Establishing a West African Science Service Center on Climate Change and Adapted Land Use (WASCAL)

Data Management Plan

This Project Communication and Data Management Plan has been produced by the WASCAL Data Management Executive Group at the outset of the WASCAL project, and has been approved by the WASCAL International Advisory Board. The Plan is intended to help research teams to think about their project's organization, activities and responsibilities, in particular communication and dissemination activities, engagement with stakeholders, and data management needs and responsibilities.

The plan aims to describe the approaches necessary to implement a WASCAL data management infrastructure. It concerns the management of data and data products to be created by the Core Research Program, the Graduate Program and the research teams in the Competence Center. The data will be hosted and managed at the Competence Center. The plan acts as an agreement between data users and data providers/creators in the framework of WASCAL, on how to use, store, disseminate and publish data generated by the project.

The target audience for this plan is all project members, in particular scientists from German and African research institutions as well as partner organizations providing and using data and data products.

The Plan also provides a basis for the overall planning and development of the central WASCAL data management by:

- providing information that can be used to co-ordinate communication activities and stakeholder relations across WASCAL
- highlighting data management and custody issues at an early stage so that data is managed in a way that meets the requirements of the WASCAL project management, and enables the Project to respond to common data needs
- providing a basis for quality assurance within the project
- providing a basis from which WASCAL project partners and the project management can report and monitor project and overall WASCAL project progress.

Authors:

- Dr. Ralf Kunkel, Research Centre Jülich, 52425 Jülich, Germany, E-Mail: r.kunkel@fz-juelich.de
- Antonio Rogmann, Zentrum für Entwicklungsforschung, 53113 Bonn, Germany, E-Mail: arogmann@uni-bonn.de
- Jürgen Sorg, FZJ, Research Centre Jülich, 52425 Jülich, Germany, E-Mail: j.sorg@fz-juelich.de

Content Scheme:

- Data Management Planning Tool "DMP online" (https://dmponline.dcc.ac.uk/) by the Digital Curation Centre (http://www.dcc.ac.uk/)

    Version: 06

    Date: 2013-01-06

Document version

    0.2

Document history:

2012-06-30	First draft 0.1 (Antonio Rogmann, Ralf Kunkel)
2012-11-30	Reviewing and extension (Ralf Kunkel, Jürgen Sorg)
2012-12-06	Reviewing (Antonio Rogmann)
2013-01-06	Reviewing (Ralf Kunkel)
2013-01-15	Reviewing (Antonio Rogmann)
2013-01-18	Reviewing (Ralf Kunkel, Antonio Rogmann)
2013-01-27	Reviewing (Antonio Rogmann, Ralf Kunkel, Joe Hill)
2013-01-31	Second draft 0.2, submitted to Project Management

Content

1 Basic Project Information

WASCAL (West African Science Service Center on Climate Change and Adapted Land Use) is an integrated research-focused program designed to help tackle the challenge of future climate and land use changes to enhance the resilience of human and environmental systems. It does so by strengthening the research infrastructure and capacity in West Africa related to climate change and by pooling the expertise of ten West African countries and Germany. Scientific competences allocated to the Science Service Center (called Competence Center hereafter) will play an important role in regional land management planning and policy advice. The project consists of three subprojects:

- The Competence Center (CC) develops local research capacity, and is a service center for the partner countries, building on the national research communities of the West African countries participating in WASCAL. The center connects regional partners in data-collection networks (observation networks), and offers the infrastructure and expertise necessary for analyzing the impacts of climate change and for developing strategies and policies to cope with them.

- The Core Research Program (CRP) is a joint research program on adapted land use and management under changing climatic conditions, which brings together a German and an equivalent West-African research consortium.

- The Graduate Research Program (GRP) supports and facilitates academic education amongst West African universities in association with German counterpart institutions. The activities focus on the training of doctoral students.

The project is funded by the German Federal Ministry for Education and Research (BMBF) within the time period 2012-10-01 to 2016-02-29. The coordinating organization is the Center for Development Research (ZEF) at the University of Bonn. A list of contributing organizations is given in Annex 2.

Due to the political conditions relevant for the establishment of the Competence Center, policies with regard to the protection of ethical standards and intellectual property rights are still to be defined and negotiated by the WASCAL Steering Committee together with the parties concerned. A preliminary draft is included in this Data Management Plan.

2 Capture Methods and Management of Research Data

2.1 Overview of the WASCAL Data Infrastructure Setup

The WASCAL Data Infrastructure (WADI) within the Competence Center will serve as a database node providing scientists and decision makers in West Africa with reliable and well accessible data and data products related to climate change and adapted land use.

Furthermore, the WASCAL Data Infrastructure will play a substantial role as a new and innovative node in the West-African network where research programs, projects and institutions working on climate change and related topics share data and knowledge.

By following the concepts of a Spatial Data Infrastructure (SDI), where spatial (and non-spatial) data, metadata, users and tools are interactively connected[1], a decentralized data infrastructure will be established within the framework of WASCAL. This type of data infrastructures is already used successfully in several other projects, e.g. the interdisciplinary and long-term research program TERENO, which involves six Helmholtz Association Centers (see http://www.tereno.net).

This decentralized setup aims at interconnecting regional data and metadata infrastructures as far as they are already established at partner institutions, and under the "umbrella" of WASCAL. A central data portal application and a central metadata calatogue (CMC) in WADI will act as a "one-stop" entry-point to the data sharing network, where, in addition to all data being hosted by WASCAL, data and metadata held by other West-African and also international partner institutions could be localized, analyzed and - depending on the data owners - retrieved.

Data publication and exchange will be facilitated predominantly through web services as standardized by the International Standardization Organisation (ISO) and the Open Geospatial Consortium (OGC)[2], operated from the Competence Center or from the data providers. WADI contains the basic features:

a) A WASCAL portal application, which allows one to query, find and access data from WASCAL databases or external data infrastructures according to the data policy.

b) Defined interfaces for data exchange between WADI and regional data infrastructures provided by standardized, OGC-conformal web-services operated in WADI and, as far as existent, in regional data infrastructures maintained by partners. The following web-service specifications will be in use:

   i. OGC Sensor Web Enablement (SWE) standards and interfaces will be used to provide bio-physical time series data from the WASCAL observation networks together with contextual details such as instruments used, intended application, valid time, contact details to responsible persons, security level, geographic location.

   ii. OGC compliant Web Map Services (WMS), Web Feature Services (WFS) and Web Coverage Services (WCS) will be implemented by using a web-map server (Geoserver) to provide map layers generated from vector data or raster data to the users. Non-OGC compliant web services may be implemented to provide additional functionalities not covered by the OGC standards and interfaces specifications. Common data file download services (based on http, FTP, etc.) will be connected to the aforementioned services where appropriate.

   iii. Metadata will be published by OGC compliant Web Catalogue Services (CSW) based on the ISO 19115/19139 metadata standard. As a central entry point to WASCAL metadata a Central Metadata Catalogue (CMC) will be created. The Catalogue will use the GeoNetwork Opensource[3] framework, which is a CSW application developed by FAO, WFP and UNEP, will be used to query, find, ac-

---

[1] http://en.wikipedia.org/wiki/Spatial_data_infrastructure (accessed 12.03.2013)

[2] http://www.opengeospatial.org/ (accessed 12.15.2012)

