Mitglied der Helmholtz-Gemeinschaft

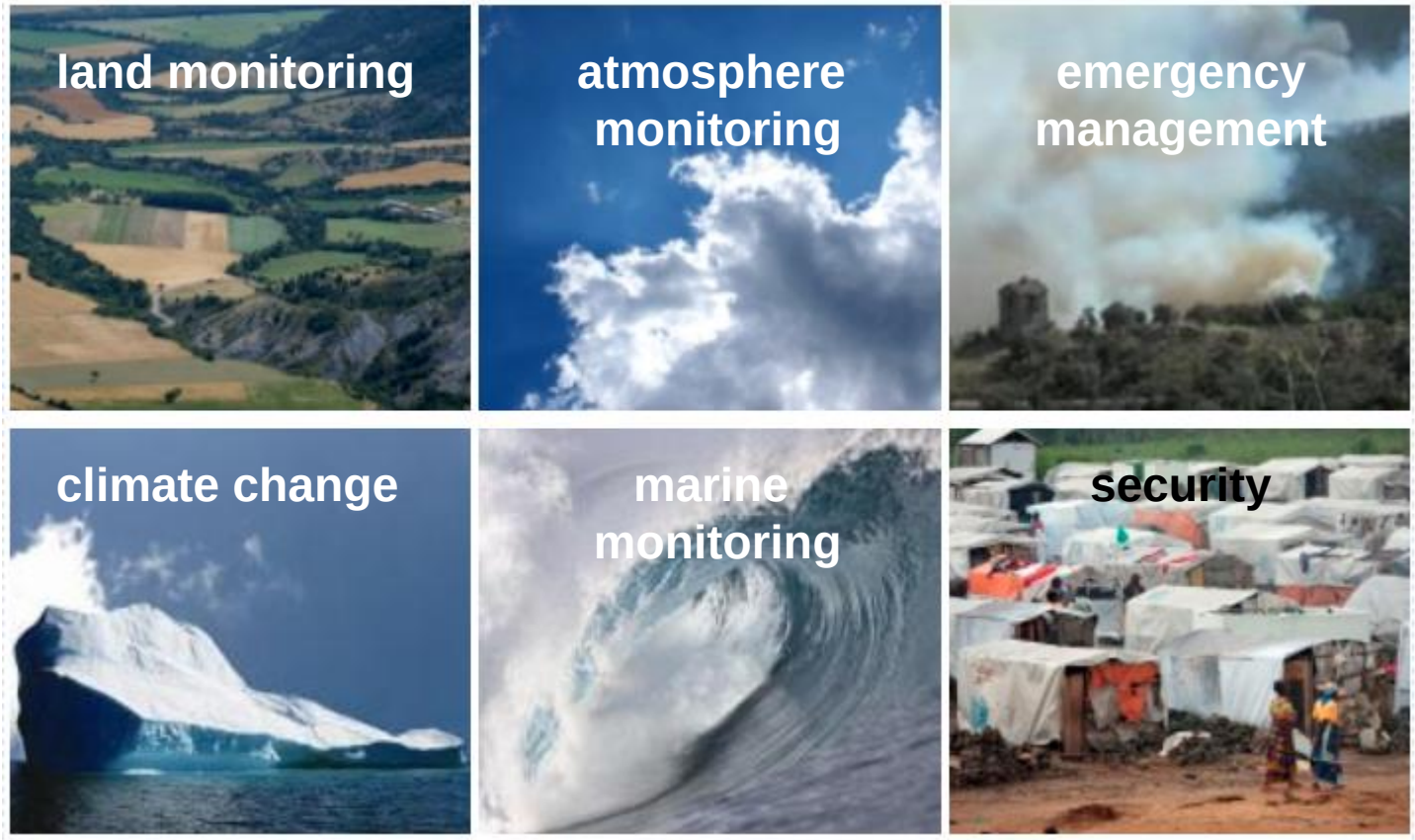
Managing and Sharing Atmospheric Composition Data

Martin Schultz[1], Vincent-Henri Peuch[2], Richard Engelen[2] and the MACC management team

[1] Institute for Energy and Climate research 8: Troposphere (IEK-8), Forschungszentrum Jülich, Germany
[2] European Centre for Medium Range Weather Forecast, Reading, UK

