Seminar on Proposed and Revised set of indicators

June 4-5, 2014 - Belgrade (Serbia)

The ORIENTGATE data platform

WP2, Action 2.4

Alessandra Nuzzo, Sandro Fiore, Giovanni Aloisio

Scientific Computing and Operations Division, CMCC

Data platform: goals and main tasks

The data platform represents a single entry point to the data produced by the ORIENTGATE partners, both in terms of ***climate simulations*** *and* ***impact indicator datasets****.*

Main tasks:

- *Task 1: Design of the data management platform*
- *Task 2: Setup of a virtual machine based environment*
- *Task 3: Data subsetting functionality and data compression*
- *Task 4: Deployment of the main data services*
- *Task 5: Enhanced configurations for the data services*
- *Task 6: Design and implementation of a dashboard based monitoring and browsing tool for the data platform*

Requirements and main activities of the data platform

General requirements

- *Heterogeneity of data*
- *Heterogeneity of the user requirements*
- *Heterogeneity of the software components (e.g. services)*
- *Integration and global view of the produced data*

Activities regarding the integrated platform

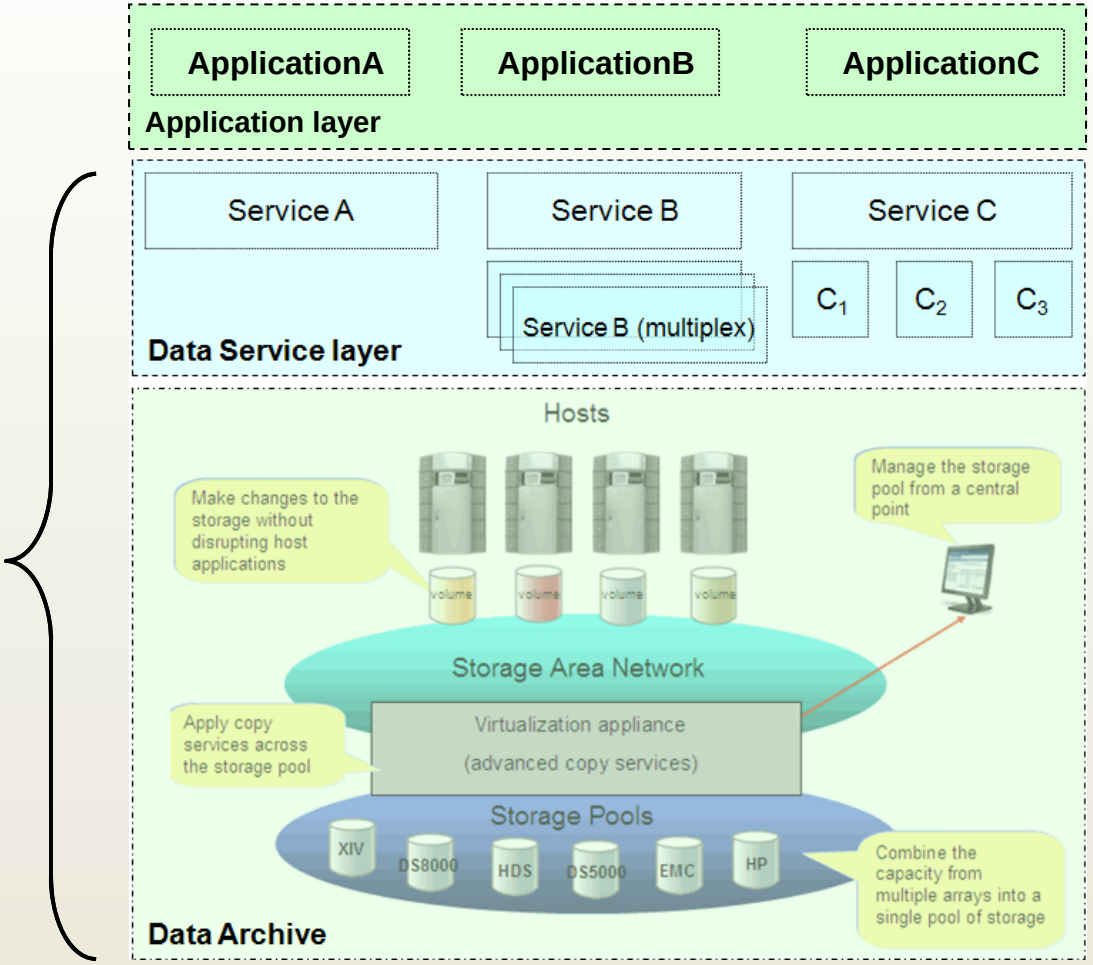
- *Design of the integrated platform*
- *Identification of the set of services that meet the project needs*
- *Test and validation activities*

Activities regarding the data storage

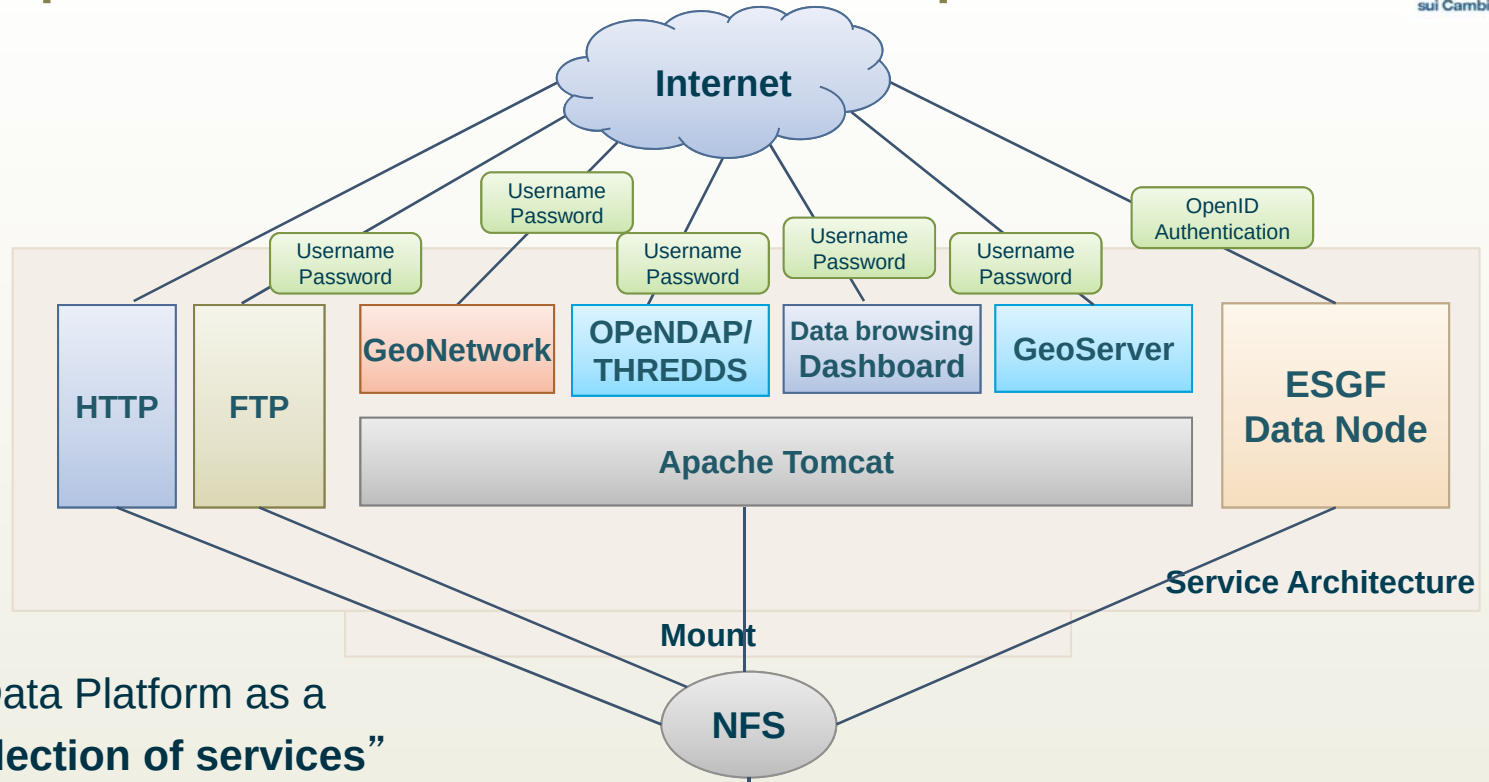
- *Definition of a directory structure*
- *Filename, datasetname encoding rules*
- *Preliminary tests and validation of the publication guidelines on some WP3, WP5 datasets.*

