Decision Provenance

Capturing data flow for accountable systems

Jatinder Singh, Jennifer Cobbe, Chris Norval
Department of Computer Science & Technology
University of Cambridge, UK
firstname.lastname@cl.cam.ac.uk

Abstract

Demand is growing for more accountability in the technological systems that increasingly occupy our world. However, the complexity of many of these systems—often systems of systems—poses accountability challenges. This is because the details and nature of the data flows that interconnect and drive systems, which often occur across technical and organisational boundaries, tend to be opaque. This paper argues that data provenance methods show much promise as a technical means for increasing the transparency of these interconnected systems. Given concerns with the ever-increasing levels of automated and algorithmic decision-making, we make the case for *decision provenance*. This involves exposing the 'decision pipeline' by tracking the chain of inputs to, and flow-on effects from, the decisions and actions taken within these systems. This paper proposes decision provenance as a means to assist in raising levels of accountability, discusses relevant legal conceptions, and indicates some practical considerations for moving forward.

1 Introduction

Technology is increasingly the subject of public discussion and regulatory attention. In line with this discourse is a demand for increased accountability for the technologies that now affect many aspects of contemporary life. This demand will likely grow as technology increasingly pervades society, particularly as visions such as of smart-cities and of the Internet of Things (IoT) come to be realised. Transparency is key for holding those responsible to account, as it enables identification, audit, and oversight. Indeed, current discussion in the public sphere often focuses on the transparency of major tech platforms, such as Facebook or Google [1].

From a technical perspective, a dominant research focus is on 'algorithmic accountability' [2], where much discussion concerns issues of fairness, transparency, and explainability in the use of machine learning (ML).[1] Generally less considered is the technology's broader operational contexts [3], which often comprise several different technical components being managed by different entities (organisations) which come together to produce a particular function. In practice, this represents an interconnected *system-of-systems*, of which data is a driver. The complexity of these systems-of-systems is set to increase as more and more advanced technologies are developed, deployed, interconnected, and automated, by a range of different entities. The visions of the future smart city are a case in point (§3).

The complexity of such interconnected environments poses significant challenges for accountability. The data flows that drive these interconnected systems are often opaque, making it difficult to exercise oversight and to determine where something went wrong and who is responsible, or in some cases, even to identify the entities involved. At the same time, these concerns are compounded by increasing levels of automation – including the use of ML – where certain happenings can result in decisions and actions with potentially immediate and far-reaching effects.

Technical measures can assist in making systems more transparent, improving the accountability of the organisations and individuals responsible for them. Data provenance methods show real promise, by providing a means to capture information that makes visible the flow of data between, through, and across these interconnected systems. Given the concerns around automation and algorithmic decision-making, we argue that it is important to *expose the decision pipeline*, in addition to ongoing work focusing on the decision-making aspects of these systems (i.e. accountability as it regards ML).

As such, we make the case for **decision provenance** as a means for exposing the data flows and interconnections leading up to a decision or action and that decision or action's cascading consequences. We explore how transparency over the inputs, entities involved, and flow on effects of any decision or action makes it easier to identify the entity responsible for a given component within a system – and how this provides benefits in terms of increasing user agency, facilitating compliance, and assisting in regulatory oversight. We also discuss some of the practical considerations for implementation and some areas for work which are necessary for moving forward. In all, our aim is to highlight the potential of exposing the decision pipeline as a key technical measure for realising more transparent, accountable, and compliant systems.

[1] For an indicative reading list, see: https://www.fatml.org/resources/relevant-scholarship .

2 Accountability

Accountability involves apportioning responsibility for a particular occurrence and determining from whom any explanation for that occurrence is owed. Generally speaking, the entities which are held accountable are natural and legal persons (i.e. people and organisations), whether for their own actions or for those of people, organisations, or machines which they have designed, which are under their control, or for which they are otherwise answerable. Therefore in a systems context, facilitating accountability involves making it easier to determine which person or organisation is responsible for a particular decision/action, its effects, and from (and to) whom an explanation is owed for that happening. This may or may not involve exploring the inner workings of particular technologies, as is the focus of much of the technical research community. While there is debate about the degree to which exposing the details of code and models actually helps accountability [4], our focus of discussion here is on making transparent the connections driving systems so to show the context in which they are operating, their effects, and to indicate the entities involved.

2.1 Legal impetus

From a legal point of view, accountability is closely related to *transparency*, and transparency is often a regulatory requirement for facilitating accountability. In this context, accountability may involve determining liability for an (automated) decision/action and, where harm arises as a result of that, what restitution is owed by who and to whom for that harm [5]. Importantly, and as is much discussed, data protection and privacy law also requires a focus on accountability.

