Containing Multitudes

Aggregating Personal Data Access Contracts to Create a Bottom-Up Data Trust

Dataswift™

This work is funded in part by the School of Technology, University of Cambridge through sub-award G109655 as a contribution to the Cambridge Data Trust Initiative.

Containing Multitudes

Aggregating Personal Data Access Contracts to Create a Bottom-Up Data Trust

Dataswift White Paper Series 02/2021-DS

Date 29 Nov 2021

Issued by Dataswift Ltd

Authored by James Kingston & Paul Tasker, Dataswift Platform Policy team

A position paper prepared by Dataswift Ltd to review a new approach to settling data in trust enabled by Dataswift's proprietary technology platform. The work is funded in part by the School of Technology, University of Cambridge through sub-award G109655 as a contribution to the Cambridge Data Trust Initiative.

Contents

Containing Multitudes

Aggregating Personal Data Access Contracts to Create a Bottom-Up Data Trust

Data Trusts: An Evolving Concept

Data: A problem of rights, a problem of property

Data-driven technologies and digital business models have proved transformative to the economy, creating great wealth and reshaping entire industries. However, while data has been a fundamental resource of this digital transformation, data cannot be considered property.

This presents a problem regarding personal data – the data generated by or on individuals as they interact with digitally enabled systems. This information can be both valuable and highly sensitive. Despite this, individuals are unable to directly control the flow of this data, or receive value for its use. Subject to the commercial decision-making of data collecting institutions, individuals are ill-equipped either to protect their data privacy, or participate on equal terms in the data marketplace.

According to Ng[1], the inability of individuals to express rights over this data prevents economic optimality. Should clear property rights exist in personal data, negative externalities around privacy and data exploitation can be solved more efficiently. However, given the right technical framework, the necessary individual legal entitlements over information can be expressed and optimality achieved.

This paper examines how digital property rights for individuals can be used in the context of the 'data trust' to generate collectivist solutions to asymmetries in the personal data market.

[1] Market Design for a Property Rights System with Entitlements for Individuals. Ng, Irene C L, WMG Service Systems Research Group, Working Paper, Series Issue number: 01/21, ISSN: 2049-4297 November 2021

What is a data trust?

‘Data trusts’ are an emerging concept. The term is commonly used to describe frameworks for sharing, aggregating, managing and accessing data in ways that are deemed to have a public benefit, or to benefit the position of individual data generators. Data trusts vary with use case and data type.

A core distinction is to be made between those who envisage data trusts as legal structures, and those who see them as more informal institutions for the mutual sharing of data. Whereas the London Economics[2] define a data trust as “a legal structure that provides independent stewardship of data”[3], O’hara[4] argues that a data trust is not a trust in a legal sense, and is instead a group of partners who trust each other for data sharing.

Alternatively, Hall and Pesenti define a data trust as, “proven and trusted frameworks and agreements...to ensure exchanges are secure and mutually beneficial”. They elaborate further that “trusts are not a legal entity or institution, but rather a set of relationships underpinned by a repeatable framework[5].”

Data trusts can have a range of different actors, institutions and data types. Local authorities may seek, for instance, to place all their environment-relevant data into some form of trust in order to enable researchers and entrepreneurs to work on decarbonisation solutions. There appears to be no dominant model, meaning that there is a significant overhead in designing and establishing a trust for any particular context.

For the purposes of this paper, we will focus on data trusts over personal data.

The potential role of trust law

A core question around data trusts is whether they can or should be formally based on the law of trusts. Many commentators believe that English trust law provides the basis by which individual data subjects can best safeguard their interests in the digital economy.

