

COMMENTARY

Fostering trustworthy data sharing: Establishing data foundations in practice

Sophie Stalla-Bourdillon¹, Laura Carmichael^{1,*} and Alexis Wintour² 

¹Law School, University of Southampton, Southampton, United Kingdom

²Lapin Ltd, Jersey, Jersey

*Corresponding author. E-mail: L.E.Carmichael@soton.ac.uk

Received: 17 November 2020; **Revised:** 00 ; **Accepted:** 25 November 2020

Key words: citizen representation; data foundations; data governance models; data institutions; guardian; independent data stewardship

Abstract

Independent data stewardship remains a core component of good data governance practice. Yet, there is a need for more robust independent data stewardship models that are able to oversee data-driven, multi-party data sharing, usage and re-usage, which can better incorporate citizen representation, especially in relation to personal data. We propose that data foundations—inspired by Channel Islands’ foundations laws—provide a workable model for good data governance not only in the Channel Islands, but also elsewhere. A key advantage of this model—in addition to leveraging existing legislation and building on established precedent—is the statutory role of the guardian that is a unique requirement in the Channel Islands, and when interpreted in a data governance model provides the independent data steward. The principal purpose for this paper, therefore, is to demonstrate why data foundations are well suited to the needs of data sharing initiatives. We further examine how data foundations could be established in practice—and provide key design principles that should be used to guide the design and development of any data foundation.

Policy Significance Statement

How to increase responsible data sharing is one of the most significant questions of the 21st century. Access to rich datasets benefits economies and societies by facilitating not only timely and well-informed policies and decisions, but also the development of new and improved data-based products and services. Data Foundations, as one of the new types of Data Institutions, offer some unique qualities which create the opportunity for a cost-effective and trustworthy mechanism for multiple data providers to share more data, even if they are competitors. They use specific design principles offering a data governance model which includes independent oversight, a rulebook, and the ability to represent data subjects in collective decision-making regarding the data.

Introduction

Data sharing initiatives involve two or more data providers who agree to pool or otherwise link together specified data in order to make such data available to each other and/or third parties (i.e., data users). Such data sharing initiatives can take many forms, including data collaboratives and data marketplaces (Smart

© The Author(s) 2021. Published by Cambridge University Press in association with Data for Policy. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike licence (<http://creativecommons.org/licenses/by-nc-sa/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the same Creative Commons licence is included and the original work is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use.

Dubai and Nesta, 2020). While the amount and demand for data continues to intensify, there is a considerable deficiency of such data sharing initiatives in practice that is “the data pooling problem” (Mattioli, 2017).

We approach this issue through the design and development of data foundations inspired by Channel Islands’ foundations laws.¹ Data foundations provide a robust data governance model that

“supports responsible and sustainable non-personal and personal data usage, sharing, and re-usage by means of independent data stewardship.” (Stalla-Bourdillon et al., 2019)

In this paper, we aim to demonstrate why data foundations are well suited to the needs of data sharing initiatives in not only the Channel Islands, but also elsewhere. We consider how data foundations may contribute toward practical advancement of independent data stewardship as part of a wider data institution movement. We then focus on opportunities for functional data foundations to be established via legal structures under English law, and introduce key design principles that should be used to guide the design and development of any data foundation.

Data Institutions

The Open Data Institute (ODI) uses the phrase “data institution” as an umbrella term to describe

“organisations whose purpose involves stewarding data on behalf of others, often towards public, educational or charitable aims.” (Dodds et al., 2020)

A core component for both data institutions and data foundations, therefore, is the concept of independent data stewardship. While this concept is not new, there is a need for more robust independent data stewardship models that are able to oversee data-driven, multi-party data sharing, usage and re-usage; especially those which increase “*citizen control*” over personal data (Mac Manus, 2020).² For instance, according to Verhulst (2018), one of the five key functions of the data steward is “*partnership and community engagement*.” The Royal Society and the British Academy (2017) have also called for a new stewardship body to oversee “*the entire data governance landscape*” with the ability to “*foster dialogue and deliberation*” between stakeholders, including the wider public.

