users 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 (ibid.)."

WP29 [8] also proposes a number of practical measures to facilitate the application of EU legal requirements to the IoT. These include providing users with "granular choices" over data collection, including the time and frequency at which data are captured, and scheduling options to "quickly disable" data capture. Users should also be in a position to administrate IoT devices and easily export their data in a "structured and commonly-used format". Furthermore, settings should be provided that enable users to distinguish between different

individuals using shared devices so that they cannot learn about each other's activities. Data portability aside, these recommendations complement the dashboard approach recommended by the FTC, insofar as they are concerned to put computational mechanisms in place that allow end-users to specify privacy settings and thereby constrain the flow of data between one another and the wider world.

Thus accountability essentially becomes a matter of enabling individual control over the flow of personal data through the design of computational mechanisms that make data collection transparent, enable consent, and permit fine-grained data flow management, data portability and access. The transparency specification makes it necessary to surface and articulate hidden aspects of the IoT. Not only M2M actions and interactions implicated in data processing, which are largely invisible at this moment in time [9], but also the social arrangements those device interactions are embedded in [10]. This is necessary as it is not only the data collected by Internet-enabled 'things' that must be made accountable and subject to individual control, but also what is done with that data and by whom. In short, it is necessary to make devices, data controller's and their processors accountable too.

The WP29 recommendations add one further specification to the accountability requirement: "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 [8]." One implication of this is that a great deal of the IoT data processing that currently takes place in the cloud is moved to the edge of the network to enable *local* control, thus minimising if not entirely dispensing with the distribution of personal data and the privacy threat that accompanies data distribution. In doing so, there is the added benefit of resilience in actuation, low computing and communications latency, and a significant reduction in data processing costs.

