Impact and Evaluation of Research Data Management

Stephen M Mutula
University of KwaZulu Natal
Context of Research Data Management

- Research, as we all know, is the heart beat of the academy. For this reason, management consultant, Arthur D. Little’s comment that “research makes building stones out of stumbling blocks” speaks to the value and importance of our pursuits.

- Global academic ranking of universities - visibility of research output; Institutional statistics; articles in citation indexes; contribution to knowledge, use of ICT; volume of research output.

- A report of the US President’s Council of Economic Advisors reveal that 50% of the growth in the American economy is attributed to investments in research and development (Atkinson 1997).
During the 1990s, California suffered the worst recession from which it was the last among American States to overcome. Today investment in university research has transformed California to become an economic powerhouse of high technology enterprises such as biotechnology and digital telecommunications because of research done at its Universities (Atkinson 1997).

Without reliable data to show how research drives economic growth and development, it is difficult to effectively convince governments to invest in R&D.
Walters in (Ray JM 2014: RDM: Practical Strategies for Information Professionals) uses the word **ecosystem** to describe the complex, dynamic technology environment needed to support data generated from intensive research process.

Walters ecosystem concept talks to critical issues affecting research data management such as:

- Sheer amount of data generated that has to be managed
- Complexity of the research process
- Heterogeneity of data and formats
- Varied technologies for storing and managing data
- Ethical aspects of research
1. Sheer amount of data generated and not used:

According Microsoft Technet (2012), the US uses 30 bn documents each year of which 85% are never retrieved; 50% are duplicated, 60% are obsolete and for US$1 a company spends to create a document, US$10 are used to manage document creation process.

The following issues become apparent:

- High cost of generating content
- Lack of access to the data
- Duplication of the data
- Obsolescence of data without being used
Problem 2
2. Onion Analogy of Research Process (Saunders 2006)
3. Complexity, heterogeneity and interconnectedness of data intensive research

- Data collected through national longitudinal/cross sectional surveys on population, food, crime, health, poverty; opinion poll data from market research organisations
- Varied theoretical frames used
- Different methodologies used
- Nature and types of research- basic, applied/action, collaborative, contract and sponsored
- Diversity of data formats
4. Ethical dimension of RDM

- Access & accessibility (WSIS AL2 & 3)
- Trust
- Privacy/confidentiality
- Integrity & accuracy
- Ownership and access rights
- Physical and network security
5. Variety of technologies in RDM

Trans-institutional and trans-border cyber-infrastructure:

- Mass digitization projects
- Collaboratories (NSF 2007).
- Cloud & grid computing technology
- Digital libraries
- Digital repositories
- Content management systems
- Federated search engines, etc
- Social networking sites
- E-journals
- Mass storage systems (Data warehousing, DISK)
6. Increasing demand for researchers

By 2020 projected that (Scholz, nd):

- Europe will have 700000 researchers
- China will have 2 million researchers

Increasing investments by governments (including South Africa) on S&T (DST 2003)
Problem 7: Most of the African political leadership does not seem to understand well the importance of research and development (Mozambican President Armando Guebuza in a summit of African leaders in Addis Ababa in January 2007). Largely all African countries investment in R&D is less than 1% of GDP.

Problem 8: Quality and relevance of research output

Government and industry are critical of quality of researchers coming from the universities (Ilvesmaki et al nd).
Sawyer (2004): “explicit public policy does not exist to reinforce and ensure research receive adequate investment from both private and public sectors for infrastructure development in the form of laboratories, equipment, libraries, and a system of information storage, retrieval, and utilization”
Institutional Performance Model (IPM): institutions that perform well are likely to elicit the confidence. Institutions that perform badly or ineffectively generate feelings of distrust and low confidence. The public recognizes whether institutions are performing well or poorly and reacts accordingly (Luhman 1979). IPM advocates for competence, responsiveness, interface design and customisation, interoperability, credibility, security, communication and availability.
Benchmarks for RDM Services

- Compliance with ethical standards – rights, dignity, health, safety and privacy, welfare of animals and the integrity of the environments
- Standardized and consistent procedures to collect, process, check, validate and verify data
- Self-explanatory nomenclature of data in terms of variable names, codes and abbreviations used
- Metadata that explain meaning of data, how they were collected and the methods used to create them
- Rights management
- Anonymisation
Research Data Management Responsibilities

- Researchers who create the data
- Principal investigators who design research
- Support staff managing and administering research
- Institutional IT services
- Data archives and centres
- Libraries
- Academic & central administrative units
- Open access services
- National Archives hold data as public records.
Challenges of RDM

- Distrust of some scholars and monopolistic practices
- Scanning errors in digitisation
- Providing quality metadata remains elusive
- Cyber-infrastructure challenges including device incompatibility and bandwidth
- High cost of access - The African consumer (or university) pays 50-500 times more than an American for an equivalent connection (e.g. $3,000/month instead of $30/month for a 1 Mbps connection (Mooketsi 2010)).
Recommendations & Actions to Improve RDM

- Research data management road map
  - RDM planning services
  - Active data infrastructure – facilities to store data actively used in current research
  - Data management advocacy
  - Data management capacity building
  - Data recovery and backup services

- Acceptable use policy

- Responsibilities for what part of data management

- Metadata management services - what research data exists, why, when, how it was generated, how it to access it