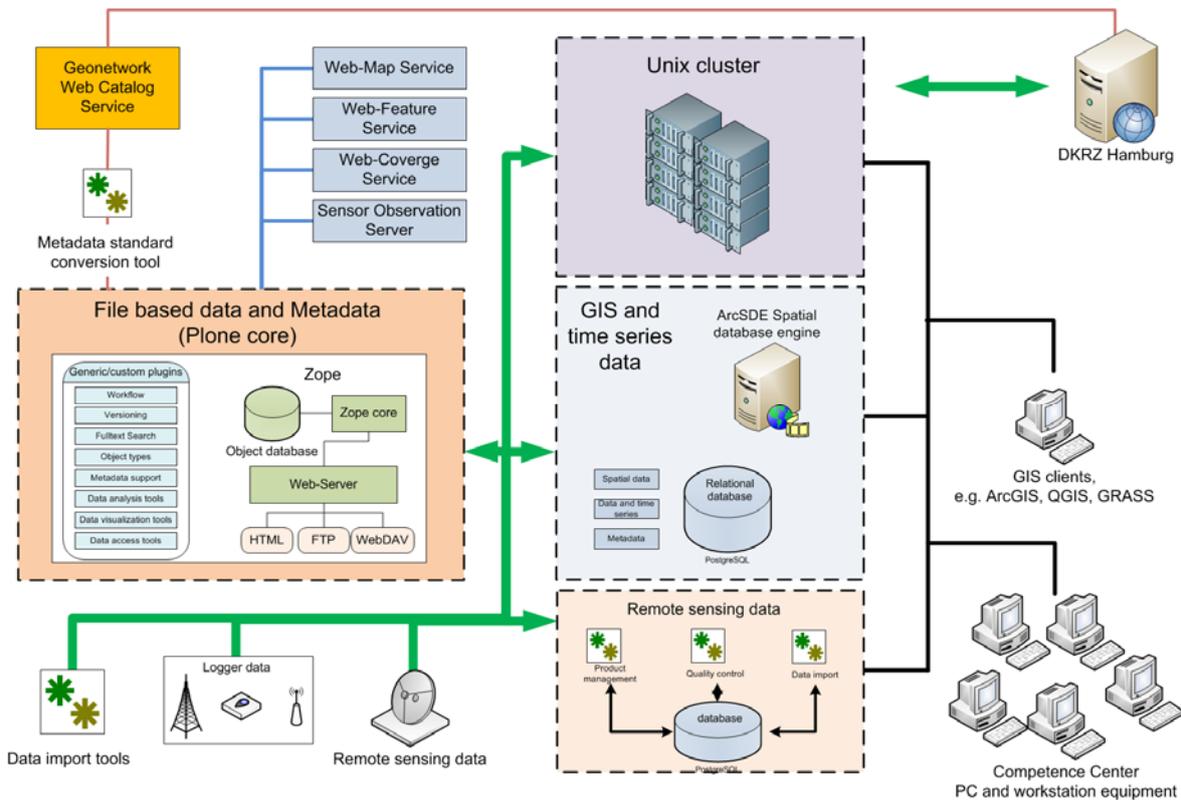[3] http://geonetwork-opensource.org/ (accessed 01.27.1013)

Figure 1: Components and functions of the WASCAL data infrastructure WADI.

cess and edit metadata and data in the WADI and to connect external data infrastructures. It supports multiple metadata standards (ISO19115, FGDC, DublinCore), contains an integrated map-server and map-viewer to access distributed map-services, is open source, free of costs, and supported by an active developer and user community. The online metadata editor offers different data sharing and security levels.

c) A central database to manage WASCAL data created, used and disseminated within the Core Research Program, Graduate Program and the Competence Center.

Figure 1 shows the main functions and components required for the central database to be implemented at the Competence Center. Basically, four different kinds of data will be managed:

a) File based data will be managed by a Content Management System based on the content management software Plone4, which combines traditional file storage with metadata support through Web- and Catalogue-services, and the aforementioned GeoNetwork metadata-catalogue (see section 4.2). Both interfaces rely on the same services to access the databases executed by WADI and provide therefore merely different views on the databases by its metadata. Workflows, user roles, versioning and extensive data search capabilities will be provided by this system. Access to data is possible by different internet protocols, e.g. HTTP, FTP and Web-DAV. An internal WASCAL communication platform based on Plone,

---

4 http://plone.org/ (accessed 01.27.13)

hosted at Research Centre Juelich, is already online since 2009 (http://www.wascal.org).

b) Time series data from the observation networks will be stored in a relational data base supported by a Time Series Management System developed at FZJ, which allows automated data input, standardized data analysis and data quality assessment routines needed to be created according to the particular requirements.

c) Geospatial data will be stored either file based or in a spatial database (PostgreSQL with PostGIS extension) and will be accessed by generic database interfaces (e.g. pgadmin, phppgadmin, …), the metadata-catalogue or by commercial GIS-Server applications (e.g. ArcGIS Server by Esri). These data are internally accessible by various GIS-clients, e.g. by ArcGIS or other open source GIS software. Data exchange between the different WASCAL partners, authorities and other users is realized by web-services following the OGC-specifications.

d) Remote sensing data may be stored either file based or in spatial databases. However, processing and creation of new products require special tools, which are provided by specialist systems for remote sensing applications. Depending on the amount of remote sensing data being collected at the Competence Center, such a system can be either incorporated into the WADI (see figure 1) or attached to the system as an additional local database connected to WADI by the OGC interfaces. The proposal of the remote sensing group contains more details.

Presently, there are no out-of-the-box systems available which can be used directly for the setup of WADI. Therefore, adaptations of existing systems, e.g. the TERENO DMI, to WASCAL's requirements will be made. These adaptations are mainly attributed to database model design and to interface tools developments between the individual systems components, which have to be developed according to the requirements identified by the WASCAL users and/or researchers. A preliminary data sharing infrastructure has already been implemented at Research Centre Juelich prior to the WASCAL main phase start in autumn 2012.

2.2 Existing Data

Existing input data sets of different types (observations/raw data, derived data, etc.), required for the work packages of the Core Research Program (CRP) and the studies to be carried out within the Graduate Program will be provided by external sources. Of main importance as data sources in this respect are collected or produced data from previous research projects, such as the IMPETUS West-Africa Project[5] and the GLOWA Volta Project[6]. Other existing datasets will be acquired from governmental institutions, non-governmental organizations and programs as well as by international programs in the region, e.g. the AMMA[7] program, the FAO[8], and others.

Access to the data of interest to WASCAL stored in the data repositories maintained by the programs and institutions, as mentioned above, will be organized by establishing a distributed, web-service based infrastructure (see previous section). Initial approaches

---

[5] http://geonetwork.impetus.uni-koeln.de/srv/en/main.home (accessed 01.27.2013)

[6] http://131.220.109.2/geonetwork/srv/en/main.home (accessed 01.27.2013)

[7] http://database.amma-international.org/ (accessed 01.16.2013)

[8] http://www.fao.org/geonetwork/srv/en/main.home (accessed 01.27.2013)

on cooperation with possible partners in this respect have already been established during the preparation phase, as with VBA (owner of the GLOWA Volta database), IMPETUS, AMMA, GEOSS, etc.

Preliminary information on the existing data stock used in WASCAL, and on the new data up-coming in the project's lifetime, are summarized in a data matrix (see section 1). The information on this data stock was gained by surveys where CRP scientists were asked to contribute to. As soon as the data is available and required by several users in WASCAL, it will be stored in WADI together with accompanied metadata.

2.3 New Data

The research region is characterized by data scarcity and existent data oftentimes is not suitable for the research purposes due to various reasons such as unknown quality, missing metadata or inaccessibility. Therefore, the generation of new data is of utmost importance to fill gaps in knowledge on regional phenomena, and to feed a variety of models and tools for scenario development which will serve as the scientific basis for policy advice to national/regional/local stakeholders. Within WASCAL new data will be captured or created by

- biophysical and socio-economic observation networks
- data collection by individual researchers from external sources and own investigations
- data creation by individual researchers by processing raw data from external sources or own collections
- observations provided by WASCAL Observation Networks and remote sensing:
  - Land cover/use data and essential climate variables (ECVs) captured by a variety of remote sensing instruments in different spatial (West-African Sudan Savannah belt to focal watersheds) and temporal resolutions (minutes, hours, days to decades, depending on parameters)
  - Hydrological data gathered by a regional observation network, distributed over Mali, Burkina Faso, Togo, Cote d'Ivoire, Benin and Ghana
  - Meteorological data gathered by a regional observation network, distributed over Mali, Burkina Faso, Togo, Cote d'Ivoire, Benin and Ghana
  - Biodiversity data collected from an observation network within and nearby the focal watersheds and from BIOTA West observation sites
- Socio-economic data generated by household surveys.