[2] *Independent Assessment of the Open Data Institute’s Work on Data Trusts and on the Concept of Data Trusts Report to the Open Data Institute,* Godel, Moritz; Natraj, Ashwini. April 2020, London Economics
[3] *Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership*. Mills, Stuart, (September 24, 2019)
[4] *Data Trusts: Ethics, Architecture and Governance for Trustworthy Data Stewardship*. O'hara, Kieron (2019) (WSI White Papers, 1) University of Southampton.
[5] *Growing the Artificial Intelligence Industry in the UK.* Hall, Wendy; Presenti, Jerome. October 2017, UK Government

English trust law is based on a separation between legal ownership and the 'equitable' rights certain persons may have over property, despite not holding legal title over it. By settling property into a trust, individuals (known as settlors) transfer legal ownership in this property to a second party, the trustee. In making decisions about that property, a trustee is bound to act in the interest of a group of persons known as beneficiaries – a set of individuals or entities whom the settlor wishes to gain the benefit of the trust property. Despite not formally owning the property, beneficiaries are afforded equitable rights over it, the nature of which may vary with the rules of the trust.

Irrespective of these rules, trustees have a range of basic fiduciary duties to beneficiaries. Core requirements include the duty to act in the best interests of the beneficiaries, the duty to avoid placing themselves in a position of any conflict of interest, and the duty to act with reasonable care and skill.

For most people, making decisions about the best use of their personal data is difficult. Individuals are not equipped to easily understand how their data may be used, nor about the potential risks and benefits of this use. With each of us generating personal data every time we use digital systems, making decisions on how this data ought to be managed and governed would impose a high cognitive load.

Just as many people will put their money in investment funds run by individuals with specialist financial knowledge, so too may they see value in relying on specialist managers of personal data to make the best decisions on how their data may be used. Should this management occur within the framework of trust law, then the managers will be bound by fiduciary obligations to the data providers – similar to those held by the managers of financial assets.

Models of data trust

Surveying models and philosophies of data ownership, Mills[6] outlines three potential forms of data trust – the collector-centric data trust, the data-centric data trust and the generator-centric data trust. Of these, both the data-centric and generator-centric trusts may be considered 'bottom up' data trusts along the lines proposed by Delacroix and Lawrence[7]. We examine all three forms in the following pages.

[6] *Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership*. Mills, Stuart, (September 24, 2019)

[7] *Bottom-up data Trusts: Disturbing the 'One Size Fits All' Approach to Data Governance*. Delacroix, Sylvie, Lawrence, Neil D, International Data Privacy Law, Vol 9, Iss 4, Nov 2019, pp 236–252

Collector-centric

Data collectors are institutions that collect data generated by individual data generators. This data may be data generated through the interaction of individuals with digital systems owned by the data collector, or it may be acquired by the data collector from a third party. Data collectors may be any kind of organisation, such as companies, hospitals, governments, and universities.

In the collector-centric data trust, the collectors collect data from their users, then agree with other data collectors to pool this data in a mutually beneficial 'data trust' structure that will enable the data collectors to more easily and compliantly access, analyse, and use one another's data. The benefits of this trust would accrue primarily to the data collectors; seen from the perspective of the data generators, little changes.

