IMPLEMENTING THE WASCAL DATA INFRASTRUCTURE (WADI)

HELMHOLTZ GEMEINSCHAFT

Open Science

Ralf Kunkel, Antonio Rogmann, Jürgen Sorg, Huaping Wang

Helmholtz Open Science Webinare zu Forschungsdaten, 2015-03-11

SPONSORED BY THE
Federal Ministry of Education and Research

Zentrum für Entwicklungsforschung
Center for Development Research
University of Bonn

JÜLICH
FORSCHUNGSZENTRUM

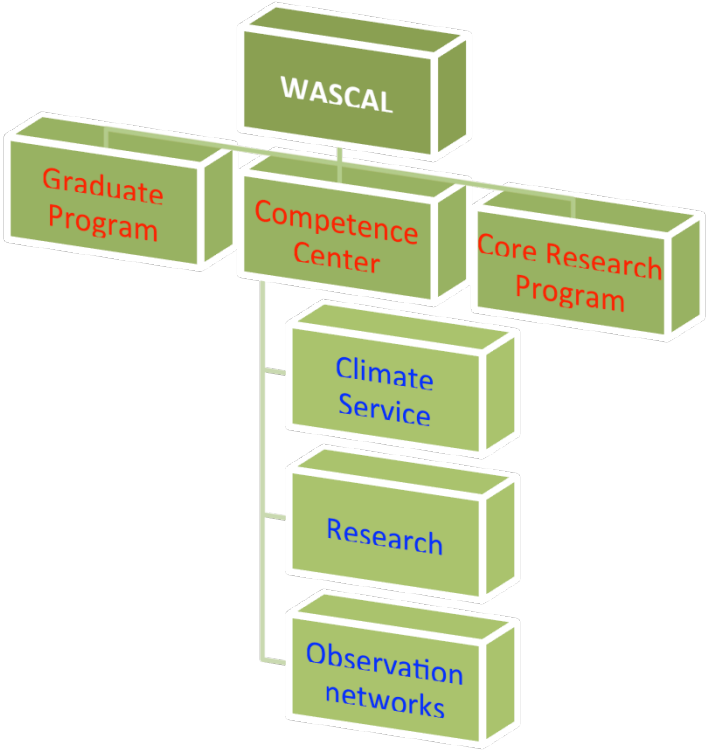
What is WASCAL?

- **West African Science Service Center on Climate Change and Adapted Land Use**
  - aims at strengthening the research, educational and policy capacity and competence of West-African countries
  - deals with issues of climate change through adapted land use on a scientific basis…
  - partnership with German institutions (research institutes, universities)

- **Implementation phase (2012-2016) funded by the German Federal Ministry of Education and Research**
- **WASCAL was recognized by ECOWAS as an international organization in May 2013**
- **9 countries signed WASCAL's constitution so far**

WASCAL Pillars

- **Graduate Program**
  - Supports and facilitates academic education (PhD and Master students) amongst West African universities in association with German counterpart institutions.
- **Core Research Program**
  - German/African joint research program on adapted land use and management of land under changing climatic conditions.
- **Competence Center (Ouagadougou)**
  - Accommodates the infrastructure for data reception, data maintenance and access, data interpretation and scenario analyses
  - Offers expertise necessary for analysing the impacts of climate change and for developing strategies and policies
  - Provides a collaborative work environment for the scientific staff, administrative and support personnel, and visiting scientists
  - Conducts training and outreach activities to involve the stakeholders
  - Connects regional partners in data collection networks (observation networks)



WASCAL data infrastructure (WADI) requirements

Common data sharing policy

Interoperability and standards

Graduate program research results

Data publication

Data provision

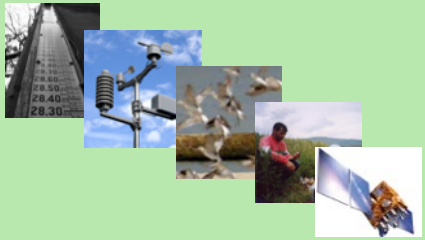
Files

Maps

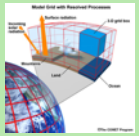
Time series

Model results

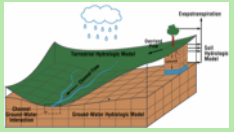
Core research program results



Observation network and remote sensing data



Climate model results



Hydrological model results

Socio-economical model results

WASCAL
West African Science Service Center on Climate Change and Adapted Land Use

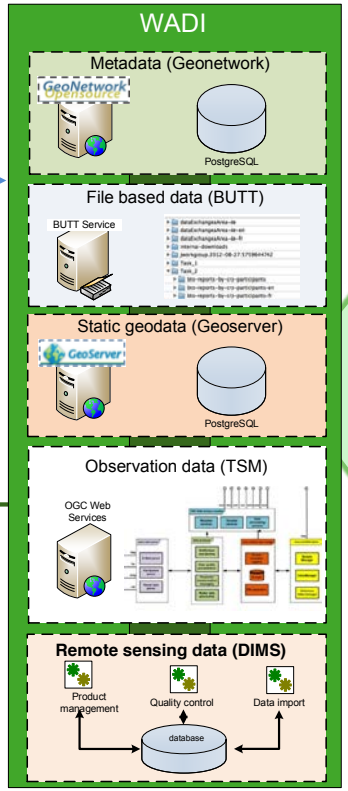
WADI components

WASCAL geoportal

Data discovery portal

OGC-compliant services

External data users

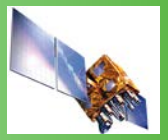


Files
Maps
Time series
Model results

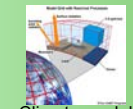
Graduate program research results

Core research program results

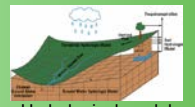
Data from the hydrological, meteorological, socio-economical, and biodiversity observation networks



