Brockmann Consult
Environmental Informatics • Geoinformation Services

xcube - Python package for Earth Observation data cubes

Pangeo Showcases Fall 2021

Gunnar Brandt, Norman Fomferra
Brockmann Consult GmbH

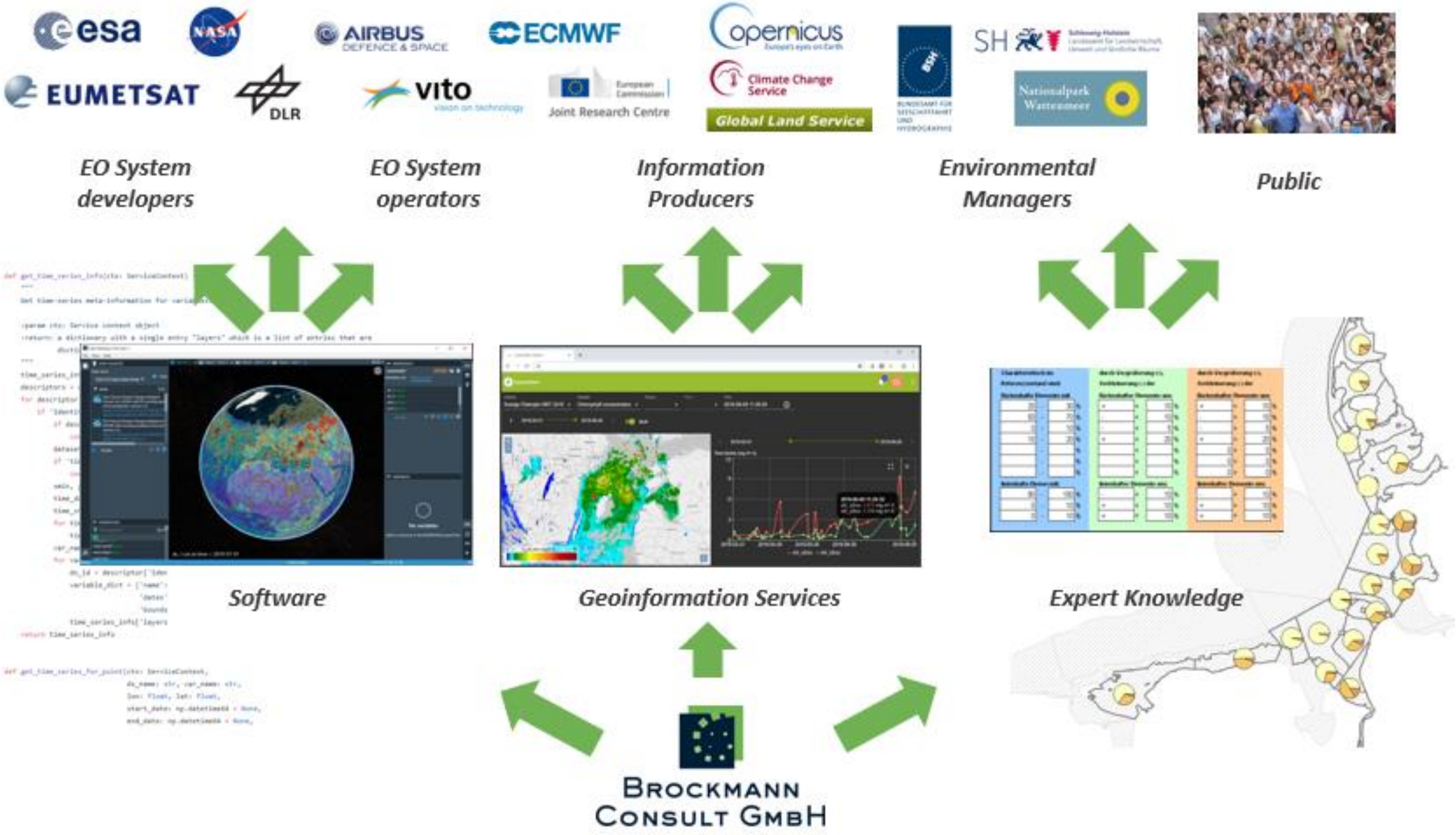
03rd Nov 2021

www.brockmann-consult.de

About Us

Brockmann Consult GmbH

- Small enterprise, located in Hamburg, Germany
- Develops open source software for exploitation of environmental data
- Provides consultancy and geoinformation services to public and private customers
- Diverse customer base including large Space Agencies and European institutions

