

UPSCALING THE BENEFITS OF PUSH-PULL TECHNOLOGY FOR SUSTAINABLE AGRICULTURAL INTENSIFICATION IN EAST AFRICA



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 861998.



D9.3:

Data Management Plan

04/2021



Author(s)/Organisation(s)	UWUE
Contributor(s)	Jie Zhang (UWUE), Melanie Schulte (BayFOR), Emily Poppenborg
	(LUH), Volker Soßna (LUH), Adewole Olagoke (LUH)
Work Package	WP9
Delivery Date (DoA)	April 30, 2021
Actual Delivery Date	April 29, 2021
Abstract:	This report covers the data management processes to be followed, respecting the EC principles on FAIR data management, covering procedures in making data discoverable (including provisioning of metadata), making data openly accessible and interoperable, and addressing data re-use considerations and related processes. The report also covers the General Data Protection Regulation (GDPR) policies adopted in UPSCALE for the protection of personal data.

Document Revision History			
Date	Version	Author/Contributor/ Reviewer	Summary of main changes
23/02/2021	1	Jie Zhang	First draft
03/03/2021	2	Melanie Schulte	Review & inputs
24/03/2021	3	Emily Poppenborg	Review & inputs
29/03/2021	4	Volker Soßna	Review & inputs
30/03/2021	5	Adewole Olagoke	Review & inputs
07/04/2021	6	Jie Zhang	Second draft
19/04/2021	6	Emily Poppenborg	Final version

Dissemination Level		
PU	Public Deliverable	\checkmark
РР	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
со	Confidential, only for members of the Consortium (including the Commission Services)	



UPSCALE Consortium			
No	Participant Organisation Name	Short Name	Country
1	LEIBNIZ UNIVERSITY OF HANNOVER	LUH	DE
2	THE INTERNATIONAL CENTRE OF INSECT PHYSIOLOGY AND ECOLOGY LBG	ICIPE	KE
3	BAYERISCHE FORSCHUNGSALLIANZ BAVARIAN RESEARCH ALLIANCE GMBH	BayFOR	DE
4	EASTERN AFRICA FARMERS' FEDERATION SOCIETY	EAFF	KE
5	JOMO KENYATTA UNIVERSITY OF AGRICULTURE AND TECHNOLOGY	JKUAT	KE
6	KENYA AGRICULTURAL AND LIVESTOCK RESEARCH ORGANISATION	KALRO	KE
7	MASENO UNIVERSITY	MU	KE
8	FH ASSOCIATION	FH	СН
9	RWANDA AGRICULTURE AND ANIMAL RESOURCES DEVELOPMENT BOARD	RAB	RW
10	INOSENS DOO NOVI SAD	INOSENS	RS
11	UNIVERSITY OF KWAZULU-NATAL	UKN	ZA
12	LUNDS UNIVERSITET	ULUND	SE
13	SVERIGES LANTBRUKSUNIVERSITET	SLU	SE
14	UNIVERSITAT ZURICH	UZH	СН
15	TANZANIA AGRICULTURAL RESEARCH INSTITUTE	TARI	TZ
16	NATIONAL AGRICULTURAL RESEARCH ORGANISATION	NARO	UG
17	INSTITUTE FOR SUSTAINABLE DEVELOPMENT	ISD	ET
18	JULIUS-MAXIMILIANS-UNIVERSITAT WURZBURG	UWUE	DE

LEGAL NOTICE

The information and views set out in this application form are those of the author(s) and do not necessarily reflect the official opinion of the European Union. Neither the European Union institutions and bodies nor any person acting on their behalf may be held responsible for the use which may be made of the information contained therein.

Funding Scheme: Research and Innovation Action (RIA) ● Theme: H2020-SFS-2019-2 Project Start Date: 01 November 2020 ● Duration: 60 months

> © UPSCALE Consortium, 2021 Reproduction is authorised provided the source is acknowledged.



Table of contents

1.	Infra	frastructure summary		
2.	2. Roles and responsibilities			
3.	Data	ta Summary7		
3	.1	Purpose of data collection7		
3	.2	Data types and formats7		
3	.3	Reuse of existing data		
3	.4	Expected size of the data		
4.	FAIR	R Data Management		
4	.1	Making data findable, including provisions for metadata9		
4	.2	Making data openly accessible10		
4	.3	Making data interoperable12		
4	.4	Increase data re-use12		
5.	Allo	cation of resources13		
6.	Data	a security		
6	.1	Research data security14		
6	.2	Project management data security14		
7.	7. Ethical aspects15			
Арр	endix	17 1: Data publishing and sharing policy		

