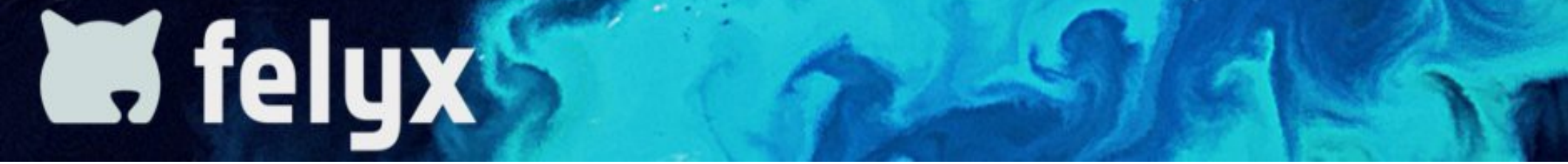


free open-source software to analyse large datasets of Earth observations

Jean-François Piollé

Veronica Guidetti (ESA),
Sylvain Herlédan (Ifremer)
David Poulter (Pelamis)
Jamie Shutler (PML)
Craig Donlon (ESA)

Review of activities | GHRSST XV | Cape Town | 2nd June 2014

esa

Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

applications

- Tool to explore large collections of Earth observations

- Processing algorithms comparison
- Datasets performances inter-comparison
- Identification for processing
- Long term survey of variables (e.g. climate applications..)
- Trend estimation (e.g. variables cross dependency over sites)

- Education, users engagement

felyx

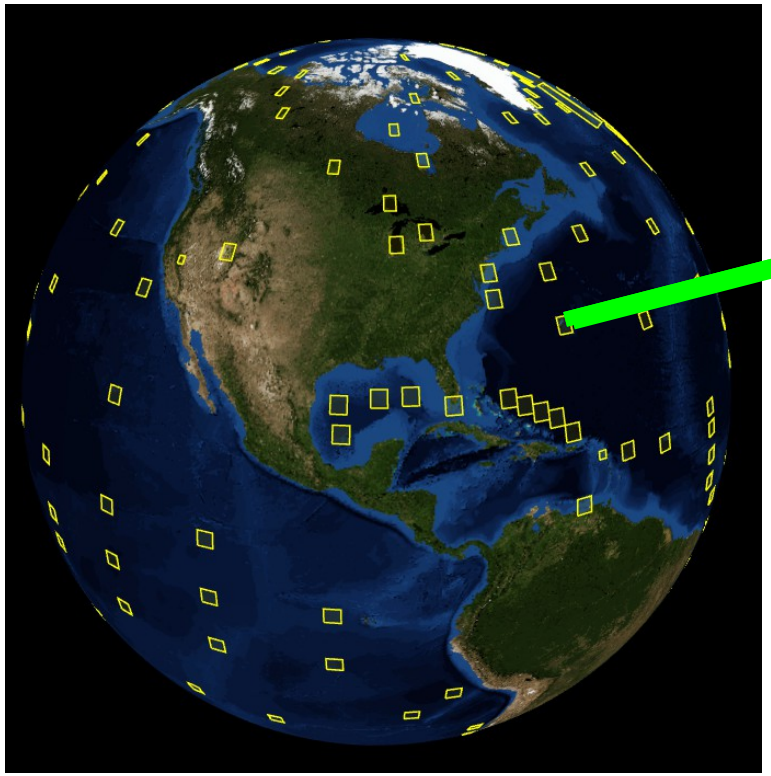
key characteristics

- data download + subsetting
- metrics generation
- cross sensor match-up
- diagnostic reports production
- distributed processing
- networked production sites
- file format and science domain agnostic
- social media

