



# Managing DRS vocabularies in the CLIPC project

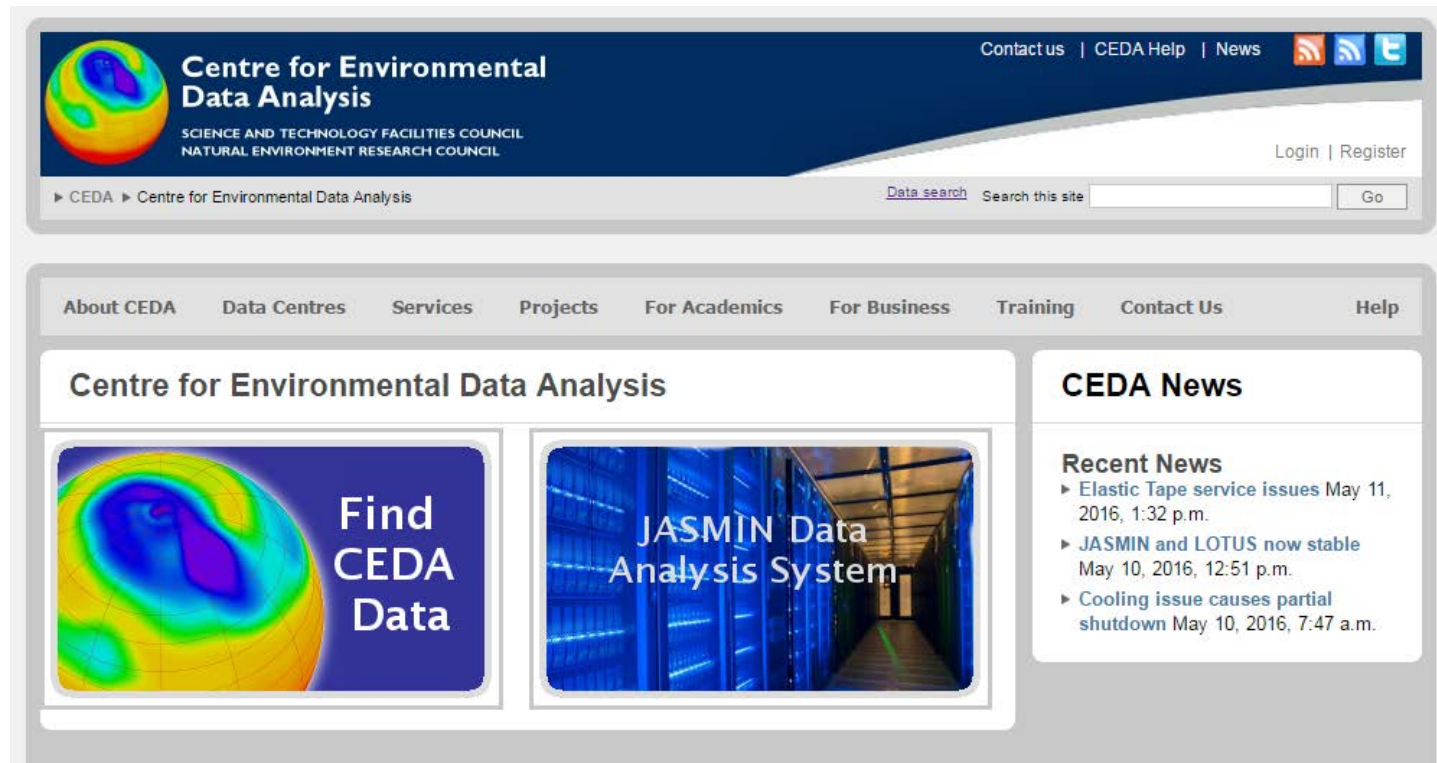
Ruth Petrie, Phil Kershaw, Antony  
Wilson and Ag Stephens