As preparatory work for the management of data to be collected or created by the Core Research Program, data surveys have already started to be carried out during the preparation phase of WASCAL. An analysis of the research proposals with regard to data and applications (models, software-tools, etc.) to be created and used during the project (results, see section 1) was carried out at the beginning of the requirements analysis. Based on a questionnaire delivered to the scientists of the CRP, detailed information on data are being gathered and compiled into data matrices. The matrices provide detailed information on dataset and parameter characteristics, and are structured in input- and output-data per work-package. An important detail is the expected time of availability of the data, because in time data flows among the CRP members and the CC research teams rely on a clear scheduling. From the results a database is being created, which allows

queries and reports by the data management executive group for its data management planning regarding data supply and demand.

Not every work-package was able to specify its data provisions over the complete lifetime of the project, at the project outset, therefore the above-mentioned survey will be repeated from time to time. In addition the questionnaire asks which software is to be used by scientists in the framework of WASCAL (first results, see Annex 4). This information will be used to plan software procurement and installation in the Competence Center. Demands on data supply in the Competence Center are not specified in detail as yet, but listings with thematic, data type and application related groupings of data to be used and produced in the CC during the project's lifetime will be prepared as far as necessary.

2.4 Data Integration and Standardization

In order to ensure compatibility in data integration by several applications/models, detailed standards and regulations for particular data groups which are not covered by common standardization rules, will be specified in cooperation with the respective scientific experts in their fields. Integration between the data being gathered in the project and pre-existing data sources will be made on several levels, as far as appropriate:

- Observable property level: where needed, transformation of units and recalculations of values
- Format level: reformatting to standardized formats (see next section)
- Geographic reference system level: transformation to standardized coordinate systems
- Documentation level: use of standardized metadata profiles
- Temporal and spatial scales (by metadata).

The Data Management Executive Group will provide reference tables with mandatory standard vocabularies to be used in WADI to the WASCAL community.

2.5 Data Formats

Data storage and data formats highly depend on the types of data within a data stock, which can be very heterogeneous. Basically, data of interest in WASCAL can be classified as follows:

- Structured (quantitative) data:
  - o Time series data provided by online in-situ or remote sensors, to be stored in relational databases
  - o Tabular (statistical) data derived from measurements, surveys or generated by simulations, to be stored in files or relational databases
- Unstructured (qualitative) data such as descriptive documents, audio- and video data, pictures, stored in files
- Georeferenced data stored in files or relational geodatabases
  - o Vector-based and attributed geodata
  - o Raster based geodata

Except for unstructured information, all data is being interpreted by computer programs to make it understandable and is - by its very nature - software dependent. Data are thus endangered by the obsolescence of the hardware and software environment on which access to data depends. The best option to warrant interoperability of data between varieties of applications and over the long term is to convert data in non-proprietary and standardized open-formats. Nevertheless, archiving of data in proprietary formats[9] can be required in cases where meaningful content would get lost when storing in open formats, e.g. formulas and diagrams in MS Excel tables when exporting in comma-separated values (ASCII/CSV) files. In order to save storage space on the hard-drives of the server, uploaded data may be stored as compressed data file packages (e.g. .zip, .7z, .rar).

The following file formats for data will be supported by WADI:

Type of data	Preferable formats for exchange and reuse	Acceptable formats
Qualitative data textual	Rich text format (rtf), OpenDocument Text (.odt), Plain Text (.txt), eXtensible Mark-up Language (XML) text according to an appropriate Document Type Definition (DTD) or schema (.xml) Portable Document Format (.pdf),	Hypertext markup language (.htm), Widely used office formats (e.g. MS Word)
Quantitative tabular data (with or without column labeling, variables names and metadata)	Comma Separated Values (.csv), Tab-delimited File (.tab), OpenDocument Spreadsheets (.ods), Extensible Markup Language (.xml), SPSS portable format (.por) NetCDF (.netcdf), Hierarchical Data Format (.hdf)	Microsoft Excel (.xls, .xlsx), dBase (.dbf), SPSS (*.sav) MS Access (.mdb, .accdb)
Digital image data	Bitmap (.bmp) Joint Photografic Experts Group (.jpg, .jpeg) Portable Network Grafics (.png) Tagged Image File (.tif, .tiff)	
Georeferenced vector data	ESRI shape file (essential - .shp, .shx, .dbf, optional - .prj, .sbx, .sbn) Geographic Markup Language (.gml) CAD data (.dwg)	Scalable Vector Graphics (.svg) Keyhole Mark-up Language (KML) CAD data (.dxf) MapInfo Interchange Format (.mif)
Georeferenced ras-	GeoTIFF (Tagged Image File Format/.tif, .tiff),	

---

ter data	NetCDF (.netcdf)
	Esri ASCII or binary Grid (.asc, .flt)
	Hierarchical Data Format (.hdf)
Digital audio data	Free Lossless Audio Codec (FLAC) (.flac)
	# MPEG-1 Audio Layer 3 (.mp3)
	Waveform Audio Format (WAV) (.wav)
Digital video data	MPEG-4 (.mp4), motion JPEG 2000 (.mj2)

Due to the extensive data volumes and for reasons of accessibility, time series data from in-situ and remote sensors will not be stored in simple file based formats but in relational database systems. For data storage in the database and registration of the sensor metadata, an underlying data model for time series data will be used, which stores environmental observations along with sufficient metadata to provide traceable heritage from raw measurements to usable information.

2.6   Naming Convention

Data file names should reflect file content as best as possible. Therefore all filenames managed by WADI will use a consistent naming scheme defined as follows:

**<Project>_<ScientificDiscipline/MainTopic>_
<ThematicContent/Parameter>_<SpatialCoverage>_<TimeFrame(-TimeStep)>_
(<SpatialGeometryType>)_<VersionNo>.<extension>**

Within one <element> several items are to be separated by a hyphen (-), while <elements> are to be separated by an underline (_). Therefore, the data files may have the following names:

WSC_hyd_soil-moisture_bf-dano-watershed_20000101-20061231-h_v1.csv

WSC_soc_household-income_burkina-dano-watershed-diebougou_2013.xls

The content of the tags are:

<Project>:	denotes the project the data is attributed to (e.g. WASCAL = "WSC")
<ScientificDiscipline/MainTopic>:	abbreviates the discipline and/or main topic the data is attributed to. Coding examples may be: adm = Administrative hyd = Hydrology clim = Climatology
<ThematicContent/Parameter>	denotes either the thematic content or the parameter, which is described by the data (e.g. reservoirs, temperature, regional-capitals, food-security)

<SpatialCoverage>	denotes the spatial extent of the data (e.g. = ghana, ghana-tamale, burkina-dano-watershed)
<TimeFrame(-TimeStep)>	denotes the temporal extent of the data (e.g. 2010, 20000101-20061231-h)
<SpatialGeometryType>	specifies the type of spatial geometry in case of geodata. Coding examples may be: pol = Polygon pnt = Point lin = Line ras = Raster
<VersionNo>	specifies the version of the data set (e.g. v1)
<extension>	file extension according to the type of data (e.g. csv, xls, see section 2.5)

When uploading data to WADI, data files may automatically be renamed according to the common file names scheme using common metadata entries provided by the user/uploader of the data. The possible content of the individual tags used for file renaming are provided by common vocabularies specified in reference tables, which can be accessed and downloaded from the WASCAL data portal.

