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 | [Be aware of the University [Researcher Code of Conduct](https://www.auckland.ac.nz/en/about/the-university/how-university-works/policy-and-administration/research/conduct/code-of-conduct-policy.html), specifically “4.5 Research Findings: Researchers should share data and findings openly and as promptly as possible, as soon as they have had an opportunity to establish priority and ownership claims and subject to any intellectual property requirements and contractual obligations.” as you produce your DMP.  List any other relevant funder, institutional (University [Policy Hub](https://www.auckland.ac.nz/en/about/the-university/how-university-works/policy-and-administration.html), [Research](https://www.auckland.ac.nz/en/about/research.html) [Open Access](http://www.library.auckland.ac.nz/guides/open-access)), departmental or group policies on data management, data sharing and data security. Some of the information you give in the remainder of the DMP will be determined by the content of other policies. If so, point/link to them here.]  Still researching this | |
| 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? | [Consider: What type, format and volume of data? Do your chosen formats and software enable sharing and long-term access to the data? - Are there any existing data that you can reuse?  Give a brief description of the data, including any existing data or third-party sources that will be used, in each case noting its content, type and coverage. Outline and justify your choice of format and consider the implications of data format and data volumes in terms of storage, backup and access.]  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  Geometric morphometric analysis will create a .tps file. It is currently being determined whether this can be converted into a .csv format which may be more suitable for ensuring long term accessibility by anyone in the future. The programs used to get the data are ‘tpsUtil64’, ‘tpsDig232’ and ‘tpsRelw32’. These are open source software created by the University of Manchester and can be run on Windows machines. The total data collected is expected to be under 1GB of data. | |
| How will the data be collected/ created? | [Consider: What standards or methodologies will you use? What quality assurance processes will you adopt?  Outline how the data will be collected/created and which community data standards (if any) will be used. This may include processes such as calibration, repeat samples or measurements, standardised data capture or recording, data entry validation, peer review of data or representation with controlled vocabularies.]  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? | [Consider: Digitisation or recording (e.g. photographing, transcribing) of any data or materials originally created/ collected in non-digital forms (e.g. original surveys, interviews). Where will non-digital data assets be stored?]  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? | [Consider how the data will be organised during the project, mentioning for example file and folder naming conventions, version control, folder structures, use of database – schema, tables and relationships.] | |
| **Storage locations** | | |
| How will the data be stored and backed up during the research? | [Consider: Do you have sufficient storage or will you need to request additional services? Where do you intend to store your data? How will the data be backed up? Who will be responsible for backup and recovery? How will the data be recovered in the event of an incident?  Describe the primary location of your data. State how often the data will be backed up and to which locations. How many copies are being made? Storing data on laptops, computer hard drives or external storage devices alone is very risky. The use of robust, managed storage provided by the University is preferable. Similarly, it is normally better to use automatic backup services provided by ITS Services than rely on manual processes. If you choose to use a third-party service, you should ensure that this does not conflict with any funder, institutional, departmental or group policies, for example in terms of the legal jurisdiction in which data are held or the protection of sensitive data.]  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. | |
| METADATA & DOCUMENTATION | | |
| What documentation and metadata will accompany the data to support its discovery, (re)use and increase impact? | [Consider: What information is needed for the data to be to be read and interpreted in the future? How will you capture / create this documentation and metadata? What metadata standards will you use and why?  Describe the types of documentation that will accompany the data to help secondary users to understand and reuse it. This should at least include basic details that will help people to find the data, including who created or contributed to the data, its title, date of creation and under what conditions it can be accessed.  Documentation may also include details on the methodology used, analytical and procedural information, definitions of variables, vocabularies, units of measurement, any assumptions made, and the format and file type of the data. Consider how you will capture this information and where it will be recorded. Wherever possible you should identify and use existing community standards. ]  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 | [If applicable, state the spatial or geographic extent of your data with places names AND coordinates (N: E: S: W:) to enable bounding box / spatial searching.]  Not applicable | |
| Temporal extent | [If applicable, state the period(s) of time over which your data is associated.]  Not applicable | |
| Links | [Links to files (e.g. DMP, README.txt, geospatial (.kml), etc.) to be published as a metadata package to accompany data publication and facilitate data use.]  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 / Restricted/ Embargoed – delete as necessary.]  [If embargoed, provide details and date for release.]  Unrestricted | |
| How will you manage access and security? | [Consider: Whether you are the only person that will have access to the active, unpublished data, or data will be shared internally (specify with who) /externally (specify with who)? What are the risks to data security and how will these be managed? How will you control access to keep the data secure? How will you ensure that collaborators can access your data securely? If creating or collecting data in the field how will you ensure its safe transfer into your main secured systems?  If your data is confidential (e.g. personal data not already in the public domain, confidential information or trade secrets), you should outline any appropriate security measures and note any formal standards that you will comply with e.g. ISO 27001.]  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 | [State licence(s) under which you plan to make the data publically accessible. Be as open as possible. [Creative Commons Aotearoa New Zealand – Licences explained](http://creativecommons.org.nz/licences/licences-explained/) ]  This is still to be decided following completion of the licencing seminar | |
| Outline how data will be prepared and where it will be published. | [Consider: How will potential users find out about your data? With whom will you share the data, and under what conditions? Will you share data via the [University Data Publishing and Discovery Service](https://www.library.auckland.ac.nz/databases/record/index.asp?record=DatPubandDisSer) or a disciplinary data repository ([listing](http://service.re3data.org/search))? Will you publish a metadata only file with the institutional repository? Will you handle requests directly or use another mechanism? When will you make the data available? Will you pursue getting a persistent identifier for your data?  The methods used to share data will be dependent on a number of factors such as the type, size, complexity and sensitivity of data. If possible, mention earlier examples to show a track record of effective data sharing.]  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? | [Publishing with DOI should offer persistent access over years, not necessarily decades. Nor does a DOI actively address any needs to transform or migrate data from obsolete formats over time.  Consider: Does your data offer something irreplaceable (actually or practically) that is of value to society over decades to come? If unsure, please contact the Library.  Provide information or guidance on how datasets that have long-term value will be preserved and curated beyond the lifetime of the project. Indicate likely areas of risk (i.e. proprietary formats). If you do not propose to use an established repository, the data management plan should demonstrate that resources and systems will be in place to enable the data to be curated effectively beyond the lifetime of the project. Document any discussion of long–term archive with academic units or other relevant staff.]  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? | [If your project has several contributors and/or institutions, and/or the ownership is complex, consider: Who is responsible for implementing the DMP, and ensuring it is reviewed and revised? Who will be responsible for each data management activity? How will responsibilities be split across partner sites in collaborative research projects? Will data ownership and responsibilities for research data management be part of any consortium agreement or contract agreed between partners?  Outline the roles and responsibilities for all activities e.g. data capture, metadata production, data quality, storage and backup, data archiving & data sharing. Consider who will be responsible for ensuring relevant policies are respected. Individuals should be named where possible.]  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? | [Consider: Contacting the Centre for eResearch. Is additional specialist expertise (or training for existing staff) required? Do you require hardware or software which is additional or exceptional to existing institutional provision? Will charges be applied by data repositories?  Carefully consider any resources needed to deliver the plan, e.g. software, hardware, technical expertise, etc. Where dedicated resources are needed, these should be outlined and justified.]  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