Data trusts³ are one form of data institution (Hardinges and Tennison, 2020) that have received considerable attention as conceptual framework for good data governance (Hall and Pesenti, 2017; Hardinges, 2020) where other mechanisms would be unsuitable (e.g., open data, commercial agreements) (Stalla-Bourdillon et al., 2020). Citizen representation is also recognized as a crucial aspect of designing decision-making processes for data trusts (Bunting and Lansdell, 2019). While the notion of the data trust is valid in principle, workable data trusts are in their infancy—meaning that any significant impacts on the growth of responsible and sustainable data sharing initiatives are yet to be realized (Lewis, 2020).

Data foundations as a workable model

In our recent Web Science Institute (WSI) White Paper (Stalla-Bourdillon et al., 2019), we introduced data foundations as a good data governance model that could be used to create independent entities that foster

¹ Foundations (Jersey) Law (2009) and Foundations (Guernsey) Law (2012).

² For instance, the Data Stewards Network has been set up (by the GovLab at the NYU Tandon School of Engineering with support from the William and Flora Hewlett Foundation) to bring together private organisations as part of a “*community of practice*” (Lewis, 2018). Furthermore, Mons (2020) has called on research institutions to invest around “*5% of overall research costs*” in data stewardship.

³ The ODI have recently updated their working definition of “data trust” as follows: “*a data trust provides independent, fiduciary stewardship of data*” (Hardinges, 2020).

responsible and sustainable data usage, management, and sharing, in a manner that will hopefully help incentivize organizations to share (more) data. We therefore define the term “data foundation” as follows:

“An entity incorporated under the Channel Islands’ foundation laws, which supports responsible and sustainable non-personal and personal data usage, sharing, and re-usage by means of independent data stewardship.” (Stalla-Bourdillon et al., 2019)

We maintain that data foundations provide a valuable framework, because they satisfy the following six fundamental components required for any data governance model (Stalla-Bourdillon et al., 2019):

- **A comprehensive rulebook** can be founded through the charter as well as the rules/regulations of the foundation.
- **A strong, independent governance body** can be established through the role of guardian for the foundation.
- **An inclusive decision-making body** can be set up via the council of the foundation.
- **A flexible membership** can be facilitated through amendments to the charter of the foundation.
- **A trust-enhancing technical and organizational infrastructure** can be established before any data are transferred to the foundation.
- **A well-regulated structure** is provided, as foundations are regulated entities.

A key advantage of the data foundations model therefore is that it leverages existing legislation and builds on established precedent (Stalla-Bourdillon et al., 2019). Whereas, there is still a lack of consensus on the most appropriate legal structure for data trusts, for example, trust law debate (BPE Solicitors et al., 2019; Delacroix and Lawrence, 2019). Furthermore, the statutory role of the guardian is a unique requirement in the Channel Islands, which interpreted in a data governance model provides the independent data steward (Stalla-Bourdillon et al., 2019).

The statutory role of guardian

As aforementioned, data stewardship is a vital aspect for all data institutions. A key part of any robust data governance model therefore is a strong, independent governance body composed of independent data stewards with interdisciplinary expertise that are able to oversee the decision-making body.⁴ We define the term “guardian” as follows:

“A legal person, who is independent from the council and who oversees the administration of a foundation to ensure it achieves its purposes in accordance with its charter and regulations/rules. It can, for example, be granted the power to veto council’s decisions. Interpreted in a data governance model this role forms a critical element in providing the independent data steward.” (Stalla-Bourdillon et al., 2019)

The role of the guardian is a unique requirement in the Channel Islands, and is peculiar to these types of structures. All foundations incorporated under Jersey foundations law must have a guardian—although this guardian does not need to be connected with Jersey (Le Cornu, 2009).

Creating Data Foundations in Practice

There are two legal routes for establishing a data foundation:

- (i) **A registered data foundation**—incorporated as a legal entity in the Channel Islands.

Alternatively, where such incorporation would be impractical.

⁴ For background, levels of data stewardship may vary dependent on the type and maturity of the data sharing model in question (Gregory, 2011; Peng et al., 2015; Mills, 2019).

- (ii) **A functional data foundation**—established through a legal structure outside the Channel Islands, which mirrors the key elements of these foundation laws (e.g., companies limited by guarantee [CLG], English law).