List of Abbreviations and Acronyms		
DMP	Data Management Plan	
DPO	Data Protection Officer	
EC	European Commission	
WP	Work Package	
RS	Remote Sensing	
GIS	Geographic Information System	
GDPR	General Data Protection Regulation	



Executive summary

This deliverable describes the data management procedures applied in UPSCALE for safe management and exchange of administrative and research data, in accordance with the GDPR and principles of FAIR data management (Findable, Accessible, Inter-operable, Re-usable). We first describe the overall infrastructure of how data is managed in UPSCALE, associated roles and responsibilities of the partners, and the types and size of the data to be used and collected in UPSCALE. UPSCALE upholds each pillar of the FAIR principle using the provisions and capacities of an online database tailored to the project needs for research data to be transferred to a long term repository after project end, the project management platform EMDESK for administrative data, and clear open access policies. We also describe the allocation of resources to data management in UPSCALE, and the measures taken to ensure data security and compliance with ethical data principles.

1. Infrastructure summary

In the UPSCALE funding period, all research data collected or generated will be stored in the online UPSCALE database, which will be hosted and maintained by the University of Würzburg (UWUE). Personal data will be anonymized before upload. This database accepts access from only UPSCALE partners via validated user accounts. For external reviewers, an account is approved case by base. Access of research data requires a request-review-grant-procedure provided by the database, where only authors have the control to grant access. At the end of the funding period, validated research data will be migrated to long term data archive centres (e.g. Pangaea, GFBio, INSPIRE) to guarantee long term preservation and accessibility. Data related to project management is managed by BAYFOR via a cloud storage system EMDESK provided by Telekom. After the funding period, migration of project management data to individual institutes of the UPSCALE partners will be carried out upon request. The data management infrastructure of USPCALE is summarized in Figure 1.



Figure 1: Data management infrastructure of UPSCALE project



2. Roles and responsibilities

Data Security Manager (DSM, UWUE)

The DSM (Jie Zhang, UWUE) ensures a correct and secure handling of all collected data, in particular: compliance with local and European data security laws; verifies that personal data have been appropriately blinded; and uses adequate encryption technology for the transmission of data. The DSM reviews all technical components to ensure secure and correct data handling. The DSM and <u>Data Exploitation Manager</u> (see below) will meet once a year to ensure that data security measures are up to date with regard to any ongoing data exploitation issues.

Data Exploitation Manager (DEM, icipe)

The DEM (Frank Chidawanyika, icipe) pays attention to the protection of the results. Tasks include: paying attention to any intellectual property arising from UPSCALE; paying attention to the protection of the results; advising the partners how to exploit the commercial potential of the results; leading the tasks for exploitation activities (WP8.5); monitor the WP and tasks dedicated to demonstration activities; supervising project progress milestones concerning demonstration issues; preparing the exploitation part of the reports to be submitted to the EC; presenting and discussing exploitable results at the annual meetings.

Data Protection Officers (DPO)

Each beneficiary must confirm that it has appointed a Data Protection Officer (DPO) and the contact details of the DPO are made available to all data subjects involved in the research via Deliverable 10.2 (Ethics). For beneficiaries not required to appoint a DPO under the GDPR, a detailed data protection policy for the project is kept on file on the UPSCALE Project Management Platform EMDESK and available to be submitted to the Agency upon request.

DPO check whether data protection measures are put into practice throughout the project; revise the drafts of consent forms and information leaflets etc. for GDPR compliance; train and counsel all project members involved in handling personal data, especially the interviewers; function as contact persons for informants with questions related to privacy and data protection; ensure that data is deleted immediately in case consent is withdrawn retrospectively.

Principal Investigators (PI)

PI ensure that every member of their working group knows, understands and follows this DMP. They communicate with the DSM and DEM whenever questions arise on practical issues regarding data handling.



3. Data Summary

3.1 Purpose of data collection

UPSCALE will test under what landscape, climatic and soil conditions the push-pull system provides the best economic and environmental support over short and long time scales for individual farmers in East Africa and their communities, using novel experiments, modelling and data. We will collect field experiment data and socioeconomic data, generate Geographic Information System (GIS) data, Remote Sensing (RS) data as well as modelling data. Using these data, we will a) address the connection between push-pull approaches and climate resilience, b) assess the links between pushpull technology adoption and aspects of land tenure, gender equality, farmer incomes, value chains, and sociocultural aspects of communities, c) expand knowledge of push-pull applicability to better target and upscale the existing technology to different cropping systems, as well as to expand the establishment of push-pull fields into other regions using integrated soil, water and pest management approaches. This report provides the Data Management Plan (DMP) that UPSCALE follows throughout the project lifetime.