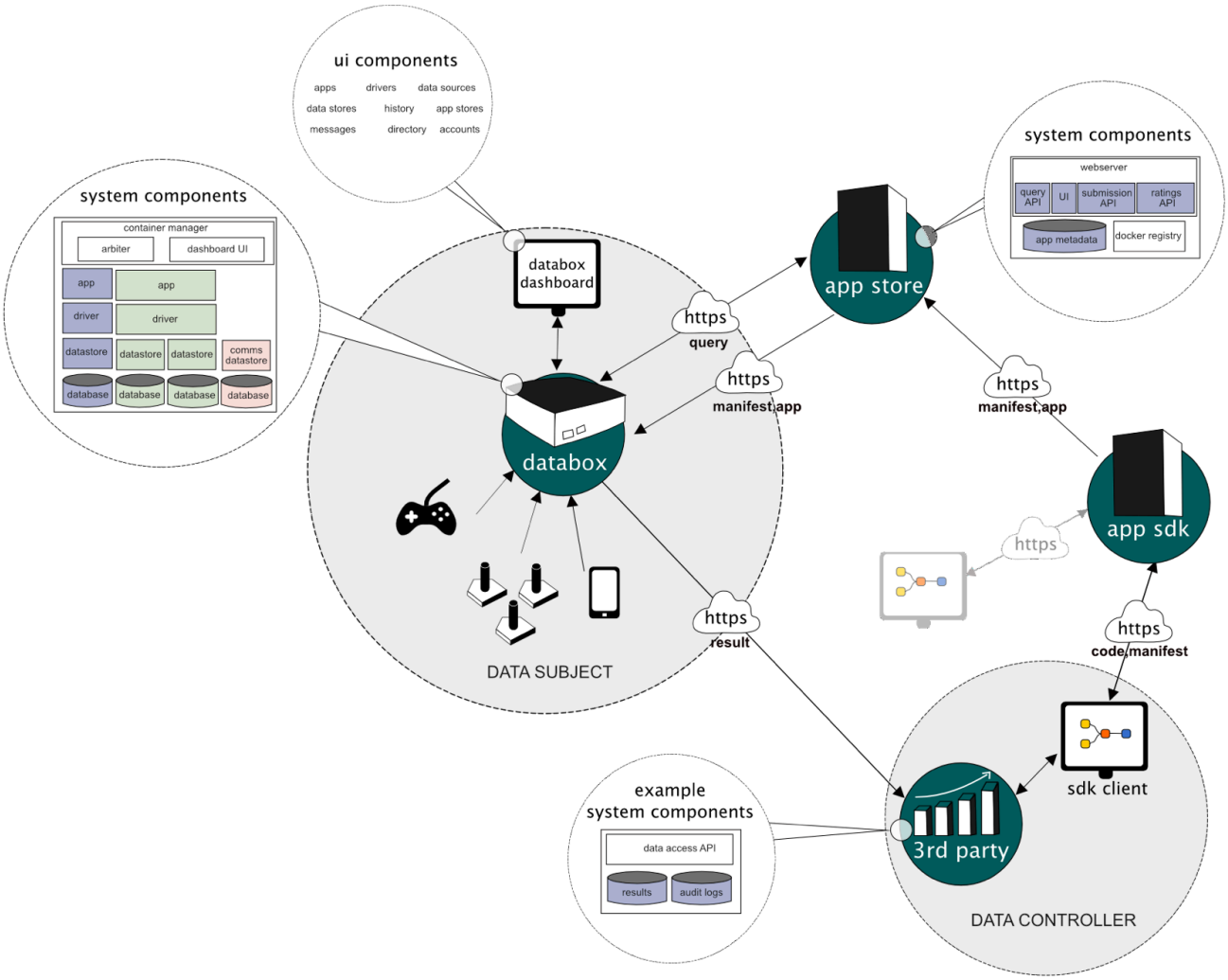

Fig. 1. Implementing the Accountability Requirement: The IoT Databox Model.

IV. The IoT Databox

The IoT Databox model provides an in principle means of implementing the accountability requirement. The model extends the Databox concept [11] to incorporate the IoT. The Databox concept posits a physical device as a gateway to a distributed platform and is predicated on the 'Dataware model' [12]. The model is a socio-technical model which implicates *the user* (by or about whom data is created), *data sources* (e.g., connected devices, which generate data about the user), a *personal container* (which collates the data produced by data sources and can be accessed via APIs), a *catalogue* (which allows the user to manage access to the personal container), and *data processors* (external machines exploited by data controllers who wish to make use of the user's data in some way).

The Dataware model is a logical entity formed as a distributed computing system. Data processing involves requests being sent to the catalogue, which are approved or rejected by the user. If approved, the catalogue issues a processing token to the data processor for permitted requests. The processor presents the token to the personal container, which accepts the token, runs the processing request on the relevant data sources, and then returns processed results to the data controller. The Dataware model represents a distinctive approach to personal data processing, that not only seeks to enable user control but also data minimisation. Thus, the Dataware model takes a significant step towards implementing the local control recommendation, minimising data sharing to the results of processing. The raw data remains 'on the box' under the users control.

The Dataware model is currently being reconfigured around the Databox concept, which embeds the Dataware model in a physical object situated in the physical environment (e.g., a networked mini-computer in the home) under the direct control of the individual. It allows the individual to collate data from an array of data sources in a single place and allows the individual to control access to them. Data from individual data sources is stored in 'unikernels', i.e., containerised, application-specific, virtual machines that reduce the attack surface and management problems associated with general purpose operating systems.

