Policy 2 - Data Preservation Policy

This document should be read in conjunction with the HSRC Data Sharing Policy, the HSRC Data Curation Standard Operating Procedure (SOP), Project Management Framework, applicable IT related policies and procedures, as well as the HR Termination Policy.

1.   Introduction

The focus on social science and humanities data has been articulated as an objective in the Human Sciences Research Council Act (Act No. 17 of 2008) that states that the HSRC has to develop and make available data sets underpinning research, policy development and public discussion of developmental issues. Data preservation as a prerequisite for data sharing and proper data management has a high priority in the HSRC as data is fundamental to the research undertaken. In doing so the HSRC recognises the risks associated with digital objects, such as data sets, relating to technology obsolescence, physical damage, unauthorised access and changes, as well as reduced usability over time.

2.   Purpose of the Data Preservation Policy

This policy aims to provide a broad framework for the preservation of HSRC-generated research data. It contains a set of guidelines to ensure the authenticity, reliability, logical integrity and usability of research data over time.

3.   Scope the Data Preservation Policy

This policy applies to:

- HSRC-generated research data;
- Data yielded by publicly-funded research;
- Data that is privately funded.

Data refers to the digitally recorded factual material which enables and also validates research findings (OMB 2011). It can be both quantitative and qualitative in nature. As data is only usable within a particular context, the preservation of data implies the preservation of data files together with its associated metadata, e.g. information about the study and data related documentation, such as data collection instruments and consent / assent forms, training manuals and ethics applications, amongst others. Documents which do not specifically relate to the data are excluded from this policy.

This policy does not cater for the preservation of web pages, internal administrative documents and correspondence, software, hardware, hard-copy material (paper) and physical objects, as well as analogue recordings.

Preservation refers to a series of managed activities which ensure the authenticity, reliability and logical integrity of digital objects to enable meaningful continued access, for as long as it is required (UKDA 2012, JISC 2013).

This policy excludes guidelines on data sharing which is governed by the HSRC Data Sharing Policy.

4. Policy statements

4.1 Data must be cared for by following best practice data management during the research process

- Data preservation planning in terms of the HSRC Data Preservation and Sharing Plan is mandatory for every proposed research project. It forms part of the ethics approval application process as stipulated by the HSRC Research Ethics Committee.
- Collected research data must be recorded and stored in digital form. Data must be collected, cleaned, processed and described according to accepted best practices. If personal data is collected the data must be handled in accordance with the stipulations articulated in the Protection of Personal Information Act (Act No. 4 of 2013) and the Electronic Communications and Transactions Act (Act No. 25 of 2002).

During the research process procedures must be followed to protect data from unauthorised access, damage or loss.

4.2 Data must be safe-guarded from loss

Data emanating from all projects, as well as the associated metadata and related documentation must be deposited for preservation as soon as the first final version of the data becomes available after data collection or within 6 months after data collection.

On resignation, employees responsible for research data must ensure handover of data files and related information to the Executive Director or delegated staff member.

4.3 All deposited data which conforms to the HSRC's data deposit standards will be preserved

Acceptable digital resource types include, but are not limited to text documents, data sets designed for use in spreadsheets, databases and statistical packages – both micro and macro / aggregated data and derived data, digital still images, audio recordings and

Social science that makes a difference

moving image (video) recordings, as well as GIS data. All information packages will be retained for the lifetime of the HSRC's data repository.

4.4 Data is preserved according to a set processing protocol

Data is processed according to one of two set standards (A or B) which determine the level of curation that is undertaken.

The "A" preservation standard applies to data sets which the HSRC owns or has the right to re-use or share. These data sets meet the objective of the HSRC Act in terms of the preservation and sharing of data sets and will be subjected to long term preservation procedures.

The "B" preservation standard applies to data sets which the HSRC does not own or does not have the right to re-use or disseminate. This implies that the curation actions of level A are scaled down for record keeping purposes or to cater for a possible change in preservation status to level "A" sometime in the future. This information will not be subjected to long term preservation procedures.

4.5 Data is preserved according to accepted international standards as adopted by the HSRC

The HSRC follows the broad guidance provided by the Open Archival Information System (OAIS) reference model within the context of its organisational objectives, but implemented according to the assigned processing protocol.

4.6 The preservation of data is a shared responsibility

Preservation spans the data life cycle in its entirety and is therefore required during the execution of research projects, as well as after project completion. Various parties are required to participate in the data preservation process, including the DCEO: Research, Executive Directors, researchers / data managers, principal investigators / project leaders, research data curators, the directors of IT and HR, as well as third party preservation service providers. The responsibility of HSRC staff may be transferred to a designated alternative.

5. Support for this policy
- The HSRC Act (Act No. 17 of 2008)
- Managing electronic records in governmental bodies: policy, principles and requirements - National Archives and Records Service of South Africa, DAC 2007
- Electronic Communications and Transactions Act (Act No. 25 of 2002)
- Protection of Personal Information Act (Act No.4 2013)

Social science that makes a difference

6. Definitions

Data
Data refers to computerised data sets and may be

- **quantitative data files** consisting of a matrix of numbers or words and its related metadata, such as variable labels, code labels and missing value definitions, including data with spatial references and maps based on these references, or summary data tables. The final data set might include both raw data and derived variables which would be described in the documentation associated with the data set.
- **qualitative data sets** that might include transcripts, thematic coding, conceptual maps, images, audio and video recordings amongst others.

Data collection
Data collection refers to the systematic gathering of records or measurements used by researchers to undertake their research or provide an evidential record of their research from various sources, including questionnaires, interviews, observation, existing records and electronic devices. It includes gathering, capturing, cleaning and processing activities in order to produce data sets which are ready for analysis.

Data curation
Data curation is the active management of data over the corresponding life cycle of scholarly and scientific interest (DCC 2004). It embraces and goes beyond that of enhanced present-day re-use and of archival responsibility, adding value through the provision of context and linkage by publishing data in ways that facilitates re-use and promotes accountability and integration (IEEE 2005).

Data management
Data management occurring within the life cycle of a research project carried out by the research team includes all activities concerned with planning, overseeing, directing and organising data related digital objects, e.g. data files, documentation, code / syntax files and metadata which contextualises the data. Data management performed by curation staff is specifically those activities related to the preservation and dissemination of data for secondary use following the original research project.

Data sharing
Data sharing includes the actions to provide access to and facilitate the re-use of data.

Social science that makes a difference

Information packages	This is the unit of exchange within an OAIS and between an OAIS and its surrounding environment. It is a container for content and description information. Depending on the function to be achieved, different types of information packages are used, e.g. a Submission Information Package (SIP) received from a data depositor, Archival Information Package (AIP) used for preservation purposes and a Dissemination Information Package (DIP) to be shared with data users.
Metadata	Information that describes significant aspects of a digital object. In terms of research data it spans the study, the data collection, data files and individual variables and values in data files.
Preservation	Preservation refers to all actions which ensure enduring access to the full content of digital objects over time. Preservation typically occurs after a research project has been concluded and is usually carried out by a data centre or archive. "Long term preservation" specifically includes actions to ensure the usability of electronic objects over an extended period of time. These actions minimise the risk related to technology obsolescence and degradation of objects, hardware and software.
File format	A file format is a pre-established layout of data within a digital file.

8. Authorisation

This Policy Document was approved by the HSRC Board and the CEO on 20 August 2014 after consultation with interested parties.

Prof O. Shisana   **Date 20 August 2014**
CEO

Social science that makes a difference