EarthServer-2

654367 EarthServer-2
Agile Analytics on Big Data Cubes
H2020 INFRA-1-2014 topic (7)

Deliverable: D 2.2

Data Management Plan

Nature of deliverable: R
Dissemination level: PU
Scheduled Delivery Data: 31st August 2015
Status: Final

Prepared by: **Peter Walker**
Verified by: **Stephan Siemen**
Approved by: **Angelo Rossi**

Table of Contents

Executive Summary

The objective of EarthServer-2 is to establish Agile Analytics on Petabyte Data Cubes as a simple, user-friendly, and scalable paradigm. EarthServer-2 adds to Copernicus/Sentinel data by creating a single 3D x/y/t datacube per product so that millions of images form a single, simple data space, irrespective of its size. Likewise, a weather simulation would form a single 4D datacube. Uniform mix-and-match access on spatio-temporal sensor, image, simulation, and statistics data will make any-size space/time data cubes a commodity for non-experts and experts alike.

This document will set out the Data Management Plan (DMP) for the EarthServer-2 project. The DMP will include a data register which will be maintained as a "live" document and updated as new data become available to the project. Periodic iterations of the DMP will be produced, providing a snapshot of the latest data register.
Data management is a major focus of the EarthServer-2 project which addresses problems such as searching, filtering and analyzing huge (100TB+) datasets that, due to their size, cannot be processed on the client side. The tools and standards developed during EarthServer-2 will improve the accessibility and reuse of these datasets.

List of acronyms and abbreviations

C3S	Copernicus Climate Change Service
CDS	C3S's Climate Data Store
CSDS	Climate Science Data Service
CITE	Communication & Information Technologies Experts S.A.
DoA	Description of Action
DMP	Data Management Plan
EODS	Earth Observation Data Service
GIS	Geographical Information Systems
GRIB	GRIdded Binary format mandated by the WMO
EC	European Commission
ECMWF	European Centre for Medium-Range Weather Forecasts
EFAS	European Flood Awareness System
ESA	European Space Agency
ESAC	European Space Astronomy Centre
ESFRI	European Strategy Forum on Research Infrastructures
FP7	EU funding programme for 2007-2013
H2020	Horizon 2020 - EU Research and Innovation programme
HCMR	Hellenic Centre for Marine Research
HDF	Hierarchical Data Format
JUB	Jacobs University Bremen
LDCS	Landsat Data Cube Service
MARS	Meteorological Archive and Retrieval System
MEEO	Meteorological and Environmental Earth Observation S.r.l.
MetOcean DWG	Meteorological & Oceanographic Domain Working Group of the OGC
MSDS	Marine Science Data Service
NASA	National Aeronautics and Space Administration of the United States
NCI	National Computational Infrastructure of Australia
NetCDF	Network Common Data Form
OGC	Open Geospatial Consortium
PML	Plymouth Marine Laboratory
PSDS	Planetary Science Data Service
WCS	Web Coverage Service
WCPS	Web Coverage Processing Service
WMO	World Meteorological Organisation
WMS	Web Map Services

1 Introduction

The EarthServer-2 project is itself built around concepts of data management and accessibility. Its aim is to implement enabling technologies to make large datasets accessible to a varied community of users. The intention is not to create new datasets but to make existing datasets (identified at the start of the project) easier to access and manipulate, encouraging data sharing and reuse. Additional datasets will be added during the life of the project as they become available and the DMP will be updated as a "live" document to reflect this.

2 Data Organisation, Documentation and Metadata

Data will be accessible through the Open Geospatial Consortium (OGC) Web Coverage Processing Service[1] (WCPS) and Web Coverage Service[2] (WCS) standards. EarthServer-2 will establish data/metadata integration on a conceptual level (by integrating array queries with known metadata search techniques such as tabular search, full text search, ontologies etc.) and on a practical level (by utilizing this integrated technology for concrete catalogue implementations based on standards like ISO 19115, ISO 19119 and ISO 19139 depending on the individual service partner needs).

3 Data Access and Intellectual Property

Data access restrictions and intellectual property rights will remain as set by the dataset owners (see Section 6). The datasets identified for the initial release have no access restrictions.