Establishing functional foundations under English law

There are various legal structures under English law that are potential candidates for establishing functional foundations outside the Channel Islands. A useful starting point for possible legal frameworks that could mirror and/or incorporate some of the elements of Channel Islands' foundations laws is the report commissioned by ODI (BPE Solicitors et al., 2019). In this report, BPE Solicitors examine five possible legal structures for data trusts referring to UK or English law: “*traditional legal trust model*,” “*contractual framework model*,” “*corporate model*,” “*public model*,” and “*community interest companies model*” (BPE Solicitors et al., 2019). Our approach is slightly different in that we are using Channel Islands' foundations laws as an existing legal benchmark—in order to find the most comparable legal structures.

The key benefits of establishing a legal entity

Since the creation of a legal entity to oversee data sharing makes it possible to simplify the management of the asset, we need first to focus on structures under English law that give rise to a legal entity. It is therefore important to note some key reasons why creating an independent, legal entity⁵ may be of benefit to data sharing initiatives in particular:

- **To widen the range of stakeholders** in the decision-making process to include not only data providers and data users, but also representatives of civil society and the wider community.
- **To offer greater assurances to members of the data sharing initiative** that data are shared, managed, used and reused sustainably—especially for data providers who may be incentivized to participate within a data sharing initiative on the condition that data usage and re-usage is monitored.
- **To provide greater assurances to the public** in that there is no conflict of interest, best practices for data governance are followed—and to make data subject rights more effective.
- **To give certainty over the legal responsibilities and liabilities** of those involved in the data sharing initiative.
- **To make sure that the whole data environment is taken into consideration** by the data governance team.
- **To incentivize independent data stewards** to fulfil their responsibilities in practice.

Given that it is not possible to create a legal entity in its own right through English trust law—and thereby to shift (at least partially) the liability burden from data owners and data users to the functional data foundation—we do not consider the traditional trust model to be an appropriate candidate in these circumstances.⁶

It is further important to note that the statutory role of guardian has more power than the comparable role of enforcer or protector of a trust, because the guardian is able to veto and/or authorize the council's decisions (Le Cornu, 2009).⁷ In the words of Le Cornu:

“This is in my view an extremely important power of the Guardian and takes the Foundation beyond any power that could be given to a protector of a trust. The Council may as a result be able to take actions which have otherwise not been foreseen, which in the case of a company or trust would require applications to court in order to sanction.” (Le Cornu, 2009)

⁵Note that legal entities can be incorporated or unincorporated under English law.

⁶For further background information on the key differences between trusts law and foundations laws in the Channel Islands see Le Cornu (2009) and Guthrie and Moore (2015).

⁷Unless such actions are prohibited by the charter and regulations of the data foundation in question.

The guardian, therefore, is likely to have more input into the day-to-day operation of a data sharing initiative than an enforcer or protector. Moreover, the guardian is likely to require multi-disciplinary skills to oversee the data foundation effectively, and therefore may be a legal entity in its own right with the ability to provide expertise in a variety of areas, including risk and compliance as well as law and information technology.

In terms of structures that give rise to legal entities under English law, there are various potential options for functional data foundations, including: CLG, co-operative societies (COOP), and community benefit societies (CBS).

Example: establishing a functional foundation through a CLG

A CLG appears to be a strong candidate for establishing functional foundations outside the Channel Islands, as it seems to mirror key elements of Channel Islands' foundations laws. For instance:⁸

- **A comprehensive rulebook** could be founded through its articles of association on incorporation.⁹
- **An inclusive decision-making body** could be set up via its board of directors, which could include an independent chairperson.
- **A flexible membership** could be facilitated through membership processes outlined by its articles of association (e.g., for granting and terminating membership).
- **A trust-enhancing technical and organizational infrastructure** could be established before any data are transferred to the CLG, as it does not require any initial capital or endowment.
- **A well-regulated structure** is provided, as CLGs are regulated entities (e.g., under the Companies Act, 2006).