## (Centre for Environmental Data Analysis)

<http://www.ceda.ac.uk>

- Data centres
- JASMIN platform



The screenshot shows the CEDA website homepage. The header features the CEDA logo (a globe with a blue and yellow swirl) and the text "Centre for Environmental Data Analysis" and "SCIENCE AND TECHNOLOGY FACILITIES COUNCIL NATURAL ENVIRONMENT RESEARCH COUNCIL". Navigation links include "Contact us", "CEDA Help", "News", and social media icons. A search bar is present with the text "Data search" and "Search this site". Below the header is a navigation menu with links: "About CEDA", "Data Centres", "Services", "Projects", "For Academics", "For Business", "Training", "Contact Us", and "Help". The main content area is divided into two columns. The left column has a large image of a globe with a blue and yellow swirl and the text "Find CEDA Data". The right column has a large image of a server rack and the text "JASMIN Data Analysis System". To the right of the main content area is a "CEDA News" section with the heading "Recent News" and three bullet points: "Elastic Tape service issues May 11, 2016, 1:32 p.m.", "JASMIN and LOTUS now stable May 10, 2016, 12:51 p.m.", and "Cooling issue causes partial shutdown May 10, 2016, 7:47 a.m."

# What is CLIPC?

- Climate Information Portal for Copernicus  
<http://www.clipc.eu>
- Enable discovery and access, through ESGF, to European climate data from different sources:
  - Observational (satellite and in-situ)
  - Climate modelling (global and regional)
  - Climate Impact Indicators (derived products)



# Putting EO (ESA CCI) Data into ESGF Faceted Search

- Heterogeneous data
  - Data come from different communities with different ways of doing things
    - Different formats / vocabularies....
- Does a standardised term always mean the same thing?
- Data is being made available outside its traditional user community (so can't assume knowledge of EO)



# Developing new DRSs

- New data types required new DRS patterns:
  - New facets
  - New structures - for dataset identifiers
- New challenge?
  - Datasets with multiple facet values, e.g.:
    - Multi-sensor
    - Multi-platform
  - How to represent in ESGF Dataset ID?
  - How to represent in Solr/Search/CoG?

# New facets

- A standards document was developed to control/manage facets in CLIPC, introducing facets such as:

data type

platform

sensor

essential climate  
variable

processing  
level



# DRS for different data types

## 1. Satellite remotely sensed data (e.g. ESA-CCI)

<activity>	esacci
<cci_project>	FIRE
<time_frequency>	day
<processing_level>	L4
<geophysical_data_type>	BA
<sensor_id>	<b>multi-sensor</b>
<platform_id>	<b>multi-platform</b>
<product_string>	MERIS
<product_version>	<b>v4-1</b>
<realization>	r1
<version>	v20161101

## 2. In-situ observational data (e.g. MOHC HadOBS)

<activity>	clipc
<product>	insitu
<institute>	MOHC
<framework>	<b>HadOBS</b>
<collection>	<b>HadISD</b>
<time_frequency>	subdaily
<realization>	r1
<product_version>	v1-0-3-2014f
<version>	v20160416

# DRS for Climate Impact Indicators

## Model based

<activity>	clipc
<product>	rcm-derived
<package>	icclim-4-1-2
<institution>	SMHI
<GCMName>	ICHEC-EC-EARTH
<ExperimentName>	historical
<EnsembleMember>	r1i1p1
[<RCMName>	SMHI-RCA4
<RCMRealisation>	v4
<domain>]	EUR-11
[<BcName>	DBS42
<BcObsName>	EURO4M-Mesan
<BcRefPeriod>]	bcref-1989-2010
<frequency>	yr
[<Reference_period>]	1981-2010
<version>	v20151112

## Observation/Multi-sources

<activity>	clipc
<product>	multi-derived
<package>	R-3-1
<domain>	EUR-05-Med
<institution>	SKYE
<sourceDataID>	multi-model
<frequency>	day
[<Reference_period>]	1981-2010
<version>	v20161009

Difficult:  
multiple "sub-  
facets"



# Vocabulary Service (1)

Vocabulary service enables central management of:

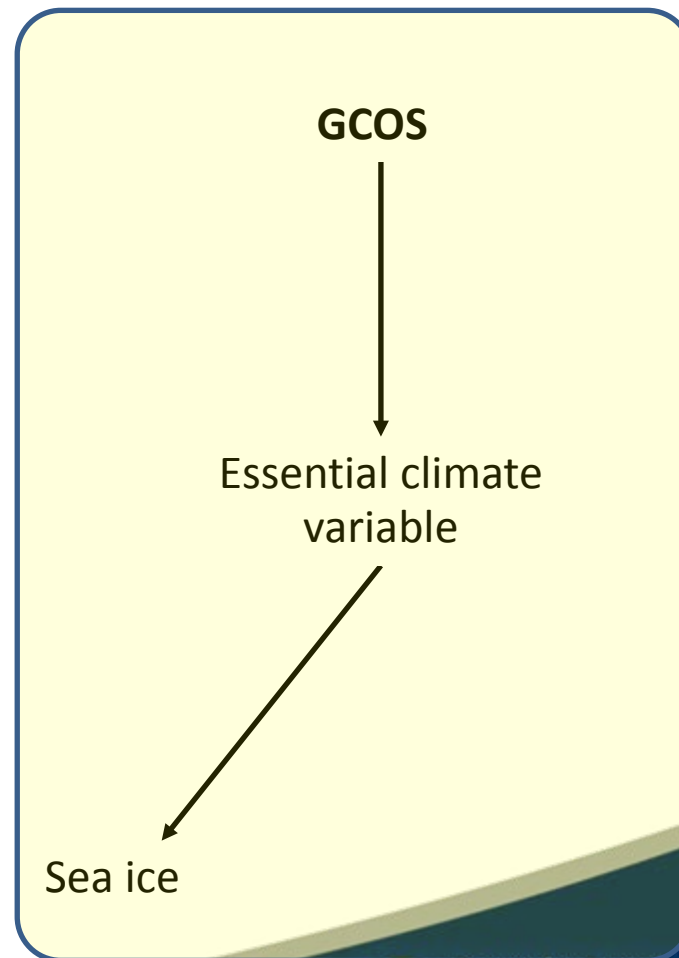
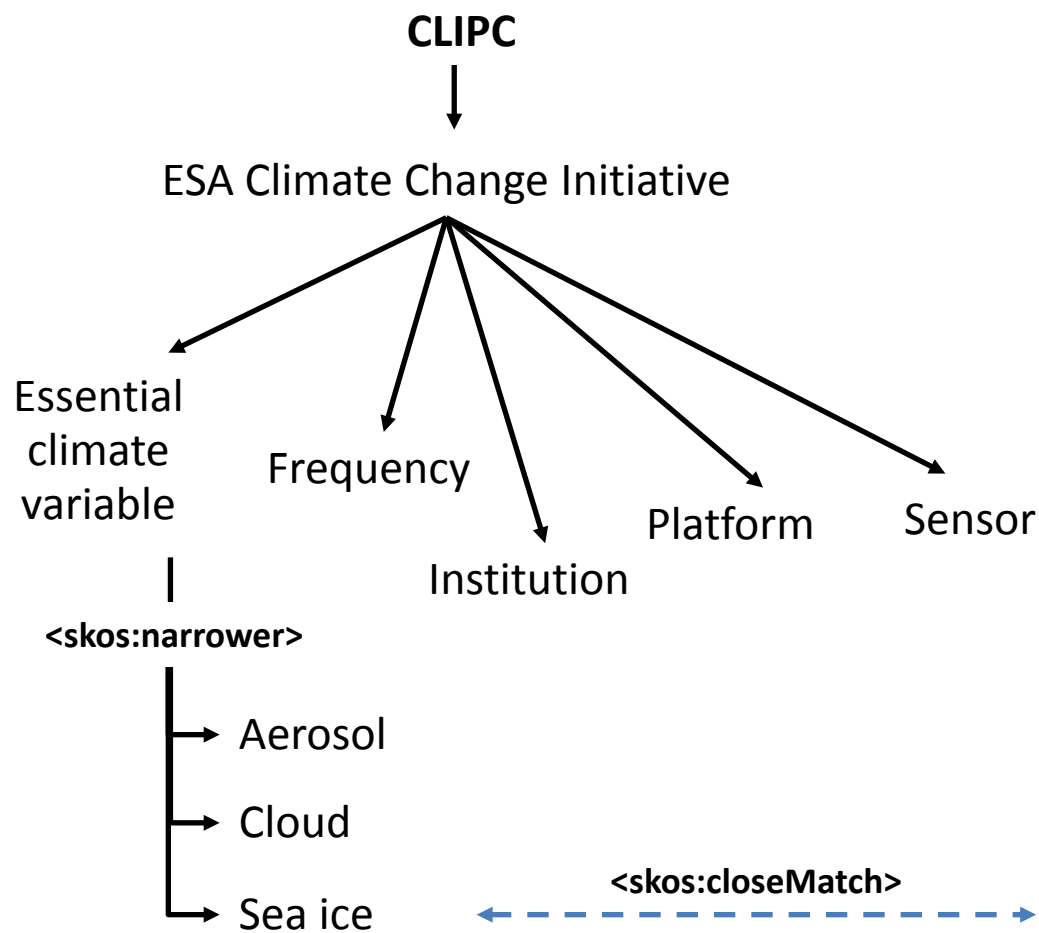
- Controlled vocabularies
  - Definitions of terms
  - Relationships between term, internally and externally
- 
- Assists in integrating multiple services, providing richer content.
  - Developed for CCI / re-used by CLIPC project

