



**open data
portal**
cci

DEL 15 Tech Specification Annex: Archive and Metadata Catalogue Implementation

Prepared by : Philip Kershaw (STFC-CEDA)
Alison Waterfall (STFC-CEDA)
Antony Wilson (STFC-CEDA)
Victoria Bennett (STFC-CEDA)

Approved by : Victoria Bennett (STFC-CEDA)

Customer:	ESA	Author:	CEDA
Project Name:	CCI Open Data Portal	Contract No.:	4000113814/15/I-NB
Project Ref:	Task 3	Document Ref:	<i>CDAP.REP.023</i>
Issue:	<i>Version 2.2</i>	Issue Date:	<i>07 July 2017</i>



AMENDMENT RECORD SHEET

The Amendment Record Sheet below records the history and issue status of this document.

ISSUE	DATE	REASON
0.1	11 May 2015	First draft issue for review
1.0	19 May 2016	First formal release
1.3	24 June 2016	Addressing RIDs following review by ESA: <ul style="list-style-type: none">• RID D.15 v1.0 001• RID D.15 v1.0 002• RID D.15 v1.0 003• RID D.15 v1.0 004
2.0	28 April 2017	2017 annual update
2.1	30 June 2017	Addressing RIDs following review by ESA: <ul style="list-style-type: none">• RID_D.15_v2.0_001• RID_D.15_v2.0_002
2.2	07 July 2017	Section 4.1.3.3 – Additional information on download stats, in particular in relation to WCS added

DISTRIBUTION LIST

The Amendment Record Sheet below records the history and issue status of this document.

NAME	AFFILIATION
Ed Pechorro	ESA
Simon Pinnock	ESA
Pascal Lecomte	ESA
Cat Downey	ESA
Richard Lowe	TVUK
Kevin Halsall	TUVK
Anna Corlyon	TVUK
Victoria Bennett	STFC
Philip Kershaw	STFC



NAME	AFFILIATION
Anne Chadwick	STFC
Alison Waterfall	STFC
Debbie Clifford	University of Reading
Jon Blower	University of Reading
Jane Lewis	University of Reading
Clive Farquhar	CGI
Alex Wood	CGI
Katy Hurst	CGI
Carsten Brockmann	Brockmann Consult
Tom Block	Brockmann Consult
Norman Fomferra	Brockmann Consult



TABLE OF CONTENTS

1. EXECUTIVE SUMMARY	7
2. INTRODUCTION	8
2.1 Purpose and Scope	8
2.2 Structure of the Document	8
2.3 Referenced Documents	8
3. PLATFORM	9
3.1 Data access routes.....	12
4. CENTRAL DATA ARCHIVE	14
4.1 Implementation updates.....	14
4.1.1 Functionality and Integration with Earth System Grid Federation	14
4.1.2 Data Compression	14
4.1.3 Data Download	14
4.1.3.1 THREDDS Data Server – support for OPeNDAP, WMS and WCS protocols.	15
4.1.3.2 ProFTP – FTP	15
4.1.3.3 Download Monitoring and Reporting.....	15
4.1.4 Processes	16
4.1.4.1 Data Ingestion	16
4.1.4.2 DOI Management.....	16
4.1.4.3 Quicklook Generation	16
4.1.5 Data Version Control	16
4.1.6 Performance and Scalability.....	16
4.1.6.1 Performance Goals and Analysis Assumptions.....	16
4.1.6.2 System Resource Use	17
4.1.6.3 Resources Deployed and Scaling	17
4.1.7 Long Term Data Preservation.....	17
4.1.8 Data Product Selection	18
4.1.9 Dataset Rollout.....	18
5. METADATA CATALOGUE	20
5.1.1 Functionality	20
5.1.1.1 Metadata Data Model.....	20
5.1.1.2 Inclusion of externally hosted datasets	20
5.1.2 Processes	20
5.1.2.1 Scanning of Archive to Catalogue Records	20
5.1.2.2 Generation of CSW Metadata Content.....	21
5.1.2.2.1 Export MOLES Records via Dedicated CSW	21
5.1.2.3 Generation of ESGF Search Metadata Content	22
5.1.2.4 Web Presence Dashboard and Faceted Catalogue Search	22
5.1.3 Performance and Scalability.....	22
5.1.3.1 Performance Goals and Analysis Assumptions.....	22
5.1.3.2 System Resource Use	22
5.1.3.3 Resources Deployed and Scaling	22



