

DATA MANAGEMENT PLAN

0. Proposal name
1. Description of the data
<p>1.1 Type of study Molecular, biochemical and imaging techniques will be used to determine the protein kinase pathways that regulate cytokinesis in <i>T. brucei</i>. Kinase assays will be developed, and kinase inhibitors (from the GSK HAT Box) sought, for future exploitation in drug discovery.</p> <p>1.2 Types of data Qualitative (e.g. phenotyping data for cell lines), quantitative (e.g. cell counts), mass spectrometry and image data will be generated. Raw data will be analysed and expressed as graphs, tables and annotated images, some of which, it is expected, will be published.</p> <p>1.3 Format and scale of the data Data generated will be in various formats and sizes of datasets, all of which will be accessible using common software allowing easy access and long term validity during and after the project, thus facilitating data sharing. The format/types of data include:</p> <ul style="list-style-type: none"> i) Cell images e.g. phase and fluorescence, and electron micrographs (~5,000 images over project). Software used includes OpenLab, Softworx and IN Cell Investigator, with data saved as software-specific files e.g. liff and lg3 files, as well as generic formats such as jpeg, tiff etc. ii) mass spectrometry spectra (from <50 samples). MS data will be analysed using Bruker Data Analysis or Thermo Excalibur software (generating xml and raw files) and proteins will be matched to the <i>T. brucei</i> genome dataset using the Matrix Science Mascot search engine. Each LC-MS data file is between 1-2GB. iii) Cell line phenotyping data including growth curves and DAPI counts (Excel and GraphPadPrism files) and flow cytometry data (FlowJo and jpeg/tiff files) (~200 data sets). iv) HAT Box screening data and EC50 curves which will be saved as Excel spreadsheets (~20 data sets) v) enzyme assay data which will be recorded as Excel spreadsheets or jpeg/tiff scans of gels etc (<50 data sets) vi) DNA and protein sequencing data, which will be saved as abi or MASCOT files (~200 data sets). Plasmid maps will be generated and oligonucleotide sequences stored using software such as Vector NTi or CLC.
2. Data collection / generation
<p>2.1 Methodologies for data collection / generation Data will be generated by the experiments described in the Case for Support. Some data will be collected manually by project staff e.g. DAPI counts while others will be collected automatically by the equipment used e.g. flow cytometry or mass spectrometry profiles. Some data collected automatically e.g. In Cell High Content images will then be analysed manually e.g. to analyse the stage of cytokinesis in 2N2K cells. Data files generated will be labelled appropriately and placed in suitably labelled/organised folders and sub-folders, so that it is obvious which cell line, sample, experiment etc the data originated from, and to allow the project team to easily find files. Different versions of a data subset will be distinguished via a subscript (v1, v2, final etc) attached to the file name.</p> <p>2.2 Data quality and standards Standard protocols will be optimised and used to collect data to ensure they are reliable and consistent. All experiments will incorporate appropriate positive and negative controls to ensure validity; biological and technical replicates will be used to assess consistency of data. Project staff will be adequately trained in the techniques they use to ensure they generate high quality data. Data generated and the methods used will be scrutinised in weekly lab meetings to ensure procedures have been carried out correctly, that appropriate controls have been used, that all information is suitably recorded and that therefore there can be a high level of confidence in the data generated. These are routine procedures for conducting high quality research, which allow it to be judged and published in fully peer-reviewed journals, as well as discussed at relevant international conferences.</p>
3. Data management, documentation and curation
<p>3.1 Managing, storing and curating data.</p>

All laboratory experimental details and data will be recorded daily and dated in dedicated laboratory notebooks and scrutinised at weekly lab meetings. Electronic records of the recorded data will be generated (e.g. by scanning or photographing gels/blots, by plotting cell counts and DAPI data in Excel spreadsheets etc) and saved to the University server (which is automatically backed up daily) and also, where appropriate, on the [PI's] lab website. File names/locations will be recorded in lab notebooks to allow electronic records to be linked to the raw data; likewise, file names will have an appropriately descriptive title, including the date the data were generated to allow the corresponding raw data records to be easily found. Details of all oligonucleotides, plasmid constructs and parasite cell lines etc used/generated during this project will be recorded in existing electronic searchable databases located on the University server. Additionally, regular electronic progress reports summarising key data will be compiled by project staff.

3.2 Metadata standards and data documentation

Standard operating procedures will be saved on the University server. Details of SOPs used and any deviations from them will be recorded in laboratory notebooks alongside records of the data generated. Project data will be published, with suitable annotation and experimental details, in open access journals and/or on the [PI's] lab website to allow it to be easily accessed and used by researchers in the field. Phenotypic information obtained for any protein studied will be released to the field by open access publication and/or by submitting appropriately annotated data to TriTrypDB (www.tritrypdb.org) where they can be easily accessed by the research community. Large research data sets e.g. proteomic data and High Content images will be deposited with suitable annotation in the University of Glasgow Data Registry (Enlighten – Research Data; DataCite standard).

