RESEARCH DATA MANAGEMENT PLAN

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| PROJECT | | | | | | | | | |
| Title | Diptera geometric morphometric analysis | | | | | | | | |
| Description | 1. Variation among species in morphology is frequently complex (various attributes of shape, rather than size or simple linear measurements).  2. The current analyses I use to handle this variation rely on comparing the relative position of "landmarks" which can be placed on homologous features in images of a series of specimens.  3. Multivariate analyses can tease apart the dominant forms of shape variation among a group of samples based on the set of x,y-coordinates of these landmarks and you can then ask questions like: does the shape of the wing vary between these two species?, and you can get a quantified answer, rather than "they look different".  4. All of the software exists for doing all of these steps, either in R or through a windows interface  The next step is to take a set of existing data on shape variation in a nicely quantifiable structure such as an insect wing, and figure out a way to quickly see where a new specimen falls in respect to known variation, therefore classifying it as belonging to one species or another.  So you might have an ordination that shows how 100 specimens cluster into 10 recognisable species. You collect a new specimen, take an image, place landmarks on the correct features of that image and hope to discover how your new specimen fits with the 10 species. Does it fall nicely within the morphological range of an existing species, or does it fall out somewhere new in the ordination, and perhaps represent a new species. | | | | | | | | |
| Field of Research | **060899; Morphometrics** | | | | | | | | |
| DMP created | | | Last updated | | Project start | | | Project end | |
| 20160728, 1244 | | | 20160820 | | 20160721 | | | [dd/mm/year/Ongoing] | |
| PROJECT CONTRIBUTORS | | | | | | | | | |
| Role | | Name | | Affiliation | | Email | Username | | ORCiD ([*i*](http://www.library.auckland.ac.nz/services/research-support/orcid)) |
| PI/ Lead researcher | | Neil Birrell | | School of Biological Sciences, Faculty of Science | | nbir012@aucklanduni.ac.nz | nbir012 | | <http://orcid.org/0000-0002-7961-1626> |
| Data contact | | Neil Birrell | | School of Biological Sciences, Faculty of Science | | nbir012@aucklanduni.ac.nz | nbir012 | | <http://orcid.org/0000-0002-7961-1626> |
| [Supervisor, etc.] | | Gregory Holwell | | School of Biological Sciences, Faculty of Science | | g.holwell@auckland.ac.nz | g.holwell | | Unknown |

