Plan Overview

A Data Management Plan created using DMPonline

Title: Developing a Data-Driven Approach to Evaluating Library Opening Hours at the University of Pretoria

Creator: Paballo Mamabolo

Principal Investigator: Paballo Mamabolo

Data Manager: Paballo Mamabolo

Affiliation: Other

Template: DCC Template

ORCID iD: 04978502

Project abstract:

The following document specifies protocols for the acquisition, curation, preservation, and exchange of research materials pertaining to an evaluation of library operating hour effectiveness at the University of Pretoria. Methodologically informed by Sarah Maule's earlier work at the University of Westminster, the study triangulates quantitative occupancy data with qualitative user experience feedback to generate actionable intelligence for library service enhancement.

ID: 203119

Start date: 28-04-2026

End date: 03-05-2026

Last modified: 03-05-2026

Copyright information:

Developing a Data-Driven Approach to Evaluating Library Opening Hours at the University of Pretoria

Data Collection

What data will you collect or create?

- The data collection plan aims to quantify library usage, service delivery, and user needs across four library sites, while introducing innovative data sources to capture evolving usage patterns.
- Data sources include hourly site usage (from logs), service delivery and user needs (surveys and feedback channels), qualitative data (interviews and planning documents), and digital analytics (chat transcripts and online interactions). Where applicable, existing institutional data will be reused to complement new data.

In addition to traditional methods, innovative data collection approaches such as real-time occupancy tracking and Wi-Fi usage analysis will be used to capture detailed patterns of library usage during different opening hours:

1. Real-Time Occupancy Tracking

Use sensors or Wi-Fi-based tracking to measure how many students are in the library at different times

Tracks peak vs low usage across opening hours

Why it’s significant:

Gives real-time evidence of whether extended hours are actually used

2. Wi-Fi Login Data Analysis

Analyse when students connect to library Wi-Fi networks

Shows patterns of presence and duration of stay

Value:

Helps measure actual usage without relying only on surveys

3. Mobile Check-In / QR Code System

Students scan a QR code when entering/exiting

Can include a quick “why are you here?” or satisfaction rating

Value:

Combines usage + user intent + feedback in one dataset

4. Digital Feedback Kiosks

Install tablets or kiosks inside libraries

Ask quick questions like:

“Are the opening hours suitable?”

“Would you stay longer if hours were extended?” Value:

Captures real-time user sentiment

Data Sources: Type, Format, Coverage, Volume, and Reuse 1. Library Usage Data (Opening Hours & Occupancy)

- **Type:** Time-series data (hourly entry counts, occupancy levels, dwell time where available)

- **Format:** CSV for raw data; Parquet for processed and analytics-ready data
- **Coverage:** All selected University of Pretoria library sites during term time, including historical data for baseline comparison
- **Volume:** Approximately 24 records per day per site; increases over the study period
- **Reuse:** Existing access control logs and library usage records will be reused to establish baseline patterns

Rationale:
This data directly supports the evaluation of whether current opening hours align with actual usage patterns.

2. Service Delivery and User Needs Data

- **Type:** Structured survey data (Likert-scale responses), open-ended feedback, and usage metrics from digital platforms
- **Format:**
  - Surveys: CSV with accompanying data dictionary
  - Feedback data: JSON/CSV formats
  - Transcripts: TXT or JSON
- **Coverage:** Surveys conducted periodically during term time; continuous feedback collection
- **Volume:** Hundreds of survey responses per cycle; potentially large volumes of digital interaction data
- **Reuse:** Existing survey instruments and institutional analytics dashboards will be reused where applicable

Rationale:
Captures user perceptions and satisfaction with library opening hours, supporting decision-making.

3. Qualitative and Contextual Data

- **Type:** Interview transcripts, focus group discussions, meeting notes, and planning documents
- **Format:** TXT, Word, or PDF documents
- **Coverage:** Selected students and staff; institutional planning records
- **Volume:** Moderate and manageable within project scope
- **Reuse:** Previous institutional documents may be used to inform analysis, with anonymisation where necessary

Rationale:
Provides deeper insight into user needs and operational considerations.

4. Digital Analytics Data

- **Type:** Online engagement data (chat logs, website usage, virtual service interactions)
- **Format:** JSON, CSV, or TXT
- **Coverage:** Ongoing data collection across digital platforms
- **Volume:** Variable, potentially large; requires scalable storage
- **Reuse:** Existing analytics systems and dashboards will be utilised

Rationale:
Captures off-site usage patterns, which are important when evaluating extended or reduced opening hours.

Data Formats and Software for Sharing and Long-Term Access

To ensure accessibility and long-term usability, the following formats will be used:

- **CSV:** For structured tabular data (widely accessible)
- **Parquet:** For large datasets requiring efficient storage and analysis
- **JSON/NDJSON:** For semi-structured data such as logs
- **TXT/PDF:** For transcripts and documentation

Justification:
Open, non-proprietary formats are selected to ensure interoperability, long-term preservation, and ease of sharing.

Data Organisation and Naming Conventions

Data will be organised using a structured folder system:

- /Raw_Data
- /Processed_Data
- /Documentation
- /Reports

File naming convention:
UP_LibraryUsage_Hatfield_2026-05-01_v1.csv

Version Control

- Version numbers (v1, v2, v3) will be used for all datasets
- A version log will document all changes
- Raw data will remain unchanged to preserve integrity
- Cloud platforms (e.g., OneDrive) will maintain version history

Quality Assurance

To ensure accuracy and consistency:

- Standardised survey instruments will be used
- Pilot testing will be conducted before full deployment
- Data validation rules will be applied during entry
- Regular data cleaning and verification processes will be implemented
- Supervisor review will be conducted where necessary

Storage, Backup, and Access

- Data will be stored on secure University of Pretoria servers
- Daily incremental backups and weekly full backups will be performed
- Cloud replication will be used for redundancy
- Role-based access control will be applied
- Sensitive data will be anonymised and protected

How will the data be collected or created?

Quantitative data will be collected through:

- **Access control system logs** (gate counts and usage patterns) to analyse how frequently libraries are used during different opening hours
- **Online surveys** distributed to students and staff to measure satisfaction levels and preferences regarding library opening times

Qualitative data will be collected through:

- **Interviews and focus groups** with students and library staff to gain deeper insights into user experiences, challenges, and expectations

This combination of methods supports a **data-driven evaluation** by linking actual usage patterns with user perceptions and institutional insights.

Data Standards and Organisation

To ensure consistency, interoperability, and reuse:

- Data and metadata will follow recognised standards such as **Dublin Core**
- Controlled vocabularies will be used in surveys to standardise responses (e.g., satisfaction levels, usage frequency)
- Standardised survey instruments will be used across all participants

Data will be organised using clear naming conventions:
UP_LibrarySurvey_2026-05-01_v1.csv

This ensures files are easily identifiable, organised, and retrievable.

Version Control

Version control will be managed through structured file naming and storage practices:

- Each dataset will include version numbers (e.g., v1, v2, v3)
- A version log will document all significant changes
- Original raw data will remain unchanged to preserve data integrity
- Updated datasets will be saved as new versions
- Cloud storage platforms (e.g., OneDrive) will provide automatic version history tracking

Quality Assurance Processes

To ensure the accuracy, reliability, and consistency of the data, the following measures will be implemented:

- Use of standardised survey questions and data collection instruments
- Pilot testing of survey instruments to ensure clarity and relevance
- Validation checks during data entry (e.g., required fields, response limits)
- Regular data cleaning and verification before analysis
- Cross-checking of survey data with usage logs where applicable
- Continuous review of collected data to identify errors or inconsistencies
- Documentation of all data collection procedures and any changes made

Documentation and Metadata

What documentation and metadata will accompany the data?

Metadata will be created to describe datasets, including title, creator, date of collection, methodology, and file format. Standard naming conventions will be used, and datasets will be documented using **Dublin Core elements** to ensure consistency, discoverability, and reuse. Given UP's existing data infrastructure (Figshare repository with 169 datasets, UPSpace institutional repository), metadata will align with both international standards and UP-specific requirements.

1. Information Required for Future Interpretation To ensure the overnight opening usage data can be read, understood, and reused by secondary users in the future, the following contextual and structural information must be documented: **Basic Discovery Metadata:**

- **Creator/Contributor:** [Your name as researcher], UP Library Services staff (Head: Library Services, site managers), security/cleaning staff who verify after-hours usage, Student Representative Council representatives who provided feedback
- **Title:** University of Pretoria Library Opening Hours: Cost-Effectiveness and User Needs Analysis (2026)
- **Date of Creation:** Data collection period [insert semester dates aligned with UP academic

calendar]; analysis conducted [insert dates]
- **Access Conditions:** Institutional dataset governed by UP Research Ethics Committee clearance and Library Services data sharing agreements. Student usage data requires anonymisation. Cost data may be subject to institutional confidentiality restrictions.

Methodological Documentation:

- **Data Collection Method:** Mixed methods approach adapted from Maule's framework but tailored to UP's multi-campus context:
  - *Quantitative:* Automated entry log data (swipe cards/turnstiles) or manual headcounts at UP library sites: Merensky 2 (Hatfield main), Oliver R Tambo Law Library, Education Library, BMS & Dentistry Library, Mamelodi Library, Jotello F. Soga Veterinary Science Library, Health Sciences Library
  - *Qualitative:* Student feedback via UP Student Experience surveys, Student Representative Council input, focus groups, comment cards, and potentially the South African National Student Survey (SNS) equivalent if available
  - **Procedural Information:**
    - Hourly/daily entry data extraction from UP access control systems or manual observation protocols
    - Cost calculation methodology developed in consultation with UP Library Services administration (staffing, security, utilities, maintenance, technology)
    - Cross-referencing with UP academic calendar (Semester 1: February–June; Semester 2: July–November; including examination periods)
  - **Temporal Coverage:** Minimum two full semesters to establish patterns; vacation periods (December–January, June–July) excluded or analysed separately as usage patterns differ significantly
  - **Spatial Coverage:** All UP library sites across six campuses (Hatfield, Groenkloof, Prinshof, Mamelodi, Onderstepoort, Hillcrest)

Technical & Structural Metadata:

- **Variable Definitions:**
  - *Hourly entry count / headcount:* Number of users present or entering at specific times (method depends on UP's access control infrastructure)
  - *Site capacity:* Total seating/study space capacity per library (from UP Estates or Library Services records)
  - *Percentage capacity:* Usage divided by capacity
  - *Cost per user hour:* Total operational cost (staffing + overheads) divided by total user hours
  - *Term time week:* UP academic calendar categorisation (e.g., Week 1–13 of semester, examination weeks)
  - *Peak/off-peak hours:* Defined by usage thresholds relative to site capacity (to be determined during analysis)
  - *Cost-effectiveness threshold:* To be defined in consultation with UP Library Services (e.g., minimum percentage capacity required to justify extended hours)
  - **Units of Measurement:** Student count (whole numbers); South African Rand (ZAR) for costs; percentage (%); time (24-hour format); academic week categorisation
- **Assumptions Made:**
  - Automated entry data accurately represents unique users (adjustment needed if same user enters multiple times)
  - Term-time usage patterns are representative of overall academic year needs (vacation patterns may differ)
  - Cost allocations for shared services (security, cleaning, utilities) can be reasonably apportioned per library site

        - "User needs" can be adequately captured through survey feedback and usage patterns (behavioural proxy for stated preferences)

Data Format & File Types:

- **Original Format:** Raw data extracted from UP access control systems (.csv, .xlsx) or manual logbooks; cost spreadsheets from UP administration
- **Processed Format:** Single master dataset with standardised headings, enabling pivot table analysis, cross-site comparison, and temporal trend analysis (cloud-based or secure UP server)

Specific Capture Mechanisms:

- **README.txt:** Plain text file providing overview, context, and quick-start guidance for any future user, including UP-specific context (11 libraries across six campuses, approximately 38,161 undergraduates and 16,805 postgraduates served)
- **Data Dictionary:** Structured document defining all variables, codes, and measurement units (e.g., site codes for all UP libraries, UP term week categories aligned with academic calendar, ZAR cost categories)
- **Methodology Narrative:** Following Maule's reflective approach, a written account of decisions made, challenges encountered (e.g., extracting data from multiple access control systems, reconciling different cost accounting methods across sites), and rationale for exclusions
- **Research Journal:** Individual reflective journal documenting cognitive and metacognitive processes, biases encountered, and analysis techniques used (as required by your module's ANU learning journal framework)
- **Version Control:** Documentation of transformations from raw (e.g., swipe card logs, manual headcounts, fragmented cost spreadsheets) to processed (master dataset) states, ensuring reproducibility

Justification for Selection:

- **Dublin Core** is selected because the research is situated within a library and information science context and must integrate with existing UP Library metadata infrastructures, including UPSpace.
- **DataCite** is selected because UP already maintains a Figshare data repository with 169 datasets, indicating institutional capacity and precedent for citable data publication.
- **DDI** is selected because the dataset shares characteristics with social science observational data (repeated measurements across multiple sites, temporal structure, cost variables) and DDI excels at documenting methodology and variable-level metadata.
- **CSV on the Web** is selected to ensure the processed tabular data is FAIR-compliant and can be automatically parsed by analysis tools without losing semantic meaning.
- **NRF requirements** are selected to ensure compliance with South African national research governance and UP's institutional policies.

UP-Specific Technological Context:

- UP Library Services uses automated systems including self-checkout, information kiosks, and the UP mobile application for renewals and inquiries. Documentation must record whether entry data is extracted from these integrated systems or requires manual collection.
- The presence of "Libby" (service robot) for short enquiries indicates UP is investing in technology-mediated services—documentation should note whether such innovations affect usage measurement or user expectations.

Limitations to Document (UP-Specific):

- **Multi-campus complexity:** Unlike Maule's four-site London model, UP has libraries across six campuses with potentially different usage patterns, cost structures, student demographics, and

transport accessibility. Documentation must capture these structural differences.
- **Automated vs. manual data collection:** If UP uses automated entry systems, documentation must note any system limitations (e.g., counts all entries rather than unique users, excludes staff/faculty if using different access cards, misses group study room usage if not counted separately)
- **Cost data sensitivity:** Financial data from UP Library Services may be commercially or politically sensitive; documentation must record anonymisation or aggregation protocols applied and any restrictions on sharing
- **South African academic calendar:** Unlike Northern Hemisphere patterns, UP's academic year runs from February to November with extended December–January vacation. "Term time" must be clearly defined in local context (Semester 1: approx. February–June; Semester 2: approx. July–November, with examination periods)
- **Equity and access considerations:** UP serves a diverse student body including part-time workers, distance students, and students from different socioeconomic backgrounds. Documentation should note whether these factors are captured in "user needs" analysis or represent a limitation.

Ethical and Privacy Considerations:

- Usage data linked to student IDs (if obtained via swipe cards) requires careful anonymisation before secondary use; documentation must record anonymisation protocols
- Student feedback and focus group data require informed consent documentation and secure storage protocols compliant with POPIA (Protection of Personal Information Act)
- Cost data from UP administration may be subject to institutional confidentiality agreements; documentation must record any redaction or aggregation applied
- Given UP's public university status, there may be transparency expectations around library service provision—documentation should balance openness with privacy protection

Ethics and Legal Compliance

How will you manage any ethical issues?

- Ethical clearance required from UP
- Informed consent from participants
- Anonymisation of personal data
- Compliance with Protection of Personal Information Act (POPIA)

Ethics, Consent, and Sensitive Data Management 1. Consent for Data Preservation and Sharing Consent Framework: My research involves **two distinct data types**, each requiring different consent approaches: **Quantitative usage data**

- Source: UP Library Services automated entry logs / manual headcounts
- Consent Required: Not individually required — aggregate, non-identifiable usage statistics are institutional operational data
- How Consent Will Be Obtained: Formal data sharing agreement with UP Library Services (Head: Library Services) authorising access to anonymised aggregate data for research purposes

Qualitative feedback data

- Source: Student surveys, focus groups, comment cards, Student Representative Council input
- Consent Required: Explicit informed consent required — human participants providing opinions on

library services
- How Consent Will Be Obtained: Written consent forms approved by UP Research Ethics Committee; opt-in consent with clear explanation of data preservation, sharing, and reuse intentions

Specific Consent Provisions: For **qualitative data involving human participants**, consent forms will explicitly address:

- **Purpose:** That data is being collected to assess library opening hours cost-effectiveness and user needs
- **Preservation:** That anonymised responses may be preserved beyond the immediate research period for longitudinal comparison or institutional benchmarking
- **Sharing:** That de-identified data may be shared with UP Library Services for service improvement, and potentially deposited in UP's Figshare repository or UPSpace for secondary research use
- **Reuse:** That future researchers may access anonymised datasets to study academic library service provision trends
- **Withdrawal:** That participants may withdraw consent up to [specify date, e.g., data analysis commencement] without penalty

For **quantitative institutional data**, a **memorandum of understanding** with UP Library Services will specify:

- Data may only be used for the stated research objective
- Raw identifiable data (e.g., student ID-linked entry logs) will not be shared externally
- Aggregate outputs may be published in theses, journal articles, or institutional reports
- UP Library Services retains right to review outputs before publication if commercially sensitive cost data is included

2. Protection of Participant Identity Anonymisation Protocols: Entry log / swipe card data

- Identifying Risk: Student/staff ID numbers linked to entry times
- Anonymisation Method: Direct anonymisation: Remove all personal identifiers (student numbers, staff IDs) before analysis; replace with randomly generated participant codes; aggregate to hourly/daily/site-level counts
- When Applied: At point of data extraction from UP systems

Survey / focus group responses

- Identifying Risk: Names, student numbers, course details, direct quotes
- Anonymisation Method: Direct anonymisation: Remove names and contact details; replace with pseudonyms; contextual anonymisation: Remove or generalise identifying details in quotes (e.g., "third-year BCom student" becomes "undergraduate student"; specific course codes become general faculty)
- When Applied: During transcription (for audio) and coding (for text); checked before any quotation in outputs

Cost data

- Identifying Risk: Salary information, vendor contracts, specific staff names
- Anonymisation Method: Aggregation and redaction: Report costs at site or service level only; redact individual salary figures; use percentage allocations where individual identification is possible
- When Applied: During analysis; before any publication or presentation

Additional Identity Protection Measures:

- **Data separation:** Identifying information (consent forms, contact sheets) stored separately from

research data in password-protected files
- **Pseudonymisation:** All qualitative participants assigned research IDs (e.g., P001, P002); key linking pseudonyms to real identities stored separately and destroyed after data verification
- **K-anonymity check:** Quantitative data aggregated to ensure no individual can be identified by combining variables (e.g., ensuring no cell in cross-tabulations contains fewer than 5 individuals)
- **Quote review:** All direct quotations from qualitative data reviewed to ensure no inadvertent identification through context, slang, or specific reference to unique circumstances

Limitations on Anonymisation: I recognise that **complete anonymisation may not always be possible**:

- In focus groups, participants may know each other's identities; consent forms will note that confidentiality cannot be guaranteed for disclosures made within the group
- Small campus libraries (e.g., Mamelodi, Onderstepoort) with low usage may make aggregation difficult while preserving analytical value; in such cases, data may be reported at broader temporal or categorical levels, or excluded from fine-grained site comparisons if identification risk is high
- Cost data involving specific contracts or unique staffing arrangements may require institutional approval for any publication; UP Library Services will review outputs before dissemination

3. Secure Storage and Transfer of Sensitive Data Data Classification and Handling: High sensitivity data

- Data Examples: Raw entry logs with student IDs; individual cost spreadsheets with salary data; signed consent forms with names and signatures
- Storage: Encrypted local storage: Password-protected, encrypted hard drive or UP-approved secure server; access limited to researcher and supervisor only
- Transfer: Encrypted transfer only: UP secure file transfer system or encrypted email; no personal cloud services (Google Drive, Dropbox)
- Retention: Raw identifiable data: destroyed after anonymisation and verification (maximum [specify, e.g., 6 months post-collection]); consent forms: retained for [UP policy period, typically 5 years] then securely destroyed