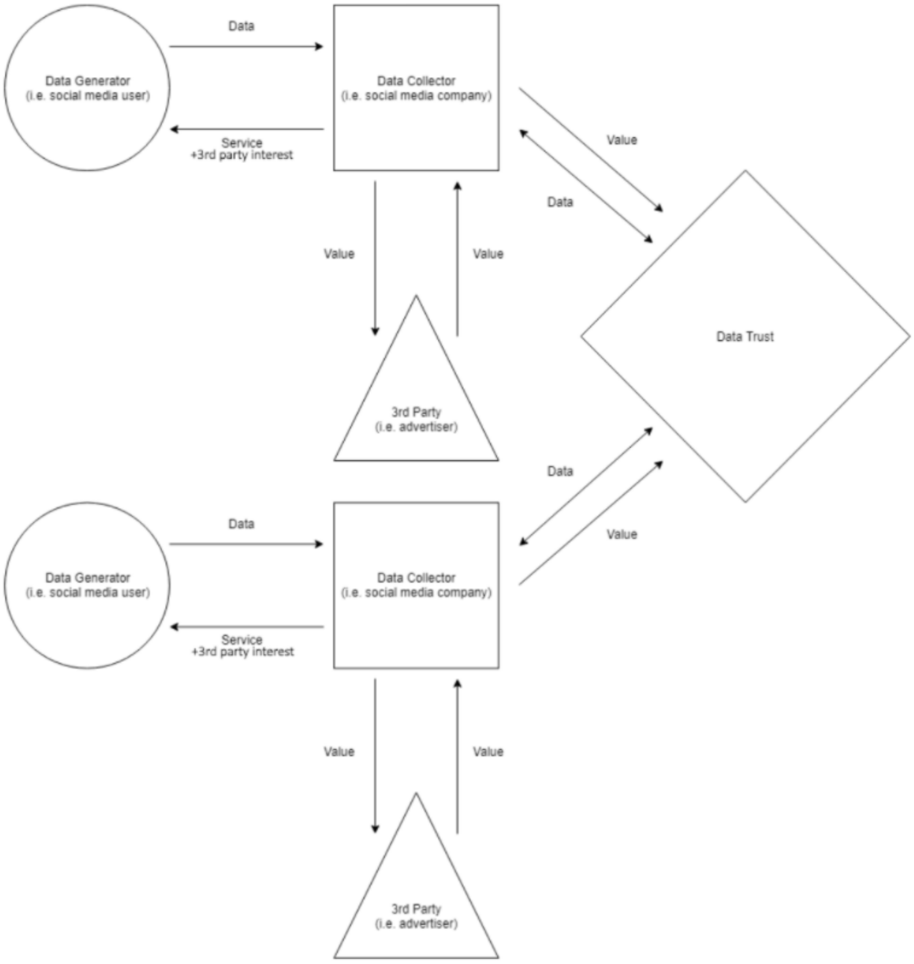

Figure 1: Collector-centric trust (Mills 2019)

Generator-centric

In generator-centric trusts, data is conceived as an individual resource held by the data generators. The data itself is not pooled in a central 'data trust' entity. Rather, the data trust sets the terms by which data collectors have access to the data of individual trusts over data contracts.

Mills[8] envisages that in a generator-centric data trust, the trust does not itself steward data – rather, its stewardship obligations involve the protection of individual members.

Notably, the efficacy of this model requires individuals to have some form of personal access, control, and/or storage capacity in relation to their own data.

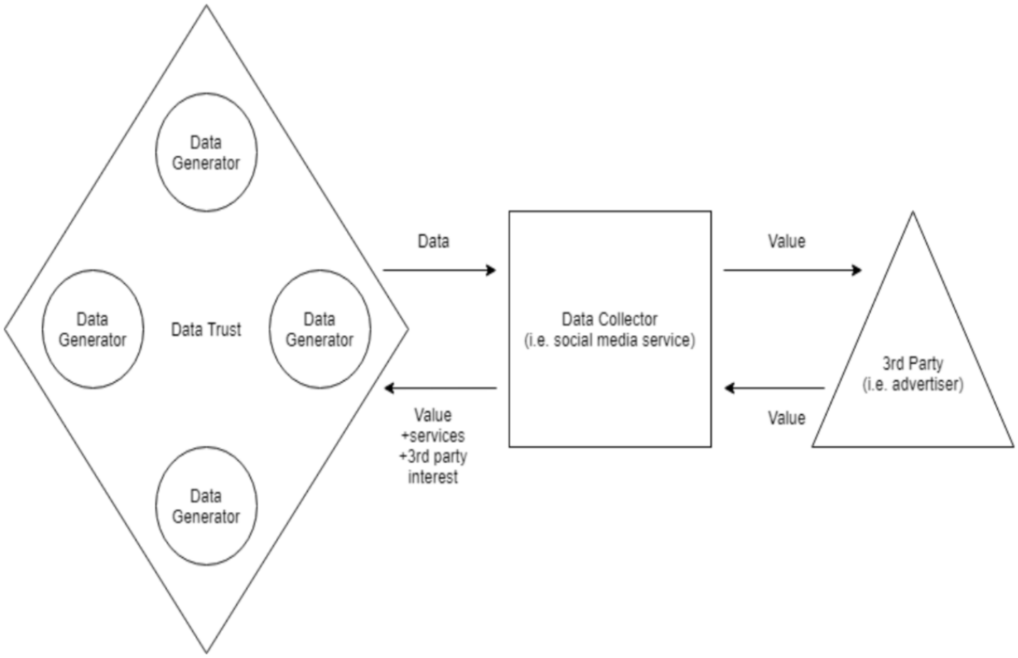

Figure 2: Generator-centric trust (Mills 2019)