4 Data Sharing and Reuse

The aim of EarthServer-2 is to make data available for sharing and reuse without requiring that users download the entire (huge) dataset. Data will be available through the OGC WCPS and WCS standard, allowing users to filter and process data at source before transferring them back to the client. Access will be simplified by the provision of data services (Marine, Climate, Earth Observation, Planetary and Landsat) that will web portals with a user friendly interface to filtering and analysis tools as required by the application domain.

5 Data Preservation and Archiving

EarthServer-2 will not generate new data; preservation and archiving will be the responsibility of the upstream projects from which the original data was obtained.

---

[1] http://www.opengeospatial.org/standards/wcps
[2] http://www.opengeospatial.org/standards/wcs

6 Data Register

The data register will be maintained as a "live" document; a snapshot will be created for each DMP release (see 6.1 and following sections).

The data register will be based upon information and restrictions supplied by the upstream data provider matched to Horizon 2020 guidelines as below (in *italics)*:

- **Data set reference and name**
  Identifier for the data set to be produced.

- **Data set description**
  Descriptions of the data that will be generated or collected, its origin (in case it is collected), nature and scale and to whom it could be useful, and whether it underpins a scientific publication. Information on the existence (or not) of similar data and the possibilities for integration and reuse.
  - *Standards and metadata*
    Reference to existing suitable standards of the discipline. If these do not exist, an outline on how and what metadata will be created.
  - *Data sharing*
    Description of how data will be shared, including access procedures, embargo periods (if any), outlines of technical mechanisms for dissemination and necessary software and other tools for enabling re-use, and definition of whether access will be widely open or restricted to specific groups. Identification of the repository where data will be stored, if already existing and identified, indicating in particular the type of repository (institutional, standard repository for the discipline, etc.). In case the dataset cannot be shared, the reasons for this should be mentioned (e.g. ethical, rules of personal data, intellectual property, commercial, privacy-related, security-related).

- **Archiving and preservation (including storage and backup)**
  Description of the procedures that will be put in place for long-term preservation of the data. Indication of how long the data should be preserved, what is its approximated end volume, what the associated costs are and how these are planned to be covered.
  Within EarthServer-2 currently, the original data are held by upstream providers who have their own policies. In this case archiving and preservation responsibility will remain with the upstream project.

6.1   Marine Science Data Service

Data set reference and name	ESA OC-CCI v2
Organisation	**ESA**
Data set description	The ESA Climate Change Initiative (CCI) programme is generating a set of validated, error characterised, Essential Climate Variables (ECVs) from existing satellite observations. The Ocean Colour ECV is providing ocean colour data, with a focus on Case 1 waters, which can be used by climate change prediction and assessment models. The dataset is created by band-shifting and bias-correcting MERIS and MODIS data to match SeaWiFS data, merging the datasets and computing per-pixel uncertainty estimates. See http://www.esa-oceancolour-cci.org/?q=webfm_send/496 for full details.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	1981-2013
Project Contact	petwa@pml.ac.uk
Upstream Contact	help@esa-oceancolour-cci.org
Limitations	None
License	Free
Constraints	None
Data Format	NetCDF-CF
Access URL	TBD
Archiving and preservation (including storage and backup)	Data is part of long term ESA CCI project and the original copy is maintained there.

Table 1: Data set description for the MSDS.

6.2   Climate Science Data Service

Data set reference and name	ECMWF ERA Reanalysis
Organisation	**ECMWF**
Data set description	FP7 Era-Clim2
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	1900-2010
Project Contact	Stephan Siemen (ECMWF)
Upstream Contact	Dick Dee (ECMWF)
Limitations	None
License	Free, but no redistribution
Constraints	None
Data Format	GRIB
Access URL	http://apps.ecmwf.int/datasets/
Archiving and preservation (including storage and backup)	Stored in MARS archive - original data will be kept without time limit

Table 2: Data set description for the CSDS.

