



Birth to Twenty Data Curation Project

2nd African Digital Curation Conference

12 - 14 May 2009

Alison Bullen & Lucia Lotter (HSRC)

Social science that makes a difference



BT20 Data Curation Project

- Birth to Twenty obtained funding from the Wellcome Trust to curate its data collection, starting with the data sets from the first wave in 1989/90.
- The HSRC was employed to assist in this process, setting up systems and developing best practice as this is an ongoing project.
- This paper will present an overview of the framework developed for this project, and then discuss some of the challenges in collecting metadata for a complex study that started 20 years ago.

BT20 Study

- Longitudinal birth cohort study
- Started tracking the development of 3,273 subjects/participants born between April and June 1990
- Studies child and adolescent health and development in Soweto
- Provides a unique source of information on public health issues in an urban setting of a developing country.

BT20 Activities

Web site

NEWSFLASH

Fast food consumption among 17 year olds

News releases

Infant, child and adolescent physical and mental health

- Norris SA, Roeser RW, Richter LM, Lewin N, Ginsburg C, Fleetwood SA, Taole E, van der Wolf K.(2007, in press) South African-ness Among young adolescents: The Emergence of a collective identity within the Birth to Twenty cohort study. *The Journal of Early Adolescence*; 28; 51-69
- Ramchandani PG, Richter LM, Stein A, Norris SA. (2007) Predictors of Postnatal depression in a urban South African cohort. *Journal of Affective Disorders*. June 19. (Epub).
- Grantham-McGregor S, Cheung YB, Cueto S, Glewwe

Publications & Editorial releases

Editorial Releases

- 2005
 - Birth To Twenty - Fatherhood
 - Birth To Twenty - Sibusiso Vilane makes history again
 - Birth To Twenty - Sibusiso Vilane

Postgraduate students

Tertiary education

1. Linda Vidulich, PhD, Department of Paediatrics, University of the Witwatersrand, *Factors that influence bone mass in a group of 10 year old urban black and white South African children.*
2. Jo- Anne McVeigh, PhD, Department of Paediatrics, University of the Witwatersrand, *The association of environmental and lifestyle factors with bone mass acquisition in South African children by sex, ethnicity and age.*
3. Barbara Willey, PhD, Loughborough University, UK, *Household socio-economic status and child growth.*

Heroes

Events

- Birth to Twenty 14th Birthday Celebration
- The World Health Sports Day

Birth to Twenty 14th Birthday Celebration

On the 8th May 2004, Birth To Twenty celebrated its 14th Birthday. In 1990, 3723 children were born in the Soweto-Johannesburg metropolitan area and were enrolled into the Birth to Ten to Twenty Study. The birthday bash was held to celebrate the children's birthday, and more especially to honour the BTT families for their continued participation in the study.

Approximately 200 of the families gathered at the Wits Education



Events

Newsletters

Heroes: Newsletters

Childline (<http://www.childline.org.za>) offers a 24 hour toll-free Helpline (dial 08000 55555) to assist abused children & their families.

Parents:

- Vol.2 No.1 newsletter
- Vol.1 No.2 newsletter
- Vol.1 No.1 newsletter

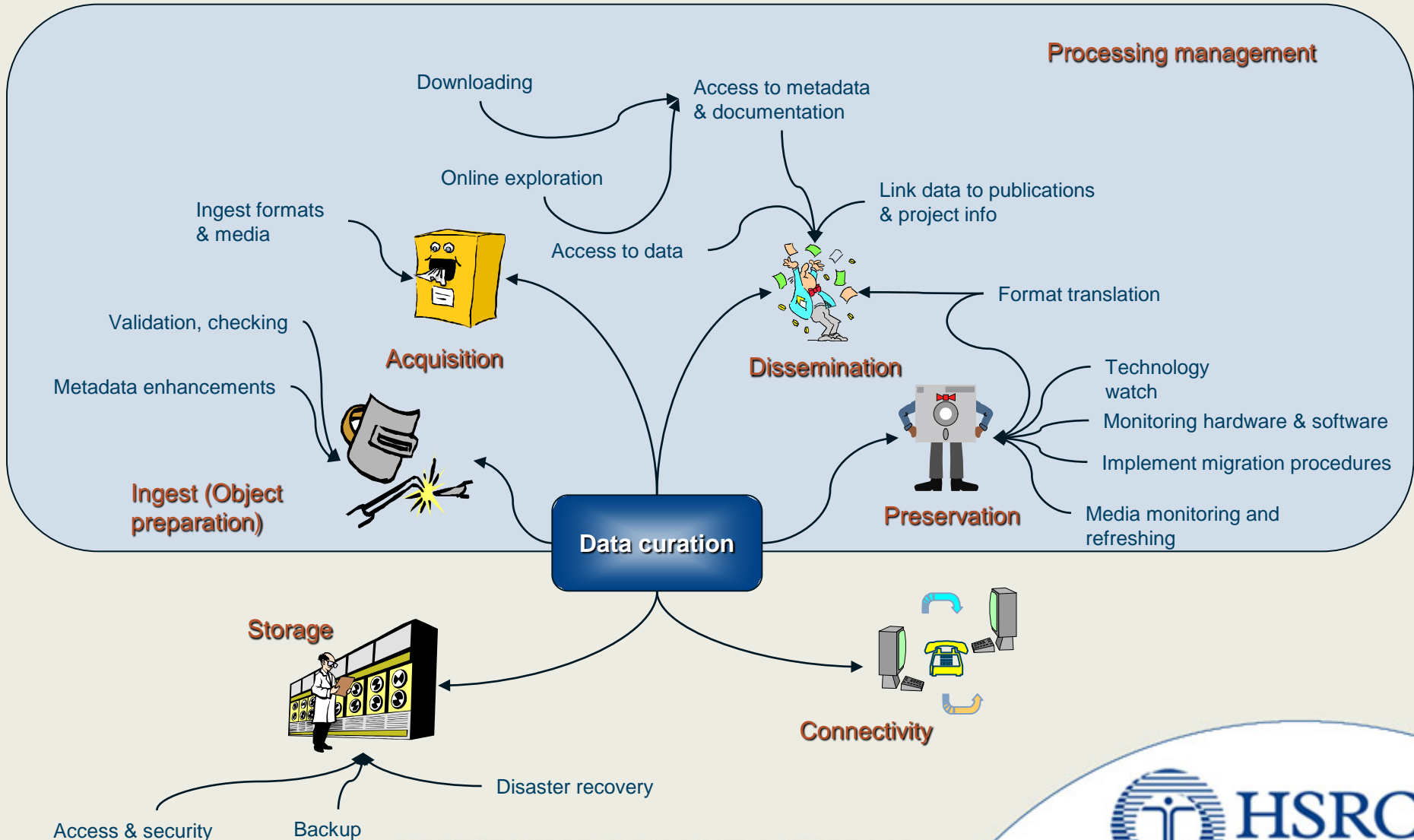
BT20 Data - Characteristics

- Almost 20 years of data collected – annual waves of collection with year 18 data collected end of 2008
- Various levels of cleaning done
- Mostly quantitative data with some qualitative
- Different themes covered over time
 - Infant, child and adolescent physical and mental health
 - Influence of home, school and family environment
 - Sexual and other risk behaviours during adolescence
 - Body composition, obesity, emerging non-communicable disease risk
 - Nutrition, bone health through childhood and adolescence
 - Methodological issues

