Data Cubes for Water Quality

Tiago A. M. Silva , Veronique Creach, Cherry Jones,

Kate Collingridge, Tiago Santos, Mark Baldry

DCS4COP

http://dcs4cop.eu

European Commission

DataBee

Brockmann Consult · Cefas · NIVA Norwegian Institute for Water Research · museum Operational Directorate Natural Environment OD Nature I OD Natuur I DO Nature · Starlab Living Science · University of Hull · LOV · Sorbonne Université · cnrs · vito

xcube

Centre for Environment Fisheries & Aquaculture Science

Cefas

Data Cubes for Water Quality

1. What is a data cube
2. Examples of datacubes at Cefas

DCS4COP

http://dcs4cop.eu

European Commission

EDataBee

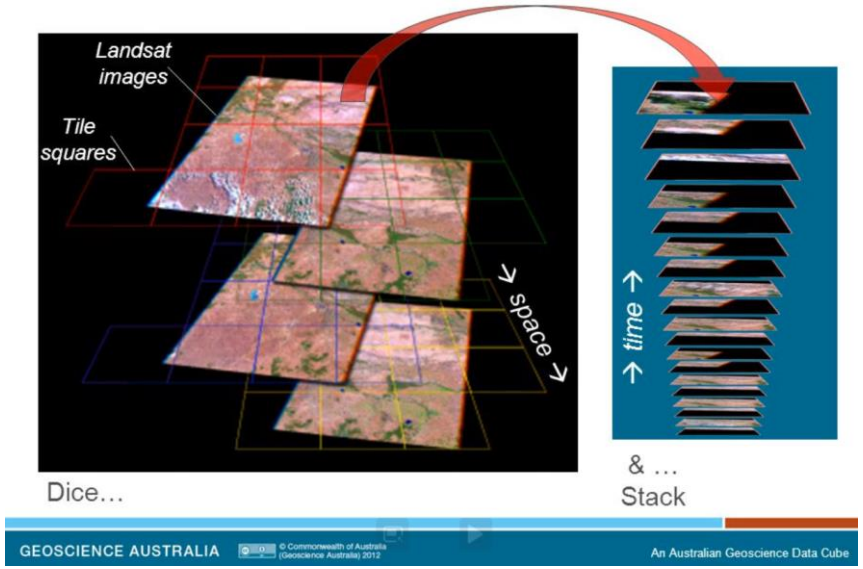
Brockmann Consult · Cefas · NIVA Norwegian Institute for Water Research · museum Operational Directorate Natural Environment OD Nature I OD Natuur I DO Nature · Starlab Living Science · University of Hull · LOV Sorbonne Université cnrs · vito

xcube

What is a Data Cube?

- A data repository on a shared grid
- Each variable is N-dimensional
- Operators can be applied in any dimension just a as easily
  - `Average('time')`, `Average('latitude')`, etc

Dim.	Array of point sensors	Satellite imagery	Ocean Model
1	Time	Time	Time
2	Location	Latitude	x
3		Longitude	y
4			Depth



