Plan Overview

A Data Management Plan created using DMPonline

Title: Testing User Experiences of the Good With Financial Health App

Creator: Rory Spanton

Principal Investigator: Rory Spanton

Data Manager: Rory Spanton

Project Administrator: Rory Spanton

Affiliation: University of Plymouth

Template: DCC Template

ORCID iD: 0000-0002-4050-2166

Project abstract:

It is clear from recent research that psychological variables can both influence and be influenced by financial outcomes. However, widely used credit-scoring methodologies focus exclusively on an individual’s financial records and credit history, neglecting their psychological attitudes. This is a problem for both lenders and customers. Competent customers are often denied credit because of a lack of credit history, and lenders lose out on these hidden prospects, or risk lending to individuals whose financial history belies a lack of psychological readiness to borrow. Good With aims to provide an interface for individuals to provide psychometric and banking data to generate Readiness ScoresTM - holistic scores indicating their current levels of financial capability. Individuals can also receive psychoeducation about financial topics and their intersection with mental health through this service. This project explores user experiences of using the Good With app to receive a financial Readiness ScoreTM, and tests whether the Good With Readiness Score can predict psychological and behavioural factors relating to credit risk and repayment.

ID: 156684

Start date: 15-08-2024

End date: 15-08-2025

Last modified: 05-09-2024

Testing User Experiences of the Good With Financial Health App

Data Collection

What data will you collect or create?

Participants will give their date of birth, gender, postcode, and employment status using a web form. In this form, they will also have the option to report their sexual orientation and employment status, although participants do not have to provide this information to participate in the study. All of these demographic variables will be used for score validation purposes, specifically to monitor any bias in the Readiness Score. Participants will also connect their primary bank account and respond to psychometric questionnaires (see accompanying document for copies of all questionnaires) on the Good With web app.

We are using TrueLayer as our Technical Service Provider (TSP) and Account Information Service Provider (AISP). TrueLayer is regulated by the FCA (FCA ref: 901996) and is authorised to act as an intermediary between customers and banks to share customers' banking information with third parties like Good With Ltd (https://register.fca.org.uk/s/firm?id=0014G00002UxKoUQAV). Open banking data is collected via TrueLayer. Users authorise access to accounts and cards individually and maintain full control over which accounts they choose to share with Good With. Specifically, data acquired from TrueLayer includes:

- Account details, including name, account number, and sort code
- Debit and Credit transactions for each authorised account
- Account balances
- Transaction descriptions e.g. 'Sainsbury's Petrol 270122'
- Transaction classification, e.g. 'Shopping, Car'
- Timestamp

This data is read-only. Good With cannot cannot alter the data in a way that can cause monetary consequences for participants. No third parties (including smartphone companies) will be able to access participant data.

All raw data from the Good With web app will be stored in a Google Cloud Firestore database, in NoSQL format. Processed data (that which has undergone statistical analysis or transformation) will be stored across the same Google Cloud Firestore database and a Google BigQuery database.

Data from qualitative user interviews will be stored in audio recordings with common file formats (.wav, .mp3). Processed data from the thematic analysis of these recordings will be stored in .docx files. These files will be stored in password-protected Google Drive folders that are only accessible to the research team. Consent forms will be administered by a web form, with results also stored in the above Google Drive folders.

How will the data be collected or created?

Data will be collected using the Good With web app running on lab computers. Participants will enter data and respond to questions in the web app using the mouse and keyboard, and the data will be written to a Google Cloud Firestore database.

Firestore databases are structured into nested collections and documents. User personal data will be

stored in the document 'users/{user_identifier}/'. Psychometric answers will be stored in the document 'users/{user_identifier}/psychometric-answers/{psychometric_test}'. Banking data will be stored in nested documents and collections inside 'banking/{user_identifier}/'. Field names inside each Firestore document are named in camelCase, with collection names in dash-case and document names in snake_case.

Processed data will be stored in Google BigQuery. These data will be stored in SQL tables that are themselves nested within datasets. For instance, raw and processed psychometric data will be stored in different tables inside a ‘psychometrics’ dataset. However, banking data will be stored in tables inside a different ‘bank_transactions’ dataset. All dataset, table, and column names will be named in snake_case.

For quality control, we will implement various rules to ensure participants complete the psychometrics in a valid manner. For instance, we will require participants to respond to each questionnaire item, and will implement an attention check to test whether participants are carefully considering written instructions. We will also require participants to reconsider their responses if we detect they are responding according to a pattern that signifies a lack of engagement with the instructions (for instance, responding in a straight line down the matrix of questionnaire items). We will also require users to authenticate via their bank accounts when giving their Open Banking data, to ensure the data being passed through is their own. Users will be directed to and from bank authentication screens via our web application.

Documentation and Metadata

What documentation and metadata will accompany the data?

Due to the sensitive nature of the data being collected and its commercial value to the funder, the data from this study will not be accessible to secondary users. The research team will maintain internal documentation that contains contextual information about the data, how it was collected, definitions of key variables, assumptions, units of measurement and analytical information. All analysis code will include inline comments that detail data cleaning and analysis steps in plain language.

Ethics and Legal Compliance

How will you manage any ethical issues?

Participants will give fully informed consent to sharing their data, and for their data to be preserved. They will do this by reading a brief and consenting prior to the start of the study, and by reading and accepting consent agreements in the Good With web app. Participants will also be able to withdraw from the study at any time without penalty and will be provided with the FoH ethics administrator’s details in the event that they have ethical concerns about the study.

As previously stated, we must keep a record in Cloud Firestore of participant names against their unique identifiers. However, in all other storage locations, all participants’ data will be associated with a randomly assigned user reference code only. This code will identify the participant in all subsequent data analyses. All databases will require secure authentication, and sensitive data (e.g. participant identifying information, demographics, and banking data) will never be copied outside of these databases.