Medium sensitivity data

- Data Examples: Anonymised but granular usage data (hourly counts by site); coded qualitative data with pseudonyms
- Storage: Password-protected storage: UP network drive or institutional cloud with two-factor authentication; access controlled
- Transfer: Secure institutional channels: UP email with encrypted attachments; shared via UP-approved collaboration platforms
- Retention: Retained for duration of research plus [UP policy period, typically 5 years] for verification; then deposited as restricted dataset or destroyed per participant withdrawal requests

Low sensitivity data

- Data Examples: Fully aggregated usage statistics; fully anonymised qualitative themes; published outputs
- Storage: Standard institutional storage: UPSpace, Figshare, or open access repository
- Transfer: Open dissemination: Publication, conference presentation, institutional reporting
- Retention: Permanent preservation in UP repositories where appropriate; published outputs retained indefinitely

Technical Security Measures:

- **Encryption:** All devices storing research data encrypted at rest (BitLocker for Windows, FileVault for Mac); files transferred using encrypted protocols (SFTP, HTTPS)

- **Access control:** Unique user accounts; no shared passwords; principle of minimum necessary access
- **Backup:** Regular encrypted backups to UP-approved institutional storage; no backups to personal devices or unapproved cloud services
- **Physical security:** Devices containing research data not left unattended in public spaces; screen locked when unattended; printed materials stored in locked cabinets
- **Disposal:** Secure deletion of electronic files (not just recycle bin); physical documents shredded using UP-approved disposal services

POPIA Compliance (Protection of Personal Information Act 2013): As research involving South African data subjects, this study must comply with POPIA. Key obligations addressed: **Accountability**

- Researcher registered as responsible party; UP Research Ethics Committee oversight; DMP documents compliance measures

Processing limitation

- Only necessary personal information collected; lawful basis (consent for qualitative; legitimate institutional interest for operational data, with data sharing agreement)

Purpose specification

- Clear research objective stated; data not used for unrelated purposes without renewed consent/agreement

Further processing limitation

- Any secondary use (e.g., longitudinal comparison, benchmarking) approved by UP Library Services and, where involving qualitative data, re-consented or fully anonymised

Information quality

- Participants may request correction of their data; UP Library Services data verified for accuracy

Openness

- Privacy notice provided to all participants; data sharing agreements transparent with institutional partners

Security safeguards

- Technical and organisational measures documented above; incident response plan in place (report to UP Information Security and Research Ethics Committee)

Data subject participation

- Participants informed of right to access, correct, or request deletion of their data; withdrawal procedures communicated

4. Institutional Ethics Review Ethics Committee Engagement: Pre-application

- Action: Consultation with UP Research Ethics Committee (or Faculty of Humanities/Natural and Agricultural Sciences Ethics Committee, depending on research affiliation) to determine review pathway
- Timeline: Before data collection begins

Application

- Action: Full ethics application submitted including: research proposal, consent forms, participant information sheets, data management plan, anonymisation protocol, POPIA compliance statement

- Timeline: Minimum 6–8 weeks before intended data collection

Conditional approval

- Action: Address any committee feedback; revise protocols as required
- Timeline: As requested by committee

Final approval

- Action: Ethics clearance certificate obtained; clearance number recorded in all documentation
- Timeline: Before any participant contact or institutional data access

Ongoing reporting

- Action: Annual progress reports if research extends beyond one year; adverse event reporting if any data breaches or participant complaints occur
- Timeline: Throughout research period

Ethics Application Documentation: The following will be submitted to the ethics committee:

- Detailed description of all data types and sources
- Sample consent forms (for qualitative participants) and data sharing agreements (for institutional partners)
- Anonymisation protocol with worked examples
- Data storage and security plan
- POPIA compliance checklist
- Risk assessment for participant identification and mitigation strategies

5. Summary of Ethical Safeguards Concern: Lack of consent for data sharing

- Safeguard: Explicit opt-in consent for qualitative participants; formal data sharing agreement for institutional data
- Responsible Party: Researcher; UP Research Ethics Committee

Concern: Participant identification

- Safeguard: Multi-layer anonymisation (direct removal, pseudonymisation, contextual generalisation, aggregation); k-anonymity checks
- Responsible Party: Researcher; verified by supervisor

Concern: Sensitive data breach

- Safeguard: Encryption at rest and in transit; access controls; secure backup; physical security; POPIA-compliant incident response
- Responsible Party: Researcher; UP Information Security

Concern: Institutional cost data exposure

- Safeguard: Aggregation and redaction; pre-publication review by UP Library Services; confidentiality clauses in data sharing agreements
- Responsible Party: Researcher; UP Library Services

Concern: Focus group confidentiality limits

- Safeguard: Explicit warning in consent forms and pre-session briefing that group disclosures cannot be fully controlled
- Responsible Party: Researcher

Concern: Long-term data preservation without consent

- Safeguard: Consent forms explicitly address preservation and reuse; withdrawal options communicated; fully anonymised datasets for open preservation
- Responsible Party: Researcher; UP repository administrators

Concern: Cross-border data transfer (if collaborating internationally)

- Safeguard: POPIA requirements for transfer to third countries; adequacy decisions or binding corporate rules; participant notification if applicable
- Responsible Party: Researcher; UP Legal Services

6. Reflection: Ethical Learning from the Maule Case Study Maule's Westminster case study offers limited direct ethical discussion—her focus was operational data (aggregate headcounts) and institutional decision-making rather than human participant research. However, several lessons apply:

- **Institutional data is not automatically "safe":** Maule linked headcount data to Student Record System (SRS Web) data for capacity calculations. At UP, similar linkage risks re-identification if not carefully managed. My protocol explicitly separates raw identifiable data from analysis datasets.
- **Qualitative data carries higher ethical obligations:** Maule coded NSS comments and feedback cards but did not discuss consent for this secondary use of existing feedback. My research proactively seeks consent for any primary qualitative collection and anonymises all qualitative outputs.
- **Stakeholder trust depends on transparency:** Maule's success relied on presenting data clearly to Student Union and registry managers. My ethical transparency—documenting how data is protected, who can access it, and for how long—builds similar trust with UP Library Services and student participants.
- **Data culture includes ethical culture:** Maule's project catalysed a broader data culture at Westminster. Embedding ethical data handling from the outset ensures that any emerging "data culture" at UP is responsible and sustainable, not merely efficient.

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

All personal data will be anonymised, and participants will provide informed consent. The study will comply with institutional ethical guidelines and POPIA requirements to ensure privacy and confidentiality.

1. Ownership of Data This research generates and uses multiple data types, each with distinct ownership arrangements: **Institutional operational data (quantitative usage data)**

- Data examples: Automated entry logs, swipe card records, manual headcount logs, site capacity figures, library opening hours schedules
- Data creator: University of Pretoria Library Services (generated through routine operations)
- Ownership: University of Pretoria, held by the Department of Library Services under the oversight of the Head: Library Services
- Researcher status: Licensed user under formal data sharing agreement; no ownership claims over raw institutional data
- Rationale: This data is created by UP as part of its statutory educational and administrative functions; it constitutes institutional records subject to UP records management and information governance policies

Research-generated analytical data

- Data examples: Cleaned and aggregated datasets, calculated cost-effectiveness metrics, percentage capacity analyses, trend visualisations, coded qualitative themes
- Data creator: [Your name] as researcher, with supervisory guidance

- Ownership: Shared between the researcher and the University of Pretoria, governed by UP intellectual property policy for postgraduate research
- Rationale: Under UP's standard research policy, the university typically retains ownership of research outputs created using institutional resources, facilities, or data, while the researcher holds moral rights (attribution) and may share in commercial exploitation if applicable. The exact distribution depends on whether the research is funded (funder terms may apply) or unfunded (standard UP policy applies)

Qualitative primary data

- Data examples: Survey responses, focus group transcripts, interview recordings, comment card entries
- Data creator: Research participants (students, staff) who provided the information; researcher who collected and processed it
- Ownership: Participants retain moral rights over their personal contributions; researcher and UP hold the compiled dataset as an anonymised research output
- Rationale: Original expressions belong to participants, but once anonymised and aggregated, the structured dataset becomes a research output governed by consent agreements specifying reuse permissions

Cost and financial data

- Data examples: Staffing costs, utility expenditures, security contracts, operational budgets for library sites
- Data creator: University of Pretoria Finance Department and Library Services administration
- Ownership: University of Pretoria; commercially and administratively sensitive
- Researcher status: Licensed access for research purposes only; no ownership or unrestricted reuse rights

2. Licensing for Reuse The following licensing framework applies to different data outputs from this research: **Fully anonymised aggregate usage datasets**

- Licence: Creative Commons Attribution 4.0 International (CC BY 4.0)
- Rationale: These datasets contain no identifiable information and represent valuable benchmarking resources for other South African and international academic libraries. CC BY maximises reuse while ensuring attribution to UP and the researcher. This aligns with open science principles and UP's Figshare repository practice of making datasets citable and discoverable.
- Conditions: Users must attribute the original source; may adapt, redistribute, and build upon the data for any purpose including commercial use

Anonymised qualitative datasets (coded themes, aggregated survey responses)

- Licence: Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0)
- Rationale: Qualitative data, even when anonymised, carries higher ethical sensitivity than quantitative operational data. The non-commercial restriction protects against exploitative reuse; the share-alike requirement ensures derivative works remain open. This balances openness with participant protection.
- Conditions: Attribution required; non-commercial use only; derivatives must use identical licensing terms

Research outputs (thesis, journal articles, conference presentations)

- Licence: Copyright retained by researcher and UP; publication agreements with individual publishers will determine specific licensing terms
- Rationale: Standard academic publishing practice applies. The researcher typically grants

publishers a licence to publish while retaining certain rights (e.g., self-archiving in UPSpace after embargo period). UP's open access policy may require deposit in UPSpace.

Institutional cost-effectiveness models and methodologies

- Licence: University of Pretoria retains ownership; may be licensed to other institutions under separate agreement
- Rationale: If the research develops a novel methodology for assessing library cost-effectiveness, this may have commercial or consultancy value for UP. The university may choose to protect or exploit this IP rather than releasing it openly.

Raw identifiable data and intermediate processing files

- Licence: No public licence; restricted internal use only
- Rationale: Raw data containing personal identifiers or sensitive institutional information cannot be shared under any open licence. Access is governed by the data sharing agreement with UP Library Services and POPIA compliance requirements.

3. Restrictions on Reuse of Third-Party Data This research incorporates or accesses several categories of third-party data with reuse restrictions: **UP Library Services operational data (entry logs, headcounts)**

- Source: University of Pretoria Library Services
- Restrictions: Cannot be shared in raw identifiable form; aggregate outputs only; UP Library Services retains right to review research outputs before publication if commercially sensitive; data sharing agreement specifies research purpose only
- How managed: Anonymisation and aggregation before any external sharing; memorandum of understanding with UP Library Services governs all use

UP Finance Department cost data

- Source: University of Pretoria Finance Department
- Restrictions: Highly sensitive; individual salary figures and vendor contracts cannot be disclosed; aggregate site-level or service-level costs only; pre-publication review required; no commercial use without explicit institutional approval
- How managed: Aggregation and redaction during analysis; confidentiality clauses in data sharing agreement; UP Library Services and Finance Department review outputs before dissemination

Student Record System data (if accessed for capacity calculations or demographic analysis)

- Source: University of Pretoria Student Administration
- Restrictions: Personal information protected under POPIA; cannot be linked to library usage data in any shareable output; aggregate faculty or campus-level enrolment figures only
- How managed: Strict separation from identifiable usage data; aggregate figures obtained through official channels rather than direct system access where possible

South African National Student Survey or equivalent external survey data (if used for comparative context)

- Source: Higher Education Data Analytics (HEDA) or Department of Higher Education and Training
- Restrictions: Subject to survey provider's terms of use; may require separate data licence; typically restricted to institutional benchmarking and research; public release of institutional-level data may be prohibited
- How managed: Verify licence terms before incorporation; cite appropriately; use only for internal analysis or with provider permission for publication

Published comparative data from other universities (e.g., Maule's Westminster case study)

- Source: Stubbing, A. (2022). Data-driven decisions: A practical toolkit for librarians and information professionals. Facet Publishing.
- Restrictions: Copyrighted material; quotations and data reproduction limited by fair dealing/fair use provisions; cannot reproduce entire case study or datasets without publisher permission
- How managed: Cite all quotations and paraphrased findings; seek publisher permission for any extensive reproduction; use as contextual reference rather than direct data reuse

4. Postponement or Restriction of Data Sharing The following scenarios may require postponement or restriction of data sharing: **Publication embargo**

- Dataset affected: All research-generated analytical datasets and qualitative findings
- Duration: Maximum 12 months from thesis submission or first publication, whichever is earlier
- Rationale: Protects the researcher's ability to publish findings without pre-emption; standard practice for postgraduate research at UP
- Conditions: During embargo, data held in restricted access in UP Figshare or secure institutional storage; metadata record created with embargo notice; automatic release after embargo period unless extension requested and justified

Institutional review requirement

- Dataset affected: Any outputs containing cost data or operational recommendations
- Duration: Until UP Library Services and relevant administrative departments have reviewed and approved publication
- Rationale: UP may have legitimate interests in reviewing research that evaluates institutional cost-effectiveness before public release, particularly if findings could affect procurement, staffing, or competitive positioning
- Conditions: Researcher submits draft outputs to UP Library Services; review period of 30 working days; reasonable amendments incorporated; UP cannot unreasonably withhold approval for academically rigorous research

Commercial sensitivity hold

- Dataset affected: Novel cost-effectiveness methodologies, detailed cost breakdowns, vendor-specific information
- Duration: Indefinite unless UP elects to release or commercialise
- Rationale: If the research develops methodologies or reveals financial information with potential commercial value or competitive sensitivity, UP may choose to restrict rather than openly license
- Conditions: Governed by UP intellectual property policy; researcher consulted on commercialisation options; revenue sharing per UP policy if applicable

Participant re-consent requirement

- Dataset affected: Qualitative datasets where participants consented to research use but not necessarily to open data release
- Duration: Until re-consent obtained or data fully anonymised beyond recognition
- Rationale: Original consent may not have explicitly covered open data deposition; ethical practice requires specific consent for open sharing
- Conditions: Contact participants if traceable; seek explicit opt-in for open dataset inclusion; exclude non-consenting participants' data from open release (retain in restricted form for thesis verification)

Funder requirements (if applicable)

- Dataset affected: All research data if funded by external body
- Duration: Per funder policy (e.g., NRF typically requires data sharing within 12 months of project completion; some international funders may have shorter or longer requirements)
- Rationale: Funder terms and conditions take precedence where they exceed institutional

minimums
- Conditions: Record funder requirements in DMP; ensure compliance timeline; negotiate with funder if UP policies create conflict

5. Institutional and Funder Policy Compliance University of Pretoria policies:

- UP Open Access Policy: Requires deposit of research outputs in UPSpace; may mandate open licensing for certain output types
- UP Intellectual Property Policy: Governs ownership, commercialisation, and revenue sharing for research-generated IP
- UP Research Data Management Policy: Sets minimum standards for data retention, sharing, and security
- UP Records Management Policy: Governs retention and disposal of institutional records, including operational data used in research

National policies:

- South African National Research Foundation (NRF) Data Management Plan requirements: If NRF-funded, must comply with open data expectations and specified retention periods
- Protection of Personal Information Act (POPIA) 2013: Governs all handling of personal information; restricts sharing of identifiable data; mandates lawful processing

Departmental/group policies:

- Department of Information Science / Library and Information Science (depending on your faculty affiliation): May have specific expectations about data management for information science research
- UP Library Services internal data governance: Operational data access governed by internal policies; researcher must comply with any additional requirements imposed by data provider

6. Reflection: IPR and Licensing Learning from the Maule Case Study Maule's case study offers implicit but important lessons on data ownership and reuse:

- **Institutional data ownership is assumed, not negotiated:** Maule accessed five years of headcount data that had "never been interrogated or analysed" despite comprehensive collection. This suggests Westminster viewed the data as operational record-keeping rather than a research asset. At UP, my explicit data sharing agreement prevents this ambiguity—ownership is clear, and my licensed use is documented from the outset.
- **No open licensing was applied:** Maule's outputs were published in a commercial edited volume (Stubbing, 2022), meaning the case study itself is copyrighted and not openly reusable. My decision to apply Creative Commons licences to anonymised datasets represents a different approach—treating research data as a public good while protecting sensitive elements.
- **Third-party data integration requires care:** Maule used National Student Survey (NSS) comments, which are subject to survey provider terms. My research must similarly respect any external survey data boundaries, particularly if comparing UP findings to national benchmarks.
- **The value of methodology versus data:** Maule's real contribution was not the raw headcounts but the analytical process—the data journey, stakeholder consultation, and change management. This suggests that at UP, the IPR value may lie in the cost-effectiveness assessment methodology I develop rather than in the usage data itself. UP's IP policy should be consulted early if methodological innovation emerges.

Storage and Backup

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

- Stored on UP secure servers
- Cloud backup (e.g., OneDrive / Google Drive)
- Weekly backups
- Password-protected files

1. Storage Requirements and Capacity Estimated data volumes: Quantitative usage data

- Raw entry logs / swipe card extracts: Approximately 5–10 MB per semester per site (depending on granularity and system format); across 7 UP library sites and 2 semesters, estimated 70–140 MB raw data
- Processed master dataset: Approximately 20–40 MB (cleaned, aggregated, with derived variables and visualisations)
- Cost spreadsheets: Approximately 5–10 MB (aggregate figures, redacted from larger institutional finance datasets)
- Total quantitative storage requirement: Approximately 100–200 MB

Qualitative data

- Survey responses (text): Approximately 2–5 MB
- Focus group audio recordings: Approximately 100–200 MB per focus group (assuming 60–90 minutes, MP3 format); estimated 3–5 focus groups = 300 MB – 1 GB
- Focus group transcripts: Approximately 1–2 MB per transcript
- Comment card scans / photographs: Approximately 10–20 MB
- Total qualitative storage requirement: Approximately 500 MB – 1.5 GB

Research documentation and outputs

- Thesis draft and versions: Approximately 50–100 MB
- Analysis scripts / spreadsheets: Approximately 20–50 MB
- Presentation files and visualisations: Approximately 50–100 MB
- Research journal and reflective documentation: Approximately 10–20 MB
- Total documentation storage requirement: Approximately 150–300 MB

Overall estimated project storage requirement: Approximately 1–2 GB Assessment of sufficiency: The University of Pretoria provides institutional storage through several channels. The estimated 1–2 GB requirement is well within standard allocations:

- UP network drives (typically 5–50 GB per staff/postgraduate student): Sufficient for active project storage
- UP Figshare data repository: Unlimited storage for publishable datasets; currently hosts 169 datasets
- UPSpace institutional repository: Sufficient for thesis and published outputs
- UP Microsoft 365 cloud storage (OneDrive): 1 TB per user; more than adequate for collaboration and backup

Conclusion: No additional storage charges are anticipated. The project falls within standard UP postgraduate research storage provisions. If the project expands unexpectedly (e.g., video focus groups instead of audio, large-scale survey expansion), the researcher will apply for additional allocation through the supervisor and UP IT Services. **2. Backup Strategy Backup frequency and locations: Active research data (daily working files)**

- What is backed up: Raw data extracts, processed datasets, analysis files, thesis drafts, research journal
- Backup frequency: Continuous automatic sync during active work sessions; full backup at end of

each working day
- Primary location: UP network drive (H: drive or departmental research folder)
- Secondary location: UP OneDrive cloud storage (automatic sync)
- Tertiary location: Encrypted external hard drive (weekly manual sync as cold backup)
- Number of copies: 3 (live working copy + 2 backups)

Audio and sensitive qualitative data

- What is backed up: Focus group recordings, raw survey responses with potential identifiers, consent forms
- Backup frequency: Immediate backup after collection; daily sync during analysis phase
- Primary location: Encrypted folder on UP network drive (access restricted to researcher and supervisor)
- Secondary location: Encrypted external hard drive (stored separately from primary device)
- Tertiary location: Not cloud-synced due to sensitivity; physical secure storage only
- Number of copies: 2 (live encrypted copy + 1 offline encrypted backup)