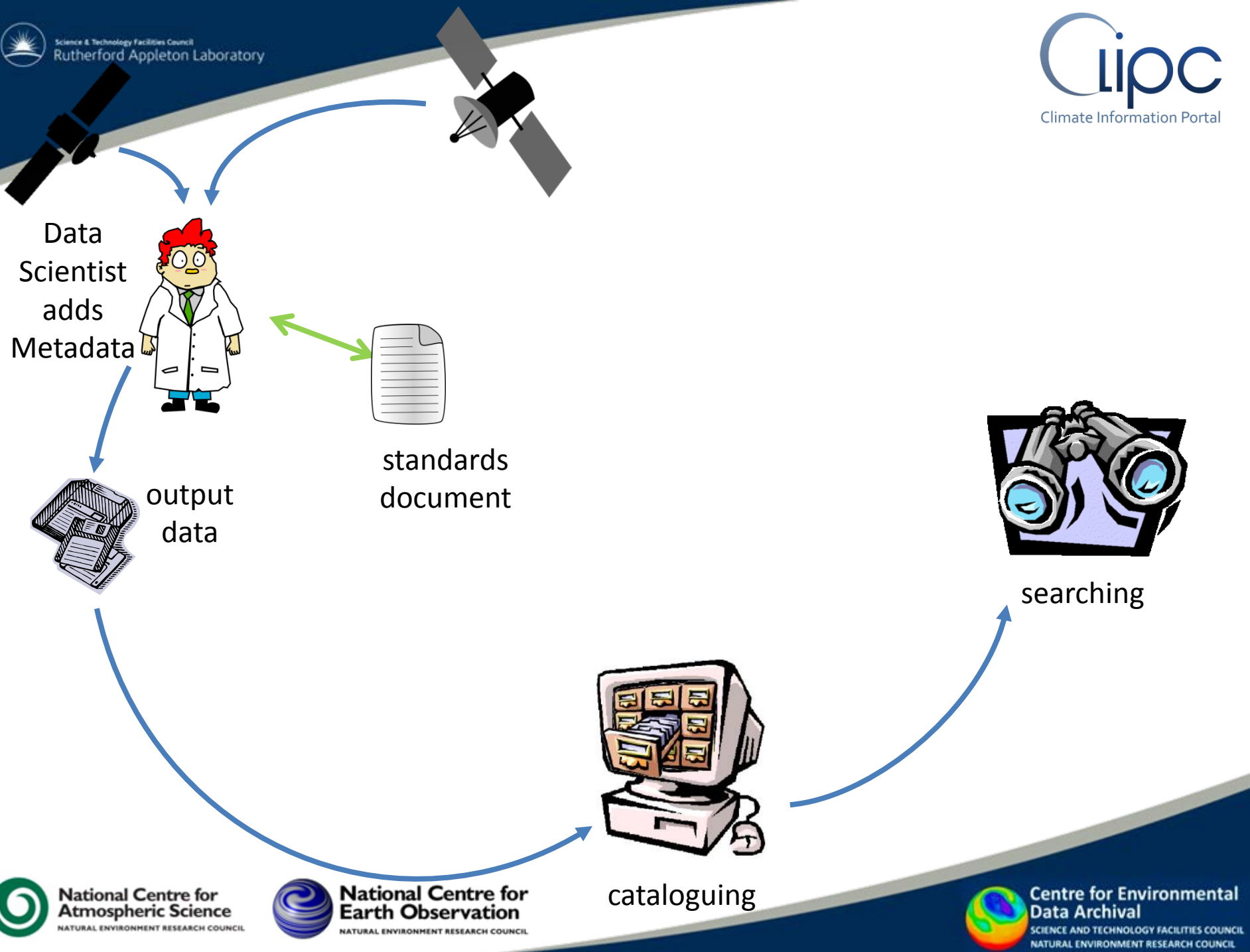
# Vocabulary Service (2)