Architecturally the IoT Databox model consists of three key components: the Databox, the app store, and a controller's processors (Fig.1). The Databox is a small form factor (x86 or ARM) computer consisting of a web server and webapp containing the *dashboard* (Fig.2), which enables individuals to:

- Create *User Accounts* on the Databox and activate sharing permissions (e.g., that consent from all users of shared resources is required for delete actions).
- Add *Data Sources* to the box, including assigning ownership to data sources, annotating data sources (e.g., smart plug XYZ is 'the kettle'), and sharing data sources with other Databox users.
- Configure *Drivers* to enable data sources to write to data stores.
- Manage *Data Stores*, including sharing stores with other Databox users, and redacting, clearing, or deleting stores.
- Access *App Stores*, apps are recommended by the box based on available data sources but individuals can also search for, download, and rate apps.
- Share *Apps*, with other Databox users within the home and between distributed Databoxes in other homes; the Dashboard also allows apps to be updated and deleted.
- Receive *Notifications*, including sharing requests, app updates, resource contention, etc.
- *Audit* data processing operation, including all accesses to data stores, and any data transactions.

The app store is a cloud-based service, interacted with using standard internet protocols (principally HTTPS). It consists of a web server that provides the *app store UI* supporting human interaction, and a *query API* providing for programmatic (machine-based) interaction. The app store manages a *docker repository* of apps, which are uploaded via the *app submission API* and are indexed by associated metadata. App developers are free to create their own containerised apps as they wish, but the apps store provides a dedicated *app SDK* supporting the app building and publication process. This is a cloud-hosted visual code editor based on IBM's open source Node-RED, which utilises a flow-based programming paradigm in which black-box processes called 'nodes' are connected together to form applications called 'flows'.

There are three principle node types: *data sources*, *processes* and *outputs*. *Process* nodes are functions that operate on data; they typically have a single input connection and one or more output connections. *Output* nodes typically perform an action, such as actuation, visualisation, or data export. Figure 3 depicts a flow taking the output from a microphone on a mobile phone, visualises its amplitude on a graph and turns a plug on or off when the amplitude exceeds a particular value. It is composed of a single *data source* (the yellow node), two *processes* (the two blue nodes) and two *outputs* (the orange nodes).

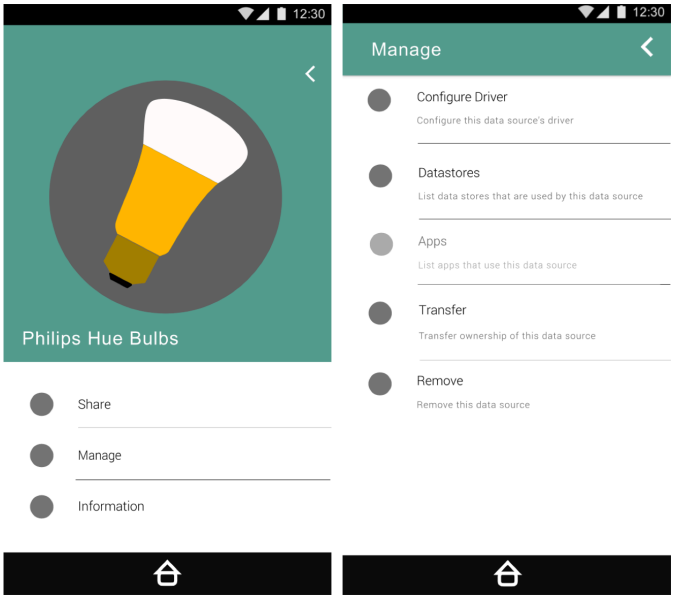

Fig. 2. The IoT Databox Dashboard.

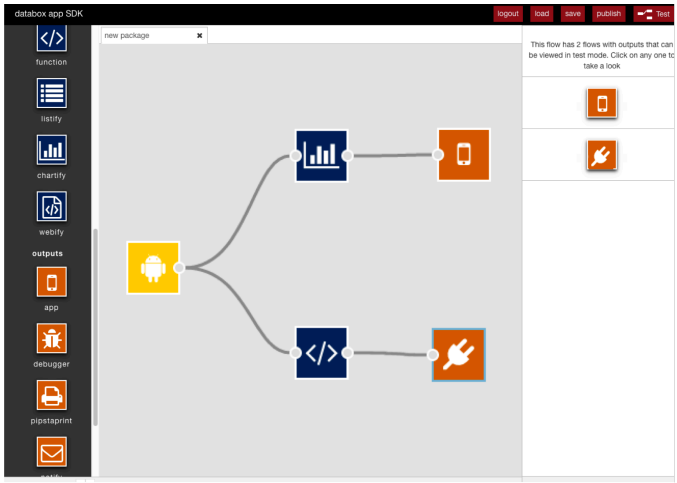

Fig. 3. The App SDK.