DOCUMENT TABLES

Table 1: summary of archive volume by ECV at the end of year 2..... 16

DOCUMENT FIGURES

Figure 1: Deployment architecture for CCI Open Data Portal 10

Figure 2: Component Diagram for CCI Open Data Portal showing the logical arrangement of services and their interfaces..... 11

ABBREVIATIONS AND ACRONYMS

The table below shows all abbreviations and acronyms used within the document

Acronym	Description
CCI	Climate Change Initiative
CEDA	Centre for Environmental Data Analysis
CSW	OGC Catalogue Services for the Web
ESGF	Earth System Grid Federation
FTP	File Transfer Protocol
GridFTP	Extended and enhanced version of FTP for Grid computing
HTTP	Hyper Text Transfer Protocol
MOLES	Metadata Object Linking
OGC	Open Geospatial Consortium
OPeNDAP	Open-source Project for a Network Data Access Protocol
TDS	THREDDS Data Server
THREDDS	Thematic Real-Time Environmental Distributed Data Services
WCS	OGC Web Coverage Server
WGET	An open source HTTP client tool. ESGF uses it to provide simple script-based batch HTTP download
WMS	OGC Web Map Service





1. EXECUTIVE SUMMARY

This report describes the current implementation and any deviations/developments from the preceding Technical Specification, including a plan for any necessary future developments of the Data Archive and Metadata Catalogue, where already identified. This report will be annexed to the DEL-9: Technical Specification document. This report is updated annually and this issue relates to year 2 of the project. Information from year 1 that is still relevant has been retained.

In the first year, significant progress was made, from the deployment of the initial ftp download service to the integration of search and download services from the ESGF software suite. An innovative solution was implemented to integrate the two search web service interfaces: ESGF search capability is augmented with the CSW search services in a system that supports a common set of search facets for both interfaces. This links together the two sources of search information. This has been made possible via vocabularies defined using Semantic Web technologies and served from a central vocabulary service.

The second year saw the release of the portal dashboard and search. The search interface builds on the capabilities linking the ESGF and CSW search services initially set up in the first year and further refinements have been made, including the release of a new CSW which will support refinement in the faceted search. More data download protocols have come on line via ESGF and include http, OPeNDAP, WMS and WCS. Progress has been made on ingesting more datasets into ESGF despite non-compliances in some datasets, meaning that data has had to be manually tagged with the search facets and ESGF DRS identifier. Problems were encountered in publishing datasets with large numbers of files to ESGF, and a number of solutions were investigated. A feasible solution has been adopted to split the dataset into smaller sub-sets when publishing in ESGF.

The priorities for the coming months are to continue to build on these capabilities, in particular refining the faceted search as seen in the front end of the data portal, and continuing to streamline the publication process with regular data releases; increased monitoring of services will also be implemented.



2. INTRODUCTION

2.1 Purpose and Scope

The purpose of this document is to provide details of any changes to the technical architecture for the archive and metadata catalogue with respect to the original specification in RD.1.

2.2 Structure of the Document

The document attempts to follow the structure of RD.1 highlighting changes where appropriate.

- Platform – changes related to the deployment architecture for the system
- Central Data Archive – processes associated with management of the archive and hosting of data access services
- Metadata Catalogue – processes for the creation and cataloguing of metadata and exposure to the outside through search services.

2.3 Referenced Documents

- RD.1 CCI Open Data Portal DEL-09: Technical Specification Document, CDAP.REP.011, Issue 1.0, 30 October 2015
- RD.2 CCI Open Data Portal DEL-15: Archive Catalogue Implementation Annex1, CDAP.REP.023, Issue 1.1



3. PLATFORM

This section outlines changes to the deployment architecture from that in the original technical specification. These can be summarised as follows, and include changes that were implemented in year 1:

1. THREDDS Data Server (TDS) configuration. As has been reported in year 1 (RD.2), two TDS instances have been deployed, one as part of an ESGF node and an additional one to provide WMS and WCS capability. Further details are given in section 4.1.3.1.
2. Quicklook VM was deployed in the Unmanaged Cloud JVO in year 1 (RD.2). This allows more convenient access for the installation of software for the project partner responsible (Telespazio Vega)
3. GridFTP service was withdrawn as a requirement in year 2. The functionality that it provides is met with ftp and http download services.