Remote sensing data



Climate model results



Hydrological model results

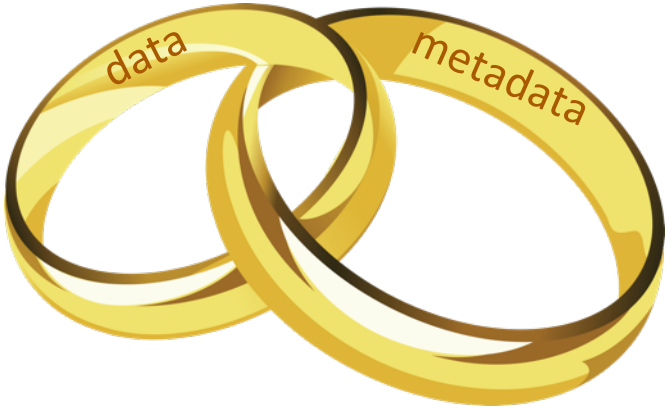
Socio-economical model results

Data and Metadata

- **Metadata are data providing information about one or more aspects of the data, such as:**
  - Means of creation of the data
  - Purpose of the data
  - Time and date of creation
  - Creator or author of the data
  - Location where the data were created
  - Standards to create or describe the data
- **Required to describe, discover and specifying the location of available data**
- **No data without metadata**
  - Each data set is described by a metadata set
- **No metadata without data (if possible)**
  - Each metadata set contains the location of the data

Data and Metadata

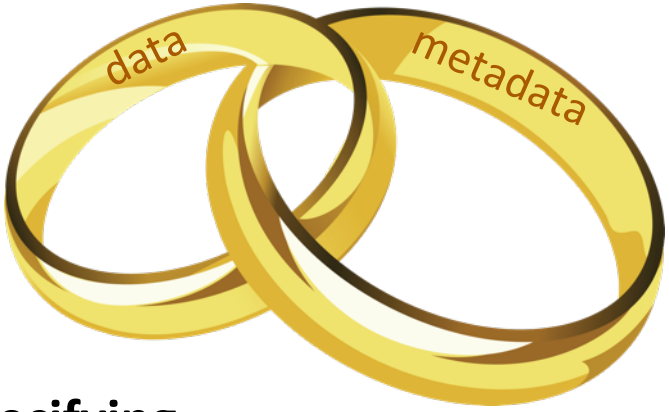
- **Metadata are data providing information about one or more aspects of the data, such as:**
  - Means of creation of the data
  - Purpose of the data
  - Time and date of creation
  - Creator or author of the data
  - Location where the data were created
  - Standards to create or describe the data



- **Required to describe, discover and specifying the location of available data**
- **No data without metadata**
  - Each data set is described by a metadata set
- **No metadata without data (if possible)**
  - Each metadata set contains the location of the data

Data and Metadata

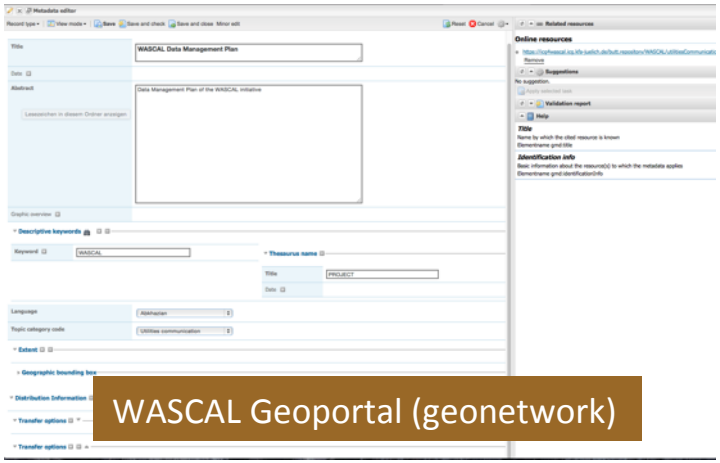
- **Metadata are data providing information about one or more aspects of the data, such as:**
  - Means of creation of the data
  - Purpose of the data
  - Time and date of creation
  - Creator or author of the data
  - Location where the data were created
  - Standards to create or describe the data



- **Required to describe, discover and specifying the location of available data**
- **No data without metadata**
  - Each data set is described by a metadata set
- **No metadata without data (if possible)**
  - Each metadata set contains the location of the data