3.2 Data types and formats

The UPSCALE project outputs will provide a set of components based on storage and modelling of several types of data such as the following: 1) Maps of chemical and optical traits, and species distribution and interactions in and around fields; 2) Maps of projected yields under push-pull (and conventional practices), dynamics of push-pull adoptions and socioeconomic effects, given different Shared Socio-economic Pathway (SSP) and Representative Concentration Pathway (RCP) scenarios for climate change; 3) Household socioeconomic baseline and mid-term evaluation data, adoption profiles, demographic and economic data on production, costs, incomes, livelihoods, map of the sampled households in the project sites; 4) Very high resolution maps of the distribution of push-pull-fields in the focus areas and at lower resolution in agricultural areas for all study countries; 5) Estimates of ecosystem functions in a range of current push-pull and monocrop fields, including available water holding capacity, water infiltration capacity, potential nitrogen mineralization, soil texture, total soil nitrogen, organic carbon and Mehlich-extractable phosphorus, biodiversity of invertebrates, pressure of invertebrate pests and Striga weeds, natural pest control levels, crop yields, and predator-prey food web structure.

Data formats created, acquired and collected will include: 1) tabular data in ASCII format (*.txt, *.csv); 2) spatial data, including vector data in ESRI shapefile/geodatabase format, raster data in Geo-tiff format; 3) Survey questionnaire and related quantitative data; 4) Audio files (.mp3 etc.), multimedia files (video recordings in *.mp4), images (*.jpg); 5) model scripts; 6) project reports, presentations, documentations in text (*.txt), Pdf (*.pdf), MS Word (*.docx), MS PowerPoint (*.pptx) format.



3.3 Reuse of existing data

We will use various existing RS and GIS data such as Worldview 2 remote sensing data (https://earth.esa.int/web/eoportal/satellite-missions/v-w-x-y-z/worldview-2), data from the Copernicus Program (Sentinel-1/2) (https://sentinels.copernicus.eu/web/sentinel/home), digital elevation model from SRTM (http://srtm.csi.cgiar.org), land cover map from CCI Land cover (http://2016africalandcover20m.esrin.esa.int/), soil map from European Soil Data Centre ESDC (https://esdac.jrc.ec.europa.eu/content/soil-map-soil-atlas-africa), total N and total P from World Soil Information ISRIC (https://data.isric.org/geonetwork/srv/eng/catalog.search#/home), topographical GADM (https://www.gadm.org) and Digital data from Chart of the World (https://worldmap.harvard.edu/data/geonode:Digital Chart of the World), Climate data from WorldClim (https://www.worldclim.org/). These data will be analysed and synthesized to provide a comprehensive assessment of soil, climate, and landscape characteristics of the study region. We will also use existing data from previous projects and networks of the UPSCALE partners to extract and organize information and indicators on push-pull spread and socioeconomic parameters. Furthermore community and regional survey data, and agro-economic data will be derived and compiled for community- and regional-level assessments. These data derives from sources like publically available National Panel and Socioeconomic Surveys performed by national or international organizations including the World Bank in corresponding areas, and from the e-Granary app.

3.4 Expected size of the data

The initial estimation of total size for research data is 4TB. Collected and generated data will be curated and preserved in the UPSCALE database that will follow data quality, security, and privacy standards and relevant EU legislation (i.e., GDPR). Each WP as well as local partners will use the UPSCALE database for research data preservation and exchange.

The estimated size for project management data (project reports, presentations, documentations, images and multimedia files) is 500 GB. These data will be preserved under EMDESK as described in chapter 2.1 of this document.



4. FAIR Data Management

4.1 Making data findable, including provisions for metadata

Data of the UPSCALE project can be classified into 2 categories as demonstrated in Figure 1: 1) research data, which includes ecological data from field experiments, data generated from loggers, spatial data, socioeconomic data from surveys and interviews, model scripts, qualitative interview data, reused data from other open sources; 2) project management data, which refers to project reports, meeting minutes, conference presentations and posters, specification documents and tutorials, videos and photos from meetings.

