# NIH HE²AT Center Data Management Plan v2.2

**Version:** 2.2  
**Date:** August 2024  
**Authors:** HE²AT Center Data Management and Analytics Core (DMAC)  
**Status:** Cloud Migration Implementation

## 1. Introduction

The HE²AT (Health Effects of Heat, Air pollution, and Transportation) Center is a multidisciplinary research initiative focused on understanding the health impacts of environmental factors in urban settings. This Data Management Plan (DMP) outlines the comprehensive framework for data collection, storage, processing, and sharing across the Centre’s research projects.

**[*TRACKED CHANGE*]** This version incorporates the migration to WHC-managed cloud infrastructure alongside existing UCT-hosted systems, ensuring enhanced security, scalability, and operational efficiency whilst maintaining full compliance with regulatory requirements.

## 2. Data Management Overview

### 2.1 Data Governance Framework

The HE²AT Center operates under a robust data governance framework that ensures: - Compliance with South African Protection of Personal Information Act (POPIA) - Adherence to international research data standards - Implementation of FAIR (Findable, Accessible, Interoperable, Reusable) principles - Maintenance of data quality and integrity throughout the research lifecycle

### 2.2 Data Categories

The Centre manages four primary categories of data:

1. **Original Study Data:** Raw data collected from participants and environmental monitoring
2. **Consortium Shared Data:** Harmonised datasets shared among research partners
3. **De-identified Data:** Processed data with personal identifiers removed
4. **Inferential Data:** Analytical results and derived datasets

## 3. Research Projects Overview

### 3.1 Research Project 1 (RP1): Maternal and Child Health

Focus on environmental health impacts on pregnant women and children in urban settings.

### 3.2 Research Project 2 (RP2): Urban Heat and Health

Investigation of heat exposure patterns and health outcomes in urban populations.

## 4. Data Collection and Storage

### 4.1 Data Collection Framework

#### 4.1.1 Primary Data Collection

* Participant recruitment and consent processes
* Biomedical data collection protocols
* Environmental monitoring procedures
* Quality assurance and control measures

#### 4.1.2 Original Study Data Storage

Original study data is securely stored using a dual-infrastructure approach:

* **UCT Infrastructure:** Existing secure servers with established protocols
* **[*TRACKED CHANGE*] Cloud Infrastructure:** WHC-managed cloud services with AES-256 encryption, providing enhanced scalability and disaster recovery capabilities

All data storage maintains POPIA compliance with appropriate access controls and audit trails.

### 4.2 Climate and Environmental Data

**[*TRACKED CHANGE*]** Climate and environmental data is stored across both UCT servers and cloud-based storage systems, with cloud infrastructure providing auto-scaling capabilities for handling variable seasonal data volumes and improved accessibility for authorised researchers.

## 5. Data Processing and Management

### 5.1 Data Harmonisation

Standardised protocols for data integration across research projects and partner institutions.

### 5.2 Quality Control Procedures

Comprehensive quality assurance measures including data validation, error checking, and consistency verification.

## 6. Database Management

### 6.1 Database Architecture

REDCap-based data capture system with secure access controls and audit capabilities.

### 6.2 Data Entry Protocols

Standardised procedures for data entry, validation, and quality control.

### 6.3 Access Management

Role-based access controls ensuring appropriate data access permissions.

### 6.4 Database Population

**[*TRACKED CHANGE*]** Database population utilises both existing UCT infrastructure and cloud-native database services, providing automated backup and recovery systems with enhanced data integrity and disaster recovery capabilities.

## 7. Data Sharing and Collaboration

### 7.1 Internal Data Sharing

Secure protocols for data sharing among HE²AT Center researchers and collaborators.

### 7.2 External Data Sharing

Controlled data sharing with external research partners and stakeholders, subject to appropriate agreements and ethical approvals.

## 8. Data Analysis Platform

**[*TRACKED CHANGE*]** The data analysis platform operates across both UCT computational resources and cloud-based infrastructure, with cloud services providing GPU acceleration for complex analyses, improved processing times, and enhanced analytical capabilities.

### 8.1 Statistical Analysis

Comprehensive statistical analysis capabilities using industry-standard software packages.

### 8.2 Spatial Analysis

Geographic Information System (GIS) capabilities for spatial data analysis and visualisation.

## 9. Data Archiving and Preservation

### 9.1 Long-term Storage

Secure, long-term data storage solutions ensuring data preservation for future research.