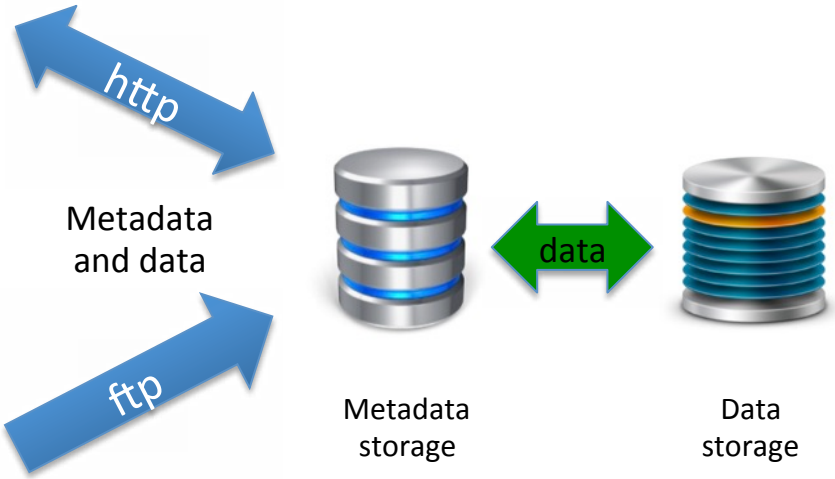
No divorce allowed

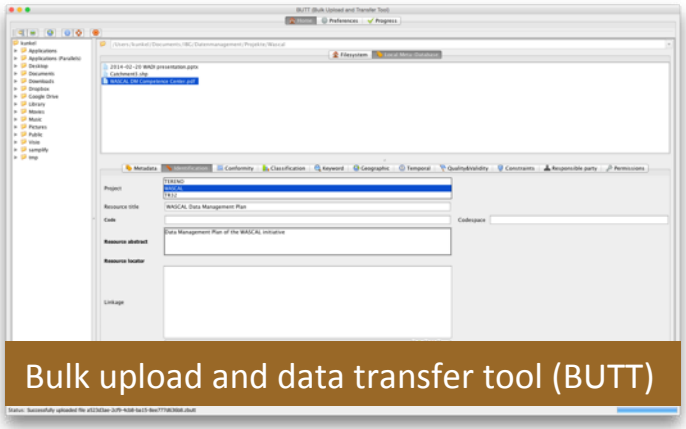
Putting metadata and data files into WADI



WASCAL Geoportal (geonetwork)

Online Metadata editing
File data upload and download
Metadata query and discovery
Advanced visualization of geospatial layers





Bulk upload and data transfer tool (BUTT)

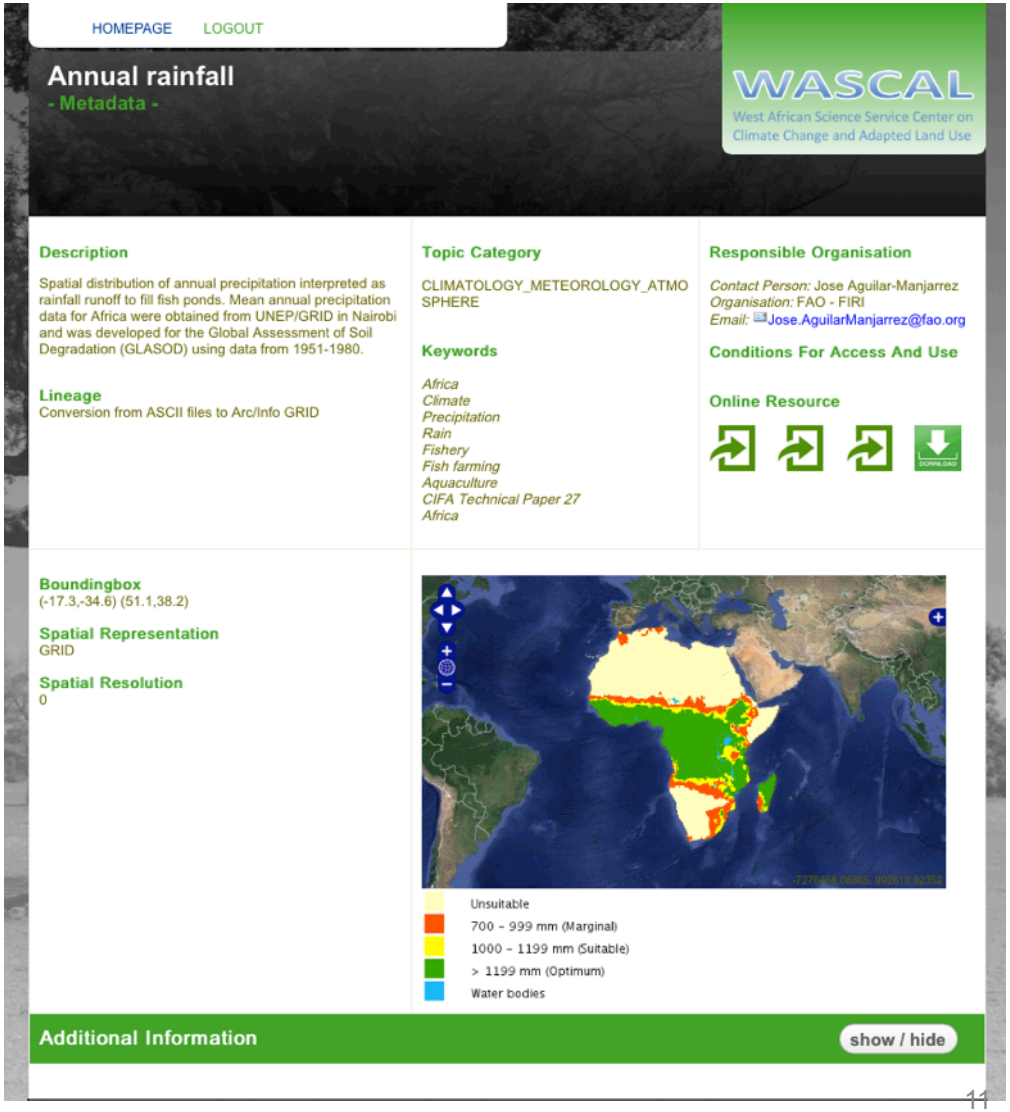
Offline metadata editing
Desktop application for **large** size
file data upload