6.3   Earth Observation Data Service

Data set reference and name	MOD 04 - Aerosol Product; MOD 05 - Total Precipitable Water; MOD 06 - Cloud Product; MOD 07 - Atmospheric Profiles; MOD 35 - Cloud Mask
Organisation	**NASA**
Data set description	There are three MODIS Level 3 Atmosphere Products, each covering a different temporal scale: Daily, 8-Day, and Monthly. Each of these Level 3 products contains statistics de-rived from over 100 science parameters from the Level 2 Atmosphere products: Aerosol, Precipitable Water, Cloud, and Atmospheric Profiles. A range of statistical summaries (scalar statistics and 1- and 2-dimensional histograms) are computed, depending on the Level 2 science parameter. Statistics are aggregated to a 1° x 1° equal-angle global grid. The daily product contains ~700 statistical summary parameters. The 8-day and monthly products contain ~900 statistical summary parameters.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	2000 - today
Project Contact	mantovani@meeo.it
Upstream Contact	http://modaps.nascom.nasa.gov/services/user/
Limitations	None
License	Free
Constraints	The distribution of the MODAPS data sets is funded by NASA's Earth-Sun System Division (ESSD). The data are not copyrighted; however, in the event that you publish data or results using these data, we request that you include the following acknowledgment: *"The data used in this study were acquired as part of the NASA's Earth-Sun System Division and archived and distributed by the MODIS Adaptive Processing System (MODAPS)."* We would appreciate receiving a copy of your publication, which can be forwarded to MODAPSUSO@lists.nasa.gov.
Data Format	GeoTIFF (generated from HDF)
Access URL	TBD
Archiving and preservation (including storage and backup)	Data is part of Level-2 MODIS Atmosphere Products

Table 3: First data set description for the EODS.

Data set reference and name	MOD 08 - Gridded Atmospheric Product; MOD 11 - Land Surface Temperature and Emissivity
Organisation	**NASA**
Data set description	There are three MODIS Level 3 Atmosphere Products, each covering a different temporal scale: Daily, 8-Day, and Monthly. Each of these Level 3 products contains statistics de-rived from over 100 science parameters from the Level 2 Atmosphere products: Aerosol, Precipitable Water, Cloud, and Atmospheric Profiles. A range of statistical summaries (scalar statistics and 1- and 2-dimensional histograms) are computed, depending on the Level 2 science parameter. Statistics are aggregated to a 1° x 1° equal-angle global grid. The daily product contains ~700 statistical summary parameters. The 8-day and monthly products contain ~900 statistical summary parameters.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	2000 - today
Project Contact	mantovani@meeo.it
Upstream Contact	http://modaps.nascom.nasa.gov/services/user/
Limitations	None
License	Free
Constraints	The distribution of the MODAPS data sets is funded by NASA's Earth-Sun System Division (ESSD). The data are not copyrighted; however, in the event that you publish data or results using these data, we request that you include the following acknowledgment: *"The data used in this study were acquired as part of the NASA's Earth-Sun System Division and archived and distributed by the MODIS Adaptive Processing System (MODAPS)."* We would appreciate receiving a copy of your publication, which can be forwarded to MODAPSUSO@lists.nasa.gov.
Data Format	GeoTIFF (generated from HDF)
Access URL	TBD
Archiving and preservation (including storage and backup)	Data is part of Level-3 MODIS Atmosphere Products

Table 4: Second data set description for the EODS.

Data set reference and name	SMOS Level 2 Soil Moisture (SMOS.MIRAS.MIR_SMUDP2); SMOS Level 2 Ocean Salinity (SMOS.MIRAS.MIR_OSUDP2)
Organisation	**ESA**
Data set description	ESA's Soil Moisture Ocean Salinity (SMOS) Earth Explorer mission is a radio telescope in orbit, but pointing back to Earth not space. Its Microwave Imaging Radiometer using Aperture Synthesis (MIRAS) radiometer picks up faint microwave emissions from Earth's surface to map levels of land soil moisture and ocean salinity. These are the key geophysical parameters, soil moisture for hydrology studies and salinity for enhanced understanding of ocean circulation, both vital for climate change models.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	12-01-2010 - today
Project Contact	mantovani@meeo.it
Upstream Contact	EO-Support (https://earth.esa.int/web/guest/contact-us)
Limitations	None
License	Free
Constraints	https://earth.esa.int/web/guest/data-access/how-to-access-eo-data/earth-observation-data-distributed-by-esa
Data Format	GeoTIFF (generated from measurements geo-located in an equal-area grid system ISEA 4H9)
Access URL	TBD
Archiving and preservation (including storage and backup)	Data is part of Level-2 SMOS Products

Table 5: Third data set description for the EODS.