Final datasets and outputs (preservation copies)

- What is backed up: Anonymised master datasets, final thesis, published articles, deposited datasets
- Backup frequency: At completion of each milestone (data collection end, analysis completion, thesis submission, publication)
- Primary location: UP Figshare (for datasets) or UPSpace (for thesis and publications)
- Secondary location: UP institutional archive
- Number of copies: 2 institutional repositories + researcher's personal preservation copy

3. Responsible Parties for Backup and Recovery Researcher (primary responsibility)

- Daily: Ensure active files are saved to UP network drives / OneDrive (automatic sync verification)
- Weekly: Verify external hard drive backup integrity; check sync logs for errors
- After each data collection event: Immediately transfer and back up qualitative recordings; verify file integrity
- Monthly: Test restore process from backup to ensure recoverability
- Upon project completion: Deposit final datasets in UP Figshare; ensure thesis uploaded to UPSpace

Supervisor (oversight responsibility)

- Quarterly: Review backup logs and storage organisation with researcher
- Ad hoc: Verify access to shared backup locations in case researcher unavailability
- Upon project completion: Confirm all outputs deposited in appropriate repositories

UP IT Services (infrastructure responsibility)

- Continuous: Maintain network drive and OneDrive infrastructure; monitor for hardware failures
- Scheduled: Perform institutional-level backups of server infrastructure (daily incremental, weekly full)
- On request: Assist with data recovery; provide guidance on storage best practices
- Incident response: Manage institutional infrastructure recovery in event of major system failure

UP Library Services (data provider responsibility)

- Initial: Provide raw operational data in agreed format
- Ongoing: Maintain original source systems (entry logs, swipe card databases) according to institutional records management policy
- Note: UP Library Services retains original operational data; researcher backups are working copies for analysis purposes only

4. Incident Recovery Procedures Scenario 1: Local device failure (laptop crash, theft, damage)

- Impact: Loss of working files on local machine
- Recovery process:
  1. Obtain replacement device through UP IT Services or personal arrangement
  2. Restore working files from UP OneDrive (automatic sync means minimal loss, typically <1 day of work)
  3. Verify file integrity against last known good version on UP network drive
  4. Resume work from cloud-synced version
  5. If cloud sync was interrupted, restore from encrypted external hard drive weekly backup (maximum 7 days potential loss)
- Expected recovery time: 2–4 hours for file restoration; 1 day to reconfigure software environment

Scenario 2: Corruption of processed dataset or analysis files

- Impact: Loss of derived data, requiring reprocessing from raw data
- Recovery process:
  1. Identify last known uncorrupted version from version history (UP OneDrive retains 30 days of version history; UP network drives may have longer snapshot retention)
  2. Restore clean version from version history
  3. If version history insufficient, restore from encrypted external hard drive weekly backup
  4. Reapply any processing steps performed since last backup (documented in research journal and version control log)
  5. Verify integrity against raw source data
- Expected recovery time: 4–8 hours depending on reprocessing complexity

Scenario 3: Loss of qualitative audio recordings (irreplaceable primary data)

- Impact: Permanent loss of focus group data; cannot be recreated
- Prevention: Multiple backups immediately after collection
- Recovery process:
  1. Check primary encrypted folder on UP network drive
  2. If unavailable, restore from encrypted external hard drive (stored separately from primary device specifically for this scenario)
  3. Verify audio file integrity (playback test)
  4. If both copies lost, data is irrecoverable; this would constitute a serious incident requiring ethics committee notification if participant confidentiality cannot be verified
- Expected recovery time: 1–2 hours if backup available; irrecoverable if not

Scenario 4: Ransomware or malicious encryption

- Impact: All digitally accessible files encrypted or locked
- Recovery process:
  1. Immediately disconnect affected device from network to prevent spread
  2. Contact UP IT Services incident response team
  3. Restore from clean backup taken before infection (encrypted external hard drive offline backup is protected as it was not connected during infection)
  4. Scan all restored files for malware before reintroduction to network
  5. Review and strengthen security practices
- Expected recovery time: 1–3 days depending on infection scope and IT Services response

Scenario 5: Major institutional infrastructure failure (UP servers, network drives)

- Impact: Loss of access to primary and secondary backup locations
- Recovery process:
  1. Await UP IT Services restoration of institutional infrastructure (institutional disaster recovery

plan activated)
2. During outage, work from encrypted external hard drive if available
3. After restoration, verify integrity of restored institutional data against researcher's external backup
4. Report any discrepancies to UP IT Services

- Expected recovery time: Dependent on institutional disaster recovery capabilities; researcher maintains offline copy to ensure continuity

5. Third-Party Service Considerations UP-approved services (preferred): UP OneDrive (Microsoft 365)

- Jurisdiction: Data stored in Microsoft cloud infrastructure; UP's enterprise agreement ensures South African or EU data residency (verify with UP IT Services)
- Compliance: Covered by UP's institutional agreement; POPIA compliance addressed through Microsoft's South African data centre commitments and UP's data processing agreement
- Suitability: Appropriate for non-sensitive working files; automatic sync; version history; collaboration features

UP Figshare

- Jurisdiction: Figshare operates globally; UP's institutional subscription ensures data residency compliance
- Compliance: Designed for academic research data; supports embargo periods; DOI assignment; long-term preservation
- Suitability: Appropriate for final anonymised datasets; not for sensitive or raw identifiable data

UPSpace

- Jurisdiction: University of Pretoria institutional repository; local hosting
- Compliance: UP-controlled; fully compliant with South African legal requirements
- Suitability: Appropriate for thesis, publications, and final outputs

Services to avoid: Personal cloud services (Google Drive, Dropbox, iCloud, personal OneDrive)

- Reason for exclusion: Not covered by UP data processing agreements; data may reside in foreign jurisdictions (primarily USA); POPIA compliance uncertain; potential conflicts with institutional policy and funder requirements
- Risk: Loss of control over data residency; potential for subpoena or access by foreign authorities; institutional disciplinary consequences if sensitive data exposed

Consumer-grade external drives without encryption

- Reason for exclusion: Physical loss or theft exposes unencrypted data; violates POPIA security safeguard requirements for personal information
- Risk: Data breach; ethics committee censure; legal liability under POPIA

Unapproved collaboration platforms (WhatsApp, Telegram, personal email for data transfer)

- Reason for exclusion: No institutional oversight; insecure transmission; data retention uncontrolled
- Risk: Participant confidentiality breach; data integrity compromise; irreversible dissemination

6. Verification and Testing Monthly backup verification checklist:

- Confirm UP OneDrive sync is active (check sync icon status)
- Verify network drive accessibility and file listing

- Check encrypted external hard drive for recent backup dates
- Open and verify integrity of one randomly selected file from each backup location
- Review UP IT Services announcements for any infrastructure changes affecting storage

Quarterly recovery testing:

- Perform full restore of thesis draft from backup to test device
- Verify all files open correctly and are uncorrupted
- Time the restore process to confirm recovery time estimates
- Document any issues in research journal and notify supervisor

Annual comprehensive review:

- Review storage capacity usage against projections
- Assess whether backup strategy remains adequate for project phase (collection vs. analysis vs. write-up)
- Update incident recovery procedures based on any encountered issues or institutional policy changes
- Confirm all responsible parties remain available and informed of their roles

7. Reflection: Storage and Backup Learning from the Maule Case Study Maule's case study contains a cautionary tale about storage and data integrity:

- **Fragmented storage created analysis barriers:** The original Westminster data was "scattered among different tabs and documents," "incomprehensive," and stored in poorly formatted Excel files across multiple locations. This fragmentation was "off-putting and ultimately made it very time consuming to collate the data and get it into a usable format." At UP, my strategy of creating a single cloud-based master document from the outset, with clear version control, prevents this fragmentation.
- **No backup culture was evident:** Maule spent significant effort amalgamating "legacy spreadsheets" that had accumulated over five years without apparent systematic backup or version control. The risk of data loss through file corruption, accidental deletion, or staff turnover was high. My multiple-copy strategy with institutional and offline backups ensures no single point of failure.
- **Manual processes introduced risk:** Data was "input into a template that could not be easily analysed," requiring "one dedicated staff member on each library site" to manually transfer figures. This manual handling created multiple opportunities for error and loss. Where possible, my research will extract data directly from UP's automated systems rather than relying on manual transcription; where manual collection is necessary (e.g., headcounts), immediate digital backup replaces paper logbooks.
- **The cloud transition was transformative but late:** Maule's team moved to "one cloud-based master document" only after years of problematic local storage. My research begins with cloud-based, automatically synced storage, embedding good practice from project initiation rather than retrofitting it after problems emerge.

How will you manage access and security?

- Only researcher + supervisor access during study
- Controlled access after project
- Sensitive data restricted

1. Risks to Data Security and Management Strategies This research handles multiple data types with varying security risk profiles. The following risks have been identified with corresponding

mitigation strategies. **Risk Category: Unauthorised access to identifiable student usage data**

- Description: Raw entry logs or swipe card data linked to student IDs could expose individual library usage patterns, revealing sensitive behaviours such as late-night study habits, campus movements, or academic stress indicators
- Likelihood: Medium (institutional systems have some protection, but researcher access creates additional exposure point)
- Impact: High (POPIA violation; potential harm to students if patterns revealed; reputational damage to UP; ethics committee censure)
- Management strategy:
    - Extract only aggregate data from UP systems where possible; if individual records required for analysis, immediate anonymisation upon extraction
    - Store raw identifiable data only on encrypted institutional drives with access limited to researcher and supervisor
    - Never download raw identifiable data to personal devices or unencrypted media
    - Destroy raw identifiable data immediately after anonymisation and verification (maximum 6 months retention)
    - Document all access in audit log

Risk Category: Disclosure of sensitive institutional cost information

- Description: Detailed salary figures, vendor contracts, or operational budgets could be misused competitively or cause industrial relations issues if disclosed
- Likelihood: Low (data held within UP; researcher is internal)
- Impact: Medium-High (financial and political sensitivity; potential procurement disadvantage; staff relations impact)
- Management strategy:
    - Aggregate cost data to site or service level before analysis; never retain individual salary figures in research files
    - Redact vendor names and contract specifics; use generic descriptors (e.g., "security services provider" rather than company name)
    - Pre-publication review by UP Library Services and Finance Department for any outputs containing cost information
    - Mark all cost-related files as "UP Confidential – Research Use Only"
    - Store in access-controlled folder separate from other research data

Risk Category: Breach of participant confidentiality in qualitative research

- Description: Focus group participants, survey respondents, or comment card authors could be identified through direct quotes, contextual details, or combination of demographic information
- Likelihood: Medium (qualitative data inherently rich in detail; focus groups reduce individual control over disclosure)
- Impact: High (breach of trust; potential harm to participants; ethics violation; legal liability under POPIA)
- Management strategy:
    - Obtain explicit consent for recording, transcription, and use of anonymised quotations
    - Apply pseudonymisation immediately after transcription; destroy audio linking keys after verification
    - Contextual anonymisation: generalise identifying details in transcripts and quotes (specific courses become faculties; rare combinations of characteristics are suppressed)
    - K-anonymity review: ensure no combination of reported variables in qualitative themes identifies fewer than 5 individuals
    - Store qualitative data in encrypted, access-controlled location separate from quantitative data
    - Informed consent explicitly notes that focus group confidentiality cannot be guaranteed for disclosures made within the group

Risk Category: Physical loss or theft of data collection devices

- Description: Laptop, external hard drive, or mobile recording equipment containing research data lost, stolen, or damaged during fieldwork or travel
- Likelihood: Medium (researcher moves between campuses, libraries, and home/institutional locations)
- Impact: Variable (high if device contains unencrypted identifiable data; low if only encrypted or cloud-synced working copies)
- Management strategy:
    - All devices encrypted at rest (BitLocker for Windows, FileVault for Mac, device encryption for mobile)
    - No identifiable raw data stored solely on portable devices; immediate transfer to institutional secure storage
    - External hard drives encrypted and physically stored separately from laptop when not in use
    - Mobile recording devices (if used for focus groups) password-protected; audio transferred immediately after session; device wiped after transfer
    - Report all loss/theft immediately to UP IT Services and supervisor; activate incident response

Risk Category: Cyber attack (ransomware, phishing, malware)

- Description: Researcher's device or institutional account compromised, leading to data encryption, exfiltration, or destruction
- Likelihood: Medium (ubiquitous threat; academic institutions frequent targets)
- Impact: High (potential loss of all digital data; exposure of sensitive information; irreversible breach of participant confidentiality)
- Management strategy:
    - UP-managed antivirus and endpoint protection on all devices
    - Regular software updates and security patches via UP IT Services
    - Phishing awareness training; verification of unexpected communications
    - Multi-factor authentication on all UP accounts (OneDrive, network drives, email)
    - Offline encrypted backup (external hard drive) not connected to network except during scheduled backup (protects against ransomware propagation)
    - No clicking of suspicious links or downloading unverified attachments on research devices

Risk Category: Insecure data transfer from field sites

- Description: Data collected at UP library sites (manual headcounts, focus group recordings, comment cards) transferred to main secure systems through insecure channels, exposing data in transit or on intermediary devices
- Likelihood: Medium (field collection inherently creates temporary data vulnerability)
- Impact: Medium-High (interception of personal information; loss of irreplaceable qualitative data)
- Management strategy:
    - Immediate digital capture where possible (tablet-based headcount forms rather than paper; digital audio recorders rather than analogue)
    - Transfer via encrypted channels only (UP VPN for remote access; encrypted email for small files; secure file transfer for large files)
    - No use of personal email, WhatsApp, or unencrypted messaging for data transfer
    - If paper comment cards collected, scanned immediately to encrypted institutional storage; originals stored securely and shredded after scanning verification
    - Mobile devices used for collection encrypted and password-protected; remote wipe capability enabled

Risk Category: Accidental disclosure through collaboration platforms

- Description: Shared documents, cloud folders, or collaborative tools inadvertently expose sensitive data to unauthorised viewers

- Likelihood: Low-Medium (collaboration requires sharing; misconfiguration possible)
- Impact: Medium (depends on data sensitivity exposed)
- Management strategy:
  - Share only anonymised or aggregated data through collaboration platforms
  - Explicit permission settings on all shared folders (restricted to named individuals, not "anyone with link")
  - Regular review of sharing permissions; revoke access when collaboration phase ends
  - Supervisor access granted to specific folders only, not entire project directory
  - No sharing of raw identifiable data through any platform

2. Access Control Framework Access to research data is controlled through a tiered system based on data sensitivity and role necessity. **Tier 1: Raw identifiable operational data (highest restriction)**

- Data covered: Swipe card logs with student IDs; individual cost spreadsheets with salary data; signed consent forms with names and contact details
- Access granted to: Researcher only (primary access); supervisor (emergency/oversight access with documented justification)
- Authentication required: UP network credentials plus multi-factor authentication; encrypted device access passwords
- Storage location: Encrypted folder on UP network drive; not synced to cloud
- Access logging: Yes – UP IT Services can audit network drive access; researcher maintains manual log of any supervisor access
- Retention limit: Raw identifiable data destroyed after anonymisation and verification (maximum 6 months)

Tier 2: Processed working data with pseudonyms (restricted access)

- Data covered: Coded qualitative transcripts; anonymised but granular usage data (hourly counts by site); working analysis files
- Access granted to: Researcher; supervisor (for guidance and review)
- Authentication required: UP network credentials plus multi-factor authentication
- Storage location: Password-protected folder on UP network drive; encrypted external hard drive backup
- Access logging: Informal – supervisor access noted in research journal
- Retention limit: Retained for duration of research plus standard UP policy period (typically 5 years); then deposited as restricted dataset or destroyed

Tier 3: Anonymised aggregate datasets (controlled access)

- Data covered: Fully aggregated usage statistics; anonymised qualitative themes; final visualisations; thesis drafts
- Access granted to: Researcher; supervisor; UP Library Services (for verification and service improvement); external examiners (thesis only)
- Authentication required: UP network credentials for institutional access; examiner access via secure institutional channel
- Storage location: UP network drive; UP OneDrive for collaboration; UPSpace for thesis; UP Figshare for datasets (with appropriate embargo)
- Access logging: Repository-level logging (UPSpace, Figshare track downloads)
- Retention limit: Permanent preservation in UP repositories where appropriate; published outputs retained indefinitely

Tier 4: Public outputs (open access)

- Data covered: Published thesis; journal articles; conference presentations; fully anonymised datasets released under Creative Commons licence
- Access granted to: Public

- Authentication required: None for open access materials; repository registration may be required for some Figshare features
- Storage location: UPSpace; UP Figshare; publisher platforms; conference repositories
- Access logging: Standard repository analytics
- Retention limit: Permanent

Access control technical measures:

- **Multi-factor authentication (MFA):** Required on all UP accounts (OneDrive, email, network access, VPN)
- **Password policy:** Unique strong passwords for all research accounts; no password reuse across personal and institutional accounts; password manager recommended (UP-approved if available)
- **Device encryption:** Full disk encryption on all devices accessing research data (laptop, external drives, mobile collection devices)
- **Screen locks:** Automatic activation after 5 minutes of inactivity; manual lock when leaving device unattended
- **Physical security:** Devices not left visible in vehicles or public spaces; external drives stored in locked cabinet when not in use; home office storage secured

3. Secure Collaboration Access Collaboration on this research involves limited but necessary data sharing. The following protocols ensure secure access for each collaborator type. **Supervisor**

- Data accessed: All tiers as appropriate for guidance (Tier 1 only with documented justification and researcher notification)
- Method of access: UP network drive shared folder (Tier 2-3); specific file sharing via encrypted email for review (Tier 1 with redaction where possible)
- Security measures: MFA on supervisor's UP account; confidentiality agreement implicit in supervisory relationship; explicit discussion of data sensitivity at project commencement
- Duration of access: Throughout research period; access revoked upon project completion except for thesis repository access
- Revocation process: Shared folder permissions removed; supervisor notified of any Tier 1 data destruction

UP Library Services (data provider and stakeholder)

- Data accessed: Tier 3 and 4 only (aggregate findings, recommendations, final report); no access to Tier 1 or 2 raw or processed participant data
- Method of access: Presentation of findings in meetings; shared summary documents via UP email; final report deposited in institutional repository
- Security measures: Data sharing agreement specifies no onward sharing without researcher consent; UP institutional confidentiality obligations apply
- Duration of access: During consultation phases and for service improvement implementation; ongoing access to published outputs
- Special arrangement: Pre-publication review of any outputs containing cost data or operational recommendations

UP Finance Department (if cost data consultation required)

- Data accessed: Their own original data only (not researcher's compiled datasets); aggregate research outputs showing cost-effectiveness calculations
- Method of access: Meetings with researcher presenting aggregate figures; no direct system access to research files
- Security measures: Existing institutional confidentiality obligations; no external sharing of UP financial information
- Duration of access: Consultation period only

External examiners (thesis examination)

- Data accessed: Tier 3 and 4 only (thesis, anonymised appendices, public datasets)
- Method of access: UPSpace thesis repository; or secure institutional file transfer if embargoed material requires pre-publication examination
- Security measures: Examiner appointment through UP formal process; confidentiality obligations in examination contract; no onward sharing permitted
- Duration of access: Examination period only

Potential future collaborators (if research expands)

- Data accessed: Tier 3 and 4 only; new data sharing agreement required for any Tier 2 access
- Method of access: Formal collaboration agreement through UP research office; ethical review if human participant data involved; POPIA compliance verification
- Security measures: Institutional affiliation and ethics clearance verification; signed confidentiality agreement; access limited to specific agreed datasets

Prohibited collaboration practices:

- No sharing of Tier 1 or Tier 2 data with peer researchers, family, or friends for "advice" or "proofreading"
- No screen sharing of sensitive data during informal video calls (Teams or Zoom with institutional account only for formal supervision)
- No forwarding of data-containing emails to personal accounts or non-UP addresses
- No granting of "anyone with link" access to cloud documents containing anything above Tier 3