EUDAT User forum, Prague, Czechia, 23-24 April 2014

http://www.copernicus.eu/



Establish a European capacity for Earth Observation

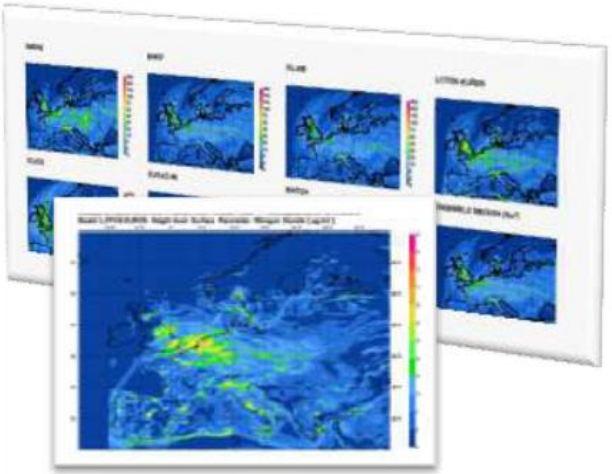
Copernicus consists of a complex set of systems which collect data from multiple sources: earth observation satellites and in situ sensors such as ground stations, airborne and sea-borne sensors. It processes these data and provides users with reliable and up-to-date information through a set of services related to environmental and security issues.

Helping Europe respond to climate change and poor air quality

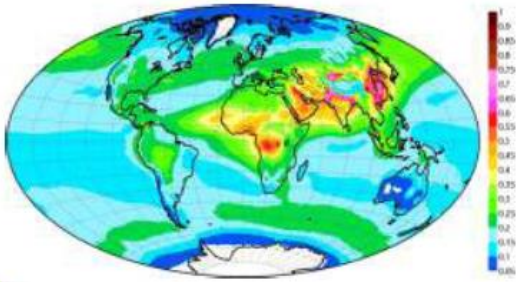
macc
Monitoring atmospheric composition & climate – II