Yet, as the unique two-layered board structure of the data foundation—that is the separate layer for management (i.e., the council of the foundation) and further layer for oversight (i.e., the guardian)—cannot be replicated under English law, a key aspect that needs further consideration is how a CLG, or other legal structure, could provide a strong, independent governance body. Given English law does not give rise to an equivalent statutory role to that of the guardian (as specified by foundations laws in the Channel Islands), the key question therefore is how this key role could be generated through the CLG or another legal structure? One option is as follows:

- **A strong, independent governance body** could be generated and specified through the rulebook of the functional data foundation, for example, via the articles of a CLG, the rules of a COOP or CBS.

To ensure that such prescription would be sufficient to give rise to a guardian is a crucial area for further research—and one that requires expertise from specialists in legal structures under English law as well as foundations laws in the Channel Islands.

Citizen representation

A further key aspect that necessitates further deliberation is how a CLG, or other legal structure, could provide robust mechanisms for citizen representation. Building on the assessment of BPE Solicitors et al. (2019), concerning the representation of data subjects in the context of data trusts, one option is that robust mechanisms for citizen representation could be generated and specified through the rulebook of the functional data foundation, for example via the articles of a CLG, the rules of a COOP or CBS. Such mechanisms could include:

- **A citizen representative** who sits as a mandatory member of the inclusive decision-making body (as per BPE Solicitors et al., 2019).

⁸Note we gratefully acknowledge the work by Womble Bond Dickinson (2020) on this area.

⁹Section 18 of the Companies Act (2006).

- **An advisory group** that reports to the inclusive decision-making body that, as part of its remit, focuses on citizen representation (as per BPE Solicitors et al., 2019).
- **Citizen panels** that have influence within key decision-making processes—that is the incorporation of “*a learning data governance model*” (Banner, 2020)—such as the evaluation of data re-usage outcomes.
- **The independent guardian has well-defined responsibilities** for both oversight of and direct involvement with citizen representation.

Key design principles

So far, we have focused on the rationale for data foundations as good data governance models for data sharing initiatives—and provided an example of how a functional foundation could be established outside the Channel Islands as a CLG under English law. However, it is also important that we look beyond the minimum standards prescribed by law, and outline some universal principles that unite all data foundations—whether they are incorporated in the Channel Islands, or are established as functional data foundations under English law or elsewhere.

We therefore propose all decisions related to the construction and operation of a data foundation should be guided by the following eight key design principles:

- **All data are relevant**—personal as well as nonpersonal data. This is because personal data and nonpersonal data are not binary concepts¹⁰—data shared, used, and reused as part of data sharing initiatives are likely to be personal data or become personal data, for example through purpose of use, data linkage, result of use (Stalla-Bourdillon et al., 2020).
- **Data stewards are independent**—the guardian generates necessary tensions in order to ensure legal and ethical compliance.
- **Expected standards of good practice for data governance specified by a code of conduct**—a code of conduct should be adopted that sets out core principles, practices, and responsibilities for data governance, which is driven by data ethics.¹¹ This code of conduct is essential given that only personal data are heavily regulated and the line between both nonpersonal and personal data is not always clearly drawn.
- **Self-regulation**—internal compliance mechanisms are crucial for ensuring that expected standards for data governance are achieved.
- **Monitoring is the heartbeat**—audit and assurance processes create transparency and accountability.
- **Sustainability**—data foundations must be able to maintain themselves at an accepted level for their lifespan.¹²
- **Accreditation stimulates market growth**—certified data foundations are likely to have a competitive advantage.¹³
- **Stakeholder approvals need to be maintained**—data foundations must work to attain and retain a social license,¹⁴ proactive and meaningful mechanisms for citizen representation are especially important to build trust and confidence.

¹⁰ For further background on the fluidity between personal and non-personal data see: Graef et al. (2018).

¹¹ Note that we are involved with the ongoing development of a Code of Conduct for Data Foundations (CCDF) that can be utilised by any organisation irrespective of its corporate structure. For information about the CCDF, please contact the authors of this paper via hello@lapin.limited.

¹² For example, this is a key focus of the wider data institution movement (Dodds et al., 2020).

¹³ Note that the development of a certification scheme for data foundations remains a key objective for the authors of the paper—certification is likely to be tied to adherence with a code of conduct. Furthermore, the Oxford Insights report (Martin and Pasquarelli, 2019) recommended that “*adequate certification models and processes*” should be established for data trusts.