The European Union's General Data Protection Regulation (GDPR),[2] for example, obliges data controllers (i.e. the entities responsible for determining the means and purposes of processing personal data) to be able to demonstrate compliance with its various requirements, including the data processing principles, if personal data is being processed. GDPR affords several rights to data subjects (i.e. those whose personal data is being processed) which controllers are tasked with meeting, including the right to erasure (the 'right to be forgotten') and the right to object to further processing of personal data, among others. It follows that transparency over data – for example, data inventories that record where data came from, the subjects that it refers to, and where it goes to – will be important for fulfilling data controllers' obligations in relation to these rights. Further, GDPR emphasises the accountability aspects of automated decision-making, particularly that which produces legal or similarly significant effects, with the so-called 'right to an explanation' seeking to provide data subjects with transparency rights regarding decisions.[3]

GDPR also gives regulators the power to conduct data protection audits, and will require data controllers to establish binding corporate rules for auditing. And, under GDPR, contracts between data controllers and processors must include provisions relating to the auditing of processors by controllers. Auditing, which necessarily requires transparency, will therefore become a key aspect of data protection regulation and compliance. Moreover, there are strong incentives for compliance, as GDPR non-compliance risks serious financial penalties – regulators are empowered to impose fines of up to the greater of €20m or 4% of annual global turnover. Further, the EU's proposed ePrivacy Regulation[4], in its current form, looks to extend data management obligations to non-personal data, e.g. including around information relating to user devices, which becomes particularly relevant as the IoT proliferates.

It follows that the ability to fulfil data subject's rights, to demonstrate compliance, to allow for auditing, and to meet other legal and regulatory obligations (beyond data protection) will be a consideration for anyone designing, implementing and using technical components and systems. And, in an interconnected systems context, being able to identify which organisation is responsible for a given system component will be of significant importance in determining liability should harms arise.

2.2 Technical requirements

In light of the above, there is a clear role for technical mechanisms to aid in increasing levels of transparency. Indeed, it is generally said that accountability at a technical level is grounded in transparency and control [9]. But transparency helps with with more than just fulfilling legal rights. As well as allowing those responsible for systems to be held to the standards required by law, transparency also helps systems themselves to be evaluated for correctness, bias, fairness, and for whether they are operating within the parameters that are expected or desired by their designers, operators, or end-users. Transparency thus also facilitates control through intervention where appropriate so that where problems or potential issues have been identified within or resulting from a system, the designer or operator of the system can intervene (automatically or manually) to address them. However, achieving transparency over systems in a highly interconnected context can be difficult, as we will explore next.

[2] Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) [2016] OJ L119/1, which comes into force 25 May 2018.

[3] There is on-going debate about the extent and utility of this right [6–8].

[4] Proposal for a Regulation of the European Parliament and of the Council concerning the respect for private life and the protection of personal data in electronic communications and repealing Directive 2002/58/EC (Regulation on privacy and electronic communications)

3 The challenge of interconnectedness

Accountability is often discussed in relation to discrete systems, where the entities involved are known or predefined. However, the increasingly interconnected nature of systems means that in many cases they do not operate discretely and in isolation but are employed as part of a *system-of-systems* [10], with potentially many entities involved [11, 12]. One system, for example, may take inputs from a range of sensors and may then produce an automated decision in the form of a data output that itself forms the input to another system, which, in turn, may produce an output through a device that creates a physical interaction with a real-world object or a person. In other words, the individual component systems interact and combine to bring functionality through a system-of-systems, where components may be potentially managed by multiple different entities.

The general direction towards more connectivity can be seen, for example, in relation to smart cities [13–15]. As Fig. 1 shows, there are already a number of smart systems which are being deployed to manage cities. And, as the vision of the smart city is increasingly realised, more and more complex systems will be deployed and will interact with each other. At a more granular level, we already see systems which are composed of interconnected systems – cloud *([something]-as-a-service)* a case in point.

Smart living	Sonitus sound sensing	Network of sound sensors monitoring noise levels
	Monitored sheltered housing	Remote monitoring of movement sensors and panic buttons in sheltered homes
	Smart Stadium	Sensor network monitoring different facets of stadium use
Smart mobility	Smart parking	Transponder payment system; park-by-text; display around city; API feed
	Information display signs	Traffic (crash/delay) alerts; airport queue counters
	Bliptracker displays	Bike counters; car parking spaces counters; airport queue counters
Smart environment	Sensor flood monitoring	Use of sensor network to monitor river levels by the Environmental Protection Agency (EPA) and local authorities
	Public building energy use	Real-time monitoring of energy use with local authority buildings; publicly displayed on screens
	Big Belly Bins	Networked compactor bins that use sensors to monitor waste levels; waste collection route optimisation

Figure 1. Some smart city technologies implemented in Dublin as of 2016 [14]. The broader visions for smart cities is for systems like these, as well as new systems, to interact with each other as well as with various commercial and consumer-oriented services.