Design of ORIENTGATE data platform

Data Platform as a **"collection of services"**



Implementation of ORIENTGATE data platform



Data Platform as a
"**collection of services**"

orientgate2 Server Features
Nr core: 2 RAM: 4 GB SWAP: 4GB HD: 24 GB

Export

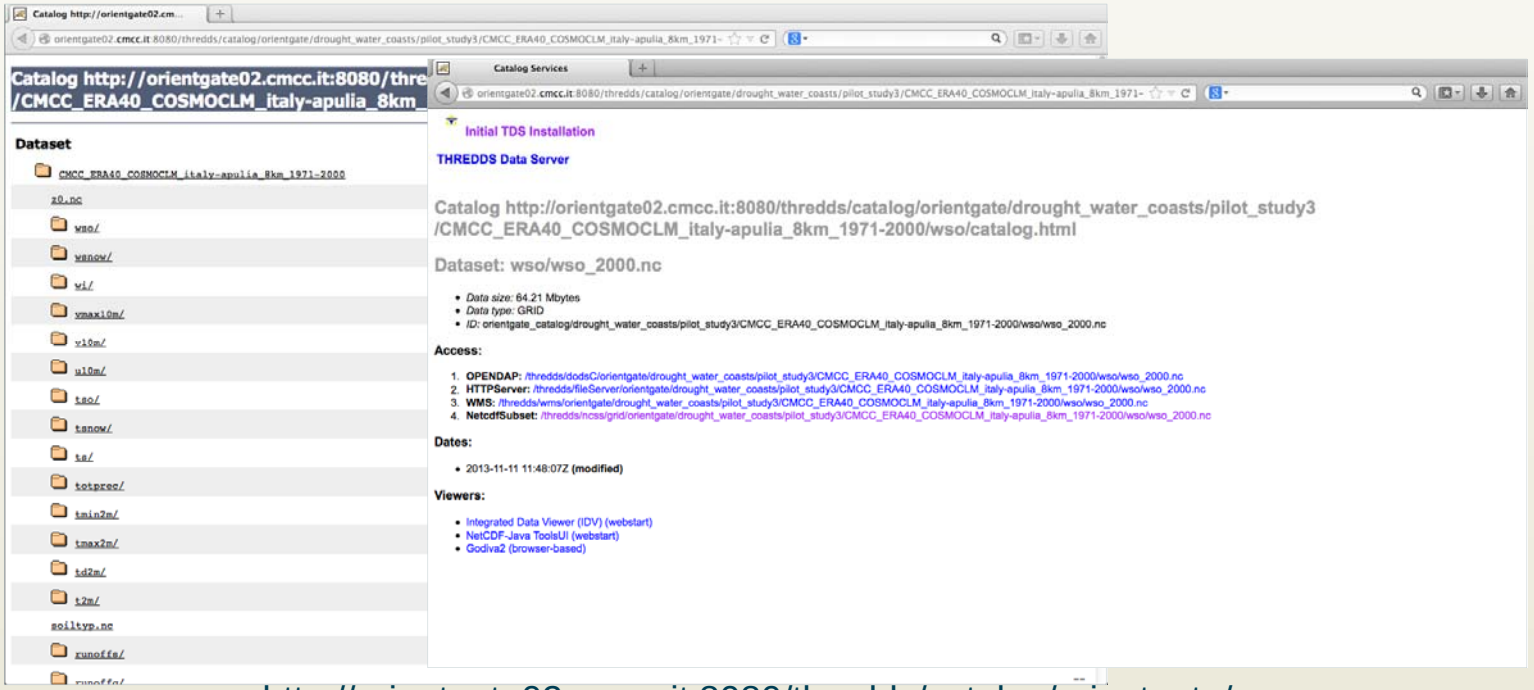
Orientgate Storage

ESGF node Server Features
Nr core: 2 RAM: 4 GB SWAP: 4GB HD: 24 GB

THREDDS (Thematic Realtime Environmental Distributed Data Services)

The THREDDS service aims at bridging the gap between data providers and data users. The goal is to simplify the discovery and use of scientific data and to allow scientific publications and educational materials to reference scientific data.

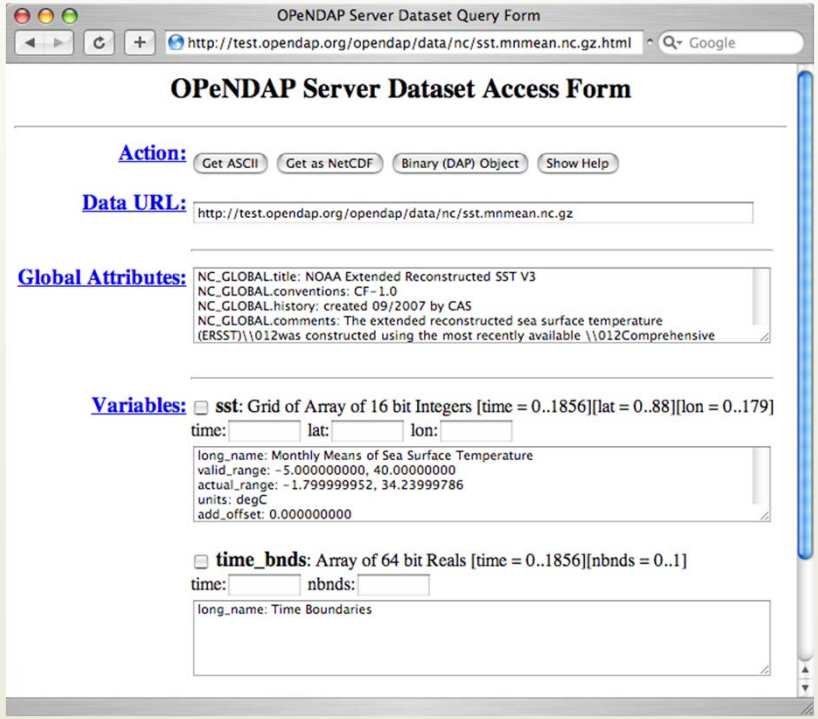
It's a web server providing features of metadata and data access using HTTP, OPeNDAP, WMS, NetCDF subset service.



http://orientgate02.cmcc.it:8080/thredds/catalog/orientgate/

OPeNDAP (Open-source Project for a Network Data Access Protocol)

OPeNDAP is a framework designed to easily share scientific data on web, making accessible local data from remote connections.



Features

- *Sharing of scientific data*
- *Remote access*
- *Different data formats*
- *Data subsetting*

Activities

- *Deploy, configuration and tuning on test VMs*
- *Test with compressed data*
- *Test on multiplexed configurations to increase throughput and fault tolerance*
- *First service available for test on a preliminary set of ORIENTGATE data*
- *Security added to protect data*
- *Logging enabled to keep track of accesses to the data/service.*

http://orientgate02.cmcc.it:8080/thredds/dodsC/orientgate/

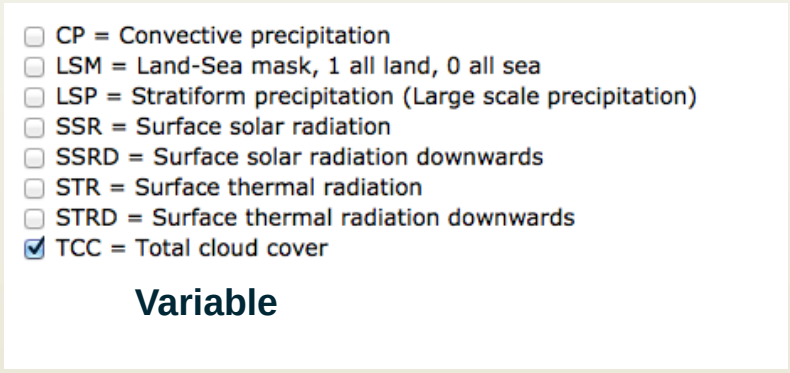
NetCDF Subset Service

Web service for subsetting CDM scientific dataset. The subsetting is specified using earth coordinates. The data arrays are subsetted but not resampled or reprojected, and preserve the resolution and accuracy of the original dataset.