[8] *Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership*. Mills, Stuart, (September 24, 2019)

Data-centric

Data-centric data trusts involve data generators pooling their data into a data trust. The data trust then sits between the generators and other organisations who might wish to make use of their data. It aggregates the data of users and negotiates with external parties who might wish to access this data.

A challenge faced by potential data-centric data trusts is that of data collection - how do the generators themselves collect their data and contribute that data to the trust so that it may negotiate with third party organisations that wish to use the data?

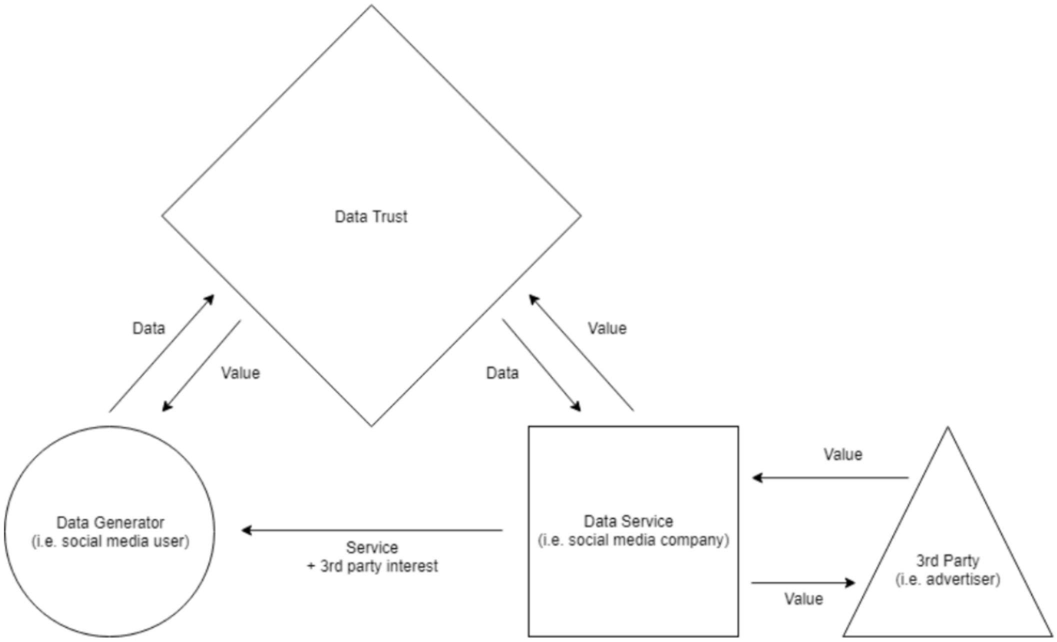

Figure 3: Data-centric trust (Mills 2019)

Dataswift's distributed infrastructure contributes to this debate in that it provides a mechanism for individual data generators to gain access to their data from multiple data collectors, and centralise this data onto their own 'Personal Data Accounts' (PDAs). Moreover, the infrastructure includes proprietary legal code so that the data collector settles rights-in-use of the data onto the individual data generator (the data subject) allowing them, the data generator, to transact the data with whomever they choose. This means that as PDA owners, the individual data generators can contribute to the creation of both data-centric data trusts and generator-centric data trusts as outlined in the above typology.

Some key presumptions

For the purposes of this paper we propose to examine a data trust that is firstly generator-centric and secondly a valid trust. We therefore presume that individual data generators can be equipped with control over their personal data, and that they can make decisions regarding whether or not to contribute some or all of their data to a trust. We also presume the existence of property rights pertaining to this data.

We make these presumptions based upon our knowledge of the possibilities inherent to Dataswift's technology.