Accessing WASCAL data
http://www.wascal.org

- Gateway to data for stakeholders, politicians and scientists
- Search and discover data catalogues
  - Predefined queries (e.g. “soil”)
  - Custom multi-conditional queries

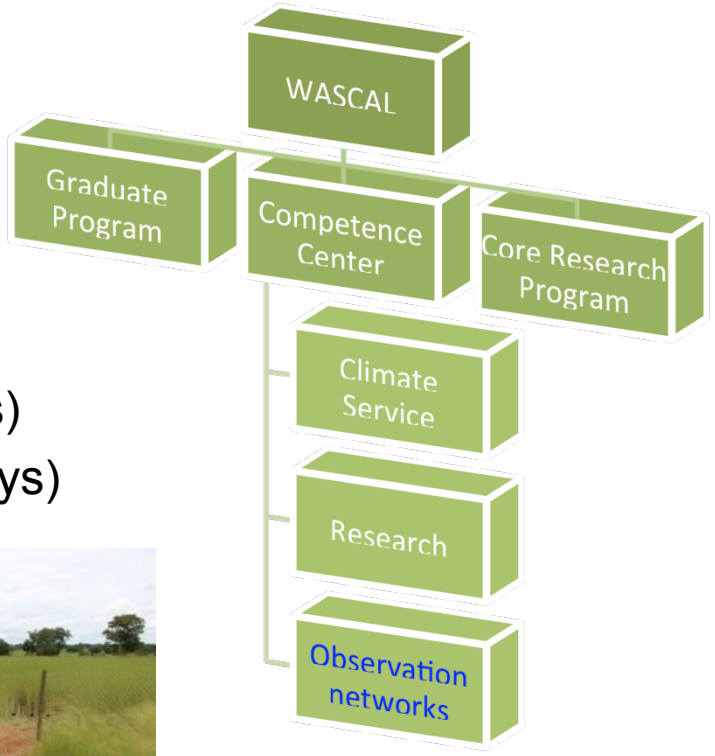
Accessing WASCAL data
http://www.wascal.org

- Gateway to data for stakeholders, politicians and scientists
- Search and discover data catalogues
  - Predefined queries (e.g. “soil”)
  - Custom multi-conditional queries
- Provides detailed information on data sets
- Advanced visualization of geospatial layers through WASCAL geoportal
- Online resources provides direct access to data

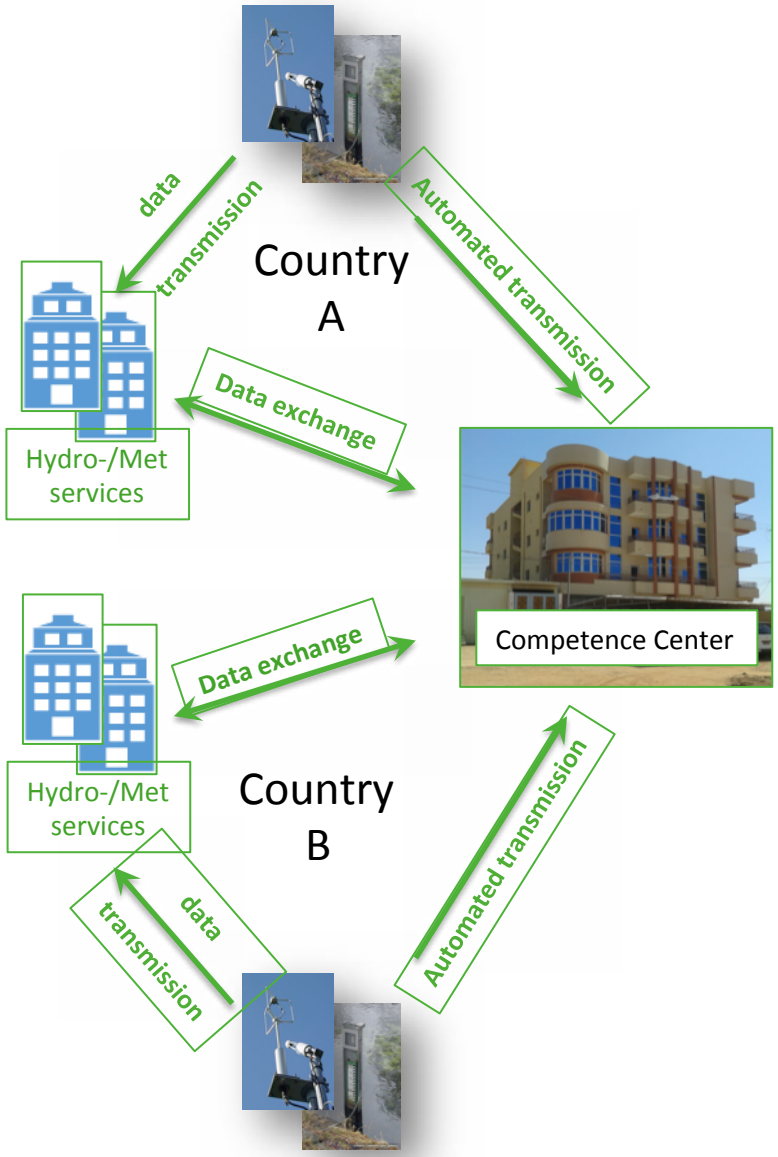


Observation networks and databases

- Within WASCAL observation networks and databases will be established in cooperation with national services:
  - Meteorology (station network)
  - Hydrology (station network)
  - Land (satellite receiving station)
  - Biodiversity (observatories, surveys)
  - Households (socio-economic surveys)