Importantly, in these interconnected systems **data flow drives everything**, in that the flow of data between, through, and across systems brings their functionality [16]. A given data flow might encapsulate, for example, a sensor reading, the inputs to or outputs (decision or actions) from a machine learned model, an actuation command, a database query or result, and more. However, in most cases visibility over the flow of data is lost once it moves beyond a particular boundary, whether that is technical (e.g. between software components and services) or administrative (e.g. between jurisdictions).

The complexity and general opacity of interconnected systems therefore poses accountability issues, as it becomes difficult to discern the technical components involved, and then who is responsible for these. Moreover, decisions and actions somewhere within an assemblage might propagate widely, making it difficult to trace the source (organisational or technical) when concerns arise, and all the consequential (flow-on) effects. As an example, a reading from a faulty sensor could lead to cascading knock-on effects causing significant disruption, though its relationship to this may not be readily evident given the gaps in 'time and space'.

4 Decision provenance: exposing the decision pipeline

From the above, we can see that greater visibility over the assemblage of systems is important for increasing accountability for them. Given that data drives systems, there appears to be real potential in applying data provenance methods as a means of working towards this. Put simply, *data provenance* is metadata about data which records the lineage of data, including where it came from and where it moves to [17]. A common application is in a research context, where it can be used to assist in research reproducibility [18], though its potential for assisting information compliance has been considered [16, 19–22].

Given the concerns around accountability for automated and algorithmic decision-making and the increasingly interconnected nature of system deployments, we believe that there is a real need for **decision provenance**.[5] This involves using provenance techniques to capture the chain of data flows and associated interconnections leading up to a decision or action, as well as its flow on effects. The purpose of decision provenance is to expose the *decision pipeline* in order to make visible the inputs to and cascading consequences of any decision or action (e.g. Fig. 2), as well as the entities involved.

This involves tracking the flow of data between, through, and across component systems in a conceptually similar way to the tracking of physical items through product supply chains, which helps manufacturers gain a better understanding of the provenance of their products and of the materials therein [23]. This allows factories and warehouses to be accountable to manufacturers, which in turn allows manufacturers to be accountable to regulators and consumers.

In a similar way, information about the flow of data between, through, and across (interconnected) systems, in relation to both *ex post* (i.e. that following a decision, e.g. consequences) and *ex ante* processes (i.e. that proceeding a decision, e.g. inputs and design), helps facilitate evaluation and control. In this way decision provenance provides accountability opportunities, both ex post and ex ante, as follows.

[5] Distinct from the decision provenance (DECProv) ontology for modelling decision-making processes: https://promsns.org/def/decprov/decprov.html.

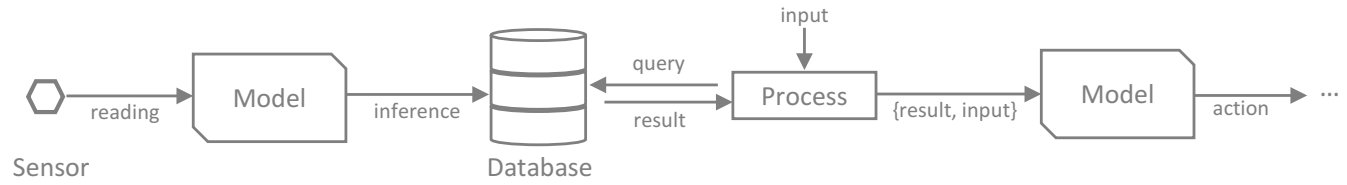

Figure 2. An example decision pipeline. Sensor data drives the inferences of a predictive model which are recorded in a datastore. This is then queried by a process which feeds the result, alongside another input, into another model that affects an action. One can imagine similar arrangements involving, for instance, a smart home management system, which automatically controls the environment, being influenced by systems that forecast weather, wearables inferring a resident's mood and well-being, and so forth.

4.1 Ex post opportunities

As we have outlined, the opacity of complex systems arrangements poses significant accountability challenges. Provenance that records the decision pipeline, both in terms of the technical components and the entities involved, naturally helps support compliance with data protection obligations and other ex post accountability measures such as audit and investigation.