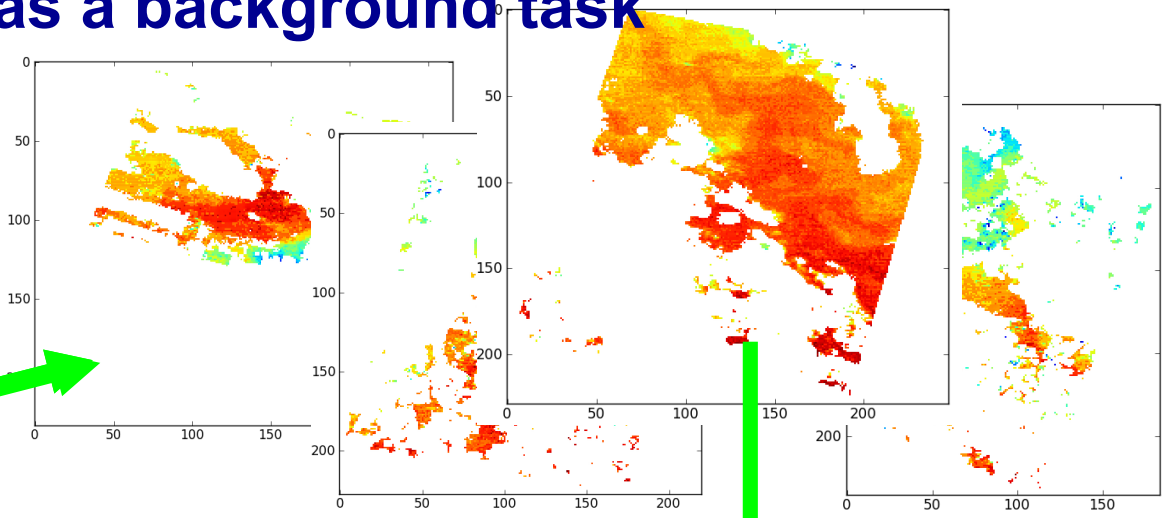
felyx

back-end concept

- extract MINIPRODs over static and dynamic sites
- process quant, qual, stat metrics over MINIPRODs
- Processed on the fly or as a background task



NetCDF files



source: 20130101-IFR-L4_GHRSST-SSTfnd-ODYSSEA-GLO_010-v2.0-fv1.0.nc
felyx_dataset_name: ifr-l4-sstfnd-odyssea-glob_010_v2.1
percentage_coverage_of_site_by_miniprod: 100.0
date_modified: 2014-04-18T10:30:21
felyx_site_identifier: ukm005
date_created: 2014-04-18T10:30:21
time_coverage_start: 2013-01-01T00:00:00
time_coverage_stop: 2013-0101T00:00:00

sst_standard_deviation : 1.34
mean_sst : 286.289
ice_presence: 0
cloud_presence": 46.80
day_or_night: "night"
mean_wind_speed: 4.8388

JSON files
indexed in a search
engine (ElasticSearch)

