WASCAL DATA SHARING POLICY AND GUIDELINE 2.0

COVER PAGE

West African Science Service Center on Climate Change and Adapted Land Use
WASCAL, Headquarters, CSIR Office Complex, Agostino Road, Airport Residential Area, PMB CT 504, Cantonments-Accra, Ghana
www.wascal.org | Tel: +233 0501 529 000 | +233 302 777 137 | 0501 528 999

Content

Definition of terms and abbreviations

CoC	WASCAL Competence Center
Copyright	"Copyright (or author's right) is a legal term used to describe the rights that creators have over their literary and artistic works. Works covered by copyright range from books, music, paintings, sculpture, and films, to computer programs, databases, advertisements, maps, and technical drawings"[1]. "In general, a work is copyrighted when it is created, and it is not necessary to apply for copyright. In any case, when a work is copyrighted, others may not use or redistribute the work without the permission of the author"[2].
Data	Distinct units of information such as facts, numbers, letters, symbols, usually formatted in a specific way, stored on a computer device and suitable for processing by a computer
Data management	Development and implementation of architectures, policies, practices and procedures to properly manage the full data lifecycle needs of an institutional unit or research project
Data provider	Individual or institution, which distributes data to other parties
Data rights management	Regulations of access to data by consideration of data ownership, intellectual property rights and other legal restrictions
Data user	Individual or institution authorized to access and use data
DMD	Data Management Department
DMC	Data Management Council to discuss general strategic issues with regard to data management and pending decisions of major relevance like data openness and licensing, data rights management . It consists of scientists from WASCAL SAC, Competence Centre, and representatives from partner institutions named by WASCAL. The scientific staff of DMD are by default members of this committee and are in charge of implementing its recommandations. .
DMP	Data Management Plan: formal document that outlines how project data are managed during research time and after project completion. The goal of a data management plan is to consider the many aspects of data management, metadata generation, data preservation and analysis before the project begins. This ensures that data are well-managed at present, and prepared for preservation in the future

[1] From WIPO: http://www.wipo.int/copyright/en/
[2] From http://www.iusmentis.com/ip/

Funding institution	Institution funding projects and programs in WASCAL
Institutional data manager	Responsible Person for the operational management and processing of data. Data managers are nominated for each institution contributing data to WADIREP (see WASCAL DMP)
IP	Intellectual property (IP) refers to creations of the mind, such as inventions; literary and artistic works; designs; and symbols, names and images used in commerce. IP is protected in law by either copyright, patent, or trademarks, which enable people to earn recognition or financial benefit from what they invent or create. (Definition from WIPO[3]). For more details, see what the notions of Copyright and Patent stand for in this document.
License	A data owner or copyright holder can grant licenses for the use of his/ her data. A license is a usage agreement (= terms of use) stating the scope of the granted use rights
Metadata	Data providing descriptive information on data to enable users to query valuable information to decide about the usability of the data for its own use and purposes
MoU	"A memorandum of understanding (MoU) describes a bilateral or multilateral agreement between two or more parties. It expresses a convergence of will between the parties, indicating an intended common line of action"[4]. "Companies and organizations can use MOUs to establish official partnerships. MOUs are not legally binding but they carry a degree of seriousness and mutual respect, stronger than a gentlemen's agreement"[5].
Partner institution	Organization in cooperation with WASCAL
Staff member	Person, who holds a work contract with a WASCAL institution
Patent	"A patent is an exclusive right granted for an invention, which is a product or a process that provides, in general, a new way of doing something, or offers a new technical solution to a problem"[6]. "In order to get this right, the inventor must apply for a patent at his patent office"[7]. "If no measures are taken to protect an invention, an inventor may find themselves the object of a patent enforcement action by a subsequent inventor. This could result in having to prove prior invention, or paying royalties"[8].

[3] From WIPO : http://www.wipo.int/about-ip/en/
[4] https://en.wikipedia.org/wiki/Memorandum_of_understanding
[5] http://whatis.techtarget.com/definition/memorandum-of-understanding-MOU-or-MoU
[6] From WIPO: http://www.wipo.int/patents/en/
[7] From http://www.iusmentis.com/ip/
[8] From http://www.defensivepublications.org/defensive-pubs-faqs

WADIREP	WASCAL Data Infrastructure Repositories and Portals. It refers to the set of repositories, portals and applications continuously developed and maintained by the DMD to provide online open accessibility to the data alongside long-term storage, management, licensing, sharing and preservation of the data.
WADI-Sci-DataP	WASCAL Scientific Research Data Catalog. It is one of the repositories and portals composing WADIREP. It stores and publishes online metadata and data files on research activities carried on by WASCAL-related scientists or from research/surveys/publications made by partner institutions. It provides services to search, access and create metadata, and to upload data files
WASCAL-HydroMet	WASCAL Hydrological and Meteorological Data Portal. It is an online data repository and portal providing open access to ground hydrological and meteorological data mainly coll3ected from WASCAL hydro-meteorlogical observation networks. It is another repository and portal composing WADIREP.
WASCAL-related scientist	Any person working in the frame of scientific projects / activities (research, services development, capacity building) funded, or led by WASCAL. This does not include only researchers, visiting scientists, postdocs, research assistants, and interns working at the competence centre and the headquarters,, but also researchers, supervisors and students from WASCAL GSPs and any other scientific person collaborating with WASCAL in projects funded or led by WASCAL.