4.1.1 Legal compliance

Records of information flow can assist those responsible for systems in complying with their legal obligations [20]. As discussed in §2, many of the rights afforded to individuals in data protection law could be facilitated through mechanisms for tracing the paths of data so as to identify where data resides, with whom it was shared, and how it was processed [22]. Data inventories can be of use in meeting obligations such as the right to erasure, subject access requests, and others.

4.1.2 Evidence, audit and investigation

Data provenance can assist audit and investigation by providing evidence of who is or might be responsible for a system or a decision, thereby assisting in identifying which entities to hold account. Information could also be obtained from provenance data that absolves responsibility by providing evidence demonstrating that the right decisions were made and actions were taken.

Provenance information can also help facilitate technical audit and investigations in complex systems by helping to identify the points of failure or further investigation. For example, provenance regarding data flow would make it easier to determine whether it was a particular data source such as a sensor producing erroneous readings that led to decisions with poor outcomes being made.

4.2 Ex ante opportunities

The potential for provenance methods to allow one to 'see' across the systems and its interconnections can also enable proactive steps to be taken in order to address problems before they arise. This works towards increasing the overall quality, security, and reliability of complex, interconnected systems and services. Generally speaking, more information regarding the nature of systems could allow those building and deploying systems to improve their quality, while helping users make better-informed decisions about which systems they engage with can help incentivise those producing and operating technology to improve their practices at all steps.

4.2.1 Proactive compliance

Provenance data can be leveraged to help enable more proactive approaches to legal and regulatory compliance. Generally, knowledge of the nature of the data flowing into and within a system, in terms of where it has come from and how it has and is being used, can help systems designers and operators (manually or through automatic means) monitor, maintain and improve the quality of their systems. There is, for example, research on using provenance information for detecting system faults [24]. In a similar vein, and in line with reactive, event-based mechanisms that automatically take actions to assist obligation management [16], provenance information could be used trigger particular compliance operations; such as to automatically report data breaches to the relevant authorities, to screen and filter out data based on compliance criteria (e.g. data past an expiry date), or to not act on inputs that may have come from, or through, an unreliable entity [20].

4.2.2 Increased user agency

There are also potential benefits for individuals in terms of increased agency. Decision provenance information could help users make better-informed decisions about the services that they use. If a user could see in advance that giving data to a particular system could, for example, (a) result in certain decisions being made that have particular undesired consequences; or (b) flow to an undesired entity (such as an advertisement network), then this helps that user to make a more an informed choice about whether to use (send their data) that system. This is important given the ever-increasing prominence of data-gathering systems, and might facilitate a more informed and effective exercise of data subject rights.

5 A focus on machine learning

A key driver for discussions of algorithmic accountability, automated decision-making, and indeed, decision provenance, is the increasing prevalence of machine learning (ML). We now briefly discuss ML in a decision provenance context.

ML works to uncover patterns in data so as to build and refine representative mathematical models of that data which can be used to make predictions, decisions, and gain knowledge and insight [25]. These models can be used to solve problems that would otherwise be challenging to program specific rules for, such as object classification in images, by instead having the model derive these through trends in the training data. In a systems context, these models are applied to new ('live') input data, with the outputs being a prediction, a decision, an action, an inference, and so on.

ML raises interesting considerations in a systems context. First, ML is part of the 'big data' trend, where vast quantities of data which can originate from a wide variety of sources is relevant for both the building and use of ML models. Second, we are increasingly seeing ML models being offered as a service,[6] allowing users to pass input data to these services and receive their predicted outputs. In practice this means that ML models can integrated together with other software processes and indeed other models to form complex, interconnected chains of systems (see Fig. 2).

As is the case with interconnected systems in general, data flow is the enabler. This includes, for example, the data flows relating to the training of models (learning), the use of models that make decisions (i.e. applying models to data), and the flow on effects of those decisions throughout the system. As such, we argue that provenance, at every stage in the process, is becoming ever more important for assisting accountability in these potentially complex systems arrangements.

5.1 Data for learning

Machine learning models are built (trained) on data, and therefore reflect the nature of that data. Training models on data of a poor quality (e.g. unrepresentative, biased, erroneous, etc.) can lead to issues being encoded within these models and reflected in the model's application. Indeed, models can encode issues of bias, discrimination, and unfairness [26–29]. In an interconnected context, a training dataset may comprise data from a range of sources, including sensors, human input, system logs, data brokers, and so on, and can also include outputs from calculations, analytics, or other ML models (§3).

[6] See, for example, services currently offered by Google: https://cloud.google.com/products/machine-learning, Microsoft: https://azure.microsoft.com/en-gb/services/cognitive-services, and Amazon: https://aws.amazon.com/machine-learning.