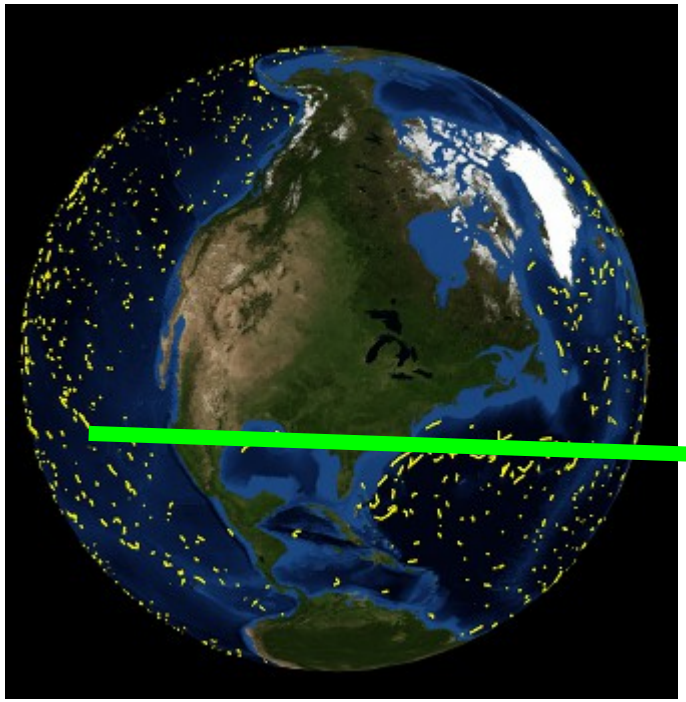
esa

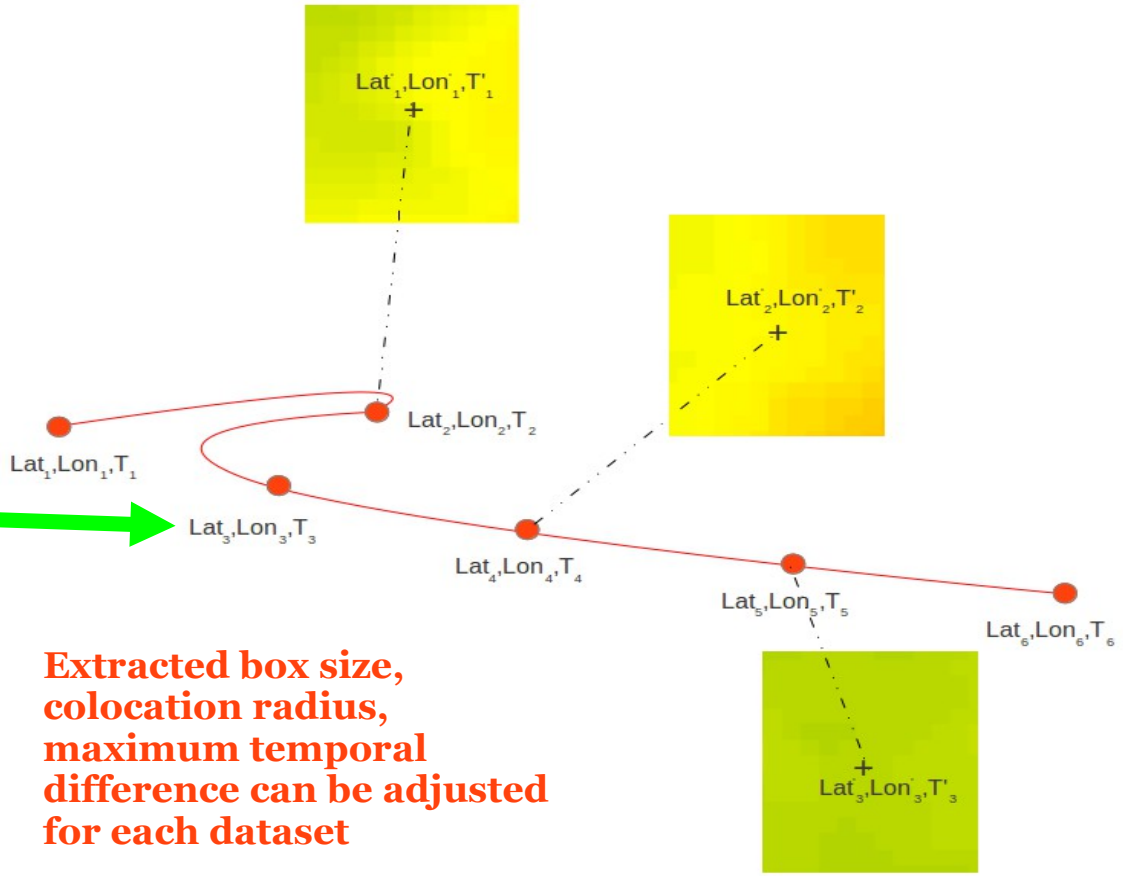
uth Marine
tory

Match-up database concept

- sites may be trajectories (buoys, cruise, hurricane)
- MINIPROD's centred on trajectory locations closest in time

trajectory files ingested through import web service (CSV file)





Extracted box size, colocation radius, maximum temporal difference can be adjusted for each dataset

esa

uth Marine
atory

Recombining matchups with in situ server

Miniprods extracted over trajectories can be recombined with in situ data.

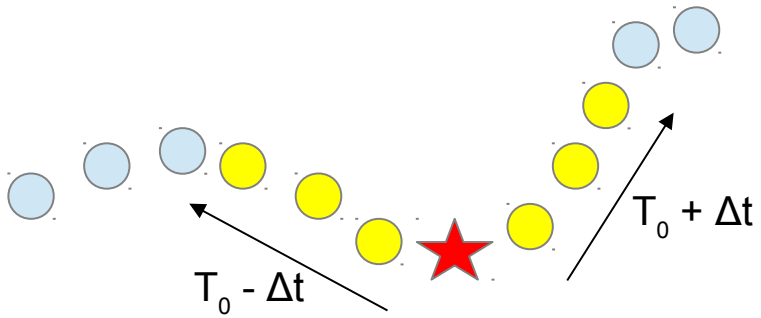
The in situ data server is a third party tool provided together with felyx.