Data set reference and name	Landsat8 L1T
Organisation	**ESA**
Data set description	Level 1 T- Terrain Corrected
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	European
Temporal extent	2014 - today
Project Contact	mantovani@meeo.it
Upstream Contact	EO-Support (https://earth.esa.int/web/guest/contact-us)
Limitations	Terms and Conditions for the Utilisation of Data under ESA's Third Party Missions scheme
License	Open and Free
Constraints	Acceptance of Terms and Conditions

Data Format	GeoTIFF
Access URL	TBD
Archiving and preservation (including storage and backup)	ESA is an International Co-operator with USGS for the Landsat-8 Mission. Data is downlinked via Kiruna and Matera (KIS and MTI) stations whenever the satellite passes over Europe, starting from November 2013. Typically the station's will receive 2 or 3 passes per day each and there will be some new scenes for each path, in accordance with the overall mission acquisition plan. The Neustrelitz data available on the portal from May 2013 to December 2013 Data will be processed to either L1T or L1Gt product format as soon as it is downlinked. The target time is for scenes to be available for download within 3 hours of reception. https://landsat8portal.eo.esa.int/faq/

Table 6: Fourth data set description for the EODS.

Data set reference and name	Sentinel2
Organisation	**ESA**
Data set description	Level-1C[3]
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	Q3 2015
Project Contact	mantovani@meeo.it
Upstream Contact	eosupport@copernicus.esa.int
Limitations	None
License	Free and Open
Constraints	Registration A maximum of 2 concurrent downloads per user is allowed in order to ensure a download capacity for all users.
Data Format	Sentinel Standard Archive Format for Europe (SAFE) format, including image data in JPEG2000 format, quality indicators, auxiliary data and metadata
Access URL	Sentinels Scientific Data Hub: https://scihub.esa.int
Archiving and preservation (including storage and backup)	Data is part of ESA Earth Observation Long Term Data Preservation (LTDP) Programme

Table 7: Fifth data set description for the EODS.

---

[3] https://earth.esa.int/web/sentinel/sentinel-2-msi-wiki/-/wiki/Sentinel+Two/Level+1c+Products

6.4   Planetary Science Data Service

Data set reference and name	MGS MOLA GRIDDED DATA RECORDS
Organisation	**Jacobs University**
Data set description	Mars Orbiter Laser Altimeter (MOLA)
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Global
Temporal extent	Not Applicable (gridded from multiple experiment data records)
Project Contact	an.rossi@jacobs-university.de
Upstream Contact	geosci@wunder.wustl.edu
Limitations	None
License	Free
Constraints	None
Data Format	PDS standard (GDAL-compatible .IMG or alike)
Access URL	http://ode.rsl.wustl.edu/mars/pagehelp/quickstartguide/index.html?mola.htm
Archiving and preservation (including storage and backup)	Data is part of long term NASA PDS archives and the original copies are maintained there

Table 8: First data set description for the PSDS.

Data set reference and name	MRO-M-CRISM-3-RDR-TARGETED-V1.0
Organisation	**Jacobs University**
Data set description	TRDR - Targeted Reduced Data Records contain data calibrated to radiance or I/F.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Local
Temporal extent	Variable
Project Contact	an.rossi@jacobs-university.de
Upstream Contact	geosci@wunder.wustl.edu
Limitations	None
License	Free
Constraints	None
Data Format	PDS standard (GDAL-compatible .IMG or alike)
Access URL	http://ode.rsl.wustl.edu/mars/pagehelp/quickstartguide/index.html?crism.htm

Archiving and preservation (including storage and backup)	Data is part of long term NASA PDS archives and the original copies are maintained there

Table 9: Second data set description for the PSDS.