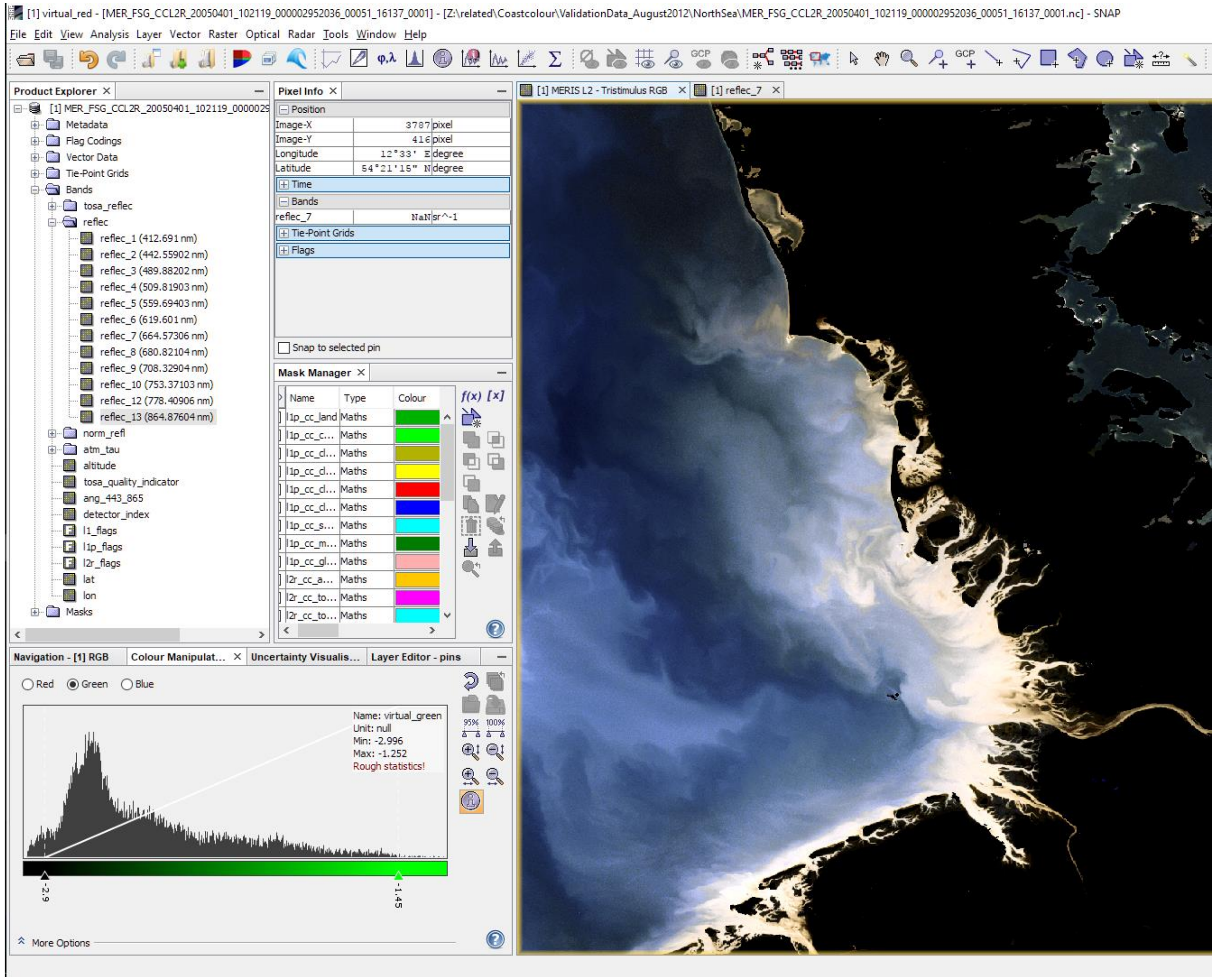


About Us

SNAP

- Lead developer of Sentinel Application Platform (SNAP) for ESA since 2014
- Standard software for processing and analysis of Earth Observation data with 190k downloads for current v8.0.
- Visualisation, analysis, graph processing framework, reprojecting, ortho-rectification
- Large and active user community, 8100 registered forum users

https://step.esa.int/main/toolboxes/snap



[1] virtual_red - [MER_FSG_CCL2R_20050401_102119_000002952036_00051_16137_0001] - [Z:\related\Coastcolour\ValidationData_August2012\NorthSea\MER_FSG_CCL2R_20050401_102119_000002952036_00051_16137_0001.nc] - SNAP

File Edit View Analysis Layer Vector Raster Optical Radar Tools Window Help
Product Explorer
Pixel Info
[1] MERIS L2 - Tristimulus RGB
[1] reflec_7

[1] MER_FSG_CCL2R_20050401_102119_0000029
- Metadata
- Flag Codings
- Vector Data
- Tie-Point Grids
- Bands
  - tosa_reflec
  - reflec
    - reflec_1 (412.691 nm)
    - reflec_2 (442.55902 nm)
    - reflec_3 (489.88202 nm)
    - reflec_4 (509.81903 nm)
    - reflec_5 (559.69403 nm)
    - reflec_6 (619.601 nm)
    - reflec_7 (664.57306 nm)
    - reflec_8 (680.82104 nm)
    - reflec_9 (708.32904 nm)
    - reflec_10 (753.37103 nm)
    - reflec_12 (778.40906 nm)
    - reflec_13 (864.87604 nm)
  - norm_refl
  - atm_tau
  - altitude
  - tosa_quality_indicator
  - ang_443_865
  - detector_index
  - l1_flags
  - l1p_flags
  - l2r_flags
  - lat
  - lon
- Masks

Position

Image-X	3787	pixel
Image-Y	416	pixel
Longitude	12°33' E	degree
Latitude	54°21'15" N	degree

Time

Bands

reflec_7	NaN	sr^-1

Tie-Point Grids

Flags

□ Snap to selected pin

Mask Manager

Name	Type	Colour
l1p_cc_land	Maths	🟩
l1p_cc_c...	Maths	🟩
l1p_cc_cl...	Maths	🟨
l1p_cc_cl...	Maths	🟨
l1p_cc_cl...	Maths	🟥
l1p_cc_cl...	Maths	🟦
l1p_cc_s...	Maths	🟦
l1p_cc_m...	Maths	🟩
l1p_cc_gl...	Maths	🟥
l2r_cc_a...	Maths	🟧
l2r_cc_to...	Maths	🟪
l2r_cc_to...	Maths	🟦

○ Red ● Green ○ Blue

Name: virtual_green
Unit: null
Min: -2.996
Max: -1.252
Rough statistics!

About Us

SNAP

- Lead developer of Sentinel Application Platform (SNAP) for ESA since 2014
- Standard software for processing and analysis of Earth Observation data with 190k downloads for current v8.0.
- Visualisation, analysis, graph processing framework, reprojecting, ortho-rectification
- Large and active user community, 8100 registered forum users

https://step.esa.int/main/toolboxes/snap

Calvalus

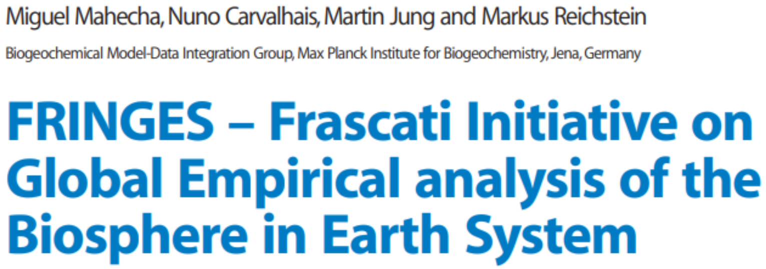
- Batch processing system for Earth Observation data, based on Apache Hadoop
- Full archive processing
- On-premise or cloud deployments also for clients
- Pluggable processors: SNAP graphs, SNAP operators, Python programs, Unix executable, docker containers
- Supports as outputs NetCDF, Zarr, GeoTIFF, Dimap, and others

xcube History

Continuous history of activities with similar requirements:

Build a multi-variate, analysis-ready data cube from several, heterogeneous data sources and enable users to exploit it.

- **FRINGES (2011),** Frascati Initiative on Global Empirical analysis of the Biosphere in Earth System
- **CAB-LAB (2015-2017):** Coupled Atmosphere Biosphere Laboratory, funded by ESA, first data cube application project by Brockmann Consult
- **ESDL (2018-2021):** Earth System Data Lab, ESA project, successor of CAB-LAB
- **DCS4COP (2017-2021):** Data Cube for Copernicus Services, European Commission project; **first version of xcube** based on lesson's learnt in previous projects, providing Copernicus Service data and added-value products to customers.

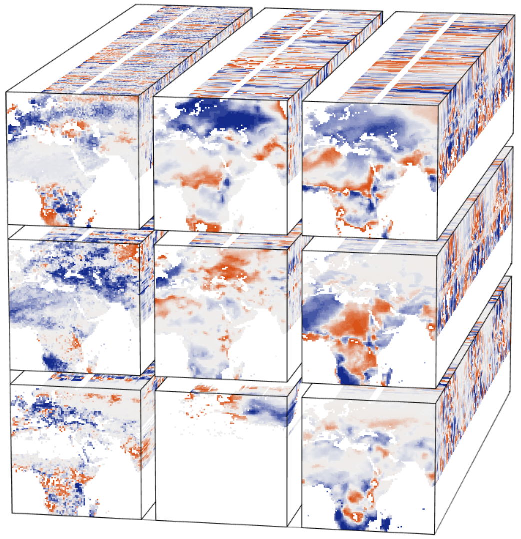


Miguel Mahecha, Nuno Carvalhais, Martin Jung and Markus Reichstein

Biogeochemical Model-Data Integration Group, Max Planck Institute for Biogeochemistry, Jena, Germany

FRINGES – Frascati Initiative on Global Empirical analysis of the Biosphere in Earth System

http://www.ileaps.org/sites/default/files/iLEAPS_Newsletter_11_September_2011.pdf



M. Mahecha (2017)
https://figshare.com/articles/figure/Earth_Data_Cube/4822930/2

Common Project Objectives

- Convenient access to as many gridded data sources as possible
- Generate analysis-ready data (ARD) cubes
- High-level tools and functions to exploit and manage data cubes
- Serve data, directly from (Zarr) data cubes
- Visualise and analyse served data
- Provide Toolbox-as-a-Service