Activities

- *Several actions have been carried out to **install** the software and **test** the **robustness** as well as the **capabilities** provided by this service*
- *Specific settings have been setup on the thredds service to enable and perform some tuning*

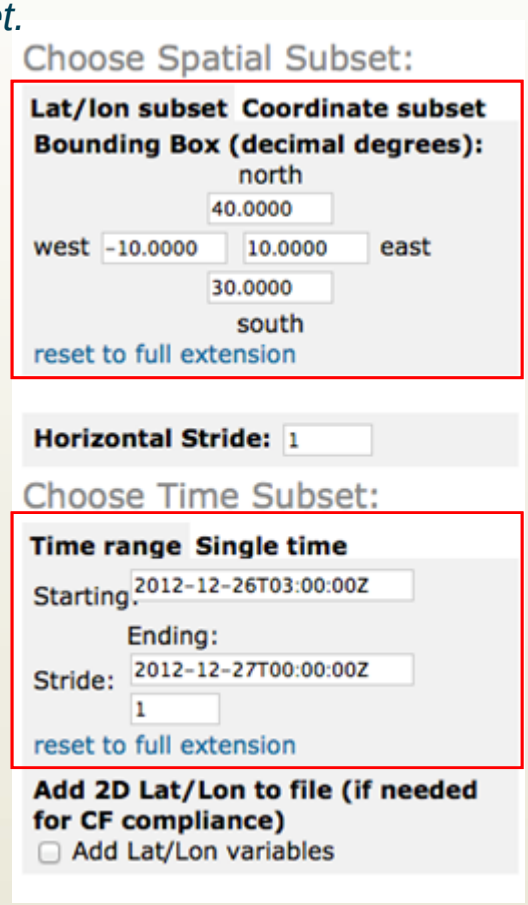
GeoQuery := (GeoBox, Variable, Date)



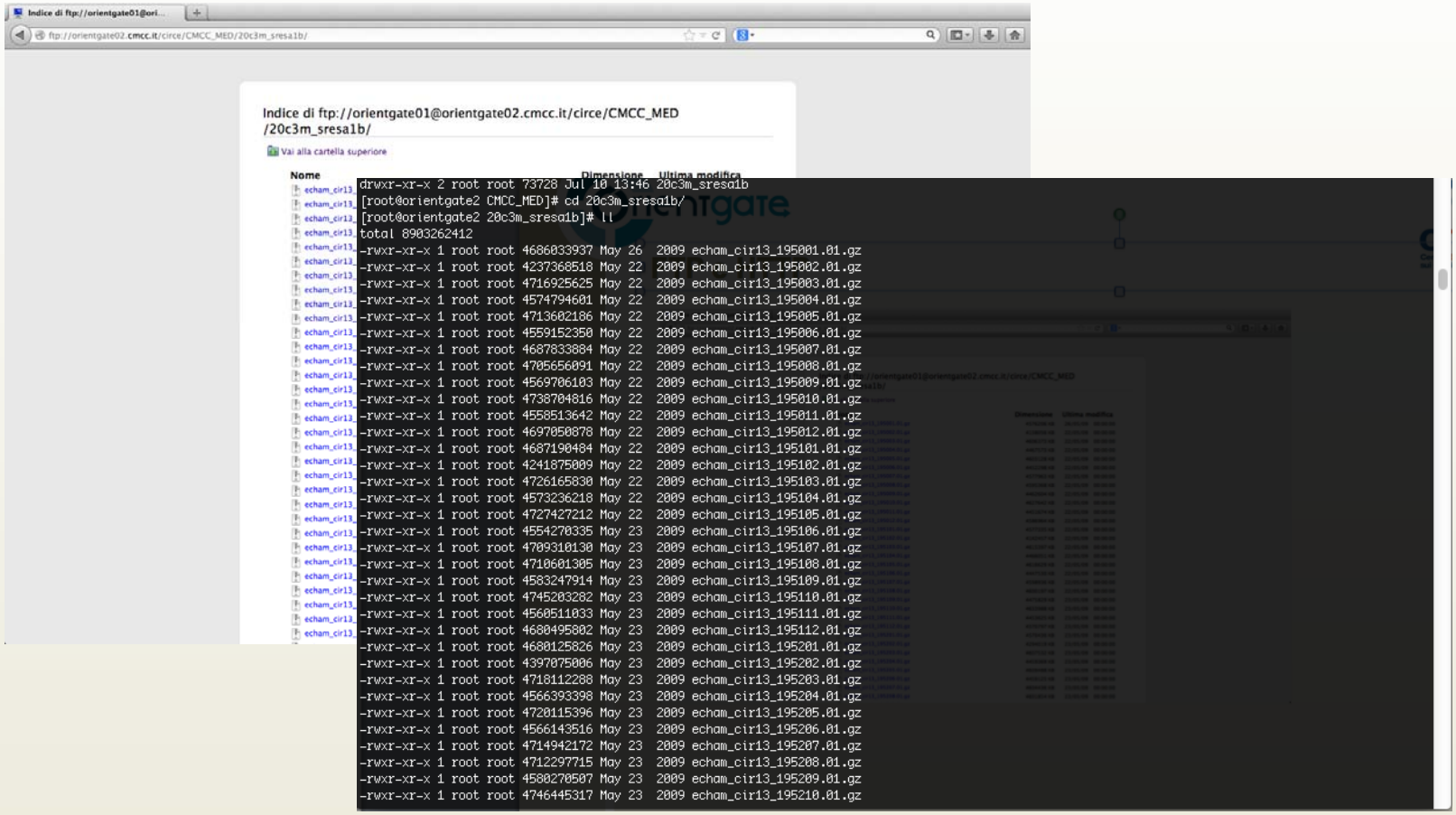
Variable

GeoBox

Temporal extent



FTP and HTTP


Indice di ftp://orientgate01@orientgate02.cmcc.it/circe/CMCC_MED
/20c3m_sresa1b/
Vai alla cartella superiore
Nome
drwxr-xr-x 2 root root 73728 Jul 10 13:46 20c3m_sresa1b
[root@orientgate2 CMCC_MED]# cd 20c3m_sresa1b/
[root@orientgate2 20c3m_sresa1b]# ll
total 8903262412
-rwxr-xr-x 1 root root 4686033937 May 26 2009 echam_cir13_195001.01.gz
-rwxr-xr-x 1 root root 4237368518 May 22 2009 echam_cir13_195002.01.gz
-rwxr-xr-x 1 root root 4716925625 May 22 2009 echam_cir13_195003.01.gz
-rwxr-xr-x 1 root root 4574794601 May 22 2009 echam_cir13_195004.01.gz
-rwxr-xr-x 1 root root 4713602186 May 22 2009 echam_cir13_195005.01.gz
-rwxr-xr-x 1 root root 4559152350 May 22 2009 echam_cir13_195006.01.gz
-rwxr-xr-x 1 root root 4687833884 May 22 2009 echam_cir13_195007.01.gz
-rwxr-xr-x 1 root root 4705656091 May 22 2009 echam_cir13_195008.01.gz
-rwxr-xr-x 1 root root 4569706103 May 22 2009 echam_cir13_195009.01.gz
-rwxr-xr-x 1 root root 4738704816 May 22 2009 echam_cir13_195010.01.gz
-rwxr-xr-x 1 root root 4558513642 May 22 2009 echam_cir13_195011.01.gz
-rwxr-xr-x 1 root root 4697050878 May 22 2009 echam_cir13_195012.01.gz
-rwxr-xr-x 1 root root 4687190484 May 22 2009 echam_cir13_195101.01.gz
-rwxr-xr-x 1 root root 4241875009 May 22 2009 echam_cir13_195102.01.gz
-rwxr-xr-x 1 root root 4726165830 May 22 2009 echam_cir13_195103.01.gz
-rwxr-xr-x 1 root root 4573236218 May 22 2009 echam_cir13_195104.01.gz
-rwxr-xr-x 1 root root 4727427212 May 22 2009 echam_cir13_195105.01.gz
-rwxr-xr-x 1 root root 4554270335 May 23 2009 echam_cir13_195106.01.gz
-rwxr-xr-x 1 root root 4709310130 May 23 2009 echam_cir13_195107.01.gz
-rwxr-xr-x 1 root root 4710601305 May 23 2009 echam_cir13_195108.01.gz
-rwxr-xr-x 1 root root 4583247914 May 23 2009 echam_cir13_195109.01.gz
-rwxr-xr-x 1 root root 4745203282 May 23 2009 echam_cir13_195110.01.gz
-rwxr-xr-x 1 root root 4560511033 May 23 2009 echam_cir13_195111.01.gz
-rwxr-xr-x 1 root root 4680495802 May 23 2009 echam_cir13_195112.01.gz
-rwxr-xr-x 1 root root 4680125826 May 23 2009 echam_cir13_195201.01.gz
-rwxr-xr-x 1 root root 4397075006 May 23 2009 echam_cir13_195202.01.gz
-rwxr-xr-x 1 root root 4718112288 May 23 2009 echam_cir13_195203.01.gz
-rwxr-xr-x 1 root root 4566393398 May 23 2009 echam_cir13_195204.01.gz
-rwxr-xr-x 1 root root 4720115396 May 23 2009 echam_cir13_195205.01.gz
-rwxr-xr-x 1 root root 4566143516 May 23 2009 echam_cir13_195206.01.gz
-rwxr-xr-x 1 root root 4714942172 May 23 2009 echam_cir13_195207.01.gz
-rwxr-xr-x 1 root root 4712297715 May 23 2009 echam_cir13_195208.01.gz
-rwxr-xr-x 1 root root 4580270507 May 23 2009 echam_cir13_195209.01.gz
-rwxr-xr-x 1 root root 4746445317 May 23 2009 echam_cir13_195210.01.gz

GeoNetwork

Several testing activities are been carried out:

As ADMINISTRATOR:

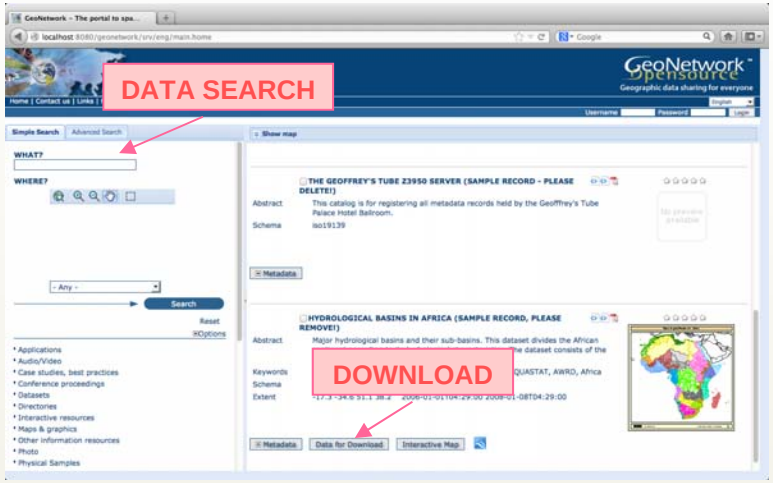
- *Installation of the platform*
- *Users and Group management*

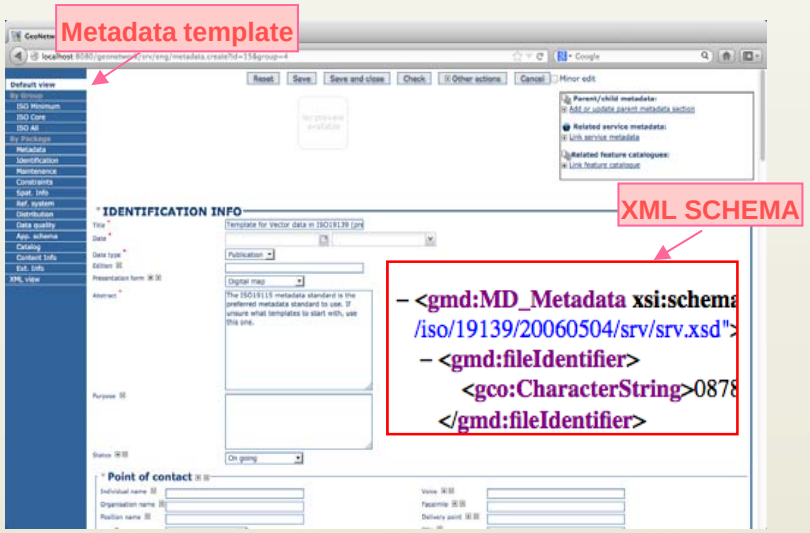
As USER:

- *Search&Discovery in multiple catalogs through a website*
- *Access to interactive maps*
- *Data download: depending on the privileges that have been set for each record, the dataset is available and downloadable*
- *Support for multiple metadata standards*

As DATA PROVIDER:

- *Metadata editing tool*
- *XML metadata import*
- *Set different sharing levels*





GeoServer

GeoServer is an open-source software allowing users to share, process and edit geospatial data.
It allows data publication data from any major spatial data source using open standards, such as:

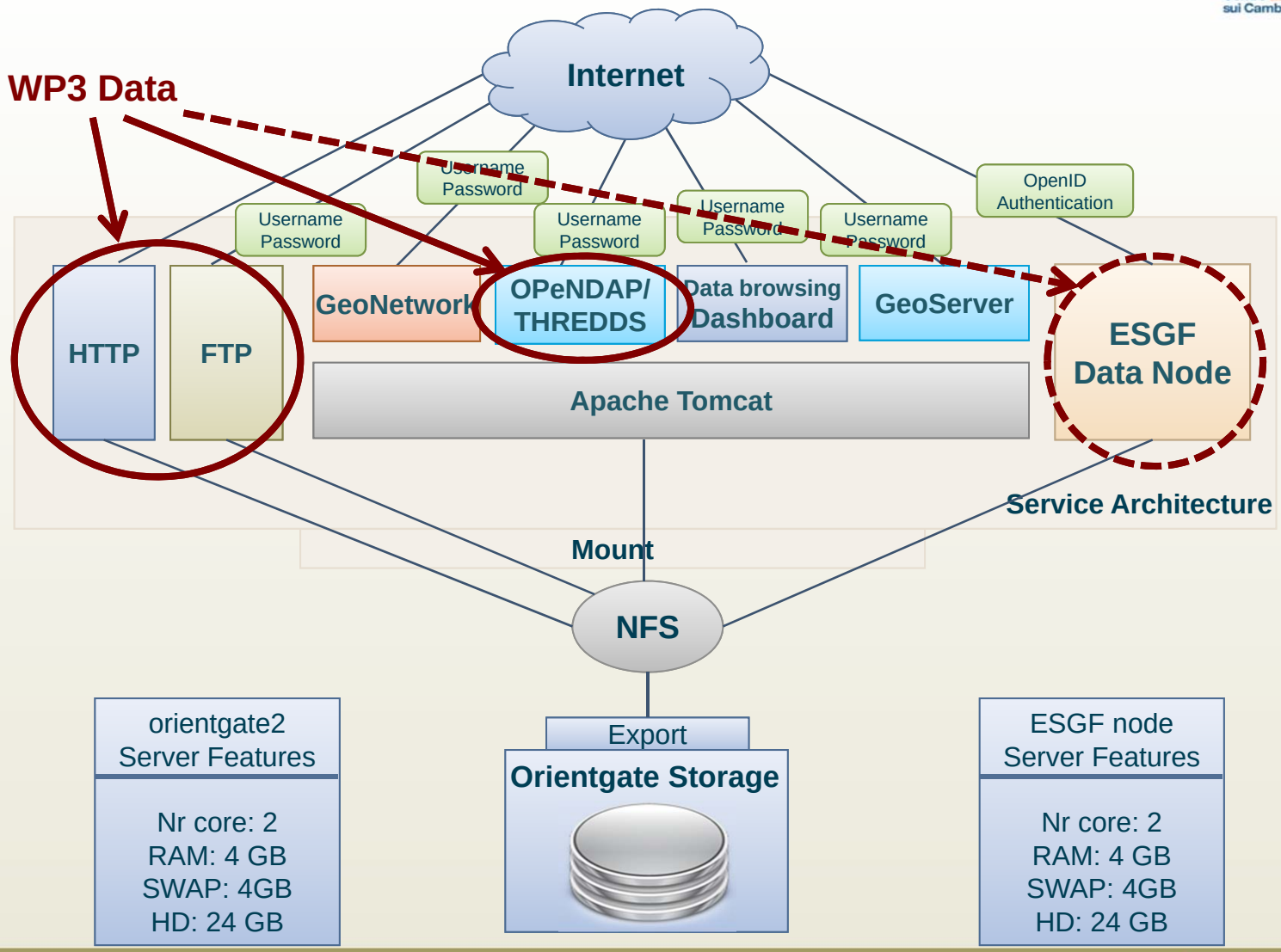
- Web Map Service (**WMS**) allows for requests of images generated from geographical data.
- Web Feature Service (**WFS**) supports requests of geographical feature data (vectors)
- Web Coverage Service (**WCS**) supports requests for coverage data (rasters)

These standards allow web clients to query and receive geographic information in the form of image, vector, or coverage data.

GeoServer

- GeoServer reads a variety of data formats, including:
  - ❑ Shapefile
  - ❑ GeoTIFF
  - ❑ GTOPO30
  - ❑ ECW, MrSID
  - ❑ JPEG2000
  - ❑ Post GIS
  - ❑ MySQL
  - ❑ DB2
  - ❑ ArcSDE
  - ❑ Oracle Spatial
- Output formats: KML, GML, Shapefile, GeoRSS, PDF, GeoJSON, JPEG, GIF, SVG, PNG and other more formats.
- Integrated OpenLayers client for previewing data layers.
- Efficient publishing of geospatial data to Google Earth, using KML language.