http://atmosphere.copernicus.eu
Online catalogue, quick-looks and data

European Air Quality

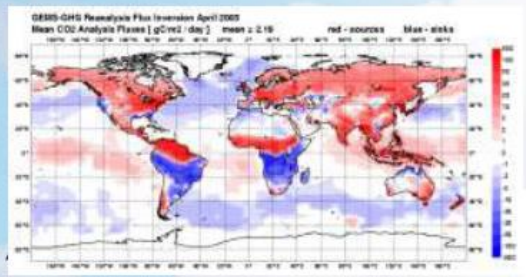


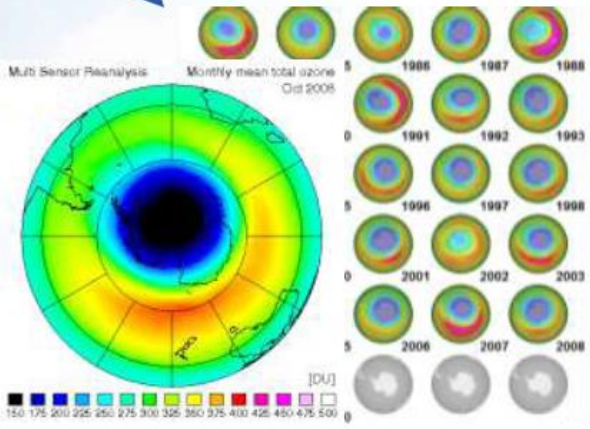
Global atmospheric composition

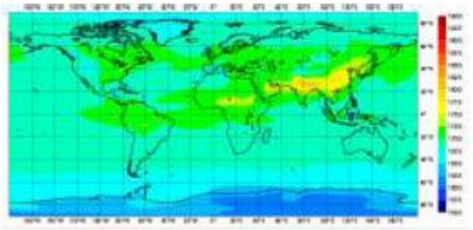


Radiation and ozone layer

Surface fluxes: greenhouse gases, fires, emissions







SEVENTH FRAMEWORK PROGRAMME
European Commission
copernicus

The global observing system for atmospheric composition

Balloons

Airplanes

Satellites

Trains

Ground-based stations

Ships

GEO

GLOBAL ATMOSPHERE WATCH

ACAP

ICOS integrated carbon observation system

IAGOS

ACTRiS

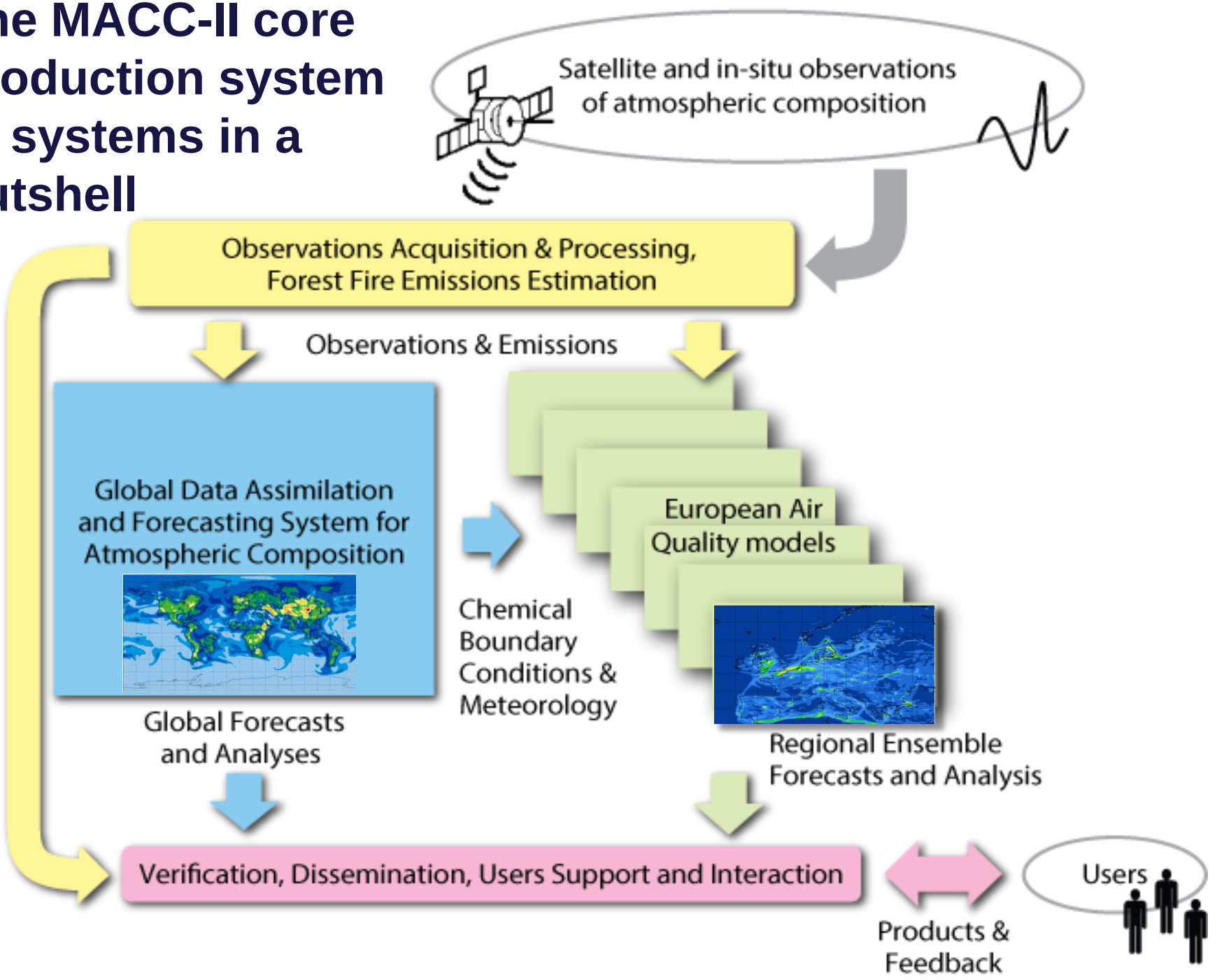
CAPMoN RCEPA The Canadian Air and Precipitation Monitoring Network

U.S. EPA AIRNOW airnow.gov

EEA

AERONET Aerosol Robotic Network

The MACC-II core production system of systems in a nutshell



Atmospheric (Composition) Data

Example - ECMWF: ERA-20C

- Assimilating billions of observations
- Producing 2,400 global forecasts per day
- Generating 1 PB of reanalysis data in 200 days
- Currently serving reanalysis products to 20,000 users

Atmospheric (Composition) Data Properties

Two main types of data:

Model output (and satellite data):
- huge volume
- relatively homogeneous (gridded or swath)
- formats: GRIB, netCDF, HDF

GBytes → TBytes*

Observation data:
- (usually) low volume
- very heterogeneous (time series, profiles, swaths)
- formats: ASCII, Excel, netCDF

kBytes → MBytes*

* per file

Documentation/metadata:
several standards = hardly standardized

Timeliness:
data often required in near realtime (NRT)

More properties

Geolocation:
- global sphere → local coordinates

Calendars and times:
- UTC vs. local time
- verification time, forecast time, model time, actual time
- climatological calendars …

Data flags:
- missing values
- detection limit
- measurement problems

Traceability:
- version control
- calibration history
- work flow

ECMWF's Meteorological Archival and Retrieval System

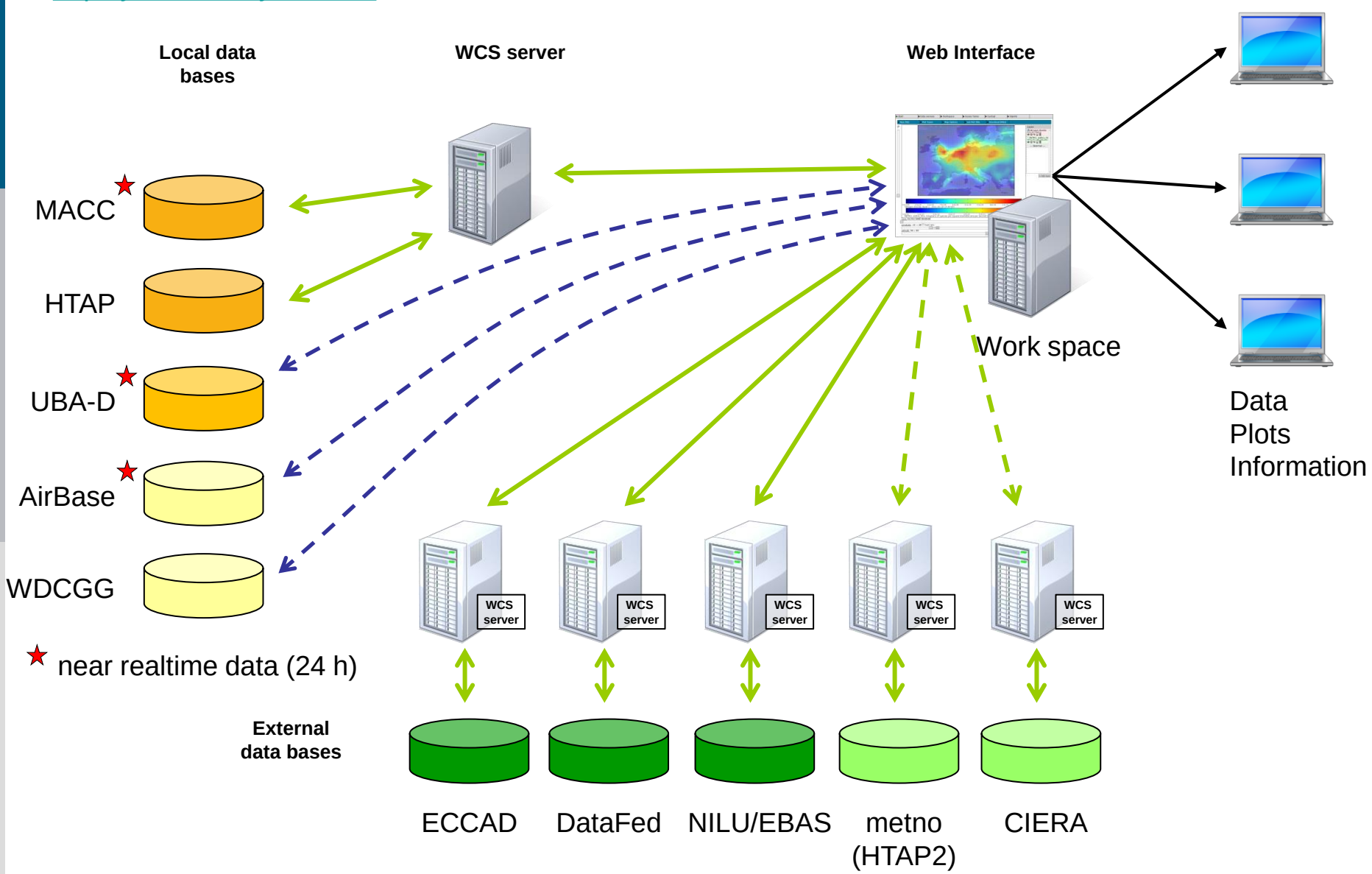
- 10^{14} directly addressable objects
  - Unique hypercube-based indexing
- Data is kept forever:
  - For many studies, a dataset becomes useful once enough data has been accumulated
  - Deleting old data in an exponentially growing archive is meaningless
- 200 million objects/65 TB added daily
- 7000 registered users
- 650 active users, 100 TB retrieved per day, in 1.5 million requests

