ERA-CLIM2 M48 Review meeting

December 2017

WP5: Service Developments

Manuel Fuentes
Products Team – Forecast Department - ECMWF

ECMWF
EUROPEAN CENTRE FOR MEDIUM RANGE WEATHER FORECASTS

ERA CLIM 2

SEVENTH FRAMEWORK PROGRAMME

WP5 Service Development

- Development of data services and visualisation tools for coupled reanalysis output.
- Identification of user requirements; monitoring and analysis of usage statistics.

This work package addresses development of efficient and sustainable data and web services that are essential to the provision of full and open access to the data and information products developed in the project.

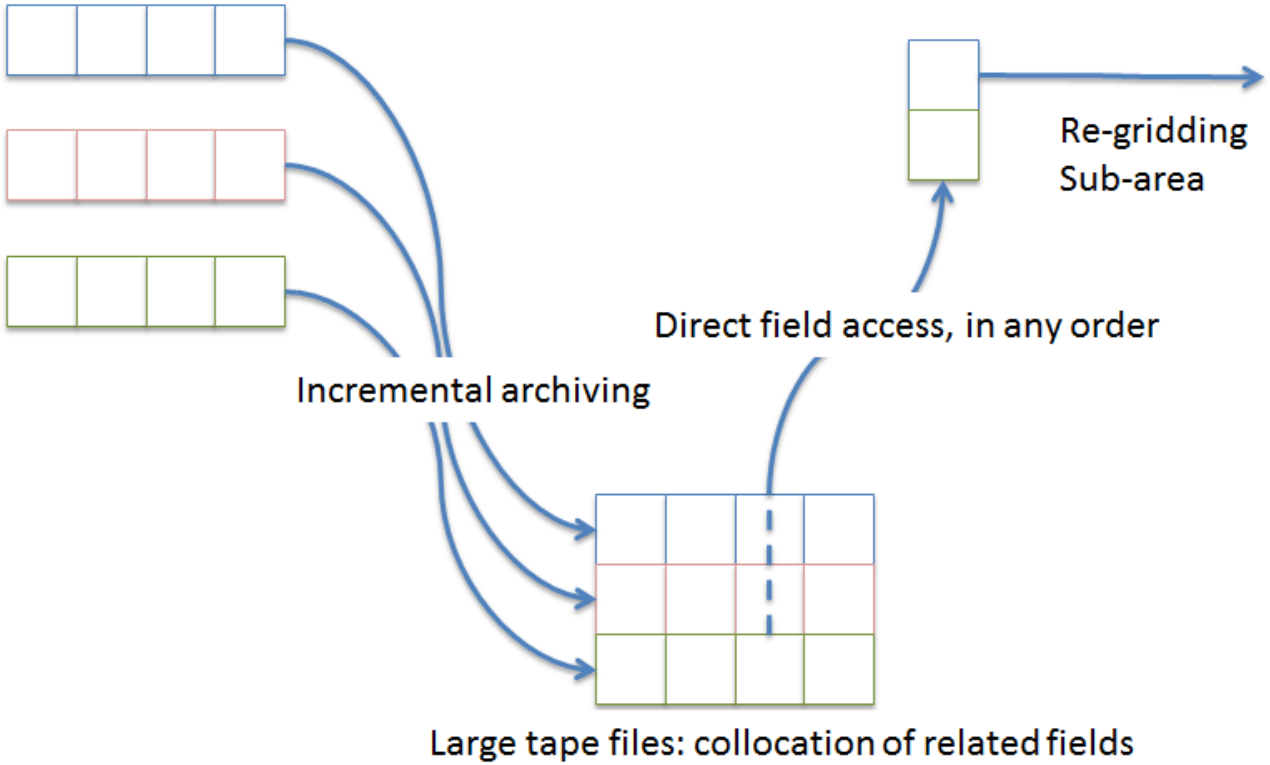
WP5 Deliverables

- D5.1 : Technical developments in MARS to **support archiving and retrieval of data in NetCDF** format. This is needed to allow use of MARS for the ocean component of the coupled reanalysis output. [***M36***]

- D5.2 : Implementation of **public data services** for gridded output from coupled climate reanalyses (CERA) [***M48***]

- D5.3 : Report on **data services usage** and requirements for climate reanalysis [***M48***]

Service offered by MARS



- MARS archives atmospheric field of CERA-20C (handles 2D fields in GRIB)
- MARS scans archived files, extracts metadata from GRIB headers and keeps an index that tracks where each GRIB field is
- GRIBs are reorganised into larger files, to minimise the total number of files and collocate related fields to speed up retrievals
- On retrieval, MARS find the required fields, reads them from tape, and re-assembles them according to the user's request

D5.1 Support for NetCDF in MARS

Original NEMO output files contain many variables (2D, 3D), feedback files, restart files, ocean observations, all annotated with NEMO's own convention

Solution considered:

- NetCDF files are split into individual NetCDF files, 2D or 3D
  - Resulting NetCDF files must follow an agreed convention based on CF
  - Resulting NetCDF files are annotated with MARS specific information. These attributes are used by MARS to index the NetCDF files, and treat them as simple binary records
- On retrieval, those records will be assembled in a single NetCDF file to be delivered to the user

Challenges:

- Define what variables to archive, focusing on user service
  - not all output above is interesting to users
- Find CF standard names
  - About 60% variables have a CF standard name
- Find sound metadata to enable assembling records on retrieval

CERA20C Ocean data in NetCDF: Challenges for an archive

- Define what variables to archive, focusing on user services:
  - 53 Variables from the monthly means
- Many variables did not have an official CF standard name
  - Requested to the CF list Sep 2016, approved June 2017
- No adequate description of time in CF: monthly means of daily means
  - Nothing found in the literature
  - Experts of the CF community still debating how best to encode them
- Finding an accurate description of the tripolar ORCA grid

- During 2017, engaged with CF-NetCDF experts to produce metadata recommendations for encoding NetCDF products using the CF convention
  - Using examples from CERA20C ocean, UERRA, C3S products
  - Forms the basis of the Common Data Model for the CDS Toolbox
  - … and many other datasets: S2S, HRES, ENS, ORAS4, ORAS5, C3S seasonal,