ORIENTGATE Data Platform – services for WP3 datasets



Publication of a WP3 datasets

The example dataset:

- *is produced by **Republic hydrometeorological Service of Serbia**;*
- *the **model** is **NMMB;***
- *the **forcing** used during the experiment is **ERA40**;*
- *refers to the **Balkan** area;*
- *the **resolution** of the data is **8 km**;*
- *the **temporal subset** goes from 1971 to 2000.*

Directory structure and datasetname encoding

The datasetname encoding will adhere to the following convention:

institutename_ forcinginfo_ modelinginfo_ geographicalinfo_ resolution_temporalsubset

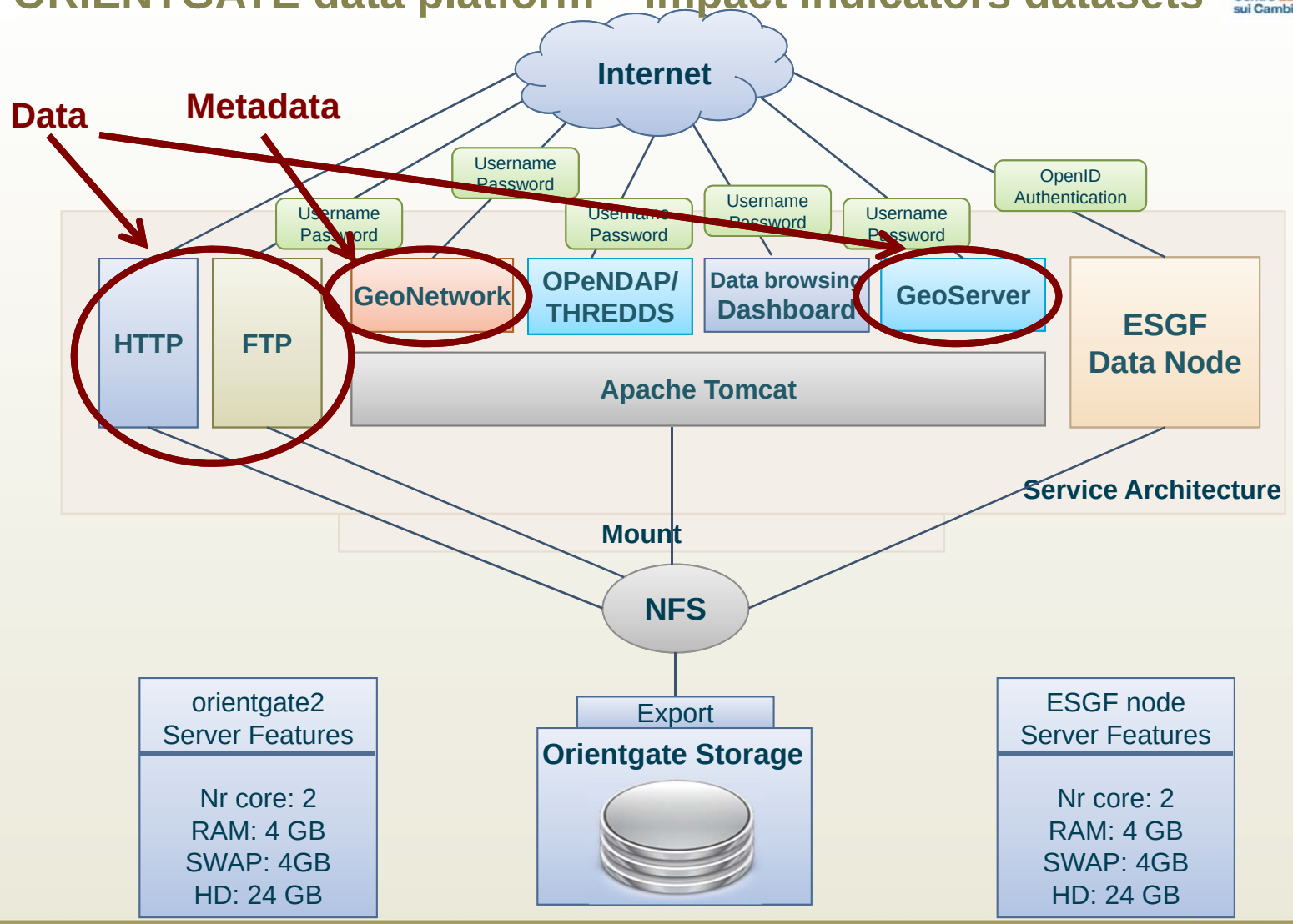
where:

- ***institutename*** represents the name of the institute producing the data
- ***forcinginfo*** indicates the forcing used during the experiment
- ***modelinginfo*** provides information about the used model and a possible scenario
- ***geographicalinfo*** indicates the geographic area
- ***resolution*** indicates the resolution of the produced data
- ***temporalsubset*** is the temporal period of the data

Example of dataset:

RHMSS_ERA40_ NMMB_ Balkan_8km_1971-2000

ORIENTGATE data platform – Impact indicators datasets



Publication of a WP5 dataset

***The first WP5 dataset published/tested on the data platform**:*

- *is produced by **Pilot Study** 3, within the **Thematic Centre** 2*
- *the **category_framework** of the indicator is "hazard_UN-DRR"*
- *refers to a **single indicator**.*

*The **identifier** of the indicator is SDI, the **time frequency** is 30 years, the **spatial resolution** is 350000 (meaning 1:350000 scale) and the **temporal subset** refers to the 1971-2000 period.*

The entire period of the dataset ranges from 1971 to 2070.

*The corresponding file is a shapefile**.***

Directory structure and datasetname encoding

- The ***directory structure*** is been already setup on the platform and with the information the data producers will provide we will put the data under the corresponding folder.
- The ***datasetname encoding*** adhers to the following convention:

<indicator_identifier>_<time_frequency>_<spatial_resolution>_<temporal_subset>

The datasetname will be composed by:

- an identifier of the indicator (***indicator_identifier***)
- the time frequency (***time_frequency***);
- the spatial resolution (***spatial_resolution***);
- a time interval (***temporal_subset***).

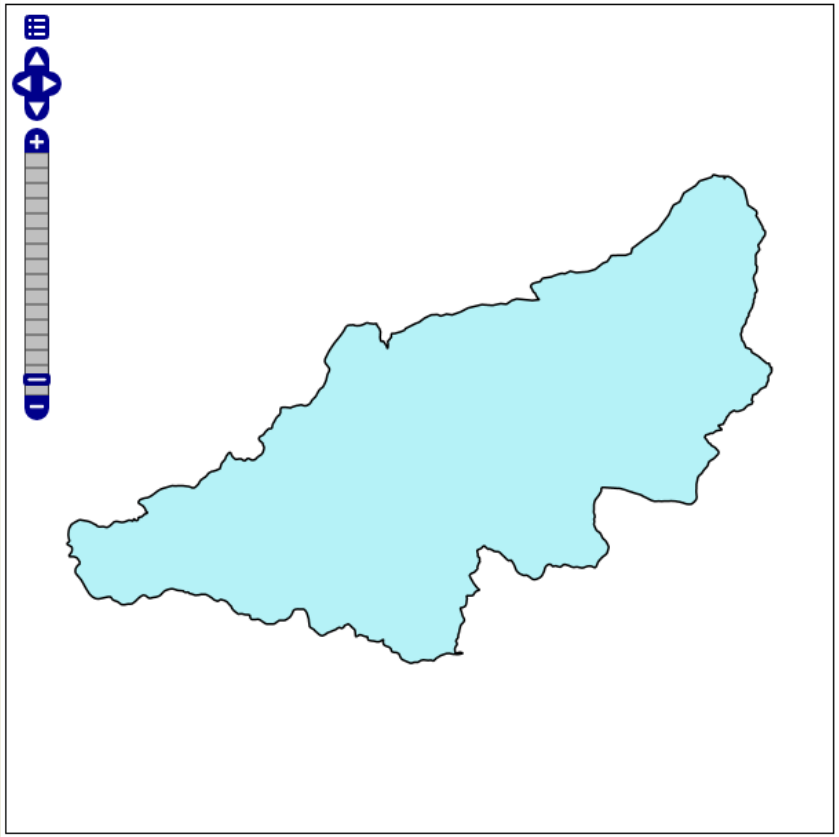
Example:

SDI_ 30y_ 350000_19712000

Visualization of a WP5 dataset (shapefile)

WMS version: 1.1.1 Tiling: Single tile Transition effect: None Antialias: Full Format: PNG 24bit Styles: Default Width/Height: Auto Auto

Filter: CQL



Scale = 1 : 697K 2609317.51089, 4551898.77290

SDI_30y_350000_19712000

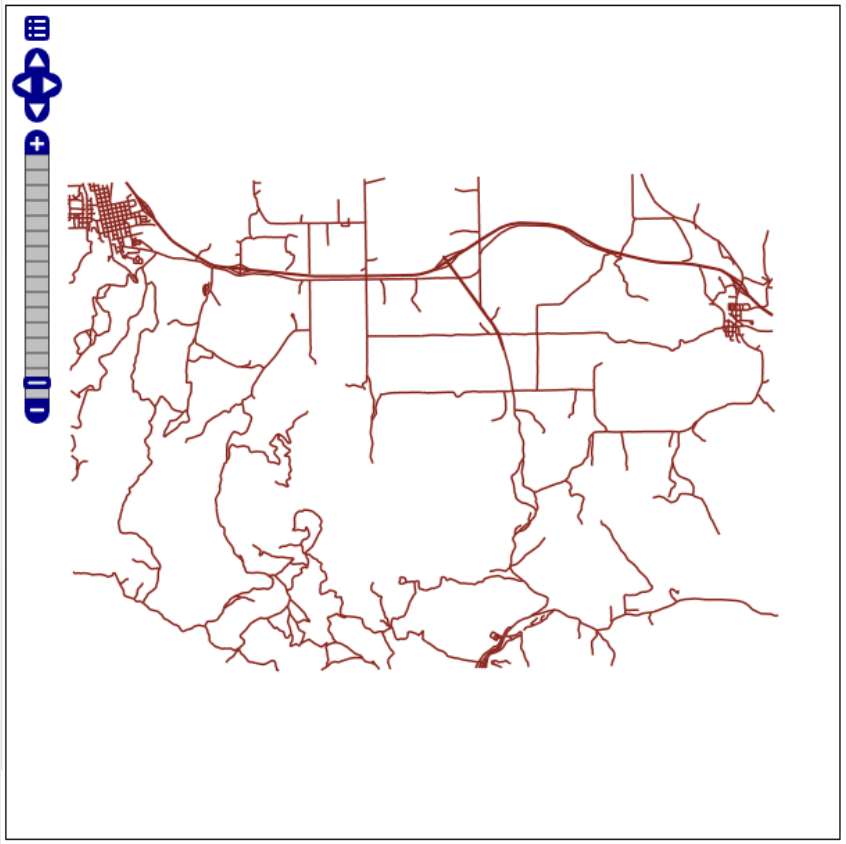
fid	NOME	SDI_1	SDI_2	SDI_3	SDI_4	SDI1_trend	SDI2_trend	SDI3_trend	SDI4_trend
SDI_30y_350000_19712000.1	Ofanto	-0.033	-0.01	-0.018	-0.017	-0.029	-0.058	-0.061	-0.063