1 Preamble

Sharing research data strengthens open scientific culture and collaboration, fosters diversity of research analyses, inspires new research horizons, helps envisage new or alternative hypotheses and methods, allows the exploration of new research topics not examined by the initial researchers, eases the education of new researchers and professionals, and can lead to better decision support by leveraging (interrelating and aggregating) data from various and multiple datasets.

In pursuit of the objective of strengthening research through data sharing, all data generated as a result of funding by WASCAL or project managed by WASCAL are to be considered for sharing and made as widely and freely accessible as possible (open access) when applicable, whilst safeguarding intellectual property, the privacy of participants and confidential data.

The following Data Sharing Policy (DSP) is a WASCAL DSP stating clearly under which rules data should be used and shared by WASCAL-related scientists.

Therefore, WASCAL staff members and any other WASCAL-related scientists shall consent to the hereby Data Sharing Policy, if not otherwise stated within bilateral agreements with individual researchers or institutional partners in the projects and programs funded or led by WASCAL.

Likewise, other persons and data providers interested in WASCAL scientific activities (research, services development, capacity building) are encouraged to acquaint with it.

2 Scope and Objectives

The WASCAL[9] Data Sharing Policy (DSP) addresses all issues related to the provision, exchange, availability, maintenance and use of data produced during WASCAL-related scientific activities (research, services development, capacity building) or acquired from third parties collaborating with WASCAL. It contains rules regarding the provision of data, the management of data rights, its storage within the WASCAL Data Infrastructure Repositories and Portals (WADIREP), and its publication[10] over the internet.

To guarantee efficiency and continuity of the research and education activities within WASCAL, it is essential that data collected from WASCAL observation networks or produced by WASCAL-related scientists are made available as soon as possible to others, including external partners or members of the worldwide scientific community to support research, capacity building and services development.

Therefore, data created, collected and produced in the framework of WASCAL-related scientific projects or activities (fully or partially funded by WASCAL, BMBF or other funders) and being regarded as valuable for verifying scientific findings, re-use in further research and integration in policy advice statements, will have to be stored in WADIREP and, possibly in connected data infrastructures as far as authorized by WASCAL and respectively agreed upon within bilateral agreements.

The objective of this data sharing policy is to define and provide a set of principles, rules and guidelines to stir up an ethical and responsible ecosystem for the enhanced sharing and access to the sharable scientific data produced in the frame of WASCAL activities.

3 Type and nature of covered data

The WASCAL Data Sharing Policy relates to all data collected under the implementation of WASCAL research, survey, service-delivery, and capacity building activities that are either fully or partially funded by WASCAL, BMBF and other project executed by WASCAL-related scientists. It also applies to environmental/climate data managed by WASCAL, where WASCAL was not the original funder and there is no other policy in place. For non-WASCAL data where there is an alternative data policy in place, then this will be applied.

3.1 Types of data

The following main categories of data types are covered:

a) **Time-series data:** include hydrological and meteorological daily and sub-daily measurements of a dozen of variables (e.g. temperature, precipitation, solar radiation, wind, pressure) captured from WASCAL mesoscale (watersheds in Dano, Dassari, Vea/Sissili, projects sites), and transboundary (in all member states) Observation Networks, WRAP 2.0 projects, etc.

[9] www.wascal.org

[10] This is currently done via the WADI Central Metadata Catalogue (WADI-CMC) with the WADI Data Portal (WADI-DP) as web-front-end for data access and the WASCAL-Geoportal (WASCAL-GP) as web-front end for metadata editing.

b) **Agricultural, fishering and livestock data:** include data on soil, vegetation, water resources, farming, breeding, and land use in West Africa

c) **Climate data:** include analytical data of weather and climate observations, climate models, projections, and simulations outputs; surveys/studies on climate extreme events and other phenomena in the region

d) **GIS and Remote Sensing data:** include topographical data, administrative boundaries, remote sensing data and images, and various maps

e) **Socio-Economic data and gender:** include surveys and studies on socio-economic aspects (crop yield, migration, demography, gender) related to climate change

f) **Biodiversity data:** include data on protected areas, species, ecosystem services, surveys and studies.

g) **Health data:** include data on Climate Sensitive Diseases and Disorders (e.g. Covid-19, Malaria, Miniagencies, Measles etc).

3.2 Nature of the data

The nature of the aforementioned data may be of the following:

a) **Primary Data** (**mostly not falling under Intellectual Property - IP**) collected directly by WASCAL-related scientists:

   i. **Raw data** collected with the use of WASCAL instruments or any other collection method and yet to be processed.

   ii. **Processed data** are raw data that have undergone certain transformation procedures.

   iii. **Quality controlled data** are processed data that has passed quality assurance procedures such as routine estimation of timing and sensor calibration or visual inspection and removal of obvious errors (exemple of data).

b) **Secondary Data**

   i. **Derived data** (**possibly Copyrighted**) are data generated by derivations from Primary Data (e.g. by aggregation, integration, recalculation, modelling) by using software tools.

   ii. **Data products** (**probably Copyrighted, possibly patentable**) generated from Primary Data or derived data by means of scientific and technical interpretation. Data products (such as drawings, maps) usually cover the major part of the data stock used within publications.

c) **Co-owned Data** refers to data collected or generated in the frame of activities conducted in collaboration with other organizations or co-funded by WASCAL. Issues related to **proprietary data** also can arise when co-funding is provided by the private sector (owner of the data if no document of collaboration?).

d) WASCAL funded Data is defined as where WASCAL has been a full or a partial funder of the activities that create/generate data. Normally, where WASCAL is a partial funder of an activity, it will be as part of a joint funding programme with other

organisations/institutions (need to have an MoU). In this case, the funding partners will agree through a signed MoU or data policy, what specific data management conditions will apply, for example, the length of the 'embargo' period. However, all specific conditions will still meet the requirements of WASCAL's policy.

e) **External/third-party Data** is data (primary as well as secondary) obtained from third parties for use in WASCAL-related research:

   i. Historical and real-time **measurement data** from the observation networks or data from any other source owned by WASCAL partner institutions (primary data);

   ii. Data from **external sources** not falling under any co-operation with WASCAL at the time of obtaining (primary / secondary data).