- Finally! Ready to start consolidation

D5.2 Data Services

Consolidation of CERA-20C into user version

- After production completes and quality of data is checked, the various CERA-20C streams are consolidated into a single version:
  - Retrieve all relevant data (GRIB Fields, ODB feedback)
  - Re-label headers with the user version (0001), filtering out un-wanted parameters/steps/cycles
  - Archive back into MARS

Extend Public Data Server

- Public interface to data stored in MARS

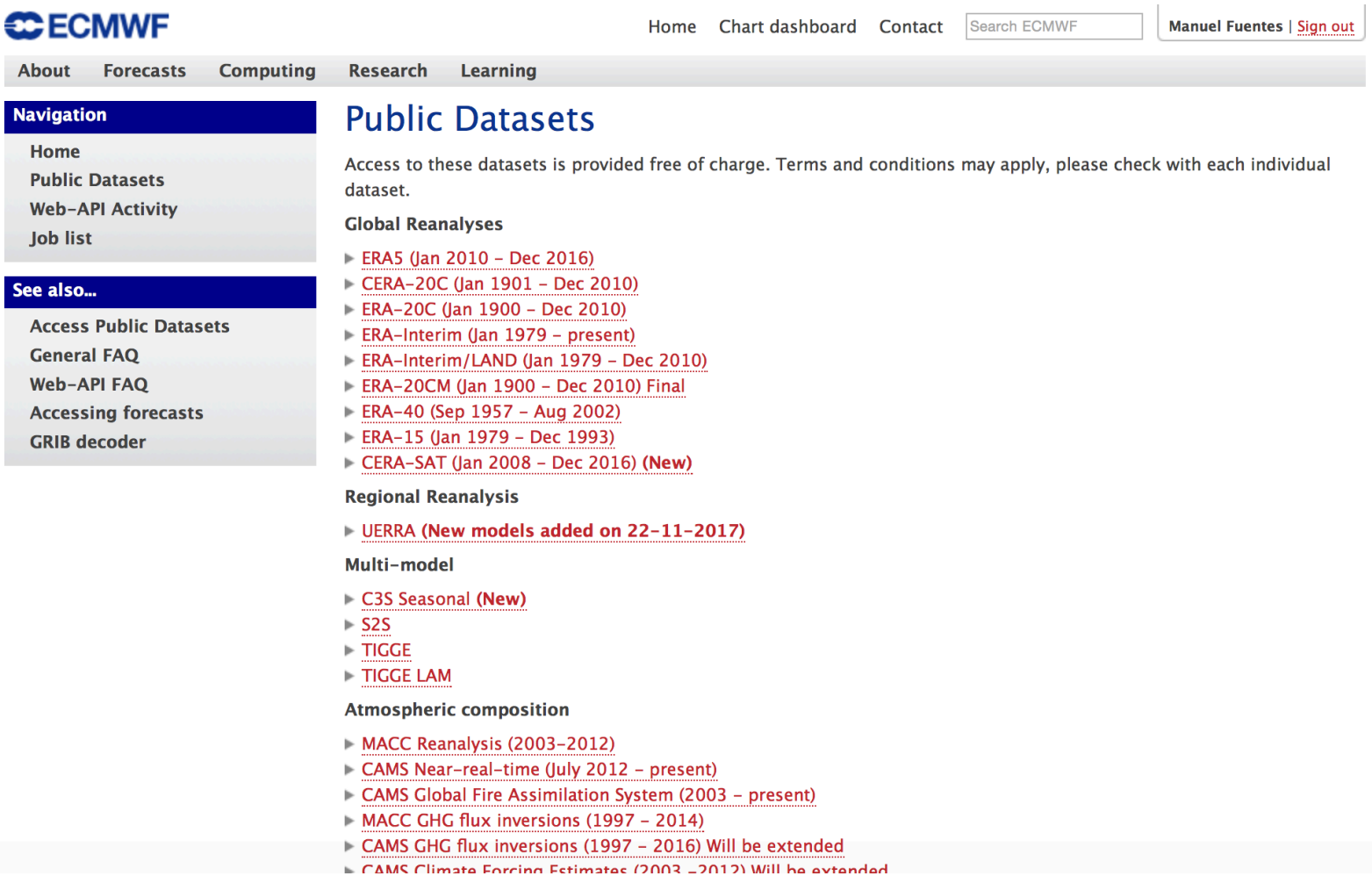
CERA-20C Consolidation

- Done in stages, priority given to most popular data, in order to facilitate access to data
  - First all monthly means
  - Then all Analysis, Ensemble means and stddev (surface, pl)
  - All Analysis, Ensemble means and stdev (ml, isentropic levels)
  - Forecast pressure level
  - Forecast surface
  - Forecast model level
  - Observation feedback
- 1 Pbyte of gridded data in GRIB (atmospheric + ocean wave fields)
- 20 Tbytes Observational feedback in ODB
- Consolidation took 5 months (Dec 2016 - Apr 2017)

