

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

Dr Elisha R T Chiware

Cape Peninsula University of Technology, Cape Town, South Africa, chiwaree@cput.ac.za

Theme/Focus Area: The Connected Researcher: National programs and Partnerships

Abstract:

ILIFU is the isiXhosa word for Cloud, an apt representation of the first node in the data infrastructure funded by the Department of Science and Technology (DST) in South Africa to support the National Integrated Cyberinfrastructure System of South Africa (NICS).

This presentation will provide an overview of the ILIFU model of shared infrastructure as a data-intensive research facility for big data management, storage and analysis. It will focus primarily on the research data management (RDM) research component of the project from conception to current operationalization and more especially on how opportunities are being created to develop a data science workforce through a collaborative Work Integrated Learning (WIL) programme across the science domains of the project.

The overall project has a strong emphasis on human capital development, to develop the workforce required to operate the regional data node. Engagement with researchers in designated domains of astronomy and bioinformatics enable the collaborative development of policies and guidelines to support users of the infrastructure. Significantly, the RDM research project fulfils an important function in support of institutional capacity building across participating institutions, where project findings will serve to promote institutional policy implementation and service development.

A working group on advocacy and training has been set up, to investigate and recommend a broad Work Integrated Learning (WIL) programme for postgraduate students of Computer- and Data Sciences, as well as in for Data Curation students in Library and Information Studies programme. The guidelines provide WIL officers in the participating universities with information to enable placement of postgraduate and mid-career professionals with various research projects to enable them to gain experience, opportunities and exposure on various aspects of data science and research data management. Through these placements, which will provide increased human capacity to meet deadlines and comply with RDM guidelines, the ilifu partners will remain up to date with current (RDM) practices, gain further industry exposure, and also action relevant curriculum developments.

In balancing personal career goals with ilifu aims, the Work Integrated Learning 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, and the acceleration of science visibility through common platforms and practices.

Presentation Outline

1. ILIFU project overview

Cloud Infrastructure

Science – Astronomy and Bioinformatics

RDM research project

2. Research Data Management project methodology

Structure

Implementation plan – currently in execution, deliverables

Communication plan – between work groups and external awareness raising

Researcher engagement

3. Developing the WorkforceWork Integrated Learning Programme

Capacity Building

Participation

Expected outcomes

4. Opportunities and Challenges

New multi-institutional collaboration

Management of the Work Integrated Programme

Opportunities for Students and Existing Data Science Work Force

Synergies among hosting and feeder institutions

5. Conclusion

Work in progress – assigning responsibilities, enabling growing areas of specialization

National benefits – Figshare brokering service model, data management guidelines for relevant domains