3.3 Data preservation strategy and standards

Electronic data generated and written records will be retained for at least 10 years after the project ends. Open access publication of data and deposition in the Enlighten repository will ensure longevity of the data in the long-term.

4. Data security and confidentiality of potentially disclosive information

4.1 Formal information/data security standards

No human participants will be used in this research. The University of Glasgow is ISO compliant (certification no. ISO27001).

4.2 Main risks to data security

The main risks to data security are loss or damage to laboratory notebooks and loss or corruption of electronic data. Data will be safeguarded by the following measures:

- a) Data in lab notebooks will, as described above, also be recorded in electronic form that is backed up daily to secure against loss or damage of the notebook.
- b) Access to electronic data (prior to publication as described above) will be limited to the members of the research group and relevant collaborators via limiting access to shared drives on the University server.
- c) Access to laboratories and offices are controlled by card access to reduce the likelihood of malicious loss/damage; all computers used in this project will run Standard Staff Desktop, whereby firewalls and antivirus software are automatically upgraded and secure remote access to data is enabled; staff will lock their workstation whenever they are away from it.

5. Data sharing and access

5.1 Suitability for sharing

Yes – data generated from screening HAT Box compounds for effects on cytokinesis, and proteomic data sets are likely to be of interest to the trypanosome community.

5.2 Discovery by potential users of the research data

Project data (with accompanying metadata) will be made available to all interested researchers via open access publication, the Enlighten: Research Data Repository or via the [PI's] lab website. Data in Enlighten: Research Data will be issued with a Digital Object Identifier (DOI). This can be included as part of a data citation in publications, allowing the datasets underpinning a publication to be identified and accessed. DOIs will also be linked with appropriate records in Enlighten: Publications, the University's publication repository, to enhance visibility of datasets.

Metadata about datasets held in the University Registry will be publically searchable and discoverable and will indicate how and on what terms the dataset can be accessed.

Information about datasets from the Registry will be displayed on the PI's University of

Glasgow webpage which will also increase the visibility of datasets. The team's approach to data sharing will be outlined in publications and on the lab website.

5.3 Governance of access
The PI will make the decision on when to submit data to Enlighten: Research Data (publically available; see above) and whether to supply research data to a new user.

5.4 The study team's exclusive use of the data
Data will be made available at the time of publication, at the latest. Depending on the nature of the data itself, data may be made available earlier, either on an individual basis to interested researchers and/or potential new collaborators, or (for e.g. negative data) publically via deposition in e.g. Enlighten: Research Data.

5.5 Restrictions or delays to sharing, with planned actions to limit such restrictions
Restrictions to data sharing (e.g. with competitors) will be in place where necessary to ensure novelty for publication or where appropriate, to protect IP but otherwise data will be shared as widely as possible.

5.6 Regulation of responsibilities of users
Where data or resources are provided to an external user ahead of their publication, it will be stipulated that the external user, prior to publishing any work using the data/resources, must consult the PI to determine whether it would be justified for the PI and project team members to be included as authors on that publication.

6. Responsibilities
Along with the PI, members of the project team will have responsibility for study-wide data management, metadata creation, data security and quality assurance of data. Research colleagues within the department will assist with quality assurance by criticising data presented at joint (confidential) lab meetings. The University Research Data Management team will be able to advise on best practice in data management and security.

7. Relevant institutional, departmental or study policies on data sharing and data security

Policy	URL or Reference
Data Management Policy & Procedures	Draft UoG Research Data Management policy http://www.gla.ac.uk/media/media_232425_en.docx
Data Security Policy	IT Security Policies http://www.gla.ac.uk/services/it/informationsecurity/policies Confidential Data Policy http://www.gla.ac.uk/services/it/informationsecurity/confidentialdata Data security issues are also covered to some extent by the UoG's Data Protection policy http://www.gla.ac.uk/services/dpfoioffice/policiesandprocedures/dpa-policy Policy on Confidential Data in the UoG http://www.gla.ac.uk/media/media_180727_en.pdf
Data Sharing Policy	Data sharing is covered in the UoG's Code of Good Practice in Research http://www.gla.ac.uk/media/media_227599_en.pdf Data sharing is covered in a number of Ethics related policies. The policies are accessible via the Ethics Committee homepage http://www.gla.ac.uk/research/aims/ourpolicies/ethics
Institutional Information Policy	The UoG lists all policies relevant to research conduct and interaction with the wider community http://www.gla.ac.uk/research/aims/ourpolicies
Other:	Details on relevant RCUK policies and draft UoG EPSRC roadmap available at http://www.gla.ac.uk/services/datamanagement/rdm-atgu

8. Author of this Data Management Plan (Name) and, if different to that of the Principal Investigator, their telephone & email contact details