2.7 Quality Management

The level of data quality, following the common definition of data quality i.e. "The state of completeness, validity, consistency, timeliness and accuracy that makes data appropriate for a specific use"[10] , is highly dependent on the real world phenomena represented by the data and data products. To carry out data quality assessments specialist knowledge on the field of data generation is demanded. Therefore, the

a) responsibility for data quality assessments and assurance measures is by the scientists generating or working on the data.

b) data creators will be requested by the data management responsible to describe the data quality assurance measures and to give an evaluation on the trustworthiness of the data within the Central Metadata Catalogue.

All WASCAL data must be attributed by the responsible scientists with respect to its data quality and data processing levels. This is to be done by a combination of a data qualifier and a processing status flag assigned to each data set. These quality information are either stored in the metadata for entire data sets (geodata or other file based data) or on the level of individual observations for times series data.

The data qualifier provides qualifying information that can note validity information about the data (like "visually checked") or anything unusual or problematic about individual observations (such as, for example, "holding time for analysis exceeded" or "incomplete or inexact daily total"). A list of data qualifiers will be elaborated by the Data Management Executive Group in collaboration with the responsible scientists in the first phases of WASCAL, published as a reference table by WADI, and will continuously be revised on demand.

---

[10] http://en.wikipedia.org/wiki/Data_quality (accessed 01.05.2013)

2.8 Intermediate and Persistent Data

All data created, collected and used by the WASCAL project and being regarded as valuable for verifying scientific findings, reuse in further research and integration in policy advice statements, will stored in WADI or in connected data infrastructures as far as authorized by WASCAL. However, in order to avoid overflows in the data repositories and the accumulation of unnecessary and unneeded data versions, scientists should carefully evaluate if a dataset should be shared by integration into the databases or not.

In the data processing chain a data set might have an intermediate character.

The processing status characterizes the general processing steps that the data have been subjected to. Depending on the parameters being controlled, several processing status levels are defined:

- Processing level "L-1": Raw Data
  Raw data is defined as unprocessed or preliminary pre-processed data and data products that have not undergone quality control. Depending on the data type and data transmission system, raw data may be available within seconds or minutes after real-time.

  Raw data

- Processing level "L-2": Quality Controlled Data
  Quality controlled data have passed quality assurance procedures such as routine estimation of timing and sensor calibration or statistical/visual inspection and removal of obvious errors.

- Processing level "L-3": Derived Products
  Derived products require scientific and technical interpretation and include multiple-sensor data.

  Intermediate results

- Processing level "L-4": Interpreted Products
  These products require researcher (PI) driven analysis and interpretation, model-based interpretation using other data and/or strong prior assumptions.

- Processing level "L-5": Knowledge Products
  These products require researcher (PI) driven scientific interpretation and multidisciplinary data integration and include model-based interpretation using other data and/or strong prior assumptions.

  Final research results
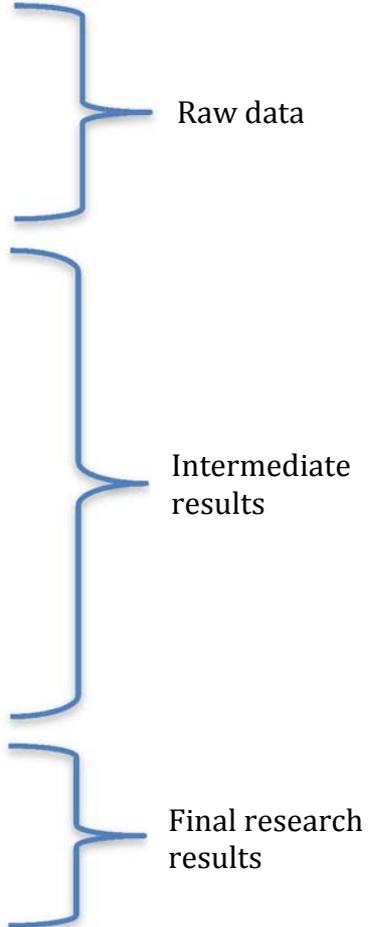
With the intention to describe phases in creating scientific and publishable findings, the state of data processing can be related to specific research domains. Research Domains define the group of users with whom data will be shared[11]. The listing below serves as an orientation guide for the decision as to whether data should be up-loaded into WADI or not. If in doubt, the WASCAL Data Management Executive Group or the project coordinators should be consulted.

---

[11] Treloar & Harboe-Ree, 2008 (http://www.valaconf.org.au/vala2008/papers2008/111_Treloar_Final.pdf accessed 10.24.2012)

a) Private Research Domain: Involves members of a core research team, e.g. a work-package in the Core Research Program. Data used in this domain usually represent raw, preliminary or intermediate rather than final research findings. Data management and sharing occurs within the team, and is pre-defined as members only.

→ Therefore, data do not have to, but may be integrated in WADI or in an external Data Management Infrastructure connected to WADI. In case data is stored in the research databases accessible by WADI, it becomes accessible with restrictions, such as for instance for work-packages or Core Research Program team members only (mapped as a specific „user group in the system)

b) Shared Research Domain: Involves researchers collaborating with a core research team, often across institutions. Data shared mostly represent intermediate research results, which are not the subject of a scientific publication as yet.

→ Therefore, data have to be integrated in WADI or in an external Data Management Infrastructure connected to WADI, but the data access is restricted to members of a research cluster, the research groups in the Competence Center or to WASCAL staff as a whole (mapped as "user groups" in the system).

c) Public Domain: Involves the public sphere. Data may have all processing states and may be used in or referenced by publications.

→ Therefore, data have to be integrated in WADI or in an external Data Management Infrastructure connected to WADI, and will be accessible by the scientific community, decision makers, politicians and the interested, non-commercial public.

2.9 Versioning

Within a data processing chain several versions of a data set may be created. A version can represent a status in the processing chain, or a small modification made to the data due to other reasons. Versioning might also occur in scenarios being results of modeling.

Versions can be marked simply by adding a version number to the file-, database-name, database-record or to a data-identifier, e.g. <file/database/database-record-name><version-number>, respectively, by adding a time stamp. In addition, parent/child relationships may be included in the Central Metadata Catalogue to document several states in the processing chain as versions.

Where required server-client based version control software will be used, e.g. Apache Subversion (SVN). Subversion management software is commonly used in software development environments, where several developers work on the software code locally, and code extensions or modifications are merged to a software code repository on a server. SVN clients are connected to the server by common data transfer internet protocols as like FTP, SFTP or SCP.

The possible usefulness of the above, and implementation for other applications e.g. climate modeling, in the framework of WASCAL will be discussed with the respective specialists in the Data Management Council (see 7.1 b) at the project outset.

2.10 Data Documentation and Metadata

Data documentation through metadata is an essential requirement for the assessment and evaluation of data within a data infrastructure. Moreover, the task of a data infrastructure to disseminate and to promote the usage of the data usually fails if the data are not properly described and, as a consequence, users are not able to judge whether the data content is relevant and the data quality is sufficient to their specific needs.

In WASCAL, metadata will be managed and published and, in case of file-based data manually provided to the WADI by scientists, be created by using the tools provided by a Central Metadata Catalogue (CMC) application integrated in the WASCAL data portal. These tools provide the user with facilities to upload research data with its metadata in one working step. A metadata set consists of several elements to describe the data, e.g. author, abstract, lineage, data quality, etc. In order to exchange metadata with other metadata catalogues or comparable applications, international approved standards and protocols will be used (see section 2.10.1).

2.10.1 Metadata Standards

The principal metadata standard for file based data and geodata to be used is the ISO 19115 "Geographic Information - Metadata" standard, in xml-encoding following ISO 19139. These standards are widely established within distributed spatial data infrastructures, where georeferenced data represent a large proportion of the data stock. The set of metadata elements in the ISO standards is capable to cover most data-types, data-topics and data–processes, even for data without coordinate-based spatial reference. If necessary, the ISO standard 19110 to describe feature types (e.g. in tabular data on attribute level) will be used. For digital documents (qualitative) and any other type of qualitative digital web-content, Dublin Core will be used as the metadata standard.

