OXFORD INSIGHTS

Report

Exploring Data Trust Certifications

April 2019

OXFORD INSIGHTS

Acknowledgements

This report has been researched and produced by Oxford Insights and published in April 2019. Its lead authors were Dr Sabrina Martin and Walter Pasquarelli (OI), with thanks to Leigh Dodds (ODI), Richard Stirling (OI), Peter Wells (ODI), and Dr Jeni Tennison (ODI) for comments. Please contact research@oxfordinsights.com for more information.

This project was 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. It builds on research from the ODI's Innovation programme funded by Innovate UK. The views in this report are those of the authors.

Executive summary

Data trusts are a relatively new concept, championed both by the UK Government and the Open Data Institute (ODI), that propose to increase access to data whilst maintaining trust. Organisations and individuals that use data provided by a data trust will want to understand how it is governed and whether access to that data will be sustainable and ethical. Thus, the need arises for a certification procedure evaluating a data trust's processes against specific standards and norms.

This report explores potential certification models for data trusts, as well as standards that certification models should seek to assess. It additionally looks at a number of organisations providing certification processes for related elements surrounding trust in data trusts. Depending on the specific use case of the data trust, we will need to consider which organisations might be most suitable for providing certifying standards as there might be conflicting interests of stakeholders that will need balancing. Working with open standards could provide a potential means for ensuring transparency and safeguarding trust in modelling certification.

This report also identifies a number of existing organisations who set standards. It is unclear whether an individual data trust should be responsible for setting the standards, or even whether this would be desirable, as different stakeholders might have different needs and interests. However, these standard setting bodies could provide a general guidelines for facilitating the sharing of data that a data trust could use.

No one body which was identified covers all the areas which would be needed to certify a data trust. For future work, people who want to create an enabling environment for data trusts should contact certifiers in order to assess their capabilities and interests, and develop initial prototypes of certification models. These prototypes should then be tested with potential stakeholders in order to identify areas of improvement and inform future research.

Introduction

As one expert recently noted, participation in the digital economy is currently a ‘zero-sum game’:[1] a person either shares data about themselves and reaps the benefits of the digital age, or they maintain data privacy but do not participate in the world’s biggest industry. The Open Data Institute (ODI) believes in the potential of increasing access to data as a means for tackling social problems, stimulating economic growth, and boosting innovation. Yet, there are growing concerns about how to maintain trust in the digital age: who has access to this data, where it is stored, and how it is used - to name just a few.

One proposed solution to these issues is the establishment of data trusts, which are institutions or bodies that help to give “people and organisations confidence when enabling access to data in ways that provide them with some value (either directly or indirectly) in return.”[2] In partnership with the UK government, the ODI has recently undertaken a pilot programme exploring potential benefits and application of data trusts.[3]

What are data trusts?

The idea of a data trust became prominent in the UK in 2017 following a UK Government report that recommended the establishment of data trusts in order to “to improve trust and ease around sharing data.”[4] Because the idea of a data trust is so new, there is little common agreement on what exactly one is, how it might be established, or even what specifically it would do. The ODI has since settled on a specific definition of a data trust as a “legal structure which provides independent third-party stewardship of data for the benefit of a group of organisations or people.”[5]

Stewardship Approaches

As a steward of data, a data trust makes decisions on who has access to data, conditions attached to it and determining who the main beneficiaries are. Where ordinarily an organisation that collects and holds data will automatically be the one to steward it, one or more organisations might allow a data trust to make decisions about how that data is used and shared. The data trust will

[1] Bots and Artificial Intelligence NYC Meetup (2019) *Introduction to Data Trust*. Available at https://www.youtube.com/watch?v=Ru5So7NunZQ.
[2] The ODI (2018) *What is a data trust?*. Available at https://theodi.org/article/what-is-a-data-trust/.
[3] https://theodi.org/article/uks-first-data-trusts-to-tackle-illegal-wildlife-trade-and-food-waste/
[4] UK Government (2017) *Growing the artificial intelligence industry in the UK: Executive Summary.* Available at https://www.gov.uk/government/publications/growing-the-artificial-intelligence-industry-in-the-uk/executive-summary
[5] The ODI (2018) *Defining a data trust*. Available at https://theodi.org/article/defining-a-data-trust/.

make decisions about its use for a previously agreed purpose and scope whilst taking into consideration all relevant stakeholder's interests.