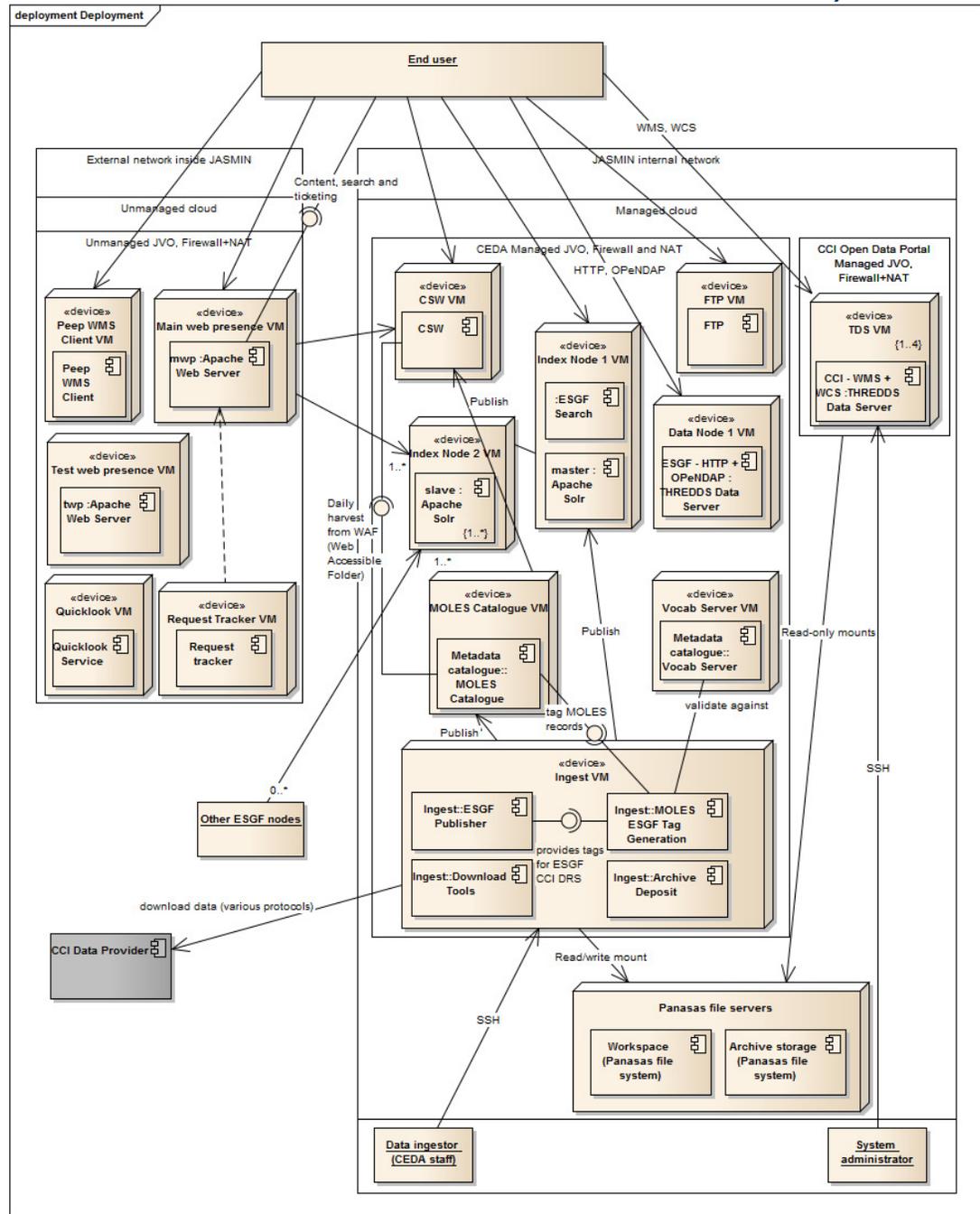


Figure 1: Deployment architecture for CCI Open Data Portal

The above diagram (Figure 1) shows the services in the context of their deployment on virtual machines and cloud tenancies (JASMIN Virtual Organisations). There has been no update to this since year 1.