Data set reference and name	MRO-M-CRISM-5-RDR-MULTISPECTRAL-V1.0
Organisation	**Jacobs University**
Data set description	MRDR - Multispectral Reduced Data Records contain multispectral survey data calibrated, mosaicked, and map projected.
Standards	Data will be made available through the OGC WCPS and WCS standard.
Spatial extent	Regional/global
Temporal extent	Not Applicable, derived data from multiple acquisition times
Project Contact	an.rossi@jacobs-university.de
Upstream Contact	geosci@wunder.wustl.edu
Limitations	None
License	Free
Constraints	None
Data Format	PDS standard (GDAL-compatible .IMG or alike)
Access URL	http://ode.rsl.wustl.edu/mars/pagehelp/quickstartguide/index.html?crism.htm
Archiving and preservation (including storage and backup)	Data is part of long term NASA PDS archives and the original copies are maintained there

Table 10: Third data set description for the PSDS.

Data set reference and name	LRO-L-LOLA-4-GDR-V1.0
Organisation	**Jacobs University**
Data set description	LRO LOLA GRIDDED DATA RECORD
Standards	Data will be made available through the OGC WCPS standard.
Spatial extent	Global
Temporal extent	Not Applicable (gridded from multiple experiment data records
Project Contact	an.rossi@jacobs-university.de
Upstream Contact	geosci@wunder.wustl.edu
Limitations	None
License	Free
Constraints	None
Data Format	PDS standard (GDAL-compatible .IMG or alike)

| Access URL | http://ode.rsl.wustl.edu/moon/pagehelp/quickstartguide/index.html?lola.htm |
| Archiving and preservation (including storage and backup) | Data is part of long term NASA PDS project and the original copies are maintained there |

Table 11: Fourth data set description for the PSDS.

Data set reference and name	MEX-M-HRSC-5-REFDR-DTM-V1.0
Organisation	**Jacobs University**
Data set description	Mars Express HRSC topography
Standards	Data will be made available through the OGC WCPS standard.
Spatial extent	Local
Temporal extent	Variable
Project Contact	an.rossi@jacobs-university.de
Upstream Contact	psahelp@rssd.esa.int
Limitations	None
License	Free
Constraints	None
Data Format	PDS standard (GDAL-compatible .IMG or alike)
Access URL	ftp://psa.esac.esa.int/pub/mirror/MARS-EXPRESS/HRSC/MEX-M-HRSC-5-REFDR-DTM-V1.0/DOCUMENT/
Archiving and preservation (including storage and backup)	Data is part of long term ESA PSA project and the original copies are maintained there

Table 12: Fifth data set description for the PSDS.

6.5   Landsat Data Cube Service

Data set reference and name	Landsat
Organisation	**ANU/NCI**
Data set description	The Australian Reflectance Grid (ARG) http://geonetwork.nci.org.au/geonetwork/srv/eng/metadata.show?id=24&currTab=simple
Standards	Data is available at OGC WCS standard.
Spatial extent	Longitude: 108 – 155, Latitude: -10 - -45, Universal Transverse Mercator (UTM) and Geographic Lat-Lon
Temporal extent	1997-now
Project Contact	Ben.Evans@anu.edu.au
Upstream Contact	datacollections@nci.org.au
Limitations	None
License	Commonwealth of Australia (Geoscience Australia) 2015. Creative Commons Attribution 4.0 International Australia License. https://creativecommons.org/licenses/by/4.0/
Constraints	Commonwealth of Australia (Geoscience Australia) 2015. Creative Commons Attribution 4.0 International Australia License. https://creativecommons.org/licenses/by/4.0/
Data Format	GeoTIFF [NetCDF-CF conversion currently underway]
Access URL	http://dap.nci.org.au/thredds/remoteCatalogService?catalog=http://dapds00.nci.org.au/thredds/catalog/rs0/catalog.xml
Archiving and preservation (including storage and backup)	This data collection is part of the Research Data Storage Infrastructure program, which aims for long-term preservation.

Table 13: Data set description for the LDCS.