→ Develop integrated cube ecosystem based on popular Python stack (xarray, zarr, dask)

xcube Components and Features (1/2)

- xcube Data Store Framework (`xcube.core.store`)
  - Data cube access Python API for existing data web APIs and storage types
  - Create lazy data cube views
  - Data stores can serve cubes, multi-resolution cubes, and vector data
  - Plugin-based
- xcube Generator (`xcube.core.gen` and `.gen2`)
  - Create data cubes from NetCDF time-slices (gen) or any data stores (gen2)
  - Write Zarr cubes to local FS (gen) or S3 (gen2)
  - Write multi-resolution Zarr cubes (gen2)
  - AWS web service (gen2)
  - Run user code: inline, from source Zip, or GitHub release

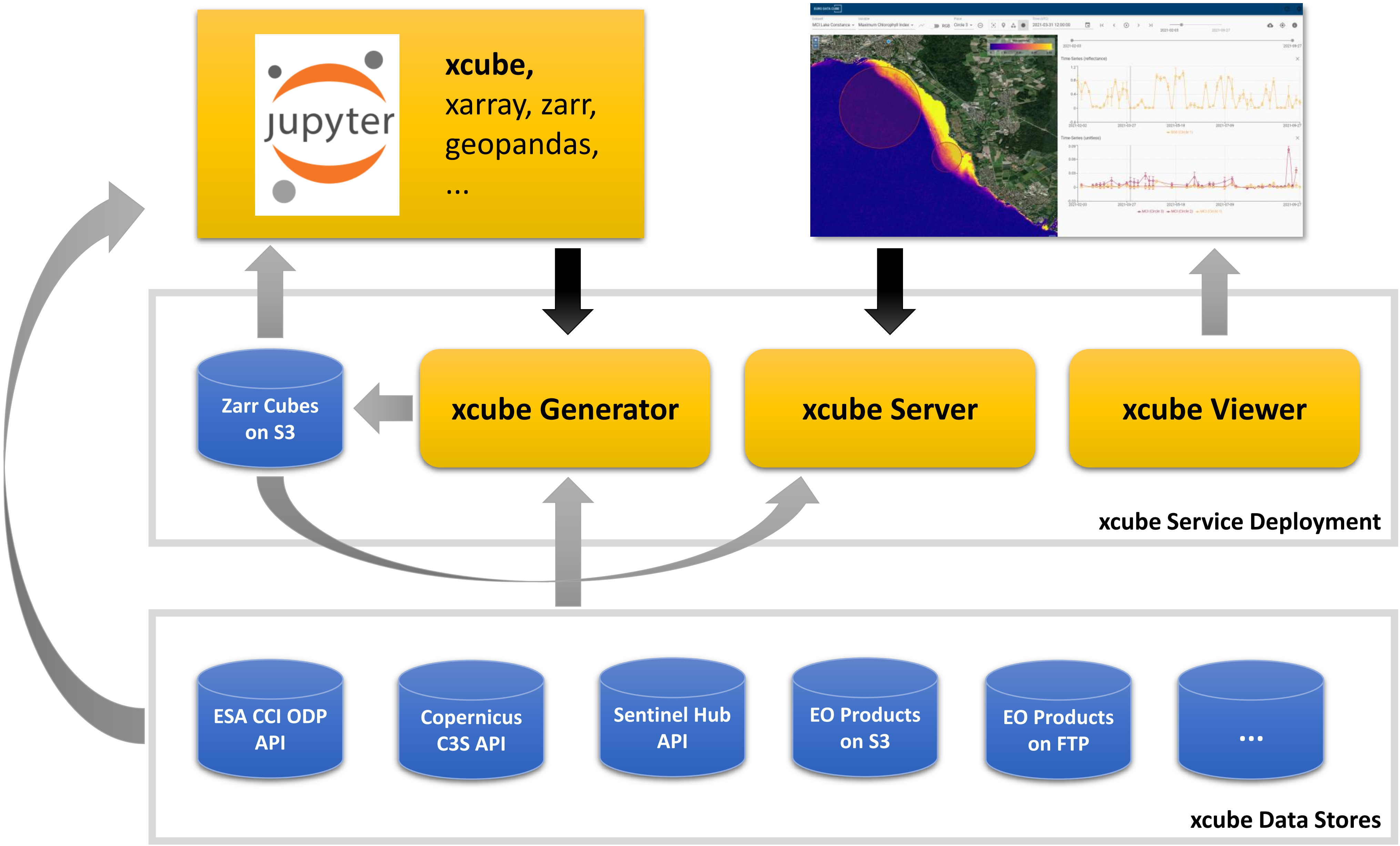
xcube Components and Features (2/2)

- xcube Server (`xcube.webapi`)
  - Serve data from local FS, Zarrs in S3, or from any data stores
  - Tile, catalogue and time-series extraction REST API (OGC STAC in prep.)
  - Web Tile Service: OGC WMTS 1.0
  - Provide raw data access via S3-compatible REST API
- xcube Viewer
  - Uses xcube Server
  - Display cubes and vector data on map
  - Animate map display through time
  - Time-series diagrams
  - Cube and vector data metadata display

- xcube (`xcube.core.*` and `xcube.cli`)
  - High-level cube computation, analysis and management tools
  - Python API and CLI

- Cate – A cloud-enabled Toolbox for the ESA Climate Change Initiative (CCI)
  - Scientific exploitation of ESA CCI Essential Climate Variables (ECV) datasets
  - Uses xcube data stores
  - xcube plugin that provides data stores for the ESA CCI Open Data Portal (OPeNDAP) and Object Storage (S3/Zarr)