Hydrological and meteorological time series data gained from the observation network will be distributed by web-services, which provides the data together with basic metadata. The metadata scheme provided by these services corresponds to the SensorML specification based on the Sensor Web Enablement (2.0) initiative of the OGC.

Metadata is retrieved from the relational database in which the data is stored, and will be integrated and/or converted into the service configuration by the Data Management Executive Group in cooperation with the scientists responsible for the Observation Network in the Competence Center.

2.10.2 Metadata Profiles

Every user group will be provided with a range of metadata profiles/templates tailored to the requirements of specific data groups (e.g. remote sensing data, sociological surveys, statistical data) under their custody. The profiles are subsets of the basic standard ISO 19115, which consists of around 400 metadata elements, and which help the user to work more economically with the metadata editors in the CMC. Already established ISO 19115 profiles will be evaluated against the needs of WASCAL (e.g. WMO-Core). Own profiles, will be developed by the Data Management Executive Group, together with responsible persons from the WASCAL research groups, and implemented in the CMC as far as is suitable.

2.10.3 Metadata Creation and Management

Creation and capturing of metadata depends on the type and origin of the research data. Some applications (e.g. ESRI ArcGIS) are capable to create and export metadata into a

standard, which is compliant to the WASCAL metadata standards. Standalone metadata editors based on the WASCAL standards are also available (e.g. the INSPIRE Metadata editor[12] or CATMDEdit[13]) and can be used alternatively. As soon as it is connected to WADI the metadata can be imported together with the data into the CMC. In general, however, metadata for file-based data will be created online through editing tools within the WASCAL Data Portal application. The life cycle of metadata, from creation and content controlling of a metadata set, to its publication in the web, is managed by dedicated user profiles in the system. The members of the Data Management Executive Group will train the users to use the metadata-editing tool. The process of establishing these metadata creating and data sharing workflows will be lead and supported by the Data Management Executive Group as well. In the creation and management of metadata the following rules apply:

a) Every data set stored or archived in WADI must be described by meaningful metadata, compliant to the defined metadata standards and profiles.

b) It is the responsibility of the scientists respectively the WASCAL staff members up-loading data to WADI to provide proper metadata with the data. She/he will be granted the "editor" role in the user account management system of WADI, which enables him/her to create, edit and delete his/her own metadata in the CMC.

c) For every working group at least one responsible person will be nominated (see section Annex 1), who will be in charge to control the upload and the distribution of data the content and the compliance to the standards of the metadata contributed by colleagues in the group. This person should be member of the Research and Institutional Data Manager Working Group (RIDMG) (see section 7.1 c)). She/he will be granted the "Content Reviewer" or the "User administrator" role in the system by which metadata content and data distribution can be controlled on working group basis.

d) The metadata must contain either a URL to the online resource for accessing the data directly or at least the contact details where the data can be retrieved when stored locally.

e) Metadata is allowed to be modified on any occasion by the owner, but it must contain a basic set of metadata elements – mostly part of by a dedicated metadata profile (see previous section) - to be classified as "completed".

f) Data may be rejected from being imported into the infrastructure if proper and completed metadata are not delivered within a certain time period after uploading the data.

In addition, controlled vocabularies such as keyword-thesauri, reference tables, fixed categories etc. will be implemented in coordination with the users, to support the easy finding of data by filtering the metadata. The General Multilingual Environmental Thesaurus (GEMET) is the first choice for thematic thesaurus implementation, together with a thesaurus for spatial locations. Furthermore, for example, more scientific discipline oriented thesauri, may be included in the catalogue if required by the users.

---

[12] http://inspire-geoportal.ec.europa.eu/editor/ (accessed 01.15.2013)

[13] http://catmdedit.sourceforge.net/ (accessed 01.15.2013)

2.11 Georeferencing

In order to create maps by overlaying geospatial layers generated from geodata of different sources it is important that the layers share the same geographic coordinate systems (GCS) and projections. Specific requirements on the accuracy of spatial localization may demand the use of regional or local coordinate systems and projections. However, when submitting geospatial layers to WADI the geodata should be referenced by the World Geodetic System 1984 (WGS 1984). Therefore it might be required to transform the geodata, e.g. from UTM into WGS 1984, before uploading into WADI or publishing the geodata as Web Map Service layers. The layers based on the original GCS can be added to the upload as it deems necessary.

3 Data Rights Management: Privacy and Intellectual Property

Data sharing and re-use is subject to legal regulations. Therefore, the data rights management is of crucial importance in the data management framework, and of high priority in each and every data management plan. This section in the data management plan will be developed by the WASCAL Data Management Executive Group, in coordination with the responsible bodies in the WASCAL project consortium, and with the help of legal experts. As soon as it is legally approved and adopted by all relevant parties in WASCAL, it will be published officially and complement the data management plan.

In the following a preliminary draft for a data rights management statement, and respectively a data policy, is provided. As mentioned, it might be adapted to the legal constraints resulting from the political and therefore juristic conditions in which WASCAL, and respectively the Competence Center, is embedded.

3.1 Ethical and Privacy Issues

In general, personal data, which is data characterizing an individual human being, is subject to particular protection. In research, personal data containing privacy-sensitive information (e.g. household data with private/personal material or family details, which lead back to the individuals) can be used and processed only with the consent of the individuals supplying the data or information. Correspondingly, in publications personal data can be used only with the written consent of the persons concerned in the data. In WASCAL the following privacy rules will be applied:

a) Data with privacy content will be stored in WADI only with a document containing the terms of use being signed by the individuals concerned.

b) Provided that the research purpose is not affected, the data will be anonymized as soon as possible by, e.g.

    i.   removing direct identifiers, e.g. names, addresses or spatial localizations

    ii.   generalization of values, e.g. ranges of household incomes

    iii.   storing the data by separating characteristics of personal or material circumstances which could be attributed to an identified or identifiable person. This can be done, for example, by storing the data in separated files/tables connected by identifiers (primary/foreign keys). Approaches in this respect will be developed with those scientists who produce data for which privacy, e.g. from local stakeholders within household surveys, needs to be maintained.

c) Data privacy protection statements have to be included in the metadata by the creator of data containing privacy-sensitive information.

d) The creator of data containing privacy-sensitive information is responsible to restrict the access to this data by using the means, which are provided by the WASCAL Central Metadata Catalogue (CMC).

e) If a researcher depends on using data with sensitive personal information in his studies, privacy protection statements, based on the terms of use (see under a) in this section) integrated in the data retrieval process would have to be signed by the data user before accessing the datasets.

3.2 Intellectual Property

Intellectual property does not protect bare facts/data, but rather protects only the form in which facts and/or data are presented. With regard to scientific outcomes literary works, 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 are protected, only in as far as that they meet the two requirements: originality (an intellectual creation) and individuality (i.e. having a personal feature).

As a consequence, not every type of data is regarded as the intellectual property of the concerned scientist, neither is every type of data protected by copyright law. An example is raw data from biophysical measurements collected by instruments or by standardized scientific methodologies. Data not subject to copyright can be subject to ownership or use right however, for example by the owner of the measuring instrument or the donor of a data collection activity.

Although intellectual property rights are non-transferable in general, a right holder can or is committed to give the license to other parties to use the data. A license can contain all sorts of regulations (see section 3.3). In WASCAL the following intellectual property rules will apply:

a) Data created by WASCAL staff members that are not considered intellectual property, are owned by the WASCAL project.

b) By signing a data sharing agreement (reflecting the rules as stated in section 4.1) every regularly employed WASCAL staff member, who is the holder of data considered to be intellectual property, grants a data use license (see section 3.3) to the WASCAL project.

c) With the data use license, WASCAL is authorized to make the data accessible by means of WADI, and in turn to grant use licenses to stakeholders for scientific, policy-advice and decision making purposes respectively, by putting the data in the public domain.

d) Data use licenses to be granted for data in the custody of WASCAL can be limited to individual users or group of users, or temporarily be rejected if restrictions as defined in section 4 come into effect.

e) Data under the ownership of partner institutions but also in custody of WASCAL due to its storage in WADI will be protected against unauthorized access on the basis of agreements between WASCAL and the respective partner institutions.

