A Design Framework for a Blockchain-Based Open Market Platform of Enriched Card-Based Transactional Data for Big Data Analytics and Open Banking

Trevor Toy[a] and Josef Langerman[b]
Academy of Computer Science and Software Engineering, University of Johannesburg, Auckland Park, Johannesburg, South Africa

Keywords: Blockchain, Big Data, Open Banking, Data Markets, Transactional Data, Cloud Data, Data Analytics, Platform Design, Personal Data Management, Data Economy.

Abstract: Around a quarter of the world's data is generated by financial institutions. The Capgemini 2022 World Payments Report predicts a 28% increase in transaction volumes from 2021 to 2026, to an estimated total of 2.122 trillion global non-cash transactions. There is a growing demand for accessible transactional data for analytical purposes and to support the rapid global adoption of Open Banking. Open banking is a collaborative business model involving customer-authorised transactional data sharing with other unaffiliated parties to allow for enhanced service and product offerings to the marketplace. This research explores utilising distributed ledger technology to facilitate the market mechanism of securely sharing data through an integrated and decentralised platform that conforms to the expected regulatory and compliance standards of the financial industry from which the data is generated. Scalable and accessible access is a core requirement of a marketplace platform for its data consumers and producers. To enable customer-authorised transactional data sharing, an incentive mechanism is proposed, which includes the data subject in the process to empower them to control access and earn money from the related transactional data that they generate. A proposed framework is defined for the development of a marketplace platform that can ultimately support the growth, prosperity and development of economies, businesses, communities and individuals, by providing accessible and relevant transactional data for big data analytics and open banking.

1 INTRODUCTION

The global trend of leveraging data to evolve businesses towards digitisation is moving from a phase of simply generating, storing and processing data to now being able to extract the real business value from it that was initially promised to many business stakeholders (Early Adopter Research, 2019). Data marketplaces are a pivotal solution by providing external supplementary analytical capabilities to enhance the extraction of value from existing internal data. Businesses realise that external data providers offer significant competitive advantages that would not be possible by only looking purely at their internal data. These additional insights can include gauging industry benchmarks, regional trends and unidentified market opportunities. The accessibility and capabilities to procure and apply external data are still challenges for most organisations seeking supplementary external data. Two-thirds of companies must use external service providers or specialist consultants to support their data-sourcing requirements (Belissent, 2020).

Several open data markets exist, with many utilising distributed ledger technology such as the Datapace and ArcBlock distributed ledger platforms (Arcblock, n.d.; *Datapace - Data Marketplace Powered by Blockchain*, n.d.). However, none are domain-specific and can guarantee the industry compliance requirements necessary for the finance sector. Moreover, none have yet developed the necessary scale and trust to consolidate the sector to be considered as a standardised Big Data repository

[a] https://orcid.org/0009-0002-3300-393X
[b] https://orcid.org/0000-0003-1984-0205

Toy, T. and Langerman, J.
A Design Framework for a Blockchain-Based Open Market Platform of Enriched Card-Based Transactional Data for Big Data Analytics and Open Banking.
DOI: 10.5220/0012143400003541
In *Proceedings of the 12th International Conference on Data Science, Technology and Applications (DATA 2023)*, pages 583-589
ISBN: 978-989-758-664-4; ISSN: 2184-285X

for card-based transactions (Hassani et al., 2018). This research explores utilising distributed ledger technology to transparently and immutably generate exchange agreements to facilitate the transactional interactions of participants in a marketplace platform. The proposed solution leverages the growing interoperability of cloud-based data products and services, blockchain and other modern system technologies to demonstrate an end-to-end market trade process for transactional card data.

An opportunity exists for a solution that provides an accessible and scalable market platform specifically for trading card-based transaction data and the related peripheral data to those transactions to stimulate cross-business data sharing desired in the Open Banking economy. In predominately non-cash-based environments, card transactions are the best representative of multiple levels of economic activity. They provide economic and business insight into the nature and volume of spend at the individual, demographic and state levels. Companies increasingly demand external data sources to supplement and broaden the context of their existing internal analytical scope (Belissent, 2020). This application can also extend beyond commercial business use. Banco Bilbao Vizcaya Argentaria (BBVA) utilised its data analytics capabilities by applying over four million anonymised credit card transactions to help city planners create strategies to stimulate economic growth. Governments could also use this data to assist in deciding where to deploy aid responses after natural disasters (Wixom & Farrell, 2019). Another example is when researchers applied retail payment data to forecast economic activity in Italy. As a result, they could show a close correlation between retail payment series and macroeconomic aggregates to demonstrate how data improved forecast accuracy of not only gross domestic product (GDP) but also overall consumption, investment, and value-added services in specific sectors (Aprigliano et al., 2019).