4. Safe Transfer of Field Data into Secured Systems Data collection occurs at multiple UP library sites across six campuses. The following protocols ensure safe transfer from field to main secured systems. **Field data type: Manual headcounts (if automated systems unavailable or supplementary)**

- Collection method: Tablet or smartphone with encrypted device; pre-loaded digital form (Excel, Google Forms with institutional account, or UP-approved survey tool); no paper logbooks unless digital failure
- Immediate security: Device password-protected; form autosaves to cloud if connectivity available; local encrypted storage if offline
- Transfer protocol:
    1. At end of each counting session (or hourly if extended), verify data saved
    2. If offline, connect to secure UP Wi-Fi or VPN before sync
    3. Sync to UP OneDrive or approved cloud form platform
    4. Verify sync success by checking file on network drive from second device
    5. Clear local device cache after verified transfer (if using personal device for collection)
- Prohibited practices: Email headcount sheets to personal email; screenshot and send via WhatsApp; save to unencrypted personal device storage without immediate transfer

Field data type: Focus group audio recordings

- Collection method: Dedicated digital audio recorder (institutional or research-purchased) with password protection; or encrypted recording app on institutional tablet
- Immediate security: Recorder password-protected; kept on researcher person at all times; not left unattended in library spaces
- Transfer protocol:
    1. Immediately after focus group conclusion (before leaving site), verify recording integrity (playback check of first 30 seconds)
    2. Transfer to encrypted laptop or institutional tablet via cable (not wireless if avoidable; if wireless, use secure UP Wi-Fi with VPN)
    3. Upload to encrypted folder on UP network drive within 2 hours of collection
    4. Verify upload by accessing from different device
    5. Delete recording from portable recorder/device only after verification

6. Store recorder in locked bag/cabinet when not in use
- Backup protocol: If network upload impossible from field site, store on encrypted external hard drive (carried separately from recording device) and transfer within 24 hours
- Prohibited practices: Store recording on unencrypted device; leave recorder in library or vehicle; delay transfer beyond 24 hours; share recording with transcription service without confidentiality agreement

Field data type: Paper comment cards or feedback forms

- Collection method: Physical cards in sealed collection boxes at library sites; researcher empties boxes on scheduled rounds
- Immediate security: Boxes positioned in staffed areas where possible; researcher carries collected cards in sealed envelope
- Transfer protocol:
  1. Transport cards directly to secure office/study space; no intermediate stops
  2. Scan cards immediately upon arrival using UP network-connected scanner
  3. Save scans to encrypted folder on UP network drive
  4. Verify scan quality and completeness
  5. Shred original paper cards after scanning verification (within 48 hours)
  6. If scanning impossible immediately, store in locked cabinet until scanning possible
- Prohibited practices: Leave cards in unattended bag or vehicle; transport to home office if institutional scanning available; retain paper originals unnecessarily

Field data type: Informal observations or researcher notes

- Collection method: Field notebook or encrypted note-taking app on institutional device
- Immediate security: Notebook kept on person; device password-protected
- Transfer protocol:
  1. Transcribe or sync notes to UP network drive within 24 hours
  2. If notebook used, scan or photograph pages to encrypted storage
  3. Store notebook securely when not in use
- Prohibited practices: Leave notebook containing observational details about identifiable individuals in public spaces; share informal observations on social media or with peers in identifiable form

Remote or off-campus data collection (if surveying distance students or online participants)

- Collection platform: UP-approved survey tool (e.g., Qualtrics if available through UP; or REDCap for research data)
- Security: Platform covered by UP institutional agreement; data encrypted in transit and at rest; hosted in compliant jurisdiction (verify with UP IT Services)
- Transfer: Data exports directly to UP network drive; no download to personal device as intermediate step
- Prohibited practices: Use of SurveyMonkey free tier, Google Forms personal account, or other non-institutional platforms for identifiable or sensitive responses

5. Formal Standards and Compliance This research will comply with the following formal standards and institutional policies. **ISO/IEC 27001: Information Security Management Systems**

- Application: While UP as an institution may or may not hold full ISO 27001 certification, the researcher will apply its principles: risk assessment, security controls, access management, incident response, and continuous improvement
- Specific alignment: Asset management (data inventory); access control (tiered framework); cryptography (encryption at rest and in transit); operations security (backup and malware protection); communications security (secure transfer protocols)

Protection of Personal Information Act (POPIA) 2013

- Application: Legal requirement for all processing of personal information in South Africa
- Compliance measures documented: Accountability (researcher as responsible party); processing limitation (minimum necessary collection); purpose specification (research objective only); further processing limitation (no unrelated use); information quality (verification and correction rights); openness (privacy notices); security safeguards (all measures above); data subject participation (access, correction, deletion rights)

University of Pretoria Information Security Policy

- Application: Institutional policy governing all use of UP IT resources and handling of institutional data
- Compliance: Use only UP-approved services; comply with access controls; report incidents; participate in security training; adhere to records management requirements

University of Pretoria Research Data Management Policy

- Application: Governs retention, sharing, and security of research data
- Compliance: DMP approved by supervisor; data retained per policy periods; appropriate access controls; repository deposition for long-term preservation

University of Pretoria Research Ethics Policy

- Application: Governs ethical conduct in research involving human participants
- Compliance: Ethics clearance obtained; consent procedures followed; participant rights respected; adverse events reported; data handling meets ethical standards

6. Incident Response and Reporting Security incident classification and response: Minor incident (suspected unauthorised access attempt, phishing email received, minor device malfunction)

- Immediate action: Do not click links or enter credentials; disconnect device if malware suspected; report to UP IT Services helpdesk
- Documentation: Record in research journal; note date, time, nature of incident, and response taken
- Reporting: Inform supervisor within 24 hours; IT Services handles technical response
- Recovery: Follow IT Services guidance; change passwords if credentials potentially compromised; verify no data accessed

Moderate incident (lost encrypted device, accidental email to wrong recipient containing aggregate data, temporary loss of access to systems)

- Immediate action: Attempt recovery if possible (recall email, remote wipe device); change passwords; notify affected parties if data disclosure occurred
- Documentation: Detailed incident log including what data was involved, potential exposure scope, and timeline
- Reporting: Inform supervisor immediately; report to UP IT Services; if human participant data involved, notify UP Research Ethics Committee within 48 hours
- Recovery: Device replacement; data restoration from backup; review of access logs to confirm no unauthorised access; corrective measures to prevent recurrence

Major incident (data breach involving identifiable personal information, ransomware infection, theft of unencrypted device containing raw data, deliberate unauthorised access)

- Immediate action: Isolate affected systems; preserve evidence; do not attempt to conceal
- Documentation: Comprehensive incident report with full timeline, data involved, and response

actions
- Reporting: Inform supervisor immediately; report to UP IT Services and Information Security Office; mandatory reporting to Information Regulator under POPIA within 72 hours if personal information compromised; notify UP Research Ethics Committee; if funded, notify funder
- Recovery: Institutional incident response team activated; forensic investigation; affected individuals notified if required by POPIA; data restoration from clean backups; potential suspension of research activities pending investigation; corrective action plan required before resumption

7. Reflection: Security Learning from the Maule Case Study Maule's case study, while not explicitly focused on security, reveals vulnerabilities that inform this protocol:

- **Manual data collection created physical security risks:** Security officers carried physical logbooks during hourly patrols, manually input data into Excel files, and stored printed logbooks and digital files on "staff shared" drives. At no point was encryption, access control, or secure transfer mentioned. The risk of logbook loss, unauthorised access to shared drives, or misplacement of printed records was unaddressed. My protocol replaces paper with encrypted digital capture, eliminates intermediate manual transcription where possible, and restricts access to named individuals rather than open shared drives.
- **No incident response capability was evident:** If a logbook was lost, a file corrupted, or data inappropriately accessed, there was no framework for detection or response. My tiered access control with logging, regular verification, and explicit incident classification ensures that breaches can be identified and managed promptly.
- **Data fragmentation increased exposure surface:** Data scattered across "different tabs and documents," multiple sites, and various staff members' input responsibilities multiplied the points where data could be exposed, lost, or inconsistently protected. My centralised encrypted storage with controlled access reduces this surface.
- **The absence of formal standards was implicit:** Maule's project achieved significant outcomes despite informal data practices, but this was possible because the data was relatively low-sensitivity operational information. My research, involving identifiable student patterns and sensitive institutional finances, cannot rely on informal security. The explicit adoption of POPIA compliance, ISO 27001 principles, and UP institutional policies provides a defensible and reproducible security posture.

Selection and Preservation

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

- Long-term storage (5–10 years)
- Data archived for future research

1. Data Retention and Destruction Obligations This research generates and uses multiple data types, each subject to different contractual, legal, and regulatory retention or destruction requirements. **Raw identifiable operational data (swipe card logs with student IDs, individual cost spreadsheets)**

- Retention obligation: Destruction required under POPIA once no longer necessary for the lawful purpose for which it was collected
- Regulatory basis: Protection of Personal Information Act (POPIA) 2013, Section 14 (retention and restriction of records); UP records management policy for personal information

- Retention period: Maximum 6 months from date of collection or immediately after successful anonymisation and verification, whichever is earlier
- Destruction method: Secure digital deletion (not just recycle bin; use secure wiping software that overwrites data); physical documents shredded using UP-approved disposal services
- Verification: Researcher confirms destruction in research journal; supervisor spot-checks if raw data access was previously logged
- Exception: If anonymisation fails or data quality issues emerge requiring re-verification, retention may extend to 12 months with documented justification and ethics committee notification

Signed consent forms and participant contact details

- Retention obligation: Retention required for ethics audit and dispute resolution purposes
- Regulatory basis: UP Research Ethics Policy; POPIA (participants have right to know what data was collected and how it was used)
- Retention period: Minimum 5 years from project completion (thesis submission or publication date, whichever is later); maximum per UP policy (typically 5–7 years)
- Storage: Physical consent forms in locked cabinet in supervisor's office or secure institutional storage; digital contact sheets in encrypted folder on UP network drive with access limited to researcher and supervisor
- Destruction: After retention period, secure shredding of physical forms; secure deletion of digital contact records
- Exception: If participant withdraws consent, their individual consent form and contact details destroyed immediately upon withdrawal, with documentation of destruction in research journal

Anonymised aggregate usage datasets

- Retention obligation: Retention encouraged for validation, longitudinal comparison, and institutional benchmarking; no legal requirement for destruction
- Regulatory basis: UP Research Data Management Policy; NRF open science expectations if funded; institutional value for library service improvement
- Retention period: Permanent preservation in UP Figshare or UPSpace with appropriate metadata
- Rationale: These datasets contain no identifiable information and represent significant institutional investment in data collection; they have ongoing value for UP Library Services planning and for comparative research across South African universities

Coded qualitative transcripts and themes

- Retention obligation: Retention required for thesis verification and potential follow-up analysis; destruction of linking keys required to maintain participant confidentiality
- Regulatory basis: UP Research Ethics Policy (data must be available for audit); POPIA (pseudonymised data retained only if necessary and proportionate)
- Retention period: Minimum 5 years from project completion; pseudonym linking keys destroyed after verification period (maximum 12 months from transcription)
- Storage: Encrypted institutional storage during active research; deposited as restricted dataset in UP Figshare after project completion (access restricted to researcher, supervisor, and authorised auditors)
- Destruction: Linking keys (document connecting pseudonyms to real identities) destroyed after verification; transcripts retained in pseudonymised form only

Audio recordings of focus groups

- Retention obligation: Destruction recommended once transcription and verification complete; retention only if explicitly consented for specific reuse
- Regulatory basis: POPIA (voice recordings are personal information; retention must be justified); participant consent terms
- Retention period: Maximum 12 months from date of recording or immediately after transcription verification and pseudonymisation, whichever is earlier

- Destruction method: Secure digital deletion with overwriting; physical media (SD cards, recorder memory) reformatted and overwritten
- Exception: If participants explicitly consented to audio retention for specific purposes (e.g., language analysis, verification of transcription accuracy by external party), retention may extend to 5 years with documented consent and secure storage

Research outputs (thesis, publications, presentations)

- Retention obligation: Permanent retention required for academic record and institutional repository
- Regulatory basis: UP Higher Degrees policy; UP Open Access Policy; publisher agreements
- Retention period: Indefinite; thesis permanently archived in UPSpace; publications retained per publisher terms and institutional repository policies
- Storage: UPSpace (thesis); UP Figshare (datasets); publisher platforms (articles); conference repositories (presentations)

Institutional cost data (aggregate, redacted)

- Retention obligation: Retention governed by UP confidentiality requirements; destruction if commercially sensitive and not consented for open release
- Regulatory basis: UP intellectual property policy; data sharing agreement with UP Library Services and Finance Department
- Retention period: If approved for open release: permanent in UP Figshare with embargo if required. If restricted: retained in encrypted institutional storage for 5 years then reviewed; destruction if no ongoing institutional need and researcher does not require for further analysis
- Exception: UP Library Services may request retention of aggregate cost-effectiveness models for internal planning; this governed by separate institutional records management schedule

2. Decision Framework for Data Retention Beyond Obligations Beyond mandatory retention and destruction requirements, the following criteria guide decisions on what additional data to preserve. **Criterion 1: Validation and reproducibility**

- Data to retain: Fully anonymised master datasets; analysis scripts or spreadsheets with documented formulas; methodology narrative; variable definitions and data dictionary
- Rationale: Other researchers or examiners must be able to verify findings; UP may wish to replicate analysis in future years; researcher may need to defend findings during examination or peer review
- Retention period: Minimum 10 years; preferably permanent with repository deposition
- Effort required: Low – these are final outputs already prepared for analysis; minimal additional formatting needed for repository deposit

Criterion 2: Future research reuse

- Data to retain: Anonymised aggregate usage data by site, hour, day, and term week; anonymised qualitative themes and coded categories (not raw transcripts); cost-effectiveness calculation methodology
- Foreseeable research uses:
    - Longitudinal comparison: Future researchers at UP may compare 2026 usage patterns with subsequent years to assess trends in student study behaviour, impact of digital resource expansion, or post-pandemic changes
    - Cross-institutional benchmarking: Other South African universities (Wits, UCT, Stellenbosch) may compare their library usage patterns and cost structures against UP data
    - Methodological development: The cost-effectiveness assessment framework may be adapted for other UP services (computer labs, study spaces, sports facilities) or other institutions
    - Teaching case study: The data journey, stakeholder consultation process, and service change outcomes may be used in LIS education to teach evidence-based decision making
- Rationale: Data collection was resource-intensive (spanning multiple sites, semesters, and data

types); destroying reusable datasets wastes institutional investment; open data supports academic community
- Retention period: Permanent in repository with appropriate licensing (CC BY for aggregate data; CC BY-NC-SA for qualitative themes)
- Effort required: Medium – requires preparation of data dictionary, documentation of methodology, formatting for repository standards, creation of README files; estimated 2–3 days of additional work

Criterion 3: Institutional service improvement

- Data to retain: All anonymised datasets; stakeholder consultation records; visualisations and presentations used for decision support
- Foreseeable uses:
  - UP Library Services strategic planning: Annual review of opening hours; budget allocation decisions; new campus or library site planning
  - UP Estates and Facilities: Space utilisation planning; energy and security cost modelling
  - UP Student Experience Committee: Evidence for student satisfaction initiatives; response to student representative feedback
- Rationale: Primary purpose of research was to inform UP decision-making; retaining data ensures evidence base remains available for ongoing service management
- Retention period: Minimum 10 years; reviewed at 5-year intervals for continued relevance
- Effort required: Low-Medium – requires institutional stakeholder review of outputs before repository deposit; potential redaction of politically sensitive findings

Criterion 4: Economic viability of retention

- Assessment: UP provides institutional repository storage (UPSpace, Figshare) at no direct cost to researcher; cloud storage (OneDrive) included in institutional subscription; no incremental storage fees for datasets under 10 GB
- Conclusion: Retention of all anonymised datasets is economically viable; no cost-driven destruction justified
- Exception: If project expands unexpectedly (video focus groups, large-scale continuous monitoring), storage costs may require review with supervisor and UP IT Services

Criterion 5: Effort to prepare for sharing and preservation

- Low effort (retain): Anonymised aggregate quantitative data; final visualisations; methodology documents; thesis – already in shareable formats (CSV, PDF, DOCX)
- Medium effort (retain with preparation): Coded qualitative themes require anonymisation review, pseudonym verification, and contextual generalisation before sharing; estimated 1–2 days
- High effort (evaluate case by case): Raw qualitative transcripts require extensive anonymisation and participant re-consent for open sharing; may be retained as restricted access only rather than open dataset; if open sharing required, effort estimated 3–5 days for full preparation
- Decision: High-effort data retained in restricted form; open sharing only pursued if significant external demand demonstrated or funder mandates

Data not retained beyond minimum obligations:

- Raw identifiable swipe card extracts (destroyed after anonymisation)
- Individual salary figures and vendor contracts (never retained by researcher; aggregate only)
- Failed or superseded analysis versions (deleted during active research; version control retains only meaningful iterations)
- Informal researcher notes and observations not incorporated into methodology narrative (destroyed after write-up)
- Duplicate or redundant file copies created during analysis (cleaned up during project close-out)

3. Foreseeable Research Uses for Retained Data Immediate use (0–2 years): Thesis

examination and publication

- Examiners access thesis, anonymised appendices, and methodology via UPSpace
- Journal articles submitted using aggregate data and qualitative themes; peer reviewers may request dataset access for verification
- Conference presentations using visualisations and key findings

Short-term use (2–5 years): Institutional service implementation and review

- UP Library Services uses findings to implement opening hours changes; monitors outcomes against baseline data
- Annual reviews compare new usage patterns to 2026 baseline
- Student Representative Council references research in ongoing advocacy

Medium-term use (5–10 years): Longitudinal comparison and benchmarking

- Future UP researchers replicate study to assess whether 2026 recommendations remain valid
- Comparative studies across South African universities using UP as benchmark case
- National library association (LIASA) or higher education body uses anonymised aggregate data for sector-wide planning

Long-term use (10+ years): Historical analysis and teaching

- Historical study of academic library evolution in digital age
- Teaching case study in LIS programmes: evidence-based decision making, stakeholder consultation, data-driven service change
- Methodological reference for cost-effectiveness studies in other educational or public service contexts

4. Retention and Preservation Timeline Active research phase (Months 0–18)

- All data retained in active working storage: UP network drives, encrypted external backup
- Raw identifiable data: accessible for verification and re-analysis if needed
- Version control maintained for all analysis files
- Monthly backup verification

Thesis completion and examination (Months 18–24)

- Raw identifiable data: destroyed after final anonymisation verification and thesis submission
- Pseudonym linking keys: destroyed after transcription verification complete
- Audio recordings: destroyed after transcription verification (unless explicit consent for extended retention)
- Consent forms and contact details: transferred to secure long-term storage (locked cabinet / encrypted archive)
- Anonymised master datasets: prepared for repository deposit; metadata created
- Thesis: deposited in UPSpace

Post-examination preservation (Months 24–60)

- Anonymised datasets: deposited in UP Figshare with appropriate licences and embargo if required
- Coded qualitative transcripts: deposited as restricted dataset (access requires application)
- Consent forms: retained in secure storage per ethics policy
- Cost aggregate data: reviewed with UP Library Services for open release vs. restricted retention
- Researcher access: maintained to all deposited data for own future use

Long-term institutional preservation (5+ years)

- UPSpace thesis: permanent
- UP Figshare datasets: permanent; reviewed periodically for format obsolescence; migrated to

current formats if necessary
- Restricted qualitative data: reviewed at 5-year intervals; if no ongoing institutional need and no researcher interest, may be destroyed with ethics committee approval
- Consent forms: destroyed after minimum retention period (typically 5–7 years) unless ethics audit or dispute requires extension

Format preservation considerations:

- Current formats: CSV for tabular data; PDF/A for documents; MP3 for audio (during brief retention); MP4 for video (if applicable); DOCX for working documents; XLSX for spreadsheets
- Preservation formats: CSV maintained as primary preservation format for data (open, non-proprietary); PDF/A for final documents; migration from MP3 to WAV or FLAC if audio retained long-term; XLSX archived with CSV equivalents to ensure accessibility if Excel becomes obsolete
- Migration responsibility: UP repository services (Figshare, UPSpace) monitor format obsolescence; researcher notified if migration required; researcher responsible for providing migration-ready files if original formats become unreadable

5. Preparation Effort for Sharing and Preservation The following additional tasks are required to prepare data for long-term sharing and preservation, beyond the minimum needed for thesis completion. **Task 1: Data dictionary and metadata creation**

- Effort: 1 day
- Description: Complete documentation of all variables, codes, units, and assumptions; align with DDI and Dublin Core standards; create README.txt
- Benefit: Enables any future user to understand and reuse data without researcher involvement

Task 2: Anonymisation verification and documentation

- Effort: 2–3 days
- Description: Review all qualitative outputs for residual identification risk; verify k-anonymity in quantitative cross-tabulations; document all anonymisation decisions and transformations applied
- Benefit: Ensures ethical compliance for open sharing; protects participants; demonstrates due diligence

Task 3: Format standardisation and migration

- Effort: 1 day
- Description: Convert all final datasets to CSV; create PDF/A versions of key documents; verify file integrity; package for repository deposit
- Benefit: Ensures long-term accessibility; reduces dependency on proprietary software

Task 4: Repository deposit and licensing

- Effort: 1 day
- Description: Upload to UP Figshare; complete metadata fields; assign DOI; apply appropriate Creative Commons licences; set embargo if required; link to thesis in UPSpace
- Benefit: Makes data citable and discoverable; establishes permanent record; supports research impact tracking

Task 5: Stakeholder review and clearance

- Effort: 2–3 days (dependent on stakeholder responsiveness)
- Description: Submit aggregate cost data and operational recommendations to UP Library Services for pre-publication review; incorporate feedback; obtain written clearance for open release of institutional data
- Benefit: Maintains institutional trust; ensures accuracy of institutional representations; prevents inadvertent disclosure of sensitive information

Total estimated additional effort: 7–9 days This effort is proportionate to the research scale and the value of the datasets for future reuse. The researcher will schedule this preparation in the final months of the project, parallel to thesis writing. **6. Reflection: Retention and Preservation Learning from the Maule Case Study** Maule's case study provides powerful lessons on what happens when retention and preservation are not planned from the outset.