The Role of Dataswift

Introduction

As outlined by Ng[9], personal data itself cannot and should not be propertized, but given the right framework, individual legal entitlements over information can be expressed. Dataswift has been set up to create this framework. It is a provider of cloud-based infrastructure for the decentralised control, storage and exchange of personal data by individuals and companies. Individuals gain agency and legal ownership over their data, while companies can gain access permissions to the individual's PDA. Security and privacy features, including the need for the individual to grant access permissions in exchange for defined benefits, improve user trust and security over other solutions while also gaining for the company, consented access to more diverse datasets (as well as edge AI).

Data flows in and out of PDAs through application programming interfaces (APIs). Enabling this flow of data and providing mechanisms for trust and assurance is Dataswift One, an infrastructure for data transactions.

Individuals and companies transact via a contractual user interface (UI) through which individuals offer access to specific datasets within their PDA.

Rights and entitlements over data

The Dataswift infrastructure gives individuals property rights over their personal data held in their PDAs. They use their rights and control over these PDAs to create data transactions with companies who seek access to the data within. Through their effective control of their data and the contractual agreements by which they regulate access to their PDAs, individuals are thus given entitlements over the data that they generate. The creation of these entitlements gives them a form of property in their data.

Dataswift's technical and legal infrastructure thus enables the relevance of Coasian economics. As individual data generators are provided with a basis of entitlement, and therefore an ability to bargain over the uses of their data, economic optimality can be achieved.

[9] *Market Design for a Property Rights System with Entitlements for Individuals*. Ng, Irene C L, WMG Service Systems Research Group, Working Paper, Series Issue number: 01/21, ISSN: 2049-4297 November 2021

At the same time, Dataswift's governance infrastructure works to ensure that economic optimality does not come at the expense of more holistic considerations of value and dignity in personal data. Parties seeking to use the Dataswift One platform to transact with PDA owners for access to their data must meet Dataswift's governance standards. All potential transaction 'gateways' (applications and data plugs connecting to the PDA through APIs) are assessed and certified by Dataswift's governance team. A key purpose of Dataswift's governance structure is working to reduce information asymmetry between PDA owners and companies seeking access to their data; Dataswift ensures that these companies are open and transparent about the data they seek and what they want to use it for. Dataswift also has discretion to assess the reasonability and proportionality of the proposed data transaction.

The existence of data trusts will work to further reduce the effects of information asymmetry between PDA owners and data collectors.

The Dataswift Data Trust: Legal Basis

The trust property

In order for a trust to exist, there must be property capable of being settled into a trust. In National Provincial Bank v. Ainsworth [1965] 1 AC 1175, Lord Wilberforce provided the classic formulation of the requirements for the existence of a property right. A property right is *“definable, identifiable by third parties, capable in its nature of assumption by third parties, and have some degree of permanence or stability”.*

There are three potential loci of property rights in a Dataswift data trust:

1. The ‘namespaces’ of personal data held in the PDA;
2. The HMIC contracts (see below) that regulate access to those namespaces; and
3. The contracts that regulate access to the namespaces of ‘Contracted PDAs’.

Namespace as trust property

Personal data enters PDAs from all over the internet, and is held within the individual database of the PDA owner. This database is divided into ‘namespaces’ – folders within which data pertaining to specific data sources is held.

External parties seeking access to PDA data use ‘HAT Microserver Instruction Contracts’ (HMICs) to request access to this data from PDA owners. HMICs specify what namespaces they seek access to, and the specific fields of data within that namespace. Applying the Wilberforce test, we find:

1. *Definability*: Namespaces are defined fields of data within the PDA database. As such, they are definable.
2. *Identifiability*: Namespace data is necessarily identifiable. Specified namespaces and specific fields of data within them are the subject of HMICs. Different namespaces hold different datasets. In order to make the correct data transactions, these namespaces are distinguishable from one another.
3. *Capable of assumption*: A namespace itself cannot be ‘assumed’ by another. It sits within the PDA database and cannot be extracted. Without an HMIC, no access can be gained to it.
4. *Permanence or stability*: A namespace is a permanent component of a PDA.