Observation network (time series) data



- Data sources:
  - Sent automatically (e.g. by GSM, E-Mail)
  - Retrieved from station manually by technician
  - Historical or preprocessed data already present

Typical data file

```
"TOA5","CR1000_Soe_ClimateStation","CR1000","E4404","CR1000.Std.24","CPU:WASCAL_WP1_Cl
"TIMESTAMP","RECORD","BattV_Avg","PTemp_C_Avg","AirTC_Avg","RH","SlrW_Avg","SlrkJ_Tot"
"TS","RN","Volts","Deg C","Deg C","%","W/m^2","kJ/m^2","meters/second","Deg","unitless"
"","","Avg","Avg","Avg","Smp","Avg","Tot","WVc","WVc","Smp","Tot","Avg"
"2012-09-25 13:00:00",0,12.29,37.11,34.42,52.02,1049,68.19306,2.19,109.4,0,0,75.54
"2012-09-25 13:30:00",1,12.26,36.64,34.67,53.33,1023,1842.239,1.396,115,0,0,982
"2012-09-25 14:00:00",2,12.26,36.61,34.91,50.89,853,1535.022,1.353,78.35,0,0,981
"2012-09-25 14:30:00",3,12.25,36.67,33.91,46.61,579.1,1042.369,1.112,61.07,0,0,981
"2012-09-25 15:00:00",4,12.27,36.69,34.12,50.34,603.4,1086.099,0.962,349.1,0,0,981
"2012-09-25 15:30:00",5,12.24,36.6,33.62,47.89,368,662.3224,0.712,7.922,0,0,981
"2012-09-25 16:00:00",6,12.42,36.52,34.39,46.77,486.4,875.4666,0.688,330.5,0,0,980
"2012-09-25 16:30:00",7,12.36,37.12,34.62,47.18,355.5,639.967,0.625,333.2,0,0,980
"2012-09-25 17:00:00",8,12.29,37.05,34.3,48.85,232.2,417.9905,0.603,317.8,0,0,980
```

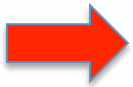
- Data will be stored in a data base rather than in individual files

Steps towards a distributed data infrastructure

Status of local data infrastructures	Support by WASCAL