The primary aim of this research is to construct a system framework for a marketplace platform called the Big Transactional Data Marketplace (BTDM), which will provide the conceptual design foundation for a decentralised, blockchain-based open market platform of enriched card-based transactional data for big data analytics and open banking

The research questions for this study were formulated to contribute to the knowledge base of transactional data markets for big data analytics and open banking from the experience and insights gained in developing this solution. Below are the four research questions identified for the BTDM derived from the initial DSR problem explication process and subsequent solution objectives defined to address those problems.

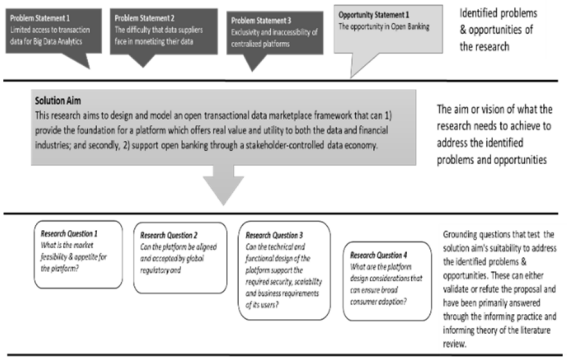

Figure 1: Research questions for the BTDM Framework.

The expected output from this research is to demonstrate the potential for a scalable card-based transactional data marketplace platform. The study would achieve this by designing and modelling a system framework that can 1.) provide the foundation for a system design that offers real value and utility to the data and financial industries and 2) support open banking through a stakeholder-controlled data economy.

2 METHOD

Design Science Research (DSR) was the selected methodology to develop the BTDM Framework. The BTDM framework guides a particular system design; therefore, various design abstraction levels that realise a deployable and functional product were considered as ideal DSR artefacts for the framework (Dresch et al., n.d.). The methodology also focuses on the iterative refinement of rigour and relevance of the solution against the existing knowledge base and the real-world application domain (Hevner, 2007).

The output artefacts required by the DSR process were presented as a set of traditional system design models, categorised within the phases of Nunamaker et al.'s System Development Research Process (Nunamaker et al., 1990). These descend in their levels of abstraction towards an implementable prototype.

- Conceptual framework
- System architecture
- Functional design (Analyse and design the system)
- System development (Build the [prototype] system)

A computer simulation/lab experiment approach was selected as the evaluation method for the DSR process according to the DSR Evaluation Method Selection Framework prescribed by Venable et al. (Venable et al., 2012). Therefore a proof-of-concept (PoC) blockchain contract was implemented to demonstrate the core functionality of the BTDM platform according to plans from subsequent system design models of the System Development Research Process. This PoC verified the process-level logic of the framework and showed that blockchain technology could successfully implement the user requirements necessary to achieve the research objectives.

The PoC demonstrated the ability to register a user linked to an Ethereum account from which they could send and receive cryptocurrency to transact on the platform. It also showed that we could create records on the blockchain that enable data producers to list their available datasets and consumers to find and subscribe to those datasets. In addition, it can act as a progressive facilitator of open banking by enabling data subjects to manage their data and any related identifiable data through the platform through a concept of "permissioning" of the current designations.

2.1 Results & Discussion

The solution design and context diagram are primary artefacts of the framework. The PoC conducted bridges the conceptual plans into code functions that can be field tested against the platform's objective and practical user requirements, reaffirming these artefacts' relevance and rigour per the DSR methodology.

2.1.1 BTDM Solution Design

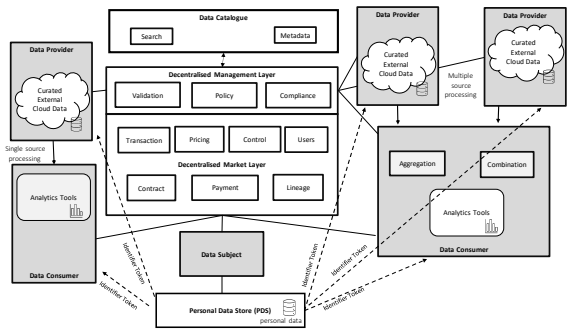

Figure 2: The BTDM Solution Design.