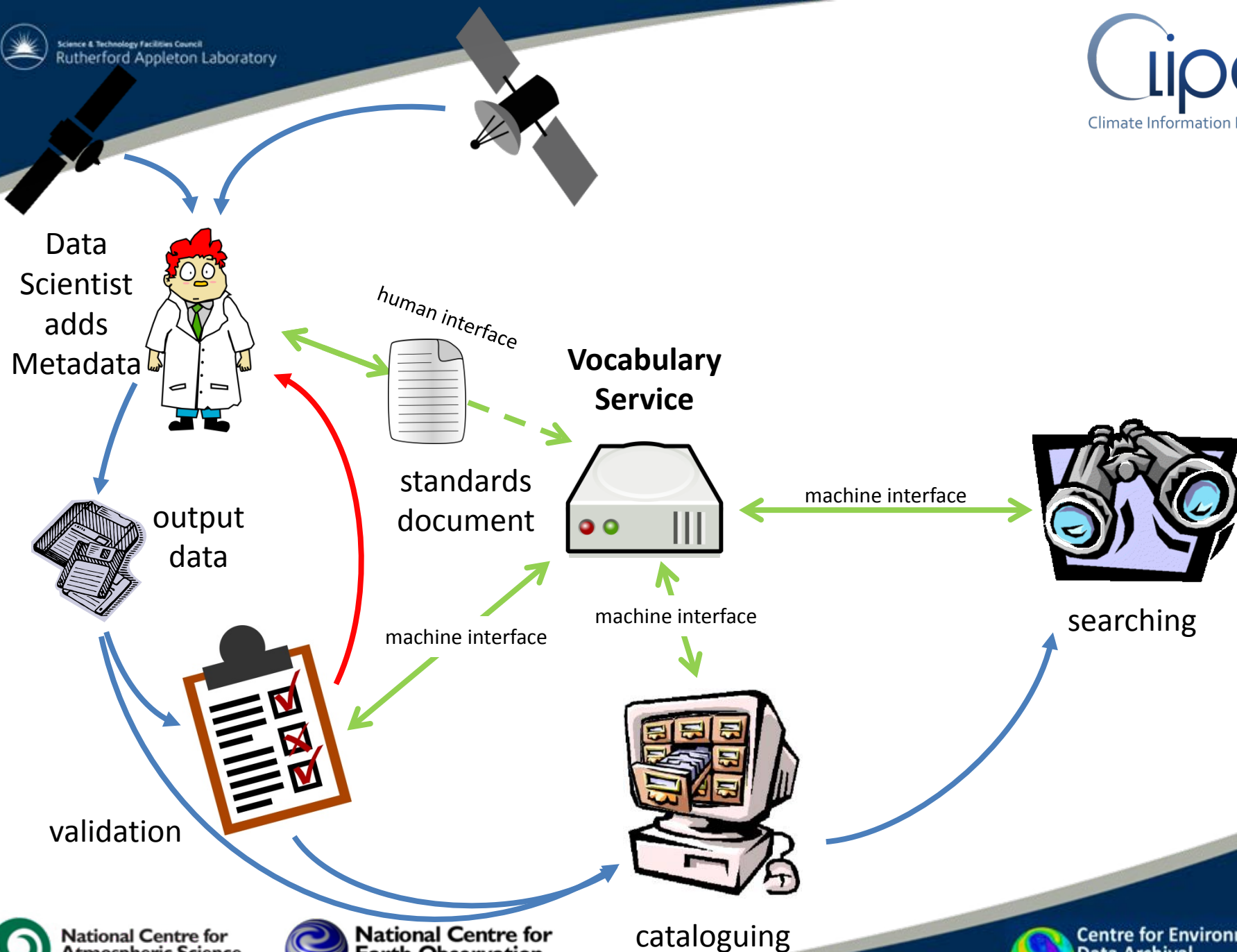
- SKOS is being used to represent these
  - well suited to deal with their complexity and variety
- SKOS provides vocabulary-building tools to:
  - represent managed definitions
  - represent relationships between similar terms used in different projects
  - represent hierarchies and navigation within hierarchies of terms
- SKOS service holds a Triple-store with:
  - SPARQL endpoint to enable querying
  - Web-interface to view concepts and browse links