Data trusts are independent from both the organisations that hold the data and the prospective data users. In order to achieve and safeguard this independence, data holders and data users may be precluded from making decisions about data access, or may be included in decision making but prevented from dominating it. Crucially, a data trust's trustees take on a (legally binding) responsibility to ensure that the data is shared and used to the benefit of a particular group of people and organisations, as well as other stakeholders affected by its use.

While data trusts need not take the form of 'trusts' in a legal sense, they are inspired by relevant legal structures. For instance, community land trusts have long been used to steward gardens, civic buildings and other community assets on behalf of a community; trust ports are similarly run by independent boards for the benefit of different stakeholders and are governed by their own local rules. It should be noted that data trusts may vary in their form and structure depending on the specific use case. In other words, there is no one-size-fits-all approach as different stakeholders may have different interests and sensitivities surrounding the sharing and use of particular kinds of datasets.

A question that remains to be explored, however, is how data holders, users, and other people who want to create an enabling environment for data trusts know how to trust a data trust. The aim of data trusts, after all, is to assure stakeholders providing, using and impacted by data that it is stored securely and used responsibly, so the next logical step is to have some means of independent assurance guaranteeing the quality of a data trust for all relevant stakeholders. There currently are numerous organisations[6] that certify data security or even data ethics, but in terms of certifying overall trustworthiness, so to speak, there is little to model data trust certification after.

Oxford Insights, in partnership with the ODI, has prepared this report to explore potential avenues of certification surrounding data trusts.

Exploring data trust certification

There are multiple models of oversight for a data trust. This report largely focuses on what it would mean to give a data trust an official stamp or seal of approval by some body, or a kind of kitemark or badge by consumer group. Who this regulatory body would be and what it might look like is also explored.

[6] CIO (2018) *26 big data certifications that will pay off*. Available at https://www.cio.com/article/3209911/big-data-certifications-that-will-pay-off.html

There are two relevant procedures in the case of data trusts, namely ‘certification’ and ‘accreditation’. Certification refers to evaluating processes or systems against certain standards. Accreditation, in turn, is done by a third party, assessing the competence of an organisation to perform specific tasks. In the context of data trusts, an independent body would *certify* that a data trust’s processes and services meet required standards and norms. Another or the same third-party, in turn would *accredit* that a potential data trust has actually the competence to be considered a data trust. For the purposes of this report we shall mainly focus on certification models of data trusts. That said, there remain questions of whether certification is the correct model for a data trust in the first place (instead of, e.g. a centralised body with oversight), and subsequently who would regulate these certifying institutions. These are questions that will need to be explored in further detail in the future, if and when data trusts begin to proliferate.

Another question that will need to be clarified is whether or not the standards for certification would be open, and several bodies would then be licensed to certify them. This open model fits with the ODI’s objectives. As discussed below, many of the existing certifying bodies do not publish their certification standards. Considering that the point of data trusts is to build *trust* for stakeholders, openness as to what elements are being certified seem to be key to establishing this trust and legitimacy.