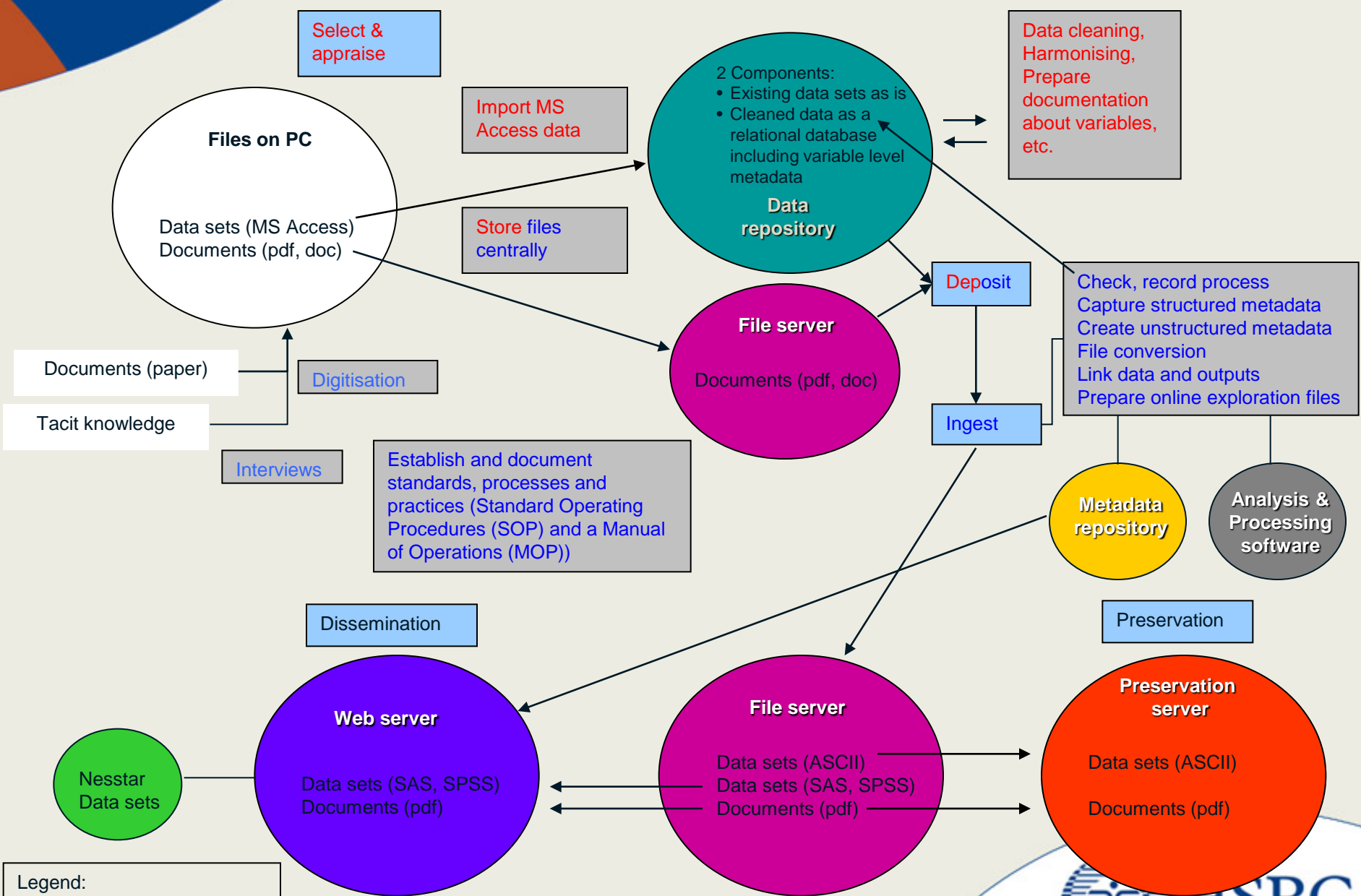
Data Curation Context

- Aim of the curation project
 - Create a repository for data & related documentation
 - Describe data & related documentation in terms of structured and un-structured metadata
 - Establish and document standards, processes and practices
 - Develop a capacity for data sharing
 - Preservation of data & documents
- Expected results
 - Better collaborative data sharing partnerships
 - Improved longitudinal data cleaning
 - Quicker analytical data set construction
 - Long term use of data & document files