The BTDM Solution Design is an extension of the Collective Multilateral Market Design set out by Koutroumpis et al. in their paper, *The (Unfulfilled) Potential of Data Marketplaces* (Koutroumpis et al., 2017). The BTDM design makes a distinction between the market layer and management layer in recognition of the need for moderation of the platform to maintain quality and control that could be lost from complete autonomy. It also introduces the Data Subject role and details the interaction between the resources providing and consuming the data products. The core market layer of the platform is provided within a blockchain contract ecosystem which connects the data providers and their data subjects to the data consumers. This market layer manages the transaction, user, pricing, payment, tagging, contract, control, and lineage features that facilitate the primary user interactions on the platform for trading data. Secondly, a management layer that enforces the validation, policy and compliance requirements needed to maintain the quality and sustainability of the platform.

A data catalogue allows for efficient search functionality and metadata management for data providers and their data. It will assist in presenting the data product available to the consumer and be integrated into the decentralised layer to complete the transaction between the parties once the product is selected.

The data subject utilises a Personal Data Store to manage and share their data with the consumer. Data providers list and sell their data through the platform but will supply it directly to the consumer. Data subjects will enter into a contract with data providers to supply transactional data related to each data subject to an interested consumer. That data supplied by the data provider is anonymised in that no one with that dataset should know to whom it pertains. Because the data subject must remain anonymous on the platform, they will reveal their identity to the consumer as part of their own contractual arrangement. This identity will then be related back to the dataset from the data provider. So they will "sell" their identity to that consumer and any other personal information they are willing to disclose.

Interested consumers identify data subjects in two ways, either through a basic demographics profile that allows them to target a general profile—for example, analytical marketing research (e.g. spending habits for middle-income white males over 50). Or directly as clients of the data providers, for instance, for customised products or service offerings from the data providers' business partners, such as life insurance or medical aid providers. They would then enter into a subscription contract with the producer to gain access to the data. A contract agreement would

be created on the blockchain along with the payment processing.

If a data consumer wanted transactional and peripheral data on a specific data subject, they would subscribe to an identifiable dataset provided by the data producer in addition to the related data subject's data provided separately by the data subject. The data subject would be paid accordingly for the personal data provided (a data subject could choose to disclose other personal details such as IoT or other sensitive demographic data). The data consumer would enter into another subscription contract with the data subject in parallel to the identifiable dataset.

Data producers would be required to house their data within accessible but secure modern commercial cloud-based repositories, where they can curate the data and control access granted to the data consumers. The data must be formatted in an efficient and standardised format (such as Parquet). The data must align with ISO 20022 definition standards for transactional card payment data (*ISO 20022 | ISO20022*, n.d.). This research investigates the ISO standards requirements in its application in BTDM andcompliance with the Payment Card Industry Data Security Standards (PCI/DSS) concerning the market trade of card-based transactional data (PCI Security Standard Council, n.d., 2008).

2.1.2 The BTDM Context Diagram

The core functions evaluated for the BTDM were based on the platform's ability to allow users to buy and sell transactional data securely and enable cardholders (data subjects) to manage and benefit from the sale of their data. The platform creates secured transactional contracts between these entities as platform users to establish these relationships between the data provider, the data consumer and the data subject. The following key business objects have been defined for the BTDM Framework to support the user interactions that facilitate the ability to make secure contractual transactions between each user role.

The interactions of these business objects are described in the following BTDM context diagram.

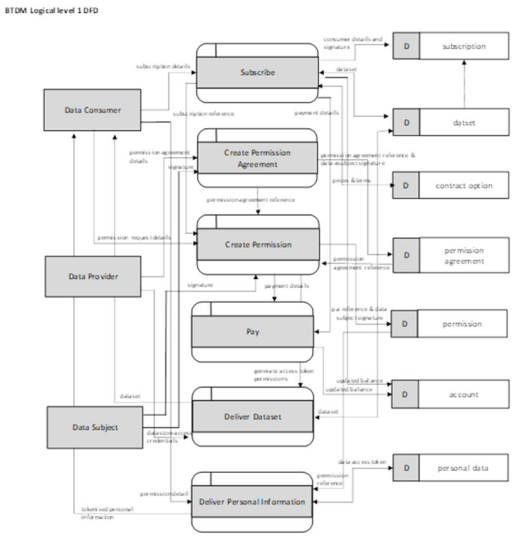

Figure 3: The BTDM Context Diagram.