Assessment: As a namespace is not capable of assumption in the absence of a HMIC, it is therefore incapable of being trust property.

HMIC as trust property

Companies that seek access to PDA data must do so via the mechanism of HMICs. Participating organisations offer HMICs to PDA owners. If the user accepts the proposed terms, then a contract between them exists.

HMICs typically specify:

- The namespace to which access is sought;
- The duration of the access;
- The consideration provided in exchange for the access;
- The specific fields of data sought from within the namespaces (only specified by certain types of HMIC).

Applying the Wilberforce test, we find:

1. *Definability*: An HMIC is definable as being the sum total of rights and entitlements conferred on the holder, pursuant to the agreement between the PDA owner and the party seeking access.
2. *Identifiability*: HMICs are identifiable – they exist between two specified parties, and have a reference number.
3. *Capable of assumption*: A 'standard' HMIC is between two parties, and only enables data flows between those two parties.
4. *Permanence or stability*: An HMIC has permanence and stability, since it continues to exist for a defined period of time until it expires.

Assessment: Standard HMICs are incapable of being property, as they are not assumable by other parties.

HMIC that pertains to 'Contracted PDAs' as trust property

Using standard HMICs, organisations must periodically re-contract with the PDA owner to renew their access to the namespace they seek. Organisations that wish to have access to specific namespaces without needing to renew their permissions with the PDA owner, or without needing the PDA owner to be online and using an app, can choose to issue 'Contracted PDAs' to users of their service. This means the organisation will have permanent access to the relevant namespace. The PDA owner is unable to alter the information held in the contracted PDA.

Applying the Wilberforce test, we find:

1. *Definability*: A Contracted PDA HMIC is definable as being the sum total of rights and entitlements conferred on the holder, pursuant to the agreement between the PDA owner and the party seeking to issue and hold contracted PDA rights over a specific namespace.
2. *Identifiability*: HMICs are identifiable – they exist between two specified parties, and have a reference number.
3. *Capable of assumption*: Contracted PDA access permissions enable the Contracted PDA contract holder to allow 3rd parties to gain access to the Contracted PDA namespace. As a result, we consider that this provides parties with the ability to assume the benefit of the contract – namely, permanent access to the namespace, including when the PDA owner is offline.
4. *Permanence or stability*: The HMIC has permanence and stability, since it continues to exist for a defined period of time until it expires.

Assessment: The benefits of Contracted PDA HMICs are capable of being settled into a trust, meeting the Wilberforce test to do so.

Type of trust

Given that the nature of the transactions between the data trust and data-collecting organisations cannot be fully anticipated, we consider that a discretionary trust is the best form of fiduciary entity for the proposed data trust. Within a discretionary trust, trustees have discretion on when to distribute trust income and capital. We consider this flexibility will enable the Data Trust Company to find the best transactions on behalf of the PDA owner.

Other forms of trust would be inappropriate. As an example, a bare trust would not work, as this form of data trust requires a trustee who is empowered to make decisions on potential data transactions that would be in the best interest of the beneficiaries.

Settling the trust

The Data Trust Company (DTC) and the individual data generator create trust property by mutually agreeing an HMIC for the creation of a Contracted PDA. This enables the DTC to have permanent access to the relevant Contracted PDA Namespace.

The data trust is created when the DTC settles the benefit of the Contracted PDA HMIC into a trust. It identifies the beneficiary as the PDA owner and holds itself to be the trustee. The DTC outlines this via a trust deed.

Documenting the trust

The trust deed is created by the DTC as it creates the trust. It sends a copy to the PDA owner, who is the beneficiary.

Ending the trust

As there is only one beneficiary, the beneficiary can end the trust under the rule in Saunders vs Vautier [1841] EWHC J82, 4 Beav 115.

Trustee or agent?

This paper has outlined the mechanism by which a bottom-up data trust may be created. However, the creation of contractually-based 'data agent' relationships may be another mechanism that enables data subjects to have their data used in their best interests. Data subjects could contract with data agents in order to let the former transact with data on behalf of the latter. However, this approach would fail to activate the higher standard of fiduciary obligations, or the flexibility inherent in discretionary trust approaches. By imposing fiduciary responsibilities on the management of personal data, data trusts have been seen as reducing scope for the principal-agent problem, thus ensuring greater alignment of intent between data generators/ PDA owners and the DTC[10].