Elements to certify

There are any number of elements that potential organisations certifying a data trust could examine and approve. As an initial, and by no means, definitive list, we propose that such body should, at a minimum, look to these following aspects of data trusts:

1. **Governance**: how the data trust is run, by whom and its decision-making processes
2. **Data access procedures**: the rules and regulations surrounding the sharing, provision, and access of data;
3. **Accountability**: whether there are appropriate auditing mechanisms in place for:
    a. Assessing whether the data is used for the purposes and scope that was defined when granting access to the data;
    b. Potentially auditing algorithms that are created as a result of the data received from the data trust; and
4. **Finance**: how the data trust is funded and whether the financial model is sustainable;
5. **Security**: this includes not just the cybersecurity surrounding the data in the repository, but also policy regulations that the trust has in place to assure stakeholders that the data is protected;
6. **Ethics and environmental sustainability**: whether or not the data trust’s rules of conduct match up to its stated ethical principles, sustainability standards, and overall purpose;

7. **Quality**: although the quality and granularity of the data itself will be relative to its applications, we should look to assure quality and consistency surrounding how data is collected, managed, and shared.

N.B. **Benefits distribution:** an independent report should be provided on the benefits distribution of a data trust evaluating how benefits are shared and whether they are distributed in an equitable way

Certification for whom?

One potential issue is that different stakeholders might have different concerns when it comes to the quality and legitimacy of the bodies guarding their data. For example, consumers might care most about data protection, whereas corporate boards might be most interested to where the finances come from. Data users might care most about the quality of data collection. Thus, it might be worth exploring the potential of having various certification schemes depending on the relevant stakeholders. Alternatively, there could be official certification stamps that are more relevant to business dealings, while organisations like Which? informally review data trusts and publish the results for consumers' information. In order to create a certification standard that takes into account the interests of all stakeholders, these different views could also be incorporated into a certification model. More research will be needed in order to establish the best model of certification for each relevant stakeholders.

Who would certify?

It remains unclear what the appropriate institutional model for certifying data trusts would be. There seem to be three broad options:

1. Government or public bodies,
2. Non-profit/non-governmental organisations, or
3. Private corporations.

Each of these actors has benefits and drawbacks when acting as a certifying body. For example, while government departments tend to have high levels of respect, they could be swayed by lobbying, so their neutrality might be brought into question. NGOs might benefit from a commonality of purpose in regulating the data trusts, but the certification process could result in a power struggle amongst major actors in the field. Finally, private corporations most commonly act as certifiers in other sectors, but when it comes to data trusts, it is worth considering if commercial interests might get in the way of independence.

All in all, different stakeholders will almost certainly have different expectations and ideals of how certification models should work. Furthermore, there could be power imbalances at play in influencing the process. Open standards could provide a means for mitigating the risks presented, by using open and transparent processes with broad participation from those being assessed. Standards that are agreed upon and maintained through open processes could additionally add a layer of trust in a certification process and the methods by which it is applied.

Sample trusted data certification schemes and organisations

Below is a list of organisations that currently offer some type of certification similar to what is being proposed in this report. This list is non-comprehensive both in terms of the individual organisations included, as well as the certification processes and standards they employ. Instead, the goal here is to highlight what is already being done in terms of certification, where there is common ground, and where we might look for improvement.

Carbon Trust

Link: https://www.carbontrust.com/client-services/certification/assurance-certification/
Founding Year: 2001
Overview: Carbon Trust offers various services in relation to assuring and certifying best environmental practices across a range of industries. Relevant to this report is their environmental verification services which, according to their website "provide the assurance necessary for organisations to report environmental data with confidence, offering the credibility and accuracy necessary to satisfy stakeholder, employee and customer expectations."[7] They offer a five-step verification process[8] based on ISO 14064 methodology, which is part of the international standard for assessing environmental management.[9]
Business Model: Carbon Trust is registered as a non-profit company. They charge an annual administrative fee of £1,499.
Organisations eligible to apply: "Suppliers and contractors with a significant track record in the design, supply and installation of energy efficient equipment and renewable energy technologies."[10] To date, they have certified 142 repositories worldwide.[11]

[7] Carbon Trust (2019). *Assurance & Certification*. Available at https://www.carbontrust.com/client-services/certification/assurance-certification/