Principles of the Data Cube Manifesto

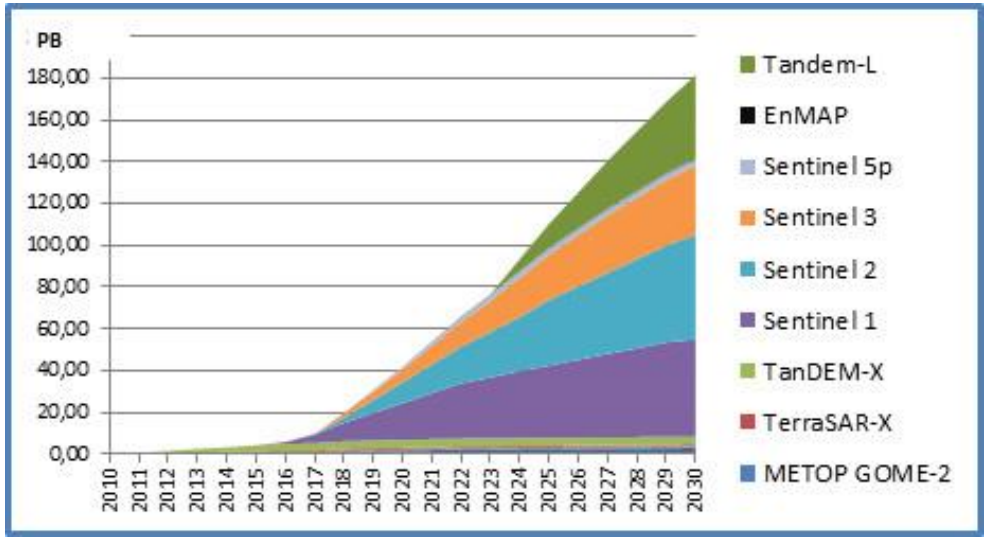
1. Gridded data with N dimensions
2. **All axes have equal treatment**
3. Efficient trimming and slicing in one request
4. **Similar performance across axes**
5. …partitioning (chunks)
6. Language that allows composite extraction, processing, filtering and fusion

https://earthserver.eu/tech/datacube-manifesto/

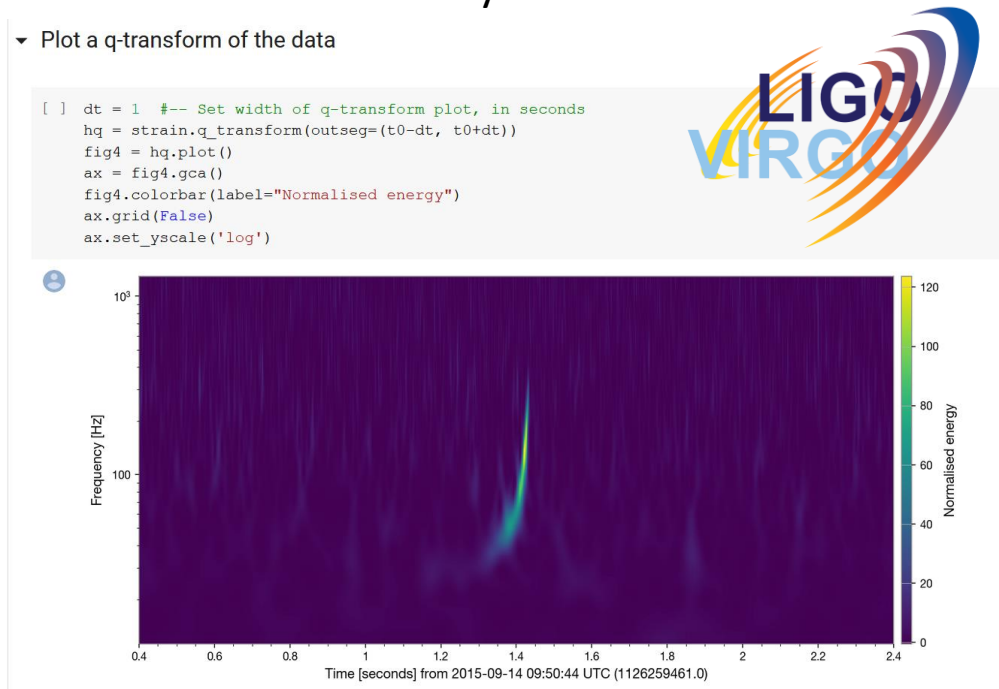
Motivation

- Keep up with data size growth.
  - Progressively harder for users to download and process locally
- Make the most of the data
  - Make it accessible *and* usable by providing the tools
- Reproducibility
  - fragmented workflow makes science impractical to reproduce
  - multiple software tools and environments