# SKOS example

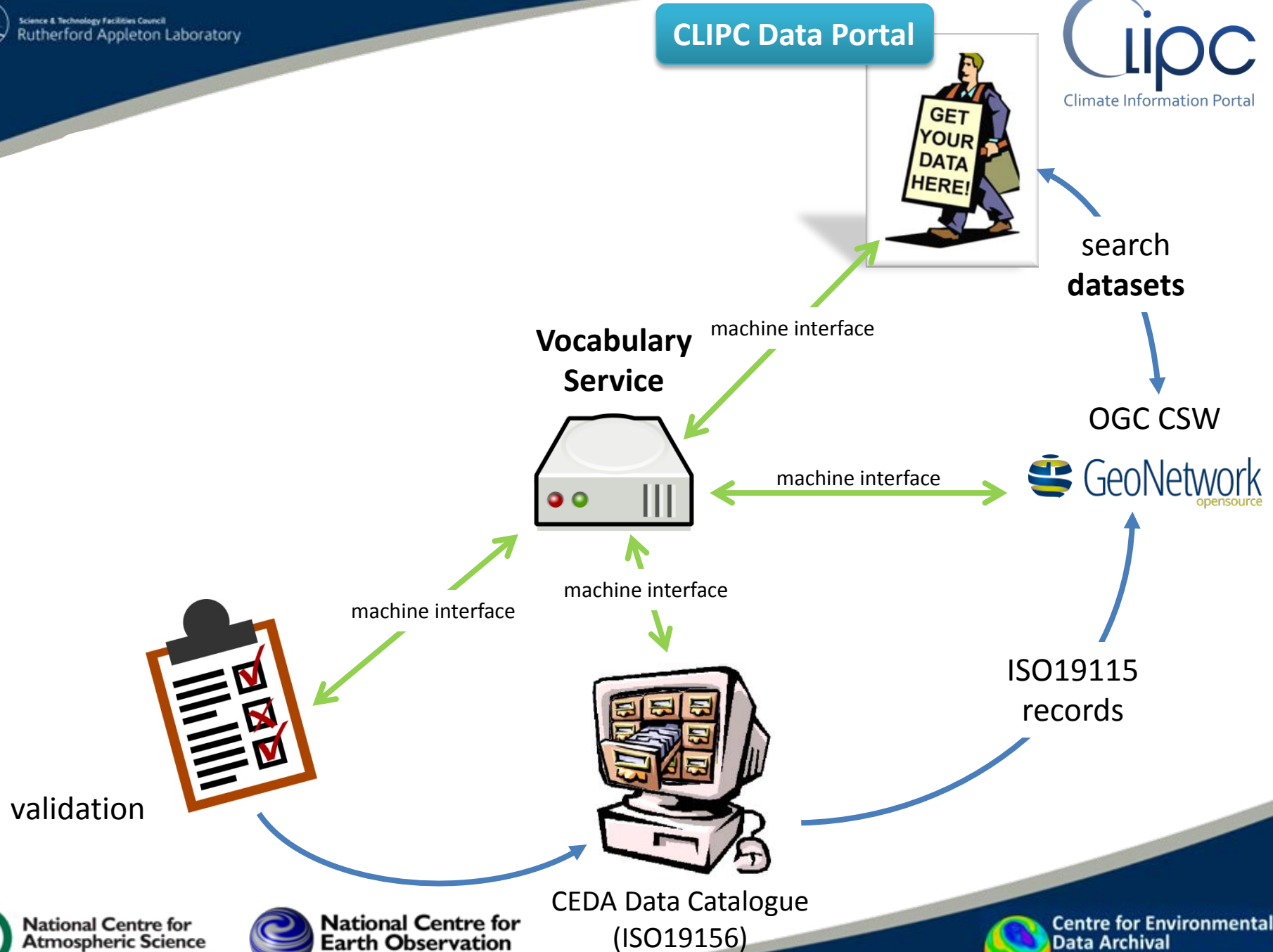








## CLIPC Data Portal







CLIPC Data Portal

search  
data files

Solr

ESGF  
Earth System Grid Federation

Publish  
to ESGF

validation



Vocabulary  
Service

machine interface

machine interface

machine interface



CEDA Data Catalogue  
(ISO19156)

search  
datasets

OGC CSW

GeoNetwork  
opensource

ISO19115  
records

# CLIPC portal: example

## Dataset details

### General

Title	ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI): Analysis long term product version 1.0
ID	485
CSW Identifier	0888e4b4-90b2-4465-84af-97728bd754a7
Source	ceda
Start date	19910901
Start time	00:00:00
End date	20101231
End time	00:00:00
Restrictions	Public; Licence: <a href="http://licences.ceda.ac.uk/image/data_access_condition/esacci_sst_terms_and_conditions.pdf">http://licences.ceda.ac.uk/image/data_access_condition/esacci_sst_terms_and_conditions.pdf</a>
Restrictions other	Public; Licence: <a href="http://licences.ceda.ac.uk/image/data_access_condition/esacci_sst_terms_and_conditions.pdf">http://licences.ceda.ac.uk/image/data_access_condition/esacci_sst_terms_and_conditions.pdf</a>
Abstract	<p>The ESA Sea Surface Temperature Climate Change Initiative (ESA SST CCI) dataset accurately maps the surface temperature of the global oceans over the period 1991 to 2010, using observations from many satellites. The data provides an independently quantified