Data Management Plan

Project:	VielFalterGarten
Principal Investigator/s:	research group at iDiv Ecosystem Services[1] Prof. Dr. Aletta Bonn, aletta.bonn@idiv.de
Project Data Contact:	Christian Nappert, christian.nappert@ufz.de
Funding :	Bundesprogramm Biologische Vielfalt[2]

Project Focus

"VielFalterGarten"[3] is a communication and education project that aims to show how a biodiversity-promoting development of urban spaces for butterflies can succeed. The project pursues four main areas of action: Education, Butterfly protection, Citizen Science and Network formation.

Data is collected especially in the third column 'Citizen Science' that has an impact on the other three areas (education: people gather hands-on data, butterfly protection: people create gardens and care for butterflies, network formation: different participants interact with a similar cause). Participants can register, mark a location in Leipzig and then observe butterflies for 15-minutes on their chosen locations regularly.

Relevant Policies and Guidelines

Research data management will follow the DFG Guidelines on the Handling of Research Data[4], the DFG Guidelines on the Handling of Research Data in Biodiversity Research[5] and the DFG Guidelines for Safeguarding Good Scientific Practice[6]. The FAIR principles[7], the requirements of the DGSVO[8] and the MyDiv Experiment – Data sharing and publication policy[9] are taken into account.

---

[1] https://www.idiv.de/en/groups-and-people/core-groups/ecosystem-services.html, https://www.ufz.de/index.php?en=34228

[22] https://biologischevielfalt.bfn.de/bundesprogramm/bundesprogramm.html

[3] https://www.vielfaltergarten.de/

[4] https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_research_data.pdf

[5] https://www.dfg.de/download/pdf/foerderung/antragstellung/forschungsdaten/guidelines_biodiversity_research.pdf

[6] https://www.dfg.de/download/pdf/foerderung/rechtliche_rahmenbedingungen/gute_wissenschaftliche_praxis/kodex_gwp_en.pdf

[7] https://www.force11.org/group/fairgroup/fairprinciples

[8] https://eur-lex.europa.eu/legal-content/DE/TXT/HTML/?uri=CELEX:32016R0679

[9] https://www.idiv.de/fileadmin/content/Files_EIE/MyDiv/Data_sharing_and_publication_policy_MyDiv.pdf

General Data Management

The research data management Unit[10] from UFZ is responsible for the coordination of the research data management. The data will be collected via a custom mobile app called "VielFalterGarten-App". The app includes different features from other apps like eBMS[11], iNaturalist[12], Schmetterlinge Österreich[13] and BigButterflyCount. A MySQL[14] database will be used for the data storing and queries. DBeaver[15] will be used for database management. QGIS[16] will be used for GIS analyses and visualisation. Web programming will be done by using MS Visual Code[17].

All data sets receive a unique identifier, which will link data throughout the project. Research data will be in origin from (1) fieldwork / observations, (2) surveys and (3) statistical evaluations. Georeferences are part of the dataset itself or of the metadata. The ID will be part of the dataset itself or it will be part of the metadata. Data processing steps will be documented by scripts and macros, or if not possible within a chronological list of the processing steps.

The project "VielFalterGarten" expects the total data volume not to exceed **1 TB** and the number of datasets will be under 100.

The data management plan is a living document and as such will be updated regularly (e.g. once or twice a year) during the project runtime.

Expected research data

I. *Observation data*

Observation data will include the butterfly counts and the coordinates of the garden segments and the butterfly locations, which are managed in tables inside the MySQL-database. The tables will be exchanged, archived and published by using the file-formats .xlsx, .csv, .xml. Additional garden descriptions and observations notes will be logged in text documents (.txt, .docx) for data exchange, archiving and publication.

II. *Multimedia data*

The "VielFalterGarten" project user will take photographs of the observed butterflies and store them as .tiff or .jpg files. For each observed butterfly the users should have the option to take photographs (each about 5 MB). The photographs will be analyzed using an open source software of the TU Ilmenau, which is comparable to Flora Incognita[18], to identify the butterflies. Author information (e.g. user name) and terms of use (e.g. notice in the legal notes, that the photos are only used for scientific purposes) are part of the metadata.

The "VielFalterGarten" project user will take video recordings of the butterflies and the interviewed persons and store them as .mp4 files. Recordings of interviews would only be evaluated and not be archived.

---

[10] https://www.ufz.de/index.php?en=45348

[11] https://butterfly-monitoring.net/de/ebms-app

[12] https://www.inaturalist.org/

[13] https://schmetterlingsapp.at/app-herunterladen

[14] https://www.mysql.com/de/

[15] https://dbeaver.io/

[16] https://www.qgis.org/de/site/

[17] https://code.visualstudio.com/

[18] https://floraincognita.com/de/apps/flora-incognita/

The "VielFalterGarten" project will take sound recordings of the interviewed persons and store them as .mp3 files. Recordings of interviews would only be evaluated and not be archived.

