A Digital Data Exchange for Australia

JOHN RUCIAK, Flinders University
THOMAS HARDJONO, MIT Connection Science & Enginering

September 16, 2021

EXECUTIVE SUMMARY

It is time for Australia to contemplate the development for a digital *marketplace* data & insights that permit relevant information to be obtained by relevant parties in a timely manner, all the time preserving the privacy of individual data-subjects. The *Data Exchange* is a technological representation and implementation of the data & insights marketplace. The *data traders* entities supply and purchase data & insights from the marketplace platform. The data trader entities can be organizations, collectives, data cooperatives or individuals. The data trader parties obtain authorized access to *data objects* in a protected manner, where data objects include: raw data, references to data, collated data and reports, metadata and insights.

The following general features are required for a successful Australian data exchange:

- Privacy and anonymity of the data subject is confirmed. The market place platform observes and enforces the *privacy first* principles.
- Access history to data objects are logged and made immutable through recording onto a public blockchain.
- The identity of any entity accessing data objects is confirmed. The default individuals identity verification mechanism must be based on their Australian +GOVAU account, which maintains the authoritative one-to-one correlation between their online digital identity and their real-world legal identity.
- Organizations and their proxies are authenticated by existing federal government authorization systems.
- The access logic to the data-objects are implemented using safe and composable smart contracts. Remunerations for data access and trades can be on-chain tokens, which may have zero economic value. The data exchange platform can be operationally self-sustaining in the long term by charging a small percentage from the value-carrying (non-zero) tokens.

The proposed Australian data exchange platform consists of three logical tiers:

(1) A top tier, where insights are shared.
(2) A middle tier, where access to data objects takes place.
(3) A foundational tier, where data is managed in a decentralized and protected manner.

1 INTRODUCTION

Modern society is current facing a dilemma with regards to data-driven decision making for individuals, organizations and communities. On one hand, individuals, organizations and communities need access to data in order to perform computations as part of decision-making. The promise is that better insights can be obtained by combining data from different domains in interesting and innovative ways. On the other hand, however, there is considerable risk to individual privacy and to commercial IP when data is shared across entities. The 2011 World Economic Forum (WEF) report [1] clearly points to inadequate care given today to personal data, with evidence abound with regards to theft or misuse of personal data reported in the media.

A new data exchange paradigm is needed to address the needs for data and insight in the data-driven society and economy. Many entities in the data-driven economy need insights in an efficient

---

Authors' addresses: John Ruciak, Flinders University, Adelaide, South Australia, john.ruciak@flinders.edu.au; Thomas Hardjono, MIT Connection Science & Enginering, Cambridge, MA, USA, hardjono@mit.edu.

and timely manner. Often these entities either do not have direct access to the data or do not possess the capacity (e.g., algorithms, compute power, know how) to derive insights from data across various verticals. Thus, the data exchange must also be a marketplace for insights, one that brings together various disciplines and expertise in computation social science, algorithm design, cloud computing management and privacy-preserving computations.

Recent advances in cloud computing, machine learning and blockchain technology [2], along with a maturing digital identity space are providing the technical framework for the development of digital data exchanges. The challenge for governments is to encourage interoperability in the digital economy with the use of open systems and networks to prevent these new exchanges becoming extensions of existing industry silos, or disconnected profit driven enterprises arising out of personal and machine data collected by telecommunications and social media organizations.

2 MOTIVATIONS: IMPROVING ACCESS TO DATA AND INSIGHTS

Today there are a number of open challenges with regards to the information sharing ecosystem:

- *Data is siloed*: Today data is siloed within organizational boundaries, and the sharing of raw data with parties outside the organization remains unattainable, either due to regulatory constraints or due to business risk exposures.
- *Privacy is inadequately addressed*: The 2011 WEF report on personal data as a new asset class finds that the current ecosystems that access and use personal data is fragmented and inefficient. For many participants, the risks and liabilities exceed the economic returns and personal privacy concerns are inadequately addressed. Current technologies and laws fall short of providing the legal and technical infrastructure needed to support a well-functioning digital economy. The rapid rate of technological change and commercialization in using personal data is undermining end-user confidence and trust.
- *Regulatory and compliance requirements*: The introduction of the EU General Data Protection Regulations (GDPR) [3, 4] will impact global organizations that rely on the Internet for trans-border flow of raw data. This includes cloud-based processing sites that are spread across the globe.
- *Lack of citizen involvement and incentive as stakeholder*: Individuals as community members are typically "out of the loop" beyond their blanket consent to access their personal data. The lack of citizen engagement for sharing data for the benefit of community is exemplified most recently by Covid pandemic and the difficulty faced by some local governments in deploying a coherent contact tracing system.
- *Lack of economic incentive*: In the absence of an efficient data marketplace, the expense of classifying, licensing and maintaining the provision of external access to data is difficult for organizations to justify. For individuals there are few opportunities available for economic participation beyond passively receiving services such as search engine access in exchange for waiving their rights over their own internet browsing history.