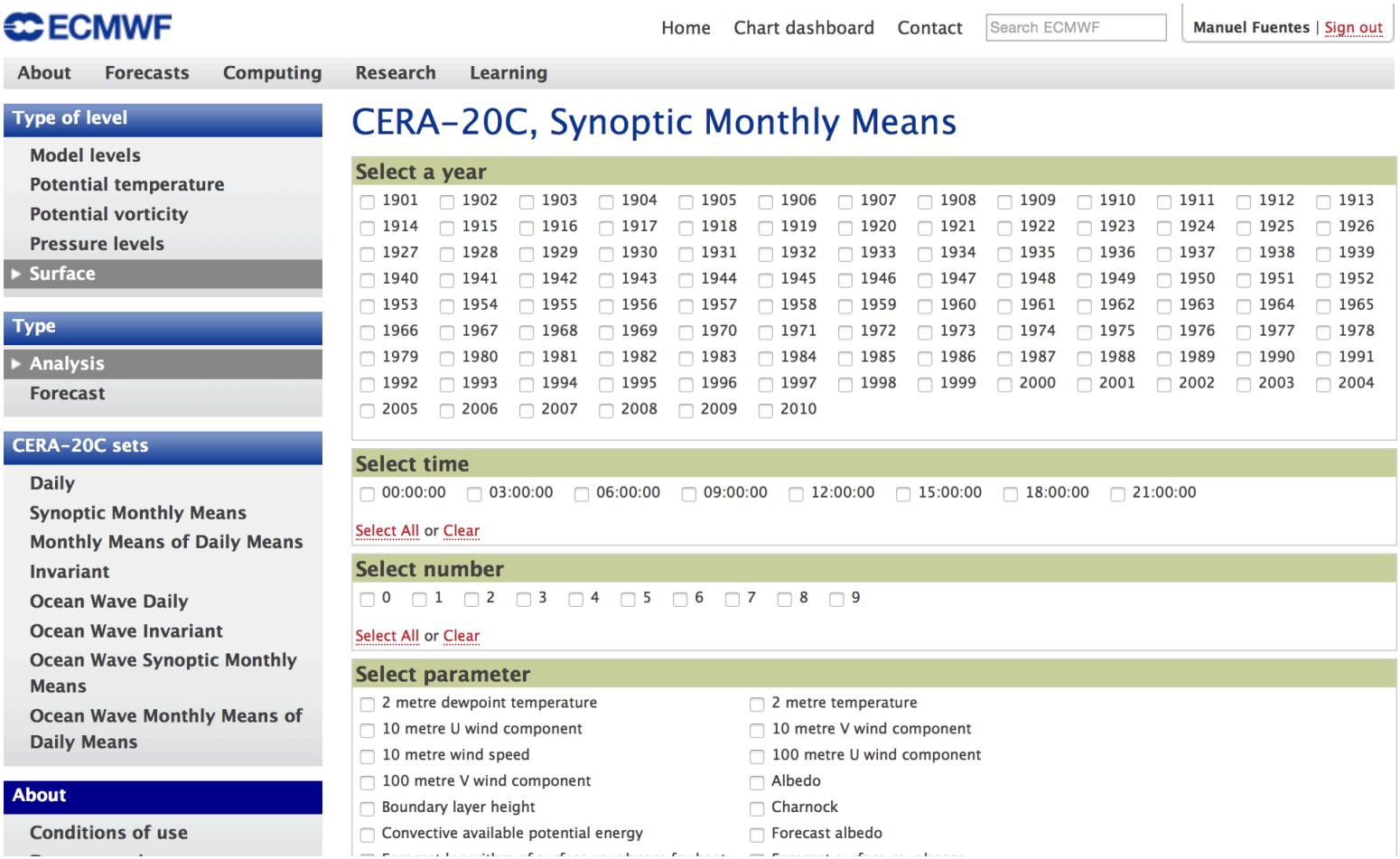
CERA-SAT Consolidation

- 450 Tbytes of gridded data in GRIB (atmospheric + ocean wave fields)
- 11 Tbytes Observational feedback in ODB
- Started November 2017
- On-going:
  - Monthly means
  - Observation feedback