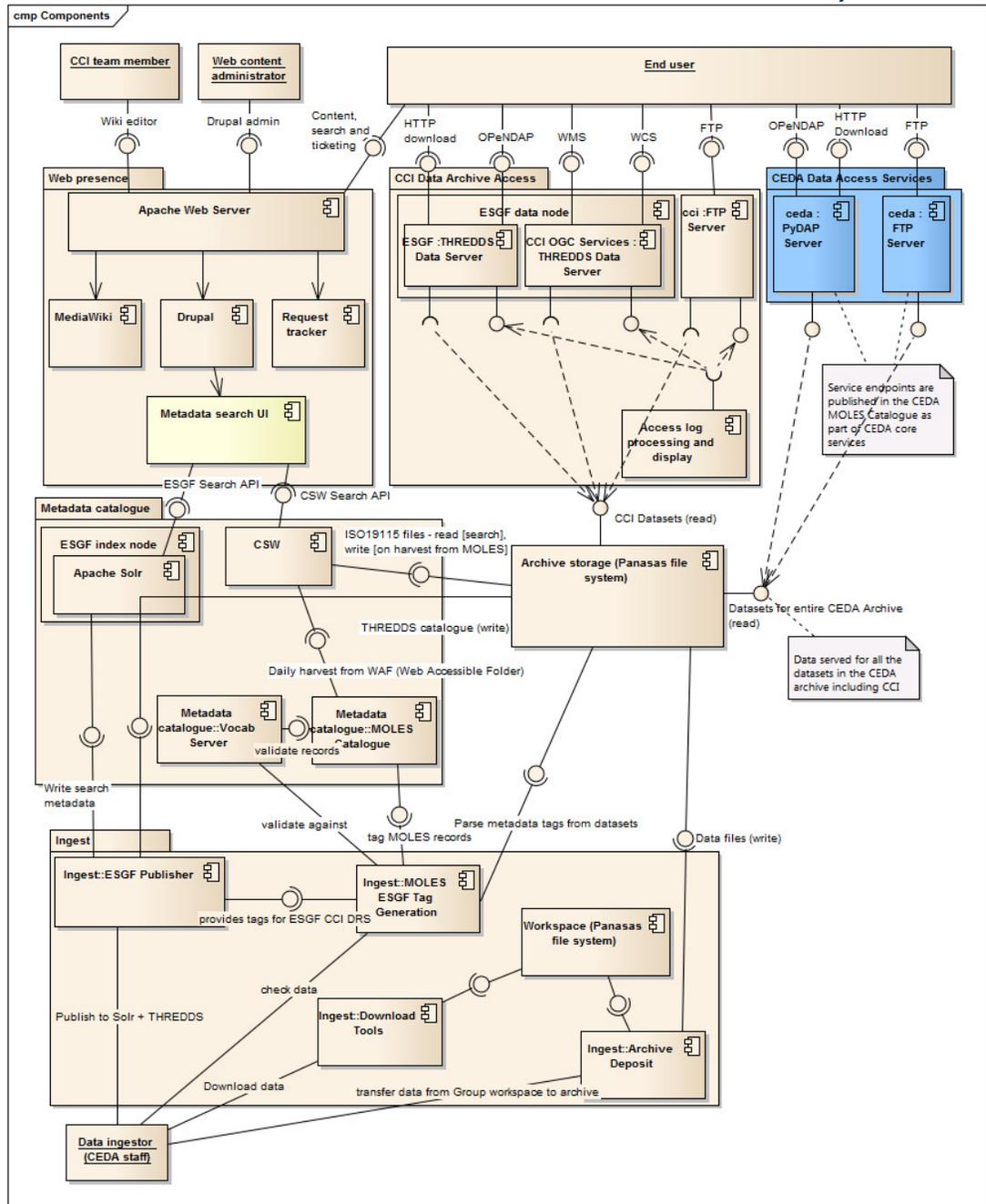


Figure 2: Component Diagram for CCI Open Data Portal showing the logical arrangement of services and their interfaces



3.1 Data access routes

The interfaces provided to the user are shown in Figure 2: Component Diagram for CCI Open Data Portal showing the logical arrangement of services and their interfaces. In summary these are:

- HTTP data download provided by ESGF THREDDS Data Server
- OPeNDAP subsetting provided by ESGF THREDDS Data Server
- Open Geospatial Consortium WMS and WCS provided by separate THREDDS Data Server¹
- Dedicated CCI anonymous FTP server

Independent of the Open Data Portal, CEDA runs core download services as part of its service to the NERC user community. These consist of:

- CEDA pydap server
- CEDA FTP server
- CEDA data browser (now retired as of end of February 2017)

The current services are shown in blue in Figure 2. They are included in this diagram because they serve all of the datasets in CEDA's archive, including CCI datasets. This means that users may download CCI data from these services in addition to the primary services for access provided by the Open Data Portal. They have relevance as another means to gather information about usage of CCI data from the user community.

The table below lists these services and the labels used to refer to them elsewhere.

Interfaces to Data	Label
HTTP data download provided by ESGF THREDDS Data Server	CEDA Managed – ESGF-HTTP
OPeNDAP subsetting provided by ESGF THREDDS Data Server	CEDA Managed – ESGF – OPeNDAP
Open Geospatial Consortium WMS	CCI ODP Managed – WMS
Open Geospatial Consortium WCS	CCI ODP Managed – WCS

¹ A separate instance of THREDDS is required for reasons explained in section 4.1.3.1. ESGF requires the latest version of THREDDS, version 5, but this release has not been updated to support the WMS and WCS extensions. A separate instance of THREDDS running an earlier version (4.6.4) is therefore needed to provide WMS and WCS. However, data is still published to the WMS and WCS services via the ESGF publisher, which is used to create the THREDDS catalogue on this separate TDS, and create and update the Apache Solr records used for ESGF search.



Dedicated CCI anonymous FTP server	CEDA Managed – FTP
CEDA pydap server	CEDA core – PyDAP
CEDA ftp server	CEDA core – FTP
CEDA data browser (HTTP access)	CEDA core - HTTP

It should be noted that the publication of datasets via ESGF (TDS HTTP, OPeNDAP, WMS and WCS) is only possible for data in suitable formats (NetCDF). All CCI datasets however, can be accessed via the anonymous FTP service, and also via the core CEDA services described above.

As shown in Figures 1 and 2, the data holdings can be searched and queried by the portal front end web presence, and also other users such as the CCI Toolbox, using the CSW (Catalogue Service for the Web) and the ESGF search API, and these are being used in conjunction with each other in the portal search interface. As not all datasets can be published via ESGF, the CSW should be regarded as the definitive source of all the latest CCI Open Data Portal data.



4. CENTRAL DATA ARCHIVE

4.1 Implementation updates

4.1.1 Functionality and Integration with Earth System Grid Federation

Since the release of RD.1, the ESGF federation has been redeployed for full operations, as has already been described in the Annex for year 1 (RD.2). The CEDA ESGF node was redeployed running release 2.2.3, which included the integration of THREDDS Data Server version 5 into the ESGF software stack. Version 5 gives significant performance benefits in memory management for THREDDS catalogues, however WMS and WCS services do not work with v5. In order to support WMS and WCS protocols it has been necessary to also deploy a second THREDDS Data Server, as is described more fully in section 4.1.3.1. .

In year 2, the decision was taken not to include GridFTP amongst the set of download services available for the Portal. GridFTP requires user authentication even where data has an open access policy. It was also deemed not to provide additional functionality needed over and above the other ftp and http services.

A problem was encountered with publishing datasets with large numbers of files to ESGF, due to memory limitations in the ESGF publishing software. Various potential solutions to improve the ESGF publisher were investigated, but it has not proved possible to overcome this in the timescale of the project since ESGF is an external collaboration and not under CEDA control. The best solution to this, in order to enable such datasets to be included in ESGF has been to publish a large dataset in smaller subsets e.g. yearly chunks. This has been done for ocean colour and SST data where needed, and has allowed publication in ESGF. This does however have implications for WMS viewing as currently these data can only be viewed in yearly chunks. The success and limitations of this approach are currently being assessed.

4.1.2 Data Compression

No change in specifications.

All data in NetCDF format have been stored without external compression. Internal compression has been used by some of the ECVs.

Currently some of the shape files produced by the Greenland Ice Sheets are stored as zipped data.

For some data products a zipped or tarred file was also included for ease of download by some users. The products were also stored unzipped.