- **Valuable data was almost lost to poor retention practices:** Five years of comprehensive hourly headcount data had been "collected for over five years, but had never been interrogated or analysed before." The data survived physically but was functionally lost due to fragmentation, poor formatting, and lack of documentation. My protocol ensures that data is preserved in immediately usable form with comprehensive metadata, preventing this functional loss.
- **No preservation format standards were applied:** The original data was trapped in poorly structured Excel files with "lots of empty rows and columns, spread across multiple tabs and documents." This format was not preservation-friendly and required extensive reformatting before analysis. My use of CSV as a preservation format, with clear standardised structure from collection, avoids this trap.
- **The effort to prepare for sharing was underestimated:** Maule noted that "compiling all the data, qualitative and quantitative, took more time than I first envisaged and was a learning curve for future projects." She was fortunate that the data gathering stage had been thorough enough that she "didn't have to go back and search for additional data." My explicit scheduling of 7–9 preparation days, with a checklist of tasks, ensures this effort is anticipated rather than discovered painfully.
- **Stakeholder consultation records were preserved informally:** Maule's detailed account of presentations to Student Union, registry managers, and the Student Experience Committee survives because she wrote the case study. But there is no indication that minutes, presentation slides, or feedback forms were systematically archived. My protocol explicitly retains consultation records as part of the institutional memory, not merely as the researcher's personal notes.
- **No long-term repository deposit was mentioned:** The Westminster data's future after Maule's analysis is unclear. My deposit in UP Figshare with DOI assignment ensures the UP data has a permanent, citable home that outlasts the researcher's affiliation or memory.

The critical insight is that **retention is not merely about keeping files**; it is about **keeping files usable, understandable, and ethically sound for future contexts that cannot be fully anticipated**. The 2026 UP library usage data may seem specific to a moment, but in 2036 it may illuminate trends in student behaviour, the impact of digital transformation, or the evolution of campus planning. Preserving it properly is an act of institutional stewardship, not merely compliance.

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

1. Repository and Archive Selection This research will use a multi-tiered repository strategy, matching different data types and preservation needs to appropriate institutional and discipline-specific repositories. **Tier 1: UP Figshare – Primary Research Data Repository**

- Description: UP Figshare is the University of Pretoria's dedicated research data repository, currently hosting 169 datasets. It provides DOI assignment, versioning, embargo capabilities, and integration with international discovery systems.
- Data deposited: Anonymised aggregate usage datasets; coded qualitative themes; cost-effectiveness calculation methodologies; data dictionaries; final visualisations; analysis scripts
- Rationale: Institutionally supported; designed for research data; citable via DOI; aligns with UP open science policy; already established with existing datasets demonstrating institutional commitment
- Access level: Open access for fully anonymised aggregate data (CC BY 4.0); restricted access for coded qualitative themes requiring application (CC BY-NC-SA 4.0); embargoed release for data under institutional review (embargo up to 12 months)

- Metadata standard: DataCite Metadata Schema for DOI registration; Dublin Core for discovery; DDI for complex methodological documentation
- Preservation commitment: UP maintains institutional subscription; data preserved indefinitely; format migration monitored by repository administrators
- URL: https://figshare.up.ac.za

Tier 2: UPSpace – Institutional Repository for Theses and Publications

- Description: UPSpace is the University of Pretoria's institutional repository for scholarly outputs, including theses, dissertations, journal articles, and conference papers.
- Data deposited: Final thesis; published journal articles; conference presentations; research summaries for institutional stakeholders
- Rationale: Mandatory deposit for UP higher degrees; permanent archiving; open access compliance; integration with national and international thesis discovery systems
- Access level: Open access for thesis (after any embargo period); publisher-determined access for articles; institutional access for pre-publication outputs
- Metadata standard: Dublin Core; OAI-PMH harvesting for interoperability
- Preservation commitment: UP Library Services maintains UPSpace; permanent preservation; regular format migration and integrity checking
- URL: https://upspace.up.ac.za

Tier 3: UP Network Drives and Institutional Cloud – Active Research Storage

- Description: Secure institutional storage for active research data and restricted-access materials not yet ready for public repository deposit.
- Data held: Raw identifiable data during active research (maximum 6 months); pseudonymised working datasets during analysis; consent forms and contact details during retention period; stakeholder consultation records
- Rationale: Secure, access-controlled environment for sensitive data; integrated with UP authentication systems; backed up by institutional IT infrastructure
- Access level: Restricted to researcher and supervisor (Tier 1-2 data); UP Library Services and authorised auditors (Tier 3 data with permission)
- Not a preservation repository: Data migrates to Figshare or is destroyed after active phase; not intended for long-term public access

Tier 4: Disciplinary and National Repositories (Conditional)

- Description: External repositories may be used for specific data types if institutional repositories are insufficient or if disciplinary standards require.
- Potential repositories:
  - LIASA (Library and Information Association of South Africa) repository or portal: For sharing library sector benchmarking data with professional community
  - DHET (Department of Higher Education and Training) research data portal: If national higher education data aggregation initiatives emerge
  - Zenodo or OSF (Open Science Framework): If international collaboration requires platform-neutral hosting; used only for copies already deposited in UP Figshare
- Rationale: Extends discoverability beyond UP; aligns with open science principles; supports national and international comparative research
- Condition: Only used for fully anonymised, non-sensitive data already cleared for open release; primary deposit always in UP Figshare to maintain institutional record

No proposed use of personal or non-institutional repositories:

- Personal cloud services (Google Drive, Dropbox, personal OneDrive): Excluded for security, legal jurisdiction, and policy compliance reasons
- Commercial data repositories without institutional agreement: Excluded to maintain UP control over data residency and preservation commitment

- Physical media only (external drives, DVDs): Excluded as sole preservation method; may supplement institutional repositories for offline backup but never replace them

2. Repository Costs UP Figshare

- Cost to researcher: None. UP maintains an institutional subscription covering all research data deposits.
- Cost to institution: Covered by UP Library Services / Research Office central budget; exact institutional cost not transparent to individual researchers but part of established service provision.
- Additional services: DOI assignment included; unlimited storage for standard research datasets; no charge for embargo periods; no charge for metadata enhancement.
- Potential future costs: If UP subscription model changes or if dataset sizes exceed standard allocations (unlikely for this 1–2 GB project), researcher would be notified and supported in seeking alternative institutional arrangements.

UPSpace

- Cost to researcher: None. Mandatory institutional repository for theses; no deposit fees.
- Cost to institution: Covered by UP Library Services as core scholarly communication infrastructure.
- Additional services: Format migration, integrity checking, and permanent preservation included.

UP Network Drives and OneDrive

- Cost to researcher: None. Included in standard UP IT provision for staff and registered postgraduate students.
- Storage allocation: Typically 5–50 GB for network drives; 1 TB for OneDrive; more than adequate for this project.

Disciplinary and National Repositories (Conditional)

- Zenodo: Free for open science deposits up to 50 GB per dataset; funded by European Commission OpenAIRE programme.
- OSF: Free for academic use; funded by Center for Open Science.
- LIASA / DHET: No costs anticipated; South African government and professional association funded.
- Condition: Only used if no additional cost to researcher or UP; primary preservation remains institutional.

Cost summary:

- Total direct costs to researcher for repository deposit and preservation: Zero
- Total indirect costs (institutional subscription, IT infrastructure): Absorbed by UP as standard research support
- No grant-funded repository charges required
- No need to cost external repository fees in project budget

3. Time and Effort Costed for Data Preparation and Preservation The following tasks represent the additional effort required beyond core research activities to prepare data for sharing and long-term preservation. These have been explicitly costed in the research timeline. **Task 1: Data cleaning and finalisation for deposit**

- Description: Ensure all datasets are complete, consistent, and free of errors; verify no residual identifiers in qualitative data; confirm all variables documented in data dictionary; validate CSV files open correctly across different software.
- Effort: 2 days
- When scheduled: Final month of analysis phase, before thesis submission

- Skills required: Data management; attention to detail; familiarity with repository format requirements
- Resources: UP network drive access; spreadsheet software; data dictionary template

Task 2: Anonymisation verification and ethical clearance for sharing

- Description: Systematic review of all qualitative outputs for residual identification risk; verify k-anonymity in quantitative cross-tabulations; confirm all direct quotes appropriately anonymised; document anonymisation decisions; obtain final ethics committee confirmation if required for open release.
- Effort: 2–3 days
- When scheduled: Parallel to Task 1; must be completed before repository deposit
- Skills required: Ethical sensitivity; understanding of POPIA requirements; qualitative data handling
- Resources: Ethics committee contact; anonymisation checklist; supervisor review

Task 3: Metadata creation and standardisation

- Description: Complete DataCite metadata for Figshare deposit (title, creator, date, description, keywords, licence, funding information); complete Dublin Core metadata for UPSpace thesis deposit; create DDI-compliant methodology documentation; write README.txt for each dataset; ensure all metadata aligns with UP, NRF, and disciplinary standards.
- Effort: 1–2 days
- When scheduled: After data finalisation; before repository deposit
- Skills required: Metadata standards knowledge; descriptive writing; technical documentation
- Resources: Figshare metadata template; UPSpace deposit guide; DDI documentation tools

Task 4: Stakeholder review and institutional clearance

- Description: Submit aggregate cost data and operational recommendations to UP Library Services for pre-publication review; present findings to Student Representative Council if required; incorporate feedback; obtain written clearance from UP Library Services and Finance Department for open release of institutional data; negotiate any embargo periods or restricted access requirements.
- Effort: 2–3 days (spread over 2–4 weeks depending on stakeholder responsiveness)
- When scheduled: After analysis complete; before public repository deposit
- Skills required: Communication; negotiation; understanding of institutional politics and sensitivities
- Resources: Contact details for UP Library Services management; presentation templates; draft summary documents

Task 5: Repository deposit and verification

- Description: Upload datasets to UP Figshare; complete all metadata fields; assign and verify DOI; apply appropriate Creative Commons licences; set embargo if required; link datasets to thesis record in UPSpace; test all download links and file integrity; verify metadata appears correctly in DataCite and discovery systems.
- Effort: 1 day
- When scheduled: After stakeholder clearance; ideally concurrent with thesis submission
- Skills required: Repository interface navigation; file management; DOI and licensing understanding
- Resources: UP Figshare account; deposit guide from UP Library Services; IT support if technical issues arise

Task 6: Thesis deposit in UPSpace

- Description: Final thesis formatting per UP requirements; PDF/A creation; metadata completion; upload to UPSpace; verification of successful deposit and public access; notification of supervisor

and faculty administrator.
- Effort: 1 day (separate from thesis writing and examination process)
- When scheduled: After examination corrections complete; before graduation
- Skills required: UPSpace deposit procedures; PDF/A creation; administrative compliance
- Resources: UPSpace deposit guide; faculty administrator support; UP Library Services repository team

Task 7: Post-deposit monitoring and maintenance

- Description: Monitor repository records for download statistics, citations, and user feedback; respond to data reuse queries; update metadata if errors discovered; manage any version updates if datasets are revised; ensure contact information remains current.
- Effort: Ongoing; approximately 0.5 day per year for 5 years, then as needed
- When scheduled: Annually for first 5 years; ad hoc thereafter
- Skills required: Communication; repository account management; responsiveness to academic community
- Resources: Repository account access; email; professional network

Total preparation effort costed:

- One-time preparation tasks (Tasks 1–6): 9–12 days
- Ongoing maintenance (Task 7): 2.5 days over 5 years
- Total: Approximately 11.5–14.5 days of researcher effort

Integration with project timeline:

- Months 1–12: Active data collection and analysis; no preservation preparation
- Months 13–16: Analysis completion; begin Tasks 1 and 2 in parallel with thesis writing
- Months 17–18: Thesis submission; complete Tasks 3, 4, and 5; stakeholder review may extend slightly beyond submission
- Month 19–20: Examination period; minimal preservation activity
- Month 21: Post-examination corrections; Task 6 (UPSpace deposit)
- Months 22 onwards: Task 7 ongoing maintenance

This timeline ensures preservation preparation is not rushed or neglected amid thesis completion pressures. The 9–12 days of preparation effort represents approximately 5–7% of total research time (assuming 18-month project), which is proportionate and economically viable. **4. Curation Beyond the Lifetime of the Project** The selected repositories provide curation mechanisms that extend beyond the researcher's active involvement or affiliation with UP. **UP Figshare curation services:**

- Automatic format migration: Repository administrators monitor file format obsolescence; notify depositors if migration required; perform batch conversions for common formats (e.g., older Excel to current CSV standards)
- Metadata enhancement: Periodic review of metadata completeness; enrichment with controlled vocabularies; alignment with evolving standards
- Link maintenance: DOIs are persistent and maintained by DataCite infrastructure; links from datasets to publications and vice versa preserved
- Integrity checking: Regular checksum verification to detect bit-rot or corruption; replacement from backup if integrity compromised
- Access policy enforcement: Embargo periods automatically lifted; restricted access applications managed by repository administrators if researcher unavailable
- Discovery enhancement: Integration with international systems (Google Dataset Search, OpenAIRE, BASE) maintained by repository platform; researcher not responsible for ongoing technical integration

UPSpace curation services:

- Format migration: PDF/A standard ensures long-term readability; migration to future archival formats managed by UP Library Services
- Metadata harvesting: OAI-PMH interface ensures ongoing discovery by national and international thesis portals; maintenance by repository technical team
- Permanent URL: UPSpace assigns persistent identifier; maintained indefinitely
- Institutional commitment: UP Higher Degrees policy mandates permanent retention; funded through Library Services operational budget

Researcher responsibilities post-project:

- Maintain current contact information in repository records for 5 years to respond to queries
- Notify UP Library Services of any discovered errors or required updates
- Respond to data reuse requests within reasonable timeframe
- After 5 years, responsibility transitions fully to institutional repository administrators unless researcher voluntarily maintains engagement

Contingency if researcher leaves UP or becomes unavailable:

- UP Figshare and UPSpace records remain accessible; institutional ownership ensures continuity
- Repository administrators handle routine queries and access requests
- For complex methodological questions, documentation in data dictionary and README files must be sufficiently detailed to substitute for direct researcher consultation
- Supervisor remains as secondary contact for academic integrity questions during retention period

5. Preparation and Documentation Plans for Sharing and Archiving The following specific preparations ensure data is shareable and archivable beyond the project lifetime. **File format preparation:**

- Tabular data: Final datasets saved as CSV (UTF-8 encoding) with clear headers; Excel versions (.xlsx) provided as supplementary access format; no proprietary formulas or macros in shared files
- Documentation: PDF/A for long-term readability; DOCX provided as supplementary format
- Visualisations: High-resolution PNG or SVG for graphics; source data included to enable reproduction
- Audio (if any retained): WAV or FLAC for preservation; MP3 for access if size constraints require
- Methodology scripts: R or Python scripts (if used) with comments; or detailed Excel formula documentation if spreadsheet-based analysis

Metadata preparation:

- DataCite mandatory fields: Identifier (DOI), Creator, Title, Publisher, PublicationYear, ResourceType, Subject (keywords)
- DataCite recommended fields: Description (abstract), Contributor (supervisor, UP Library Services), Date, Language, AlternateIdentifier, RelatedIdentifier (link to thesis DOI), Size, Format, Rights (licence), FundingReference
- Dublin Core crosswalk: Ensure all DataCite fields mappable to Dublin Core for UPSpace and OAI harvesting
- DDI elements: Study description, methodology, data collection, variables, universe, sampling, time period, geographic coverage

Documentation preparation:

- README.txt for each dataset: Overview, context, file manifest, variable explanations, usage guidance, citation request, contact information, date of deposit, version information
- Data dictionary: Variable names, definitions, data types, value labels, missing value codes, units, source, derivation method
- Methodology narrative: Written account of data journey, decisions made, challenges encountered,

stakeholder consultation, limitations acknowledged
- Codebook (for qualitative data): Coding categories, definitions, examples, inter-coder reliability (if applicable), software used

Licence preparation:

- CC BY 4.0: Applied to fully anonymised aggregate quantitative datasets; maximises reuse; requires attribution
- CC BY-NC-SA 4.0: Applied to anonymised qualitative themes; balances openness with ethical protection; non-commercial restriction; share-alike requirement
- Restricted access protocol: For coded qualitative transcripts; application-based access; applicant must agree to terms of use; researcher or repository administrator approves

Quality assurance before deposit:

- Internal review: Supervisor reviews all datasets, metadata, and documentation for completeness and accuracy
- External review: UP Library Services reviews institutional data outputs for accuracy and sensitivity
- Technical review: Verify all files open correctly on different devices and software; test download links; validate DOI resolution
- Ethical review: Confirm no residual identifiers; verify consent covers proposed sharing scope; confirm POPIA compliance

6. Reflection: Repository and Preservation Learning from the Maule Case Study Maule's case study, while not explicitly addressing repository deposit, reveals critical gaps that inform this preservation strategy.

- **No repository deposit was mentioned:** The Westminster data's long-term fate is unclear. It may remain on a shared drive, lost when staff change, or deleted during IT upgrades. My explicit deposit in UP Figshare with DOI ensures the data survives institutional memory and personnel changes.
- **No metadata standards were applied:** The original data lacked documentation that would enable future reuse without Maule's personal involvement. My comprehensive metadata preparation ensures the data is interpretable by any future researcher, not merely by those who can contact me.
- **The value of institutional infrastructure was underutilised:** Maule's project catalysed a data culture but did not apparently embed its outputs in that culture's infrastructure. My deposit in UP Figshare makes the data part of UP's permanent scholarly record, not merely a personal achievement.
- **Stakeholder consultation records were not systematically archived:** The rich process of engaging Student Union, registry managers, and committees survives only because Maule wrote the case study. My stakeholder review and clearance process, documented and deposited, preserves this institutional memory for future service change initiatives.
- **Format obsolescence was not addressed:** The original Excel files were already problematic in 2017; by 2027 they may be harder to open or interpret. My use of CSV and PDF/A as preservation formats, with institutional migration commitments, addresses this proactively.