3 THE DIGITAL DATA EXCHANGE: TOWARDS A MARKET FOR DATA ACCESS

The goal of the Data Exchange is to provide improved access for various types of data and insights across various industry verticals, while ensuring citizen data privacy is implemented and supporting the evolving market dynamics in the supply/demand of access to data.

The Data Exchange permits better matching between data consumers, data providers and algorithms providers and their agents in the following ways:

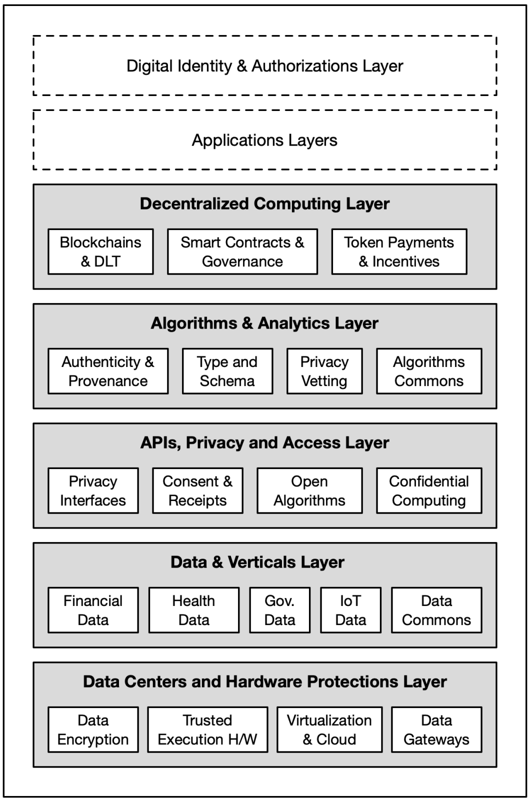

Fig. 1. Overview of layers of the data exchange

- *Data consumers*: Most organizations require information and insights that are relevant to their problem at hand and available in a timely manner. This true across different types of organizations, whether private (e.g., marketing company) or public sector (e.g., local state governments).
- *Data providers*: The holder of data have the challenge to ensure the provenance of their data is correct (provable), while at the same time permitting them to obtain revenue for access to insights derived from their data without exporting data from their current repositories. They also have the challenge of ensuring consent from data subjects [5, 6] and remunerating them for permitting access to insights derived from the subject's data.
- *Algorithm providers*: Many consumers may not have the time and skill to develop algorithms that are specific to the data at hand. In these cases, a data consumer organization may seek to outsource the design and craft of algorithms to a third party (e.g., boutique algorithm design firms). On their part, many algorithm providers may seek to retain the Intellectual Property (IP) stemming from their design and to offer these algorithms to other consumers in the market.
- *Agents*: Agents can represent the interests of data consumers, data providers or algorithm providers either individually or collectively. Agents can also provide second level services such as storytelling or report generation, and trade those outputs.

4 THE DATA EXCHANGE: DESIGN PRINCIPLES

Sound design principles should be the foundation of the Data Exchange and the services which stems from it. Some key design principles include the following:

- *Privacy first*: Preserving the privacy of individuals must be a foremost principle, else the data exchange will have little stakeholder adoption and will lead to the concentration of power in the hands of a few Privacy-preserving data processing approaches, such as MIT's Open Algorithms (OPAL) [7] ensure that data remain in their repositories and that only insights of computed from aggregate queries are delivered.
- *Data minimization*: Limit the movement, export and replication of data, and perform computations at the data endpoints by employing new distributed federated processing models.
- *Standardized access methods*: Define standard interfaces (i.e. APIs) to access insights based on computations of approved algorithms on data.
- *Technological independence*: Since technology is fast evolving around data management, artificial intelligence and privacy-preserving computation, the data exchange must be agnostic to specific technological implementations and be vendor-neutral.
- *Support for Not for Profit uses*: Revenues arising from the utilization of government and public data should be directed to support public goods (e.g. addressing the spread of diseases). One approach could be for some profits to be directed into making more open data available and for the funding data analysis initiatives that address societal issues.
- *Economically sustainable*: The data exchange should be self-funding, such as via a transactional levee, and should utilize server-less cloud and blockchain technology to be economically and technically scalable on demand, and affordable to replicate in less developed economies. Tokenized assets and the means of exchange should be directly pegged to the value of the local currency to prevent speculation.
- *Use existing open technology where appropriate*: The Data Exchange should utilize existing and emerging government sanctioned digital identity systems that are valid within a taxation context [8, 9]. In Australia this is the MyGovID system, which is evolving into the Digital Identity system [10]. Australia also provides the ability for agents to represent organizations as proxies under the RAM system [11]. There are several existing open data licensing schemes that could be adapted to include a market pricing structure.

Several general requirements should drive the technical design and implementation of the Data Exchange, including:

- *Accessibility to citizens*: The exchange should provide answers to questions to any citizen, not just API endpoints for specialized data scientists. This permits the citizen to glean the benefits of the Data Exchange as part of the broader digital infrastructure for the data-driven society.
- *Secure federation across the states*: State governments require the ability to federate their local data repositories in order to achieve national-level goals for citizens [12, 13]. Examples include the federation of local health data in order to address the spread of diseases and to address economic inequalities.

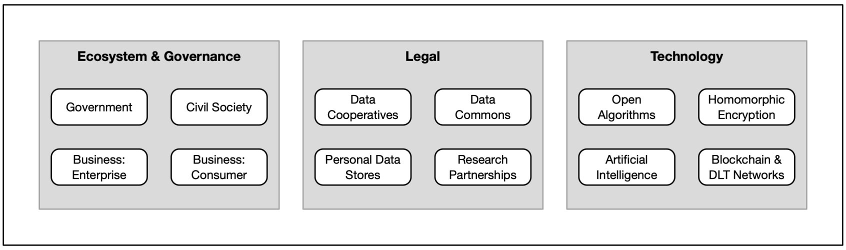

Fig. 2. Overview of governance components of the data exchange

5 THE ECOSYSTEM AND GOVERNANCE

As a data exchange platform with multiple stakeholders, service providers and users/consumers, the governance of the platform is a foundational component that must balance the requirements of citizen data privacy, economic self-sustainability of the platform, and observance of the various regulatory demands in Australia.

- *Multi-stakeholder ecosystem*: There are multiple entities that play a role within the ecosystem, and therefore a suitable governance architecture with appropriate policies must be applied.
- *Legal trust framework for data*: There are numerous data types about individuals (e.g., financial data, health data, location data, etc.) and various form of data sharing models (e.g., data commons, data cooperatives, individual data stores, etc.). A legal trust framework must underpin the data exchange to provide for usage-tracking and accountability to the data subjects (i.e., citizens and government).
- *Multiple technological tools*: There are a growing number of tools and technological solutions for data access and data sharing which enforce a privacy-first principle. This usage of these tools must be in accordance with the governance policies and legal trust framework underlying the data exchange. The tools must be selected as appropriate to the type of data and its source.

6 OPEN CHALLENGES AND RECOMMENDATIONS

There are several fundamental challenges relating to the design, deployment and management of a data exchange platform in Australia:

- *Governance model and implementation*: Governance is a complex issue, but one that needs to be explored through small proof-of-concepts. We recommend a simple policy whereby an API-based access to data is used, thereby preventing potential loss of data. The policies must identify entities permitted to access the protected APIs and the type of queries permitted to be asked. Stakeholders have visibility and transparency with regards to which APIs are most popularly accessed and which type of queries are most often asked.
- *Sustainable funding long-term*: A fundamental question for the data exchange platform is the source of funding to make it viable long term. We recommend exploring the use of tokens as a unit of payment for queriers and a unit of account for data providers to assist in tracking usage of data.

- *Exploration through pilot studies*: Various aspects of the data exchange must be the subject of exploration and study. This includes developing various proof of concepts (PoC) in collaboration with commercial service providers (e.g., data providers, cloud providers, etc.) as key stakeholders in the ecosystem. We recommend developing several small proofs of concept (PoC) software implementation using a small amount of safe data, simple APIs and using low- cost commercial services (e.g. cloud data storage, hosted API services, etc.). Technical participation by academics in Australia and by commercial service providers are crucial in these PoC projects.
- *Decentralization*: It may be that the general principles of the proposed data exchange are best accommodated by a model whereby computing power, data storage and associated exchange artifacts are as decentralized as possible. To accommodate decentralization within a nations physical, virtual and legal borders would require a Proof of Location mechanism [14].