© ECMWF

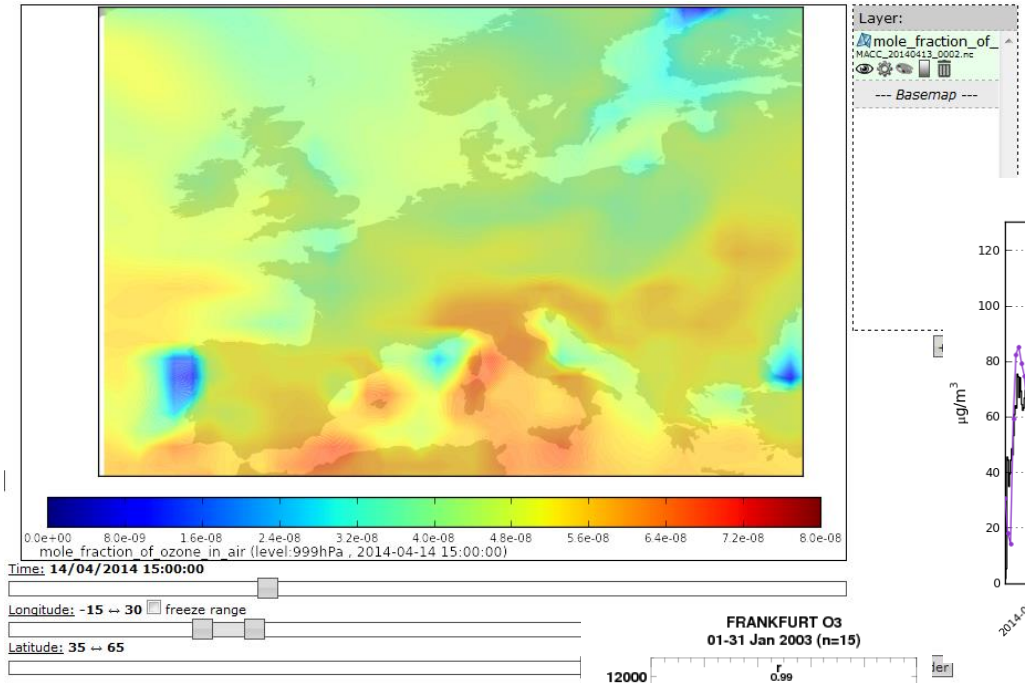
Jülich OWS INterface (JOIN)