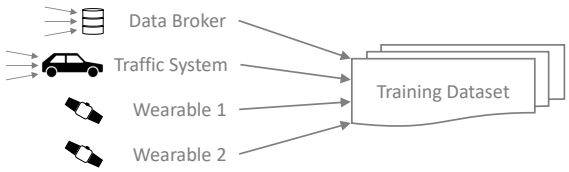

Figure 3. Training data comprised from multiple sources.

Provenance has a clear role to play in showing the nature of the training datasets, in that revealing the pipeline can allow for greater knowledge of the sources of data, assisting in assessing the veracity of the training data, as well as the potential for errors, inherent bias or unfairness present in the model. This use is in line with common applications of data provenance for research reproducibility.

5.2 Decisions and their effects

Machine learning models are applied to 'live' data which is inputted to a trained model in order to produce a decision (some action). So capturing both the model inputs and the decision itself (i.e. the model outputs), and the provenance of these, will be of particular importance when it comes to implementing technical solutions to facilitate accountability across systems. Doing so helps reveal the decision pipeline; i.e. the data flows leading up to a decision being made as well as its output and subsequent effects. Indeed, the outputs may result in physical-world actions (i.e. an actuation command) or outputs that feed into other datasets or trigger other models (see Fig. 2). And, as discussed, the effect of a decision on the wider system-of-systems may cross administrative boundaries, affecting a range of entities.

5.3 Using the decision pipeline: points for investigation

By its very nature, decision provenance helps investigate complex and interconnected systems by exposing the data flow and interconnections leading up to a decision or action. Once the decision pipeline has been used to identify the potential source(s) of an issue (e.g. a specific ML model in the chain of systems), auditors can then narrow their focus and go into detail. In an ML context, this may involve going on to inspect the workflows and processes around the construction of that model [3], e.g. by using techniques discussed in the algorithmic accountability literature (e.g. [2]). Information about the workflow of model construction and deployment itself facilitates auditing of, for example, the learning algorithm (or algorithms in ensemble models) used, hyperparameters, predictor variables, and coefficients (variable weights) [30]. As such, decision provenance complements the ML-oriented work on 'algorithmic accountability' by providing auditors information of the wider, practical context in which automated decision making processes were

defined and operate, as a means for indicating where deeper model-centric investigations can then take place.

6 Implementation considerations and challenges

Decision provenance poses technical implementation challenges. These predominately stem from the scale, federation, and complexity of what is effectively a wide-scale distributed system that can encompass a range of technologies and also a number of organisations with different and possibly competing incentives. For provenance to enable accountability, crucial is the consistency, visibility, and veracity of provenance information across this federated environment.[7]

Towards this, we now indicate some areas for consideration, focusing on the methods for gathering and representing provenance information, as that provides the foundation for the accountability opportunities earlier discussed (§4).

6.1 The mechanism: capture and management

A key consideration are the capture mechanisms, i.e. the technical means for producing provenance information.

There are broadly two categories of approach [17]: *disclosed* and *observed* provenance. Disclosed provenance tends to be application-oriented, where the details of what to record, and when, are written into application(s) code (often in APIs, as points of data exchange). This allows customisation by the designers as to what is deemed important, but will only capture what is explicitly programmed for.

Observed provenance captures information by observing what occurs; for example, by embedding the capture mechanism into the platform/infrastructure (e.g. an operating system), to capture the data flows regarding the applications running within [31–33]. By operating externally from applications, this allows the interactions between various applications (system components) to be recorded irrespective of the application specification and design – without developer intervention or even knowledge. Further, potentially capturing every data flow in the operating environment reduces the propensity for 'missing something'. However, a general capture mechanism is comparatively less-targeted, and brings overhead considerations particularly regarding the volume and complexity of the recorded information [34].

As capture mechanisms tend to focus on particular layers of the technical stack [21],[8] in practice, it is likely that decision provenance may require a combination of mechanisms to ensure that the appropriate information is recorded.

6.1.1 The challenge of systems

An interconnected system-of-systems—with its many moving parts—poses practical implementation challenges, given the requirement for the provenance regime to operate consistently throughout.

Capture Crucial is that capture mechanisms work throughout the entire decision pipeline, operating at the appropriate points within and across technical and organisational boundaries. Application-oriented capture regimes might suffice where the pipelines/supply chains are pre-known, well-defined and fairly static. However, such approaches are limited in that they apply only to pre-defined applications and organisations. In contrast, observed approaches suit capturing detailed information across applications, but only within the scope of specific operating environments (e.g. a platform/operating system instance). There is on-going work towards the challenges in maintaining consistent capture regime across boundaries, administrative/organisational and operational, e.g. as data flows from one operating system (machine) to another [16, 21, 22, 31, 33].