4.1.3 Data Download

A number of Download Services and Protocols have been deployed, updates are provided below.



4.1.3.1 THREDDS Data Server – support for OPeNDAP, WMS and WCS protocols

The CCI Open Data Portal builds on functionality provided by ESGF, including the Data Node (which provides download services including the THREDDS Data Server (TDS)). In year 1, ESGF was redeployed at CEDA as part of its ongoing involvement in the hosting of projects within the federation. These include for example CMIP5. This new version of the ESGF software, 2.2.3, included major enhancements including use of the latest version of TDS – version 5. The decision was taken by the ESGF development team to integrate into the system in order to take advantage of its improved memory management for THREDDS catalogues. Version 5 is the latest release supported by Unidata.

For the CCI Open Data Portal, OGC WMS and WCS services are required. WMS and WCS are being provided by TDS by enabling the respective extensions for these services in the TDS configuration. However, WMS and WCS services do not work with version 5 of TDS. In response to advice from the Unidata mailing list, the CEDA project team have deployed an additional instance of TDS with version 4.6.4, to enable the deployment of WMS and WCS.. The infrastructure users version 5 of TDS integrated with CEDA's ESGF data node to provide HTTP data download and OPeNDAP. This is augmented with an additional TDS instance running version 4.6.4 to provide WMS and WCS capability for the project. However, data is still published to the WMS and WCS services via the ESGF publisher, which is used to create the THREDDS catalogue on this separate TDS, and create and update the Apache Solr records used for ESGF search.

4.1.3.2 ProFTP – FTP

No update to technical specification. An anonymous ftp server was been deployed in year 1. (<ftp://anon-ftp.ceda.ac.uk/neodc/esacci/>)

4.1.3.3 Download Monitoring and Reporting

FTP and download via CEDA services continues to be monitored as per year 1. In addition the download services which form part of the ESGF data node are now also being monitored. Downloads from the ESGF data node via OPeNDAP and HTTP are currently being reported collectively as a single ESGF download method. Log files with basic access monitoring for WMS and WCS data have also been collected for the entire period of operations, and with increased details since April 2017. The full reporting method for these access methods are still under development, although basic statistics are reported in the DEL-17 Archive Catalogue Metadata Operations Annual Report for year 2. For the WMS, the online nature of the service means that like for like statistics with other download protocols are not thought to be appropriate. WCS stats are collected in the same logs as WMS, and are currently being analysed by the same process, although it is envisaged that these WCS stats could be included in the general download stats in the future. However, there is currently little usage of the WCS so this will not affect the current download stats (this is discussed further in DEL-17)

Monthly download stats and historic download trends are provided via Monthly Progress Report to ESA.

Quarterly download stats are provided for Quarterly Status Reports to ECV teams Science Leaders.



4.1.4 Processes

4.1.4.1 Data Ingestion

Process has been carried out as per RD.1.

Some of the ingestion activity was planned to be handed from CEDA to TVUK and CGI during phase 2 of the project. This has been delayed as the procedures to do this are still being determined. It is planned to hand over aspects of the ingestion activity early in year 3.

4.1.4.2 DOI Management

Process as per RD.1.

DOI's have been issued for the following datasets:

- Fire v4.1 burned area data (August 2016)
- Ocean colour products v1 and v2 (December 2016)
- Sea ice concentration data v2 (February 2017)
- SST products (v1.1 analysis product (November 2016)
- GHG Obs4mips data (October 2016)

It is out of scope in this project for CEDA to issue DOI's for (i) software and (ii) products not held on the portal. However, where such requests from an ECV team arise these will be noted by CEDA in case needed for a future new activity.

4.1.4.3 Quicklook Generation

Process as per RD.1. Quick look data were produced for each dataset from the rapid deployment phase and have been made available via the Central Data Archive.

Later datasets either a) reuse these quick looks where no significant change to the product has occurred b) use quick looks generated by the project team where this is more appropriate, or c) will have new quick looks generated by the portal project.

4.1.5 Data Version Control

Process has been carried out as described in RD.1.

4.1.6 Performance and Scalability

4.1.6.1 Performance Goals and Analysis Assumptions

Information about the current archive volume is listed below.