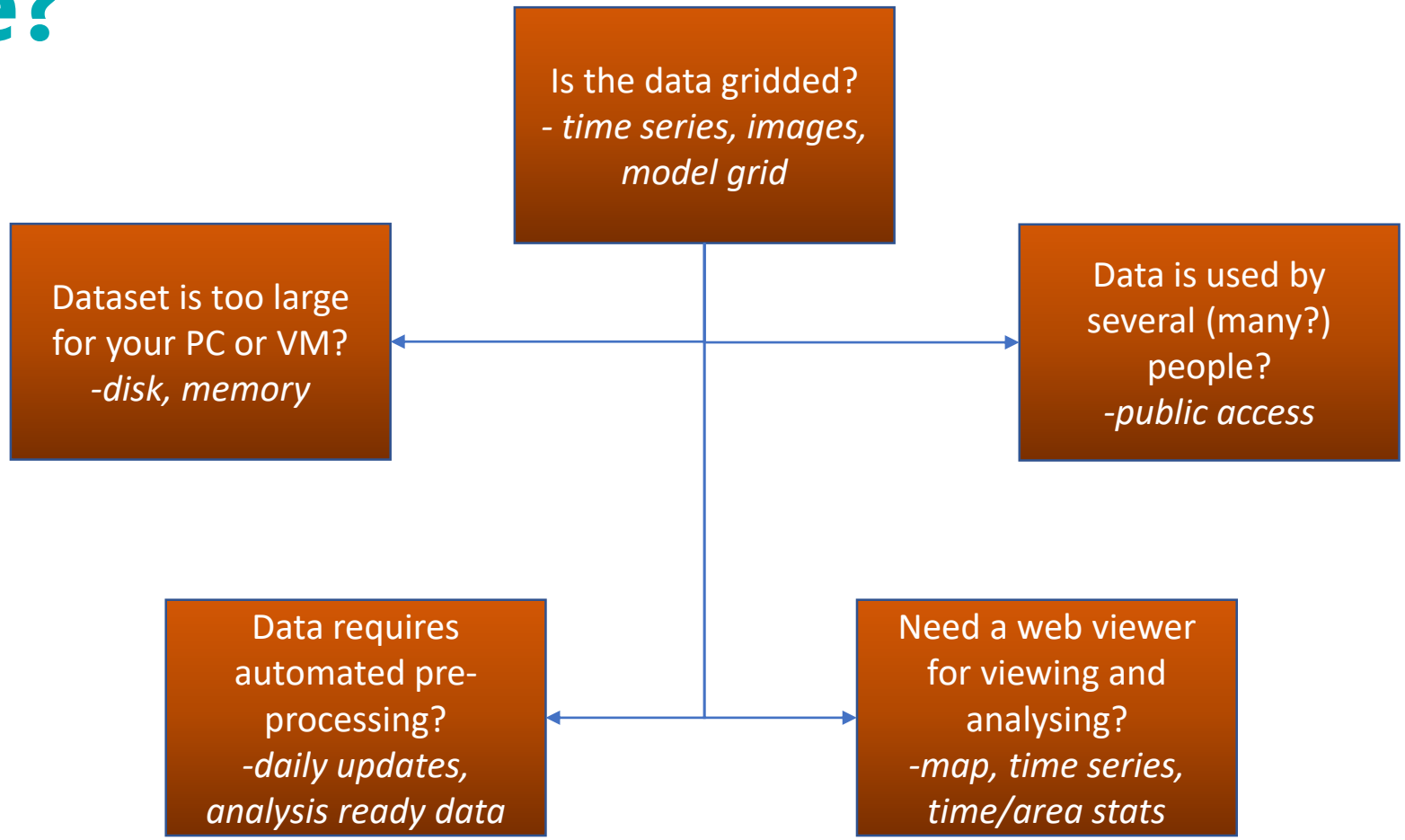
http://dlr.de



2017 Nobel Prize for Physics

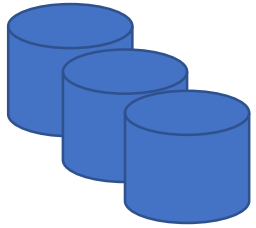


Could you benefit from using a Data Cube?



