Stewarding South African Social Science Research Data

Faye Reagon & Lucia Lötter

CODATA 22
24-27 October 2010

HSRC
Human Sciences
Research Council
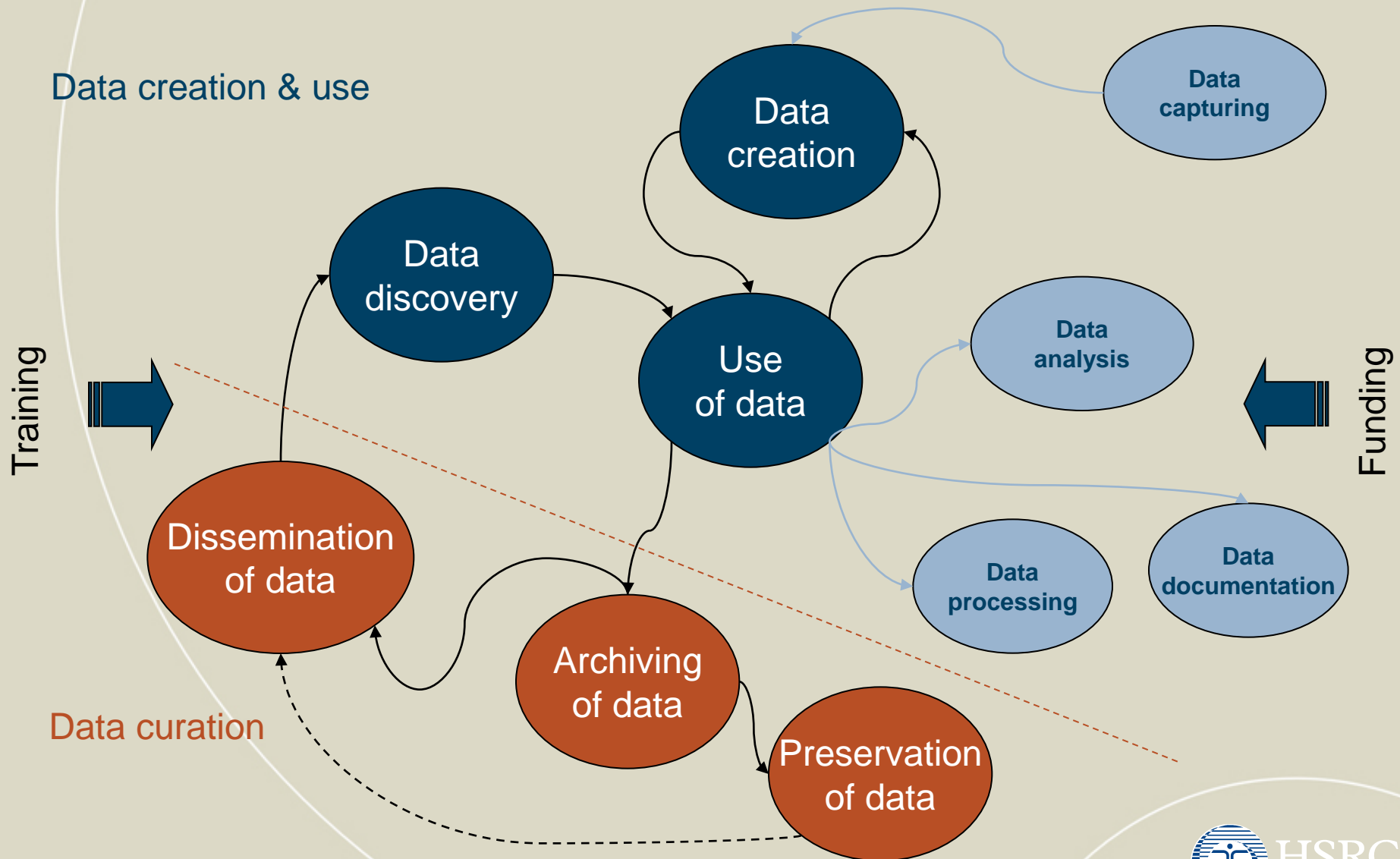
Objectives of this presentation

