Data Management Plan

Project:	**Diversity and key traits of moths (Geometridae, Arctiinae) along a large elevational gradient in a biodiversity hotspot (SE Peru)**
Principal Investigator/s:	Gunnar Brehm, info@gunnarbrehm.de, gunnar.brehm@uni-jena.de
Project Data Contact:	Gunnar Brehm, info@gunnarbrehm.de, gunnar.brehm@uni-jena.de
Funding application:	DFG Research Grants programme

Relevant Policies and Guidelines:

Research data management will follow the DFG guidelines on the handling of research data, the DFG guidelines on the handling of research data in biodiversity research and comply with general institutional guidelines of the Friedrich Schiller University Jena (Policy on the handling of research data at the Friedrich Schiller University Jena, Guidelines and recommendations on research data management at the Friedrich Schiller University Jena).

Research Data and Metadata:

In the project planned, zoological data about diversity and key traits of two extremely species-rich groups of moths (Geometridae, Arctiinae) will be collected during

- field sampling — sampling of approx. 30,000 moth individuals and corresponding collection data
- laboratory work — all individuals will be spread and prepared according to museum standards
- molecular analysis — COI gene analysis for most individuals
- photography — standardized photographs of approx. 5,000 specimens including full spectrum photography (near UV and human visible).

The volume of the data which will be submitted to GFBio has been estimated to approx. 100GB and the number of data sets to more than 30,000 (if one moth specimen represents one dataset).
Data collection will apply common standards and methodologies to take account of interoperability and reusability. GFBio also provides persistent identifiers (e.g. DOI, URI) in regard to data citation. Special attention will be paid to the requirements regarding the collection of sensitive data according to the Nagoya Protocol. These sensitive data will not be part of the data submitted to GFBio.
Relevant metadata will be collected and documented in order to make the data findable, accessible, interoperable and reusable. With respect to a sustainable metadata management we intend to seek the advice of GFBio (see service offer). As the project is set up in the form of a onetime observation, data cannot be reproduced. All data produced will be stored in an openly documented format. If proprietary formats will be necessary, documentation will include the documentation of software products if permitted under copyright.
Data backup during project runtime will be managed by the PI. Data and processed data resulting from them, as well as metadata will be stored on the local servers and thus will be part of the backup routines run by the Universitätsrechenzentrum der Friedrich-Schiller-Universität (URZ). In addition, the PI will regularly backup the data on a harddrive.
Publication data as well as primary data of high interest for future investigations (e.g. selected imaging formats) will be archived in the GFBio data center *Zoologisches Forschungsmuseum Alexander Koenig (ZFMK).*

Archiving, Publication and Licensing

GFBio

- **Collection data** will be submitted to GFBio in a pre-curated format and will be published via the GFBio portal under CC BY-NC-SA licensing. For that purpose, GFBio provides data exchange standards, protocols and formats relevant for the collection data domain within the GFBio network via the GFBio portal. Qualified staff provided by ZFMK will curate data and metadata.
- **Raw photographs** and, if reasonable, additional **selected imaging formats** will be long-term stored and published via Morph D Base and linked to the collection data in the GFBio portal (www.gfbio.org) under CC BY-NC-SA licensing. Data volume will not exceed 100GB. Data will be managed with the online database Morph D Base which is hosted and developed at the GFBio data center *Zoologisches Forschungsmuseum Alexander Koenig (ZFMK)*. The proposer is responsible for data management in Morph D Base. A personal login will be provided for this purpose. Access restrictions can be managed in Morph D Base, which enables collaboration between project members. A URI is generated for data publication; data cannot be modified after publication and are citable. All data uploaded in Morph D Base will be archived. Basic funding for the sustainable operation of Morph D Base is covered by ZFMK. GFBio guarantees the online availability and long-time archiving (at least 10 years) of the submitted data.

Other

- **Molecular and sampling data** will be uploaded to BOLDsystems according to their guidelines, as well as photographs of the photographed specimens.
- In addition to the research data collected, **specimens** fitted with individual labels (containing collection data) and unique individual identifier will be stored in the collections of the Phyletisches Museum, Jena.

Unique identifiers for both data (collection data, molecular data, multimedia data) and objects will guarantee a clear linkage between the data and the objects collected.
Data accompanying publications will be published in citable data packages under CC BY-NC-SA licence. Metadata will be published directly after the submission to and the curation by the GFBio data centers. We intend to have all datasets published one year after the projects' end. Subsequently, data will be available via the GFBio web portal. GFBio will provide persistent identifiers (e.g. DOI) in regard to data citation for each published dataset/data package.

Costs for Data Management Support

Costs and services provided by GFBio are specified in the GFBio e.V. service offer 2018001.

GFBio recommends

GFBio provides individual data management support. We give advice regarding storage, security, quality assurance and backup and help you optimizing the findability, accessibility and reusability of your research data. We highly recommend using common standards for data and metadata formats.