[10] *Independent Assessment of the Open Data Institute's Work on Data Trusts and on the Concept of Data Trusts Report to the Open Data Institute*. Godel, Moritz; Natraj, Ashwini. April 2020, London Economics

The Dataswift Data Trust: Role in the Market

We propose a generator-centric data trust over social media data. Individual persons are able to participate through becoming PDA owners. As PDA owners, they are then able to agree to HMICs and so create the basis of the trust. Rather than pooling the personal data of individual data generators into one data trust entity, our proposed mechanism provides access to the data of multiple, separate, individual PDA owners.

The Data Trust Company

The Data Trust Company (DTC) offers third parties access to the data provided in the individual Contracted PDA HMIC it holds as trustee. To highlight the specific nature of this data access mechanism, we name this the 'Aggregated Access Rights Data Trust' (AAR Data Trust).

Because the DTC is a trustee of multiple data trusts for the same type of dataset, it uses the benefits of scale to negotiate data access deals. This will achieve better outcomes for individual data generators than if they each negotiated separately[11].

As a specialist DTC the AAR Data Trust can use its knowledge of the market to find and assess data transactions that would be in the beneficiaries' best interests, achieving better outcomes than those the individual consumers can achieve.[12] At the same time, the trust can lower the transaction costs for acquiring the data held within the AAR Data Trust[13].

Who is the Data Trust Company?

As an entity, the DTC will be a limited liability company incorporated under the laws of England and Wales, likely being limited by guarantee or incorporated as a CIC (Community Interest Company) or CIO (Charitable Incorporated Organisation). A broader question arises as to the persona of this entity. Who would incorporate it, and who would be the managers?

One avenue of exploration would be data trusts established by retail banks. As banks adapt further to Open Banking, and assess their strategic position in light of fintech competition and the rising challenge posed by big tech market entrants, data trusts may be additional mechanisms to increase the prominence of their customer brands and data access, positioning the bank as trusted custodian of data assets as well as financial assets. In the

[11] *Who Owns the Future? Data Trusts, Data Commons, and the Future of Data Ownership*, Mills, Stuart, (September 24, 2019)

[12] *Designing Data Trusts Why We Need to Test Consumer Data Trusts Now*, Blankertz, Aline. Feb 2020. Stiftung Neue Verantwortung

[13] *Independent Assessment of he Open Data Institute's Work on Data Trusts and on the Concept of Data Trusts Report to the Open Data Institute*, Godel, Moritz; Natraj, Ashwini. April 2020, London Economics

case of the data assets however, if the trust is based on the Dataswift infrastructure – where the data in trust also continues to reside within and under the control of the individual customer – the bank's trust will be acting as a data broker rather than a simple trusted repository. This would make sense for markets where the aggregation of verifiable personal data can provide a better return to individuals than is available to them making personal contracts with data users (for example in exchange for merchant perks).

Energy data trusts are another potential use case. In the case of the energy data trust – focused for example on insulation and micro-grids – the aggregated data could be used for trading energy allocation based on demand anticipation using consumer personal data, or to size a communities' contribution to decarbonisation, with rewards then being shared in proportion to individual subscribers' contributions.

Risk and remuneration

The DTC would be remunerated for its activities. The nature of this remuneration remains unknown. With data transacted in exchange for a defined benefit by the data collector, it is likely that the data trust would itself hold a share of the consideration offered to the PDA owners/beneficiaries. Depending on the nature of the consideration, this could in turn be re-exchanged on the market for financial benefit to the DTC.

Aggregation

The AAR Data Trust aggregates access to the personal data of individual PDA owners, and offers this to the market. The AAR Data Trust is managed by a single trust company which acts as trustee for the multiple individual trusts over HMIC contracts that together enable the functionality of the AAR Data Trust.