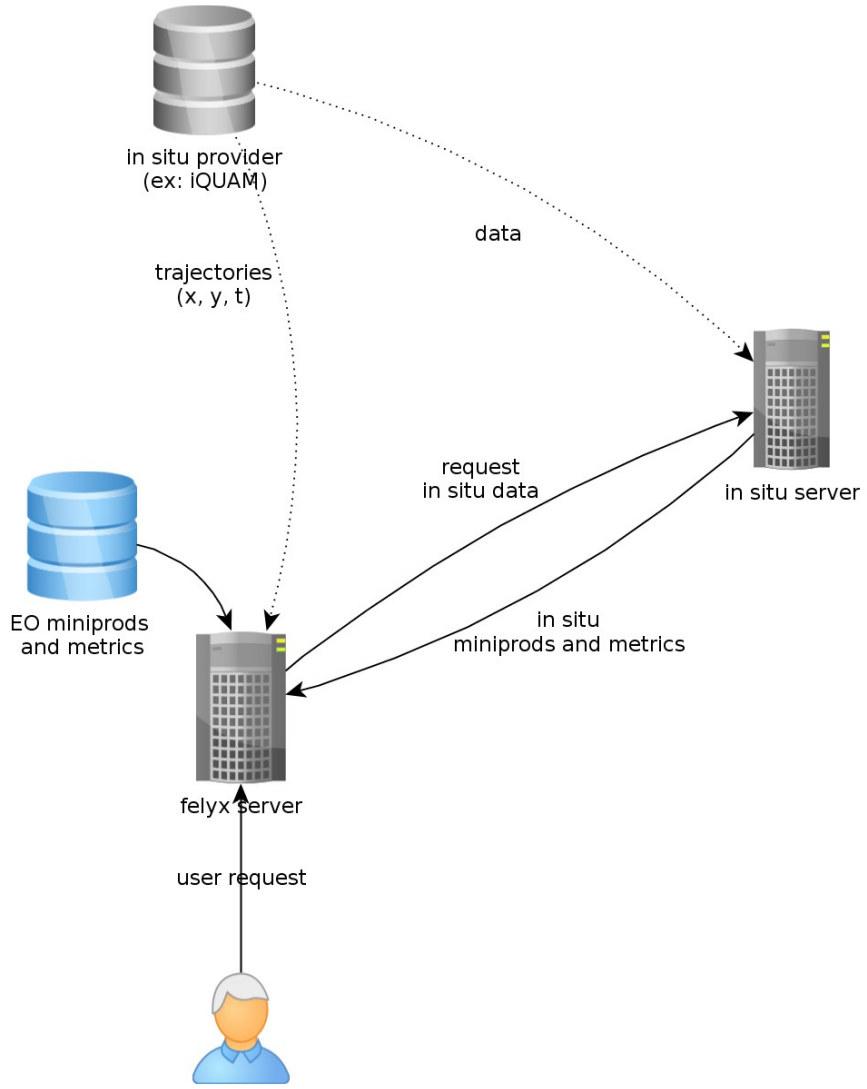
Colocation time criteria can be adjusted