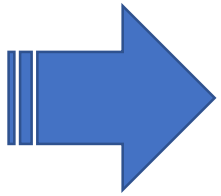
Data Cube Service

Image Files
netCDF, GeoTIFF

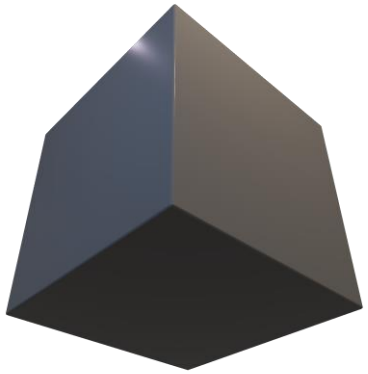


Pre-processing
to produce
Analysis Ready
Data

Generate
datacube



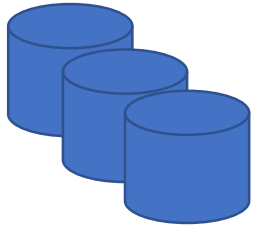
Common grid
Metadata
Chunking



File formats optimised
for cloud
Zarr, TileDB, HDF cloud
One Object has many
files

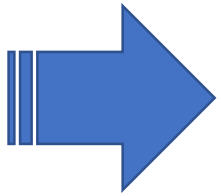
Data Cube Service

Image Files
netCDF, GeoTIFF

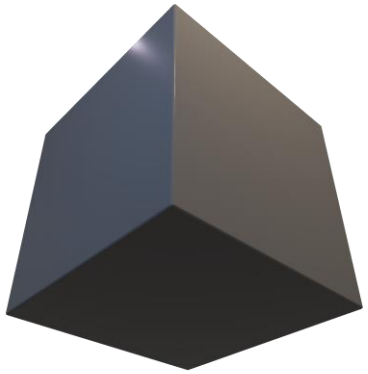


Pre-processing
to produce
Analysis Ready
Data

Generate
datacube

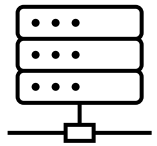


Common grid
Metadata
Chunking



File formats optimised
for cloud
Zarr, TileDB, HDF cloud
One Object has many
files

Server



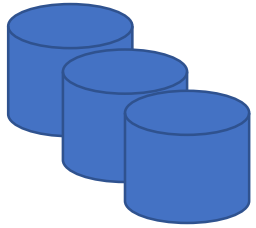
API

Data analysis

One dataset, several users

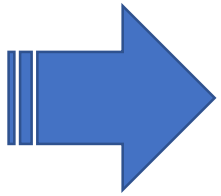
Data Cube Service

Image Files
netCDF, GeoTIFF

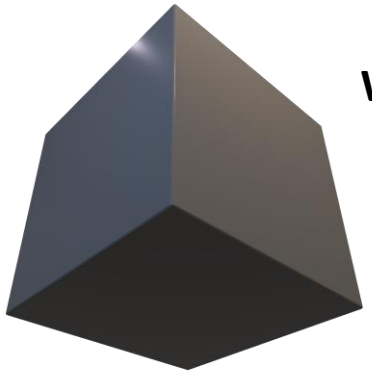


Pre-processing to produce Analysis Ready Data

Generate datacube



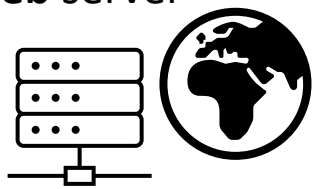
Common grid
Metadata
Chunking



File formats optimised for cloud
Zarr, TileDB, HDF cloud
One Object has many files