The app editor smooths and simplifies the build / test / deploy development workflow; it presents a high-level abstraction (e.g. an app developer can build an app without needing to be familiar with the interoperation between sources, stores and drivers); it provides 'scaffolding' to help build an app, e.g., developers can quickly inspect the structure and type of data entering and exiting a node; it provides a full testing environment, where flows are deployed (as containers) and connected to test data; and it handles the app publication process by presenting tools for building a *manifest* enabling end-user consent and granular choice (Fig.4); and, upon submission, containerising an app and uploading it to an app store. The SDK also takes care of source code management as all stages of the app development cycle are recorded in a developer's GitHub account.

On receipt of an app, the app store *reviews* and *rates* it based on its features and information provided, e.g., the absence of a data access API would result in a poor rating if data was taken off-the-box by an app (user ratings and reviews are also displayed alongside apps on the app store). An app cannot be posted on the app store or installed on the IoT Databox without a manifest being in place, and data cannot be transferred to a controller's processors without a manifest being completed by the data subject. Manifests are dynamic, user-configurable 'multi-layered notices' [13] that surface and articulate who wants to access which connected devices and what they want to process personal data for. Thus manifests make specific socio-technical data processing arrangements, implicating connected devices, data controller's and their processors accountable to individuals and available to local control.

Manifests provide an easy to read description of the *purposes* of processing and service options, the *benefits* of data processing, and the *risks* that potentially attach to particular categories of data (e.g., that occupancy can be inferred from C02 data). The manifest allows the data subject to exercise fine-grained *granular choice* over data collection, configuring which data sources may be used at which sampling frequencies. This may reduce the service options that are available to the individual, which is dynamically reflected in the manifest. Mandatory information required by data protection legislation is provided in the 'condensed' and 'full' layers of the manifest.

Once a manifest has been configured by the individual and has been 'accepted' it assumes the status of a Service Level Agreement or *SLA*, which the IoT Databox transforms into a set of machine readable policies that *enforce* a data processor's access to the particular data sources agreed upon by the individual and regulate subsequent data processing operations. Data sources implicated in an SLA may be changed, e.g., to replace a faulty sensor, and settings can be updated to enable new service options. Apps, like data stores, run within isolated containers and interact with data stores to perform a specified ('purposeful') task. Thus apps may query data stores, write to a communications data store and send query results to external machines, or write to a connected device's store to perform actuation. Data stores record all actions performed on them (queries, external transactions and actuation) in an audit log. Access to data stores and processing restrictions are determined by an app's SLA and enforced by the *arbiter*, which issues and manages the use of access tokens.