Polygon Shapefile

Other examples (shapefile)

WMS version: 1.1.1 Tiling: Single tile Transition effect: None Antialias: Full Format: PNG 24bit Styles: Default Width/Height: Auto Auto

Filter: CQL



Scale = 1 : 140K 604543.36466, 4920762.91425

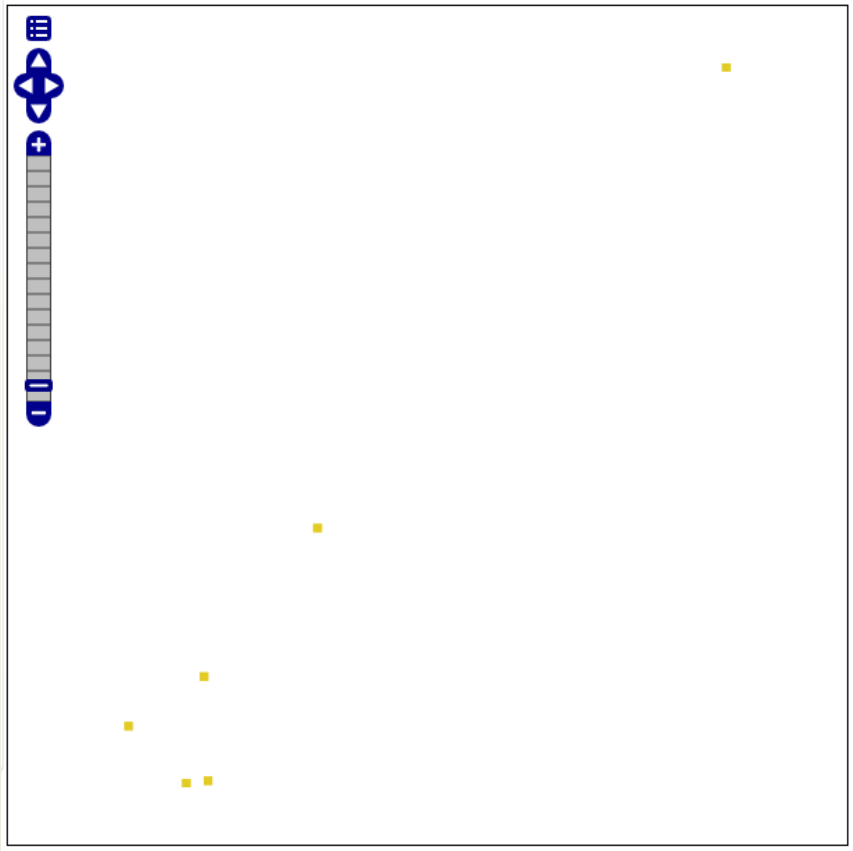
roads

fid	cat	label
roads.532	5	unimproved road
roads.555	4	light-duty road, improved surface

Polyline Shapefile

Data from www.naturalearthdata.com

Other examples (shapefile)



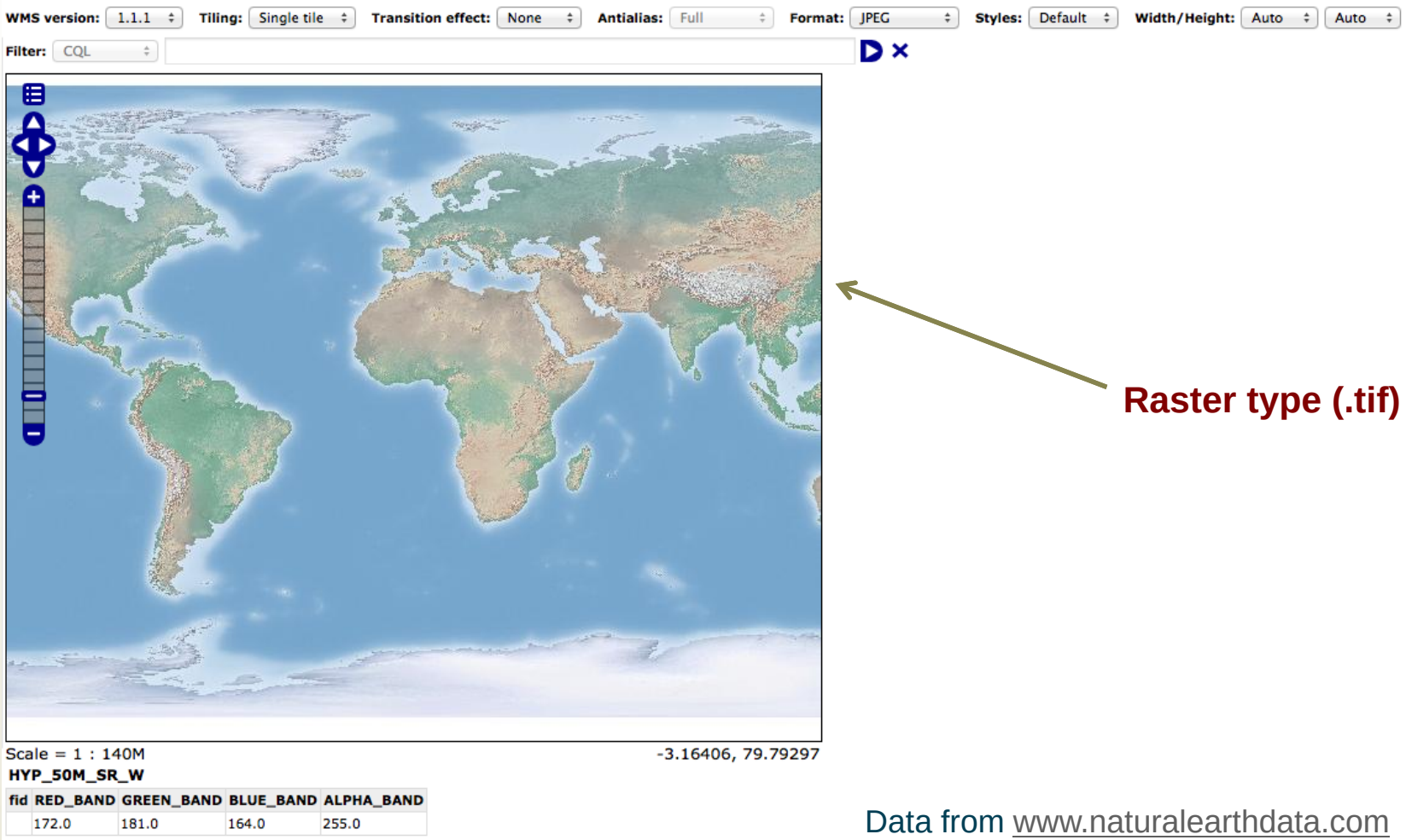
poi

fid	NAME	THUMBNAIL	MAINPAGE
poi.1	museam	pics/22037827-Ti.jpg	pics/22037827-L.jpg

Point Shapefile

Data from www.naturalearthdata.com

Other examples (raster)

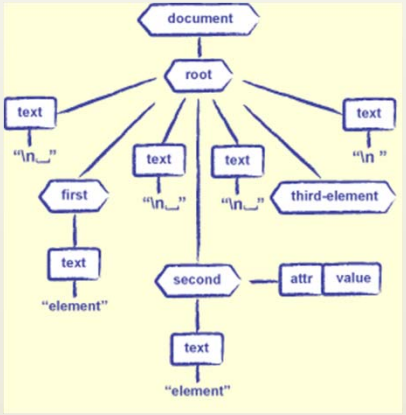


Raster type (.tif)