Table 1: summary of archive volume by ECV at the end of year 2

ECV	Volume	Note
Aerosol	992 GB	



Cloud	231 GB	New version replacing earlier products (L3C data only hence smaller)
Fire	42 GB	
GHG	52 GB	
Glaciers	0.008 Gb	
Ice Sheets – Antarctica	0	Released and awaiting addition to archive.
Ice Sheets – Greenland	6.1 GB	
Land cover	19 GB	
Ocean Colour	93000 GB	
Ozone	1.5 GB	
Sea ice	331GB	
Sea level	272 GB	
Soil moisture	48 GB	
SST	4100 GB	

Total archive size: 99 TB

An issue with slow data download rates via ESGF has been reported by the CCI Toolbox team; this is being addressed as described in section 4.1.6.3.

4.1.6.2 System Resource Use

No issues have been reported.

4.1.6.3 Resources Deployed and Scaling

Recent discussions with CCI Toolbox team indicate that there may be a need to consider redeployment of the download services into a special part of the JASMIN network called the Data Transfer Zone (DTZ) in order to increased download speeds.

4.1.7 Long Term Data Preservation

As described in RD.1.



4.1.8 Data Product Selection

As stated in Section 4.8 of RD.1 (Technical Specification Document, CDAP.REP.011, Issue 1.0), *Data products to go on the CCI Open Data Portal will be selected based on consultation with the individual CCI project teams and ESA, and informed by user engagement and literature review:*

This has been carried out following the logic described in Section 4.9 of RD.. In the second year of operations there has been adequate space to hold all of the requested data products, as the large ocean colour dataset would be separately eligible for archiving in the CEDA archive under national funding.

When new datasets have been added in general old datasets have been retained in the Central Data Archive (although not visible in the portal dashboard and search), but some ECV teams have requested their data be removed completely, and if so, this has been undertaken.

There have been further discussions with the CCI teams regarding their requirements for data registration within the constraint of maintaining open access to the data without compulsory registration.

Obs4MIPS products

All Obs4MIPs data produced within the project will be added to the Portal archive and made available to users from the CCI Open Data Portal ESGF Node [RD.1, Section 4.8]:

Greenhouse gases Obs4MIPs data have been added to the portal archive and published via ESGF. Other Obs4MIPs data will continue to be added as it becomes available.

Additional products

Additional ancillary, in-situ and model data will also be considered for inclusion in consultation with ESA and Portal users. Initial auxiliary data products that have been requested are datasets from CMUG, and mask datasets from across the CCI products. The Land Sea Mask produced by the Land Cover CCI ECV team (and others as needed) will be included in the CCI Open Data Portal archive and made available to users. A table of all the current and anticipated future products to be included in the CCI Portal will be maintained [RD.1, Section 4.8]:

The landcover water bodies mask data has been added to the Data Portal archive.

4.1.9 Dataset Rollout

It has been found that many new CCI datasets have become available at different times throughout the two years of operations, and many are expected in the coming year as the CCI teams complete their final data release, putting a high workload on the data portal. In order to better handle these demands it is proposed to adopt a regular publication cycle, with updates released to the data portal on pre-agreed dates (e.g. monthly). This will allow for better coordination as more of the operational data ingestion tasks are shared amongst the project partners and will also increase efficiency.



***DEL 15 Tech Specification Annex: Archive and
Metadata Catalogue Implementation***

Issue 2.2

07 July 2017

These scheduled publication dates will be communicated to the project teams and also to key regular users of the portal such as the CCI Toolbox.



5. METADATA CATALOGUE

5.1.1 Functionality

5.1.1.1 Metadata Data Model

Vocabularies play a key role for the CCI Open Data Portal, enabling data to be organised so that it can be readily discovered and accessed via faceted search tools. The metadata catalogue supports two classes of metadata as described in RD.1 Discovery metadata and file-level metadata:

1. Discovery Metadata - Catalogue records have been completed in the CEDA catalogue, exported via CSW (Catalogue Services for the Web).
2. File-level Metadata – CCI datasets have been ingested into the ESGF search system using the ESGF publisher. This indexes metadata content into Apache Solr, the search database used by ESGF.

The respective search services serving these metadata feed into Web Presence Dashboard and Catalogue Search.

The process of collating these vocabularies was determined in year 1, and described in RD.2 (v1.1 of this document). In year 2, vocabulary terms have been extended as required.