The AAR Data Trust's trust company – the Data Trust Company (DTC) – acts as the single point of contact for all parties that want access to the data held within the Contracted Namespace. At the same time, the DTC has a separate and individual relationship with each PDA owner, each of whom is the sole beneficiary of the individual trust created over their HMIC.

Attracting data collectors

The AAR Data Trust will attract data collectors through marketing and the pursuit of selected partnerships. Methods of outreach will vary by market, ie. by data type and use case.

It is to be anticipated that AAR Data Trusts built on Dataswift's PDA infrastructure may in many cases be scoped and defined in collaboration with data-collecting organisations, who could identify the data trust as a useful means of responsibly acquiring consented data on individuals.

In an example use case of Facebook data, academic organisations who want access to the data of individual Facebook users for research purposes could be sought. This could be established readily as a proof of concept demonstrator using Dataswift's Facebook data plug, subject to its terms.

Negotiating with data collectors

The DTC will be empowered to set the rates of access to the data held within the individual PDAs, having at all times the best interests of the PDA owner in mind. It will be left to the discretion of the DTC how best to value the data access, set the form of consideration with which to exchange the data, and consider bespoke transactions with data collectors.

Use case: social media

The AAR data trust will first be trialled using social media data, specifically Facebook data brought into PDAs via Dataswift's Data Plug technology.

Initial data collectors may be found among Dataswift's academic partners, the 'Hub-Of-All Things' (HAT) academic network of researchers from 6 UK universities with whom Dataswift's HAT Microserver technology was first developed, with Research Councils (UK) (RCUK) funding.

Conclusion

The generation, collection and analysis of data continues to redefine business models, reshape industries, and impact individual citizens. At present, individuals are disempowered, lacking both the tools to properly control their interactions with the data ecosystems around them, and the specialised knowledge to maximise their opportunities. As societies grapple with questions of data governance, digital equity, and maintaining individual dignity and autonomy online, new solutions must be developed.

The discourse on data trusts will undoubtedly continue to evolve, and it is to be expected that a variety of forms will emerge. However, in the absence of any mechanisms for creating property rights in personal data, we anticipate that most ‘data trusts’ will be data collector-centric data trusts composed of companies and other institutions pooling data between themselves, with or without the consent of the individual data generators. Most of these will be unable to comprise a formal trust under English trust law.

For ‘bottom-up trusts’ over personal data, the problem is acute. Without a mechanism for creating property claims of individuals over data, individual data subjects cannot participate in such an entity. However, such a mechanism can be provided through use of Dataswift’s PDAs and data exchange infrastructure. Together, these tools enable individual autonomy online and a way to delegate that autonomy to responsible managers. It now must be seen if these tools can attract a wider following.

Appendix

Key terms

Data generators: These are individuals who, through their interactions with a digitally enabled system, generate data. This can occur in a wide variety of contexts, from utilising online platforms to interacting with physical sensors.

Data collectors: These are the organisations that create digitally enabled systems for use by data generators, and then collect the data that the generators create.

Settlor: A person who owns property that they then settle into a trust, to be held for the benefit of one or more beneficiaries.

PDA: Personal Data Account – a space in the cloud within the HAT Microserver, legally owned and controlled by the individual.

Namespace: Personal data enters PDAs from all over the internet, and is held within the individual postgreSQL database of the PDA owner. This database is divided up into 'namespaces' – folders within which data pertaining to specific data sources is held. More technically, namespace refers to an alphanumeric attribute of a data record used for addressing data within a server database, either directly via an API endpoint or using Data Debits.

HMIC: An abbreviation for HAT Microserver Instruction Contract. These contracts are agreed between PDA owners and entities that seek access to the data held within the PDA. HMICs specify what namespaces they seek access to, and for some contracts, specific fields of data within that namespace.

Contracted PDA: A type of PDA wherein a specific namespace has different access rules than those found in standing PDAs. A contracted PDA namespace enables the company who has issued the PDA to have permanent access to that namespace. The individual PDA owner cannot alter the data within the Contracted PDA Namespace, or cancel the Contract PDA HMIC without the consent of the Contracted PDA issuer.