In practice, it is likely that exposing the decision pipeline will require several complementary capture regimes. This accords with the conceptual provenance stack, which defines a series of layers for focusing specific provenance considerations to enable a more complete regime [22, 35]. Standards, guarantees and best practices regarding capture mechanisms will also be required, for enabling ex post accountability, and as a foundation for enabling proactive (ex ante) approaches (§4.2), including technical policy enforcement regimes [16].

Data Access A related issue is how the captured provenance data is recorded and made accessible across boundaries (technical or organisational). If provenance information is federated, questions arise as to how it can be reconciled across systems and organisations in order to make the decision pipeline visible, and how this provenance data is to be aggregated (if at all), queried, and so on. Moreover, access to the provenance data itself raises accountability questions, given that, as metadata about data, it may itself be sensitive [22], particularly in a cross-organisational context. Secure management of this data is one area of consideration [36], including determining the contexts in which access controls might be restricted or opened to others, such as regulators, in certain circumstances.

Trust For provenance information to be useful it must be reliable. However, this is a particular challenge here for two inter-related reasons. First, the risks and incentives in an accountability context are complex, given that the provenance information relates to responsibilities from which onerous consequences might result.[9] Second is the inherent federation in terms of the mechanisms for capture and what is

[7] That said, there are real and immediate benefits in employing provenance methods *now*, in more local scenarios, as even internal organisational usage can help in managing systems, processes, and obligations, while assisting compliance and providing evidence demonstrating good practice [20].

[8] For instance, the nature of that recorded at lower-levels of a network infrastructure will be different to that captured by applications.

[9] There are, however, incentives to implement such provenance mechanisms, for instance, to produce evidence that aid arguments for absolving responsibility (*"I've done nothing wrong"*).

recorded. These aspects in combination raise issues of trust in multi-party scenarios, where data flows across administrative or organisational (i.e. responsibility) domains.

In all, there is a clear requirement for means of ensuring the integrity of the capture regime, recorded data, query mechanisms and any compliance mechanisms (such as policy enforcement regimes) that are built on top, and to ensure this integrity is maintained throughout any pipeline [16, 22, 37]. Towards this, standardisation, verification, attestation, and secure logging mechanisms will all be relevant.

6.2 Representing captured provenance data

In a systems context, it is important that there is a common representation of provenance data so that it can be understood and interoperate system-wide. Standards are required to support the interpretability and understanding of audit records within and across systems. The W3C PROV[10] standard provides an extensible[11] mechanism for modelling provenance data, and ontologies (description vocabularies) provide the means for describing what has been captured. W3C PROV is capable of capturing aspects relevant for compliance and accountability, for example, for recording which parties undertook various activities with regard to data.

In moving forward, one consideration is whether specialised models or vocabularies will be necessary for accountability purposes, as well as to help align the provenance information that is captured at different technical layers (§6.1). Different extensions may also be needed to cater for the specifics of a particular application domain; what needs to be captured for an automated traffic management system will differ to that of an e-commerce website.

6.2.1 Usability

A wider consideration is how stakeholders from diverse backgrounds and with varying goals can interpret and use provenance data; though this is a general challenge for provenance [2, 18, 38–40], it is exacerbated in a large-scale systems context particularly where accountability is the aim. Some (e.g. end-users) may be interested simply in the entities involved, while others (e.g. regulators) may require more information. This means that the provenance data captured may be difficult to interpret for some – for instance, making sense of data representing different levels of the technical stack may be beyond the expertise of non-technical users – and, further, those seeking to interact with such data may not be familiar with the nature of the systems involved.

Again, standards will be important. Given that provenance information may be relevant for different audiences, there may be a need for various tools and extensions. Some may describe lower-level technical details for a highly specialist audience, whereas others may, e.g., simply list the entities involved in order to assist a lay person.

There is a clear role for human computer interaction (HCI) research to assist in ensuring that decision provenance approaches enable a representation that assists end-users regarding their accountability concerns. A number of techniques have been suggested as ways in which provenance data can be made more usable, including Natural Language Interfaces for Databases (NLIDBs) [41], data visualisation techniques (such as graphs and plots) [38, 39, 41], as well as using online games [42] and comics [43] as a means for describing captured provenance information to end users. Generally, more work is required to explore the presentation of such data for accountability purposes, which supports a range of stakeholder perspectives.