[8] Carbon Trust (2019). *Verification & Assurance Services*. Available at https://www.carbontrust.com/media/677265/carbon-trust-verification-assurance-services.pdf

[9] Carbon Trust (2019). *Verification*. Available at https://www.carbontrust.com/client-services/certification/verification/

[10] Carbon Trust (2019). *Accreditation Scheme UK.* Available at https://www.carbontrust.com/client-services/certification/accredited-supplier/

[11] Carbon Trust (2019). *Green Business Directory*. Available at https://www.carbontrust.com/resources/green-business-directory/

Assessment: In terms of data, the Carbon Trust certification procedures are mostly useful in terms of a first-step certification, that is, ensuring that the data being collected and reported is accurate and secure. That said, the model might be useful in several ways. First, it seems like the certification procedures could easily scaled to include data trusts, and not just individual organisations who currently use the service. Second, they are one of the few certifiers concerned specifically with data sustainability and carbon footprints, meaning that even if their particular model is not scalable, the ideas behind the project will be important to replicate in assessing sustainability for data trust certifiers. One drawback is that they are a private company so public partnerships and/or combining it with other certification methodologies might prove difficult.

CoreTrustSeal

Link: https://www.coretrustseal.org/
Founding Year: 2017
Overview: CoreTrustSeal is a Netherlands-based non-profit data repository certification organisation. It was established in collaboration between the World Data System of the International Science Council (WDS) and the Data Seal of Approval (DSA). They have openly published the standards that they use when certifying data repositories. The full guidance notes can be found here, but key elements of the certification process include Mission/Scope, Licenses, Continuity of Access, Confidentiality/Ethics, Organizational infrastructure, and Expert guidance.[12] Additionally, "CoreTrustSeal is a legal entity under Dutch law (CoreTrustSeal Foundation Statutes and Rules of Procedure) governed by a Standards and Certification Board composed of 12 elected members representing the Assembly of Reviewers."[13] CoreTrustSeal also plans to appoint an Advisory Committee to supplement the Board and provide links to "the wider data community including other certification standards."[14]
Business Model: Core Trust Seal is registered as a non-profit organisation with the aim of "promoting sustainable and trustworthy data infrastructures." It is funded by subventions, donations accepted by their board and through in-kind contributions from the Dutch Data Archiving and Networked Services. Additionally, they charge an administrative fee of 1,000 Euros for each certification.
Organisations eligible to apply: infrastructure providers, repository software providers, bit-level replication services, national archives, "as well as commercial services designed to help preserve and protect research data and the world's digital legacy." To date, they have certified 143 repositories across the globe.

[12] Carbon Trust (2019). *Core Trustworthy Data Repositories Requirements*. Available at https://www.coretrustseal.org/wp-content/uploads/2017/01/Core_Trustworthy_Data_Repositories_Requirements_01_00.pdf
[13] Core Trustseal (2019) *About*. Available at https://www.coretrustseal.org/about/
[14] Ibid.

Assessment: The CoreTrustSeal matches up very closely to the type of certification body model that has been envisioned in this report, and indeed has the legal structure that the ODI envisions data trusts taking on in the future. As a non-profit, CoreTrustSeal's commercial interests will seem more neutral than for-profit corporations. Further, the fact that they publish their standards of assessment is a promising start for openness and trust. According to their website, "The CoreTrustSeal certification is envisioned as the first step *in a global framework for repository certification* which includes the extended level certification (nestor-Seal DIN 31644) and the formal level certification (ISO 16363)"[15] (emphasis added). CoreSealTrust seems to be an ideal place to start when looking to establish a model for wide-spread data trust certification.

