Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 52 (2015) 59 – 66

Procedia Computer Science

The 6th International Conference on Ambient Systems, Networks and Technologies (ANT 2015)

Policy-Carrying Data: A Step Towards Transparent Data Sharing

Julian Padget[a], Wamberto W. Vasconcelos[b,*]

[a]*Dept. of Computer Science, University of Bath, Bath, BA2 7AY, U.K. j.a.padget@bath.ac.uk*
[b]*Dept. of Computing Science, University of Aberdeen, Aberdeen, AB24 3LT, U.K., w.w.vasconcelos@abdn.ac.uk*

Abstract

The emerging research and application domains of the Internet-of-Things and Big Data, together with socio-technical phenomena such as social networking, as well as mobile phones (equipped with sensors, GPS, etc.) make us – companies, research centres, people in general – all (sometimes unwittingly, possibly unwillingly) producers and consumers of data. A sensitive issue for data providers concerns control over access, sharing, dissemination and use of data. By control, we mean placing limitations on *who* can access the data, *when* data can be accessed, *how* data can be accessed, and so on. We propose means to capture the expression of controls over data and to associate that indivisibly with the data via what we call "policy-carrying data" (PCD). The PCD establishes permissions for what the consumer may do to the data, but also – in a novel addition for such policies – establish what the consumer can/must do subsequently with the data. We formalise PCD and illustrate how it can meet potential stakeholder requirements.

Peer-review under responsibility of the Conference Program Chairs

Keywords: Policies; data access; Internet-of-things

1. Introduction

The ever-reducing cost of sensors, their increasing variety, fuelled by advances in smart materials, and their incorporation in every day devices are indeed realising the Internet of Things and are capable of creating the foreseen data deluge. The availability and the volume of data motivate first-order questions such as (i) what data can be sensed physically and (ii) what information can be recovered, although second-order issues such as (i) to whom the data belong and to whom information derived from the data belongs (ii) to whom the knowledge derived from this information belongs and to what risks stakeholders are exposed as a result of this information and this knowledge remain less explored. These less explored issues may hinder progress of the first-order issues through the creation of regulation, where, although there is plenty of privacy legislation, none was designed for these circumstances; legal thinking and interpretation operate at a much slower pace than technological innovation.

* Corresponding author.
E-mail address: w.w.vasconcelos@abdn.ac.uk

Peer-review under responsibility of the Conference Program Chairs
doi:10.1016/j.procs.2015.05.020

We propose means to capture the expression of controls over (derived) data and to associate that indivisibly with the data via what we call "policy-carrying data" (PCD[1]). Following policy conventions, in defining the who, the when and the how, the PCD establishes permissions for what the consumer may do to the data. A novel aspect of our proposal is the establishment of obligations concerning what the consumer should do with the (derived) data and it is these that are the foundation of transparency. Such obligations are the transactional unit for a non-pecuniary data economy, where access to and use of data may be traded for obligations that act as a form of user-definable, liquidity at-point-of-use community currency [14]. These obligations may pertain directly to actions of data consumers or – and this is another significant novelty of our approach – indirectly to the policy associated either with the extracted data or the data derived from them.

A feature of the current data landscape is the relative freedom of movement of data from individuals to the data silos used in cloud computing and thence between silos, which could be viewed as contributary to the disempowerment that individuals might feel over their own data – privacy controls aside [5]. The situation is potentially further complicated since the platform may enable the collection and interpretation of those data, thus adding value to them, as in the case of activity-monitoring devices or home energy monitors. The PCD concept associates data with bespoke policies: for example, framework policies might be defined by legislation, while specific policies for individual needs would have to satisfy the norms established at the primary level [13]. In this way, crisp but unworkable definitions of issues such as "When do data stop being private?" and "How to decide if data revelation is in the public interest?" can be blurred as distinctions are established to meet the needs of a given situation.

Legislation will necessarily lag behind practice, but also, we contend, can only ever offer a high-level framework because of the potentially huge variety of regulations that seem likely to be required. Furthermore, by its nature legislation (and its revision) tends to take a retrospective view rather than being defined with foresight in mind. The concept of data wrapped up in a policy – policy-carrying data (PCD) – forms the basis for the realization of a data-sharing economy based on the transfer of obligations to provide an achievable combination of privacy and transparency rather than unachievable – and arguably undesirable – absolute privacy. Thus, we offer a formal model of PCD and we show how this model can be put to use via reasoning mechanisms and illustrative PCDs.

The rest of the paper is organized as follows. We put forward a formal and a computational model for the combination of policies and data, presenting and justifying a reference architecture in Section 2. In Section 3 we present our formalism for PCDs – its syntax and operational semantics. In Section 4 we outline reasoning mechanisms which stakeholders can make use of when using PCDs. We discuss our approach and contrast it with related work in Section 5 and we conclude in Section 6, setting out some avenues for future work.

2. A Framework for Policy-carrying Data

In this section we set out a reference framework within which we situate and connect stakeholders, PCDs and an information model. We illustrate our framework in Fig. 1, where we show stakeholders (circles), processes (arrows) and information model (boxes within central box). The stakeholders envisaged are (i) data owners/producers who make data/information available (represented as the left-hand circle) and, in a richer version of the model, these may be separated into those that assert rights over the data and those that publish it; (ii) data consumers who want to access data (represented as the right-hand circle) and again in a richer version of the model, there may be entities that are both consumers and producers of data, either by offering aggregation services or by adding value in some way; (iii) monitor/police who are responsible for monitoring/policing the publication and access activities (represented by the upper circle in the middle).

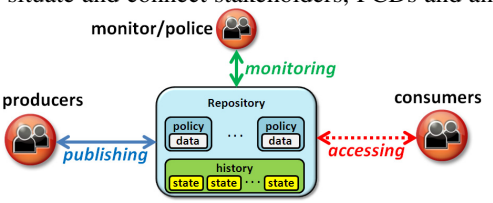

Fig. 1: A Framework for Policy-carrying Data

We note that the first two types of stakeholders can be institutions or people as well as computational entities/devices such as sensors, programs, databases, and so on. The monitor/police works as a third-party authority ensuring that activities (publishing and accessing) follow policies and dealing with violations. Each of these stakeholders has

[1] PCD also stands for "policy-carrying data collection" and we use PCDs (in the plural) to indicate a set of policy-carrying data collections.