How will you manage copyright and Intellectual Property Rights (IPR) issues?

Good With Limited are the sole owners of all data collected, created and otherwise processed through the course of this project. No data will be shared with third parties during the course of this project, however models derived from data collected in this research may form part of the commercial offering - Good With inside (GWi), which is owned in full by Good With Limited, and will be licenced to customers in the financial services industry for the purpose of determining credit applicants' creditworthiness. Copyright and IPR of all data collected, created or otherwise processed is retained by Good With Limited. Where a scientific publication is considered, data will be aggregated and all personally identifying information redacted to ensure full compliance with UK-GDPR. We will not provide open source data with any publications, however we will share, at our own discretion, samples of data for the purpose of scientific research.

Storage and Backup

How will the data be stored and backed up during the research?

All raw data from the Good With web app (participant responses by mouse or text input) will be stored in a Google Cloud Firestore database, in NoSQL format. Processed data that has undergone statistical analysis or transformation will be stored on the same Google Cloud Firestore database and in a Google BigQuery database. All cloud data will be stored in Google servers located in London, UK; the same legal jurisdiction that the data will be collected in. All data from the Good With web app used in this study will be automatically stored in the cloud and periodically backed up using an automatic process. This avoids problems associated with storing data on physical storage drives.

Data from qualitative user interviews will be stored in audio recordings with common file formats (.wav, .mp3). Processed data from the thematic analysis of these recordings will be stored in .docx files. These files will be stored in password-protected Google Drive folders that are only accessible to the research team. Consent form results will also be stored on Google Drive with the same security measures in place.

All databases used in this study will have disaster-recovery procedures in place in the rare case of cloud service provider failures or human error when maintaining data. This will allow the retrieval of previous versions of files or databases, minimising the potential for data loss. The PI will be responsible for ensuring that these systems are in place, and that collected data is being stored following the data management plan.

How will you manage access and security?

Both Google Cloud Firestore and Google BigQuery are password-protected cloud databases that can only be accessed by members of the Good With data and development team. Both databases have Identity and Access Management security measures, meaning that permissions to view, query and write to these databases must be explicitly granted by an administrator. Both databases are encrypted by default. To query or write data to BigQuery, each member of the research team must authenticate using a unique service account key that is either stored in a password-protected location on their local machine or as a secret hosted in a password-protected area in Google Cloud Platform.

Data stored on Google Drive cloud storage will also require password authentication to access. Access

permissions will also be restricted to members of the research team, meaning that unauthorised user accounts cannot access these data even with a valid password.

The transfer of data from the Good With app to cloud storage is secure and automated, using the service account mechanism detailed above. Any transfer of data from password-protected local devices to Google Drive cloud storage will take place immediately after each data collection session. Locally-stored copies of participant data will be deleted once this transfer has taken place, to reduce the risk that these data are compromised in any way.

Selection and Preservation

Which data are of long-term value and should be retained, shared, and/or preserved?

All data will be retained to provide participants with a continued service should they wish to use the Good With web or mobile apps after the study ends. These data may also be used in further research on financial behaviour and user experiences of the app. Data from qualitative interviews will also be preserved for potential use in further research and product development. Data will be held for a minimum of 6 years following Good With Ltd's privacy policy, after which it will be fully anonymised and stored indefinitely for model integrity and development. This is in line with the University of Plymouth Research Data Policy.

What is the long-term preservation plan for the dataset?

All stored data above will be preserved for a minimum of 6 years, and then fully anonymised and held indefinitely, following the University of Plymouth Research Data Policy. Data will be held in the BigQuery and Firestore databases managed by Good With Ltd for this duration of time since the data pertains to app user accounts. Any preparation and storage costs incurred will be covered by Good With Ltd as part of their existing data storage plans.

Data Sharing

How will you share the data?

The data to be collected contains some sensitive information. Namely, participant financial information, and demographic data. Records will not be fully anonymised when stored in Cloud Firestore, as records of participant names and unique identifiers must be kept to process account deletion requests and for continuity if the participant uses other Good With services. Data stored in Google BigQuery is pseudo-anonymised with a unique identifier being associated with each participant record. As such, participants may be at least partially identifiable from their data if it were shared. The data stored as part of this study also closely relate to the Good With Readiness Score in that key measures derived from the data determine the final score. Sharing the data from this study could allow secondary users to reverse engineer the scoring methodology, violating intellectual property rights.

To mitigate these anonymity and IP concerns, we have decided not to share data with any secondary users. Our results will be disseminated in scientific publications, whitepapers, internal reports, and conference presentations, but the raw data will not be made publicly available. Good With Ltd may

choose to share samples of their fully anonymised data for scientific research purposes at a later date, at their discretion.

Are any restrictions on data sharing required?

Due to intellectual property rights, Good With Ltd has ownership of any data collected through their web app. Since the data for this study will be stored exclusively on Good With Ltd databases, they maintain the right to exclusive access to these data. Details of participant consent will, however, be shared with the University of Plymouth for participant administration purposes. Accordingly, we have included a data-sharing agreement as a supplementary document in our Faculty of Health ethics application.

Responsibilities and Resources

Who will be responsible for data management?

The PI will be responsible for implementing the DMP and ensuring it is reviewed and revised in case of any changes. The PI will engage in data capture, metadata production, data quality, storage, backup, data archiving, and data sharing. The PI will share these responsibilities with the Good With Ltd data and development team.

What resources will you require to deliver your plan?

No additional staff or training will be required to deliver this data management plan. Storage, software, and hardware needs will be met by the Good With Ltd, the funding organisation. Storage charges will be paid for by the funding organisation.