- Requirements for preservation and dissemination of Social Science data

- A mapping of the current landscape of entities involved in data curation in SA

- A critique of the current landscape

- The role of a national discipline specific data archive to address existing challenges
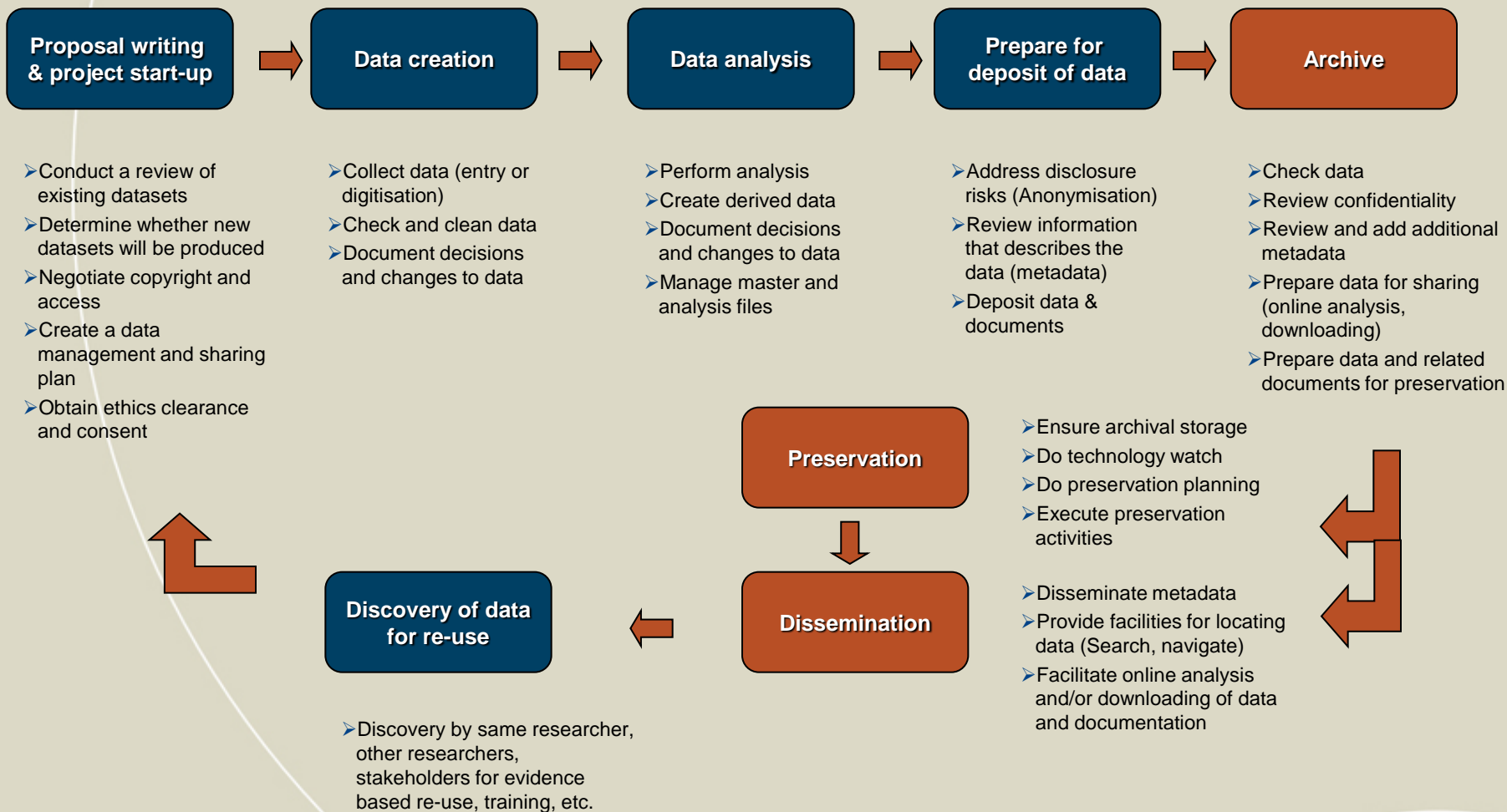
What is data curation?