Multimedia data will be analyzed with Adobe Creative Cloud[19]. The interview recordings are documented and managed as anonymous transcripts in tables (.xlsx, .csv) and text documents (.docx, .txt), which will be archived.

III. Software code

Software code will be from a developed app, which will be developed by the team[20] at the UFZ ESS. Android Studio[21] and Microsoft Visual Code[22] will be used for app programming. Additional software code will be also from VielFalterKarte, an interactive map for analysing and visualisation of the research data. The softwarecode (in Vue, TypeScript, PHP, CSS) will be published on GitLab[23] instance of UFZ.

IV. Documentation and code for data processing and simulation models

Further, upon publication code will be made public in GitLab[24] instance of UFZ. Public code will be documented, e.g. with R markdown, R notebooks, Jupyter notebooks or related tools. Where source code is not available to clearly define steps taken to process data, all other tasks will be equally documented in the same tools. Code/documentation, raw data and final data will be linked with each other for transparency and to ease reproducibility.

V. User data and coordinates

User data will be collected via the registration for the participation of the "VielFalterProject" and during the data collection via the app. 50 new users are expected per year. After the project runtime of four years there will be approx. 200 users. The user data is sensitive data and will be managed with taken account to the requirements of the DGSVO. The coordinates and the user data will be anonymized and genericized before publishing. The coordinates are given an embargo so that they are only made available for scientific purposes on request.

Metadata

Metadata will meet the requirements of the GFBio submission[25] (title, description, data collection time, dataset label, categories, dataset author, license, related publication/s, metadata schema, embargo, legal requirements). Metadata or the dataset will contain geographical coordinates for spatial reference as well as information about the respective reference system. The "VielFalterGarten" project will document metadata using open formats such as .xml or .csv following common standard Darwin Core[26], ABCD[27] or another appropriate open format if necessary, in order to make the data findable, accessible, interoperable and reusable. The "VielFalterGarten" project will create a reference documentation of all metadata and use variable names and entities consistently and globally for all project data.

---

[19] https://www.adobe.com/de/creativecloud.html?promoid=LCDWTLKJ&mv=other#

[20] Volker Grescho and Christian Robert Nappert

[21] https://developer.android.com/studio

[22] https://code.visualstudio.com/

[23] https://git.ufz.de/

[24] https://git.ufz.de/

[25] https://submissions.gfbio.org/

[26] https://github.com/tdwg/dwc

[27] https://github.com/tdwg/wiki-archive/blob/master/twiki/data/ABCD/WebHome.txt

Data storage and backup during the project

The project partners[28] are responsible for data backup until the end of the project. They perform a regular data backup, e.g. once a week. The data is stored on servers of the iDiv, which guarantee a daily backup routine. The team of iDiv IT-Support[29] will give advice and support.

Data exchange with cooperation partners

TMD[30] and the working group at the ESS are using nextCloud for data exchange. Folders and files can be shared individually.

Archiving, Publication and Licensing

All primary data, as well as selected data products (which cannot easily be reconstructed or which are related to a paper publication) and their contextual metadata will be submitted to GFBio to be long-term archived in GFBio´s associated archives. Data will be published and made available via the GFBio portal as well as directly by the archives hosting the data. Where applicable, the data will additionally be available via other connected providers (e.g. species observations via GBIF). Interlinkage between datasets will be ensured via PIDs.

GFBio provides data exchange standards, protocols and formats relevant for the submission of data within the GFBio network[31]. The curators of the associated data centers with expertise in GFBio submissions will give advice and support. Submitted data sets/packages will be provided with a persistent identifier (most likely DOI). All data will be archived for at least 10 years.

Primary data will be submitted preferably within six months after its collection but no later than the project's end date. After the submission and curation, all data will be publicly available via the GFBio portal under CC BY-NC-ND: Creative Commons Attribution-NonCommercial-NoDerivs-4.0[32].

Data linked to an article (or another kind of publication) will be submitted before or at the same time as the publication of the article. This option will allow you to refer to your dataset via unique identifier (e.g. DOI) within your article. Datasets will be submitted at the project's end.

Research data that cannot be archived and published via GFBio will be submitted to the Datenmanagementportal (DMP) of the UFZ[33] for archiving and publication.

Costs for research data management and archiving

During the runtime of the GFBio project, GFBio offers to consult and support the applicant regarding data management efforts (e.g. refining the data management plan, preparing data for submission to GFBio, providing data management training) without charging additional costs. The costs for the curation and archiving of the research data are determined by GFBio and, if necessary, collected as soon as the "VielFalterGarten" project submits research data.

---

[28] BUND Leipzig, the city of Leipzig and scientists from the Helmholtz Centre for Environmental Research - UFZ and the German Centre for Integrative Biodiversity Research (iDiv) invite citizens to participate.

[29] https://www.idiv.de/de/gruppen_und_personen/zentrales_management/it_support.html

[30] https://www.ufz.de/tagfalter-monitoring/

[31] https://www.gfbio.org/support/wiki >Forms and Assessments

[32] https://creativecommons.org/licenses/by-nc-nd/4.0/

[33] https://www.ufz.de/drp/en/