5.1.1.2 Inclusion of externally hosted datasets

Updates to our MOLES metadata catalogue were implemented to allow datasets which are hosted externally rather than within the archive to be shown in the catalogue. These can then be exported to the CCI Open Data Portal, allowing records held elsewhere to also be shown in the dashboard and search. The circumstances for which this would be deemed appropriate for a dataset has been defined and agreed with ESA.

5.1.2 Processes

5.1.2.1 Scanning of Archive to Catalogue Records

The script used in year 1 to scan all the files within a dataset was further updated to allow for scanning of a wider range of data.

As described in RD.2, the script extracts information from within the NetCDF file headers of data files as well as from information embedded in the file names. The data providers have used two different formats for file names. This extracted information was then compared against terms in the vocabulary.

The script produces three files:



- The first, a CSV file for ingest into MOLES, contained a series of records of the format dataset internal path, URL of term from vocabulary. Ingestion into MOLES is a precursor to publishing of ISO 19115 records via the CSW service.
- The second file was encoded as JSON for ingest into the ESGF publishing system. Here the records contained a DRS value and list of associated internal paths.
- The third file contained error diagnostics which enabled the CEDA team to feedback information to the CCI data providers about incorrectly formatted file names, missing terms and the use of incorrect terms.

As a number of files contained legacy information that could not be updated to conform to the vocabularies, a series of mappings were put in place to map from the legacy term to the term in the vocabulary enabling these files to be tagged with the new terms.

In addition, many of the datasets were non-compliant with aspects of the data standards (e.g. missing some of the metadata that would allow this automatic tagging to take place). In this case, manual intervention was required to generate the appropriate files as above.

5.1.2.2 Generation of CSW Metadata Content

As reported in year 1, the CEDA MOLES catalogue application (<http://catalogue.ceda.ac.uk/>) was extended to cache information about the vocabularies, namely the URI and preferred label of the terms. This is then exported in the CSW to the portal as described below.

5.1.2.2.1 Export MOLES Records via Dedicated CSW

The CEDA MOLES catalogue provides an intermediary for the CSW service. It provides a WAF (Web Accessible Folder) interface which exposes MOLES observation records rendered as ISO 19115 records. This enables the CSW to regularly poll MOLES for new ISO records and publish them to the outside. In year 1 this interface was extended to include the search facet information in the exported ISO 19115 records. This enables clients to define search queries by search facet category much in the same way as the ESGF Search API.

A new CSW web service was deployed during the second year of operations. This will allow improvements to be made to the faceted search in the portal in the future so that search terms are correctly refined as a user narrows their search criteria. This has also been updated to address security concerns discovered in routine scans, and for which fixes have been put in place.

The new CSW web service is available here:

<https://csw.ceda.ac.uk/geonetwork/srv/eng/csw-CEDA-CCI>

Error! Hyperlink reference not valid. This same CSW is also being used to provide data records in to the GEOSS portal.



5.1.2.3 Generation of ESGF Search Metadata Content

As already reported in year 1 (RD.2), the MOLES ESGF Tag Generation script produces a JSON-formatted output of search facets keyed by dataset identifier. The facets are taken from the terms defined in the vocabulary server. This JSON file provides the input to the ESGF Publisher. Once published, CCI data is served from the ESGF data node and is discoverable from the ESGF search interface.

There remains a limitation on our ability to tag terms in ESGF as the terms have to be encoded in a Directory Reference Syntax (DRS). As a result files are tagged with the label of the term rather than the URI of the term. Also only one value can be used per facet. This impacts the use of the CCI DRS, for example where datasets are tagged as derived from more than one sensor or platform. It has not proved possible to fix this in ESGF, but an adequate workaround was implemented in year 1.

5.1.2.4 Web Presence Dashboard and Faceted Catalogue Search

As noted above, changes to the CSW in year 2 have been made that will allow refinements to the faceted search in the portal. When implemented in the portal front end, this will allow only those terms that are relevant to the subset of datasets to appear in the drop down facet lists.

5.1.3 Performance and Scalability

5.1.3.1 Performance Goals and Analysis Assumptions

As described in RD.1.

5.1.3.2 System Resource Use

As described in RD.1.

5.1.3.3 Resources Deployed and Scaling

As described in RD.1. No load balancing has been required to date.