SST to a quality suitable for climate research. The ESA SST CCI Analysis Long Term Product consists of daily, spatially complete fields of sea surface temperature (SST), obtained by combining the orbit data from the AVHRR and ATSR ESA SST CCI Long Term Products, using optimal interpolation to provide SSTs where there were no measurements. These data cover the period between 09/1991 and 12/2010. This dataset is cited in: Merchant, C. J., Embury, O., Roberts-Jones, J., Fiedler, E., Bulgin, C. E., Corlett, G. K., Good, S., McLaren, A., Rayner, N., Morak-Bozzo, S. and Donlon, C. (2014), Sea surface temperature datasets for climate applications from Phase 1 of the European Space Agency Climate Change Initiative (SST CCI). Geoscience Data Journal. doi: 10.1002/gdj3.20 Please note that this dataset has now been superseded by the version 1.1 product, available from <a href="http://catalogue.ceda.ac.uk/uuid/c65ce27928f34ebd92224c451c2a8bed">http://catalogue.ceda.ac.uk/uuid/c65ce27928f34ebd92224c451c2a8bed</a></p>
Dataquality information	Data were processed by the ESA CCI SST project team and supplied to NEODC by the UK Met Office.
Keywords	CLIPC; ESACCI; AATSR; AVHRR; NOAA-4th; NOAA-14; NOAA-18; ERS-2; ATSR; ERS-1; NOAA-16; NOAA-15; Envisat; Level 4; OSTIA; sea surface temperature; sea water temperature; Metop-A; NOAA-5th; Environmental Satellite; day; NOAA-12; ERS; ESACCI_SST; Metop; NOAA-17
Keywords	orthoimagery