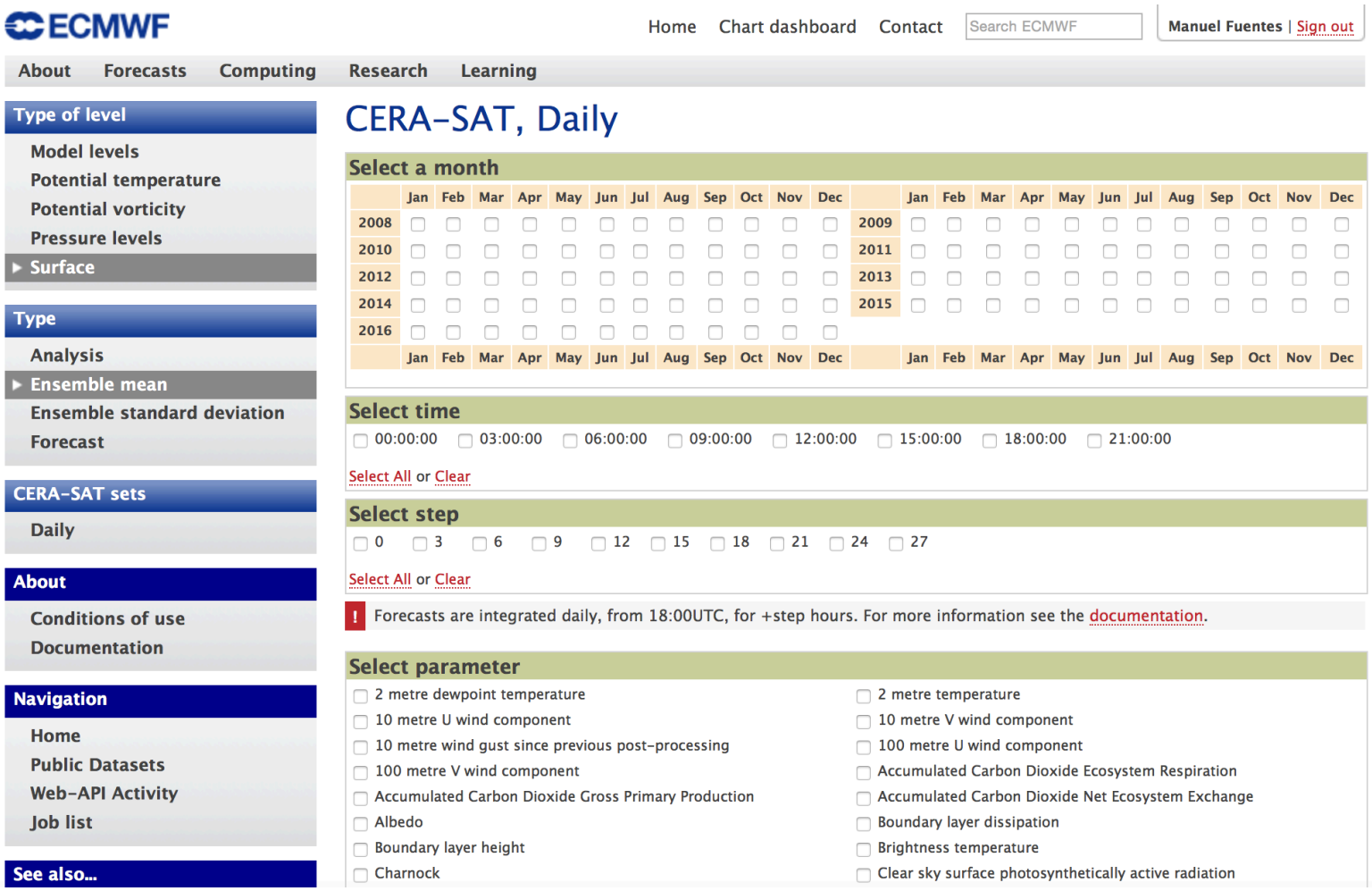
Public Datasets



Public Datasets: CERA-20C



Public Datasets: CERA-SAT



Public Datasets: ECMWF WebAPI

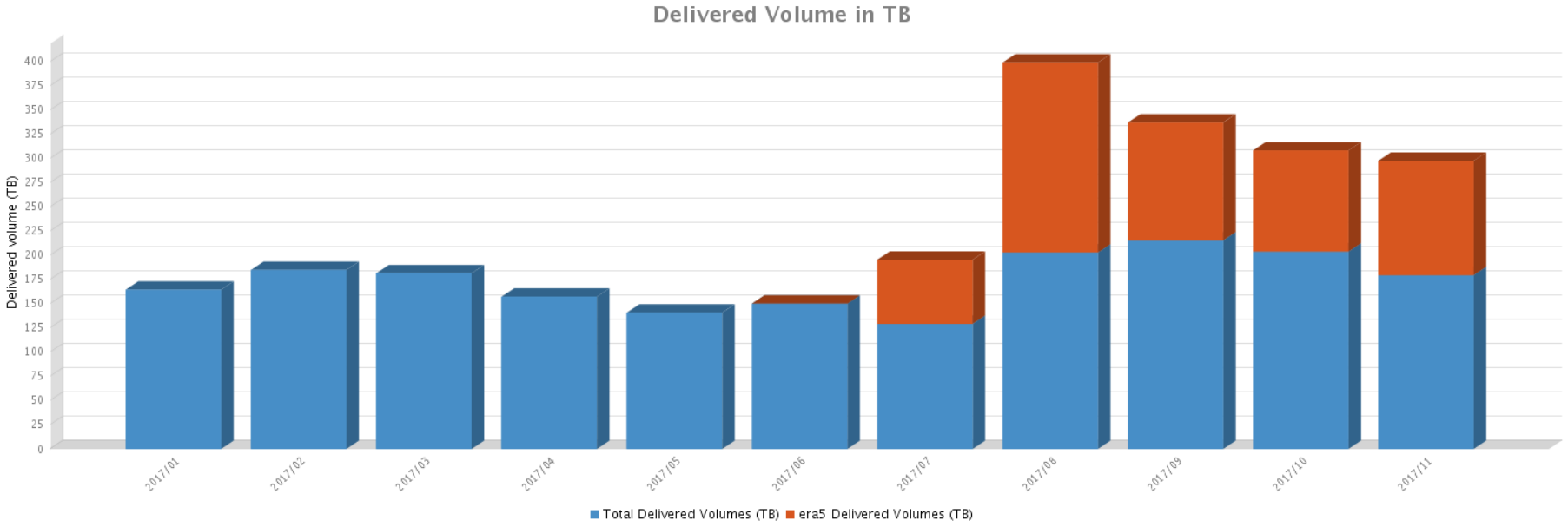
- Simple API to services using HTTPS
  - Install a simple library + token
  - Download data via batch scripts

```
#!/usr/bin/env python

from ecmwfapi import ECMWFDataServer
server = ECMWFDataServer()
server.retrieve({
    'dataset'   : "cera20c",
    'stream'    : "edmm",
    'levtype'   : "sfc",
    'date'      :
"19010101/19010201/19010301/19010401/19010501/19010601/19010701/19010801/19010901/19011001/19011101/19011201",
    'time'      : "00",
    'param'     : "2t",
    'grid'      : "1/1",
    'format'    : "netcdf",
    'target'    : "data.nc"
    })
```

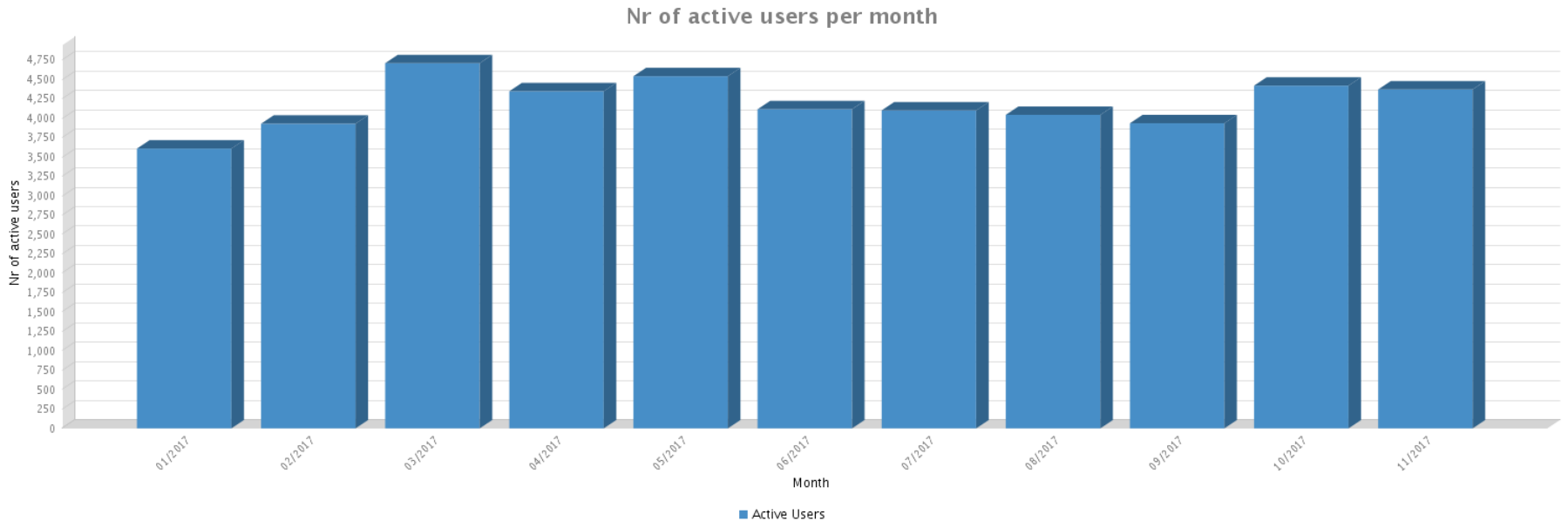
D5.3 Report on data services usage: Statistics Framework