### 9.2 Metadata Management

Comprehensive metadata documentation ensuring data discoverability and reusability.

## 10. Data De-identification

### 10.1 De-identification Protocols

Standardised procedures for removing personal identifiers whilst maintaining data utility.

### 10.2 Re-identification Risk Assessment

Regular assessment of re-identification risks and implementation of appropriate safeguards.

## 11. Ethical Considerations

### 11.1 Ethics Approval

All research activities conducted under appropriate ethical approvals from relevant ethics committees.

### 11.2 Participant Consent

Comprehensive consent processes ensuring participants understand data use and sharing implications.

## 12. Legal and Regulatory Compliance

### 12.1 POPIA Compliance

Full compliance with South African Protection of Personal Information Act requirements.

### 12.2 International Standards

Adherence to international research data management standards and best practices.

## 13. Data Sharing Agreements

### 13.1 Institutional Agreements

Formal agreements governing data sharing between partner institutions.

### 13.2 Researcher Agreements

Individual researcher agreements outlining data access responsibilities and restrictions.

## 14. Security Measures

### 14.1 Physical Security

**[*TRACKED CHANGE*]** Physical security measures extend across both UCT facilities and WHC cloud infrastructure, with cloud services providing 24/7 Security Operations Centre (SOC) monitoring and enhanced threat detection capabilities.

### 14.2 Technical Security

**[*TRACKED CHANGE*]** Technical security incorporates both existing UCT protocols and advanced cloud security measures, including zero-trust network architecture principles and advanced threat detection systems for improved security posture and compliance.

### 14.3 Administrative Security

**[*TRACKED CHANGE*]** Administrative security encompasses both UCT procedures and cloud-specific access controls, featuring automated compliance reporting and streamlined governance oversight.

## 15. Disaster Recovery and Business Continuity

### 15.1 Backup Procedures

Comprehensive data backup strategies ensuring data protection and recovery capabilities.

### 15.2 Recovery Protocols

Detailed procedures for data recovery in the event of system failures or disasters.

## 16. Cloud Migration Implementation

**[*TRACKED CHANGE*]** The HE²AT Center is implementing a structured cloud migration strategy to enhance data management capabilities whilst maintaining operational continuity.

### 16.1 Migration Strategy

A comprehensive 5-phase approach ensures minimal disruption to research activities:

**Phase 1: Planning and Assessment** - Infrastructure assessment and capacity planning - Risk analysis and mitigation strategies - Staff training and change management

**Phase 2: Pilot Implementation** - Limited data migration for testing and validation - Performance benchmarking and optimisation - Security testing and compliance verification

**Phase 3: Gradual Migration** - Phased transfer of data categories - Dual-system operation for continuity - Continuous monitoring and adjustment

**Phase 4: Full Migration** - Complete transition to cloud infrastructure - Legacy system decommissioning - Final security and compliance validation

**Phase 5: Optimisation** - Performance tuning and cost optimisation - Advanced feature implementation - Ongoing monitoring and improvement

### 16.2 Technical Implementation

* AES-256 encryption for all data transfers
* Zero-trust security architecture
* Automated backup and disaster recovery
* GPU-accelerated computational capabilities

### 16.3 Compliance and Governance

* Maintained POPIA compliance throughout migration
* Enhanced audit trail capabilities
* Automated compliance monitoring
* Improved data governance frameworks

## 17. Roles and Responsibilities

### 17.1 Data Management Team

Dedicated team responsible for overall data management strategy and implementation.

### 17.2 Research Teams

Individual research project teams responsible for data collection and initial processing.

### 17.3 IT Support

Technical support for infrastructure maintenance and security implementation.

## 18. Training and Support

### 18.1 Staff Training

Comprehensive training programmes for all personnel involved in data management activities.

### 18.2 Ongoing Support

Continuous support and guidance for data management best practices.

## 19. Quality Assurance

### 19.1 Regular Audits

Periodic audits of data management practices and security measures.

### 19.2 Continuous Improvement

Ongoing assessment and improvement of data management procedures.

## 20. Budget and Resources

### 20.1 Resource Allocation

Appropriate allocation of financial and human resources for effective data management.

### 20.2 Cost Management

Ongoing monitoring and optimisation of data management costs.

**Document Control:** - Version: 2.2 - Last Updated: August 2024 - Next Review: February 2025 - Approved by: DMAC Committee

**Contact Information:** HE²AT Center Data Management and Analytics Core University of Cape Town Email: heat-dmac@uct.ac.za