4 Intellectual Property (IP) and Copyright

Sharing and re-use of data is subject to legal regulations. Data rights management is of crucial importance in the data management framework, and of high priority in any data management infrastructure. This includes the protection of the rights of the data providers, their intellectual property and the personal privacy of data providers for all types of data.

Certain data may be intellectual property of the data author/creator and therefore protected by copyright. Though copyright laws are not harmonized at the international level,[11] most are linked to requirements of originality (an intellectual creation) and individuality (i.e. having a personal feature). In case data is subject to copyright, or other forms of ownership rights, WASCAL requires a license from the data provider to re-use such data. For reasons of legal certainty WASCAL uses standardized license schemes such as Creative Commons[12] or Open Data Commons[13]. Under specific conditions[14] more restrictive, individually tailored licenses ("bespoke licensing") can be assigned to the data, as far as approved by the DMC.

5 Provision of Data to WADIREP

The data providers contribute to the provision, sharing and exchange of data within and outside the WASCAL scientific community in order to generate data products of good quality for WASCAL stakeholders.

The WASCAL Data Management Council (DMC) has the option to discuss data provision issues with the WASCAL staff members and individual partners. Differences will be settled in a friendly, mutually understanding manner to the benefit of all stakeholders.

[11] For example: According to the "German Copyright Act", the results of scientific work, such as written works, speeches and computer programs, illustrations of a scientific or technical nature such as drawings, plans, maps, sketches, tables and three-dimensional representations may be subject to copyrights

[12] http://creativecommons.org/

[13] http://opendatacommons.org/

14 E.g. closed data sharing networks, use of confidential data

Detailed instructions on how to collect and provide data to WASCAL are outlined in the WASCAL Data Management Plan[15] or the WASCAL Geoportal User Guide[16] or by using the data management plan provided in the appendix of this document for students under the WASCAL Graduate Studies Program (GSP).

5.1. WASCAL staff members

Data created by WASCAL staff members* (give a definition is staff in the schools, stff at competence Centre, staff at Accra etc..) is owned by WASCAL, if not otherwise stated by staff funding institutions. As part of the employment procedure and documentation, WASCAL staff members must confirm the adherence to the present DSP of WASCAL by signing. This agreement includes use licenses for the data to be provided to WASCAL by the staff member.

Based on the present DSP the following rules for data provision apply:

a) **Metadata upload:** A description of all data, data products, documents, models, software code, which is of interest for in- and outside WASCAL[17], must be created by the data providers, or responsible individuals who have been assigned by the data providers, and published by means of the WADIREP/WADI-Sci-DataP within three (3) months after its generation. Data quality assessments are essential within data documentations[18];

b) **Data upload**: If data publication priority rights (section 4) do not come into effect, the data must be uploaded to WADIREP/WADI-Sci-DataP within three (3) months after its generation and by following the standards as outlined in the WASCAL DMP.

c) **Copyright**: Staff members, who are employed by WASCAL, or by one of the institutions contributing to WASCAL, hold all copyrights to the data they produce, provided the requirements of the relevant Copyright Act are met (section 2). Staff members are free in their decision to share his/ her data, with third parties as they deem necessary.

d) **Patent**: WASCAL acknowledges the need to protect patentable and other proprietary data. Any restrictions on data sharing due to co-funding arrangements or patentable work should be discussed and highlighted in a data-sharing plan submitted to the DMC. This document should provide the DMC with details regarding the arrangements related to co-funding (ex. MoUs, DSPs from other funding organizations) and how the data are planned to be shared. In the case of patentable work, the plan should highlight the chance of the researchers to get their invention patented and provide a reasonable timeframe-plan regarding the patenting process. Whilst there may be a delay in the release of data due to the application process, appropriate intellectual property protection should not hinder data sharing. Therefore, an embargo on the data should exceed no more than 24 months. The requirement of patentable data shoul be known by the data provider such as publication in order to avoid any conflict.

e) **Licenses**: For all data falling under intellectual property and subject to copyright law, staff members employed by WASCAL, or by any institution contributing to WASCAL, and

[15] https://wascal-dataportal.org/geonetwork/?uuid=fcc6b636-c0f8-4370-9e8c-074a2c138367
[16] https://wascal-dataportal.org/geonetwork/?uuid=e8bc7312-45ab-4c9d-9a73-1874f69479c4
[17] Under specific conditions primary data may be stored at the researcher locally (see DMP, section 2.8)
[18] See the DMP for further details

referring to the present DSP[19], grant WASCAL the following irrevocable Creative Commons license for open data:

Attribution-Non Commercial-Share Alike 4.0 International (CoC BY-NC-SA)[20].

With this license WASCAL and any other user is free to

- **Share** — copy and redistribute the material in any medium or format.
- **Adapt** — remix, transform, and build upon the material.