<u>Research data</u>: these data will be preserved and exchanged between project members in the UPSCALE database. This database will be developed based on BExIS (Biodiversity Exploratory Information System, <u>www.bexis.uni-jena.de</u>). It will adapt the specific requirements of the UPSCALE project and be further improved and developed with new functionalities by UWUE. We chose our own database solutions instead of other data centres to guarantee that 1) full control of research data access 2) own designed meta data structure to meet the specific need of UPSCALE project and is compatible with international standard 3) datasets are deeply integrated via the shared study sites, merging and aggregation tools are offered by the database for data synthesis 4) trainings and real time consultant is offered throughout the project phase. We follows the "as open as possible, as closed as needed" principle to restrict and regulate data access using the following approaches: 1) metadata is openly accessible while research data is restricted with user log in; 2) only project members are allowed to create log in account to the database; 3) only authors can view and download his/her datasets, others have to go through the online automatic request-review-grant-procedure where authors can decide when to grant access to request; 4) every data download is recorded and traceable; 5) users must agree to UPSCALE data sharing policy before downloading.

All the research data preserved under the UPSCALE database will follow the FAIR principle with the following details:

1. <u>Data identifier</u>: Every dataset will have a persistent and unique identifier called Dataset ID throughout the entire project.

2. <u>Metadata</u>: Metadata input is mandatory for all research data and is recorded at the time of dataset creation. User will input their metadata manually via the online metadata creation wizard of the UPSCALE database. UPSCALE metadata uses the Extensible Markup Language (XML) standard. It will be based on Ecological Metadata Language EML (https://eml.ecoinformatics.org), further simplified and extended upon specific requirements of interdisciplinarity in the UPSCALE project, and be made compatible with metadata standard from data archive centres (e.g. INSPIRE, Pangaea). The metadata is openly accessible for all audiences, and does not require user log in. It will help to attract international interests and guarantee transparent research.

3. <u>Search function</u>: Users can search for various types of keywords (author, year, research topic, or other free text) in the metadata, thus making datasets easily findable.

4. <u>Versioning:</u> Versioning is included in all structured datasets and is identifiable via Version ID. One Dataset ID can be linked with multiple version ID for multiple versions. When downloading a dataset, user has to choose a specific version ID in order to download the desired version.



<u>Project management data</u>: The web-based project management platform EMDESK is being used for project management documentation exchange among consortium partners. The EMDESK platform is fully operational and has been customized to the needs of the UPSCALE consortium (described in Deliverable D 9.1 – Web-based collaboration platform). It serves as a centralized access to the relevant information and documents of importance to monitor progress in the project and to promote communication between partners, revisions of documents and their repository. The data and document repository section allows for data annotation and content setting using tags, file grouping, commenting and adding keywords as well as document versioning. These features combined with the proper data and file-naming conventions (see Deliverable 9.2: Project Management Plan) provide an overall efficient data searching capability for UPSCALE directing to the data itself, data owners, report owners/authors as well as data contributors. To further support searching capabilities, each report owner will include some meta-data (keywords) associated to the document when uploading it to the repository, for easier searching in documents and reports/deliverables etc.

4.2 Making data openly accessible

<u>Project management related data</u> stored under EMDESK is accessible for all project members without restriction as long as no confidential information is considered. Any members who have an account under EMDESK UPSCALE group can upload/download project management data. Project management related data for example comprise contact details of partners, financial information (expenses and resources), meeting minutes and participants lists. These data can only be shared outside of the UPSCALE project if the following conditions are met: 1) no confidential, personal, financial or commercially sensitive information are included; 2) permission from the relevant stakeholders has been obtained; 3) sharing the data does not damage exploitation or IP protection prospects.

Research data is made accessible for project members via the UPSCALE database, which is hosted by UWUE. The database is made available via a website encrypted by Hypertext transfer protocol secure (HTTPS) and is accessible internationally in all types of electronic devices and software platforms. In this database the metadata of datasets are made publicly accessible for all audiences beyond the scope of the UPSCALE project. This aims at better transparency for UPSCALE research as well as attracting more public attention and possible further research collaboration. Observation data, or primary data collected/generated by UPSCALE members is closed during the project period where only authors are allowed to access and edit. Other members of UPSCALE, who are interested in accessing these data, should follow the UPSCALE data sharing policy and a request-review-grantprocedure (Figure 2) provided by the UPSCALE database as follows: 1) the requester writes a small proposal text on the database website about why and how he/she plans to use these data; 2) the database will automatically send an email to all authors with the proposed request and the requester, but only the first author will receive a link to grant or deny access; 3) when authors read this email and make a decision on this request, the first author would click on the link to either grant or deny this access request; 4) the requester will get an automatic email notification after his/her request is granted or denied. If he/she is granted with access, he/she is able to download this dataset. This way the authors will have full control of who can access their data, how and when their data should be implemented in synthetic research. UPSCALE has a data sharing and data publishing policy (see Appendix 1) which will be available as a text page in the database. Any UPSCALE member who wants to request others' data needs to agree to this policy before he/she can continue with the downloading



procedure. Documentation and tutorial of the database will be presented in text format as well as video format.



