Multidisciplinary Data Infrastructure for South African Research: Developing a Workforce for a National Node

Dr Elisha Chiware
Cape Peninsula University of Technology
South Africa
National Integrated Cyber-Infrastructure System (NICIS)
NICIS Strategic Directions

White Paper

Open Science policy framework

National Big Data Strategy
Draft White Paper on Science, Technology & Innovation

- Long-term policy direction to ensure a growing role for STI in a more prosperous and inclusive society
- National System of innovation
- Increase human capital
- Provision of Sustained long term financing for Science, Technology and Innovation
Open Science Agenda

• Mandate Open Science Policies
• Balance Open Science and IP
• Curate Confidential Information
• Long Term Funding
• Training in RI Management
• Evaluation Metrics and training for data scientists
National Big Data Strategy

• Human capital deployment – the next generation

• Cyberinfrastructure Sustainability

• R&D Ecosystem – Government, Pvt Sector and Academia

• Data Governance: sharing, reuse, and ethical reuse of big data

• Coordination of investments and initiatives
National Research Ecosystem

- Department of Science and Technology, Republic of South Africa
- NICIS (National Integrated Cyberinfrastructure System)
- CSIR (Council for Scientific and Industrial Research)
- NRF (National Research Foundation)
- RISA (Research and Innovation Support and Advancement)
- SANRen (South African National Research Network)
- DIRISA (Digital Infrastructure Research and Innovation South Africa)
- TENET (Tertiary Education and Research Network of South Africa)
Data Intensive Research Initiative of South Africa (DIRISA)

Part of National Integrated Cyber Infrastructure Systems
DIRISA

- Build data infrastructures – across the country
- Advocate and coordinate activities
- Research Data Management – policies and guidelines
- Develop Human Capital and Skills
DIRISA

• Aims to:
  • Implement Tier 1 (national) repository for research data
  • Establish Tier 2 (regional) data nodes to support thematic data intensive research
  • Formulate national strategic frameworks for data intensive research & data stewardship – to promote data sharing & sound data management
Towards a national collaborative strategy for research data infrastructure
Ilifu Project Overview

Cloud Infrastructure

Science Projects:
- Astronomy
- Bioinformatics

RDM and Open Science Research Project
ILIFU: Research Data Management and Open Science Research Project Overview
Objectives of RDM & Open Science Project

• Software platform
• RDM services
• Training
RDM & Open Science Research Methodology
Research Data Management Working Groups

- Governance, data policy and infrastructure services
- Advocacy and training
- Standards, interoperability, certification and archiving
- Open Science, Open Access, Ethics, Legal and Authorisation for reuse
Implementation Plan

Jun 2018

**D1:** Standards compliant data management policies are publicly available to inform users of the requirements for data archiving, accessibility and the reuse of data created on the infrastructure.

Nov 2018

**D2:** An advocacy programme on the benefits of data sharing and open science is hosted.

Dec 2018

**D3:** A work integrated learning programme is developed for data managers with expected outcomes aligned to relevant academic courses.

Feb 2019

**D4:** Implement end-to-end standards compliant data management polices and services.

Jun 2019

**D5:** Deposit guidelines are developed and maintained online in response to feedback on user experience.
RDM Research Project Opportunities

- Capacity Building
- New multi-institutional collaboration on RDM
- Potential publications
- Community engagement
- Early researcher engagements
ILIFU: Work Integrated Learning Programme in Data Science
Work Integrated Learning Programme Objectives

- Investigate and recommend on a broad WIL programme
- Enable the placement of post graduate students in selected institutions hosting the ilifu astronomy and bioinformatics data intensive research projects
Work Integrated Learning Programme Objectives

• Provide WIL officers with placement opportunities in data science/Research Data Management

• Students gain experience & opportunities in RDM activities

• Placement sites benefit with added human capacity in meeting ilifu policy/guidelines’ requirements
Structure of the Work Integrated Learning Programme

- Existing academic programmes (source programme) related to research data management
- RDM Related Learning Areas
- Practical Considerations
- Monitoring and Evaluation Mechanisms
Existing academic programmes (source programmes) related to Research Data Management
### Source Programmes

<table>
<thead>
<tr>
<th>Institution</th>
<th>Department/Project</th>
<th>Course/Degree</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPUT</td>
<td>Computer Science</td>
<td>PG Dip ICT (Application development OR Communications Networks OR Multimedia Applications)</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Btech: IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M.Tech: IT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhD: ICT</td>
</tr>
<tr>
<td></td>
<td></td>
<td>PhD</td>
</tr>
<tr>
<td></td>
<td></td>
<td>BSc</td>
</tr>
</tbody>
</table>
RDM-related learning areas
Data lifecycle

These stages provide various learning opportunities in combination with the ilifu RDM guidelines on the use of the collaborative infrastructure.
# Learning Outcomes

<table>
<thead>
<tr>
<th>Institution</th>
<th>Learning opportunities</th>
<th>Learning outcomes</th>
</tr>
</thead>
<tbody>
<tr>
<td>South African National Bioinformatics Institute (SANBI), UWC</td>
<td>Implementing a platform for Tuberculosis surveillance in Africa (UWC):</td>
<td>1. Data generation</td>
</tr>
<tr>
<td></td>
<td>1. Methods for storing information about biological data</td>
<td>2. Database management and data curation</td>
</tr>
<tr>
<td></td>
<td>2. Provide a scalable database for a plethora of biological and clinical information</td>
<td>3. Data analysis</td>
</tr>
<tr>
<td></td>
<td>3. Provide a platform for rapid analysis of tuberculosis data and primary data generating sites</td>
<td>4. RDM policies and services</td>
</tr>
</tbody>
</table>
Practical Considerations

• Remuneration of students

• Duration of placements

• Supervision of students at placement sites

• Assessment and credit-bearing

• MOU on placements
Monitoring and Evaluation

- **Ilifu project**: No. of placements, where they are placed, from which university successes, barriers or any comments that came from the placement reports in an ilifu annual report.

- **Institution’s WIL coordinator / placement officer**: To provide his/her report as well as the student’s report about the placement.

- **Student**: Will provide their report as required by their university to the WIL coordinator or placement officer at their university.

- **Placement site (Ilifu partner sites)**: Feedback on the placement of students.
In Conclusion

In balancing personal career goals with ilifu aims, the WIL programme goes well beyond merely implementing guidelines for the use of ilifu infrastructure, but affords all stakeholders with a multitude of opportunities for collaboration, skills and human capital development, & the acceleration of science visibility through common platforms and practices.
Thank you

For further information, see:
http://www.ilifu.uct.ac.za
chiwaree@cput.ac.za