Multiple datasets can be selected at once

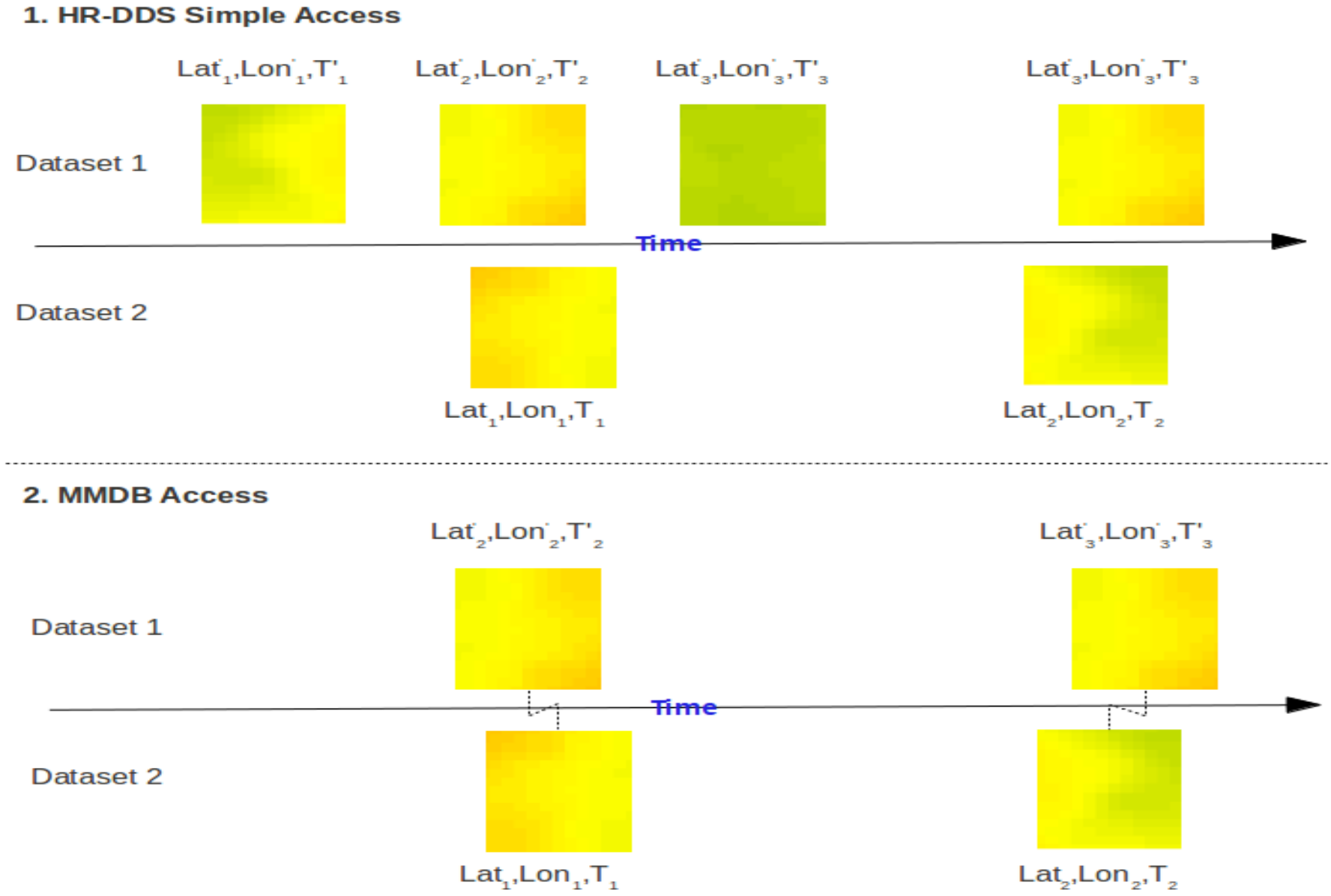
The in situ data server has the same API as felyx and can perform similar operations (and execute similar workflow)

For instance, In situ history can be extracted for each matchup.