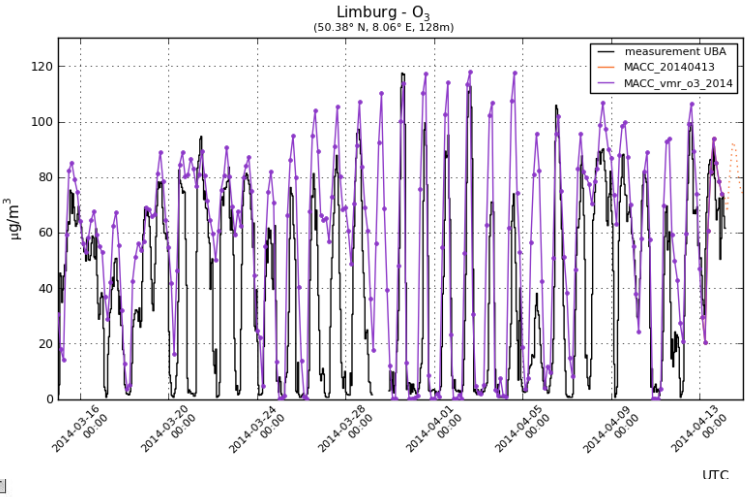
JÜLICH FORSCHUNGSZENTRUM

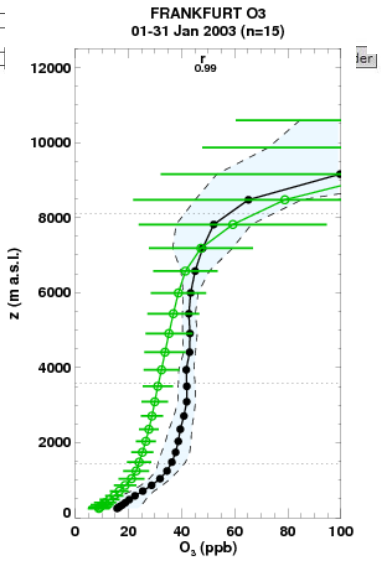
http://join.iek.kfa-juelich.de



Linking model results with observations



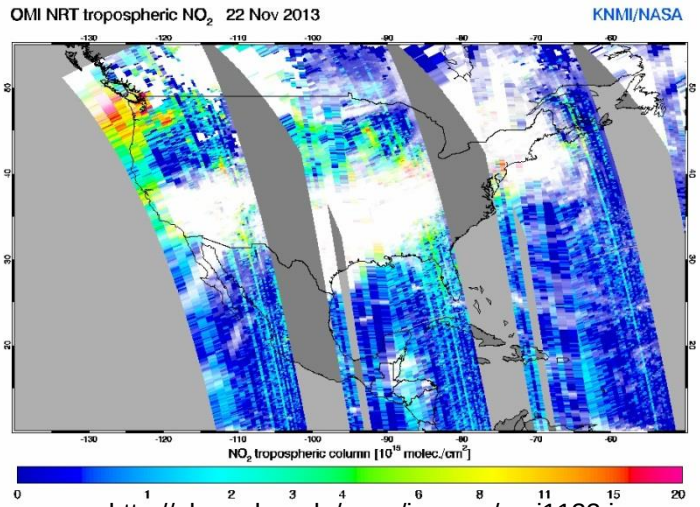




Design goals:

- online visualisation
- customized data downloads
- statistical analysis
- user profile
- data license management

http://www.iagos.fr/macc/reanalysis_profiles.php



http://alg.umbc.edu/usaq/images/omi1122.jpg

JÜLICH
FORSCHUNGSZENTRUM

The FAUL concept of interoperability

FIND
ACCESS
UNDERSTAND
LINK

without having to program anything as a user

FAUL (as a German word) means „lazy“

Content-driven	Technology-driven
Proprietary met. service infrastructure	Open data access structure/ technology-independent
Working solutions for data formats and large data volumes	**B2SHARE**: capacity? Formats?
Large diversity of formats and metadata	??
Data discovery under construction (WMO/WIGOS)	**B2FIND**: metadata standards? OGC compliance?
Near realtime data	??

Where can these communities benefit from each other?

Jülich MACC-EUDAT project

- Tool development for automated metadata generation
- Definition of „community metadata profile“
- Metadata publication (B2FIND, CSW)

OGC

ISO 19115

netCDF-CF

WCS

CSW

Seadatanet

EuroGEOSS Broker

JOIN

The time is ripe
- to define a community metadata profile for atmospheric composition and air quality
- to populate controlled vocabulary databases in order to empower search engines and web catalogues

Learn more…

Copernicus Big Data Workshop

Event detail

Copernicus Big Data Workshop
13 March 2014 - 14 March 2014, Brussels, Belgium

This workshop will focus on data availability for Copernicus users (public authorities, downstream industry, researchers, etc.), as well as on technical solutions to ensure the uptake of Copernicus data and information. Users and other stakeholders will be able to share their views on data and information formats, dissemination platform technology for Copernicus.

The workshop will be structured as follows:

- Copernicus and its data and information-policy objectives
- Copernicus data and information formats and processing levels
- Dissemination models Part 1 – Bringing the data to the users
- Dissemination models Part 2 – Bringing the users to the data
- Related dissemination infrastructures and research projects from FP7, future Horizon 2020 calls, as well as from national or international research programmes (ESA, etc.)

Presentations available at:
http://www.copernicus.eu/pages-principales/library/presentations/copernicus-big-data-workshop/