- Statistics on Public Data Servers
  - Stored all requests in a database, for further analysis
  - Performed statistics on popular datasets (ERA-Interim, etc…)
- Comparison of delivered volumes to users per month for ERA5 vs all other datasets for the year 2017
  - Clear impact of ERA5 (2010-2016) release in July 2017



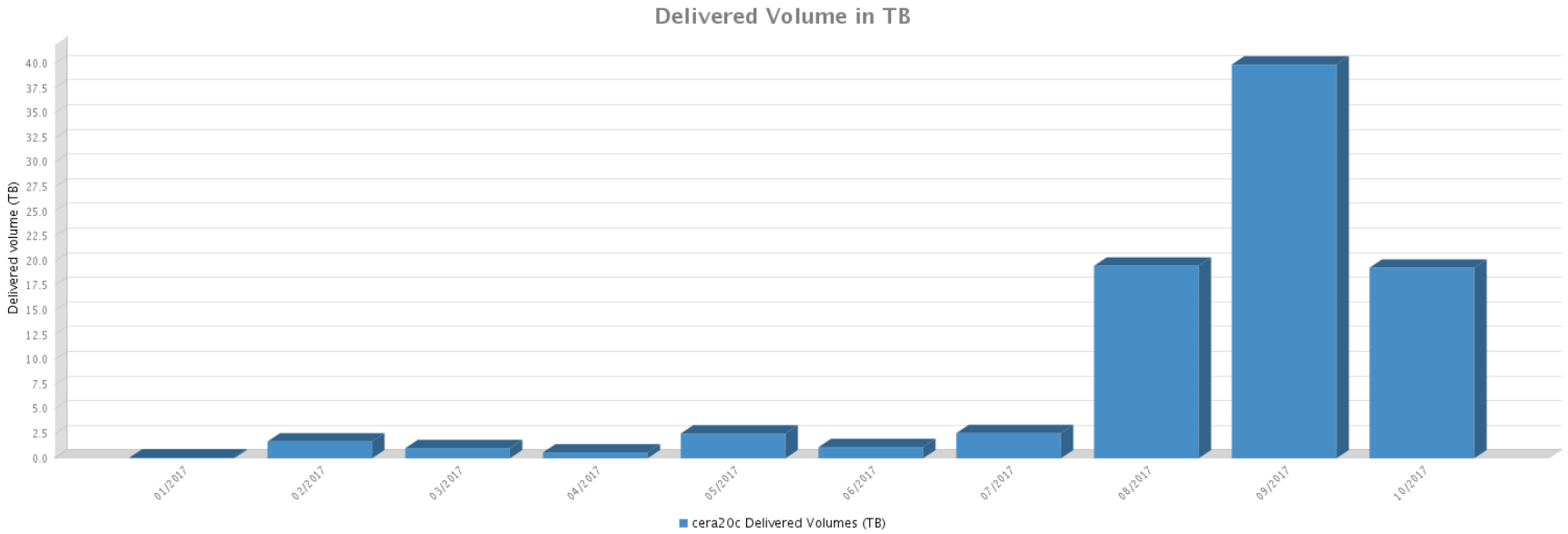
Statistics Framework: Total number of users

- Total number of active users per month for all datasets
  - Active user: executes 3 or more retrievals



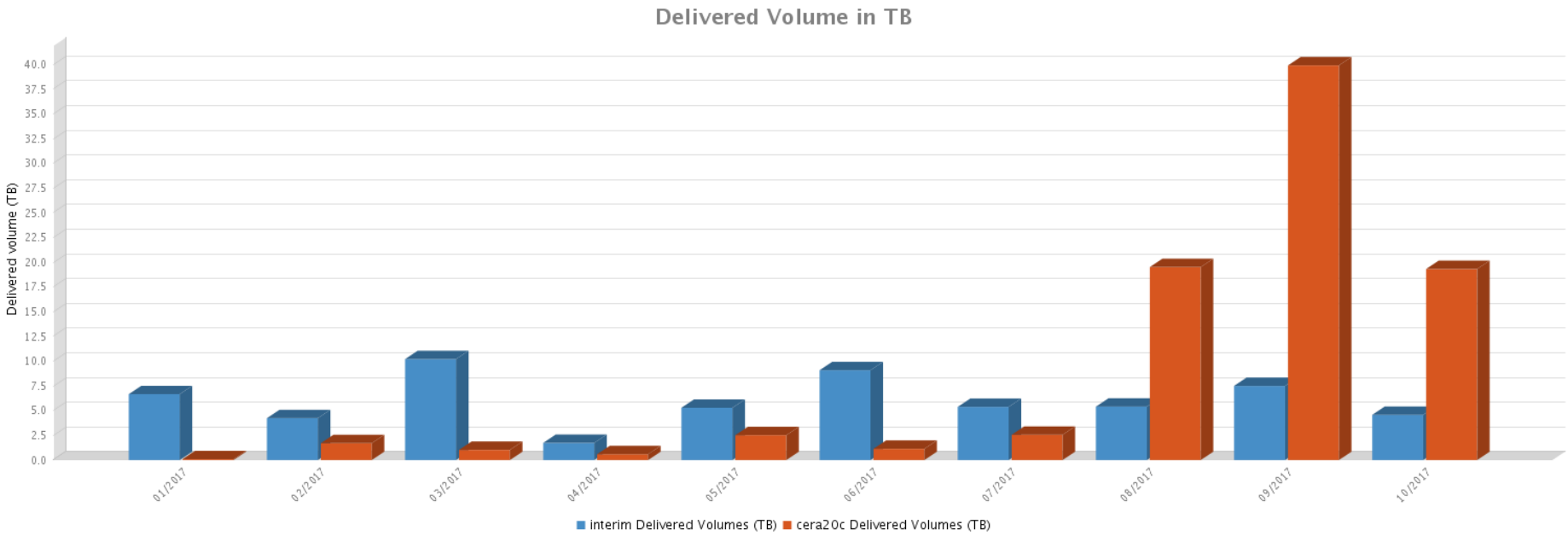
CERA-20C Data Service Usage: delivered data