Web Viewer

Web server

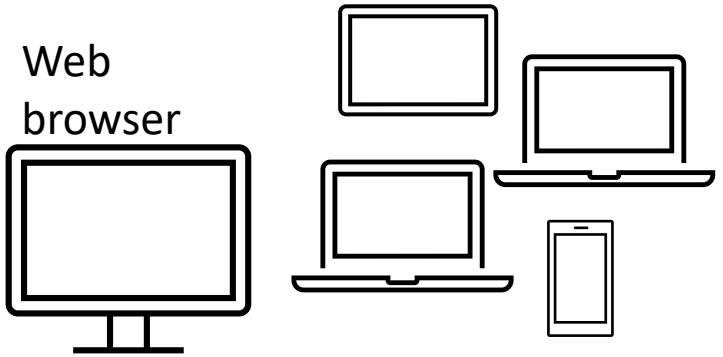


API

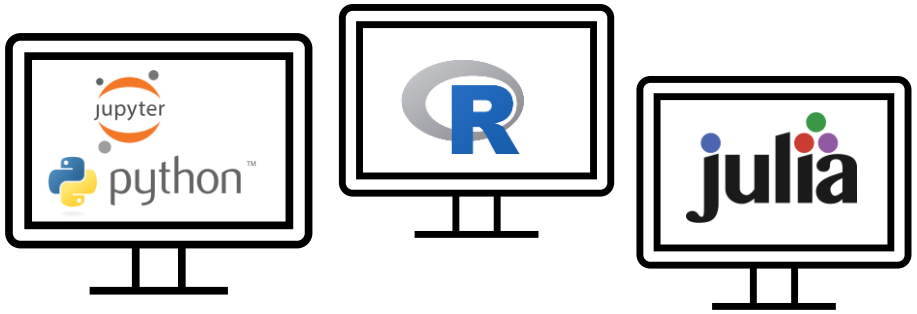
Web API

Data querying

Web browser



Data analysis



One dataset, **many** users for different use cases

earthserver.eu

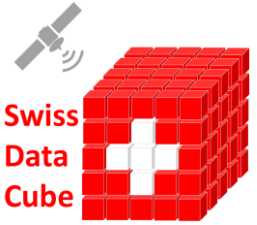
rasdaman
raster data manager

Baumann et al., 1997, 2016

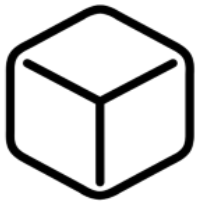
www.opendatacube.org

github.com/opendatacube

Digital Earth
AUSTRALIA



xcube



github.com/dcs4cop/xcube

EURO DATA CUBE

EARTHCUBE

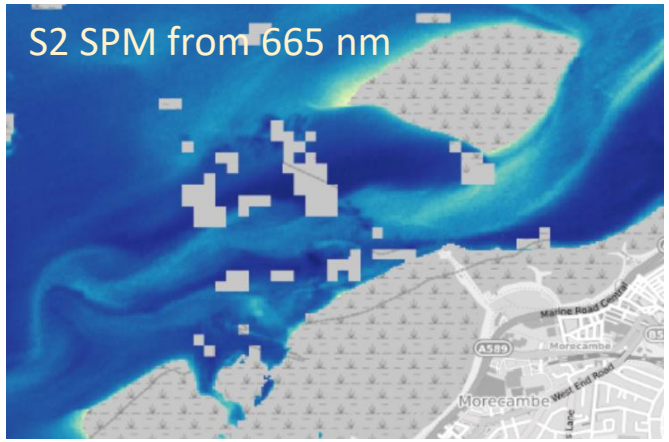
www.pangeo.io
github.com/pangeo-data
Signell and Pothina, 2019

Data Cube Usages in the Marine Domain

- Intertidal cycle, coastline change
- Deal with clouds, missing data: cloud-free mosaics, robust statistics
- Area statistics: environmental assessments
- Temporal classification: water types, river plumes
- Explore multiple algorithms on-the-fly:

  Multiband indexes, fitted inversions
- Combine multiple satellite sensors

Intertidal DEM
Digital Earth Australia



Pixels as set of independent observations

- Vary in time, space, and uncertainty
- Combine them to increase accuracy

Water Quality Data Cubes at Cefas

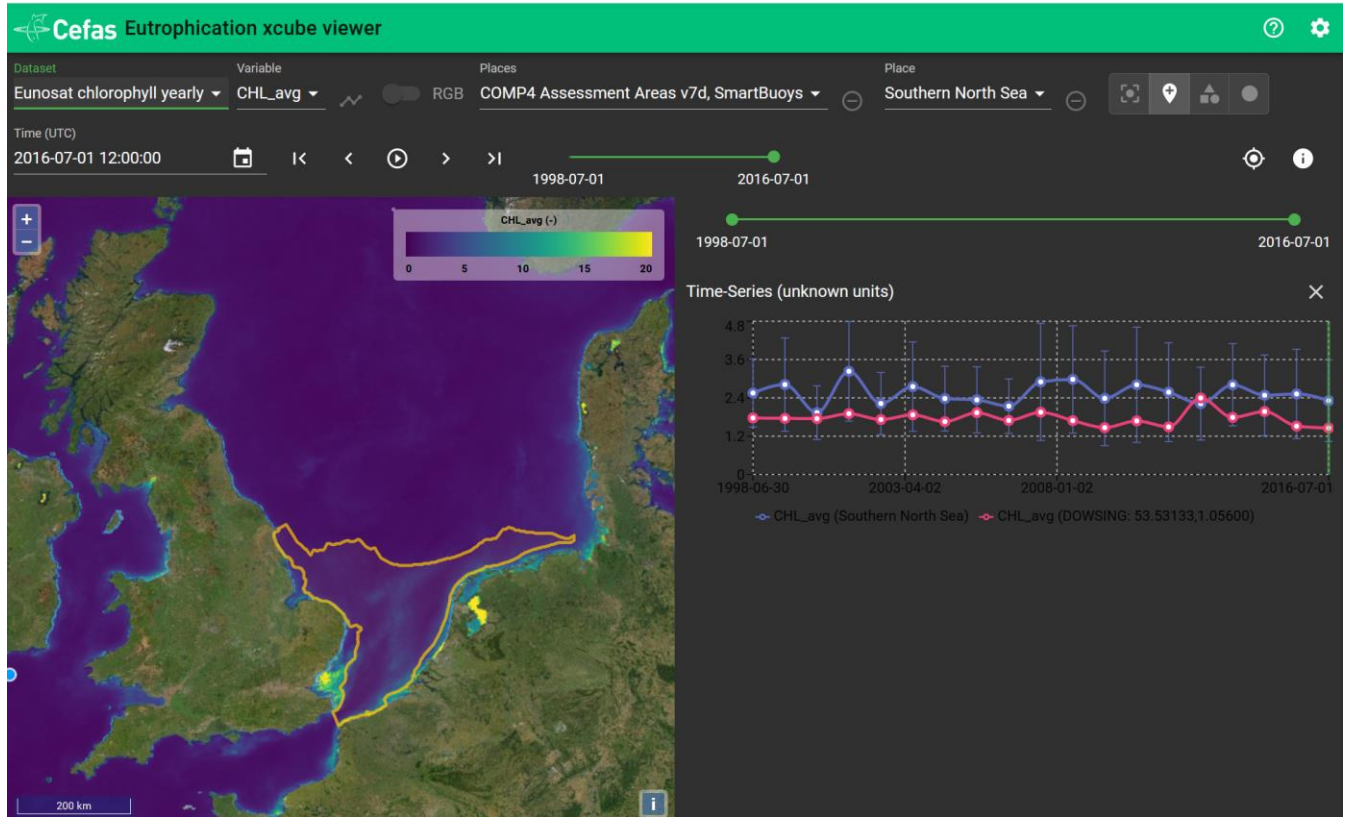
I. MSFD Eutrophication Assessment

II. WFD Eutrophication Risk

III. Marine Discharge Plumes

I . MSFD Eutrophication Assessment

- Chlorophyll observations
  - EUNOSAT coherent algorithm
  - CMEMS standard
- COMP4 Assessment Areas
- SmartBuoy locations