Level 1: No central file server, mostly hardcopy data or file based data (e.g. Excel sheets) on local machines	• Digitizing hardcopy data • Uploading data into a central WADI database • Unfied data access by WADI • (set up local data bases)
Level 2: Central file server and/or data base, no publication of data to the web	• Setting up web services to publish data via WADI (or any other external web clients)
Level 3: Central file server and/or data base, application server to publish data to the web	• Setting up own web interfaces to provide data to the web

Challenges to implement observation neworks in West-Africa

- **Current situation:**
  - Lack of technical and human resources
  - Existing stations often need to be improved to allow automated data transfer
  - Historical data exist only in analogous form or file based (Excel-sheets)
  - No common internal data infrastructure (central server), almost no metadata available
  - Sometimes necessity for governmental institutions to sell data

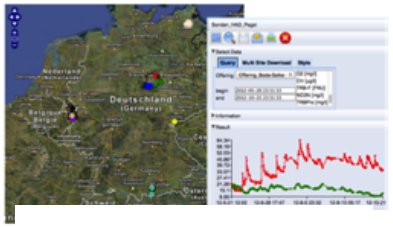


Level 1 Situation

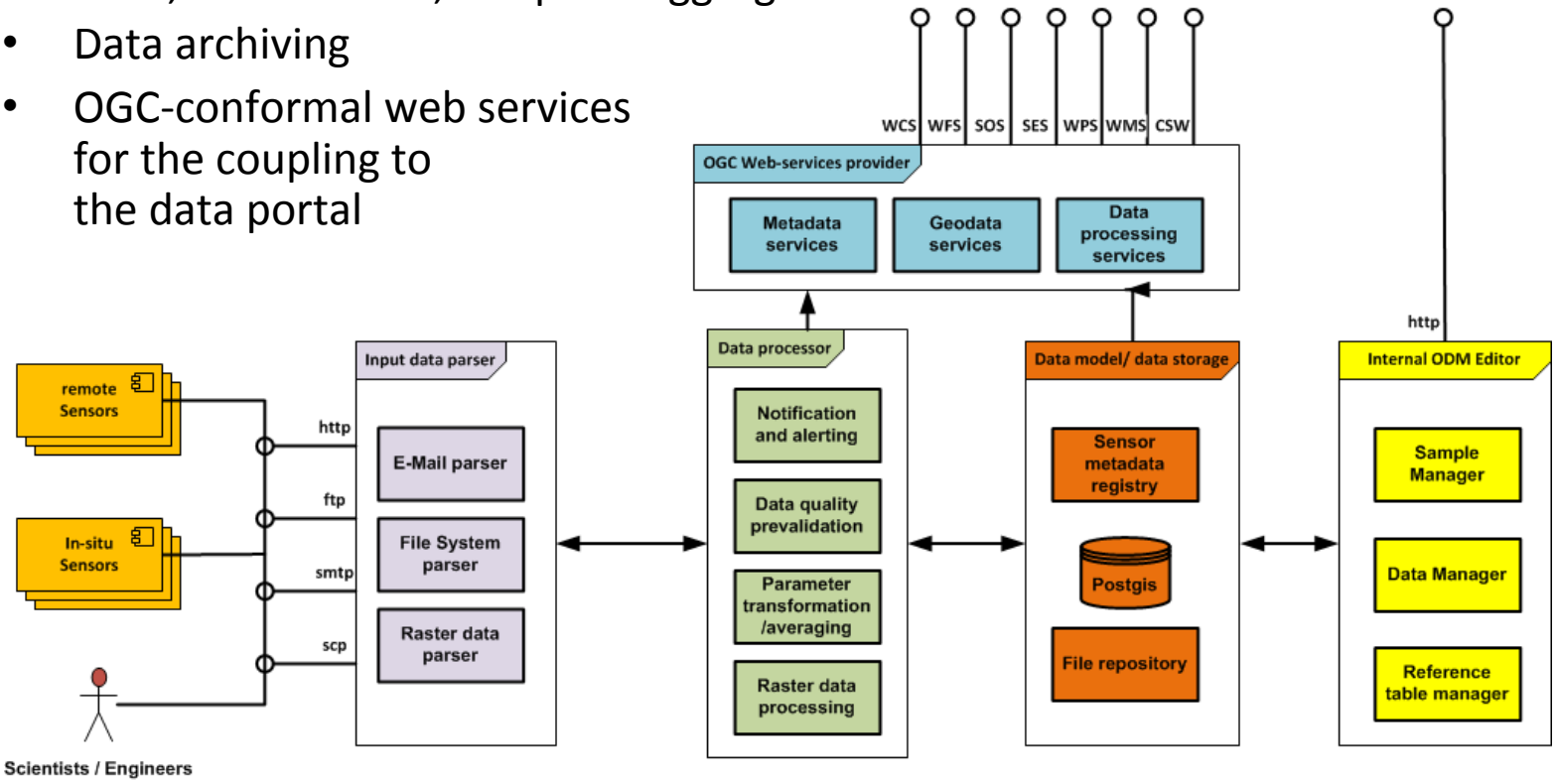
- **Measures (WASCAL ReHCON workshop 22-24.1.2015, Ouaga)**
  - Acquisition of unified equipment by WASCAL and transfer to each National Service (NHMS)
  - Training of staff such as observers, meteorological technicians and engineers
  - Support to digitize historical data
  - Implementing the transfer of present and past hydrological observations from the NHMS to the Competence Centre and vice
  - Participative implementation of regional platforms connecting NHMS and regional researches at the Competence Centre