Fair Data

Link: https://www.fairdata.org.uk/
Founding Year: 2013
Overview: Fair Data is a certification scheme that was launched by the Market Research Society (MRS)[16] in 2013 to certify companies that "handle their customers' personal data fairly."[17] Fair Data certifies consumer organisations, research and data suppliers, public and government bodies, and consumers themselves.[18] According to their website, "The Fair Data mark is a consumer facing mark which appears on corporate materials as a guarantee that an organisation meets the Fair Data principles." In this way, Fair Data accreditation is a similar model to the Fairtrade mark that consumers will recognise on coffee and bananas. Fair Data has published 10 Principles of fair data usage that organisations must subscribe to in order to receive accreditation. These principles include principles surrounding data collection, security, ethics, and supply chains, but are not as rigorous or comprehensive as, for example, CoreTrustSeal's standards.
Business Model: Fair data is registered as a non-profit company. An initial advisory visit by Fair Data is currently priced at £1,000. If organisations fail the accreditation process additional visits are priced at £500 per day. Certifications are due to annual renewal for £350.
Organisations eligible to apply: Public/ government bodies, consumer organisations consumers, suppliers of research and data
Assessment: Fair Data seems like a promising schema for certifying several aspects surrounding trust in data, but the one major drawback for this particular project is that their 10 principles do not include anything about governance. That said, MRS recently undertook a review of their principles, suggesting that they would be open to reviewing them again in the future. Consequently, the Fair Data mark might be a useful model to scale up to certify data trusts in general.

[15] Ibid.
[16] MRS is an independent market research regulator, established in 1946.
[17] MRS (2019) *Fair Data.* Available at https://www.mrs.org.uk/standards/fairdata
[18] Fair Data (2019). *Who it's for.* Available at https://www.fairdata.org.uk/who-its-for/

TrustArc’s TRUSTe Data Collection Certification

Link: https://www.trustarc.com/products/data-certification/
Founding Year: 1997
Overview: San Francisco-based TrustArc is a private company that specialises in technology compliance and security. Among several certification services they provide, TrustArc offer a TRUSTe Data Collection Certification for companies that “that act as a 3rd Party data collectors.”[19] The certification process proceeds in three phases. First, they perform an assessment where they undertake a privacy review and produce a report. Second, they provide remedial steps, if any need to be taken (though what these might be are not listed). They then give the company a Letter of Attestation and a TRUSTe Privacy Certification Seal. Third, they provide ongoing monitoring and guidance for the company, complete with a dispute resolution service and a feedback button for users.[20]
Business Model: Trust Arc’s is registered as a for-profit company. Fees undisclosed.
Organisations eligible to apply: Organisations acting as 3rd party data collectors through mobile or desktop environments. This includes companies that collect data, “such as Personally Identifiable Information (PII) and sensitive segments.”[21]
Assessment: The TRUSTe Data Collection Certification seems to include the type of *process* that would be extremely relevant for data trust certification. However, their standards of assessment are not public and so analysing the scalability or overall appropriateness of the model remains unclear. Further, it does seem that the certification is mostly for the purposes of privacy, but that does not preclude issues of governance or sustainability, for example.

Matrix of comparisons of existing bodies

	CARBON TRUST STANDARD REDUCING CO2 IN THE SUPPLY CHAIN	CORE TRUST SEAL	Fair Data™	TRUSTe Certified Privacy Powered by TrustArc
Governance		✓		
Data Access Procedures		✓	✓	

[19] TrustArc (2019). *Data Certification*. Available at https://www.trustarc.com/products/data-certification/
[20] Ibid.
[21] Ibid.

Accountability		✓	✓	✓
Finance				
Security		✓	✓	✓
Ethics and Sustainability	✓	✓		
Quality				

Who sets the standards?

One final question to explore, as discussed above, is whether or not it makes sense for these certification standards to be globally uniform (i.e. offer open standards set by some overarching respected body who then accredits individual bodies to offer certification stamps), or if it makes sense to have various certification schemes depending on the needs of the stakeholders.

If the standards are to be uniform, there are a number of models of institutions that could set the standards. Below, various examples of these standard setting bodies and certifiers are explored. It should be noted that none of these bodies is necessarily specifically related to data trusts, or even the IT sector.