- Good uptake of CERA-20C by users
- Almost 40 TBytes delivered to users in September 2017



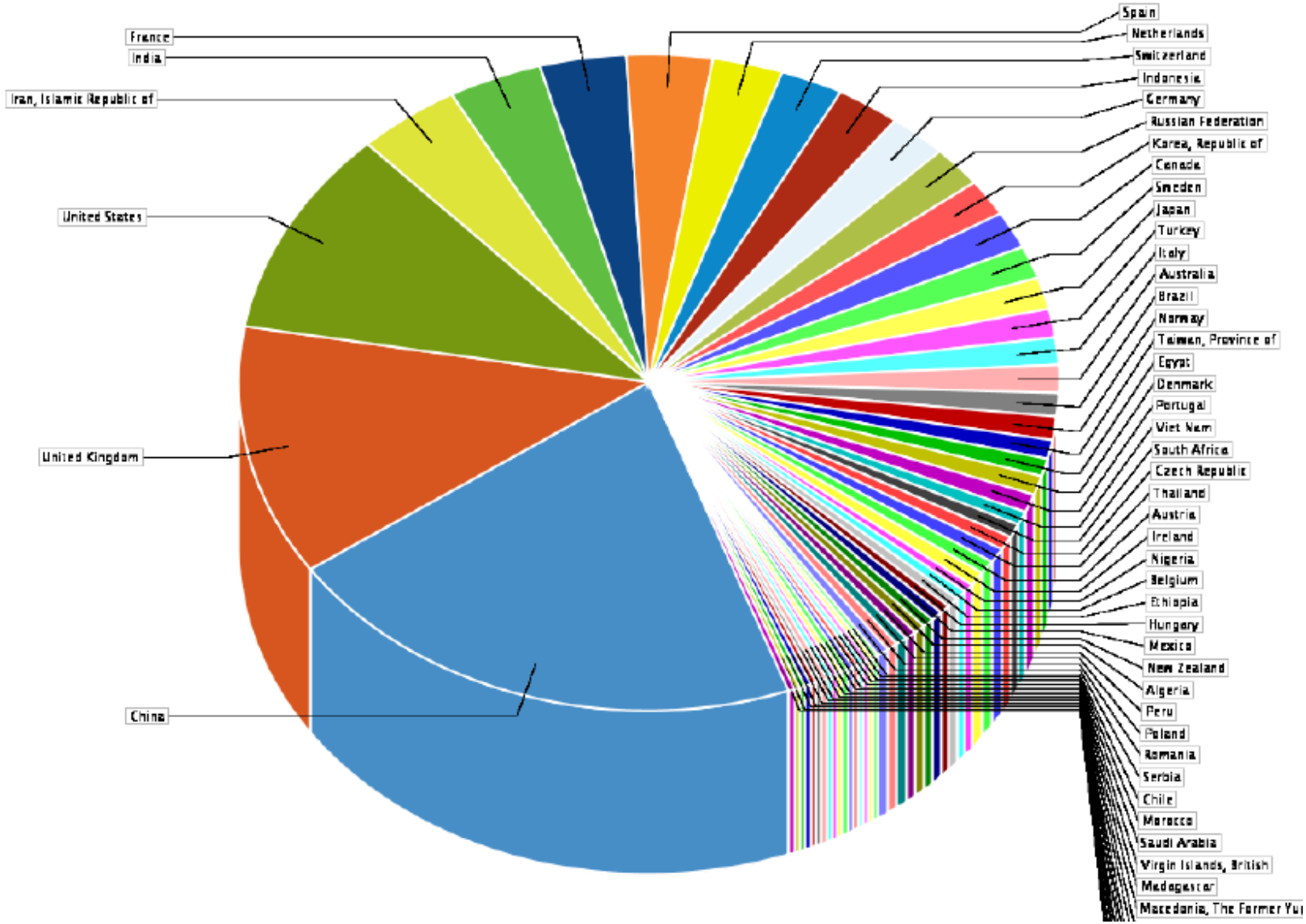
CERA-20C Data Service Usage vs ERA-20C

- Compared with ERA-20C, first indications show CERA-20C is more popular



CERA-20C Data Service Usage: User distribution by country

- Total number of users registered to CERA-20C: **259**

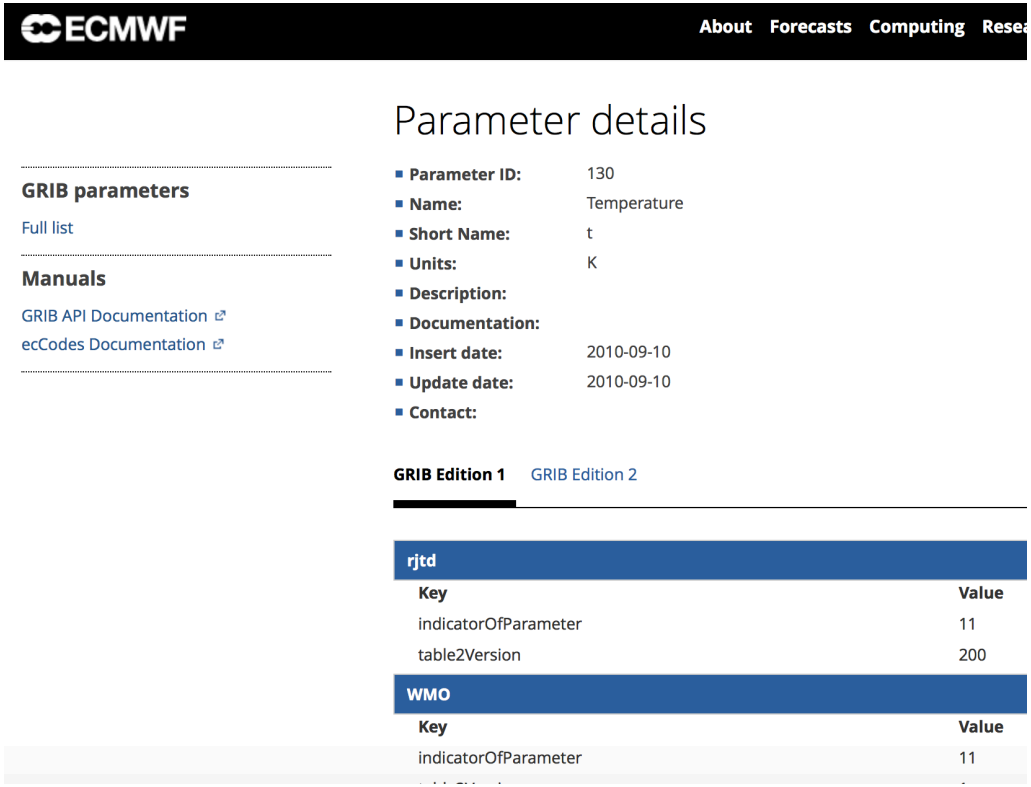


Data Services usage: Study on popular parameters

- Project that aims to improve the description of reanalysis parameters archived in MARS. Start by the top 90 parameters, with a review process including internal and external review (15 reanalyses users)
  - Find the most retrieved parameters for popular datasets (eg, ERA-Interim)
  - The following 8 variables represent more than 6% (out of ~12,000 variables)

parameter	acronym	%
10 metre u-wind component	10u	1.25 %
10 metre v-wind component	10v	1.19 %
2 metre temperature	2t	1.17 %
total precipitation	tp	0.70 %
mean sea level pressure	msl	0.52 %