Under the following terms

- **Attribution** — WASCAL as well as any other user gives appropriate credit, provides a link to the license, and indicates if changes were made;
- **Non-Commercial** — WASCAL as well as any other user may not use the material for commercial purposes;
- **Share Alike** — If WASCAL or any other user remixes, transforms, or builds upon the material, the contributions must be distributed under the same license as the original.

For all data not falling under intellectual property, the staff members, who are employed by WASCAL, or by one of the institutions contributing to WASCAL and referring to the present DSP, grant WASCAL the following irrevocable Open Data Commons License

Open Database License (ODbL) v1.0[21]

With this license WASCAL and any other user is free to

- **Share** — To copy, distribute and use the database
- **Create** — To produce works from the database
- **Adapt** — To modify, transform and build upon the database.

Under the following terms

- **Attribution** — WASCAL as well as any other user must attribute any public use of the database, or works produced from the database, in the manner specified in the ODbL. For any use or redistribution of the database, or works produced from it, WASCAL or any other user must make clear to others the license of the database and keep intact any notices on the original database
- **Share-Alike** — If WASCAL as well as any other user publicly use any adapted version of this database, or works produced from an adapted database, the adapted database must be licensed under the ODbL
- **Keep open** — If WASCAL or any other user redistribute the database, or an adapted version of it, then technological measures may be used that restrict the work (such as DRM) as long as a version without such measures is redistributed.

[19] Reference must be included in the MoU between WASCAL and the project partner institution
[20] http://creativecommons.org/licenses/by-nc-sa/4.0/. Formulation of terms of use adapted by the authors
[21] http://opendatacommons.org/licenses/odbl/summary/. Formulation of terms of use adapted by the authors

f) **Data Publication priority right**: Staff members, who are employed by WASCAL, or by one of the institutions contributing to WASCAL and referring to the present DSP[22], have priority rights to publish his/her data first in case the data is his/her intellectual property. For data not subject to intellectual property (e.g. biophysical data from own observations) first publication rights may be set upon request and approval by the DMC. If the priority right to publish is not used, then after an adequate time span of twenty-four (24) months WASCAL as well as the funding institution have the exclusive right to use the data under the licenses mentioned in section 5.1 e or similar terms of use. The DMC may decide on an extension of this time span (i.e. delay to first publication) upon request. Publication purposes have to be announced to the DMC within three (3) months after the generation of the data and do not exempt from creating metadata following 5.1 a. In the metadata the estimated date of accessibility to the data has to be declared.

g) **Secondary data**: If WASCAL staff members, who are employed by WASCAL, or by one of the institutions contributing to WASCAL and referring to the present DSP, want to store secondary data from external sources in WADIREP and make these accessible within the WADI-Sci-DataP, the corresponding copyrights must be respected e.g. by requesting a license statement to the data source, which has to be added to the metadata in the WADI-Sci-DataP;

h) **End of contract Staff:** the staff members, who are leaving WASCAL (e.g. end of contract, retirement, resignation, dismissal, etc.) or one of the institutions contributing to WASCAL and referring to the present DSP should upload to the WADIREP/WADI-Sci-DataP all the data belonging to WASCAL, or hand them over to the Data Management Unit before their departure date. After leaving WASCAL, a former WASCAL-related staff is not anymore allowed to use restricted WASCAL data without any prior authorization from WASCAL.

i) **Long-term data preservation**: Once data sets (metadata+data files/records) are created, and validated on WADIREP/WADI-Sci-DataP, a web link with a persistent URI (Universal Resource Identifier) is created to ensure the uniqueness accessibility to that specific data sets. This persistent identifier will last as along as the data portal is maintained and accessible. The whole WASCAL Data Infrastructure is designed and equipped with up-to-date technologies, servers and software to ensure the long-term preservation of the data and services (e.g. apps) over 25 years.

5.2. WASCAL-related scientists at partner institutions

Partner institutions are organizations in (a formal) collaboration with WASCAL. Where applicable, a focal person at the partner institution, preferably responsible for data management shall be nominated to act as a counterpart of the Data Management at the Competence Centre.

5.2.1. Regular research staff

WASCAL-related scientists employed by partner institutions can upload and share data through WADIREP.

To ensure that the objectives of WASCAL are reached in such cases, WASCAL partners will set up bilateral agreements (e.g. MoUs) with WASCAL describing rights and commitments on both sides

[22] Reference must be included in the MoU between WASCAL and the project partner institution

regarding the provision, distribution, and re-use of data owned by the partners, but hosted and managed in WADIREP. WASCAL affirms its strict adherence to these agreements. WASCAL partners have the option to refer to parts or the whole of the present DSP.

5.2.2. Students in the Graduate/Master Studies Programs

Data obtained by students funded by WASCAL are co-owned by WASCAL, if not otherwise stated within MoUs by co-funding institutions or the partner university (place of Lead university in case?). Students and their supervisors must confirm the adherence to the present DSP. They have to follow the regulations outlined in the present DSP, with the following additions and exceptions:

a) **Immediate data preservation**: to ensure zero data loss by students, all data, data products, documents, models, software code, which is of interest for in- and outside WASCAL, must be immediately stored locally on the student laptop under a professional OneDrive data folder that will be provided and setup for them by WASCAL Data Management Department.
b) Section 5.1 b) [Data Upload] is replaced by The data generated and used by students within the context of their study carried in collaboration with WASCAL, must be uploaded to the WADIREP/WADI-Sci-DataP not later than four (4) months after the completion of the study. In case of doctoral students (GRP) before the defence takes place more explanation. To avoid last minute rush, students must progressively upload their data under-processing with different versions reflecting their advancement.