The essential lesson is that **preservation is not an afterthought but a design requirement** . From the moment data collection begins, every decision about format, structure, documentation, and access should anticipate the moment when the data must survive without the researcher's active stewardship. The repositories selected, the effort costed, and the preparations outlined above ensure that this research on UP library opening hours contributes not merely to a thesis but to an enduring evidence base for institutional decision-making and academic inquiry.

Data Sharing

How will you share the data?

- Non-sensitive data shared in institutional repository

1. How Potential Users Will Find Out About the Data Users will discover this research data through multiple channels, ensuring broad visibility across academic, institutional, and professional communities. **Repository-based discovery**

- UP Figshare: The primary data repository provides automatic indexing and integration with international discovery systems. Datasets will be discoverable through:
  - DataCite search portal (search.datacite.org)
  - Google Dataset Search (datasetsearch.research.google.com)
  - OpenAIRE (openaire.eu)
  - BASE – Bielefeld Academic Search Engine (base-search.net)
  - Dimensions (dimensions.ai)
  - Web of Science Data Citation Index (if UP Figshare is indexed)
  - Repository browse and search functions on the UP Figshare platform itself
- UPSpace: The institutional repository ensures discovery through:
  - OAI-PMH harvesting by national and international thesis portals
  - Google Scholar indexing
  - South African National ETD Portal
  - UP Library catalogue
  - Institutional browse and search

Publication-based discovery

- Journal articles: Data availability statements in published papers will reference the UP Figshare DOI, directing readers to the underlying datasets
- Thesis: The UPSpace record will include links to associated datasets in UP Figshare
- Conference presentations: Slides and abstracts will reference dataset DOIs; audience members directed to repository

Institutional and professional networks

- UP Library Services internal communications: Findings and data availability communicated to library staff through existing channels (staff meetings, newsletters, intranet)
- Student Representative Council: Research outcomes shared through SRC channels, with data access information for student leaders
- LIASA (Library and Information Association of South Africa): Conference presentations and newsletter contributions referencing available data
- HELIOS (Higher Education Library Information Systems) or equivalent South African library consortia: Data shared for sector benchmarking if appropriate agreements established

Direct researcher engagement

- Academic social networks: Researcher profiles on ResearchGate, Academia.edu, and ORCID will list publications and associated datasets with DOI links
- Professional conference attendance: Oral presentations at LIASA conferences, IFLA regional meetings, or South African higher education research forums will publicise data availability
- Email enquiries: Researcher contact information provided in repository metadata; reasonable response commitment to data reuse queries

Metadata enhancement for discovery

- Comprehensive keyword coverage: Terms will include "academic library opening hours," "cost-effectiveness," "university library usage," "South Africa," "University of Pretoria," "student needs assessment," "library service evaluation," "higher education resource allocation"
- Subject classification: Library and Information Science (LIS); Higher Education Administration; Educational Management; Operations Research
- Geographic tagging: South Africa; Gauteng; Pretoria; University of Pretoria campuses
- Temporal coverage: 2026 academic year; term time periods explicitly defined

2. With Whom Data Will Be Shared and Under What Conditions Data sharing is tiered according to data sensitivity, with different audiences granted access under specific conditions. **Tier 1: Fully anonymised aggregate quantitative datasets**

- Shared with: General public; global academic community; library practitioners; higher education administrators; students; journalists; policymakers
- Conditions: Open access via CC BY 4.0 licence; no registration required; no access restrictions; attribution required through citation of DOI
- Mechanism: Direct download from UP Figshare; no intermediary approval
- Rationale: These datasets contain no identifiable information and represent public good value for benchmarking, teaching, and policy development

Tier 2: Anonymised qualitative themes and aggregated survey findings

- Shared with: Academic researchers; library professionals; higher education researchers; graduate students
- Conditions: CC BY-NC-SA 4.0 licence; non-commercial use; derivatives must share alike; attribution required; no registration required for basic access
- Mechanism: Direct download from UP Figshare; licence terms bind all users
- Rationale: Qualitative data, even anonymised, carries higher ethical sensitivity than quantitative operational data; non-commercial restriction protects against exploitative reuse

Tier 3: Coded qualitative transcripts (restricted access)

- Shared with: Bona fide researchers with demonstrated need; supervised graduate students; institutional auditors
- Conditions: Application required via UP Figshare; applicant must provide institutional affiliation, research purpose, and intended use; researcher or repository administrator approves; applicant agrees to terms of use including no attempt at re-identification, no sharing with third parties, secure storage, and destruction after use
- Mechanism: Application form on repository record; approval within 14 working days; download link provided upon approval; access logged
- Rationale: Protects participant confidentiality while enabling legitimate research reuse; balances openness with ethical obligation

Tier 4: Institutional cost-effectiveness models and detailed operational recommendations

- Shared with: UP Library Services senior management; UP Finance Department; approved external consultants working with UP; peer institutions under formal data sharing agreement
- Conditions: UP institutional approval required; may be subject to confidentiality agreement; no public dissemination without explicit UP Library Services clearance; aggregate reporting only
- Mechanism: Direct institutional sharing via secure channels; not deposited in public repository without clearance
- Rationale: Protects institutional commercial and administrative sensitivities; maintains trust with data provider

Tier 5: Raw identifiable data and linking keys

- Shared with: No one. Destroyed after anonymisation and verification.

- Conditions: Not applicable
- Mechanism: Secure deletion with overwriting; documentation of destruction in research journal
- Rationale: POPIA compliance; ethical obligation to minimise personal information exposure

Specific sharing arrangements:

- With UP Library Services: All tiers shared for institutional service improvement; UP may use findings internally without restriction; public dissemination of Tier 4 data requires mutual agreement
- With supervisor: All tiers during active research and retention period; after project completion, supervisor retains access to public repository records only
- With external examiners: Thesis and anonymised appendices only; no access to restricted qualitative transcripts unless specifically required for verification and approved by ethics committee
- With peer researchers requesting collaboration: Formal collaboration agreement required; ethics clearance verification; access limited to specific agreed datasets

3. Sharing Mechanisms Primary mechanism: Repository-mediated sharing

- UP Figshare serves as the primary sharing platform for all data suitable for open or restricted access
- Advantages: Persistent identifiers (DOIs); standardised metadata; automatic licence enforcement; download tracking; international discoverability; institutional backup and preservation; no ongoing researcher effort for routine access
- For restricted data: Repository handles application processing, approval workflow, and access logging; researcher notified of applications and can review if desired

Secondary mechanism: Direct request handling

- Researcher email contact provided in repository metadata for:
    - Complex methodological questions not answered in documentation
    - Collaboration proposals requiring discussion before data access
    - Institutional enquiries requiring contextual explanation
    - Technical issues with repository access
- Response commitment: Within 14 working days for academic enquiries; reasonable effort for other enquiries
- No direct sharing of data files via email or personal cloud; all file transfers directed through repository or institutional secure file transfer

Tertiary mechanism: Institutional and professional network sharing

- Presentations at UP Library Services meetings, Student Representative Council forums, LIASA conferences
- Summary reports and briefs shared through institutional channels
- Data availability referenced in all outputs; audiences directed to repository for full access

Mechanisms not used:

- Personal cloud sharing (Google Drive, Dropbox, personal OneDrive): Excluded for security, policy, and persistence reasons
- Social media file sharing: Excluded; social media used only for announcement of availability, not for direct data transfer
- Unencrypted email attachments: Excluded; all file transfers through secure channels

4. When Data Will Be Made Available Immediate availability (upon thesis submission or earlier)

- Anonymised aggregate quantitative datasets: Deposited in UP Figshare with immediate open

access
- Data dictionary and README files: Immediate open access
- Methodology narrative: Immediate open access
- Final visualisations: Immediate open access

Embargoed availability (delayed release)

- Coded qualitative transcripts: 12-month embargo from thesis submission; allows researcher to publish primary qualitative findings without pre-emption; reduces risk of qualitative data being mined by others before researcher has established publication record
- Institutional cost-effectiveness models: Embargo until UP Library Services has implemented recommendations and reviewed outcomes; embargo period negotiated with UP Library Services (estimated 6–12 months)
- Thesis itself: If publisher requirements or institutional policy dictate, thesis may have 6–12 month embargo in UPSpace; datasets in Figshare may be available earlier if independent of thesis embargo

Restricted availability (application-based, ongoing)

- Coded qualitative transcripts: Available after embargo via application; ongoing availability for legitimate research use; reviewed annually for continued relevance and ethical appropriateness

No availability (destroyed)

- Raw identifiable data: Destroyed after anonymisation; never available
- Pseudonym linking keys: Destroyed after verification period; never available
- Audio recordings: Destroyed after transcription; never available (unless explicit extended consent obtained)

Timeline summary:

- Month 0–18: Active research; no data sharing except with supervisor and institutional stakeholders under confidentiality
- Month 18 (thesis submission): Tier 1 and 2 data deposited in UP Figshare with immediate or short embargo; thesis in UPSpace with any required embargo
- Month 19–20: Examination period; embargoed data remains restricted
- Month 21 (post-examination): Embargoes lifted per schedule; restricted access activated for Tier 3 data
- Month 22 onwards: Ongoing availability; researcher monitors and responds to access requests

5. Persistent Identifiers DOI assignment for datasets

- UP Figshare automatically assigns Digital Object Identifiers (DOIs) to all deposited datasets
- DOI structure: 10.25403/UP.figshare.[unique number] (UP Figshare prefix)
- Benefits of DOI:
    - Persistence: DOI resolves to dataset record even if repository URL changes; maintained by DataCite infrastructure
    - Citability: Datasets citable in academic literature using standard citation formats
    - Impact tracking: Altmetrics and citation counts track dataset reuse and influence
    - Trust: DOI signals institutional validation and long-term commitment

DOI assignment for thesis

- UPSpace assigns a persistent identifier to the thesis record
- Linked to Figshare DOIs through RelatedIdentifier metadata field, creating bidirectional discovery

ORCID for researcher

- Researcher maintains ORCID iD (Open Researcher and Contributor ID)
- ORCID linked to UP Figshare profile, UPSpace record, and all publications
- Ensures disambiguation of researcher identity; automatic updating of publication and dataset records

No persistent identifier for restricted access applications

- Application-based access to Tier 3 data does not generate DOI for individual users
- Access logged by repository but not publicly identified

6. Acknowledging Reuse of Data Users of this data are required to acknowledge reuse through the following mechanisms. **Citation requirement**

- All users must cite the dataset using the DOI and recommended citation format provided in the repository record
- Recommended citation format: [Researcher surname], [Initial]. ([Year]). [Dataset title]. University of Pretoria. https://doi.org/10.25403/UP.figshare.[number]
- Example: Smith, J. (2026). University of Pretoria Library Opening Hours Usage Data (2026). University of Pretoria. https://doi.org/10.25403/UP.figshare.12345678

Acknowledgement in publications

- Users publishing research based on this data must acknowledge the source in their manuscript
- Suggested wording: "This research uses data from the University of Pretoria Library Opening Hours Study (Smith, 2026), deposited in the University of Pretoria Figshare repository."
- Users encouraged but not required to notify researcher of publications using the data (via email or repository feedback); researcher maintains informal bibliography of reuse

Licence compliance

- CC BY 4.0 users: Must provide attribution; may adapt but must indicate changes; may use commercially with attribution
- CC BY-NC-SA 4.0 users: Must provide attribution; must use non-commercially; must share derivatives under identical licence; must indicate changes
- Restricted access users: Bound by application terms; no publication without additional permission; no sharing with third parties; secure storage and destruction requirements

No additional restrictions

- No requirement for co-authorship in exchange for data access
- No requirement for pre-publication review of user outputs
- No fees for data access
- No discrimination based on user nationality, institution, or research topic (subject to licence terms and South African law)

7. Track Record of Effective Data Sharing As a postgraduate researcher, formal track record of data sharing may be limited. However, the following demonstrate commitment to open and responsible data practices. **During this project:**

- Research journal maintained per ANU learning journal guidelines, documenting all data management decisions, ethical considerations, and methodological choices; journal itself may be shared as exemplar of reflective research practice
- All data management decisions documented in this DMP, which may be shared as a template or example for peers
- Supervisor and UP Library Services consulted throughout on data handling; established collaborative relationships that support responsible sharing

Anticipated future track record:

- This DMP and the resulting repository deposits will establish initial track record
- Intention to publish methodology papers that explicitly discuss data sharing practices
- Intention to present at LIASA or higher education conferences on data-driven library decision making, including discussion of data sharing as professional practice
- Commitment to respond to data reuse enquiries promptly and constructively, building reputation for collaborative data stewardship

8. Reflection: Data Sharing Learning from the Maule Case Study Maule's case study reveals both positive sharing practices and missed opportunities that inform this strategy.

- **Effective institutional sharing:** Maule shared findings extensively within Westminster—Student Union, registry managers, Student Experience Committee, library staff at all levels. This internal sharing was crucial for service change implementation. My strategy extends this by also making data available externally, while maintaining robust internal stakeholder engagement.
- **No formal repository deposit:** The Westminster data's long-term availability is unclear. It may survive in institutional reports or the published case study, but there is no persistent identifier, no discoverable dataset, and no mechanism for external researchers to verify or build upon the findings. My DOI-assigned Figshare deposit ensures the UP data escapes this opacity.
- **No licence or conditions specified:** Maule's sharing was governed by informal institutional trust. My explicit Creative Commons licences and restricted access protocols provide clarity that protects all parties—users know what they can do; the institution knows what is protected; participants know their confidentiality is enforced.
- **Qualitative data sharing not addressed:** Maule coded NSS comments and feedback cards but did not discuss whether or how these could be shared. My tiered approach distinguishes between shareable themes and restricted transcripts, providing ethical clarity.
- **The value of consultation records as shared knowledge:** Maule's detailed stakeholder consultation was a significant contribution, but only survives as narrative in the published case study. My documentation and deposit of consultation frameworks and outcomes preserves this process knowledge for future institutional change initiatives.

The fundamental insight is that **sharing is not merely dissemination but contribution to a cumulative evidence base**. Maule's project advanced Westminster's library service; my aspiration is that the UP data, properly shared and discoverable, may advance library services across South Africa and beyond.

Are any restrictions on data sharing required?

1. Identified Restrictions and Their Causes This research faces several restrictions on data sharing, arising from ethical obligations, institutional sensitivities, and legal requirements. Each restriction is identified below with its underlying cause. **Restriction 1: Raw identifiable student usage data cannot be shared**

- Data affected: Swipe card logs with student IDs; individual entry records; any data linking usage patterns to identifiable individuals
- Cause: Protection of Personal Information Act (POPIA) 2013 prohibits sharing personal information without lawful basis; research consent obtained for aggregate analysis only, not for open release of individual records; ethical obligation to protect student privacy; potential harm if individual study habits, campus movements, or academic stress indicators are exposed
- Severity: Absolute prohibition; no waiver or workaround possible under South African law
- Duration: Permanent; data destroyed after anonymisation, never shared

Restriction 2: Individual cost and salary information cannot be shared

- Data affected: Specific staff salary figures; vendor contract details; individual supplier pricing; line-item operational budgets

- Cause: Institutional confidentiality; commercial sensitivity of procurement information; potential industrial relations impact if individual remuneration disclosed; competitive disadvantage if vendor pricing revealed; UP Finance Department governance requirements
- Severity: High restriction; aggregate sharing possible with approval; individual detail prohibited
- Duration: Permanent for individual data; aggregate data may be shared after institutional review

Restriction 3: Pseudonymised qualitative transcripts require controlled access

- Data affected: Coded focus group transcripts; detailed survey responses with contextual information; interview recordings (if retained)
- Cause: Even with pseudonyms and contextual anonymisation, qualitative data richness carries residual re-identification risk; participants consented to research use but not necessarily to open public release; focus group dynamics mean individual confidentiality cannot be fully guaranteed; ethical duty to minimise harm from disclosure
- Severity: Moderate restriction; sharing possible under controlled conditions with additional safeguards
- Duration: Extended; application-based access ongoing after embargo period

Restriction 4: Institutional operational recommendations require pre-publication review

- Data affected: Cost-effectiveness models; proposed opening hours changes; efficiency recommendations; staffing implications
- Cause: UP Library Services must manage service change politically and operationally; premature public release of recommendations could create staff uncertainty, student protest, or competitive positioning issues; Finance Department must verify cost calculations before institutional commitment; senior management requires control over timing of change announcements
- Severity: Moderate restriction; sharing possible after institutional implementation and review
- Duration: Temporary; estimated 6–12 months embargo, then public release with approval

Restriction 5: Third-party survey data may have provider restrictions

- Data affected: Any National Student Survey, South African equivalent, or external benchmarking data incorporated for comparative analysis
- Cause: Survey providers typically retain copyright and control over data distribution; institutional licences may prohibit public sharing of raw or aggregate results; provider terms may restrict cross-institutional comparison
- Severity: Variable; dependent on specific provider terms
- Duration: Per provider licence; may be permanent if terms prohibit sharing

2. Actions to Overcome or Minimise Restrictions For each restriction, specific actions will be taken to minimise barriers to sharing while respecting ethical and institutional obligations.
Minimising Restriction 1: Raw identifiable data

- Action 1: Immediate anonymisation upon extraction from UP systems; aggregate to hourly, daily, and site-level counts before any researcher analysis
- Action 2: Destroy raw identifiable data within 6 months of collection; document destruction in research journal
- Action 3: Verify that aggregate data meets k-anonymity standards (no cell in cross-tabulations contains fewer than 5 individuals)
- Action 4: Share only fully anonymised aggregate datasets openly; these contain no personal information and are not subject to POPIA sharing restrictions
- Outcome: Complete elimination of restriction for shared outputs; raw data restriction managed through destruction rather than withholding

Minimising Restriction 2: Individual cost information

- Action 1: Aggregate all cost data to site level or service category level before analysis; never

retain individual salary figures in research files
- Action 2: Use percentage allocations and ratios rather than absolute figures where individual identification possible (e.g., "security costs represent 15% of site operational budget" rather than "R45,678 for Security Company X")
- Action 3: Redact vendor names; use generic descriptors ("cleaning services provider" not "Company Y")
- Action 4: Present cost-effectiveness as relative metrics (cost per user hour, percentage capacity utilisation) rather than absolute expenditure
- Action 5: Submit all cost-containing outputs to UP Library Services and Finance Department for pre-publication review; incorporate feedback; obtain written clearance
- Action 6: If institutional clearance for open release denied, retain aggregate cost models as restricted access with UP institutional approval required; do not destroy if institutional value remains
- Outcome: Significant sharing enabled through aggregation and redaction; residual restriction managed through institutional review and potential restricted access

Minimising Restriction 3: Pseudonymised qualitative transcripts

- Action 1: Apply robust anonymisation: direct removal of names and identifiers; pseudonym replacement; contextual generalisation (specific courses to faculties, rare characteristics suppressed); verification of k-anonymity in reported themes
- Action 2: Create two-tier qualitative outputs: fully anonymised themes and categories for open sharing (Tier 2); coded transcripts with more context for restricted application-based access (Tier 3)
- Action 3: Obtain explicit re-consent from focus group participants for transcript sharing if original consent insufficient; offer opt-out for open release while retaining data in restricted form
- Action 4: Implement application-based access with terms of use binding applicants to no re-identification attempts, no onward sharing, secure storage, and post-use destruction
- Action 5: Review restricted access annually; if re-identification risk diminishes over time or if participants explicitly consent to wider release, upgrade to more open access
- Outcome: Balanced sharing enabled; open release of low-risk thematic data; controlled sharing of richer contextual data; ethical obligations maintained