Multi-sensor match-up



Access through RESTful web API

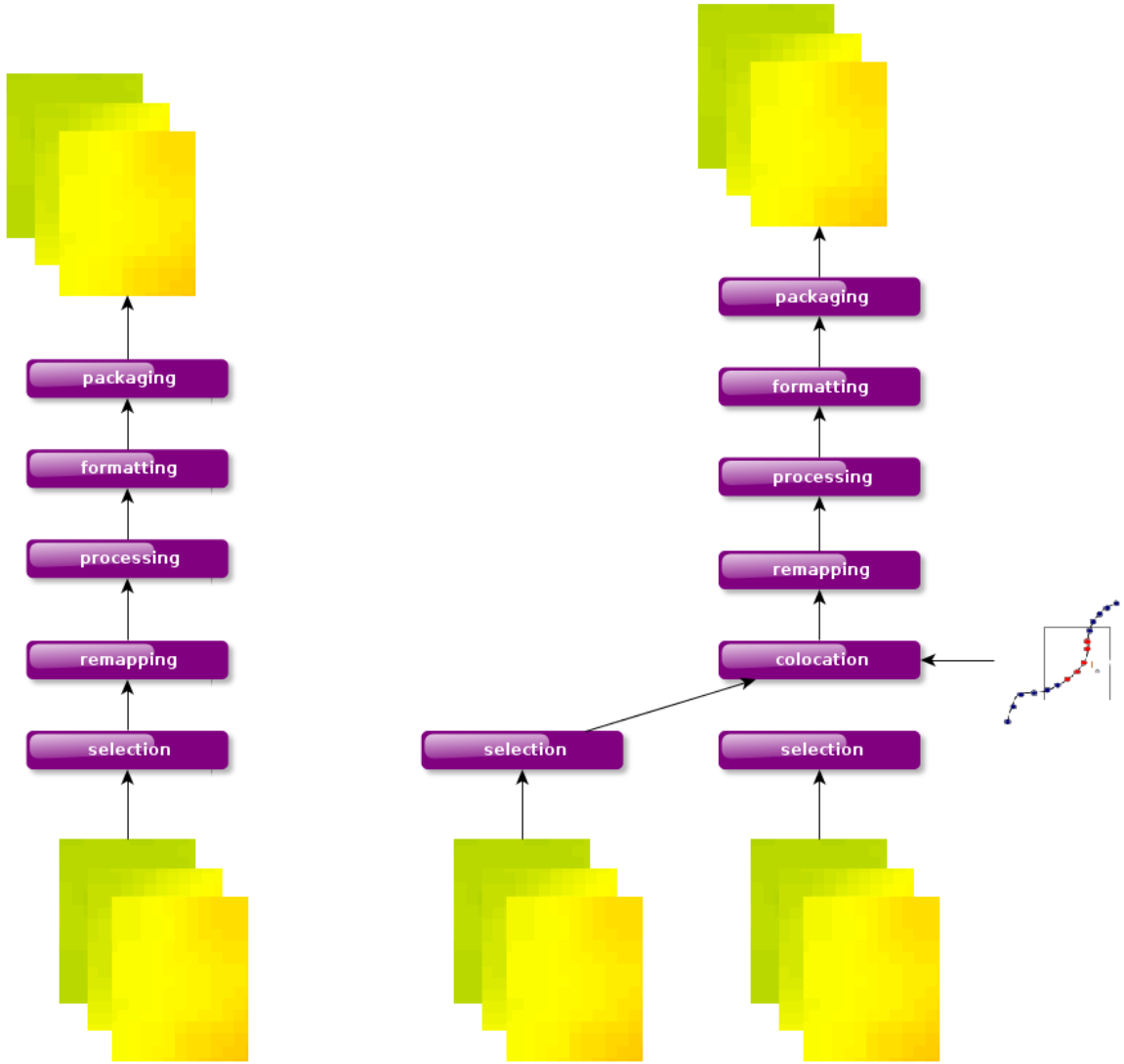
```
curl -XPOST felyx.cersat.fr/extraction/extraction/ -d '{
  "extraction": {
        "selection": {
                "dataset_list": ["arc-upa-l2p-aatsr-v2.1"],
                "start_time": "2002-01-01T00:00:00",
                "stop_time": "2012-04-31T00:00:00",
                "site_list": ["TEKK000", “LYGH001”],
         "metric_list": ["mean_sst"]
        }
  }
}
```

Access through python client API

```python
from pyfelyx.instance import Instance
from pyfelyx.metrics import Selection, Metrics

# instanciate felyx instance to query from (here Ifremer)
inst = Instance(url='http://www.ifremer.fr/cersat1/exp/felyx/')
# instanciate a selection query
result = Selection(
             sites=['TEKK000', 'LYGH001'],
             datasets=['arc-upa-l2p-aatsr-v2.1'],
             metrics=['mean_sst'],
             start=datetime(2002,1,1),
             end=datetime(2012,04,31),
             )
 # perform query
 res = Metrics.select(
   inst,
   selection)

 # from here you can start working with the returned data
 # ....

 # print columned result
 Metrics.nice_metrics_display(res)
```