Section 5.1. f) is replaced by

c) Students have the right to publish his/ her data first, e.g. as part of his/ her Ph.D. or any other scientific publication. Except patentable data until the patente obtention. If this priority right is not used within twelve (12) months after the end of the scholarship contract, this right is passed over to the university and to WASCAL by granting a license following section 5.1. e).

d) Students must be aware and follow the student data provision milestones as summarised in the flowchart below (Figure 1).

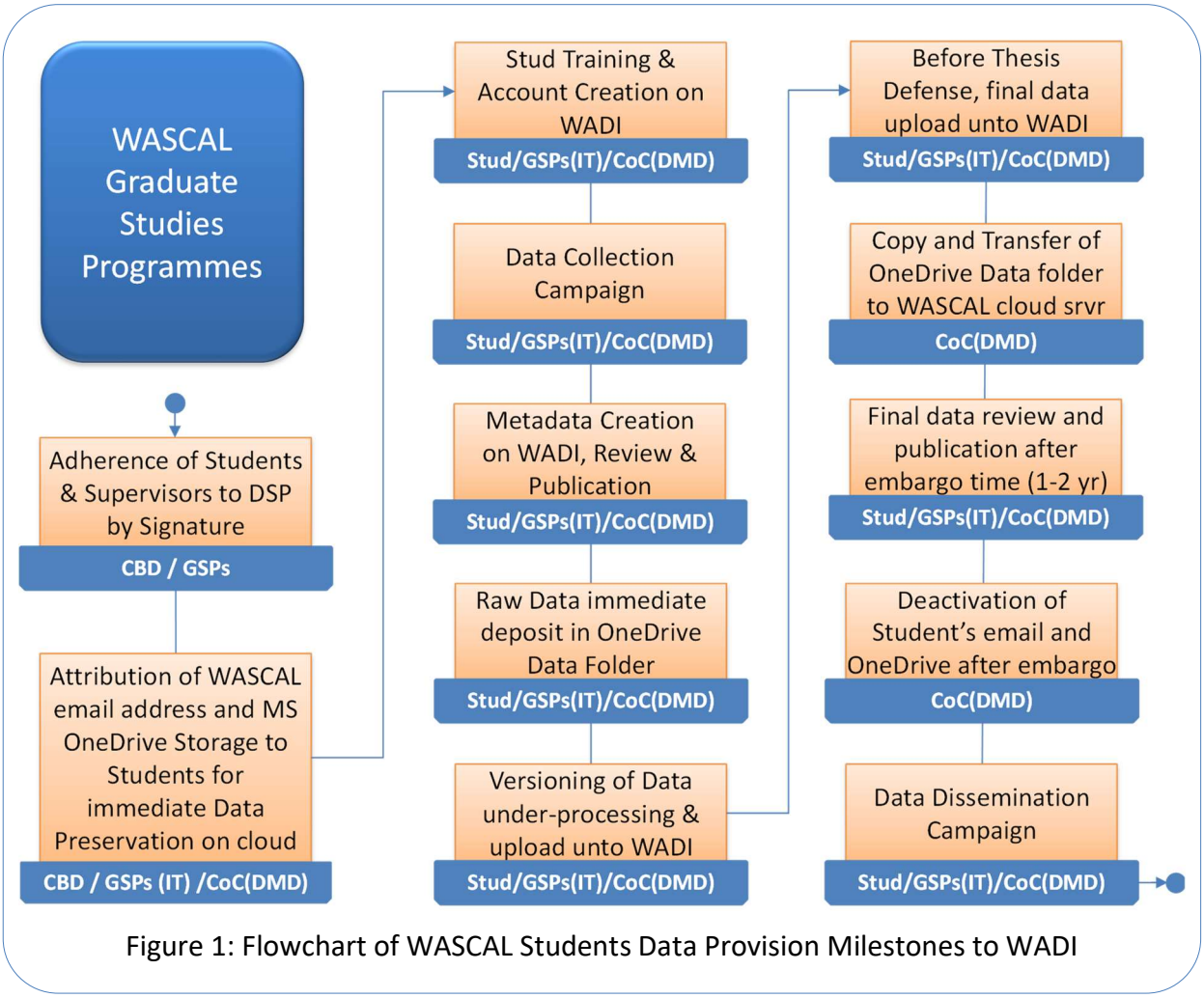

Figure 1: Flowchart of WASCAL Students Data Provision Milestones to WADI

For each step/milestone depicted in above Figure 1, the step name is provided along with the entities in charge of achieving it. For the sake of visibility of the overall flowchart, the following abbreviations are used:

- CBD: Capacity Building Directorate
- GSP: Graduate Studies Programme
- GSP (IT): IT unit/officer of the Graduate Studies Programme
- CoC: Competence Centre
- DMD: Data Management Department
- Stud: Students

7.Data Sharing

7.1. Access to Data

The entry point to WASCAL research data is provided by the WADIREP/WADI-Sci-DataP, which provides a metadata catalogue application and other data query and data access tools. Partner institutions (e.g. institutions responsible for individual observatories) may set up additional data

portals, as long as they make the data services accessible through the WADIREP/WADI-Sci-DataP (see DMP for further details).