7 Conclusion

There are strong pressures for improving the levels of accountability for technology, driven by societal demands, increasingly stringent legal requirements, and for reasons of public acceptance and adoption. In line with this, there is much discussion of algorithmic accountability, with a particular focus on automated decision-making (ML). However, there are also accountability challenges in the context of the broader systems – particularly as systems are increasingly complex and interconnected. Because these data-driven assemblages tend to be opaque, there is a clear need for greater transparency. Means for enabling accountability, technical or otherwise, will increase in importance as visions of pervasive computing: smart cities, the Internet of Things, autonomous transport systems, etc., become a reality.

We have argued that decision provenance, as a means for exposing the decision pipeline, is an important piece of the accountability puzzle, with much potential for assisting with issues of responsibility, technical compliance, and improved user agency. Though work is required, there appears a clear opportunity for decision provenance (alongside other measures) to improve accountability as it relates to technology.

Acknowledgments

We acknowledge the financial support of the UK Engineering and Physical Sciences Research Council and Microsoft, through the Microsoft Cloud Computing Research Centre.

References

[1] F. Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information.* Harvard University Press, 2015.

[2] N. Diakopoulos, "Accountability in algorithmic decision making," *CACM*, vol. 59, no. 2, pp. 56–62, 2016.

[3] J. Singh, I. Walden, J. Crowcroft, and J. Bacon, "Responsibility & machine learning: Part of a process," 2016, available at: https://ssrn.com/abstract=2860048.

[4] J. A. Kroll, J. Huey, S. Barocas, E. W. Felten, J. R. Reidenberg, D. G. Robinson, and H. Yu, "Accountable algorithms," *University of Pennsylvania Law Review*, vol. 165, no. 3, pp. 633–705, 2017.

[10] https://www.w3.org/TR/prov-overview/

[11] For instance, see the PROVOne extensions for scientific workflows: https://purl.dataone.org/provone-v1-dev.

[5] C. Reed, E. Kennedy, and S. Silva, "Responsibility, autonomy and accountability: Legal liability for machine learning," 2016, available at: https://ssrn.com/abstract=2853462.

[6] B. Goodman and S. Flaxman, "European union regulations on algorithmic decision-making and a "right to an explanation"," in *Workshop on Human Interpretability in Machine Learning at ICML*, 2016.

[7] S. Wachter, B. Mittelstadt, and L. Floridi, "Why a right to explanation of automated decision-making does not exist in the General Data Protection Regulation," *International Data Privacy Law*, vol. 7, no. 2, pp. 76–99, 2017.

[8] A. D. Selbst and J. Powles, "Meaningful information and the right to explanation," *International Data Privacy Law*, vol. 7, no. 14, pp. 223–242, 2017.

[9] D. J. Weitzner, H. Abelson, T. Berners-Lee, J. Feigenbaum, J. Hendler, and G. J. Sussman, "Information accountability," *CACM*, vol. 51, no. 6, pp. 82–87, 2008.

[10] M. W. Maier, "Architecting principles for systems-of-systems," *Systems engineering*, vol. 1, no. 4, pp. 267–284, 1998.

[11] J. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, "Internet of Things (IoT): A vision, architectural elements, and future directions," *Future Generation Computer Systems*, vol. 29, no. 7, pp. 1645 – 1660, 2013.

[12] J. Singh, T. Pasquier, J. Bacon, H. Ko, and D. Eyers, "Twenty security considerations for cloud-supported Internet of Things," *IEEE IoT*, vol. 3, no. 3, pp. 269–284, 2016.

[13] A. van Dijk, "Smart cities how rapid advances in technology are reshaping our economy and society," 2015, available at: https://www2.deloitte.com/content/dam/Deloitte/tr/Documents/public-sector/deloitte-nl-ps-smart-cities-report.pdf.

[14] R. Kitchin, *Getting smarter about smart cities: Improving data privacy and data security*. Data Protection Unit, Department of the Taoiseach, 2016, available at: http://eprints.maynoothuniversity.ie/7242/.

[15] Y. Mehmood, F. Ahmad, I. Yaqoob, A. Adnane, M. Imran, and S. Guizani, "Internet-of-things-based smart cities: Recent advances and challenges," *IEEE Communications Magazine*, vol. 55, no. 9, pp. 16–24, 2017.

[16] J. Singh, T. Pasquier, J. Bacon, J. Powles, R. Diaconu, and D. Eyers, "Big ideas paper: Policy-driven middleware for a legally-compliant Internet of Things," in *ACM Middleware*, 2016.

[17] L. Carata, S. Akoush, N. Balakrishnan, T. Bytheway, R. Sohan, M. Seltzer, and A. Hopper, "A primer on provenance," *CACM*, vol. 57, no. 5, pp. 52–60, 2014.