API – workflow concept

```
curl -XPOST 78.145.246.215/felyx/extraction/extraction/ -d '{
    "extraction": {
        "packaging": {
            "input": {
                "selection": {
                    "dataset_list": ["SST-34"],
                "start_time": "2012-01-01T00:00:00",
                    "stop_time": "2014",
                     "site_list": ["ghr274"]
                }
            },
            "operation_name": "targzip"
        }
    }
}' > data.tgz
```



Data access

- Raw data (extracted miniprods and metrics) accessible through :

  Netcdf format for miniprods

  Csv, netcdf, json for metrics

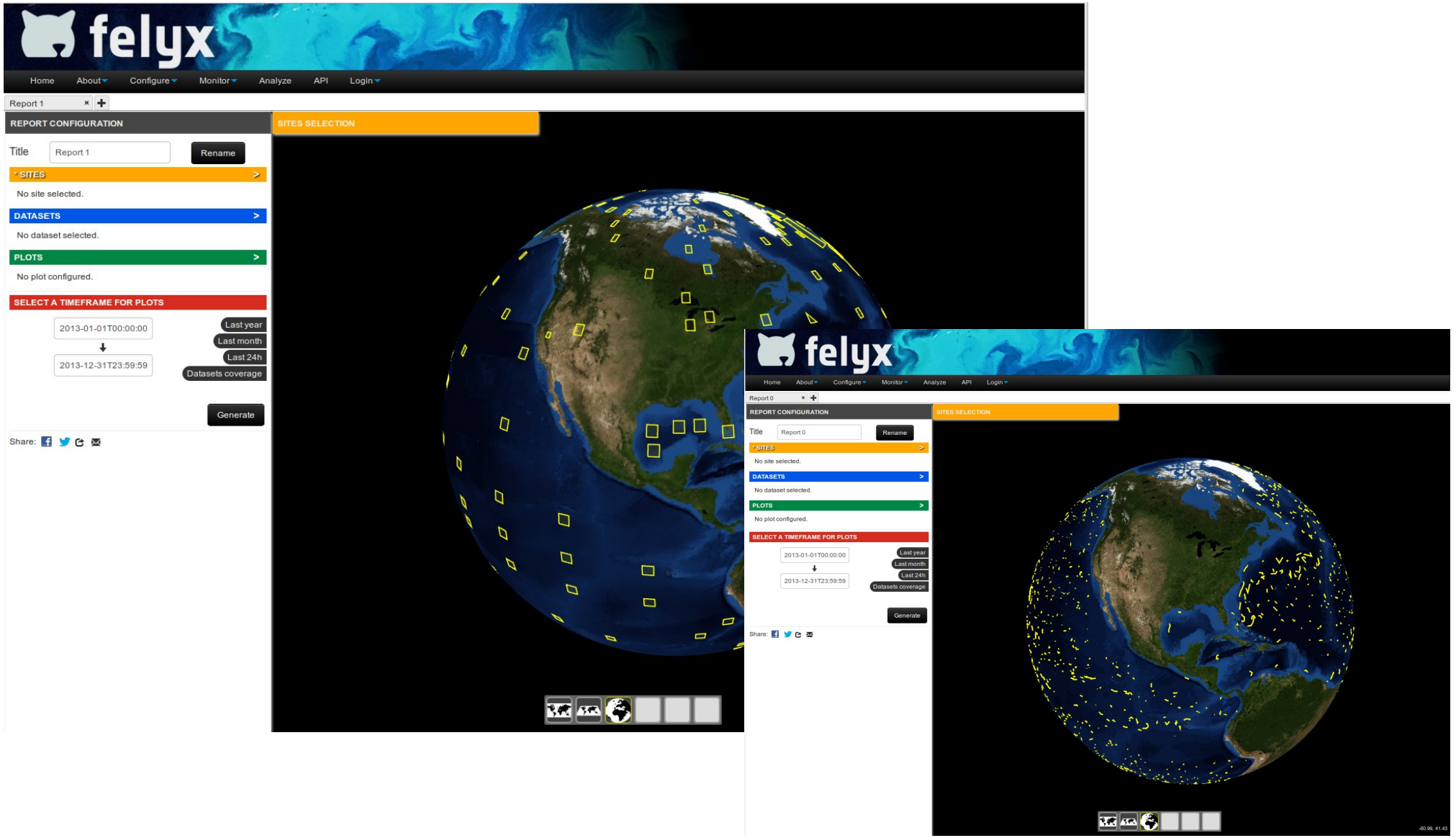
  FTP / OpenDAP

  - Query through RESTful web service (tar file or values) for more advanced selection criteria (for instance matchup)
  - Selection results (for instance MDB) could be pre-processed and packaged for specific applications