Figure 2: Request-review-grant-procedure for research data in UPSCALE

After the funding phase of UPSCALE, datasets will be reviewed by the UPSCALE Steering Committee and data owners and must be approved before becoming candidates for contribution to the open research pilot. These parties and the originating consortium partner(s) will then agree on licensing (https://creativecommons.org/publicdomain/zero/1.0/deed.en), such as CC0 CC ΒY (https://creativecommons.org/licenses/by/4.0/deed.en), CC **BY-NC** (https://creativecommons.org/licenses/by-nc/4.0/deed.en), CC **BY-NC-ND** (https://creativecommons.org/licenses/by-nc-nd/4.0/deed.en), CC **BY-NC-SA** (https://creativecommons.org/licenses/by-nc-sa/4.0/deed.en), CC **BY-ND** (https://creativecommons.org/licenses/by-nd/4.0/deed.en), CC BY-SA (https://creativecommons.org/licenses/by-sa/4.0/deed.en). Following in principle approval, UPSCALE will make the datasets available through suitable open access repositories such as GFBio and PANGAEA, following the guidelines of the Grant Agreement, in order to guarantee long term preservation of our data. In particular, we will only choose repositories providing citable persistent

identifiers such as DOI on upload and the possibility to indicate a license. We will also make the source code of the UPSCALE database publicly available so that more projects can set up a similar database on their own.



4.3 Making data interoperable

Both research data and project management data can be exchanged and re-used within the project period among all the institutions involved in the UPSCALE project.

Research data in UPSCALE can be categorized due to different curation methods into structured data and unstructured data. Structured data is preserved and indexed in database tables, which provides further manipulation possibilities such as filtering, aggregation, statistical analysis and visualisation directly from the server. Structured data uses ASCII format for uploading/downloading to guarantee maximum interoperability. Unstructured data will be stored as Binary large object (BLOB) in the database management system. It includes spatial data such as shapefiles, satellite images, videos, audios, and images. We use ESRI shapefiles for spatial vector data, Geo-tiff format for spatial raster data, MP4 format for videos, MP3 format for audios, JPEG format for images, in order to be consistent and compliant with the most common software applications across different operating platforms.

Our metadata will be in XML format, which is standardized throughout the project and is machinereadable. This metadata standard is based on Ecological Metadata Language (https://eml.ecoinformatics.org), because a large part of the data generated in our project is from earth and environmental science. We will further simplify its syntax and append it with new semantic vocabularies from socioeconomic science and other disciplines involved. We will use this metadata standard for all the data types presented in UPSCALE to allow inter-disciplinary interoperability. If ontologies are proven necessary during the data collection and documentation, we will map our vocabularies to ontologies that is compatible with ontologies offered by open data centres such as INSPIRE.

4.4 Increase data re-use

UPSCALE will participate in the Open Data Pilot of the EU Horizon 2020. Open access to research data will be provided whenever there is no conflict with the protection of results and GDPR and under the supervision of the consortium. During the project phase, collected and generated data will be curated and preserved in the UPSCALE database where project members are able to exchange their data. At the end of the project, validated datasets will be transferred to European certified research data archive centres which comply with sustainability and open access requirements. Embargo times for opening access to datasets are defined by each dataset owners in the metadata, but in theory not later than 3 years after data collection. In practice if one needs longer embargo time, he/she shall inform the UPSCALE data security manager the reason for delay and the estimated time for data publishing. All data that can encompass any personal data protection or privacy and IPR issues will not be publicly disclosed and datasets will be stripped of these issues (e.g., anonymized) before publication following the procedures described in deliverables D10.1 and D10.2 (Ethics requirements).

For the publication of scientific results, "Green Open Access" (i.e. self-archiving) and "Gold Open Access" (i.e. open access publishing) will be chosen depending on each publication. For green open access, researchers will deposit the final peer-reviewed manuscript in a public database of their choice, ensuring open access to the publication within the embargo period of on average no more than six months.