1. Account – Each platform user is registered with their private blockchain account. The account comes from the selected blockchain on which the platform is based and is used to facilitate payments to and from the holder.
2. Subscription – A contract between a data provider and a data consumer to supply/access transactional and transaction-related data.
3. Permission Agreement – A contract between a data provider of transactional data and the data subject to allow the data provider to supply the transaction details generated by that data subject to a requesting data consumer.
4. Permission – A contract between a data subject and a data consumer that allows a data consumer to access identifying data of the data subject with which they can associate transactional data from a subscription.
5. Contract Option – The contract option defines the subscription's unique payment plan and product delivery details. The data provider will predefine the contract options available to each dataset, which the user will select when subscribing.

2.1.3 The BTDM Proof-of-Concept

A proof-of-concept was deployed as the evaluation activity of the DRS process. It aimed to evaluate whether the conceptual design artefacts could be actualised on the proposed technologies. The core functionalities of the BTDM platform were implemented on an Etheruem blockchain contract to determine whether the platform's required transactional, security and accessibility requirements

could be met. The following core functions were tested:

- User Registration & Functional Roles
- Transaction handling
- Data Catalogue
- Data Subject Permission Management (Open Banking)
- Data Subscription & Contract Management

Data Subject Permission Management

This article will focus on Data Subject Permission Management, as it is one of the key functionalities of the BTDM proof-of-concept. Data Subject Permission Management differentiates the platform from other data marketplaces by utilising a concept of "permissioning", a contractual agreement implemented on the blockchain to allow a data consumer to provide information on their identity to a data consumer. This information of the data subject can then be associated with a data set of card-transaction data supplied by their card provider or financial institution. Note that a separate permission agreement must first be in place between the data subject and their card provider (as a data provider) to authorise the generation of identifiable datasets to the market. Identifiable datasets on the platform are curated by the data provider and related to a specific data subject but don't have any revealing information about who that data subject is. Because these involve the data subject, a commission fee is charged to compensate the data subject.

Once the permission agreement is in place, a consumer can generate a permission request for a data subject to reveal their identity to associate against an identifiable dataset. Consumers will identify a data subject on the BTDM they are interested in from their anonymous demographic profile via the data provider or directly from the data subject catalogue.

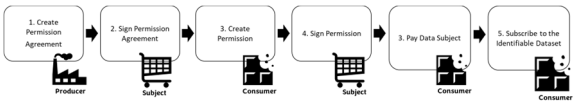

Figure 4: Permission management flow.

The permission must record the addresses of requesting data consumer and the data subject. The data provider is not directly involved, but their address is part of the permission agreement associated with this permission request. The permission also uses a status indicator to manage the changing states of the permission from being created, signed, and expired or cancelled.

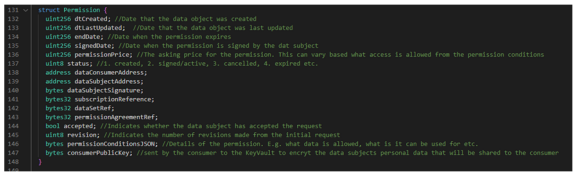

Figure 5: Permission data object structure.

A permission also represents a transaction between the data consumer and the data subject. It not only allows the consumer to associate the identity of that data subject with the identifiable dataset generated by the data provider but also allows the consumer to access additional detail about the data subject. In the data catalogue, a data subject can list various personal attributes available "for sale" that they themselves maintain in their Personal Data Store, such as race, family, income, profession, IoT data (e.g. fitness band, GPS tracking), browsing & search data etc. These attributes could be listed, each with their own individual prices, which a consumer could purchase and the total captured in the permission price. The details of each of the selected personal data request items and their prices are captured in a permission conditions json, like an invoice, which can be validated on the data subject's side to confirm the total before accepting the request. The accepted indicator shows whether the data subject has accepted or rejected the request. Each time a request is revised, it will also be logged on the blockchain.

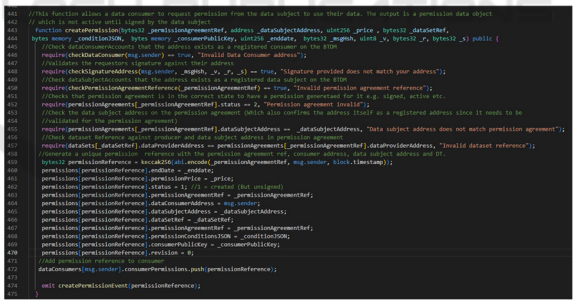

Figure 6: BTDM createPermission function.