3.3 Licenses

Copyright law provides the intellectual property rights holder with the possibility to grant a license to the data user. A license is a user agreement, signed by both parties, which states the scope for using the data. In information systems, however, where usable items under intellectual property can be downloaded, a license is included in the metadata and/or oftentimes in the order process, so that by downloading it the user accepts the license, for instance by clicking a special check-box. In WASCAL the following licensing rules will apply:

a) WASCAL staff members will sign a data sharing agreement with WASCAL (based on the regulations stated in section 4.2) at the beginning of their work in WAS-CAL. This agreement includes a data use license to be provided to the WASCAL project, and respectively the Competence Center. This use right license includes the right to disseminate the data to other parties as deemed necessary by WAS-CAL, but only in consideration of the rules as stated in section 4.2.

b) WASCAL grants data use licenses to data users based on the license framework provided by the non-profit organization Creative Commons[14] (e.g. by a license type <Attribution-NonCommercial-ShareAlike 3.0 Unported – CC BY-NC-SA 3.0)[15]>.

c) The licenses will cited as abstracts within the metadata of the data. In addition the URL to the detailed license definition at Creative Commons will be included.

d) In the metadata-catalogue the license will be included in the order process for data under intellectual property, so that by downloading it, the user accepts the license by entering his/her name and email-address, and by clicking a special check-box.

e) Data in the public domain will be marked with a Public Domain Mark. It might be restricted to non-commercial purposes.

4 Data Sharing

4.1 Access to Data

Project data hosted by WADI will be accessible to project partners, and respectively the public, as soon as this is demanded by the research teams and/or the project coordinators and the Executive Director.

In WADI legal requirements will be mapped to a fine-grained user and user group structure. This structure will regulate the dissemination of data and data products by reflecting the right management demands as delineated in section 3. Therefore the following rules apply:

a) For data not subject to intellectual property rights: Mostly L-1 to L-2 data (see section 2.8) and owned by WASCAL. Will be accessible by everyone in the internet for scientific purposes. An identification/registration of the user in WADI is

---

[14] http://creativecommons.org/ (accessed 12.17.2012)

[15] http://creativecommons.org/licenses/by-nc-sa/3.0/legalcode (accessed 12.17.2012)

not required. A public domain mark (see section 3.3 e) will be attached to the data.

b) For data falling under intellectual property rights: Mostly L-3 to L-5 data with a data use license at WASCAL:

    i. Can be accessible for everyone in the internet for scientific purposes without any identification of the data requesting user. The data use license will be attached to the data, and needs to be accepted by the user when downloading (by submitting his/her name and email-address).

    ii. Access can be restricted to WASCAL members or specific groups in and outside of WASCAL for whatever reasons (e.g. due to publication intentions). These restrictions have to be communicated to the project coordinators, and the data managers need to be informed. The user identified as a member of such a closed data sharing network, and respectively as authorized by the data provider, needs an account in WADI to access the data. A data use license (see section 3.3 b) will be attached to the data, and needs to be accepted by the user when downloading (by submitting his/her name and email-address).

c) For data owned by external project partners (e.g. Met- or Hydro-Services): access restrictions and data use licenses will applied as stated within bilateral agreements.

d) For user accounts and group memberships, which grant access to datasets: will be administered by the WASCAL Data Management Executive Group, and respectively the data managers in the Competence Center, in collaboration with the members of the Data Management Council. Information on how to get permission to download restricted data ("ordering instructions") is included in the metadata of the requested data.

e) Data downloads will be registered by notifications in the form of an email sent to a responsible member of the Data Management Council in WASCAL (section 7.1 b)) or Research and Institutional Data Manager Working Group (section 7.1c) .

The main entry point to WADI will be provided through the WASCAL homepage, which connects either to the Central Metadata Catalogue application or to other data query and data access tools.

4.2 Provision of Data

In order to guarantee continuity to the research activities within WASCAL, it is essential that data produced by one research group is made available as soon as possible to the other research groups or individuals, and respectively to external partners or members of the worldwide scientific community.

In producing datasets to be exchanged, site operators need to refer to this data management plan, in particular to sections 2.4 to 2.10.

In consideration of the requirements and procedures as fixed in section 3, the following rules for data provision apply:

a) The project coordinators are the authorities to decide on data sharing and exchange. They will settle conflicts in a friendly, mutually understanding manner to the benefit of all stakeholders. They will make sure that a scholar is contacted

when his/her data are to be used by others. In case of conflict the decision rests with the WASCAL Executive Director.

b) Internal documents or project related announcements have to be published and exchanged by using the WASCAL homepage. Al other data, data products, documents, models, software code, etc. which is of interest for in- and outside of the project will be published and exchanged by using the Central Metadata-Catalogue integrated in the data portal of WADI.

c) Senior scientists and regular staff members of WASCAL have to provide the data to WADI as soon as it is prepared for exchange, and by following the standards as explained in sections 2.7-2.10 and the recommendations in section 2.8, but not later than 2 months after its acquisition. Access restrictions to the data will be regulated as mentioned in section 4.1.

d) PhD students ("scholars"), who are funded by WASCAL have

    i. priority rights to publish his/her data first, e.g. as part of his/her Ph.D. or publication, as far as the data (L-3 to L-4, section 2.8) is subject to intellectual property rights. If this priority right is not used, then after an adequate time span (24 months), this right is passed over to the project. In practice this means that onwards from the time the right is passed to the project, the coordinators in consultation with the executive director decide how the data will be used. If possible, they will consult and inform the scholar about their decisions

    ii. the obligation to upload data which is not subject to intellectual property rights (mostly L-1 and L-2 data, section 2.8) together with adequate metadata into WADI by using the CMC application.

e) If the study is a Ph.D. study conducted at a partner university, e.g. the University of Bonn, then the rules of the University apply. At present, such rules stipulate that a data CD should be handed over to the coordinator of the Ph.D. program at ZEF or uploaded into the ZEF Data Portal (online spring 2013) at the end of the study. This CD is stored at ZEF for 10 years. However, if the study is at any other partner institution, then their rule applies.

f) Monitoring data from the observation networks are to be made available online within three months after installation.

g) Project data are to be made available by database access within three months after the WASCAL project ends, or according to other allocated submission deadlines. It will be the Institutional Data Manager's responsibility to ensure that submission deadlines are observed and to chase possibly delayed or missing data.

h) For data provided by external parties, individual agreements will be made between WASCAL and responsible persons from the data providing and using partner institutions.

4.3 Reuse of Data outside of WASCAL

WASCAL is an interdisciplinary multi-partner project, with the aim to provide policy makers and stakeholders in the area of 'land use adaptation to climate change in West Africa' with useful information based on the research work to be made by senior and

junior scientists in the WASCAL Core Research Program, the graduate schools, and the research teams in the Competence Center.

Other organizations and programs working on comparable scientific agendas, such as the AMMA project and the GEOSS program, national political bodies and NGOs, will be interested in the data to be made available by WASCAL.

The Competence Center is to be established as a permanent institution. Data managers will have to further develop data management strategies fitting up-coming requirements in research and decision making, and considering new conceptual and technological developments relevant for the management of data (as far as this falls within the responsibility of the data managers).

4.4 Citation rules

4.4.1 Citation and Acknowledgment of Data Sources in Publications

When data stored in WADI and obtained from the Data Portal is used within a publication, the source of the data is to be cited by including a persistent URL (<server-address><UUID of the metadataset>), which addresses the metadata in the CMC. The link can either be included as a footnote or in the references/bibliography. If data is obtained from other 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 person within the WASCAL project, these persons should be acknowledged, cited or even become co-authors.

b) If the data is available from the WASCAL Data Portal, a citation of the URL to the metadata set in the CMC is sufficient, because the metadata contains the details from the data creator and contributors.

c) If an institution has provided the data set (generally the case with secondary data), the institution must be cited.

d) 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.