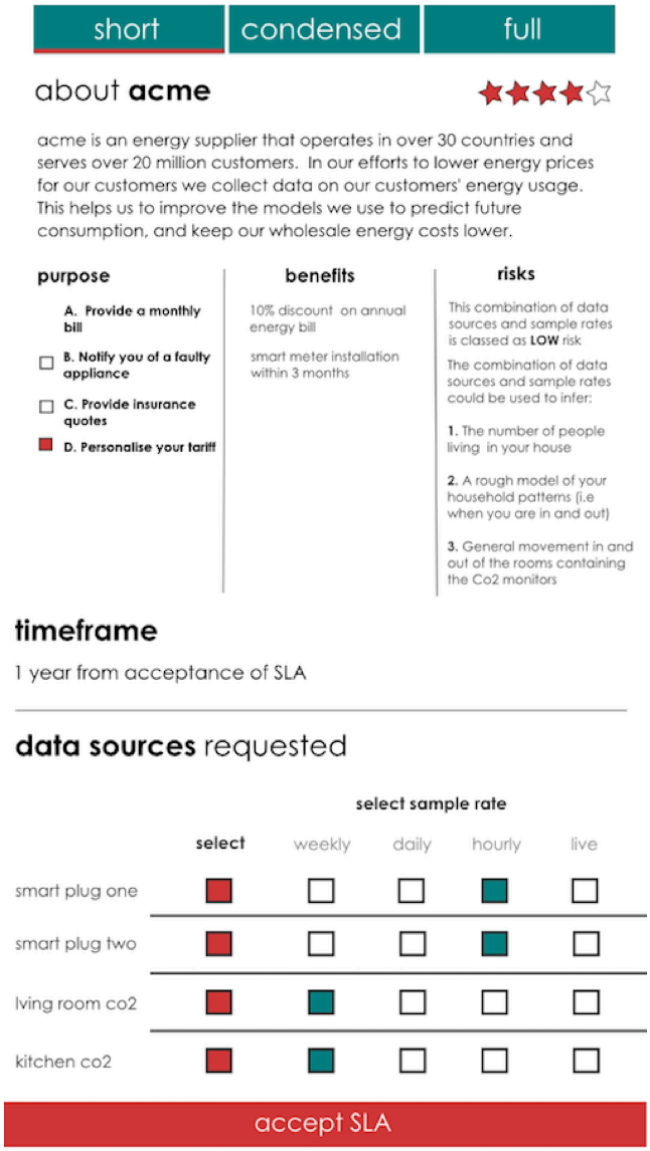

Fig. 4. The IoT Databox manifest/SLA.

An app may implement data processing, including actuation, locally and entail no transfer of data off-the-box, thus meeting the local control recommendation [8]. A manifest/SLA will still be required but no further components are needed in this scenario. Apps transferring the results of local processing meet the data minimisation principle [1]. Nevertheless, where an app exports data, even if only the results of processing, then the data controller is responsible for providing a *secure data endpoint* and an encrypted connection for data transfer, which the box will monitor. Controllers are also encouraged to provide a *data access API*. While access is a requirement of proposed legislation, it is not mandatory in the IoT Databox model as it is not enforceable. However, and as noted above, that an app does not support data access may be leveraged as a visible disincentive for individuals to use it.

The IoT Databox model thus puts in place a set of interactional arrangements and supporting system architecture that enables *demonstrable* compliance with legislation, actual and proposed, providing the transparency and consent, granular choice, data portability (data is on the box), and potential access needed to meet the external data subject accountability requirement. Insofar as apps satisfy the local control recommendation, keeping data processing on-the-box, then the IoT Databox model arguably circumvents the need for data protection regulation at all, as *no data* is handed over to an organisation for processing. Insofar as processed data might be exchanged, then the IoT Databox model enforces the data minimisation principle *and* creates incentives for data controllers and their processors to comply with the requirements of regulation in furnishing data access APIs: "Where possible, the controller should be able to provide remote access to a secure system which would provide the data subject with direct access to his or her personal data [1, paragraph 63]." Of course, in an IoT Databox world, such access will be a readily accountable and mundane feature of app use.

V. Conclusion

"Data protection must move from 'theory to practice' … accountability based mechanisms have been suggested as a way of … implement[ing] practical tools for effective data protection [14]."

This paper has sought to explore how data protection might move from theory to practice with respect to connected devices through the development of accountability based mechanisms provided by the IoT Databox model. The model seeks to respond to the external data subject accountability requirement of actual and proposed legislation in Europe and the US. This user-facing requirement seeks to make data collection transparent, enable consent, permit fine-grained data flow management, data portability, and access, and allow entities that process personal data to *demonstrate* that these measures are in place.