Data from www.naturalearthdata.com

Not only data… the Metadata concept

- *Structured information about the data*
- *Metadata are "data (information) about data (raw data)"*
- *Metadata describe the content, the quality and other features of the data (e.g. origin of the data, citations, abstracts, scope, credits, state and point of contact).*
- *Helps to search and discovery data, to better organize and use them and enables interoperability*
- *The adoption of standards allows the interoperablity and the possibility to interact with other project and share and compare results*

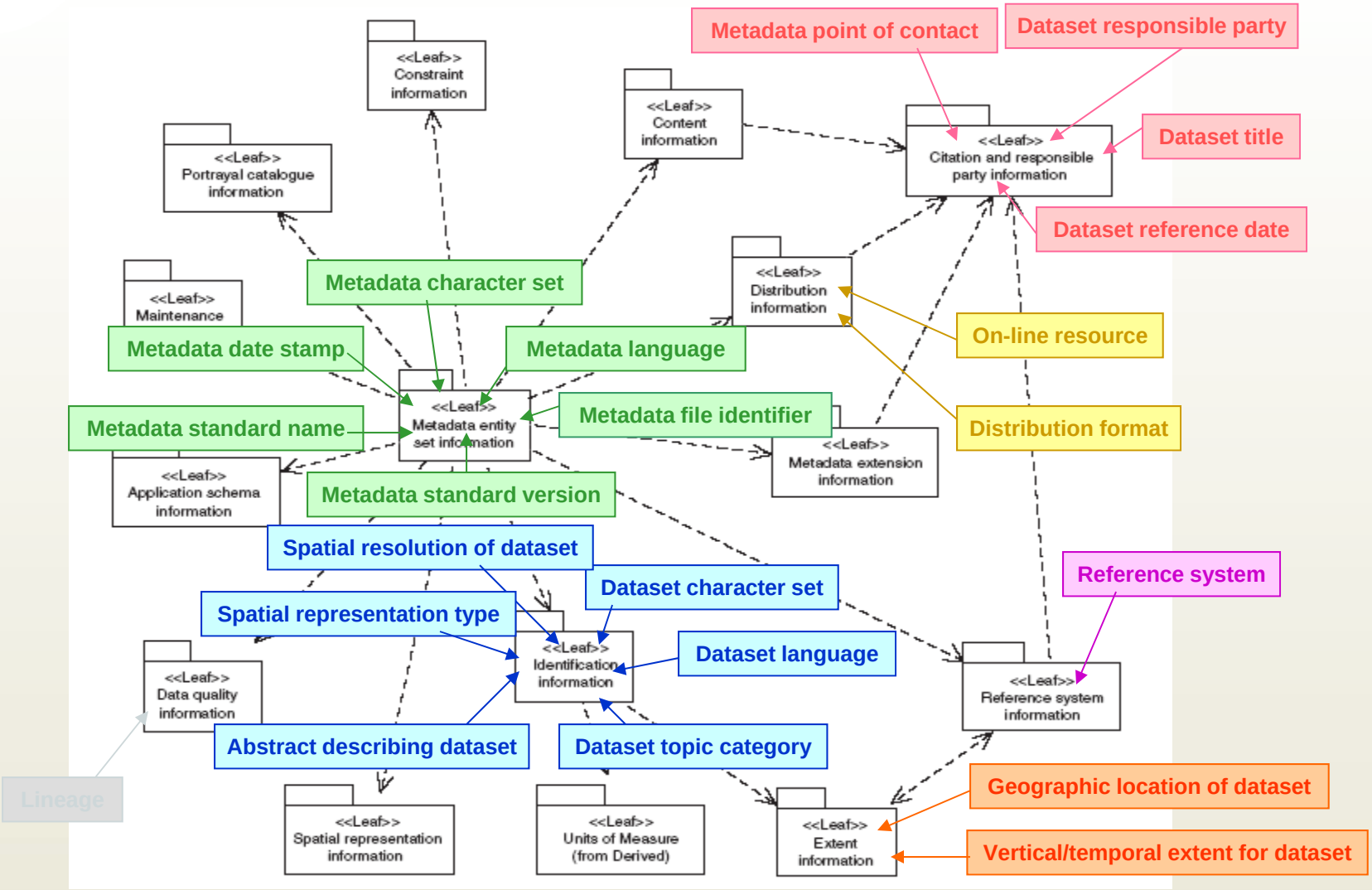


SCHEMA

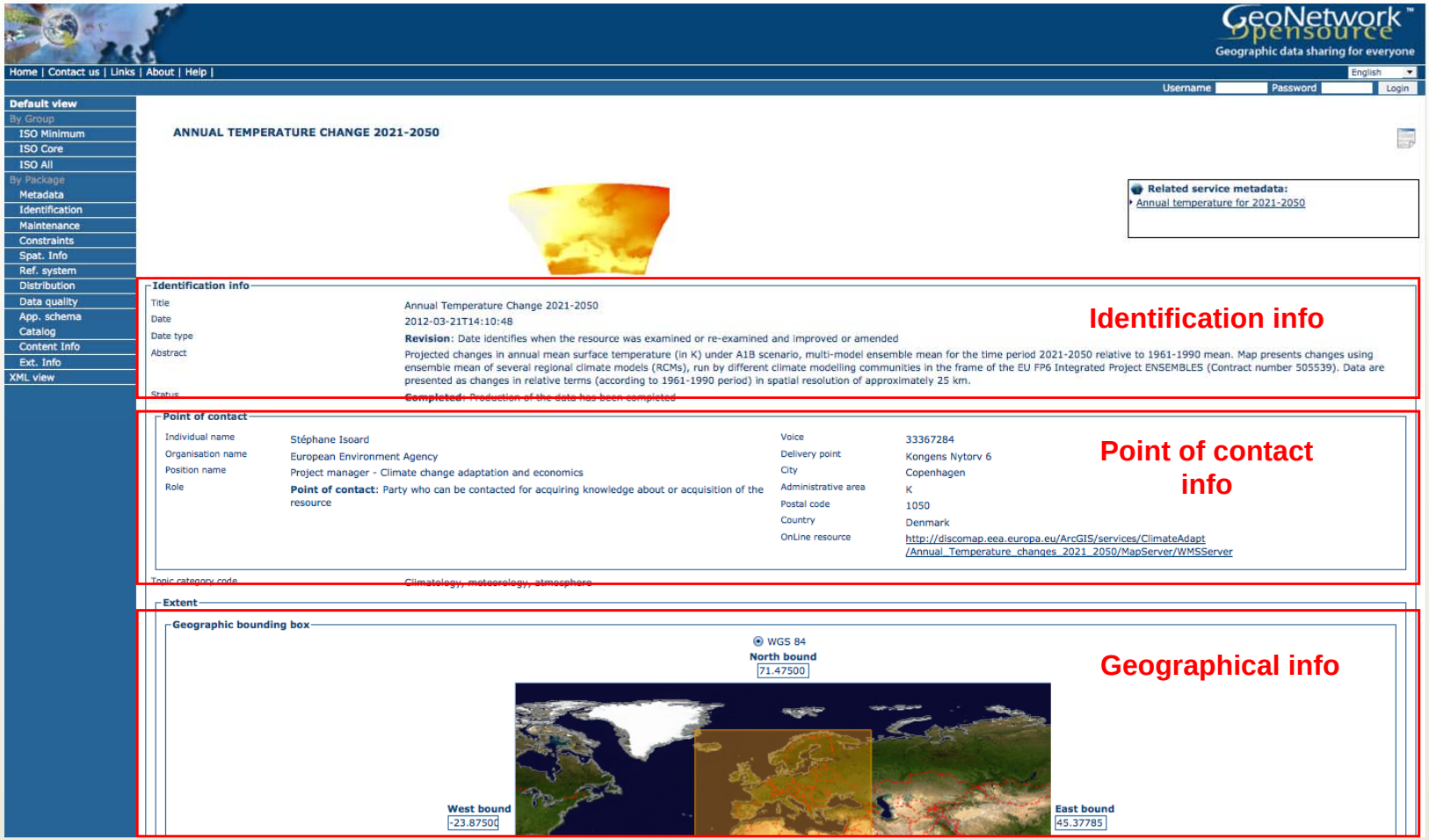
```
<METADATA>
    <AUTHOR>John Smith</AUTHOR>
    <DATE>11-06-2009</DATE>
    <DATA>
        <FILEN>simulation.nc</FILE>
        <KEYWORD>ocean</KEYWORD>
    </DATA>
    ....
</METADATA>
```

XML Documents

ISO19115 Schema



Metadata for dataset related to impact indicators (1)



Metadata for dataset related to impact indicators (2)

South bound
29.12312

Distribution Information

Transfer options

Interactive Map	Annual Precipitation Change 2021-2050 (OGC-WMS Server: http://discomap.eea.europa.eu/ArcGIS/services/ClimateAdapt/Annual_Precipitation_changes_2021_2050/MapServer/WMSServer)
View in Google Earth	Annual Precipitation Change 2021-2050
OnLine resource	1 (LegendURL)