The following rules for accessing WASCAL data apply:

a) **Registration of users accessing open data**: Data and metadata hosted by WADIREP will be made accessible via the WADIREP/WADI-Sci-DataP for scientific purposes as soon as quality control of the data has been completed (see DMP for further details). A verified[23] user registration for data download is generally not required, however the user must provide name, email address, institutional background, and the data usage purpose during the download process. Acceptance of use restrictions, in form of a license to be agreed upon by the user, is compulsory before data can be downloaded;

b) **Registration of users accessing restricted data**: Access to restricted data according to section 5.1 c) and requires a registration. A user identified as a member of a closed data sharing network needs a verified registration in the WADIREP/WADI-Sci-DataP to access the data. Quotation and use restrictions will be attached to the data as a license, and needs to be accepted by the user before downloading the data is enabled. When the restriction time period ends, the data will be made available to the public according to section 4.1 a);

c) **Closed data sharing networks**: Data providers may temporarily restrict access to data of the WASCAL community in general or specific groups in and outside of WASCAL ("closed data sharing network") for legitimate reasons[24]. These restrictions and their durations have to be communicated to the DMC and the data managers in charge;

d) **User accounts**: User accounts and group memberships, which grant access to datasets, will be administered by the data managers in the CoC, in collaboration with the members of the DMC. Information on how to get permission to download restricted data is included in the metadata (element "ordering instructions") of the requested data;

e) **Co-owned data**: For co-owned data, WASCAL researchers should negotiate and sign MoUs giving them the right to share their partners' data under conditions as close to open access licences as possible.

f) **External data**: For data provided by external partners (e.g. Met- Hydro-Services, etc.) access restrictions and data use licenses will be applied as agreed upon within bilateral agreements (MoUs).

7.2. Data Citation rules

If data are obtained from colleagues within WASCAL, these other colleagues must be cited properly according to the possible cases:

a) If the data have been produced in actual field work by one or more persons within WASCAL, these persons should be acknowledged, cited or even become co-authors;

[23] Verification means, that the user is identified by the data managers before they create an account. Users without account in the WADI-DP have to submit their contact details during the download process, but the authenticity cannot be verified.
[24] E.g. as part of the MoU

b) If considerable efforts of two or more persons go into one data set, joint publishing (i.e. the naming of both as data sources/authors in a publication) is recommended and appropriate;

c) When gathering or preprocessing primary data was substantially supported by others (e.g. student assistants), they should be acknowledged;

d) When a WASCAL staff member intends to publish data, research results and information, especially in an international context (journal, book, conference, and so on), it is mandatory that reference be made to the institutions that funded the research. For example, the acknowledgement in a publication may include the following statement:

"This research was conducted in the framework of WASCAL (PN: 01LG1202G) funded by the German Federal Ministry for Education and Research (BMBF)"

e) If data downloaded from WADIREP is used in a publication or in any other presentation to the public, the data source must be cited by following the instructions in the respective metadata record in WADIREP (metadata element: "Use constraints"). The syntax is:

<*name of authors*> <(*year of publication of the dataset in the CMC*)>: <*title of the dataset*> <URL: *host/unique ID*>

More details may be provided by the data use license or the DMP.

8. Data Ethics and Privacy

"Data ethics is about responsible and sustainable use/sharing of data. It refers and adheres to the principles and values on which human rights and personal data protection laws are based."[25]

The Data Ethics guideline provides a framework to develop ethical guidance that suits any context, whatever the project's size or scope.

8.1. Compliance with Data Ethics

WASCAL-researchers must comply with the following ethical rules addressing data collection, processing, use and sharing. Therefore, any research proposal, project, activity, or service in which WASCAL is involved as lead, co-lead, consortium member, consultant or investigator through its researchers or funded students should follow a responsible approach. It should:

i. be respectful of human rights as laid down in the Universal Declaration of human rights
ii. be aware and sensible to gender issues
iii. try as much as possible to remain neutral and unbiased. Personal preconceptions or opinions should not interfere with the research data.
iv. acknowledge ideas and work from others and run away from plagiarism
v. be legal in the study area
vi. not expose the research subjects nor the investigators to clearly known dangerous or harmful activities. When an efficient protection cannot be ensured against these threats, the activity should be canceled and at least postponed.

[25] https://dataethics.eu/data-ethics-principles/

vii. not be harmful to local communities, gender, and environment
viii. be respectful as much as possible to local beliefs

8.2. Guidelines for sharing data among external partners:

Guidelines for sharing data among external partners should comply with section 7.1 and also be guided with the following ethical rules.

1. Ongoing collaboration and mutual accountability are necessary between data-sharing partners.
2. Build common contracting procedures, but treat every contract and dataset as unique
3. Develop ethical review procedures between partners.
4. Be mutually accountable for interpretive resources.
5. Identify potential risks of sharing data within sharing agreements
6. Treat trust as a networked phenomenon.

8.3. Institutional Administrators

There are various ways research administrators can support sharing of data in WASCAL:

- Encourage data sharing by ensuring that ethics forms include relevant questions.
- Encourage an institutional approach to data management planning by creating or supporting data management planning policies.
- Support institutional data archives to develop policies for managing sensitive data.
- providing advice to researchers about designing their research so the data can later be shared.
- adding a question to your ethics application form: "Is there any reason NOT to share the data from this project?"
- adding suggested wording to your Institution's participant information sheets and consent forms that ask for permission from participants to share data.
- ensuring that researchers intending to re-use data adhere to the original ethics approval conditions for which consent from the original participants was obtained.

8.4 Data Ethics Governance Framework:

Figure 2 and table 1 show the key topics of the guideline on Data Ethics Governance adapted from data4sdgs[26] to guide and frame WASCAL scientific and research projects and activities for an efficient consideration of questions around data ethics governance.