The IoT Databox model responds to the accountability requirement's shift in functional status and provides for the demonstration by surfacing interactions between connected devices and data processors, and articulating the social actors and activities in which machine-to-machine interactions are embedded, through the construction of manifests that accompany (and must accompany) data processing apps and tools (dashboards and app stores) to manage app use. In adopting the local control principle to ensure the individual can control the flow of personal data, the IoT Databox model enhances the efficiency of data processing, makes actuation more resilient, minimises the impact of IoT traffic on the network, and negates the need for costly privacy regimes: if no data is taken off the box then there is no need for data protection. Insofar as it is possible for data processing and data to demonstrably *stay on-the-box* then the IoT Databox model also holds the promise of opening up personal data as never before, allowing data processing across manifold sources of personal data rather than single connected devices.

The IoT Databox model is not a theoretical model. It exists, albeit in nascent form [15]. It enables data controllers and app developers working on their behalf to demonstrate compliance with the external data subject accountability requirement of actual and proposed legislation. It's ability to support local computation minimises and even circumvents the widespread threat to privacy occasioned by the IoT. And in circumventing the privacy threat, it opens up new possibilities for exploiting personal data in ways that may demonstrably build consumer trust in the digital economy.

Acknowledgment

The work reported here was supported by EPSRC research grants EP/M001636/1, EP/N028260/1 and EP/M02315X/1.

References

[1] General Data Protection Regulation, in Official Journal of the European Union, vol. 59, pp. 1-88, May 2016.

[2] Chernichaw, A. and Freeman, B.C., "White House Re-introduces Consumer Privacy Bill of Rights Act", in White & Case Newsflash, April 2015.

[3] Adminstration Discussion Draft: Consumer Privacy Bill of Rights Act of 2015,https://www.democraticmedia.org/sites/default/files/field/public/2015/draft_consumer_privacy_bill_of_rights_act.pdf

[4] Consumer Data Privacy in a Networked World: A Framework for Protecting Privacy and Promoting Innovation in the Global Digital Economy, in Journal of Privacy and Confidentiality, vol. 4 (2) , pp. 95-142, March 2013.

[5] WP203 Opinion 03/2013 on Purpose Limitation, Article 29 Data Protection Working Party, April 2013.

[6] Gellman, R., Fair Information Practices: A Basic History, 2016. http://bobgellman.com/rg-docs/rg-FIPShistory.pdf

[7] Internet of Things: Privacy and Security in a Connected World, FTC Staff Report, January 2015.

[8] WP233 Opinion 8/2014 on Recent Developments on the Internet of Things, Article 29 Data Protection Working Party, September 2014.

[9] Ziegeldorf, J., Morchon, O. and Wehrle, K., "Privacy in the Internet of Things: threats and challenges", in Security and Communication Networks, vol. 7 (12), pp. 2728-2742, 2014.

[10] Robertson, T. and Wagner, I., "CSCW and the Internet of Things", in Proc. of ECSCW, pp. 285-294, Oslo, Springer, 2015.

[11] Chaudry A., Crowcroft J., Howard H., Madhavapeddy A., Mortier R., Haddadi H. and McAuley, D., "Personal data: thinking inside the box", in Proc. of Critical Alternatives, pp. 29-32, Aarhus, ACM, 2015.

[12] McAuley, D., Mortier, R. and Goulding, J., "The dataware manifesto", in Proc. of 3rd International Conference on Communication Systems and Networks, pp. 1-6, Bangalore, IEEE, 2011.

[13] WP202 Opinion 02/2013 on Apps and Smart Devices, Article 29 Data Protection Working Party, February 2013.

[14] WP173 Opinion 3/2010 on The Principle of Accountability, Article 29 Data Protection Working Party, July 2010.

[15] The Databox Project, http://www.databoxproject.uk/code/