5. Allocation of resources

WP9 is responsible for the data management life cycle monitoring for all datasets to be collected, processed or generated by the project. In order to ensure compliance with data management decisions as they relate to the DMP, the following measures apply in UPSCALE:

- WP leaders are considered responsible for adhering to the specifications above in their respective work packages.
- The principal investigators and the Data Protection Officer (see deliverable D10.2) of each beneficiary organization are considered responsible for the DMP actions. They will support the WP 9 in all issues related to research data management. The principal investigators of each beneficiary should ensure that personnel working on the project have read the Data Management Plan and apply/exercise all the principles as described in the UPSCALE DMP (this document).
- Data collectors have the ultimate responsibility of complying with the specifics of the UPSCALE Data Management Plan, as well as with the related GPDR policies.
- For the overall UPSCALE project activities, the Data Security Manager (DSM) at UWUE is the main contact point to assist compliance of all project partners with the DMP guidelines to ensure secure and correct data handling in UPSCALE, including technical support for the database. The DSM regularly reviews all technical components of secure data handling and storage and communicates on a yearly basis (whenever the need arises) with the Data Exploitation Manager (DEM, icipe) to ensure safe and correct handling of data exploitation issues.

UWUE is in charge of research data management in UPSCALE. This includes costs for a server, software, and the DSM for developing the database, providing counsel and support, organising data workshops and other training. To achieve long term data preservation we will transfer our data to European research data archive centres such as CESSDA (https://www.cessda.eu) for socioeconomic data, GFBio (https://www.gfbio.org) or PANGAEA (https://www.pangaea.de) for earth and environmental data. The data migration will be carried out by the DSM and data owners. The goal is to keep UPSCALE data accessible and re-usable for at least 10 years after the funding phase. The overall cost is ca. 64,000.00 €.

BayFOR is in charge of the project management related data. This includes providing and managing the EMDESK platform as well as training partners and organization in using data deposited on the platform. The allocated personnel resources are a percentage of the resource allocation of BayFOR to WP9 (total 30PM for the whole WP or 223,596.00 €). The further allocation is the cost of the EMDESK software license (11,062.92 € for the whole project period and 50 participating persons), noting that the software is used not only for data management but also for planning and communication purposes (see Deliverable D9.1).



6. Data security

6.1 Research data security

UPSCALE database will be hosted in a Windows server provided the computer centre of UWUE. The server is backed up automatically on a daily bases, secured by firewall of UWUE and the firewall of the server itself. UPSCALE database uses PostgreSQL as database server, which offers encryption at several levels and provides flexibility in protecting data from disclosure due to database server theft, unscrupulous administrators, and insecure networks. UPSCALE database website uses encrypted connection via HTTPS and applies encrypted Forms-Based Authentication. Each user account in the UPSCALE database is verified and validated by the Data Security Manager (DSM). In theory this database only accepts accounts from UPSCALE partners. For external users, e.g. project reviewers, their account will be discussed within the steering committee about its necessity and be approved individually.

6.2 Project management data security

All relevant team members of the UPSCALE beneficiaries have received personalized access to the EMDESK project management platform. The site is accessible with personal, password-protected accounts and therefore only for internal use within the consortium. If the need arises to invite external viewers, they can be provided with password-protected guest accounts in read-only view. All team members can obtain personalized access rights to folders and documents, however the consortium aims at complete transparency, thus every member has reading rights for all documents and editing rights for documents of all work packages the person is involved in. Solely the resources section will be limited to the Coordination Team to safeguard financial information.

To ensure the highest security of the infrastructure, EMDESK is hosted in the Open Telekom Cloud (OTC) - one of the most secure and modern cloud data centres worldwide. All services are strictly regulated and are regularly audited and certified by independent institutions to ensure they meet the latest security and data protection requirements (e.g. TISAX, Trusted Cloud, ISO 14001, ISO 22301). It provides protection from virus-attacks and paralysation. The software undergoes continuous improvement processes and updates.



7. Ethical aspects

Ethical aspects with regard to data management are described in detail in Deliverables D10.1 and D10.2 (Ethics requirements). UPSCALE will ensure that Article 34 of the Horizon 2020 Framework Programme (Grant Agreement- Ethics) is properly implemented and, in particular, that all the data sharing and re-use activities for this Horizon 2020 project comply with ethics principles and relevant national, EU and international legislation.

Ethic aspects in using material from surveys and interviews: The project will ensure strict application of ethical requirements when preparing, conducting, and evaluating interviews. Procedures of using material from surveys and interviews will be documented, and informed consent from all participants will be ensured before starting the interviews. We will confirm the content to be generated, processed and analysed, and products to be created in the scope of the project that may include anonymised data. This especially includes explicit agreement of all participants on how to document, store and use sensitive data. The consent form will include information about the project, how the data will be used (only for research purposes), and a checklist indicating the types and/or intended uses of consented material including but not limited to (1) audio recording; (2) video recording; (3) photographs; (4) transcriptions and annotations; (5) translations. Anonymity will be ensured. Personal data which are collected in third countries (e.g., during interviews) and imported in the EU will be double-checked according to professional standards in the country of origin, and according to EU standards. Declarations of confirming compliance with laws in the country of origin, and declarations of compliance with EU regulations will be provided.