u-wind component at 850 hPa	u	0.50 %
v-wind component at 850 hPa	v	0.47 %
t at 850 hPa	t	0.43 %

Data Services usage: Improve parameter description



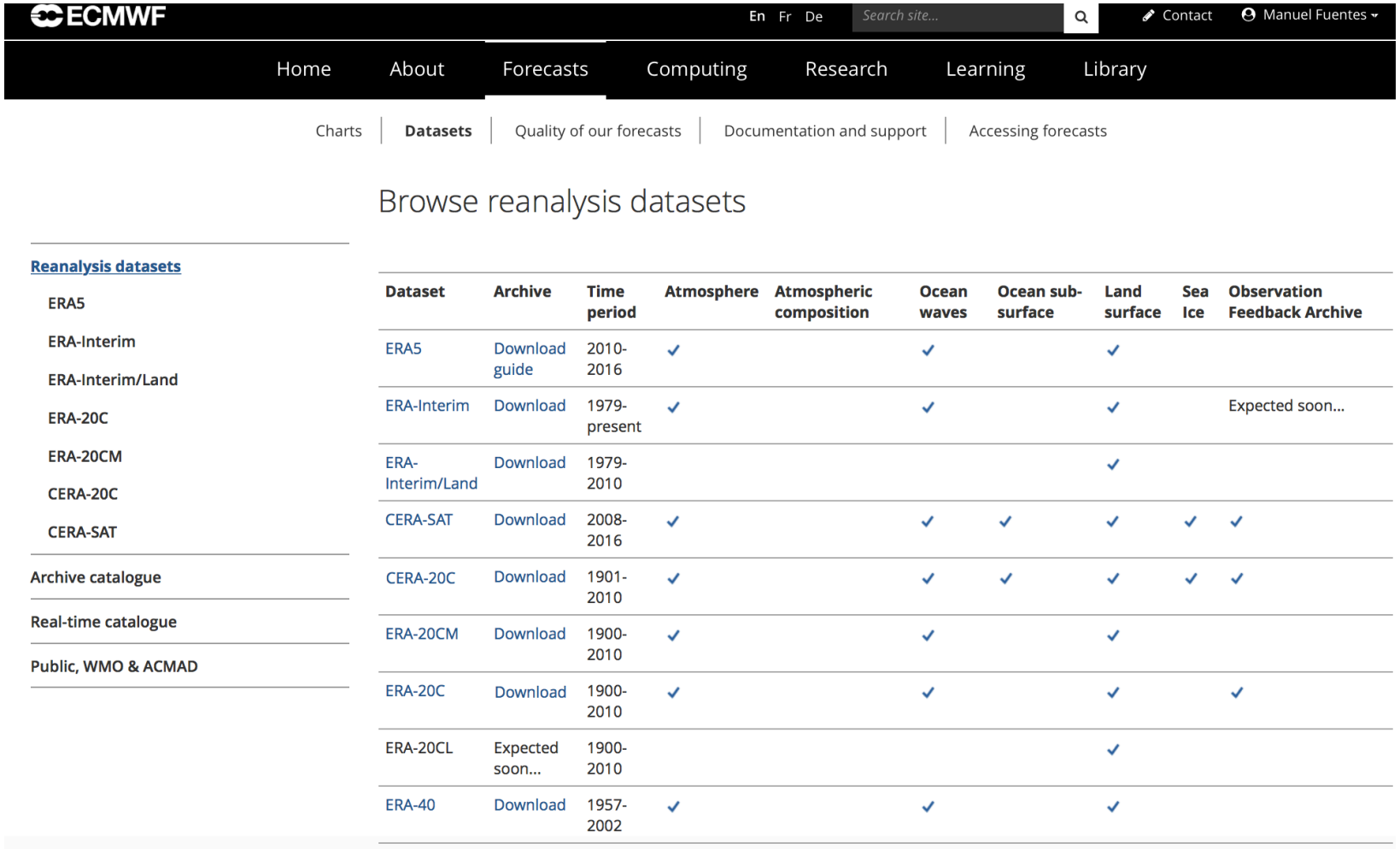
Parameter name	**Parameter ID**
Temperature	130
Draft Definition (FH, Version 5, 5 December 2017)	
The temperature in the atmosphere. This parameter is available at all model levels, from close to the Earth's surface to the top of the atmosphere. This parameter has units of kelvin (K). Temperature measured in kelvin can be converted to degrees Celsius (°C) by subtracting 273.15. Units: kelvin (K)	
Keywords: TEMPERATURE, HEAT, COLD	

ERACLIM2 WP5: Service Development - Summary

- WP5 Deliverables:
  - D5.1 MARS support for NetCDF (M36)
    - Technical development
  - D5.2 CERA Data Servers
    - CERA-20C (M40)
    - CERA-SAT (M48)
    - Ocean component Q1 2018
  - D5.3 Data services usage (M48)
    - CERA-20C
    - Framework to build customised reports

ERACLIM2 Data Services: Not the end, but the beginning!