Concept



jupyter

xcube,
xarray, zarr,
geopandas,
…

Zarr Cubes on S3

xcube Generator

xcube Server

xcube Viewer

xcube Service Deployment

ESA CCI ODP API

Copernicus C3S API

Sentinel Hub API

EO Products on S3

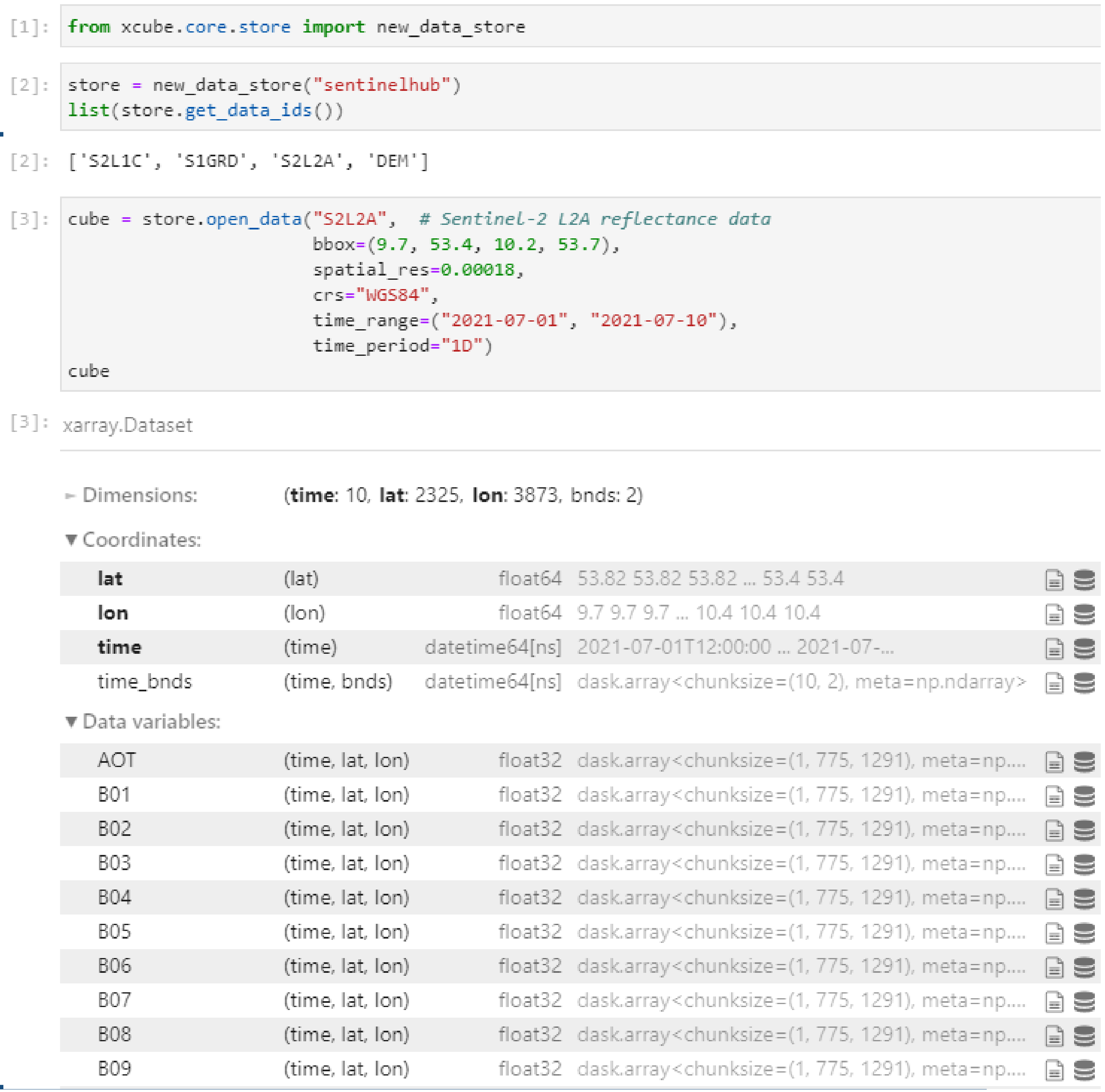
EO Products on FTP

…

xcube Data Stores

xcube Data Stores

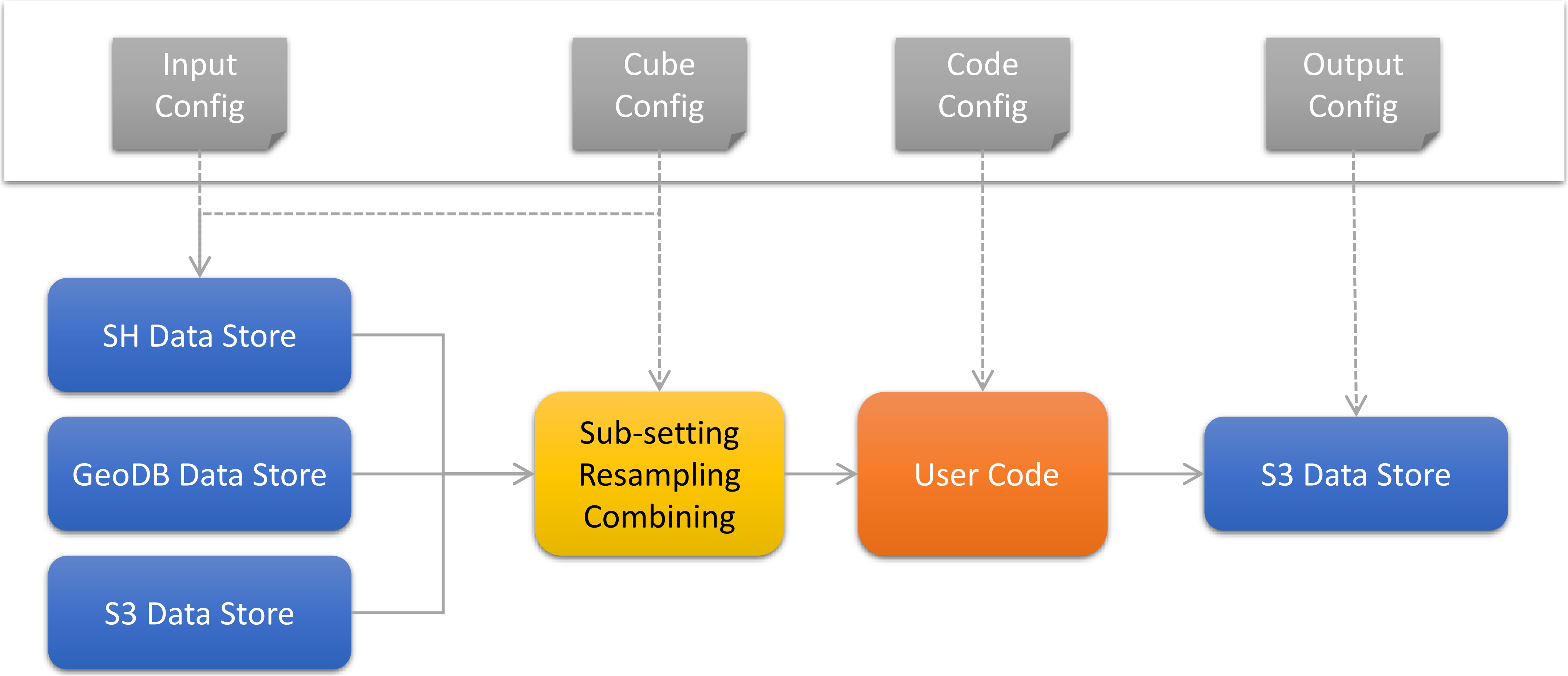
- ESA CCI Open Data Portal
  - Store ID: "cciodp"
- Copernicus Climate Data Store
  - Store ID: "cds"
- Sentinel Hub
  - Store ID: "sentinelhub"
- Generic (based on `fsspec`):
  - Store ID: "s3", "file", "memory", ...



```
[1]: from xcube.core.store import new_data_store

[2]: store = new_data_store("sentinelhub")
     list(store.get_data_ids())

[2]: ['S2L1C', 'S1GRD', 'S2L2A', 'DEM']

[3]: cube = store.open_data("S2L2A",  # Sentinel-2 L2A reflectance data
                            bbox=(9.7, 53.4, 10.2, 53.7),
                            spatial_res=0.00018,
                            crs="WGS84",
                            time_range=("2021-07-01", "2021-07-10"),
                            time_period="1D")
     cube
```