This applies to: (a) data retrieval: consent forms and information sheets will be provided, (b) data processing (oral biographies / stories): Interview partners will sign written agreements in which they indicate which data they allow to use, and under which conditions (e.g., modification of data to avoid any risk of correlation) the data may be re-used, (c) data storage: Interview partners will give their consent to location and (restricted) accessibility of data, and to potential further re-use of such data.

<u>Protection of Personal Data</u>: The collection and use of individual data & more generally private information will be reduced to a minimum on a "need to use basis" maintaining data confidentiality. All participating partners need to ensure legal requirements related to the use of private data according to EU rules and standards. It will be made sure that also possible subcontractors will adhere to these rules. For the planned socio-economic surveys data like zip-code and function (e.g., owner of a business, consumer etc.) may be required but it will be made sure that no answers can be traced back to single participants or single groups of participants. If the survey is performed online it will be made sure that the encrypted connection via HTTPS is applied, data will be stored in institute computer and the UPSCALE database instead of cloud storage offered by third party, IP-addresses of the participants are also be treated according to national and European law. Storage time of anonymized data is not limited (no longer than the specified maximum conservation time), but again according to prevalent laws, the informed consent and the time needed to reach the goals for this project.

Furthermore, the people involved will handle the data accurately, securely, confidentially, anonymously, and ensure that data are processed in accordance with the data subject's rights. All information will be stored in a confidential manner and in accordance with the EU Directive 95/46/EC and GDPR regarding use of personal data. It will be ensured that the following requirements are met:



(a) Confirmation that data will be collected on a need to know basis only (b) The guarantee of withdrawal rights and oblivion rights as made compulsory by the European court of justice in 2014 (c) Avoidance of merging data sets in order to prevent any unforeseen personal information disclosure
(d) Provision of detailed information on the procedures that will be implemented for data collection, processing, storage, protection, retention and destruction and confirmation that they comply with national and EU legislation.

Non-EU participants in UPSCALE include partners from Switzerland, Serbia, Ethiopia, South Africa, Kenya, Rwanda, Tanzania and Uganda. Regarding the ethical issues in UPSCALE, especially data protection, it will be made sure by the principal investigators and the Data Protection Officer (see Deliverable D10.2) of each beneficiary organization that all measures undertaken in the study countries (i.e. where data is generated: Ethiopia, Kenya, Uganda, Rwanda, Tanzania) comply with ethics principles and relevant national, EU and international legislation. For all personal data (e.g. transcripts of interviews), the project, more specifically the Principal Investigators and the Data Protection Officers of the corresponding beneficiary organizations will provide a declaration of confirming compliance with Chapter V of the EU General Data Protection Regulation (GDPR), as well as a declaration of confirming compliance with the laws of the country where the data was collected.

Personal data will be anonymized before it can be uploaded to the database. All the data stored in the database will be exchanged among partners regardless of EU or non-EU participants following the request-review-grant-procedure and agreement to the data publishing and sharing policy (Appendix 1).

<u>Geographic coordinates</u>: Geographic coordinates of the study sites as well as their landscape and climate characteristics will be stored as one dataset in UPSCALE database. The identifier of study sites (the primary key of this dataset) will be called Plot ID. Other datasets which uses the study sites will use Plot ID as reference but not including the actual coordinates. When we migrate our data to open access data archives, the Plot IDs will be made publically accessible while the study sites dataset will apply another license to make sure the sensitivity of the geographical locations (e.g. coordinates of a farm) are protected.



Appendix 1: Data publishing and sharing policy

Within the UPSCALE project, many publications will arise within and across subprojects. In addition, the UPSCALE project offers exciting opportunities for many collaborative results and publications. This offers the advantages of synergy and increased numbers of publications. The UPSCALE project specifically requires agreements on the documentation, quality, sharing, and publication of data. These agreements are the subject of this policy.

A. DATA POLICY OF THE UPSCALE PROJECT

§1 Coverage and definitions

a. This policy applies to all members and staff

b. "Project data" in the sense of this agreement are conventional measurements (data from measuring instruments, field surveys, generally in tabular form), images (digital photos, satellite data, maps, etc.), GIS-Data (digital maps and modelling data), genetic data as well as resulting (digital) publications, thesis, proceedings and posters.

c. "Meta data" shall mean any data describing and documenting the project data.

§2 Data Management and Publication Committee

a. The UPSCALE project forms a Data Management and Publication Committee, in which every PI has one vote. The committee sets the data standards and also serves to adjudicate possible disputes relating to this policy. In case of disputes, a majority of votes is necessary to make a decision. The committee is headed by the PI of the data base project.