[26] https://www.data4sdgs.org/sites/default/files/file_uploads/AVF-%20Data%20Ethics%20Canvas-%20for%20publishing.docx

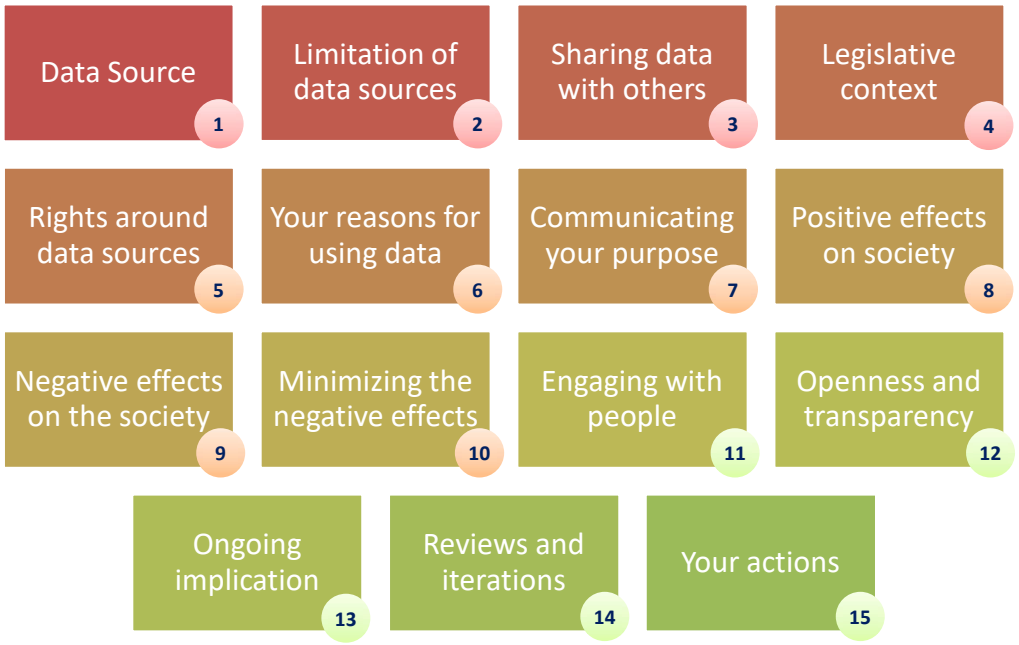

Figure 2: Data Ethics Governance Framework

Table 1: Objects of Data Ethics Governance Framework

Topic	Description
1. Data sources	Name/describe your project's key data sources, whether you're collecting data yourself or accessing via third parties. Is any personal data involved, or data that is otherwise sensitive?
2. Limitations in data sources	Are there limitations that could influence your project's outcomes? Consider: bias in data collection, inclusion/exclusion, analysis, algorithms gaps or omissions in data, provenance and data quality, other issues affecting decisions, such as team composition
3. Sharing data with others	Are you going to be sharing data with other organisations? If so, who? Are you planning to publish any of the data? Under what conditions?
4. Ethical and legislative context	What existing ethical codes apply to your sector or project? What legislation, policies, or other regulation shape how you use data? What requirements do they introduce? Consider: the rule of law; human rights; data protection; IP and database rights; antidiscrimination laws; and data sharing, policies, regulation and ethics codes/frameworks specific to sectors (eg health, employment, taxation).
5. Rights around data sources	Where did you get the data from? Is it produced by an organization or collected directly from individuals? Was the data collected for this project or for another purpose? Do you have permission to use this data, or another basis on which you're allowed to use it? What ongoing rights will the data source have?
6. Your reason for using data Engaging	What is your primary purpose for collecting and using data in this project? What are your main use cases? What is your business model? Are you making things better for society? How and for whom? Are you replacing another product or service as a result of this project?

7. Communicating your purpose	Which individuals, groups, demographics or organizations will be positively affected by this project? How? How are you measuring and communicating positive impact? How could you increase it?
8. Positive effects on people	Who could be negatively affected by this project? Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)? How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organizations using the data
9. Negative effects on people	Who could be negatively affected by this project? Could the way that data is collected, used or shared cause harm or expose individuals to risk of being re-identified? Could it be used to target, profile or prejudice people, or unfairly restrict access (eg exclusive arrangements)?How are limitations and risks communicated to people? Consider: people who the data is about, people impacted by its use and organizations using the data.
10. Avoid or minimizing negative impact	How can people engage with you about the project? How can people correct information, appeal or request changes to the product/service? To what extent? Are appeal mechanisms reasonable and well understood.
11. Engaging with people	How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements? Can you ask peers for feedback on the project? How will you communicate it internally? Will you publish your actions and answers to this canvas openly?
12. Openness and transparency	How open can you be about this project? Could you publish your methodology, metadata, datasets, code or impact measurements? Can you ask peers for feedback on the project? How will you communicate it internally? Will you publish your actions and answers to this canvas openly?
13. Ongoing implementation	Are you routinely building in thoughts, ideas and considerations of people affected in your project? How? What information or training might be needed to help people understand data issues? Are systems, processes and resources available for responding to data issues that arise in the long-term?
14. Reviews and iterations	How will ongoing data ethics issues be measured, monitored, discussed and actioned? How often will your responses to this canvas be reviewed or updated? When?
15. Your actions	What actions will you take before moving forward with this project? Which should take priority? Who will be responsible for these actions, and who must be involved? Will you openly publish your actions and answers to this canvas?