Example Standard Setting Bodies

The **International Auditing and Assurance Standards Board (IAASB)** is "an independent standard-setting body that serves the public interest by setting high-quality international standards for auditing, assurance, and other related areas, and by facilitating their adoption and implementation." The IAASB could serve as the standard setter for certification schemes that private companies or other non-profits could then participate in.

The UK's **Information Commissioner's Office (ICO)** and the EU's **European Data Protection Board (EDPB)** have jointly started a certification programme to assess bodies that are in compliance with General Data Protection Regulation (GDPR). According to the ICO's website, there are no UK-wide certification schemes, yet, but "Once the certification bodies have been accredited to issue GDPR certificates, you will find this information on ICO's and UKAS's websites."[22] The ICO and EDPB have also said there is scope for instituting a European Data Protection Seal. These same bodies could undertake a programme of accrediting data trust certification schemes.

The Institute of Electrical and Electronics Engineers (IEEE), is a professional organisation that established a **Standards Association (IEEE-SA**) that publishes various standards relevant across various technology sectors. The IEEE-SA accepts both corporate and individual members who can then influence projects and standards. The IEEE-SA is not authorised by any formal body (unlike the ISO, for example), but instead operates through a community of respect. A body like this could also serve as a standard-setter for data trust certification schemes.

The **International Standards Organisation (ISO)** is an independent, NGO comprised of 164 national standards bodies. Through its national members, it produces consensuses on relevant standards for goods and services and codifies them into International Standards. The ISO could serve as a standard-setting body for the elements and standards necessary to certify data trusts.

Which? is a registered charity based in the United Kingdom that helps consumers access information about goods and services by reviewing them and publishing their findings. As a trusted consumer body, an independent organisation such as Which? could provide assessments of data trusts. Currently, only products receive the Which? stamp of approval, the 'Best Buy' logo -- but this endorsement could be expanded to services.

Recommendations

To conclude, data trusts could potentially provide a huge opportunity to reap the benefits of increasing access to data whilst maintaining trust. As part of this it will be essential to have adequate certification models and processes in place. Oxford Insights recommends that people who want to create an enabling environment for data trusts should **reach out to relevant organisations** that could provide certifications for data trusts. As part of this exploration phase, these organisations' interests, terms and conditions, and capabilities for providing this type of service should be investigated.

A deeper exploration into certifying data trusts should include **interviews with these relevant organisations** to cross- and sense-check the standards and norms recommended here, as well as those that the organisations above also certify in addition to the ones in this report. Further, it

[22] Information Commissioner's Office (2019). *Certification*. Available at https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protection-regulation-gdpr/accountability-and-governance/certification/

will be important to hold **focus groups with stakeholders** to determine what standards and norms are important to them.

The ODI has already conducted some research on potential data trust stakeholders as part of their Data Trusts Pilot Programme. Findings from the pilots could be sufficient to **build initial prototypes of certification models** and/or **offer a list of initial standards to be used in the certification process**. Further insights could also be garnered through additional rounds of stakeholder interviews. These prototype certification models should take into account both the preferences of stakeholders within specific use cases, as well as what potential partners/organisations could offer. As part of this, they should explore whether or not certification models should be standardised across organisations and for varying stakeholders.

Finally, research could be undertaken to **test potential certification models with stakeholders** to obtain insights and feedback from relevant parties. Findings from this phase should be used to adapt and enhance prototype certification models to the needs of stakeholders, and to **identify areas for future research**.

As discussed above, there are a variety of areas that need further exploration in order to determine an appropriate framework for building certification models. As part of this, we propose a **discovery project of approximately 6 months** for conducting a minimum of 60 user interviews. These interviews should include certifiers in order to understand capabilities, experiences and interests; regulators, to understand the legal and normative landscape of creating certification models; and lastly, potential stakeholders of a data trust to test prototype certification models and understand their needs and concerns.