# Enriched metadata links

- Metadata record links back to SKOS Vocabulary Service:  
<http://vocab-test.ceda.ac.uk/>

## AATSR

**IRI:** [http://vocab-test.ceda.ac.uk/collection/cci/sensor/sens\\_aatsr](http://vocab-test.ceda.ac.uk/collection/cci/sensor/sens_aatsr)

Advanced Along-Track Scanning Radiometer

**type**

[Sensor](http://www.w3.org/2004/02/skos/core#Concept), <http://www.w3.org/2004/02/skos/core#Concept>

**is top concept in scheme**

[Sensor](#)

**is in scheme**

[Sensor](#)

**has platform**

[Envisat](#)


**see also**

<http://www.wmo-sat.info/oscar/instruments/view/2>

Relationships and definitions are available via SKOS service



# Link to authoritative sources



## OSCAR

Observing Systems Capability Analysis and Review Tool

[Home](#) | [Observation Requirements](#) | [Space-based Capabilities](#) | [Surface-based Capabilities](#) |

[Overview](#) | [Programmes](#) | [Satellites](#) | [Instruments](#) | [Frequencies](#) | [Agencies](#) | [Gap Analyses](#)


### Instrument: AATSR

#### Instrument details

<b>Acronym</b>	AATSR		
<b>Full name</b>	Advanced Along-Track Scanning Radiometer		
<b>Purpose</b>	Multi-purpose VIS/IR imagery, with emphasis on very accurate sea surface temperature for climate		
<b>Short description</b>	Two views (close-to-nadir and fore- for accurate atmospheric corrections), 7 channels, balanced VIS, NIR, SWIR, MWIR and TIR, narrow swath [see detailed characteristics below]		
<b>Background</b>	Evolution of ATSR on ERS-1 and identical to ATSR-2 on ERS-2		
<b>Scanning Technique</b>	Conical oblique (cross-nadir and 47° fore); 2000 pixel of 1 km ssp; swath 500 km; scan rate 400 rpm		
<b>Resolution</b>	1 km IFOV		
<b>Coverage / Cycle</b>	Global coverage in 3 days (IR) or 6 days (VIS)		
<b>Mass</b>	101 kg	<b>Power</b>	100 W
		<b>Data Rate</b>	625 kbps

#### Satellites this instrument is flying on

*Note: a red tag indicates satellites no longer operational, a green tag indicates operational satellites, a blue tag indicates future satellites*

- [Environmental Satellite](#) (ESA)
  -  [Envisat](#) (see instrument status) Mar 2002 - Apr 2012

#### Instrument classification

- Earth observation instrument
  - Passive optical radiometer or spectrometer
    - Moderate resolution optical imager

#### Mission objectives

**Primary mission objectives**

- Cloud cover
- Cloud optical depth



# Summary

- Linked data technologies for heterogeneous data:
  - help manage DRS-related Controlled Vocabularies
  - enrich metadata records exposed to users
  - allow common terms to be used in different infrastructures, e.g. OGC CSW or ESGF
  - can be accessed dynamically by services
- Data that has facets of “multiples” needs more thought/attention.







# Thank you, any questions?

CLIPC:

<http://www.clipc.eu>

Vocabulary service at STFC:

<http://vocab-test.ceda.ac.uk/>

Source vocabularies:

<https://github.com/cedadev/cci-vocabularies/tree/master/data>

Climate Impact Indicator metadata:

<https://github.com/cerfacs-globc/impact-indicators>