8.5. Protection of Data Confidentiality and Privacy

In this Data Sharing Policy, confidential data are of two groups: restricted data, and personal/private data. They are subject to particular protection

Restricted data are unauthorized to disclosure and are clearly stated as confidential by WASCAL executive management. This may include sensible or internal strategic and financial records, Intellectual property, Business-critical data.

Personal data, which are “information concerning the personal or material circumstances of an identified or identifiable natural person (“data subject”)”, include: social security number, phone numbers of friends/family/colleagues/students, driver's license numbers, bank account

numbers, tax information, passwords or passphrases, home address or phone numbers, employee id number, digital images, any personal electronic documents containing personal text, and alike.

In research, personal data containing privacy-sensitive information (e.g. household data with private/personal material or family details, which can be traced back to the individuals in question) can be used and processed only with the consent of the individuals supplying the data or information, with respect to local and international laws and policies and with a strong focus in meeting the General Data Protection Regulation (GDPR) standards.

Correspondingly, in publications personal data can be used only with the written consent of the persons concerned in the data ("data subject") or if allowed or required by law, e.g. if publishing the data is indispensable for the presentation of research findings on contemporary events. In this latter case, the WASCAL Management should deeply study the subject before WASCAL issuing any authorization.

Therefore, all scientists should be careful when dealing with personal data. In general, in addition to the compliance with GDPR, the following privacy rules will be applied as guidelines to safeguard data privacy:

a) Provided that the research purpose is not affected, the data should be anonymized after collecting them, as soon as possible by, e.g.
   i. removing direct identifiers, e.g. names, addresses or spatial localizations;
   ii. generalization of values, e.g. using ranges of household incomes;
   iii. splitting the personal data records in parts and storing them separately, so that an identification of natural persons is impeded. With an assignment function the data can be rejoined if required. This function may not be known to the people involved in the data processing, but only to the persons authorized to custody the sensitive data

b) Data privacy protection statements must be included in the metadata by the creator of data containing privacy-sensitive information, for example, "all names have been changed to protect research subjects' identities";

c) Moreover, no data with personal items is allowed to be stored in WADIREP. Only data relevant for re-use and sharing should be considered for upload to WADIREP.

9.0. Conclusion

Given that research requires input data as fuel and produces output data often seen as gold (e.g. by their owners), the hereby WASCAL Data Sharing Policy provides a strategic framework on data exchange that does not only encourages data sharing to stimulate research, but also fully recognizes researchers' intellectual property, organizations ownership, and reassures data providers.

Hence, as WASCAL is a regional education, service and research center that promotes research collaboration and data dissemination,, WASCAL-related scientists should strictly adhere to this Data Sharing Policy that is providing dynamic means to license their data and protect their Intellectual Property.

Appendix

Template for GSP Students' Project Data Management Plan

A Data management Plan (DMP) should describe the data from the students and what he/she will do with the data both during and after their research project. By doing so before and during data collection, it will prevent you from having problems in understanding your data and metadata in the future. It is also meant to ensure that your data will be preserved and useful both now and, in the future, by both you and other researchers.

1. Information about data

1.1 Data collection

Description of the data that will be acquired, methods of data acquisition to be specified, including the who, what, when, where of your collection.

Specify any existing data you may use, including their origins, and how new data to collected/generated will relate to those data for your project.

1.2 Data processing & files format

How will the data be processed once it is acquired?

Required data formats and data integration/organization. Hardware, software, algorithms to be used and their fit into the overall workflow of the project proposal. How do you plan to analyze your data and which software are you going to use?

1.3 Quality assurance & control

identify what data quality assurance and control measures that will be put in place during and after data collection, and during data analysis.

1.4 Short-term data versioning, preservation/backup

Describe how data will be managed in the short term?.

How will you keep track of different versions of your data and analyses?

How will you back up your data? Are servers available at your institution?

Consider both on-site and off-site backup options.

Describe how you will ensure security and no loss of your data, especially if your data are sensitive. Use WASCAL address and cod for data use.

2. Metadata format & content

Metadata is data documentation. It includes contextual details about data collection, e.g. temporal and spatial details, instruments used, parameters, units, etc. It should also mention whether a readme documentation will be accompanying the data files to describe abbreviations, formula, processing, algorithms, etc.

3. Policies for access, sharing & reuse

Describe policies for access, sharing, and reuse of your data. What obligations do you have to share your data?

Here, you can just mention that you will align with WASCAL Data sharing policy (DSP) with focus on sections dedicated to students.

Also, clearly specify any ethical or and privacy issues associated with your data. For example, if your data involves human subjects, endangered species, or locations of sensitive habitats

4. Long-term storage and data management

Describe what data will be preserved for long-term going from X to Y years. Not all data worth necessarily to be preserved. In general, any raw data should be kept. Also, any data products that were particularly expensive or time consuming to obtain should be preserved. Any data that is not easily replaceable should be kept. We need to keep also data already use for publication.

Ex: All raw data collected from the field as well as data products (e.g. maps, graphics or image used for GIS need) and output of simulations/models will be kept.

Also identify on which data platforms you will archive your data.

Kindly mention WASCAL Data platform here. You can also name other platforms where you intend to archive the data

Mention what data transformations and formats will be needed to comply with data platforms data standards.

Ex: All data in proprietary formats will be converted as much as possible to open data format to ensure their long-term usability.