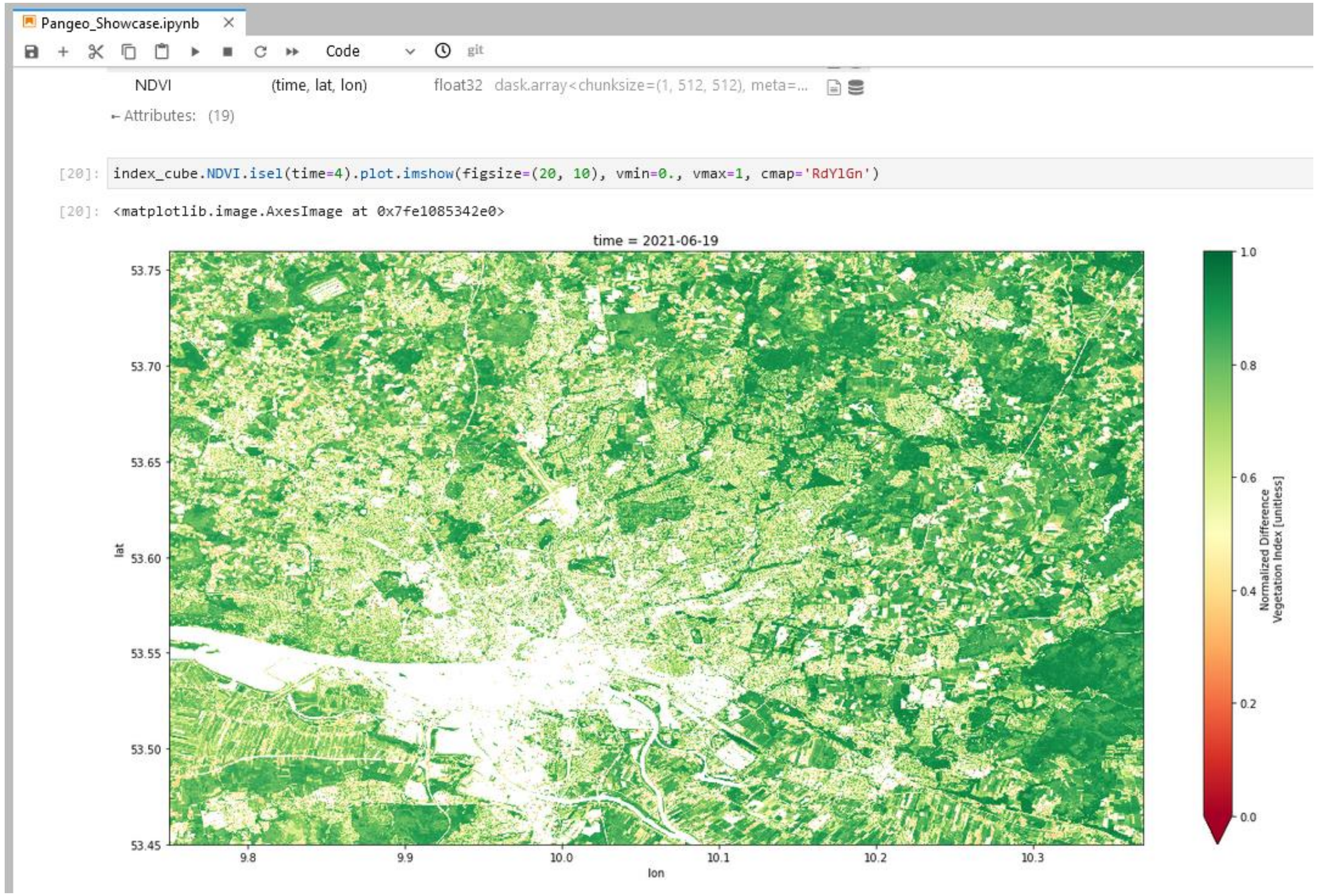
```
[3]: xarray.Dataset

► Dimensions:      (time: 10, lat: 2325, lon: 3873, bnds: 2)
▼ Coordinates:
    lat            (lat)           float64         53.82 53.82 53.82 ... 53.4 53.4
    lon            (lon)           float64         9.7 9.7 9.7 ... 10.4 10.4 10.4
    time           (time)          datetime64[ns]  2021-07-01T12:00:00 ... 2021-07-...
    time_bnds      (time, bnds)    datetime64[ns]  dask.array<chunksize=(10, 2), meta=np.ndarray>
▼ Data variables:
    AOT            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B01            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B02            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B03            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B04            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B05            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B06            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B07            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B08            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
    B09            (time, lat, lon)  float32  dask.array<chunksize=(1, 775, 1291), meta=np....
```

xcube Generator



Example I: xcube Generator

- Jupyter Notebook demonstrating data cube genaration from Earth Observation data, here Sentinel-2 MSI
- Sentinel Hub data store with access to full archive of several sensors
- Developing and testing user code
- Integrating user code into generation request
- Local generation and xcube Generator service



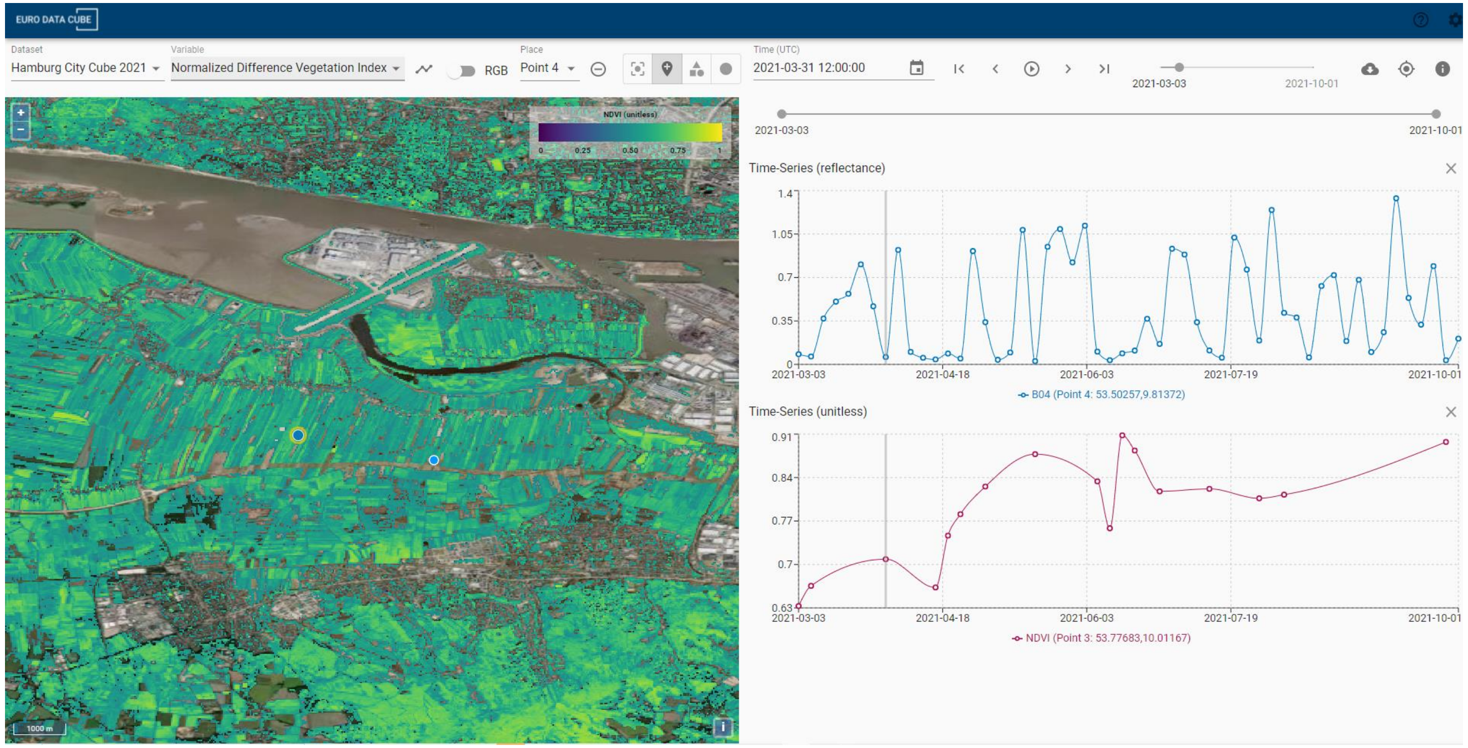
```
NDVI        (time, lat, lon)        float32  dask.array<chunksize=(1, 512, 512), meta=...
▸ Attributes: (19)
```

```
[20]: index_cube.NDVI.isel(time=4).plot.imshow(figsize=(20, 10), vmin=0., vmax=1, cmap='RdYlGn')
```

```
[20]: <matplotlib.image.AxesImage at 0x7fe1085342e0>
```

