ILIFU Research Data Management and Open Science Guidelines
V 0.5 (June 2018)

Preamble

The ILIFU Research Data Management Project is part of an overall collaborative venture on the use of data intensive handling facilities by researchers at six partner institutions in the Western and Northern Cape (University of Cape Town, Stellenbosch University, University of the Western Cape and Cape Peninsula University of Technology, Sol Plaatje University). The goal is to provide researchers and scientists with advanced research data management facilities which enables data processing, storage, curation and storage for re-use and sharing. The project is in response to global trends on Open Science, Open Access and Open Data which are calling for visibility of research data and it's sharing to enable the acceleration in the advancement of science, innovation and development.

The project is funded through the Department of Science and Technology's DIRISA (Data Intensive Research Infrastructure South Africa) to build a national cyberinfrastructure for researchers and scientists. The ILIFU project is one of the first such regional data nodes under the DST initiative and will eventually link up with additional data nodes in the national cyberinfrastructure.

The Research Data Management and Open Science Project is responsible for the development of policies and guidelines on good data management and sharing practices. It is also responsible for a Work Integrated programme for the main project. These guidelines have been developed to provide support to researchers and scientist working/using the facility wishing to follow best practices and enable data visibility and it sharing. With the accompanying tools and services to be provisioned in the course of the project, such as the ILIFU Data Catalogue/s, these generic guidelines provide initial support to aid better data discoverability. Furthermore, ongoing discipline based approaches will provide access appropriate metadata schema for enabling visibility of data.

Table of Contents

Section A. Data Governance, Data policy and Infrastructure Service

1. **Intent**
   The intent is to develop common policies and guidelines to guide researchers working on the ILIFU infrastructure [see section A.3]

2. **Scope**
   The following guidelines have been developed to provide good research data management practices for researchers working on the ILIFU infrastructure including accessibility and reuse of data. It includes coverage on Data protection, End-user agreements; Conditions of use; Intellectual Property and License agreements.

3. **Infrastructure services**
   ILIFU infrastructure is a shared and collaborative project which provides processing of big data, and its storage that requires researchers working on the platform to comply with its terms and conditions of use including data policies and guidelines.

4. **Data protection**
   The ILIFU data intensive infrastructure will ensure the protection of researchers' data in regards to data processing, publication and sharing.
   Researchers are responsible for the protection of personal data by anonymizing personal identifiers.
   Data privacy is a shared responsibility which requires that users have a basic understanding of all collection and storage systems' interfaces as far as data sharing and isolation is concerned.

5. **End User agreements**
   The reuse of data and end user agreement will depend on the choice of Creative Commons license selected by the researcher.

6. **Conditions of use**
   Researchers working on the infrastructure will ensure the protection of research subjects (human and animal) and comply with national legislative frameworks (e.g. POPI) and institutional IP guidelines.

7. **Intellectual property**
   Researchers using the ILIFU facilities are governed by their institutional IP guidelines, as set out in Technology Transfer Office Guidelines, Patent contracts, etc.

8. **License agreements**
   The researcher is required to select the appropriate license agreement for access and reuse of their data. This selection should be in-line with ILIFU compliant License agreements.

9. **Data security**
   The ILIFU Infrastructure implementation follows recognized industry standards in securing data collection and storage systems by selecting tried and tested interface systems such as Figshare.

ILIFU undertakes to ensure safety and security of both physical infrastructure and data it stores in the cloud infrastructure provided.

10. Data collection

It is expected that researchers collect data in compliance with disciplinary ethical requirements before the data is processed, deposited and archived on the ILIFU infrastructure. [see also Data protection, T&C, Data collection]

11. Access to data

The ILIFU infrastructure serves the principles of Open Science. To protect the research process, all data published on the data management platform will be made publicly available as specified by the funding agency, not exceeding a period of two years. In the interim period, access to the data may be open (unrestricted), embargoed (available after a specified period of time), restricted (to a specified group) or closed (inaccessible to anyone other than the researcher). Any requests for access to restricted or closed data are to be routed to the researcher.

Section B. Standards, interoperability, certification and archiving

1. **Intent**
   The intent is to ensure researchers comply with national and international standards on data management, deposit, publishing and sharing. The ILIFU infrastructure will ensure interoperability with other external systems. The facilities will be compliant with international certification standards on data certification.

2. **Scope**
   The coverage of this section on standards, interoperability and certification relates to metadata standards, data assurance, submission, deposit, storage, security versions, and authenticity.

3. **Research Data Management planning (DMP)**
   All ILIFU infrastructure users are required to provide a data management plan using the template of the approved funder, or the generic ILIFU DMP. The DMP should accompany the written application to the ILIFU resource allocation committee.

4. **Metadata**

4.1 Requirements
   Sufficient metadata description is needed to aid in the discovery, accessibility and effective re-use of the datasets. The metadata must adequately describe the structure of the data and how it was created, as well as details of the owner and any legal restrictions. Data that support academic publications must be adequately described according to the metadata standards provided by the data management platform.

4.2 Terms of use
   Metadata is made publicly accessible on the ILIFU infrastructure. The metadata may be re-used in any medium without prior permission for not-for-profit provided the DOI or a link to the original metadata record are given**.**

5. **Data deposit**
   Data deposits are to be guided by a metadata guide that outlines the preferred metadata schema for a specified field or institution. Uploading of data is accompanied by a data deposit form provided by the data management platform.