"Data curation is the activity of managing and promoting the use of data from its point of creation, to ensure it is fit for contemporary purpose, and available for discovery and reuse."

Philip Lord, Alison MacDonald, 2003

Data curation and the data lifecycle

Data creation & use

Training

Funding

Data curation

HSRC
Human Sciences
Research Council

Data curation and the data lifecycle - detail

Proposal writing & project start-up → **Data creation** → **Data analysis** → **Prepare for deposit of data** → **Archive**

Proposal writing & project start-up
- Conduct a review of existing datasets
- Determine whether new datasets will be produced
- Negotiate copyright and access
- Create a data management and sharing plan
- Obtain ethics clearance and consent

Data creation
- Collect data (entry or digitisation)
- Check and clean data
- Document decisions and changes to data

Data analysis
- Perform analysis
- Create derived data
- Document decisions and changes to data
- Manage master and analysis files

Prepare for deposit of data
- Address disclosure risks (Anonymisation)
- Review information that describes the data (metadata)
- Deposit data & documents

Archive
- Check data
- Review confidentiality
- Review and add additional metadata
- Prepare data for sharing (online analysis, downloading)
- Prepare data and related documents for preservation

Preservation
- Ensure archival storage
- Do technology watch
- Do preservation planning
- Execute preservation activities

Dissemination
- Disseminate metadata
- Provide facilities for locating data (Search, navigate)
- Facilitate online analysis and/or downloading of data and documentation

Discovery of data for re-use
- Discovery by same researcher, other researchers, stakeholders for evidence based re-use, training, etc.

(Based on ICPSR Guide to Social Science Data Preparation and Archiving, 2009:5)

Who does data curation?

A Trusted Digital Repository is one whose mission is to provide reliable, long-term access to managed digital resources to its designated community, now and in the future.

Trusted Digital Repositories: Attributes and Responsibilities, 2002

Audit & certification criteria

❶ **Digital object management**
- Ingest: acquisition of content
- Ingest: creation of the archivable & dissemination packages
- Preservation planning
- Archival storage & preservation/maintenance of AIPs
- Information management
- Access management
- Dissemination

❷ **Organisational infrastructure**
- Governance & organisational viability
- Organisational structure & staffing
- Procedural accountability & policy framework
- Financial sustainability
- Contracts, licenses, & liabilities

❸ **Technologies, technical Infrastructure & security**
- System infrastructure
- Appropriate technologies
- Security

Based on Trustworthy Repositories Audit & Certification: Criteria and Checklist, 2007

HSRC
Human Sciences
Research Council

Mapping of the landscape

- National organisations
  - NSO - StatsSA
  - Statutory Research Councils (HSRC)
  - Universities (DataFirst)
  - NGOs
  - Private sector
  - Government departments on various levels
  - SADA
- International organisations
  - Where are SA/African data archived, preserved and disseminated internationally? ISSP, etc.

Survey of the current landscape

- Survey on data management and data sharing **practices** in the social sciences in South Africa
  - Organisation profile - mandate in terms of dealing with Social Science research data
  - Profile of the data being housed in the organisation
  - Data management
  - Archiving & preservation
  - Data sharing
  - Governance
  - The organisation's role in the research community
  - The need for a South African Social Science data archive
- Challenges to defining the landscape of **data related services**

HSRC
Human Sciences
Research Council

Survey of the current landscape

- Survey of **datasets** that speak to the 12 outcomes for measurable performance and accountable delivery by government.
  - Reduce poverty, grow economy and create jobs
  - Improve quality of education
  - Promote human and social advancement through skills development
  - Accelerate service delivery
  - Reduce crime
  - Develop youth, families and social cohesion
  - Promote health, health systems and wellbeing
  - HIV/AIDS and STI control
  - Development of nutrition and food security
- Challenges to defining the **landscape of data**