|  |  |
| --- | --- |
| POLICIES & GUIDANCE | |
| Related policies | [IT Institutional Data Management Policy](https://www.auckland.ac.nz/en/about/the-university/how-university-works/policy-and-administration/computing/data-management/data-management.html) (particulary points 1 through 5) [Research Code of Conduct Policy](https://www.auckland.ac.nz/en/about/the-university/how-university-works/policy-and-administration/research/conduct/code-of-conduct-policy.html) (in particular 5.4 Retention of Data and Records) |
| FUNDING (if applicable) | |
| Funding agency | Vice Chancellor Strategic Development Grant, University of Auckland (2015) |
| Funding ID | (require this from Fabiana) |
| Research Office ID | (require this from Fabiana) |
| ETHICS & PRIVACY | |
| Ethics requirements | No, I have discussed with supervisor and, as we are simply photographing historical insect specimens, we will not require ethical approval. |
| How will you manage any ethical issues? | N/A |
| Are there other privacy and/or security requirements? | No, there are no privacy/security requirements. The data does not contain any confidential or sensitive information. |
| DATA ORGANISATION | |
| **Data collection/ creation** | |
| What data will you create/ collect? | A metadata file that records the taxonomic information of the specimen, the file name of the photograph that is assigned to the specimen. The date, time and location of the data being recorded.  There will be photographs of the wings of insect specimens with landmarks placed on points of the wing from which geometric morphometrics will be run  The programs used to get the data are ‘tpsUtil64’, ‘tpsDig232’ and ‘tpsRelw32’. These are freely available software created by F. James Rolf and can be run on Windows machines or a windows virtual environment. The total data collected is expected to be under 1GB of data. The output from the tpsRelw32 will be converted from a txt file into a csv file which will allow us to run a statistical analysis on the data via Python. |
| How will the data be collected/ created? | Photographs will be taken using a macro lens DSLR camera. The picture will be taken by holding the camera so that the lens is parallel to the face of the wing. This minimises any risk of distortion which might skew results of the analysis. The metadata will be stored in a spreadsheet and will consist of the tray number, species name and unique image ID of each photograph (e.g. DSC\_7851).  There are no formal community data standards for the morphometric analyses performed. Repeat samples may be able to be collected at a later stage to assist with quality assurance of the data however this will be assessed based on time allowances and workload. |
| What non-digital data/assets will you create/ collect? | There will not be any non-digital data. All assets will be composed in a digital format. |
| **File management** | |
| How will the data be organised? | The data will be organised in a github account online based on a folder structure of:  1.0 Project Folder  1.1 Code  1.2 Data  1.3 Documents |
| **Storage locations** | |
| How will the data be stored and backed up during the research? | Three copies of the data will be stored on a laptop hard drive, the University of Auckland (UoA) server andin cloud storage, namely google drive. Total data is expected to be under 1GB for which all locations have sufficient capacity. In the event of an incident it is expected that the data will be retrieved from one of the redundant storages e.g. the hard drive, UoA server or Google Drive. The university servers are backed up **<require more information on frequency of backup – unable to find in institutional data management policy>**  There is no sensitive data which are subject to legislative/jurisdiction requirements. The data will be stored in Neil Birrell’s University of Auckland google drive. |
| METADATA & DOCUMENTATION | |
| What documentation and metadata will accompany the data to support its discovery, (re)use and increase impact? | A spreadsheet has been created that captures metadata for each photo. A read me document will be created outlining the data capture process and methodology. This will be stored in all three locations where the data are stored. |
| Spatial extent | Not applicable |
| Temporal extent | Not applicable |
| Links | To be provided by end of Medsci 736 |
| OWNERSHIP, COPYRIGHT & IP | |
| **The copyright and other IP is owned/held by:** | Yes or leave blank |
| The University of Auckland (normal situation for research undertaken by university staff) |  |
| The student (research by research student in the normal course of study, which does not fall into any of the other categories.) | Yes |
| Joint ownership (research conducted in collaboration: copyright and IP ownership are documented in an agreement between the organisations) |  |
| Third party data (data owned by third party or generated under UniServices agreements. |  |
| If ownership *is* jointly held, third party or generated under UniServices contract. | [State the relationships, agreements and relative rights to use, store, publish and re-use the data.]  N/A |
| ACTIVE DATA - SHARING & ACCESS CONTROL | |
| Access to the data during the project will be: | Unrestricted |
| How will you manage access and security? | The data will be shared internally via university server. The use of Google Drive and figshare will allow for public sharing, however, for Google Drive this will be managed by providing access to the data when requested and unable to edit the original data but can create copies to work on as the person wishes.  The data is not confidential so security other than read and copy only access is not required. |
| RETENTION & DISPOSAL | |
| **Data must be retained after submission of thesis or publication of results for a minimum of:** | (select) |
| 6 years (standard minimum retention after last publication based on data) | Yes |
| 10 years (for medical research involving clinical trials from the end of the trial) | N/A |
| Until patient reaches 26 years of age, and at least 10 after last treatment (for clinical research involving children) | N/A |
| 21 years from the date of filing a patent related to this research | N/A |
| Other specified time | N/A |
| Details of other time | N/A |
| Based on the above, data must be kept until at least | [dd/mm/year] (will be determined when published. |
| DATA PUBLISHING AND DISCOVERY | |
| Licencing | MIT License or The Unlicense |
| Outline how data will be prepared and where it will be published. | Data will be prepared using Python and Juypter Notebook and will be shared using figshare |
| LONG-TERM ARCHIVE / PRESERVATION (20+years, if applicable) | |
| What is the long-term preservation plan for the dataset? | I don’t believe that the data will offer something irreplaceable as the specimens they were taken from are available in permanent storage at Landcare Research, East Tamaki, New Zealand. However, this will be further checked with the library to ensure this assumption is reasonable. Further information is required to determine if .tps is likely to become obsolete and if so whether it can be migrated to a new format easily. |

|  |  |
| --- | --- |
| RDM/DMP RESPONSIBILITIES & RESOURCES | |
| Who will be responsible for data management? | Neil Birrell is responsible for implementing the data management plan and each activity therein including, but not exclusive to, data capture, metadata production, data quality, storage and backup, data archiving & data sharing. |
| What resources will you require to deliver your plan? | I will require a laptop (provided by UoA). Further training on Github repositories, Python and licencing is required (provided by Medsci 736). Some assistance on the correct statistical analysis will be sought, most likely from Greg Holwell. Software from the University of Manchester is required and this is available for free. There are no known charges for data repositories. |

References and thanks to:

DCC. (2013). Checklist for a Data Management Plan. v.4.0. Edinburgh: Digital Curation Centre. Available online: http://www.dcc.ac.uk/resources/data-management-plans