Example II: xcube Server + Viewer

- Simple, web-based viewer for all data cubes generated by xcube
- Works directly on zarr repos, no pre-processing of image tiles
- Based on react
- xcube server required for providing image tiles processed on-the-fly
- Image and time-series functionality
- On-the-fly execution of user code possible, e.g. temporal aggregation



Dataset: Hamburg City Cube 2021

Variable: Normalized Difference Vegetation Index

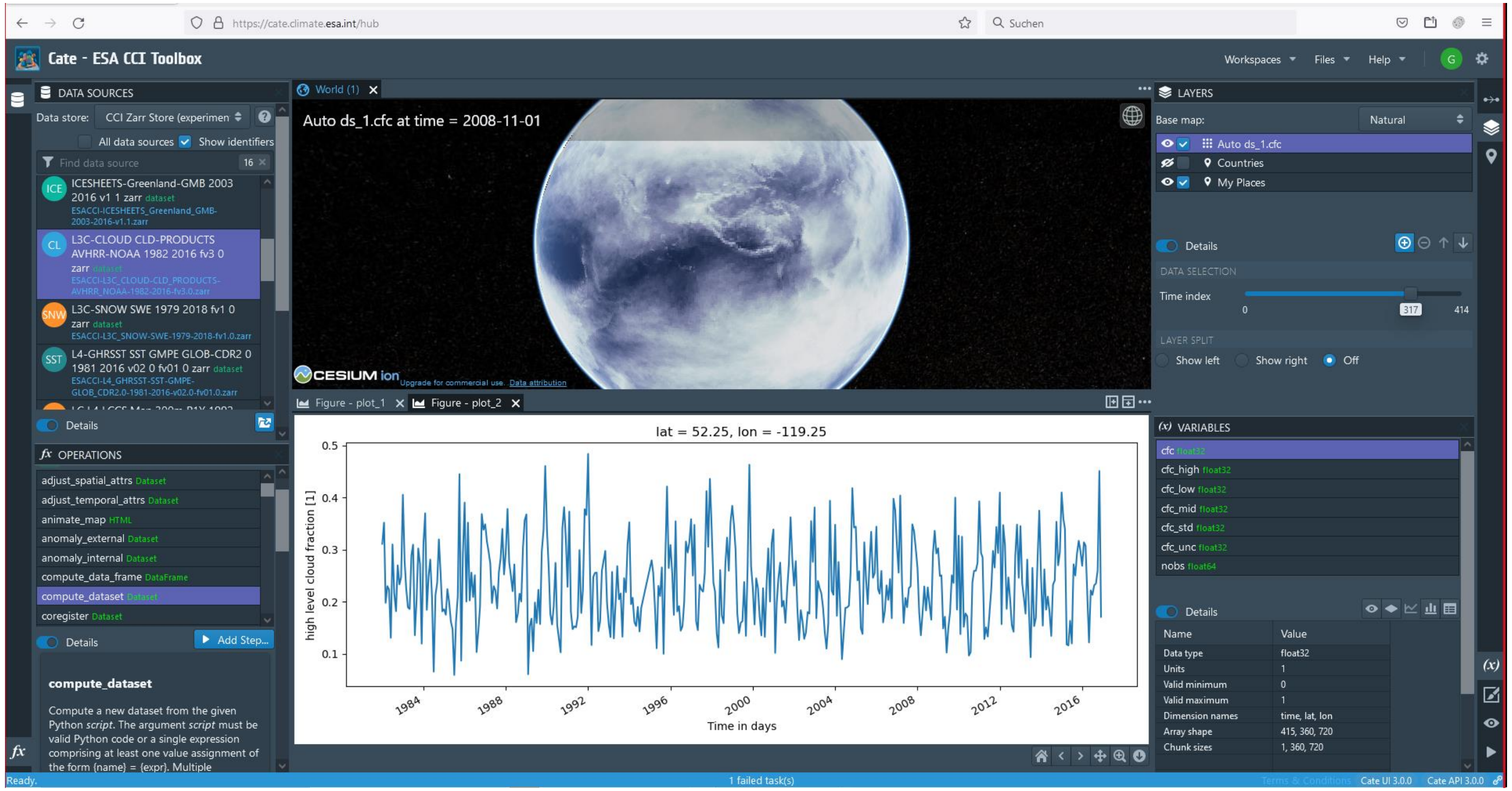
Place: Point 4

Time (UTC): 2021-03-31 12:00:00

https://edc-viewer.brockmann-consult.de

Example III: Cate – Toolbox as a Service

- Developed to analyse Essential Climate Variables provided by ESA's Climate Change Initiative
- Offers numerous analytical functions by default, additional ones can be simply added as Python functions
- Allows for developing user workflows
- Visualises Images, time-series, histograms


https://cate.climate.esa.int/hub
Suchen
Cate - ESA CCI Toolbox
Workspaces
Files
Help
DATA SOURCES
Data store: CCI Zarr Store (experimen
All data sources
Show identifiers
Find data source
16
ICESHEETS-Greenland-GMB 2003 2016 v1 1 zarr dataset
ESACCI-ICESHEETS_Greenland_GMB-2003-2016-v1.1.zarr
L3C-CLOUD CLD-PRODUCTS AVHRR-NOAA 1982 2016 fv3 0 zarr dataset
ESACCI-L3C_CLOUD-CLD_PRODUCTS-AVHRR_NOAA-1982-2016-fv3.0.zarr
L3C-SNOW SWE 1979 2018 fv1 0 zarr dataset
ESACCI-L3C_SNOW-SWE-1979-2018-fv1.0.zarr
L4-GHRSST SST GMPE GLOB-CDR2 0 1981 2016 v02 0 fv01 0 zarr dataset
ESACCI-L4_GHRSST-SST-GMPE-GLOB_CDR2.0-1981-2016-v02.0-fv01.0.zarr
Details
OPERATIONS
adjust_spatial_attrs Dataset
adjust_temporal_attrs Dataset
animate_map HTML
anomaly_external Dataset
anomaly_internal Dataset
compute_data_frame DataFrame
compute_dataset Dataset
coregister Dataset
Details
Add Step...
World (1)
Auto ds_1.cfc at time = 2008-11-01
CESIUM ion
Figure - plot_1
Figure - plot_2
lat = 52.25, lon = -119.25
high level cloud fraction [1]
0.5
0.4
0.3
0.2
0.1
1984
1988
1992
1996
2000
2004
2008
2012
2016
Time in days
LAYERS
Base map: Natural
Auto ds_1.cfc
Countries
My Places
Details
DATA SELECTION
Time index
0
317
414
LAYER SPLIT
○ Show left
○ Show right
● Off
VARIABLES
cfc float32
cfc_high float32
cfc_low float32
cfc_mid float32
cfc_std float32
cfc_unc float32
nobs float64
Details
Ready.
1 failed task(s)
Terms & Conditions
Cate UI 3.0.0
Cate API 3.0.0

compute_dataset

Compute a new dataset from the given Python *script*. The argument *script* must be valid Python code or a single expression comprising at least one value assignment of the form {name} = {expr}. Multiple