   Uploaded data should be accompanied by an access license (e.g. Creative Commons license) detailing any restrictions on access. Data licenses may also outline any institutional, legal or regional and national restrictions including intellectual property and copyright restrictions.

   The ILIFU infrastructure will maintain appropriate data schema that will enable the interoperability with respective domains of data.

5.1 Data deposit, quality and copyright
   5.1.1 Items may only be deposited by accredited members of any of the ILIFU institutions, or their delegated agents.

5.1.2 The Research Data Repository (RDR) administrator only moderates deposits for the eligibility of authors/depositors, relevance to the scope of the repository, valid layout and format, and the exclusion of spam.

5.1.3 The validity and authenticity of the content of submissions is the sole responsibility of the depositor.

5.1.4 Items can be deposited at any time, but will not be made publicly visible until the depositors releases the data, e.g. when publisher- or funder-specified embargo periods have expired.

5.1.5 In making the deposit, the depositor confirms that they are the sole rights holder and/or they have received the right/permission from the rights holder(s) to publish the materials, i.e.:

 5.1.5.1 if the depositor is not the copyright holder and the material is copyrighted, the depositor acknowledges that they have received permission from the rights holder to publish the copyright material openly and have adhered to the conditions under which the copyright holder has granted publication rights. e.g. acknowledged the original creator of the work and/or licensed the work under specified conditions

 5.1.5.2 where the depositor is not the sole IP creator, that written permission approving the deposit has been received from all the IP creators, and in the case of video material that the necessary image releases have been obtained from people appearing in the videos

5.1.6 ILIFU may take down any publication without notice if any of the items contained within the publication are in breach of any of the above, or:

 5.1.6.1 if the data includes work that is found to be fraudulent, falsified, plagiarised or against the FAIR principles of Open Science

 5.1.6.2 the depositor has requested to have the data removed from the repository provided that the request is accompanied by a reasonable motivation approved by the Research Data Management Governance Committee in terms of its mandate to approve exceptions to the public release of data for legal, ethical and commercial reasons

 5.1.6.3 if ILIFU or the depositor choose to remove the data removed from the repository, the metadata may be retained in the repository in order to trace the original DOI and indicate that the removal was deliberate. Furthermore, ILIFU may retain a copy of the data for archival and reference purposes

6. **Data assurance**

No files containing personal data will be accepted.

Data curators in the respective partner institutes will moderate deposited data to confirm usability and accuracy. Records corrected or amended in consultation with the depositor will be allocated a new version number. Changes or updates due to data correction will be recorded but previous versions will not be deleted.

File versions will be recorded in the metadata record up loaded with data files and will be reflected in the metadata record.

7. **Storage**

The Resource Allocation Committee will allocate storage as per research requests and requirements.

8. **Version control and authenticity**

   The ILIFU facility makes provision for versioning, control and authenticity of data being processed, stored and archived.

9. **Archiving**

   Users of the ILIFU infrastructure are advised to make adequate provision in grant proposals for alternative storage to ensure the long-terms archiving of data considered to be of persistent value.

Section C. Open Science, Open Access, Ethics, Legal Framework and Reuse

1. **Intent**
   The intent is to provide guidelines on the overall national open science framework requirements, open access, ethics, legal requirements for data storage and reuse.

2. **Scope**
   The coverage of this section of the guideline will relate to data discoverability, data citation, reuse, ethics, end user agreements and access to the data.

3. **Data discoverability**
   The facility will be compliant with international standards to enable discoverability eg.DOI

4. **Data citation**
   The facility encourages that data generated be discoverable and citable by means of the DOI provided by the data management platform.

5. **Reuse**
   Conditions of reuse are determined by the selection of a Creative Commons license. [see also section A.4]

6. **Ethics**
   Researchers working on the ILIFU infrastructure are subject to various institutional ethics requirements, to the requirements of any funding agencies and to national legislative frameworks, e.g. PoPI.

7. **End user agreement**
   [see also section A.5]

8. **Open science and Open access**
   [see also section A.11]

Section D: Glossary of terms

Dataset- a set of files containing both research data, and documentation.

DMP- Data management plan is a document which describes how data will be handled throughout the research project and after the completion of the project.

Documentation- any digital files such as a codebook, technical report, methodology, workflows which explain the research data's production or use.

Metadata- enables and enhances the discovery and reuse of data. Metadata is a description of the data in order to help people to obtain and use the data. The description can be background information on the data, as well as information on how the data can be used. There are standards for describing whole datasets and there are also standards for describing entities within a dataset. To ensure that you understand your own data and that others may find, use and properly cite your data, it helps to add metadata to the documents and data sets you create.

Research data- are defined as factual records (numerical scores, textual records, images and sounds) used as primary sources for scientific research, and that are commonly accepted in the scientific community as necessary to validate research findings. A research data set constitutes a systematic, partial representation of the subject being investigated.

Research data management (RDM)- Research data management concerns the organisation of data, from its entry to the research cycle through to the dissemination and archiving of valuable results. It aims to ensure reliable verification of results, and permits new and innovative research built on existing information.