¹⁴ For further background information on social licence, see Carter et al. (2015) and Jones and Ford (2018).

Conclusion

Data foundations offer a robust workable model for data governance in practice, as they provide: a comprehensive rulebook; a strong, independent governance body; an inclusive decision-making body; a flexible membership; a trust-enhancing technical and organizational infrastructure; and a well-regulated structure. There is therefore an opportunity to advance the wider data institution movement through a legal structure that is ready for use and well-suited to the needs of data sharing initiatives, in particular, since data foundations incorporate the vital element of independent data steward through the statutory role of the guardian.

There are two legal routes for establishing a data foundation in practice, either as a registered data foundation incorporated in Guernsey or Jersey, or as a functional data foundation through a compatible legal structure outside the Channel Islands. In this paper, we focused on CLG as a strong candidate for such a legal structure under English law. Another advantage of the data foundations model for the wider data institution movement therefore is that it provides a legal benchmark for data governance to support the comparative analysis of different legal structures in other jurisdictions.

We further outlined eight universal design principles to unite all data foundations irrespective of legal structure: (a) all data are relevant, (b) data stewards are independent, (c) expected standards of good practice for data governance specified by a code of conduct, (d) self-regulation, (e) monitoring is the heartbeat, (f) sustainability, (g) accreditation stimulates market growth, and (h) stakeholder approvals need to be maintained.

Moving forward, in addition to producing a prototype data foundation, a key area for further exploration is how to establish one or more standardized ways in which the role of guardian for functional data foundations could be created, which aligns with the statutory role prescribed by foundations laws of the Channel Islands. Furthermore, we need to explore how robust mechanisms for citizen representation can be integrated within the data foundation model.

Funding Statement. This work was supported by the Web Science Institute at the University of Southampton under its pump-priming program.

Competing Interests. Sophie Stalla-Bourdillon is a Professor at the University of Southampton and is the Principal Investigator for this work and received the grant for its creation. Laura Carmichael is a Research Fellow at the University of Southampton and declares no competing interests. Alexis Wintour is a Director of Lapin and declares no competing interests.

Data Availability Statement. The data that support the findings of this study are openly available on the University of Southampton Web Science Institute's website at <https://www.southampton.ac.uk/wsi/enterprise-and-impact/white-papers.page>

Authorship Contributions. Conceptualization, S.S-B., L.C., and A.W.; Formal analysis, S.B., L.C., and A.W.; Investigation, S.S-B., L.C., and A.W.; Methodology, S.S-B., L.C., and A.W.; Project administration, A.W.; Resources, S.S-B., L.C., and A.W.; Supervision, S.S-B.; Writing-original draft, S.S-B., L.C., and A.W.; Writing-review & editing, S.S-B, L.C., and A.W.

Acknowledgments. An earlier version of this article was made available for the Data for Policy conference 2020. We extend our special thanks to Krish Dholakia, Womble Bond Dickinson and again to all those that supported and contributed to the White Paper (Stalla-Bourdillon et al., 2019) on which this article is built. Please note that all views and opinions expressed in this article are those of the authors, and do not necessarily represent those involved in the wider consultation process and named above.

References

- Banner N** (2020) A new approach to decisions about data. Understanding patient data. Available at: <https://understandingpatientdata.org.uk/>
- BPE Solicitors, Pinsent Masons and Reed C, Queen Mary University of London** (2019) Data trusts: legal and governance considerations. Available at <http://theodi.org/>
- Bunting M and Lansdell S** (2019) Designing decision making processes for data trusts: lessons from three pilots. Communications Chambers and INVOLVE, commissioned and run in collaboration with the Open Data Institute as part of a project funded by the UK Government's Office for Artificial Intelligence and Innovate UK. Available at <http://theodi.org/>
- Carter P, Laurie GT and Dixon-Woods M** (2015) The social licence for research: why care data ran into trouble. *Journal of Medical Ethics* 4, 404–409.
- Companies Act** 2006.
- Data Stewards Network.** Available at <http://www.medium.com/>