Importing and processing observatory network data into WADI

- Automated data import and pre-processing
- Processing of different import sources (E-Mail, ASCII-files …)
- Min/Max-filter for data prevalidation
- Application of sensor specific transformation rules, mean values, temporal aggregates
- Data archiving
- OGC-conformal web services for the coupling to the data portal

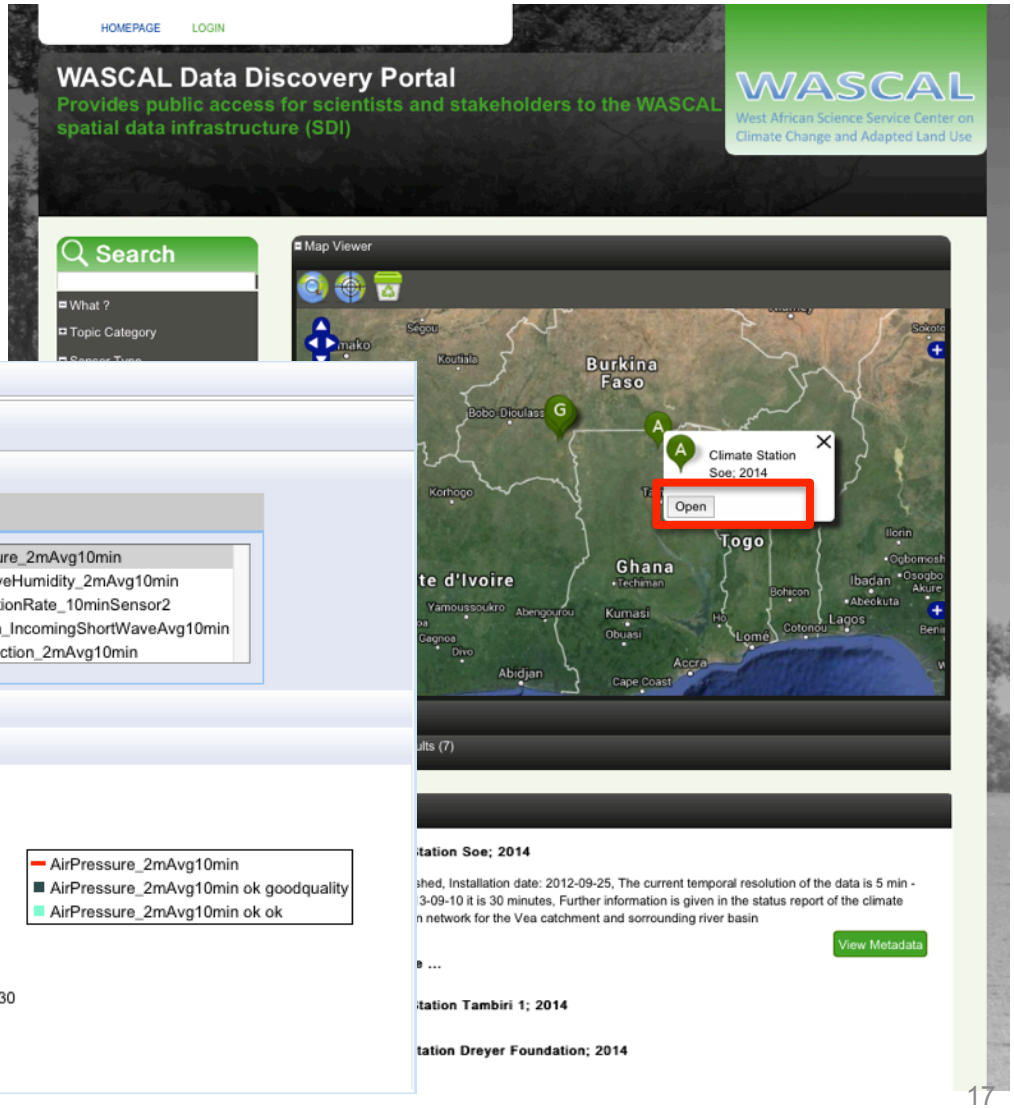


WADI Observatory Data Portal



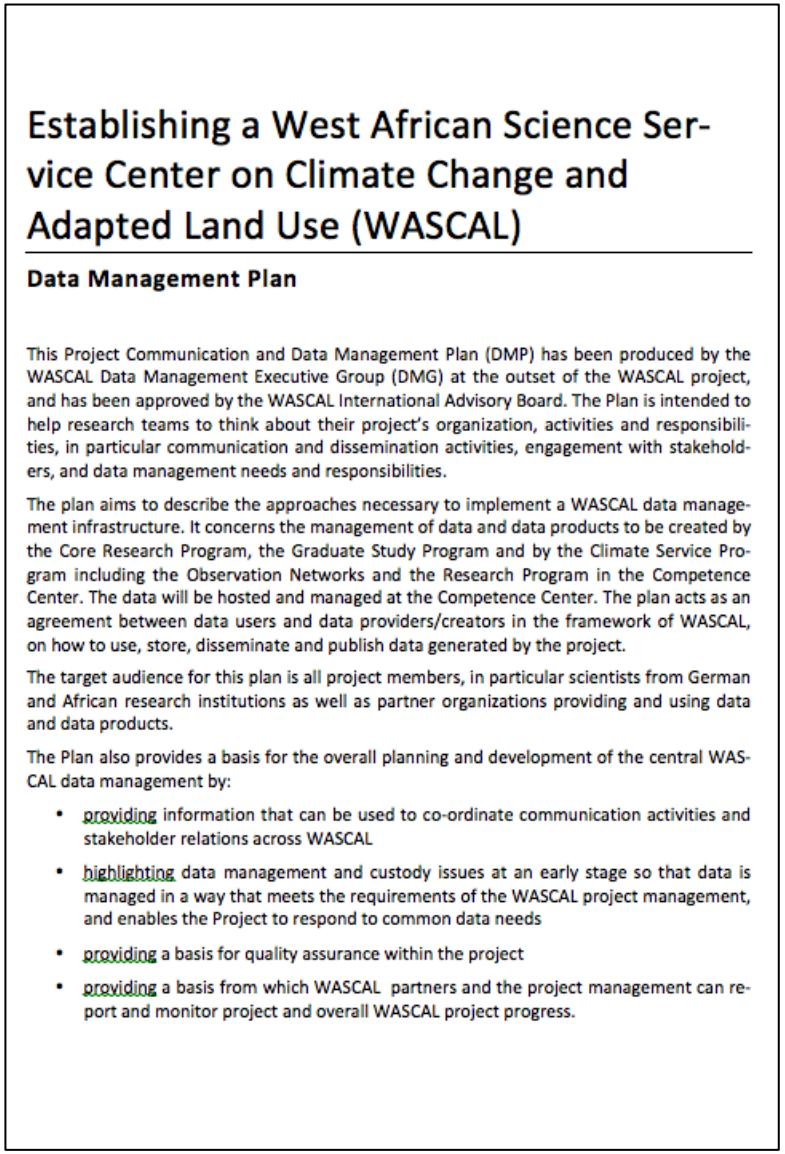
Accessing WASCAL time series data

- Standardized, multi-conditional search
- Visualization and download



Interoperability and standards: Data management plan

- Outlines how data are handled during research and after the project is completed.
- Described issues:
  - Technical implementation of the data infrastructure
  - Data to be managed
  - Used standards (services, data, quality, metadata, …)
  - Data preservation and security strategies
  - Organizational framework
- Ensures that data are well-managed in the present, and prepared for long term preservation in the future



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 (DMP) has been produced by the WASCAL Data Management Executive Group (DMG) 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 Study Program and by the Climate Service Program including the Observation Networks and the Research Program 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.