ACKNOWLEDGMENTS

We thank the following for on-going discussions, ideas and support for this work: Prof Sandy Pentland, Stephen Buckley, Mohsen Bahrami, Dan Calacci (MIT); Gavin Artz, Mark Wadewitz, Ivor Hay, Stuart Davis (DTI); Nick Reade, Peter Worthington-Eyre, Akhilesh Harsh (DPC); Michael Luchich (Optus); Ben Owen, Mike Diakomichalis (BankSA); Paul Rybicki (DSpark).

REFERENCES

[1] WEF, “Personal Data: The Emergence of a New Asset Class,” World Economic Forum, Report, February 2011. [Online]. Available: http://www.weforum.org/reports/personal-data-emergence-new-asset-class

[2] H.-Y. Paik, X. Xu, H. M. N. D. Bandara, S. U. Lee, and S. K. Lo, “Analysis of Data Management in Blockchain-Based Systems: From Architecture to Governance,” *IEEE Access*, vol. 7, pp. 186 091–186 107, 2019.

[3] European Commission, “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 (General Data Protection Regulation),” *Official Journal of the European Union*, vol. L119, pp. 1–88, 2016.

[4] OAIC, “Australian entities and the EU General Data Protection Regulation (GDPR),” Australian Government Office of the Australian Information Commissioner, Tech. Rep., 2018, accessed 23 August 2021. [Online]. Available: https://www.oaic.gov.au/privacy/guidance-and-advice/australian-entities-and-the-eu-general-dataprotection-regulation/

[5] T. Hardjono, E. Maler, M. Machulak, and D. Catalano, “User-Managed Access (UMA) Profile of OAuth2.0 – Specification Version 1.0,” Kantara Initiative, Kantara Published Specification, April 2015, https://docs.kantarainitiative.org/uma/rec-uma-core.html.

[6] E. Maler, M. Machulak, and J. Richer, “User-Managed Access (UMA) 2.0,” Kantara Initiative, Kantara Published Specification, January 2017, https://docs.kantarainitiative.org/uma/ed/uma-core-2.0-10.html.

[7] T. Hardjono and A. Pentland, “MIT Open Algorithms,” in *Trusted Data - A New Framework for Identity and Data Sharing*, T. Hardjono, A. Pentland, and D. Shrier, Eds. MIT Press, 2019, pp. 83–107.

[8] J. Saviano, J. Badenach, and M. De Ruiter, “Tax Grid: What happens when government, industry and investors seek common digital ground,” Ernst & Young, Tech. Rep., July 2021. [Online]. Available: https://assets.ey.com/content/dam/ey-sites/ey-com/en_gl/topics/tax/tax-pdfs/ey-withholding-tax-distributed-ledger-report.pdf

[9] Deloitte, “Blockchain technology and its potential in taxes,” Deloitte, Tech. Rep., December 2017. [Online]. Available: https://www2.deloitte.com/content/dam/Deloitte/pl/Documents/Reports/pl_Blockchain-technology-and-its-potential-in-taxes-2017-EN.PDF

[10] Government of Australia, “Your Digital Identity helps you prove who you are when you want to access services online,” 2021. [Online]. Available: https://www.digitalidentity.gov.au/

[11] ——, “Accessing online services with myGovID and RAM,” 2021. [Online]. Available: https://www.ato.gov.au/General/Online-services/Accessing-online-services-with-myGovID-and-RAM/

[12] T. Hardjono, “Federated Authorization over Access to Personal Data for Decentralized Identity Management,” *IEEE Communications Standards Magazine – The Dawn of the Internet Identity Layer and the Role of Decentralized Identity*,

vol. 3, no. 4, pp. 32–38, December 2019. [Online]. Available: https://doi.org/10.1109/MCOMSTD.001.1900019

[13] ——, “Owner-Centric Access to IoT Data,” in *New Solutions for Cybersecurity*, H. Shrobe, D. Shrier, and A. Pentland, Eds. MIT Press, 2017, pp. 405–422.

[14] M. Amoretti, G. Brambilla, F. Medioli, and F. Zanichelli, “Blockchain-Based Proof of Location,” in *2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C)*, 2018, pp. 146–153.