- Delacroix S and Lawrence ND** (2019) Disturbing the ‘one size fits all’ approach to data governance: bottom-up data Trusts. *International Data Privacy Law* 9(4), 236–252.
- Dodds L, Szász D, Keller J, Snaith B, Duarte S, Hardinges J and Tennison J** (2020) Designing sustainable data institutions. Open Data Institute (ODI) report. Available at <http://theodi.org/>
- Foundations (Guernsey) Law** (2012).
- Foundations (Jersey) Law** (2009).
- Graef I, Gellert R and Husovec M** (2018) Towards a holistic regulatory approach for the European data economy: why the illusive notion of non-personal data is counterproductive to data innovation. Tilburg University, TILEC Discussion Paper No. 2018-029. Available at <https://ssrn.com>
- Gregory A** (2011) Data governance—protecting and unleashing the value of your customer data assets. *Journal of Direct, Data and Digital Marketing Practice* 12, 230–248.
- Guthrie M and Moore C** (2015) Guernsey: who owes what and to whom: the roles and duties of trustees, protectors, and enforcers compared to those of counselors and guardians in respect of Guernsey law trusts and foundations. *Trusts & Trustees* 21(6), 645–649.
- Hall W and Pesenti J** (2017) Growing the artificial intelligence industry in the UK. Independent review. Available at <https://www.gov.uk/>
- Hardinges J** (2020) Data trusts in 2020. Open Data Institute (ODI) blog. Available at <http://theodi.org/>
- Hardinges J and Tennison J** (2020) What do we mean by data institutions? Open Data Institute (ODI) blog. Available at <http://theodi.org/>
- Jones KH and Ford DV** (2018) Population data science: advancing the safe use of population data for public benefit. *Epidemiology and Health* 40, e2018061.
- Le Cornu P** (2009) The Jersey Foundations Law 200[–]. *Trusts & Trustees* 15(3), 140–147.
- Lewis P** (2020) Peter Lewis’s 2020s vision: stop glibly signing over your data and take control. *The Guardian*. Available at <https://www.theguardian.com/uk>
- Lewis T** (2018) Data Stewards: data leadership to address the challenges of the 21st century. GovLab, NYU, Tandon School of Engineering. Available at <http://thegovlab.org/>
- Mac Manus S** (2020) How can data transform our health and care system? *Nesta Blog*. Available at <https://www.nesta.org.uk/>
- Martin S and Pasquarelli W** (2019) Report: exploring data trust certifications. Oxford insights report commissioned and run in collaboration with the Open Data Institute (ODI) as part of a project funded by the UK Government’s Office for Artificial Intelligence and Innovate UK. Available at <http://theodi.org/>
- Mattioli M** (2017) The data-pooling problem. *Berkeley Technology Law Journal* 32(1), 179–236.
- Mills S** (2019) Who owns the future? Data trusts, data commons, and the future of data ownership. Available at <https://ssrn.com>
- Mons B** (2020) Invest 5% of research funds in ensuring data are re-usable. *Nature* 578, 491.
- Peng G et al.** (2015) A unified framework for measuring Stewardship practices applied to digital environmental datasets. *Data Science Journal* 13, 231–253.
- Smart Dubai and Nesta** (2020) Data sharing toolkit: approaches, guidance and resources to unlock the value of data. Available at <https://www.nesta.org.uk/>
- Stalla-Bourdillon S, Thuermer G, Walker J, Carmichael L and Simperl E** (2020) Data protection by design: building the foundations of trustworthy data sharing. *Data & Policy* 2(e4), 1–10.
- Stalla-Bourdillon S, Wintour A and Carmichael L** (2019) Building trust through data foundations: a call for a data governance model to support trustworthy data sharing. Web Science Institute White Paper #2. Available at <https://eprints.soton.ac.uk/>
- The Royal Society & The British Academy** (2017) Data management and use: governance in the 21st century—priorities for data governance: discussions at a British Academy and Royal Society seminar on 16 October 2017. Available at <https://www.thebritishacademy.ac.uk/>
- Verhulst SG** (2018) The three goals and five functions of data Stewards—Data Stewards: a new role and responsibility for an AI and data age. Medium. Available at <https://medium.com/>
- Womble Bond Dickinson** (2020) Recommendations on the most appropriate English legal structure to satisfy the six fundamental components for any data governance model proposed by the University of Southampton. Internal document.