Once a permission request is generated, it is sent to a data subject for review. The consumer indicates what personal data attributes they wish to access from the subject's listing. The consumer should also provide details on their purpose for the data and how they plan to manage it (retention, distribution, outcome, etc.). This detail is captured in the permissionConditionsJSON and can be considered a binding agreement between the parties. The data subject can also reject or revise the agreement, in which case it will be returned to the consumer for acceptance.

To conclude the permission process, the data consumer must pay the data subject for the requested data. The PoC also provides an example of Transaction Handling to demonstrate this functionality. The money is then transferred to the data subject 's account, and the permission status is updated. This permission's status is what is checked when a data consumer subscribes to an identifiable dataset where they will need to submit this permission reference. The data consumer is now obliged to provide access to the data consumer through their PDS. The inputs to the PDS are obtained from the permission on what specific data attributes have been purchased by the consumer and the consumer's public encryption key to encrypt the data before exposing it to the consumer.

Now that the permission is in place for the targeted dataset the consumer wishes to subscribe to, the data consumer will proceed with subscribing to the identifiable data from the data provider. The permission would be validated for the data provider to ensure it has been authorised by the data subject for that data consumer before allowing the data consumer to subscribe. If all is successful, the subscription reference is added to the permission since the subscription can only be generated after the signed permission is in place. These interactions are all recorded on the immutable public ledger and form the contractual basis on which all the parties can proceed with the actions of the trade.

2.1.4 Blockchain as an Effective Technology for a Data Marketplace

The PoC offered positive results, proving that blockchain technology could be a successful foundation for a decentralised market platform. The following points are the key observations taken from the PoC.

1. Blockchain offers a scalable and accessible platform for developing a cross-border solution because of its decentralised design, easy user profile integration to blockchain wallets and accounts, and use of non-fiat digital currencies for transactions.
2. It provides an accessible structure for user management which is easily integrated into the built-in account features of the technology.
3. Secure authentication and authorisation can be applied through signature verification of accounts and contract code that can enforce logical conditions (smart contracts).
4. A degree of manual verification of data consumers wishing to access "identifiable data" related to a data subject is still required. Without some assurance of the consumer's intent, a data subject will not be inclined to sell their data.
5. The data subject's profile remains anonymous and secured on the platform. Once they provide permission to a consumer, that consumer can match the blockchain address to the profile of the data subject and their transactions from the data producer subscription. Once that subscription expires, the customer cannot match any subsequent transactions from that data producer without requesting a new permission from the data subject. Any static identifying data that the data consumer receives will always be known, but any dynamic data (E.g. IoT data, like fitness trackers) will stop after the contractual end date.
6. For complex and dynamic data sharing of a data subject's personal data, the Personal Data Stores (PDS) that the solution proposes are still being developed. It does not appear that any suitable solutions currently exist yet. But for basic identifying data sharing, public-private key encryption offers a feasible mechanism.
7. The PoC facilitated an initial payment on creating a new subscription; however, scheduled payment collection needs further exploration. Scheduled collections were partly handled by incorporating a billing plan data element in the subscriptions.
8. The concept of peripheral transaction data also needs further exploration. Merchants of the sale could contribute by providing detail of each transaction (E.g. the receipt of what was purchased). The proposal also establishes a permission agreement model between the primary data provider (providing transaction data) and the peripheral one. The proposed solution was to allow the primary data provider to be a consumer of the peripheral data to offer enhanced transaction data to the end consumer. This concept was not demonstrated in the PoC, and the evaluation revealed more complexity than was initially understood.

3 CONCLUSION

The results from research into the BTDM show potential for a scalable platform to address the solution aim that was defined – to design and model an open transactional data marketplace framework that can 1.) provide the foundation for a platform which offers real value and utility to both the data and financial industries; and secondly, 2) support open banking through a stakeholder-controlled data economy.