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

4.4.2 Citation and Acknowledgment in Publications

When a WASCAL staff intends to publish data, research results and information especially in an international context (journal, book, conference, and so on), several regulations should be followed. In each publication a reference to the donors and the executing institutes must be made: this is mandatory. For example, the acknowledgement in the publication may state one of:

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

b) This research project was carried out under the auspices of the agreement on scientific-technological cooperation signed by the governments of Germany and -

--. The German partner was funded by the German Federal Ministry for Education and Research (BMBF). The WASCAL Partner was sponsored by Ministry of .... The executing partners were the Center for Development Research (ZEF) of the University of Bonn, Germany and the ....

c) Prior to submitting the document for publication, permission should be requested from the editorial committee of WASCAL which consists of ..... persons: The director of ZEF-, ..., ..., ... Only after approval the document can be submitted.

d) Oral presentations and poster presentations of the scientific work in any kind of meeting must be preceded by a dry run within a group of project participants. This will serve not only as a form of training and as a means to communicate the presentation's intentions, but also as an occasion to obtain possible corrections for any misinterpretations of data.

5 Short-Term Storage and Data Management

5.1 Storage Media and Data Transfer

Data to be exchanged and published inside as well as outside of WASCAL will be stored on storage devices on servers at the Competence Center. Primary storage will be accomplished using RAID hard drive disks on a storage area network (SAN). In addition data are stored in relational database systems and Web-based content management systems, enabling data access via Hypertext Transfer Protocol (Secure) (HTTP(s)), (Secure) File Transfer Protocol ((s)FTP), Secure Shell (SSH), Server Message Block Protocol (SMB/CIFS) and Web-based Distributed Authoring and Versioning Protocol (Web-DAV). For data transfer from and to the Competence Center it is especially for large data objects necessary to provide a software tool that enables resume a transfer after an abnormal connection lost to avoid resubmitting already submitted data. This is particularly necessary because of a comparatively poor communication infrastructure within Africa.

Data from partners might be stored in their own data infrastructures and linked to the central data infrastructure by standardized web services (see section 2.1). Independently of the actual data storage devices in the WASCAL data infrastructure, a registration of data set in the metadata base together with distribution information giving instructions where to ask for data has to be made at least.

5.2 Backup

Data backup in WADI is accomplished using regular data and system (virtual machines) backups on hard disk drives and regular data archiving using tape drives. Backup frequencies will depend on the data modification rate but will be at least one per day. Archive of data to tape drive drives will be performed if necessary. The responsibility for data backup and archive is with the WASCAL data infrastructure administrators represented by the WASCAL Data Management Executive Group, and respectively, after implementation of WADI at the Competence Center by its data managers.

Institutional data managers (section 7.1c)) are responsible for the implementation and supervision of proper data backup and archive strategies in their institution.

5.3 Security

During the project's lifetime access restrictions and data security is ensured by using common IT components such as firewalls, virus-scanners, secure connections (ssl, tls,...). Access to data provided by internal and external web services may be restricted to take place via WADI based on Lightweight Directory Access Protocol (LDAP) technology. Based on data policy agreements, users need to be registered in the data infrastructure. Every user who accesses restricted data ought to sign an agreement in which the license conditions have to be confirmed (see 3.3).

Access control will be implemented on a per dataset basis in the CMC or through network access control, e.g. when accessing the server via FTP, ssh or other web-protocols.

Based on user profiles and user groups, WADI will give dedicated access and activity permissions on the system to a user. Access to data will be managed by the Data Management Executive Group in cooperation with those having data management responsibilities in the research teams (might be members of the Data Management Council), the project coordinators and the Executive Director. A user and group concept will be elaborated and will consider the following aspects:

- Role of the user with respect to data acquisition: Data provider/ producer/ consumer/ and data management related role in the research team.

- Type of individual relationship to the project: member of project staff, member of partner institution, employment conditions (funding, position).

- Scientific discipline and team-membership in working groups.

6 Deposit and Long-Term Preservation

A long-term strategy for maintaining, administering and archiving the data is to be ensured by the long-term storage of data at the Competence Center.

6.1 Long-Term Specifics

Data are to be kept beyond the lifetime of the project for as long as needed by the users in the Competence Center. How data will be managed beyond the CC establishing project (WASCAL) will depend upon how the center will continue thereafter. Assuming a continuation as a Science Service Center, data and data products will play a crucial role further on. Life cycles of data will be handled by following the demands expressed by scientists and stakeholders at the end of WASCAL. Plans for archiving and of preservation strategies (by seeing archived data as not immediately accessible data), will be developed by the Data Management Group at the Competence Center in time.

Managing datasets including sensitive data over the longer term will be accomplished taking into consideration national laws. Sensitive data will either be anonymized or the protection of data privacy will be assured by the Competence Center data managers by contracts.

Data transformation by reformatting to open formats or by transformation to an appropriate geographic reference system may be performed, as and if necessary.

6.2 Metadata and Documentation for Long-Term Preservation

Datasets will be linked by internal entries with the UIDs in the metadata base. In addition, xml files (ISO 19139 standard) are to be downloaded automatically together with the data files when a user accesses the data in WADI. This will be done by the data providers/creators by editing metadata online on the WASCAL data portal, or by using appropriate tools extracting meta information from data stored on local drives in a format (19139 xml schema), which can be uploaded to the central metadata catalogue.

Published materials and/or outcomes, preferably linked to the used data, will be included in the data infrastructures. This will be done by persistent URLs (composed of server hostname and metadata UID) to the related entry in links in the metadata catalogue, respectively by Digital Object Identifiers (e.g. from the DOI-System[16]) as far as is introduced in WASCAL.

6.3 Longer-Term Stewardship

Responsibility over time for decisions about the data once the original personnel have gone is with the Data Management Executive Group together with the headquarter of the Competence Center, and based upon data policy statements agreed with the respective institutional bodies of WASCAL and the Competence Center.

7 Resourcing

7.1 Organizational Roles and Responsibilities for Data Management

During the WASCAL project funding time period, coordination of data management at different levels of responsibility will be established in the following manner:

a) Data Management Executive Group (DMG): The DMG consists of the data managers responsible for conceptualization and implementation of the data management infrastructure in WASCAL. This group is responsible for the overall coordination, to build up and operate the WASCAL data management infrastructure and to organize and carry out user trainings. Members of the group are listed in Annex A 1.1.

b) Data Management Council (DMC): The DMC consists of the executive director respectively his deputy, one representative per Research Cluster, one leading representative of the Competence Center, representatives of the DMG and representatives of the WASCAL partner countries. In the DMC general strategic issues with regard to data management will be discussed and pending decisions of major relevance will be taken, for example with regard to data rights management. The group will meet or communicate on demand. Member of the group are listed in Annex A 1.2.

c) Research Working Group (RWG): The RWG will work together with the DMG and consists of one representative per each Work Package, research group in the Competence Center and Graduate Research Program and/or representatives of the main partner institutions contributing data to WASCAL. Within the groups/institutions they represent they have to organize the controlling, pro-