Distribution info

Reference System Information

Code	EPSG:4326

Reference system info

Reference System Information

Code	EPSG:3857

Data quality info

Hierarchy level	**Dataset**: Information applies to the dataset

Data quality info

Metadata

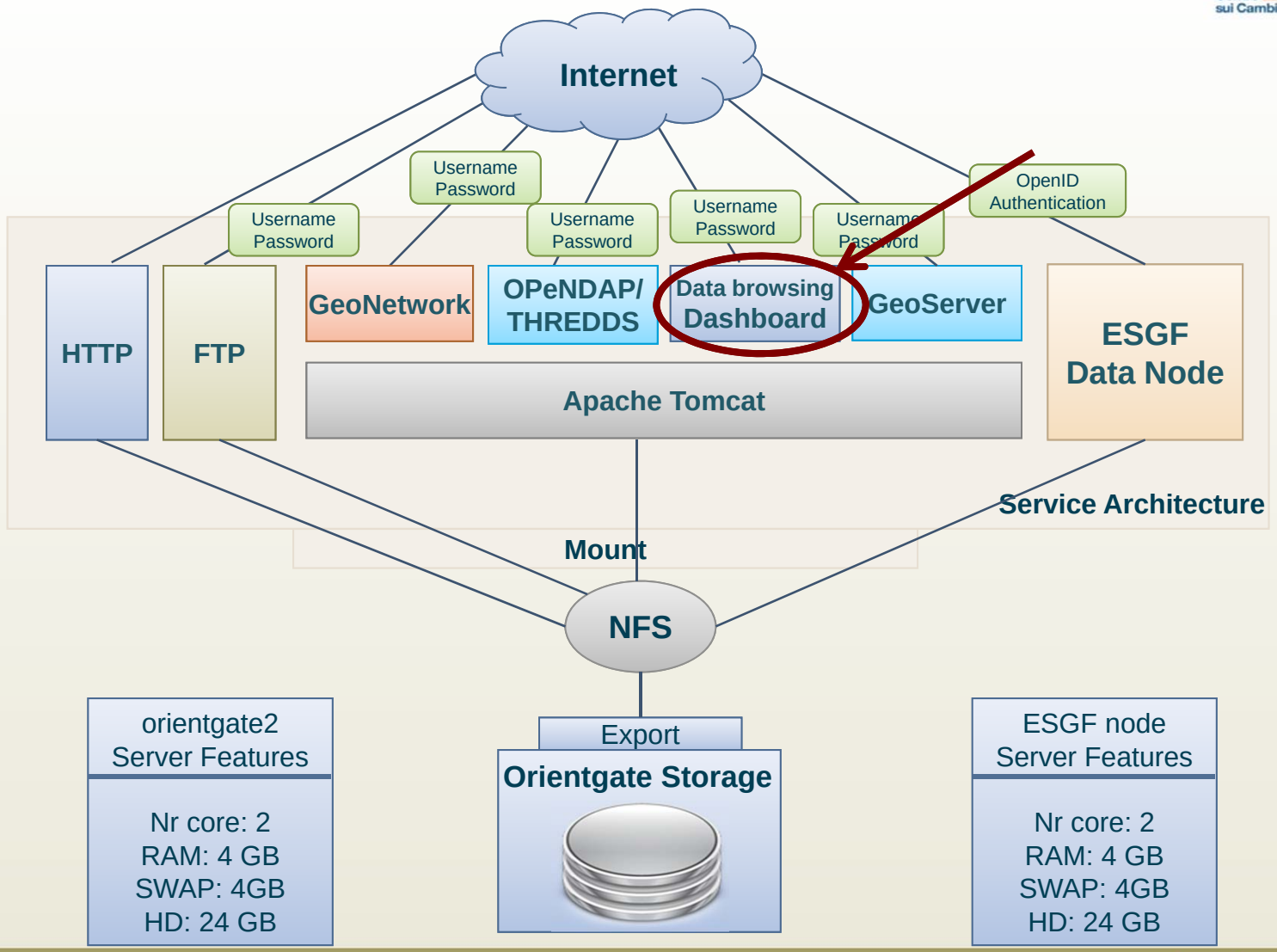
File identifier	f921b4af568c49668dedf9942789292b8ba85415
Metadata language	English
Character set	**UTF8**: 8-bit variable size UCS Transfer Format, based on ISO/IEC 10646
Hierarchy level	**Dataset**: Information applies to the dataset
Date stamp	2012-11-14T10:46:43
Metadata standard name	ISO 19115:2003/19139
Metadata standard version	1.0

Metadata info

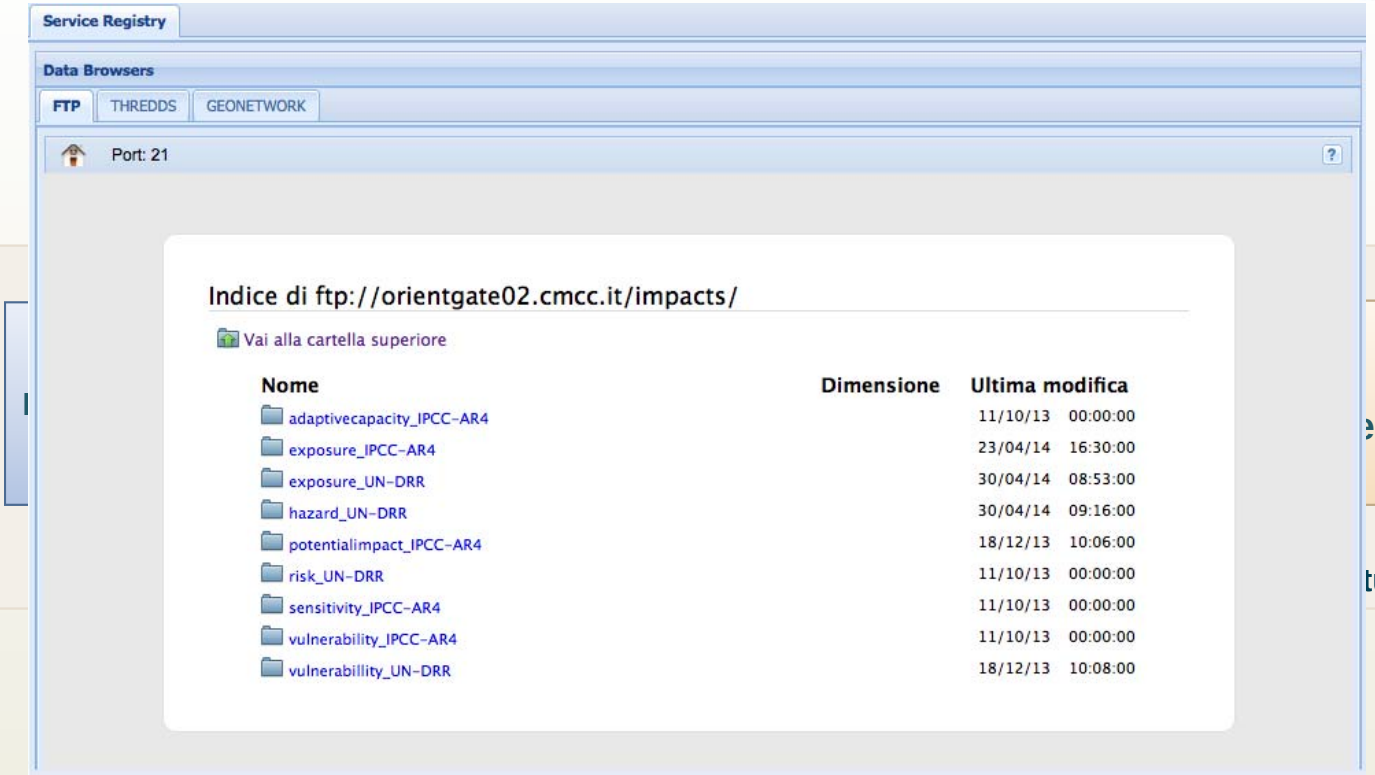
Contact

Individual name	Stéphane Isoard	Voice	33367284
Organisation name	European Environment Agency	Facsimile	-
Position name	Project manager - Climate change adaptation and economics	Delivery point	Kongens Nytorv 6
Role	**Point of contact**: Party who can be contacted for acquiring knowledge about or acquisition of the resource	City	Copenhagen
		Administrative area	K
		Postal code	1050
		Country	Denmark
		OnLine resource	http://discomap.eea.europa.eu/ArcGIS/services/ClimateAdapt/Annual_Precipitation_changes_2021_2050/MapServer/WMSServer

ORIENTGATE data platform – Dashboard gadget



ORIENTGATE data platform Interface



Access through the data platform gadget

Dashboard gadget and integrated into the ORIENTGATE website

Thank you!

Alessandra Nuzzo (alessandra.nuzzo@cmcc.it)
Sandro Fiore (sandro.fiore@unisalento.it)
Giovanni Aloisio (giovanni.aloisio@unisalento.it)