§3 Rights and responsibilities

a. Project members (PIs, Postdocs, PhD students, other staff) have a right of access to project data according to the following provisions. The right of access to project data is limited to scientific use.

b. Furthermore, project members have a right, that their intellectual input and their academic interest is respected with regard to the project data in accordance to §5.

c. Each member and staff agrees to have her/his data with metadata stored in the central database of the UPSCALE project at the University of Würzburg.

d. The data base manager and the respective PI are responsible for the management of this database.

e. In cases of a serious violation of obligations under this agreement the Data Management and Publication Committee of the UPSCALE project may impose adequate sanctions.



§4 Documentation of datasets

All datasets have to be documented with meta-information. All data sets will be stored with information on who collected them, version of the data and date of the latest update.

a. Details on deadlines and data formats will be given by the database project.

§5 Access to data in the UPSCALE project database

All members and staff have free access to the meta-information.

a. After request and approval by the respective data owner, each member or scientific staff will get access to the original data in due time. The scientists originally obtaining the data and the PI of their project will be informed when data are used in other subprojects or work packages and give their consent prior to data publication. Of course, scientists making use of data supplied by other scientists for scientific publication must acknowledge the use of the data appropriately (see below).

b. Release of data to non-members of the UPSCALE project will be an exception for which explicit permission has to be sought of the Data Management and Publication Committee.

c. All data will be made publicly available taking into account privacy issues and in accordance with standards set by DFG (https://www.dfg.de/en/research_funding/proposal_review_decision/applicants/research_data/inde x.html).

§6 Use of data collected by UPSCALE project

a. Data use must always be based on an agreement between the original data supplier and the data user. Original data suppliers are the scientists originally obtaining the data and the PIs of the respective projects. The latter are responsible for supplying the data obtained by all staff of their project to the database.

b. Data accessed by a scientist must only be used for purposes necessary to carry out his/her own work in UPSCALE project. Data accessed must only be used for scientific purposes, i.e., commercial use of data is not allowed. It is prohibited to distribute other scientist's data to a third party without the written consent of the scientist.

§7 Delivery of data and quality control

a. Data need to be deposited as fast as possible, normally within half a year, and at the latest one year after the field sampling or laboratory analysis has been completed.

Data quality is controlled by careful review of the different components saved to the database.



B. PUBLICATION POLICY OF THE UPSCALE project

The success of the UPSCALE project depends on successful publication in appropriate journals. Many collaborative papers will have several authors. The people most important in the derivation of a particular result are acknowledged by first authorship (usually the PhD students or Postdoc). To avoid uncertainties about access to data or authorship, UPSCALE project agrees on the following publication policy.

§1 General publication rules

a. The UPSCALE project adheres to the DFG guidelines on good scientific practice (https://www.dfg.de/en/research_funding/principles_dfg_funding/good_scientific_practice/)

b. The planning of an experiment or a field campaign, the planning and execution of data collection, the analysis of the data, and the writing of a manuscript all are necessary steps that can merit authorship.

c. Of course, usually the person who originally obtained the data (often PhD students) will publish the results of his/her disciplinary study. To also promote swift publication of collaborative papers in appropriate journals, the following procedure is adopted. Before writing a collaborative paper, credit to the data suppliers is discussed. Often, this credit will consist of co-authorship. Because collaborative synthesis does complement rather than forestall disciplinary publications, the access to data needed for collaborative synthesis should not be blocked by the data supplier.

d. Members and staff planning a publication distribute a tentative title of the paper, a preliminary list of authors and the anticipated journal to the internal UPSCALE project mailing list as early as possible, latest one month before the manuscript is finished. This serves for rapid information on ongoing activities and allows improvements of synthesis work at an early stage.

Members and staff planning to submit a paper to a journal distribute the final version to the internal UPSCALE project mailing list as early as possible, latest 2 weeks before submission. This serves for rapid information on ongoing activities and allows final improvements of synthesis work.

§2 Acknowledgement of original data suppliers and co-authorship

a. In the case of providing unpublished data to a paper, the credit of the original data supplier can be in form of co-authorship, mentioning in the acknowledgements or in the figure or table legend or as "personal communication". Co-authorship is appropriate if the data contribute important information to the main story of the publication. The other options are appropriate if the data are only used as additional side information. The original data suppliers should decide what form they find appropriate.

b. PhD students and PI's are encouraged to co-author where appropriate. For example, if a PhD project receives intellectual input into the design from a PhD student or a PI from another project, then co-authorship may well be appropriate.