HSRC
Human Sciences
Research Council

Critique of the current landscape

- Organisations define data curation differently in terms of what is organisationally/operationally possible – this results in gaps in the bigger picture
- Preservation
  - Risk intensive - technology obsolescence
  - Not happening optimally in terms of criteria for TDRs
  - Data scattered within organisations and not being preserved
  - Problem with interoperability – different standards if at all
- Dissemination
  - Data scattered between different organisations
  - Difficult to identify and find data
  - Many datasets not being disseminated
  - Duplication of data because people do not know about or can't access existing data
  - Limited data-sharing culture
  - No coherent comprehensive data resource to address SA's challenges
- Fragmented, uncoordinated services – no Team SA approach
- Lack of national ownership: roles, responsibilities
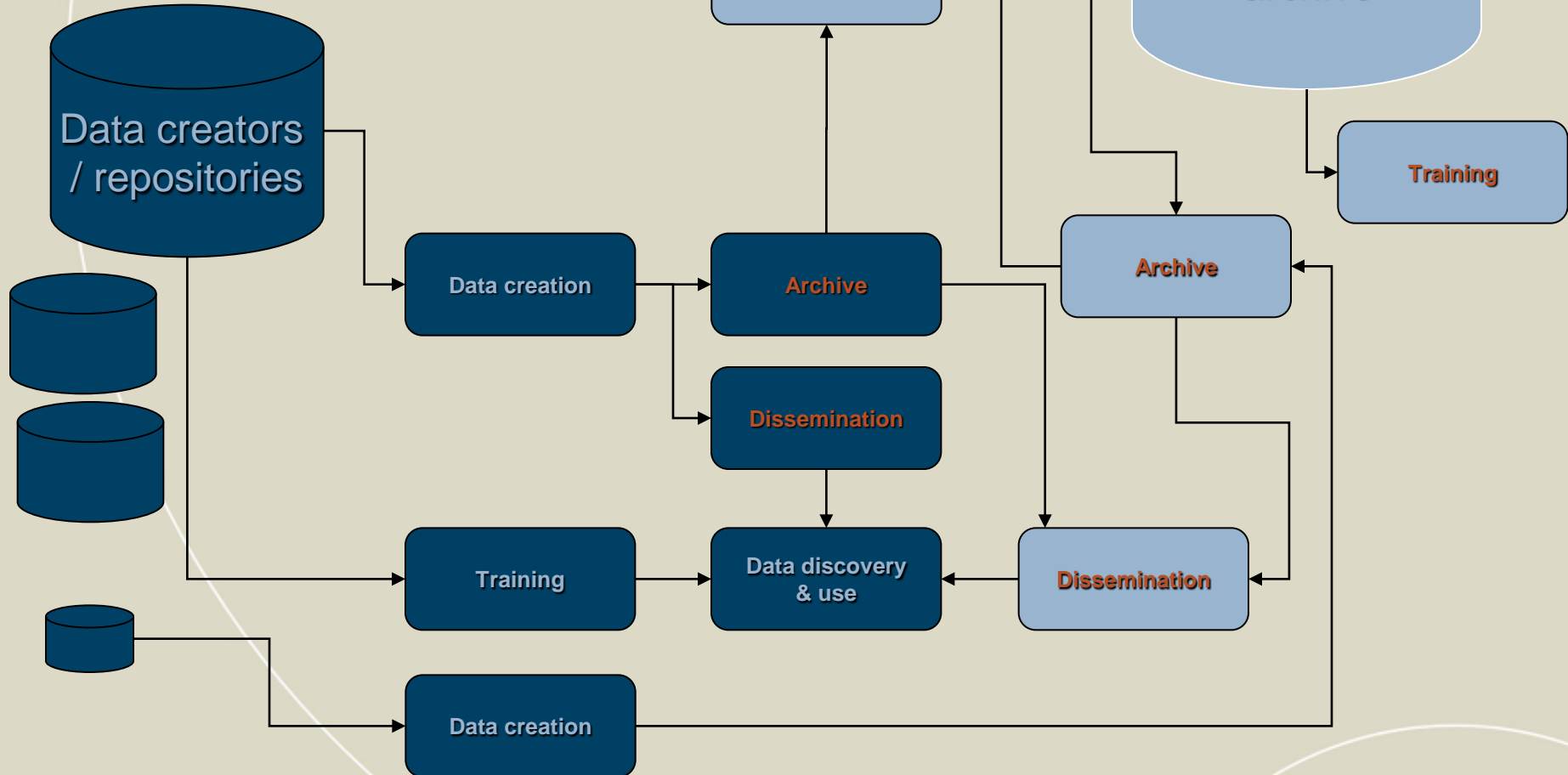- Lack of funding of essential services and support

HSRC
Human Sciences
Research Council

Why a SA Social Science Data Archive?