The PoC applied for evaluating the artefacts was ideal for this initial proposal for the BTDM. They proved that a feasible technical implementation of the design concepts could be realised. However, further development and evaluation are required since the BTDM is a heavily socio-technical system operated by many stakeholders and involves numerous interactive processes in bringing the complete solution together. Methods such as action research, field testing and focus groups on the completed platform would provide more rigour in testing end-to-end processes and relevance as the solution moves into a naturalistic state with more objective influences on the solution's outcome and relevance.

The immaturity of personal data stores, volatility of cryptocurrency, universal protocols for external access to data storage, and balancing moderation vs accessibility are some of the limitations identified for the BTDM. None of these limitations is considered severe enough to prohibit a version of this platfrom from being developed in today's business landscape. The expectation from the analysis of this research indicates that the technical challenges can be mitigated, and the relative social and regulatory challenges will subside over time to allow for broad adoption by the data and financial industries.

Lastly, as a digital platform, much of its success relies on broad adoption and sustainable usage to achieve the desired network effects and prevent disintermediation. These outcomes are greatly influenced by business factors outside the system design, such as strategic positioning and policies, the competitive landscape, changing legislation and regulation, and even the appropriate marketing strategy. These must be considered as the BTDM develops from a conceptual to a commercial product offering.

REFERENCES

Aprigliano, V., Ardizzi, G., & Monteforte, L. (2019). Using payment system data to forecast economic activity. *International Journal of Central Banking*, 15(4), 55–80.

Arcblock. (n.d.). Data Marketplace and Blockchain. Retrieved March 17, 2021, from https://www.arcblock.io/en/blockchain-data-marketplace

Belissent, J. (2020). The Insights Professional ' s Guide To External Data Sourcing The Insights Professional ' s Guide To External Data Sourcing. In *Forrester*.

Brodsky, L., & Oakes, L. (2017). *Data sharing and open banking. July*, 16–23. https://www.mckinsey.it/sites/default/files/data-sharing-and-open-banking.pdf

Datapace - Data Marketplace Powered by Blockchain. (n.d.). Datapace . Retrieved March 17, 2021, from https://datapace.io/

Dresch, A., Daniel, ·, Lacerda, P., Antônio, J., & Antunes, V. (n.d.). *Design Science Research A Method for Science and Technology Advancement*.

Early Adopter Research. (2019). The Transformative Impact of the Data Marketplace. In *Early Adopter Research*.

Hassani, H., Huang, X., & Silva, E. (2018). Banking with blockchain-ed big data. *Journal of Management Analytics*, 5(4), 256–275. https://doi.org/10.1080/23270012.2018.1528900

Hevner, A. R. (2007). A Three Cycle View of Design Science Research. In *Scandinavian Journal of Information Systems* (Vol. 19, Issue 2).

ISO 20022 | ISO20022. (n.d.). Retrieved March 26, 2021, from https://www.iso20022.org/

Koutroumpis, P., Leiponen, A., & Thomas, L. D. W. (2017). The (Unfulfilled) Potential of Data Marketplaces. *ETLA Working Papers*, 2420(53). http://pub.etla.fi/ETLA-Working-Papers-53.pdf%0Ahttp://pub.etla.fi/ETLA-Working-Papers-53.pd

Nunamaker, J. F., Chen, M., & Purdin, T. D. M. (1990). Systems development in information systems research. *Journal of Management Information Systems*, 7(3), 89–106. https://doi.org/10.1080/07421222.1990.11517898

PCI Security Standard Council. (n.d.). *Official PCI Security Standards Council Site - Verify PCI Compliance, Download Data Security and Credit Card Security Standards*. Retrieved December 28, 2021, from https://www.pcisecuritystandards.org/pci_security/glossary#P

PCI Security Standard Council. (2008). *At a Glance PCI Data Storage PCI Data Storage Do's and Don'ts*. www.pcisecuritystandards.org/pdfs/

Venable, J. R., Pries-heje, J., & Baskerville, R. (2012). A Comprehensive Framework for Evaluation in Design Science Research. *Design Science Research in Information Systems. Advances in Theory and Practice*, 7286(May), 423–438. https://doi.org/10.1007/978-3-642-29863-9

Winship, T., Mistry, D., & Colceriu, D. (2016). Banking Data Monetization. Temenos.

Wixom, B., & Farrell, K. (2019). Building Data Monetization Capabilities that Pay Off. *Mit Cisr*, 19(11), 1–8.

World Payments Report 2022 | Research & insight | Capgemini. (n.d.). Retrieved May 30, 2023, from https://www.capgemini.com/insights/research-library/world-payments-report/