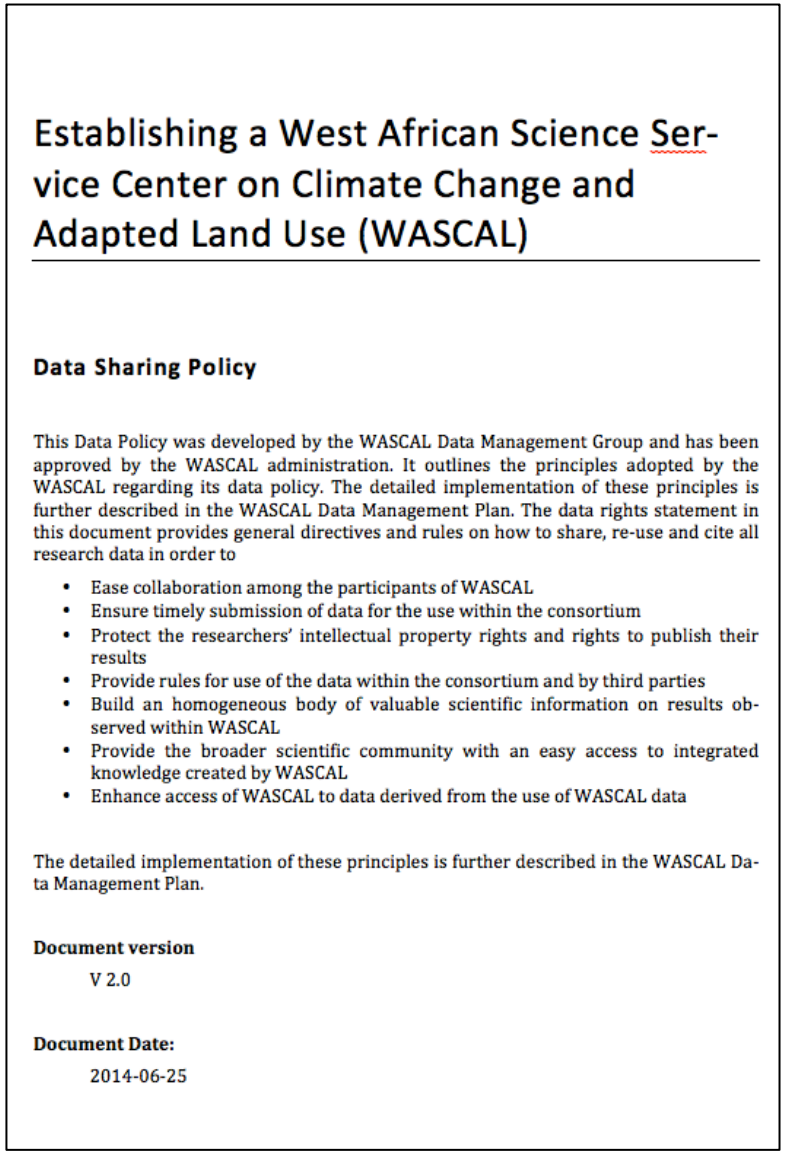
The target audience for this plan is all project members, in particular scientists from German and African research institutions as well as partner organizations providing and using data and data products.

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 partners and the project management can report and monitor project and overall WASCAL project progress.

Data sharing policy

- General directives and rules on how to share, reuse and cite WASCAL data:
  - Ethical Issues and data privacy
    - Written consents to use personal data for WASCAL research
    - Anonymisation of personal data
  - Copyrights and intellectual property (IP) protection
    - Rules for IP and Non-IP-data
    - Rules depend on work contract conditions of IP-data authors
    - Retention times for certain groups
  - Licensing model (Creative Commons)
  - Citation rules
- Provision, distribution, and reuse of externally owned data set by additional bilateral agreements



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

Data Sharing Policy

This Data Policy was developed by the WASCAL Data Management Group and has been approved by the WASCAL administration. It outlines the principles adopted by the WASCAL regarding its data policy. The detailed implementation of these principles is further described in the WASCAL Data Management Plan. The data rights statement in this document provides general directives and rules on how to share, re-use and cite all research data in order to

- Ease collaboration among the participants of WASCAL
- Ensure timely submission of data for the use within the consortium
- Protect the researchers' intellectual property rights and rights to publish their results
- Provide rules for use of the data within the consortium and by third parties
- Build an homogeneous body of valuable scientific information on results observed within WASCAL
- Provide the broader scientific community with an easy access to integrated knowledge created by WASCAL
- Enhance access of WASCAL to data derived from the use of WASCAL data

The detailed implementation of these principles is further described in the WASCAL Data Management Plan.

Document version

V 2.0

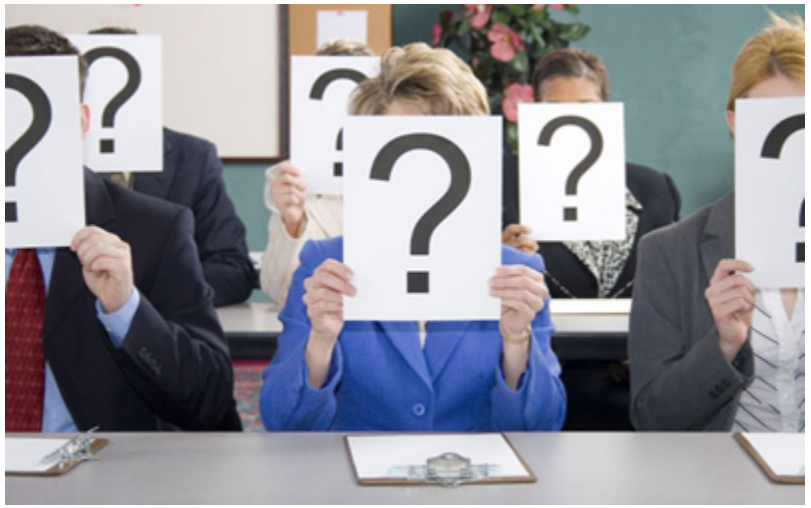
Document Date:

2014-06-25

Implementing data access rights

- **Data security and data access rights defined by**
  - The WASCAL data policy document
  - Additional documents and regulations
- **Central user/group management by LDAP directory** (WASCAL home page, Data Discovery Portal, Geoportal)
- **Only for authorized users / groups possible:**
  - Data editing and flagging
  - Data download (depending on access constraints)
- **With each download:**
  - Data use license provided
  - E-Mail notification to responsible party containing information with requesting user, downloaded data and purpose

Questions?



Contact:

- Ralf Kunkel (r.kunkel@fz-juelich.de)
- Antonio Rogmann (arogmann@uni-bonn.de)