Name	Value
Data type	float32
Units	1
Valid minimum	0
Valid maximum	1
Dimension names	time, lat, lon
Array shape	415, 360, 720
Chunk sizes	1, 360, 720

https://cate.climate.esa.int

xcube in Use – Selected Examples

CyanoAlert

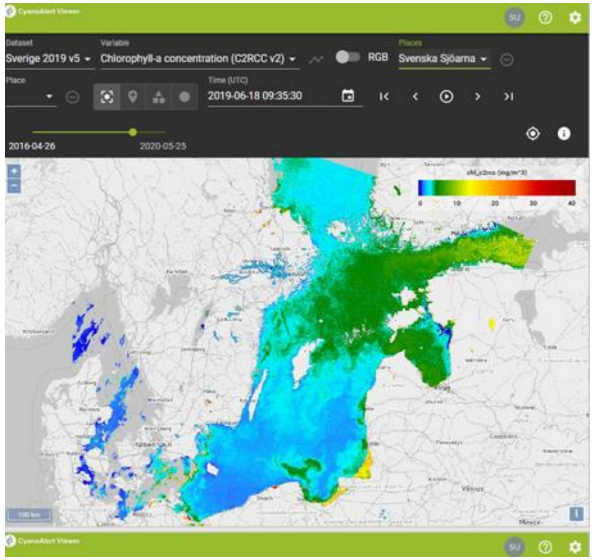
Cyanobacteria information and notification service with several public customers
https://www.cyanoalert.com

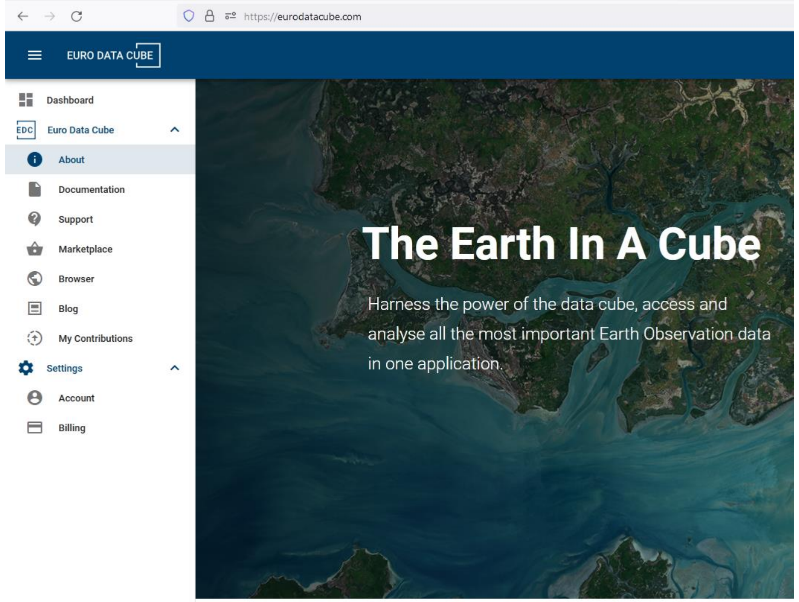
EURO DATA CUBE

Commercial service for access, processing, and selling Earth Observation Data
www.eurodatacube.com

AGRICULTURE VIRTUAL LAB

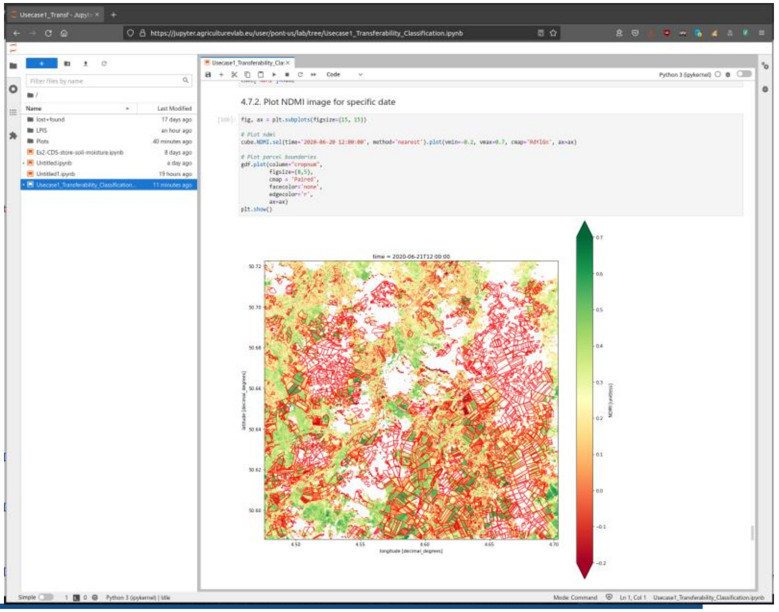
Operating virtual laboratory connecting numerous data sources for agriculture research community
https://agriculturevlab.eu





The Earth In A Cube

Harness the power of the data cube, access and analyse all the most important Earth Observation data in one application.



https://jupyter.agriculturevlab.eu/user/pont-us/lab/tree/Usecase1_Transferability_Classification.ipynb
Usecase1_Transferability_Cla
Code
Python 3 (ipykernel)
Last Modified
17 days ago
an hour ago
8 days ago
a day ago
19 hours ago

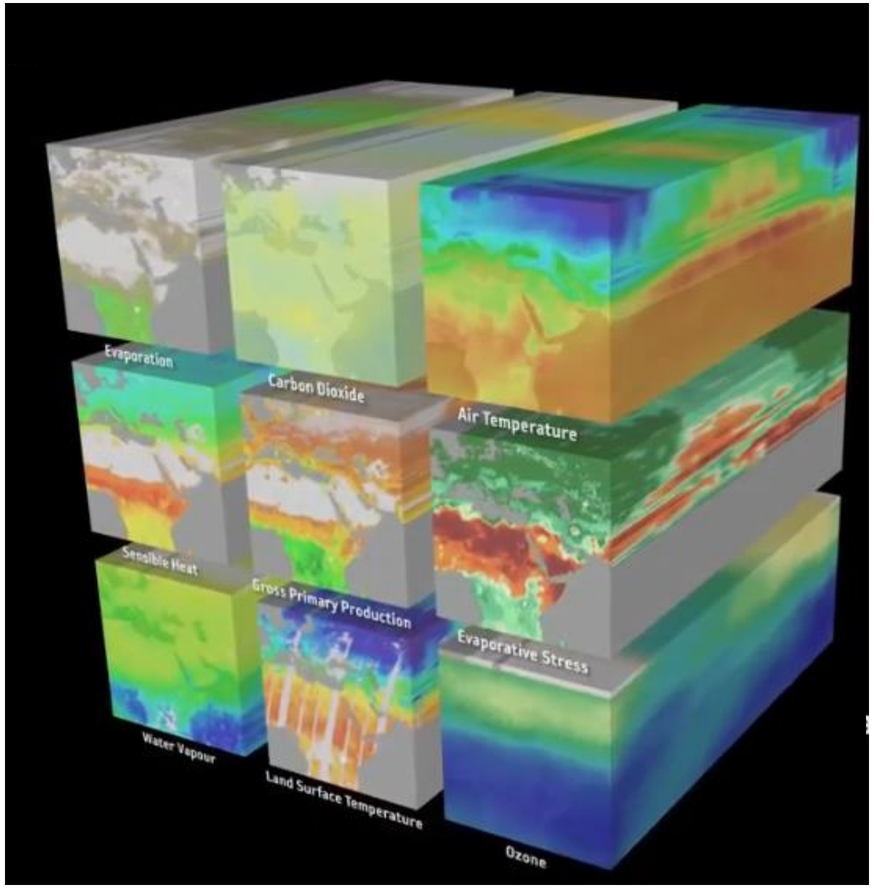
4.7.2. Plot NDMI image for specific date

```
fig, ax = plt.subplots(figsize=(15, 15))

# Plot ndmi
cube.NDMI.sel(time='2020-06-20 12:00:00', method='nearest').plot(vmin=-0.2, vmax=0.7, cmap='RdYlGn', ax=ax)

# Plot parcel boundaries
gdf.plot(column="cropnum",
         figsize=(8,5),
         cmap = 'Paired',
         facecolor='none',
         edgecolor='r',
         ax=ax)
plt.show()
```