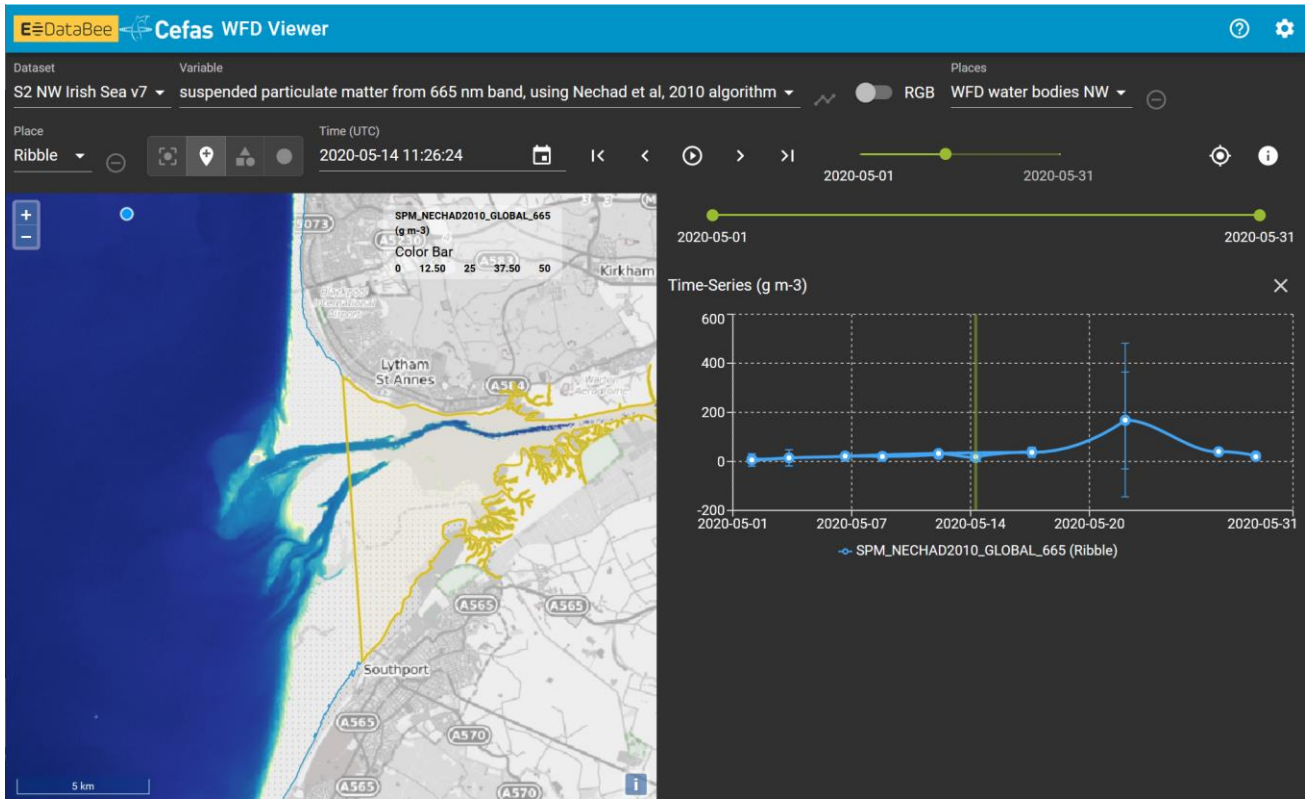
xcube

Centre for Environment Fisheries & Aquaculture Science

Cefas

II . WFD Eutrophication Risk

- Sentinel-2 (10 m) Suspended Particulate Matter/Turbidity
  - Multiple bands and algorithms
- WFD Water bodies
- EA Water Quality Data Archive
  - *Aim is to integrate with* in-situ