Minimising Restriction 4: Institutional operational recommendations

- Action 1: Engage UP Library Services early and continuously throughout research; share draft findings informally for feedback; build trust and alignment
- Action 2: Frame recommendations as evidence-informed options rather than prescriptive directives; reduce institutional defensiveness
- Action 3: Separate data sharing from recommendation sharing: make aggregate usage data available immediately; embargo only the specific operational recommendations and cost models until implementation
- Action 4: Negotiate clear timeline for institutional review and implementation; set mutual expectations for embargo duration
- Action 5: Offer to co-author or co-present findings with UP Library Services representatives, giving institutional ownership of public messaging
- Action 6: If embargo extends beyond 12 months, review with UP whether restricted access or redacted release is preferable to complete withholding
- Outcome: Temporary restriction managed through stakeholder engagement; eventual open sharing anticipated; institutional trust preserved

Minimising Restriction 5: Third-party survey provider restrictions

- Action 1: Before incorporating any external survey data, verify licence terms and sharing permissions with UP Library Services and survey provider
- Action 2: If provider terms prohibit sharing, use data only for internal analysis and

contextualisation; do not include in shared datasets; cite provider publication if available
- Action 3: If provider permits institutional-level sharing but not raw data sharing, report only UP's own aggregate results with provider attribution
- Action 4: If no external survey data is accessible under shareable terms, rely on primary qualitative collection (surveys, focus groups) where researcher controls consent and sharing conditions
- Outcome: Restriction managed through proactive compliance; alternative data sources secured if external data unavailable for sharing

3. Exclusive Use Period Duration of exclusive researcher use:

- Raw identifiable data: Exclusive use from collection until anonymisation (maximum 6 months); no sharing with anyone except supervisor under confidentiality
- Processed working datasets: Exclusive use during active analysis phase (approximately 12–18 months from project start); sharing with supervisor for guidance; no external sharing
- Anonymised aggregate datasets: Exclusive use embargoed for 0–12 months depending on data type; thesis submission triggers open or embargoed release
- Institutional recommendations: Exclusive use with UP Library Services during implementation planning (6–12 months from thesis submission)

Rationale for exclusive use:

- Academic integrity: Researcher must establish primary publication record before data is mined by others; qualitative embargo protects against pre-emption of thematic analysis
- Institutional implementation: UP Library Services requires controlled period to assess and action recommendations before public commentary
- Verification: Exclusive use allows researcher to verify data accuracy and completeness before releasing to wider scrutiny
- Ethical sequencing: Raw data must be anonymised before any sharing; this processing takes time and requires exclusive handling

Justification for embargo periods:

- 0 months (immediate release): Appropriate for fully anonymised aggregate usage data that poses no ethical or institutional risk; maximises community benefit
- 6 months (short embargo): Appropriate for anonymised qualitative themes; allows initial thesis examination and potential journal submission without pre-emption
- 12 months (standard embargo): Appropriate for institutional recommendations; allows UP Library Services to develop implementation plans and communicate internally

No indefinite exclusive use claimed:

- All data with long-term value intended for eventual sharing; embargo periods defined and time-limited
- If exceptional circumstances require extension beyond 12 months, this must be justified to UP Research Ethics Committee and repository administrators
- Restricted access (Tier 3) provides ongoing controlled sharing even during embargo; not complete withholding

4. Data Sharing Agreements Agreement 1: UP Library Services Data Sharing Agreement

- Parties: Researcher and University of Pretoria Library Services (represented by Head: Library Services)
- Purpose: Authorises researcher access to operational data (entry logs, headcounts, capacity figures) for stated research purpose; governs handling, analysis, and sharing of institutional data
- Key provisions:
    - Data use limited to research objective: determining cost-effectiveness and user needs of

library opening hours
    - Raw identifiable data to be anonymised within 6 months; destruction documented
    - Aggregate outputs may be published with appropriate attribution to UP Library Services
    - Cost data to be aggregated and redacted; pre-publication review by UP Library Services for any outputs containing operational recommendations or financial information
    - UP Library Services retains right to use findings internally without restriction
    - Researcher to share draft findings with UP Library Services before public dissemination
    - Agreement terminates upon thesis submission and repository deposit; superseded by repository licences for shared data
- Confidentiality: Mutual confidentiality for sensitive institutional information; not a one-way non-disclosure agreement
- Duration: Active research period plus 12-month post-submission review period
- Signatories: Researcher, supervisor, Head: Library Services

Agreement 2: UP Finance Department Data Access Protocol

- Parties: Researcher and UP Finance Department (represented by relevant director)
- Purpose: Governs access to cost and budget information for research analysis
- Key provisions:
    - Access limited to aggregate figures and percentage allocations approved by Finance Department
    - No retention of individual salary figures, vendor contracts, or line-item budgets in research files
    - All cost-containing outputs submitted for pre-publication review
    - Finance Department may require redaction or aggregation of specific figures
    - Researcher acknowledges institutional ownership of financial data
- Duration: Active research period; specific access sessions scheduled and logged
- Signatories: Researcher, supervisor, Finance Department representative

Agreement 3: Focus Group Participant Consent Form

- Parties: Researcher and individual student/staff participants
- Purpose: Obtains informed consent for participation, recording, transcription, and data sharing
- Key provisions:
    - Explanation of research purpose and data uses
    - Consent for audio recording, transcription, and anonymised quotation in publications
    - Separate consent tick-box for open data sharing of anonymised themes (opt-in, not assumed)
    - Explanation that focus group confidentiality cannot be guaranteed for disclosures made within the group
    - Right to withdraw consent up to data analysis commencement (specify date)
    - Contact information for researcher and UP Research Ethics Committee
    - Data retention and destruction timeline explained
- Duration: Consent valid for stated research period; withdrawal rights exercisable until specified date; long-term data retention governed by repository licences if participant consented to sharing

Agreement 4: Restricted Access Terms of Use (for Tier 3 qualitative data)

- Parties: Repository (UP Figshare acting on behalf of researcher and UP) and approved data user
- Purpose: Governs application-based access to coded qualitative transcripts
- Key provisions:
    - Applicant must provide institutional affiliation, research purpose, and intended use
    - Applicant agrees to no attempt at re-identification of participants
    - Applicant agrees to no sharing of data with third parties
    - Applicant agrees to secure storage during use and secure destruction after use
    - Applicant agrees to cite dataset and notify researcher of publications

    - Applicant agrees to non-commercial use (if CC BY-NC-SA licence applies)
    - Breach of terms results in access revocation and potential institutional action
- Duration: Per access session; re-application required for extended or new use
- Enforcement: Repository logs access; researcher or administrator approves applications; UP IT Services can audit access if concerns arise

Agreement 5: Third-Party Data Provider Licence (if applicable)

- Parties: UP (or researcher via UP) and external survey provider (e.g., national student survey body)
- Purpose: Governs any use of externally provided survey data
- Key provisions:
    - Compliance with provider terms of use
    - No sharing beyond permitted scope
    - Attribution requirements
    - Reporting of aggregate findings to provider if required
- Duration: Per provider terms
- Note: If provider terms prohibit any sharing, this agreement effectively restricts researcher from including such data in shared datasets; alternative primary data collection pursued instead

5. Non-Disclosure Agreement Assessment A non-disclosure agreement (NDA) was considered but deemed insufficient as the primary protection mechanism for this research. **Why NDA alone is insufficient:**

- One-way NDAs protect the institution but do not bind the researcher to ethical data handling standards beyond confidentiality
- NDAs do not address participant rights under POPIA; they are institutional contracts, not ethical instruments
- NDAs do not provide mechanisms for controlled sharing; they typically prohibit all disclosure rather than enabling tiered access
- NDAs do not assign persistent identifiers or ensure long-term preservation
- NDAs create adversarial framing (institution vs. researcher) rather than collaborative data stewardship

Where NDA elements are incorporated:

- Confidentiality clauses within the broader UP Library Services Data Sharing Agreement protect sensitive institutional information
- Restricted access terms of use include NDA-like confidentiality obligations for approved users of Tier 3 data
- Researcher and supervisor operate under implicit confidentiality in supervisory relationship, supplemented by explicit discussion of sensitive data handling

Preferred approach:

- Comprehensive data sharing agreements that include confidentiality provisions alongside sharing permissions, ethical safeguards, institutional collaboration, and long-term preservation commitments
- Tiered access mechanisms that enable appropriate sharing rather than blanket prohibition
- Repository-mediated sharing with licence enforcement rather than bilateral contractual restrictions

Reflection: Restriction Management Learning from the Maule Case Study Maule's case study illustrates both effective restriction management and unaddressed challenges.

- **Effective institutional negotiation:** Maule successfully navigated financial and political sensitivities by presenting data clearly, consulting stakeholders early, and framing

recommendations as evidence-based rather than prescriptive. Her Student Union and registry manager consultations built trust that enabled eventual service change. My strategy formalises this through explicit data sharing agreements and pre-publication review protocols.
- **No formal restriction framework was documented:** Maule does not describe any data sharing agreements, consent forms, or institutional clearances. The project appears to have operated on informal trust. My explicit agreements provide protection for all parties and ensure the research is reproducible and defensible if challenged.
- **Qualitative data restrictions were not discussed:** NSS comments and feedback cards were coded and used, but there is no indication whether participants consented to this research use of their feedback, or whether the coding framework was shared. My tiered consent and application-based access address this gap directly.
- **The absence of embargo or exclusive use planning:** Maule's findings were published relatively quickly in the Stubbing edited volume, but there is no discussion of whether an exclusive use period was needed or whether Westminster had time to implement changes before public commentary. My 6–12 month embargo for institutional recommendations provides this protected implementation window.
- **No persistent identifier or repository deposit:** The restriction on external sharing was effectively total because no mechanism existed for controlled external access. My repository-mediated sharing with DOI and tiered access transforms restriction from prohibition to managed enablement.

The essential insight is that **restrictions are not obstacles to sharing but parameters for responsible sharing**. Every restriction identified—ethical, legal, institutional—can be minimised through thoughtful data transformation, stakeholder engagement, and technical access controls. The goal is not unrestricted sharing but **appropriate sharing**: the right data, with the right people, under the right conditions, for the right purposes.

Responsibilities and Resources

Who will be responsible for data management?

Primary Researcher: Paballo Mamabolo

- Role: Principal investigator and data steward for all research-generated data
- Institutional affiliation: Postgraduate student, University of Pretoria, Department of Information Science
- Contact: u04978502@tuks.co.za; maintained throughout research and 5-year post-completion period

Core responsibilities: Data capture and collection

- Design and implement quantitative data collection protocols (headcount procedures, swipe card data extraction, cost data requests)
- Conduct qualitative data collection (surveys, focus groups, comment card collection)
- Ensure immediate secure transfer of field data to institutional storage
- Maintain field collection logs and research journal documenting all collection activities

Metadata production

- Create and maintain data dictionary for all datasets
- Complete metadata fields per Dublin Core, DataCite, and DDI standards
- Write README.txt files for each dataset

- Document methodology, assumptions, limitations, and transformations in research journal

Data quality

- Verify completeness and accuracy of collected data
- Clean and validate raw data; document all cleaning decisions
- Cross-check quantitative findings against qualitative insights
- Test file integrity after transfers and backups
- Verify anonymisation effectiveness before sharing

Storage and backup

- Save all working files to UP-approved locations (network drives, OneDrive)
- Perform weekly encrypted external hard drive backups
- Verify backup integrity monthly
- Maintain version control for analysis files
- Ensure no raw identifiable data stored on personal devices

Data security and access control

- Implement tiered access controls as defined in DMP
- Encrypt all devices and external media
- Manage passwords and multi-factor authentication
- Control sharing permissions on collaborative folders
- Report any security incidents immediately

Ethical compliance

- Obtain ethics clearance before any data collection
- Secure informed consent from all qualitative participants
- Maintain consent forms and contact details securely
- Ensure POPIA compliance in all data handling
- Document ethical decisions in research journal

Data archiving and preservation preparation

- Prepare anonymised datasets for repository deposit
- Create final metadata records
- Write methodology narratives and documentation
- Format files for long-term preservation (CSV, PDF/A)
- Schedule and execute repository deposits

Data sharing and access management

- Respond to data reuse enquiries within 14 working days
- Process restricted access applications for Tier 3 qualitative data
- Monitor repository download statistics and citations
- Update contact information and respond to error reports

Stakeholder communication

- Maintain regular communication with UP Library Services
- Present draft findings to institutional stakeholders
- Incorporate feedback and obtain clearance for publication
- Attend scheduled consultation meetings

DMP implementation and review

- Execute all DMP provisions as written

- Review DMP quarterly for continued relevance
- Revise DMP if research scope, methods, or institutional conditions change
- Document all revisions and justify changes

Supervisor: Dr Louise Patterton

- Role: Academic supervisor and oversight authority
- Institutional affiliation: University of Pretoria, Department of Build Environment and Engineering/ Information Science
- Contact: LPatterton@csir.co.za

Core responsibilities: Oversight of DMP implementation

- Review DMP at project commencement and approve as adequate
- Monitor DMP execution through regular supervision meetings (minimum monthly during active data phases)
- Review quarterly DMP self-assessments from researcher
- Authorise any significant DMP revisions

Quality assurance

- Review data dictionary and metadata for completeness and accuracy
- Spot-check anonymisation of qualitative data for residual identification risk
- Verify that raw identifiable data destruction has occurred as scheduled
- Review analysis outputs for methodological soundness

Ethical oversight

- Confirm ethics clearance obtained and consent procedures adequate
- Review participant information sheets and consent forms before use
- Advise on ethical dilemmas encountered during research
- Ensure adverse events reported to ethics committee

Backup and recovery verification

- Maintain emergency access to shared backup locations in case researcher unavailability
- Participate in annual recovery testing
- Verify that researcher is performing monthly backup checks

Repository deposit oversight

- Review datasets, metadata, and documentation before repository deposit
- Confirm that appropriate licences and access conditions applied
- Verify stakeholder clearances obtained before public release

Professional development guidance

- Advise on data management best practices
- Recommend training opportunities (UP research data management workshops, POPIA compliance sessions)
- Review research journal entries on data management learning

Post-project continuity

- Remain as secondary contact for data queries during retention period if researcher unavailable
- Assist repository administrators with methodological questions if researcher has left UP

UP Library Services: Head of Library Services or delegated representative

- Role: Institutional data provider and stakeholder

- Institutional affiliation: University of Pretoria Library Services
- Contact: Carike.schoeman@up.ac.za

Core responsibilities: Data provision

- Authorise researcher access to operational data (entry logs, headcounts, capacity figures)
- Provide cost data in agreed aggregate format
- Ensure data extracts comply with institutional records management policy
- Respond to data requests within agreed timeframe

Stakeholder consultation

- Review draft findings and recommendations containing operational or cost data
- Provide feedback on practical feasibility and institutional implications
- Attend scheduled presentation meetings
- Communicate with library staff about research purpose and outcomes

Pre-publication clearance

- Review outputs containing institutional cost data or operational recommendations
- Approve, request redaction, or deny publication of sensitive elements
- Provide written clearance for open release when satisfied

Service implementation

- Use research findings for opening hours planning and adjustment
- Monitor outcomes against research baseline
- Provide feedback to researcher on implementation results for potential follow-up study

Data sharing agreement signatory

- Execute formal data sharing agreement with researcher
- Ensure institutional obligations met
- Maintain copy of agreement for institutional records

UP Finance Department: Relevant director or delegated representative

- Role: Institutional cost data provider and reviewer
- Institutional affiliation: University of Pretoria Finance Department
- Contact: [Finance Department email]

Core responsibilities: Cost data access authorisation

- Approve researcher access to financial information in aggregate form
- Ensure no individual salary figures or vendor contracts disclosed to researcher
- Provide percentage allocations and site-level budgets as agreed

Pre-publication review

- Review any outputs containing institutional cost information
- Approve or request redaction of specific figures
- Ensure no competitive or industrial relations harm from publication

Data access protocol signatory

- Execute data access protocol with researcher
- Maintain records of data provided and conditions

UP Research Ethics Committee: Committee Chair or administrator

- Role: Ethical oversight body
- Institutional affiliation: University of Pretoria

Core responsibilities: Ethics clearance

- Review and approve research ethics application
- Issue ethics clearance certificate
- Monitor compliance through annual reports if required

Adverse event response

- Receive and investigate reports of ethical breaches or data incidents
- Authorise any modifications to approved protocols
- Maintain records of ethical approvals and incidents

Guidance provision

- Issue guidance on POPIA compliance in research
- Provide templates for consent forms and participant information sheets
- Advise on complex ethical dilemmas referred by researcher or supervisor

UP IT Services: Research support team or helpdesk

- Role: Technical infrastructure and security provider
- Institutional affiliation: University of Pretoria Information Technology Services
- Contact: [UP IT Services helpdesk]

Core responsibilities: Infrastructure provision

- Maintain network drives, OneDrive, and authentication systems
- Ensure backup systems operational and regular snapshots taken
- Provide encryption tools and guidance
- Issue multi-factor authentication tokens

Security incident response

- Receive and investigate reports of device compromise, ransomware, or data breach
- Perform forensic analysis if required
- Coordinate recovery from institutional backups
- Report to Information Security Office and researcher

Technical support

- Assist with repository deposit technical issues
- Provide guidance on secure file transfer
- Support device encryption and secure disposal
- Advise on software compatibility for preservation formats

UP Figshare and UPSpace Administrators: UP Library Services repository team

- Role: Repository curation and preservation
- Institutional affiliation: University of Pretoria Library Services

Core responsibilities: Repository access

- Issue researcher account for UP Figshare and UPSpace
- Provide deposit guidance and training
- Verify metadata completeness at deposit

Curation and preservation

- Monitor format obsolescence and notify depositors of migration needs
- Perform integrity checks on deposited files
- Maintain DOI resolution and repository infrastructure
- Enforce access conditions (embargoes, restricted access applications)

Post-researcher continuity

- Assume primary contact role for data queries if researcher unavailable
- Manage restricted access applications if researcher uncontactable
- Maintain repository records indefinitely

3. Responsibility Allocation by Data Management Activity Data capture and collection

- Primary: Researcher
- Support: UP Library Services (provides operational data); UP IT Services (provides technical infrastructure for digital collection)
- Oversight: Supervisor

Metadata production

- Primary: Researcher
- Review: Supervisor
- Standards guidance: UP Figshare administrators (DataCite, Dublin Core); disciplinary norms (DDI for social science data)

Data quality

- Primary: Researcher
- Spot-check: Supervisor
- Institutional verification: UP Library Services (confirms operational data accuracy); UP Finance Department (confirms cost data accuracy)

Storage and backup

- Primary: Researcher (daily execution)
- Infrastructure: UP IT Services (maintains systems)
- Verification: Researcher (monthly); Supervisor (annual recovery testing)

Data security and access control

- Primary: Researcher
- Policy enforcement: UP IT Services (MFA, encryption tools); UP Information Security Office (incident response)
- Compliance monitoring: Supervisor; UP Research Ethics Committee

Ethical compliance

- Primary: Researcher
- Approval: UP Research Ethics Committee
- Guidance: Supervisor; UP Research Ethics Committee

Data archiving and preservation

- Primary: Researcher (preparation and deposit)
- Curation: UP Figshare and UPSpace administrators
- Clearance: UP Library Services; UP Finance Department (for cost-containing outputs)

Data sharing and access management

- Primary: Researcher (enquiries, applications, collaboration proposals)

- Platform management: UP Figshare administrators (handles technical aspects of restricted access workflow)
- Policy enforcement: Repository terms of use; Creative Commons licence terms

DMP implementation and review

- Primary: Researcher
- Approval and oversight: Supervisor
- Institutional compliance: UP Research Ethics Committee; UP Library Services

4. Collaborative and Multi-Site Considerations This research does not involve multi-partner collaborative sites or external consortium arrangements. However, the multi-campus nature of UP requires clarity on site-level coordination. **UP multi-campus coordination:**

- Researcher is responsible for data collection across all UP library sites (Hatfield, Groenkloof, Prinshof, Mamelodi, Onderstepoort, Hillcrest)
- UP Library Services site managers at each campus are not formal collaborators but operational contacts
- Researcher must coordinate with site managers for access, headcount scheduling, and local context
- Site managers have no data management responsibilities but may provide local logistical support
- All data collected across sites flows to researcher as single data steward; no distributed data ownership

If future collaboration emerges: Should the research expand to include comparative analysis with other South African universities (e.g., University of the Witwatersrand, University of Cape Town, Stellenbosch University), the following would be required:

- Formal consortium agreement or memorandum of understanding between institutions
- Named data stewards at each partner site
- Agreed metadata standards and data formats
- Cross-institutional ethics clearance
- Data sharing agreement between partner institutions
- Designated lead institution for repository deposit and DOI assignment
- This DMP would be revised to reflect distributed responsibilities

No such collaboration is currently planned; this section serves as contingency planning. **5. Data Ownership and RDM in Agreements** Data ownership and research data management responsibilities are explicitly addressed in the following formal instruments: **Data Sharing Agreement with UP Library Services**

- Institutional ownership of operational data affirmed
- Researcher granted licensed use for stated research purpose
- Researcher obligations for ethical handling, anonymisation, and destruction specified
- Institutional rights to review outputs and use findings internally reserved
- Researcher moral rights (attribution) acknowledged
- Repository deposit and licensing terms agreed

Data Access Protocol with UP Finance Department

- Institutional ownership of financial data affirmed
- Aggregate-only access granted to researcher
- Researcher obligations for redaction and pre-publication review specified
- No researcher ownership claims over institutional financial information

Research Ethics Application and Consent Forms

- Participant ownership of personal contributions acknowledged (moral rights)

- Researcher and UP ownership of compiled anonymised dataset affirmed
- Participant rights to access, correction, and deletion specified
- Consent for sharing and reuse obtained where applicable

Thesis and Publication Agreements

- UP retains ownership of thesis as institutional requirement for higher degree
- Researcher retains moral rights (attribution)
- Publisher agreements govern article-specific copyright and licensing
- Dataset ownership and licensing governed by repository deposit terms (Creative Commons)

No consortium agreement required as no multi-partner collaboration exists. **If researcher becomes unavailable (illness, withdrawal, completion and departure):**

- Supervisor assumes primary contact role for data queries during retention period
- UP Figshare administrators manage repository records and restricted access applications
- UP Library Services retains institutional copy of findings for service improvement
- Raw data already destroyed; no recovery issues
- If thesis incomplete, supervisor and UP Research Office determine whether data can be used by replacement researcher (governed by ethics committee and data sharing agreement)

If supervisor becomes unavailable:

- Researcher consults alternate supervisor or departmental postgraduate coordinator
- UP Research Ethics Committee provides ethical guidance directly
- DMP review continues with new supervisory assignment
- Repository deposit proceeds with administrator support

If institutional conflict arises (UP Library Services disputes findings or denies clearance):

- Researcher and supervisor engage with UP Library Services senior management to resolve
- If unresolved, UP Research Office or Faculty Dean mediates
- Researcher retains right to publish aggregate data not subject to institutional clearance; operational recommendations may be embargoed or redacted pending resolution
- Ethics committee consulted if participant rights or research integrity at stake

8. Reflection: Responsibility Learning from the Maule Case Study Maule's case study reveals both effective personal responsibility and systemic gaps in distributed accountability.