[18] S. B. Davidson and J. Freire, "Provenance and scientific workflows: Challenges and opportunities," in *ACM SIGMOD*, 2008.

[19] R. Aldeco-Pérez and L. Moreau, "A provenance-based compliance framework," in *FIS*, 2010.

[20] J. Singh, J. Powles, T. Pasquier, and J. Bacon, "Data flow management and compliance in cloud computing," *IEEE Cloud Computing*, vol. 2, no. 4, pp. 24–32, July 2015.

[21] Y. S. Tan, R. K. L. Ko, and G. Holmes, "Security and data accountability in distributed systems: A provenance survey," in *IEEE EUC*, 2013.

[22] T. Pasquier, J. Singh, J. Powles, D. Eyers, M. Seltzer, and J. Bacon, "Data provenance to audit compliance with privacy policy in the Internet of Things," *Personal and Ubiquitous Computing*, vol. 22, no. 2, pp. 333–344, 2017.

[23] S. New, "The transparent supply chain," *Harvard Business Review*, vol. 10, no. 88, pp. 76–82, 2010.

[24] X. Han, T. Pasquier, T. Ranjan, M. Goldstein, and M. Seltzer, "FRAPpuccino: Fault-detection through runtime analysis of provenance," in *USENIX HotCloud*, 2017.

[25] G. James, D. Witten, T. Hastie, and R. Tibshirani, *An introduction to statistical learning*. Springer, 2013.

[26] S. Hajian, F. Bonchi, and C. Castillo, "Algorithmic bias: From discrimination discovery to fairness-aware data mining," in *ACM KDD*, 2016.

[27] F. Kamiran and T. Calders, "Data preprocessing techniques for classification without discrimination," *KAIS*, vol. 33, no. 1, pp. 1–33, 2012.

[28] K. Kirkpatrick, "Battling algorithmic bias: How do we ensure algorithms treat us fairly?" *CACM*, vol. 59, no. 10, pp. 16–17, 2016.

[29] D. Pedreshi, S. Ruggieri, and F. Turini, "Discrimination-aware data mining," in *ACM KDD*, 2008.

[30] S. Schelter, J.-H. Böse, J. Kirschnick, T. Klein, and S. Seufert, "Automatically tracking metadata and provenance of machine learning experiments," in *Machine Learning Systems Workshop at NIPS*, 2017.

[31] T. Pasquier, X. Han, M. Goldstein, T. Moyer, D. Eyers, M. Seltzer, and J. Bacon, "Practical whole-system provenance capture," in *ACM SoCC*, 2017.

[32] T. Pasquier, J. Singh, D. Eyers, and J. Bacon, "Camflow: Managed data-sharing for cloud services," in *IEEE TCC*, 2015.

[33] A. Gehani and D. Tariq, "SPADE: Support for provenance auditing in distributed environments," in *ACM Middleware*, 2012.

[34] T. Pasquier, J. Singh, J. Bacon, and D. Eyers, "Information flow audit for PaaS clouds," in *IEEE IC2E*, 2016.

[35] R. K. L. Ko and T. W. Phua, "The full provenance stack: Five layers for complete and meaningful provenance," in *SpaCCS*, 2017.

[36] S. Xu, Q. Ni, E. Bertino, and R. Sandhu, "A characterization of the problem of secure provenance management," in *ISI*, 2009.

[37] A. Bates, D. J. Tian, K. R. Butler, and T. Moyer, "Trustworthy whole-system provenance for the linux kernel," in *USENIX Security*, 2015.

[38] P. Chen, B. Plale, Y. W. Cheah, D. Ghoshal, S. Jensen, and Y. Luo, "Visualization of network data provenance," in *IEEE HiPC*, 2012.

[39] W. Oliveira, L. M. Ambrósio, R. Braga, V. Ströele, J. M. David, and F. Campos, "A framework for provenance analysis and visualization," in *ICCS*, 2017.

[40] Q. Wang, W. U. Hassan, A. Bates, and C. Gunter, "Fear and logging in the internet of things," in *ISOC NDSS*, 2018.

[41] F. Li and H. Jagadish, "Usability, databases, and HCI," *IEEE Data Eng. Bull.*, vol. 35, no. 3, pp. 37–45, 2012.

[42] K. Bachour, R. Wetzel, M. Flintham, T. D. Huynh, T. Rodden, and L. Moreau, "Provenance for the people: An HCI perspective on the W3C PROV standard through an online game," in *ACM CHI*, 2015.

[43] A. Schreiber and R. Struminski, "Visualizing provenance using comics," in *USENIX TAPP*, 2017.