- Support the National System of Innovation

- Focus on evidence-based research and the importance of "Research into Use"

  - Evidence to inform human and social development in South Africa in terms of monitoring and evaluation, further analysis, as well as debate, advocacy and decision-making. The objective will be to speak to the social priorities of SA.

- The HSRC's imperatives expressed in Act as objectives:

  - (a) initiate, undertake and foster strategic basic research and applied research in human sciences, and to gather, analyse and publish data

  - (g) develop and make publicly available new data sets to underpin research, policy development and public discussion of the key issues of development

HSRC
Human Sciences
Research Council

Principles

- Data should be easy to find and use
- Data should be preserved for long term use
- Build on current initiatives
- Foster partnerships
- One source, multiple uses
- Standards (for curation, social science data)
- Cater for different kinds of data
- Protect of the rights of data producers & users
- Mechanisms must be efficient and cost-effective
- Avoid duplication, upscale benefit
- Cross disciplinary linkages
- Open access when at all possible
- High quality data, high re-use value

HSRC
Human Sciences
Research Council

The position of the archive in the landscape

HSRC
Human Sciences
Research Council

A SA Social Science Data Archive

Vision

To be a trusted digital repository for Social Science research data that informs human and social development in South Africa.

Mission

To make research data accessible and to ensure its future survival and usability by managing the datasets as an effective information source, and

- ensuring the long time preservation of these digital resources, and
- disseminating the data for secondary use

to inform monitoring and evaluation, further analysis, as well as debate, advocacy decision-making and training.

HSRC
Human Sciences
Research Council

A SA Social Science Data Archive

- Objectives

  - Establish a Trusted Digital Repository for quantitative and qualitative data in the Social Sciences in SA

  - Develop a data collection of high re-use value that speaks to the priorities of a developing country

  - Promote best practice in terms of data management and curation standards

  - Promote the use and impact of the data collection in the Archive

  - Build meaningful and enduring relationships with stakeholders – nationally and internationally

  - Demonstrate excellence in accountability and operations

HSRC
Human Sciences
Research Council

A SA Social Science Data Archive

- Benefits

    - Collections of high value can be created as datasets grow into collections and become readily available for re-use

    - Transform the disparate collections of research data into a cohesive corpus of research resources

    - Provide a readily accessible resource of data for research that inform human and social development in SA

    - Provide capacity to the rest of the scientific community (teaching, secondary use)

HSRC
Human Sciences
Research Council

The way forward

- Obtain executive custody, buy in and support from the Social Science community

- Resource mobilization

- Advocacy and raising awareness

- Build capacity

- Develop of infrastructure, policies & processes

- Pilot, use-cases

- Monitoring and evaluation

- Implementation

HSRC
Human Sciences
Research Council

Thank you

References

Research Libraries Group (2002). *Trusted Digital Repositories: Attributes and Responsibilities: An RLG-OCLC Report.* California, USA: RLG Inc.

Lord, P., Macdonald, A., Lyon, L., Giaretta, D. (no date). *From Data Deluge to Data Curation.* no place.: The Digital Archiving Consultancy Limited and the Digital Curation Centre.

OCLC and CRL (2007). *Trustworthy Repositories Audit & Certification: Criteria and Checklist.* Ohio, USA: CRL, The Center for Research Libraries.

Inter-university Consortium for Political and Social Research (ICPSR). (2009). *Guide to Social Science Data Preparation and Archiving: Best Practice Throughout the Data Life Cycle* (4th ed.). Ann Arbor, MI: Inter-university Consortium for Political and Social Research (ICPSR).