---

[16] http://www.doi.org/ (accessed 12.09.2013)

cessing, preparation for sharing (e.g. by transformations), quality management, and providing of data and metadata to WADI as far as demanded by the WASCAL data management policies. This could be done e.g. by the designation of data management responsibles (data managers) in the respective institution or WP. The DMG will work closely together with these data managers by supporting and training them in data management tasks within their responsibility e.g. the submission of metadata and data to the central databases, and will organize further processing as far as it falls under their responsibility. Further processing, aside of data sharing, could be the creation of portrayal and other visualization services. The group will meet and communicate on demand. Members of the group are listed in Annex A 1.2.
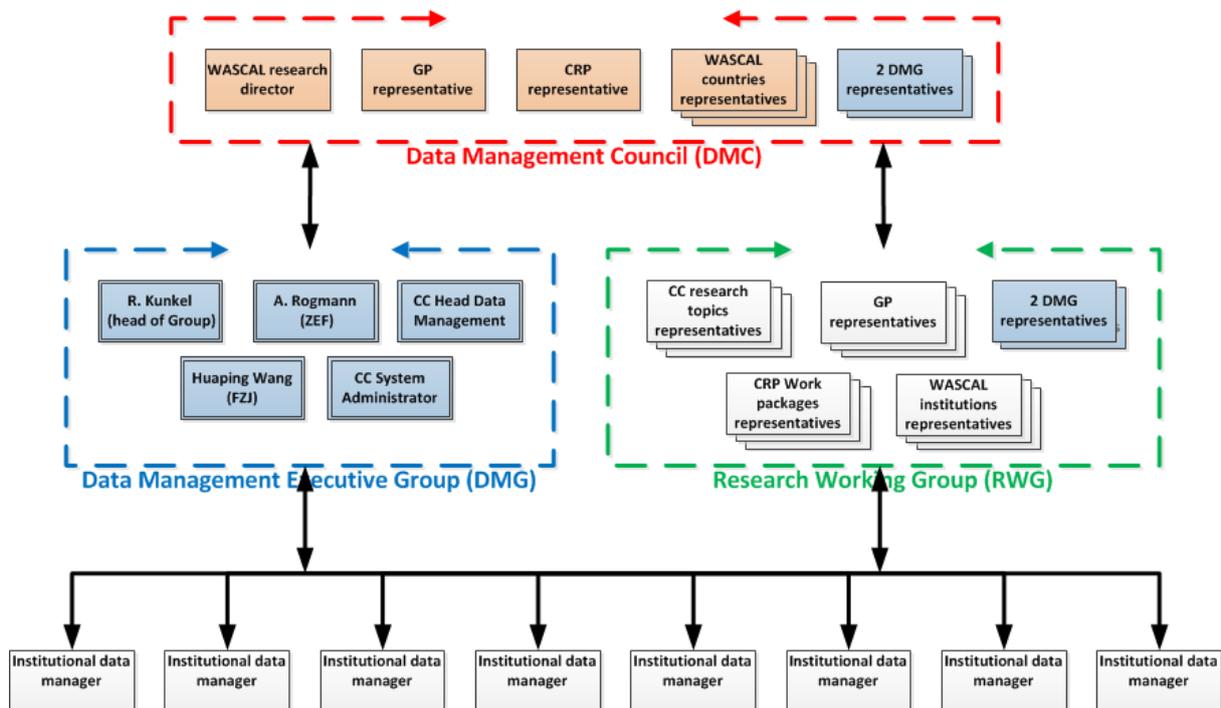

Figure 2: Organizational structure in the WASCAL Data Management

7.2 Data Management Activities Funding

Funding of data management within the project lifetime is part of the proposed project budget. Longer term funding after the project's end is not yet specified.

8 Adherence and Review

This data management plan will be signed by the WASCAL project management and eventually by the staff members of the WASCAL competence center. The degree of adherence to the plan will be observed in the daily work of the Data Managers and checked by the Data Management Executive Group.

Currently, there are no detailed plans for reviewing this data management plan.

9 Statement of Agreement

Annex 1   Contact details of nominated data managers / named individuals

A 1.1   Data Management Executive Group

- Dr. Ralf Kunkel (head of group)
  Senior researcher and data management coordinator at Research Center Jülich
  52425 Jülich, Germany
  Phone: ++49-2461-613262
  E-Mail: r.kunkel@fz-juelich.de

- Antonio Rogmann
  Senior researcher and data manager at Center for Development Research, University of Bonn
  Walter-Flex-Str. 3, 53113 Bonn, Germany
  Phone: ++49-228-734904
  E-Mail: arogmann@uni-bonn.de

- Dr. Christine Fürst
  Senior researcher at Center for Development Research, University of Bonn
  Walter-Flex-Str. 3, 53113 Bonn, Germany
  Phone: ++49-228-734922
  E-Mail: cfuerst@uni-bonn.de

- Jürgen Sorg
  Data manager and computer scientist at Research Center Jülich
  52425 Jülich, Germany
  Phone: ++49-2461-615535
  E-Mail: j.sorg@fz-juelich.de

- Huaping Wang
  Data manager and computer scientist at Research Center Jülich
  52425 Jülich, Germany
  Phone: ++49-2461-618055
  E-Mail: h.wang@fz-juelich.de

- N.N., Data Manager at WASCAL Competence Center
  Address: ?
  Phone: ?
  E-Mail: ?

- N.N., System Administrator at the Competence Center
  Address: ?
  Phone: ?
  E-Mail: ?

A 1.2   Data Management Council

A 1.3   Research Working Group

A 1.4   Institutional Data Manager

Annex 2   List of contributing organizations to WASCAL

Annex 3 Glossary

CC	WASCAL Competence Center: Service center for the WASCAL partner countries, building on the national research communities of the West African countries participating in WASCAL. The center connects regional partners in data-collection networks (observation networks), and offers the infrastructure and expertise necessary for analyzing the impacts of climate change and for developing strategies and policies to cope with them.
CMC	Central Metadata Catalogue, which stores all metadata managed by WADI.
CMS	Content Management System: Computer program that allows publishing, editing and modifying content as well as maintenance from a central interface. Such systems of content management provide procedures to manage workflow in a collaborative environment.
CRP	WASCAL Core Research Program: Joint research program on adapted land use and management under changing climatic conditions, which brings together a German and an equivalent West-African research consortium.
CSW	Catalog Service for the Web (sometimes seen as Catalog Service - Web): Standard for exposing a catalogue of geospatial records on the Internet (over HTTP). CSW is one part (or "profile") of the OGC Catalog Service, which defines common interfaces to discover, browse, and query metadata about data, services, and other potential resources.
DMI	Data Management Infrastructure
DMP	Data Management Plan: Formal document that outlines how you will handle project data during research time and after the project is completed. 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 ensure that data are well-managed in the present, and prepared for preservation in the future.
GIS	Geographic Information System: System designed to capture, store, manipulate, analyze, manage, and present all types of geographical data. In the simplest terms, GIS is the merging of cartography, statistical analysis, and database technology.
GRP	WASCAL Graduate Research Program: Program to support and facilitate academic education amongst West African universities in association with German counterpart institutions. The activities focus on the training of doctoral students.
OGC	The Open Geospatial Consortium (OGC), an international voluntary consensus standards organization, originated in 1994. In the OGC, more than 400 commercial, governmental, nonprofit and re-

search organizations worldwide collaborate in a consensus process encouraging development and implementation of open standards for geospatial content and services, GIS data processing and data sharing.

SDI	Spatial Data Infrastructure: Data infrastructure implementing a framework of geographic data, metadata, users and tools that are interactively connected in order to use spatial data in an efficient and flexible way.
SWE	OGC's Sensor Web Enablement framework defines a suite of web service interfaces and communication protocols abstracting from the heterogeneity of sensor (network) communication.
UID	Unique Identifier: any identifier which is guaranteed to be unique among all identifiers used for those objects and for a specific purpose.
WADI	WASCAL Data Infrastructure: Spatial data infrastructure created to manage the data from the WASCAL project
WASCAL	West African Science Service Center on Climate Change and Adapted Land Use
WCS	The Open Geospatial Consortium Web Coverage Service Interface Standard defines Web-based retrieval of coverage
WFS	Web Feature Service: Interface standard of the Open Geospatial Consortium, which provides an interface allowing requests for geographical features across the web using platform-independent calls.
WMS	Web Map Service: Standard protocol for serving georeferenced map images over the Internet that are generated by a map server using data from a GIS database. The specification was developed and first published by the Open Geospatial Consortium in 1999.

Annex 4   Data matrix

Excel File: WASCAL_CRP-OUT-IN-Data-Matrix.xlsx