Data Curation Activities

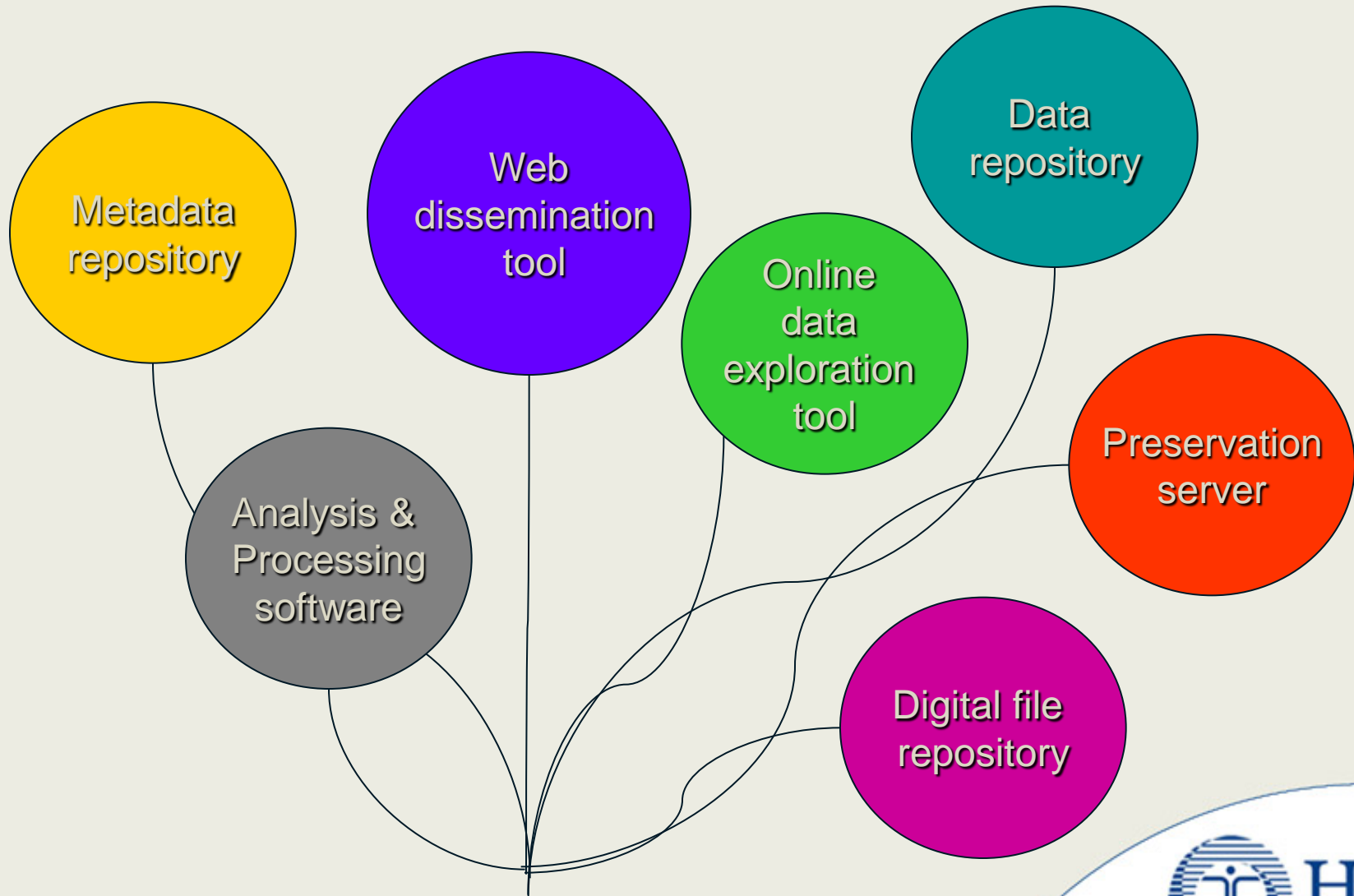


Social science that makes a difference



Legend:
 Data scientist (BTT)
 Information scientist (HSRC)

Technology – What is needed?



Social science that makes a difference

People and Processes

Researchers	Information scientist	Data scientist	IT / Systems Development
Appraise, select and prioritise data sets	Do data inventory	Design and develop database	Provide server & desktop infrastructure services:
Provide required context and meta data	Digitise printed documents	Check, clean & harmonise data	Hardware and software installation & maintenance
Attend to data curation throughout the research life cycle	Develop a document repository architecture & archive files	Anonomise data	Security & access management, backup and disaster recovery
Conduct data management according to best practices	Create metadata for data and documents & capture the historical context of the project (interviews)	Import data into database	Create systems for meta data, storage, dissemination, archiving management
	Do file conversions to dissemination and preservation formats	Create variable level meta for data sets	Preservation
	Produce a Standard Operating Procedures (SOP) and a Manual of Operations (MOP)		
	Train and develop capacity to maintain and continue the data archive system		
	Document and evaluate the data archival process		

Data Description

- Consider from perspective of end user
 - Researchers who are part of the research project
 - Collaborators
 - Scientific community at large

This will influence how much and what type of meta data you need to add.
- Data description includes
 - Provenance or context – in order to understand the data properly – both tacit and explicit from formal documents as well as informal communications
 - Metadata: Tools for retrieval – key wording and abstracting

Standards and Best Practice Manuals

- What standards do you use
 - For data – Data Documentation Initiative (DDI)
 - For data description – Dublin Core
 - Classification – HASSET
- Processes - Need for Data curation manual of operations and standard procedures
 - Workflow and responsibilities
 - Creation of data files best practice – data management
 - Minimum supporting documentation required
 - How to organise documentation (File plans; file naming conventions)
 - Metadata best practice
 - Format and conversion of documents
 - Privacy and confidentiality

Challenges

- Data description
 - Missing documents
 - Different versions of the same document
 - Tacit knowledge
 - Turnover of PI's
- Data
 - Data audit
 - Qualitative data
 - Missing data
 - Version control
 - Data sets not harmonised

Suggestions

- Data curation requirements need to be considered at all stages of the research process
 - Proposal phase
 - Project start up
 - Data collection
 - Data cleaning and analysis
 - Data sharing
- Ensure data management happens according to best practices
 - Make sure that master files are updated regularly and in a controlled manner
 - Keep good records of decisions and issues related to the data
 - Prepare metadata throughout the research life cycle
 - Deposit data and documents into a proper storage facility at strategic times, but before project closure
 - Use proper version control on data and document files
 - Use appropriate storage with adequate security and backup procedures

Social science that makes a difference