time = 2020-06-21T12:00:00
50.72
0.7
0.6

The Road Ahead

- Address user and project requirements for additional features:
  - Support for arbitrary spatial CRSs
  - Large-scale, batch processing
  - Catalogue service
  - Multi-resolution pyramids
  - Improve resampling
  - Enhance OGC support
  - Add ML/AI capabilities
- Evaluate other initiatives and software projects, e.g. Pangeo, Open Data Cube et al. with respect to using existing functionalities and to ensure compatibility
- Improve documentation
- Be more actively engaged in community developing and sharing xcube's tech stack



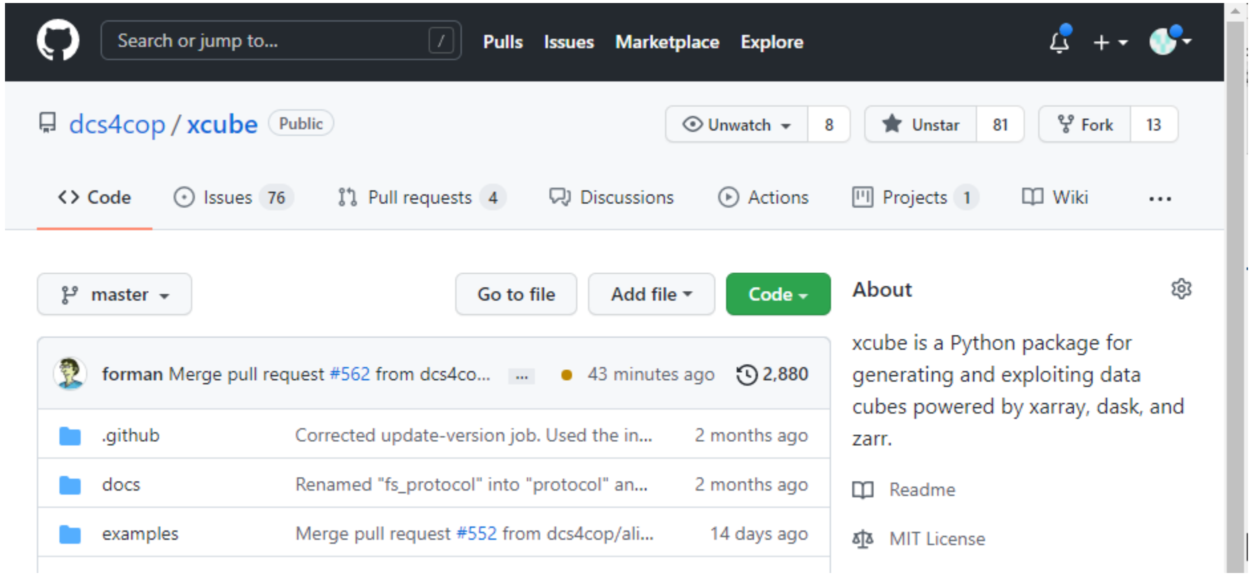
https://www.earthsystemdatalab.net

How to get, use, and contribute to xcube

- Source code on github:

https://github.com/dcs4cop/xcube

- Permissive MIT License, free use
- Submit pull requests to contribute or report issues
- Install with conda, channel *conda-forge*
- Contact us directly via email (see next slide), we are happy to hear from users and potential contributors!



forman Merge pull request #562 from dcs4co... · 43 minutes ago · 2,880

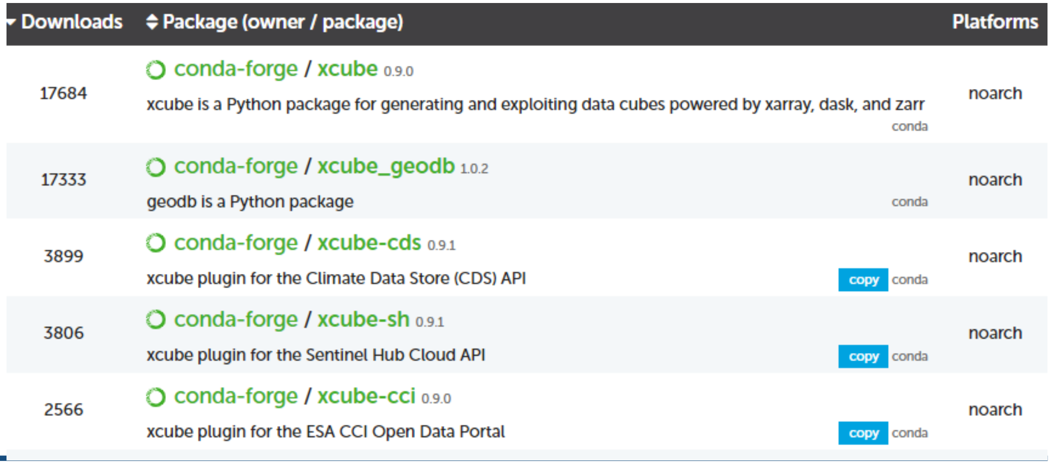
.github	Corrected update-version job. Used the in...	2 months ago
docs	Renamed "fs_protocol" into "protocol" an...	2 months ago
examples	Merge pull request #552 from dcs4cop/ali...	14 days ago

About

xcube is a Python package for generating and exploiting data cubes powered by xarray, dask, and zarr.

Readme

MIT License



Downloads	Package (owner / package)	Platforms
17684	conda-forge / xcube 0.9.0 xcube is a Python package for generating and exploiting data cubes powered by xarray, dask, and zarr conda	noarch
17333	conda-forge / xcube_geodb 1.0.2 geodb is a Python package conda	noarch
3899	conda-forge / xcube-cds 0.9.1 xcube plugin for the Climate Data Store (CDS) API copy conda	noarch
3806	conda-forge / xcube-sh 0.9.1 xcube plugin for the Sentinel Hub Cloud API copy conda	noarch
2566	conda-forge / xcube-cci 0.9.0 xcube plugin for the ESA CCI Open Data Portal copy conda	noarch

Thanks!

Do not hesitate to contact us directly!

gunnar.brandt@brockmann-consult.de or norman.fomferra@brockmann-consult.de

We are grateful to the Open Source Python community, particular to the developers of xarray, zarr, numba, and dask. xcube would not exist without these great packages!

www.brockmann-consult.de