- **Concentrated responsibility enabled agility:** Maule personally drove the entire data journey —from identifying the need, to extracting data, to cleaning, analysing, presenting, and implementing change. This concentration meant rapid decision-making and coherent vision. My allocation of primary responsibility to myself as researcher mirrors this, while adding supervisory oversight to prevent unilateral errors.
- **No named institutional data steward:** The security officers collected data but had no data management responsibility; senior managers used data but did not steward it; the data existed in a responsibility vacuum for five years. My explicit naming of UP Library Services as data provider with formal agreement prevents this vacuum.
- **No documented backup or security responsibility:** Maule does not mention who would recover data if her laptop failed, who managed the cloud transition, or who ensured the master document remained accessible. My explicit allocation to UP IT Services for infrastructure and to myself for execution, with supervisor verification, addresses this gap.
- **No ethics or compliance oversight named:** The Westminster project appears to have operated without formal ethics review or POPIA-equivalent oversight, presumably because it used operational data rather than human participant research. My inclusion of UP Research Ethics Committee and explicit ethical responsibilities is essential given my qualitative data collection.
- **Post-project continuity was unaddressed:** Maule moved on; the data culture she catalysed

survived or faded depending on subsequent management. My repository deposit and institutional agreements ensure that responsibility for preservation and access transitions formally to UP infrastructure rather than depending on individual memory.

The critical lesson is that **responsibility must be named, allocated, and documented**. Vague assumptions that "the institution" or "someone" will manage data lead to the very fragmentation and loss that Maule's project overcame. My DMP names every responsible party, specifies every activity, and provides contingency for every absence.

What resources will you require to deliver your plan?

- Tools:
  - Excel
  - Survey platforms
  - Data storage systems
- **1. Specialist Expertise and Training** This research requires specific expertise across data management, ethics, analysis, and institutional engagement. The following assesses whether existing capabilities are sufficient or whether additional training is required. **Data management and curation expertise**
  - Current status: Basic understanding of file organisation, spreadsheet management, and backup procedures; awareness of metadata importance but limited practical experience with standards (Dublin Core, DataCite, DDI)
  - Gap identified: Need for systematic training in research data management best practices, metadata creation, and repository deposit procedures
  - Training required: UP Research Data Management workshop or equivalent online training (e.g., UK Data Service training modules, MANTRA online course)
  - Provider: UP Library Services research support team; or self-directed online learning
  - Duration: 1–2 days
  - Cost: Free if UP-provided; free for self-directed online modules
  - Justification: Ensures DMP implemented to institutional and disciplinary standards; reduces risk of metadata errors that compromise discoverability; builds transferable professional skill

  POPIA and research ethics compliance expertise
  - Current status: General awareness of ethical research principles; basic understanding of consent requirements; limited specific knowledge of South African POPIA requirements for research
  - Gap identified: Need for detailed understanding of lawful processing of personal information, anonymisation standards, breach notification procedures, and cross-border data transfer restrictions
  - Training required: UP Research Ethics Committee orientation session; POPIA compliance workshop for researchers; review of Information Regulator guidance documents
  - Provider: UP Research Ethics Committee; UP Legal Services or Information Governance office; Information Regulator online resources
  - Duration: Half day to 1 day
  - Cost: Free if UP-provided
  - Justification: Legal compliance is non-negotiable; incorrect handling of personal information exposes researcher, supervisor, and institution to regulatory action and reputational harm; training reduces risk of inadvertent breach

  Qualitative data analysis expertise
  - Current status: Familiarity with basic thematic analysis; experience with spreadsheet and document analysis; limited experience with qualitative data analysis software
  - Gap identified: Need for systematic coding, inter-coder reliability procedures, and software-assisted qualitative analysis if dataset is large
  - Training required: Introduction to qualitative analysis software (NVivo, Atlas.ti, or open-

source alternative Taguette); or manual coding framework development if software not used
    - Provider: UP postgraduate skills programme; software vendor tutorials; self-directed learning
    - Duration: 2–3 days for software proficiency; 1 day for manual coding framework
    - Cost: NVivo/Atlas.ti licences may require purchase if not institutionally provided; open-source alternatives free. Check UP software licence availability through IT Services or department
    - Justification: Qualitative data quality depends on systematic and transparent analysis; software enhances audit trail and reproducibility; manual coding requires rigorous framework development training to avoid bias

  Data visualisation and presentation expertise
    - Current status: Basic chart creation in Excel; standard presentation skills
    - Gap identified: Need for effective visual communication of complex usage and cost data to diverse stakeholders (students, library staff, senior management, finance)
    - Training required: Data visualisation best practices workshop; stakeholder communication training; potentially specific software (Tableau Public, Power BI, or R ggplot2)
    - Provider: UP postgraduate skills programme; online platforms (Coursera, DataCamp); self-directed learning
    - Duration: 2–3 days
    - Cost: Free to low cost depending on provider; Tableau Public free; R free; DataCamp subscription may apply
    - Justification: Maule's success depended significantly on clear visualisations that made data accessible to non-specialist stakeholders; effective presentation builds institutional support for recommendations

  Statistical analysis expertise (if advanced analysis required)
    - Current status: Descriptive statistics competence; basic inferential statistics; spreadsheet-based analysis
    - Gap identified: May need regression analysis, time-series analysis, or cost-modelling techniques depending on analytical depth required
    - Training required: Statistical analysis software (R, SPSS, or Stata); or advanced Excel functions; specific techniques as needed
    - Provider: UP statistics support service; department statistical consultant; online learning
    - Duration: Variable; 3–5 days for software proficiency; ongoing as techniques needed
    - Cost: R and RStudio free; SPSS/Stata may require licence. Check UP institutional licence availability
    - Justification: Cost-effectiveness analysis may require more than descriptive statistics; robust analytical methods strengthen findings and examiner confidence

- **Total training time estimated: 9–14 days Training scheduling:**
    - Months 1–2: Research data management; POPIA and ethics (before data collection begins)
    - Months 3–4: Qualitative analysis software; data visualisation (parallel to early data collection)
    - Months 5–6: Statistical analysis as analytical needs become clear
    - Ongoing: Self-directed learning and peer consultation as specific challenges emerge

  2. Hardware Requirements Existing institutional provision assessment:
    - Laptop or desktop computer: Assumed available to researcher as registered UP postgraduate student; standard institutional or personal device
    - Mobile device (smartphone/tablet): Assumed available for field data collection and communication
    - External storage: UP does not typically provide encrypted external hard drives as standard; researcher may need to acquire

  Additional or exceptional hardware required: Encrypted external hard drive
    - Purpose: Offline encrypted backup for sensitive data; protection against ransomware; cold storage for irreplaceable qualitative audio recordings before network transfer
    - Specification: Minimum 1 TB capacity; hardware encryption (not just software password); USB 3.0 or later interface; ruggedised if field transport required
    - Justification: UP network drives and OneDrive provide primary and secondary backup, but an offline encrypted copy is essential for ransomware protection and as recovery source if

institutional systems fail. Maule's project benefited from comprehensive data collection over five years; losing such data to hardware failure or cyber attack would be catastrophic.
    - Cost: Approximately R800–R2,000 depending on capacity and brand (e.g., Samsung T7 Shield, SanDisk Extreme Pro, or equivalent)
    - Funding: Researcher personal funds; departmental postgraduate support if available; supervisor research funds if allocated
    - Alternative: If funding unavailable, researcher must ensure exceptionally rigorous institutional backup verification and accept higher ransomware risk

  Digital audio recorder (if focus groups conducted)
    - Purpose: High-quality audio recording for focus group transcription; backup to smartphone recording
    - Specification: Minimum 16 GB internal storage; MP3/WAV format; USB connectivity; password protection if available
    - Justification: Smartphone audio may be adequate but dedicated recorder provides reliability, longer battery life, and better audio quality for transcription accuracy. Clear audio reduces transcription time and cost, and improves qualitative data quality.
    - Cost: Approximately R1,500–R3,500 depending on specifications (e.g., Zoom H1n, Olympus WS series, or equivalent)
    - Funding: Researcher personal funds; departmental equipment loan if available
    - Alternative: Use institutional tablet or laptop with recording software if available; or smartphone with high-quality recording app and external microphone

  Tablet for field data collection (optional but recommended)
    - Purpose: Digital capture of manual headcounts; electronic survey administration; immediate sync to institutional storage
    - Specification: Minimum 10-inch screen; Wi-Fi connectivity; institutional device management compatible; rugged case if field use
    - Justification: Paper logbooks introduce manual transcription errors, delay data entry, and create physical security risks (loss, damage, unauthorised access). Digital capture with immediate cloud sync eliminates these risks and aligns with Maule's finding that manual processes were a primary barrier to analysis.
    - Cost: Approximately R3,000–R6,000 for basic tablet; more for institutional-grade ruggedised device
    - Funding: UP Library Services may loan device if research benefits their operations; departmental equipment pool; researcher personal funds
    - Alternative: Use personal smartphone with institutional app (e.g., Microsoft Forms, REDCap mobile) if screen size adequate for data entry
- **3. Software Requirements Existing institutional provision assessment:**
    - Microsoft Office suite (Word, Excel, PowerPoint): Available through UP Microsoft 365 licence to all registered students
    - UP OneDrive cloud storage: 1 TB per user through Microsoft 365
    - UP network drives: Available via institutional login
    - Antivirus and endpoint protection: Provided by UP IT Services
    - VPN access: Provided for remote institutional resource access

  Additional or exceptional software required: Qualitative data analysis software
    - Options: NVivo (Windows/Mac); Atlas.ti (Windows/Mac/Web); Taguette (open-source, free); QualCoder (open-source, free)
    - Justification: Manual coding of qualitative data is feasible for small datasets but becomes unwieldy and error-prone with larger volumes. Software provides systematic coding, query functions, visualisation, and audit trails that enhance rigour and transparency.
    - Cost: NVivo individual licence approximately R8,000–R12,000 (one-time or subscription); Atlas.ti similar; Taguette and QualCoder free
    - Institutional availability: Check with UP IT Services or department whether institutional licences available for postgraduate use. Many universities site-licence NVivo or Atlas.ti for research.

- Recommendation: If institutional licence unavailable, use Taguette or QualCoder (free, open-source, adequate for most needs) rather than personal purchase of commercial software. Alternatively, manual coding with rigorous documentation if dataset small.

Statistical analysis software

- Options: R and RStudio (free, open-source); SPSS (commercial); Stata (commercial); Python with pandas/statsmodels/scikit-learn (free)
- Justification: Excel is adequate for descriptive statistics and basic visualisation but may struggle with large datasets, complex models, or reproducible analysis workflows. R or Python provide powerful, free alternatives with strong academic community support.
- Cost: R and RStudio completely free; Python free; SPSS approximately R15,000–R30,000 per year subscription; Stata similarly priced
- Institutional availability: Check UP statistics department or IT Services for SPSS/Stata site licence availability
- Recommendation: Use R with RStudio unless specific institutional or supervisory preference for SPSS/Stata. R is free, has extensive library support for library research (bibliometric packages, visualisation), and builds transferable skill. Learning curve is moderate but manageable with training.

Data visualisation software

- Options: Excel (existing); R ggplot2 (free with R); Tableau Public (free for public use); Power BI (free desktop version; UP may have Pro licence)
- Justification: Effective stakeholder communication requires professional visualisations. Excel is limiting for complex multi-site, multi-temporal data.
- Cost: Tableau Public free; Power BI Desktop free; Power BI Pro may require licence if institutional sharing needed
- Institutional availability: Check UP IT Services for Power BI Pro availability
- Recommendation: Start with Excel and R ggplot2 (free, integrated with analysis workflow). Upgrade to Tableau Public or Power BI if specific interactive dashboard requirements emerge for stakeholder engagement.

Reference management software

- Options: Mendeley (free basic version); Zotero (free, open-source); EndNote (commercial); RefWorks (institutional subscription)
- Justification: Systematic management of literature, sources, and citations is essential for thesis and publications.
- Cost: Zotero completely free; Mendeley free with limited storage; EndNote approximately R4,000–R6,000
- Institutional availability: UP Library Services may provide EndNote or RefWorks access; check current subscriptions
- Recommendation: Zotero if no institutional EndNote/RefWorks available; free, powerful, integrates with Word, and has strong community support.

Repository deposit and metadata tools

- Options: UP Figshare web interface (free through institutional subscription); DataCite metadata generator (free); DDI tools (free online)
- Justification: No additional software needed; institutional repository provides web-based deposit interface
- Cost: None

- **4. Data Repository Charges UP Figshare**
  - Charge to researcher: None
  - Basis: UP maintains institutional subscription as part of research infrastructure
  - Services included: Unlimited standard dataset deposits; DOI assignment; embargo functionality; basic usage statistics; metadata support
  - Potential additional charges: None anticipated for this project scale (estimated 1–2 GB total)

UPSpace

- Charge to researcher: None
- Basis: Mandatory institutional repository for higher degree theses; funded through UP Library

Services
- Services included: Permanent thesis archiving; OAI-PMH harvesting; Google Scholar indexing; basic usage statistics

Disciplinary or national repositories (conditional)
- Zenodo: Free for open science deposits up to 50 GB per dataset
- OSF: Free for academic use
- LIASA / DHET portals: No charges anticipated
- Note: Only used for copies already deposited in UP Figshare; no additional cost

Third-party commercial repositories
- Not used for this research
- Rationale: Institutional repositories provide adequate service without cost; commercial repositories (e.g., Dryad, Figshare individual plan if UP subscription unavailable) would incur charges but are unnecessary given UP provision

Total repository charges: R0 5. Other Resource Considerations Transcription services (if focus groups conducted)
- Purpose: Converting audio recordings to text for qualitative analysis
- Options: Manual transcription by researcher (time-intensive but free); professional transcription service (costly but faster and potentially more accurate); automated transcription software (Otter.ai, Trint, Descript—subscription costs apply but lower than professional services)
- Cost estimate: Professional service approximately R10–R20 per audio minute (R600–R1,200 per 60-minute focus group); automated software approximately R200–R500 per month subscription
- Justification: Maule's project did not involve audio transcription, but if focus groups are primary qualitative method, accurate transcription is essential. Researcher time for manual transcription is substantial (estimated 4–6 hours per hour of audio) but builds intimate data familiarity.
- Recommendation: Manual transcription by researcher for first 1–2 focus groups to build analytical immersion; assess whether professional or automated assistance needed for remainder based on time constraints and accuracy requirements.

Printing and presentation materials
- Purpose: Stakeholder consultation presentations; printed reports for UP Library Services and committees; thesis printing and binding
- Cost estimate: R2,000–R5,000 depending on consultation frequency and thesis requirements
- Justification: Effective stakeholder engagement requires professional presentation; thesis submission requires bound copies per UP requirements
- Funding: Researcher personal funds; departmental conference/travel fund if available; supervisor research support if allocated

Travel between UP campuses
- Purpose: Data collection at multiple library sites (Hatfield, Groenkloof, Prinshof, Mamelodi, Onderstepoort, Hillcrest)
- Cost estimate: UP shuttle services may be free for registered students; private transport or public transport costs variable depending on campus and frequency
- Justification: Multi-campus data collection is essential for representative findings; cannot rely solely on main campus data
- Funding: UP shuttle if available; personal transport funds if shuttle insufficient; departmental fieldwork support if available

Internet connectivity
- Purpose: Remote access to institutional systems; cloud sync; online research; video conferencing for supervision and stakeholder meetings
- Assumption: Researcher has adequate home or campus connectivity as standard for registered student
- Exceptional need: If conducting fieldwork at campuses with poor Wi-Fi, mobile data may be required for immediate sync

    - Cost: Standard mobile data plan; no exceptional requirement anticipated
- **Reflection: Resource Planning Learning from the Maule Case Study** Maule's case study reveals resource implications that inform this planning.
    - **No training investment was mentioned:** Maule and colleagues appear to have learned data management through trial and error, noting that compiling data "took more time than I first envisaged and was a learning curve." My explicit training allocation reduces this learning curve, improving efficiency and reducing error.
    - **Manual processes were resource-intensive:** Manual data input by dedicated staff members across four sites consumed significant staff time. My investment in digital field capture (tablet or smartphone) and automated systems reduces manual labour and associated error.
    - **No hardware or software investment was discussed:** The Westminster project operated within existing institutional provision (Excel, shared drives, manual logbooks). My assessment identifies where existing provision is adequate and where modest additional investment (encrypted drive, potential audio recorder) significantly improves security and data quality.
    - **The cost of poor data structure was high:** Maule spent considerable effort reformatting "poorly laid out" spreadsheets. My investment in training and potentially software (R for reproducible data cleaning scripts) prevents this costly rework.
    - **Institutional repository value was unexploited:** No mention of DOI, persistent identifiers, or long-term preservation. My zero-cost repository deposit (UP Figshare) provides immense long-term value without financial burden.

  The critical lesson is that **modest, targeted resource investment in training and hardware yields disproportionate returns in data quality, efficiency, and long-term value**. The free availability of high-quality software (R, Zotero, Taguette, Tableau Public) and institutional repository services means that financial constraints need not compromise research data management standards.