Environment Agency

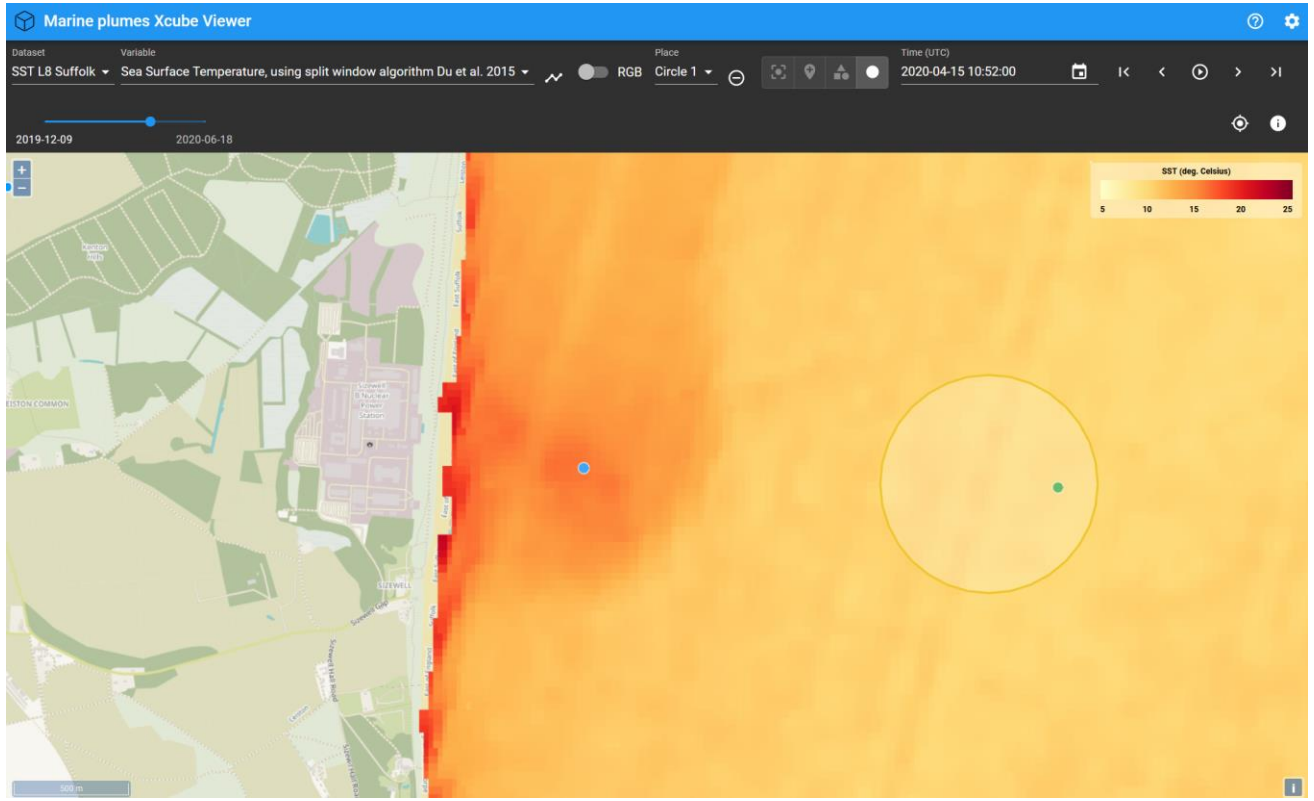
xcube

Centre for Environment Fisheries & Aquaculture Science

Cefas

III . Marine Discharge Plumes

- High Resolution Sea Surface Temperature (100 m)
  - *Atmospheric correction*
- Power station Intakes and outfalls
- Protected Areas
- *Automatic detection of plume*
- *In situ* SST



xcube

Thank you for your attention

tiago.silva@cefas.co.uk

Together we are working for **a sustainable blue future**

Centre for Environment Fisheries & Aquaculture Science

Cefas