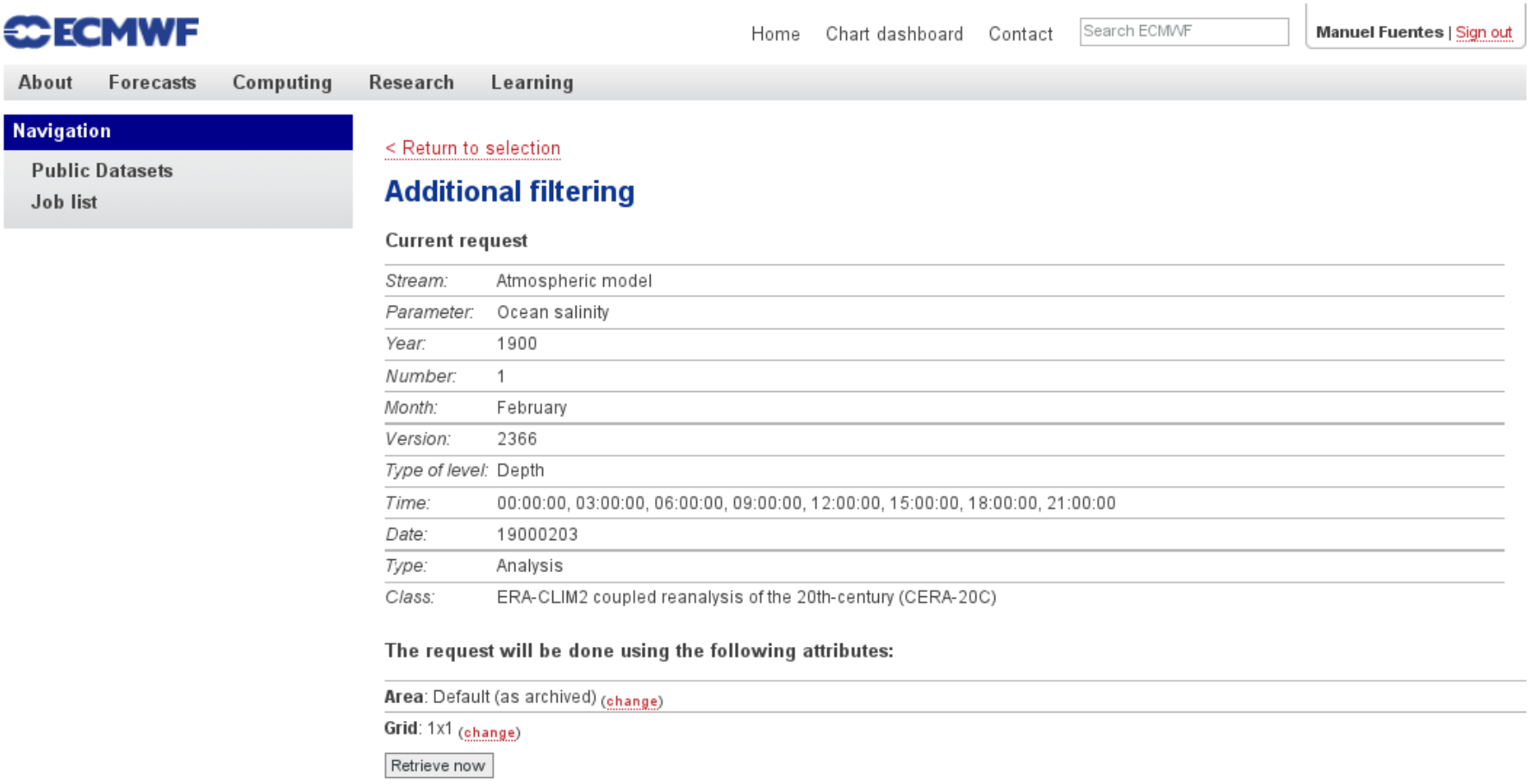
- Two Coupled Earth System Re-analyses available for the scientific community
- Foundations for Copernicus Climate Change Services



Dataset	Archive	Time period	Atmosphere	Atmospheric composition	Ocean waves	Ocean sub-surface	Land surface	Sea Ice	Observation Feedback Archive
ERA5	Download guide	2010-2016	✓		✓		✓		
ERA-Interim	Download	1979-present	✓		✓		✓		Expected soon...
ERA-Interim/Land	Download	1979-2010					✓		
CERA-SAT	Download	2008-2016	✓		✓	✓	✓	✓	✓
CERA-20C	Download	1901-2010	✓		✓	✓	✓	✓	✓
ERA-20CM	Download	1900-2010	✓		✓		✓		
ERA-20C	Download	1900-2010	✓		✓		✓		✓
ERA-20CL	Expected soon...	1900-2010					✓		
ERA-40	Download	1957-2002	✓		✓		✓		

Thank you for your attention!

Prototype of NetCDF Ocean output in MARS



ECMWF's commitment to support NetCDF

- Other projects require NetCDF support
  - All data being served from ECMWF Data Portals
  - Sub-seasonal to Seasonal project (S2S) requires archiving in MARS of Ocean output from 11 production Centres (near real-time + reforecast)
  - Other forecasting systems producing ocean output (HRES, ENS, ….)
  - Seasonal forecast being delivered in NetCDF to the C3S (real-time + reforecast), and will need to be served by the Climate Data Store (CDS)
  - Climate Predictions (CMIP6) will be part of the C3S, will be served by the CDS
- Define common metadata that will allow tools to seamlessly handle all the above
  - Define a MARS/NetCDF convention, on top of CF, CMIP5/6, SPECS

- ERA-CLIM2 D5.1 will enable support for NetCDF at ECMWF
  - Transfer the development outcomes into preoperational phase of C3S