- Reporting and visualization through front-end or user designed scripts or applications

felyx

sites selection



esa

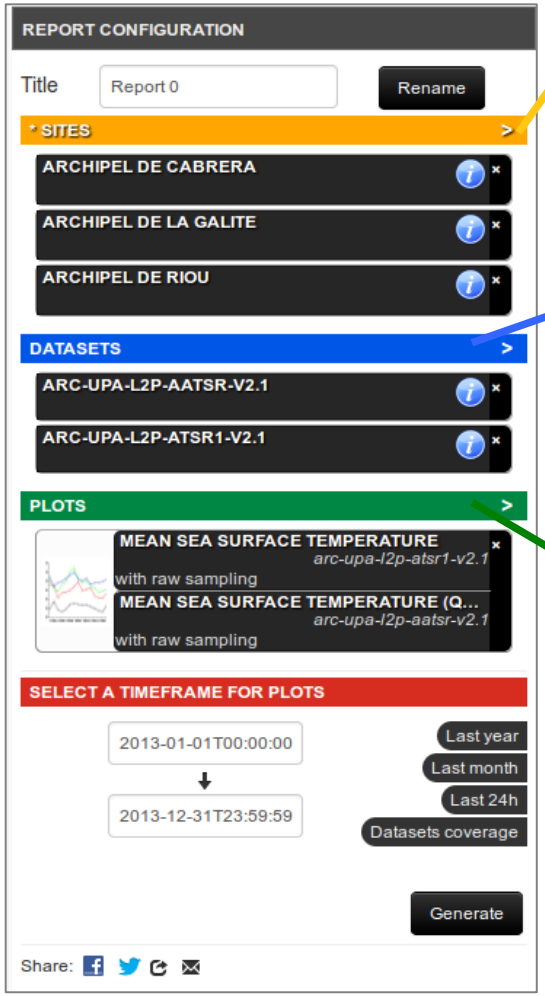
Ifremer

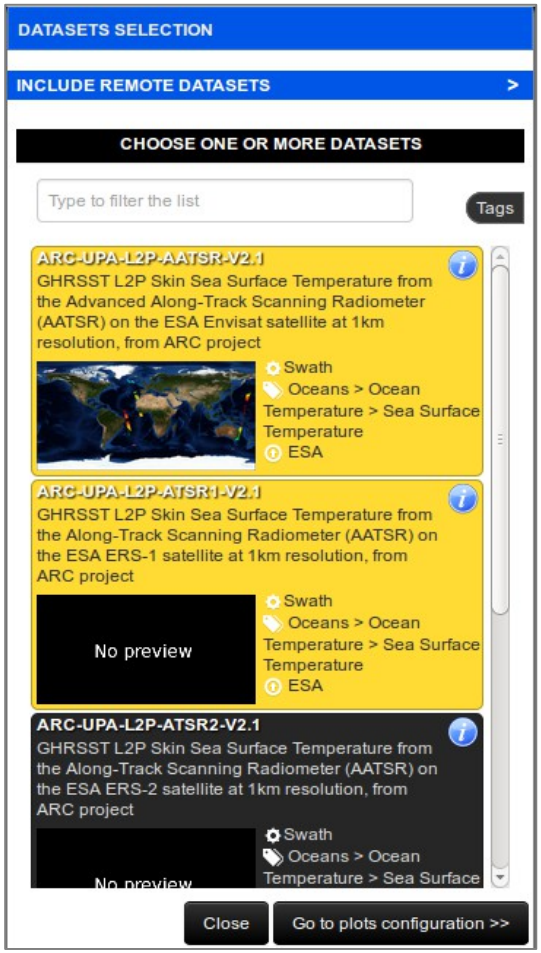
pelamis scientific software

PML Plymouth Marine Laboratory

felyx

config your report

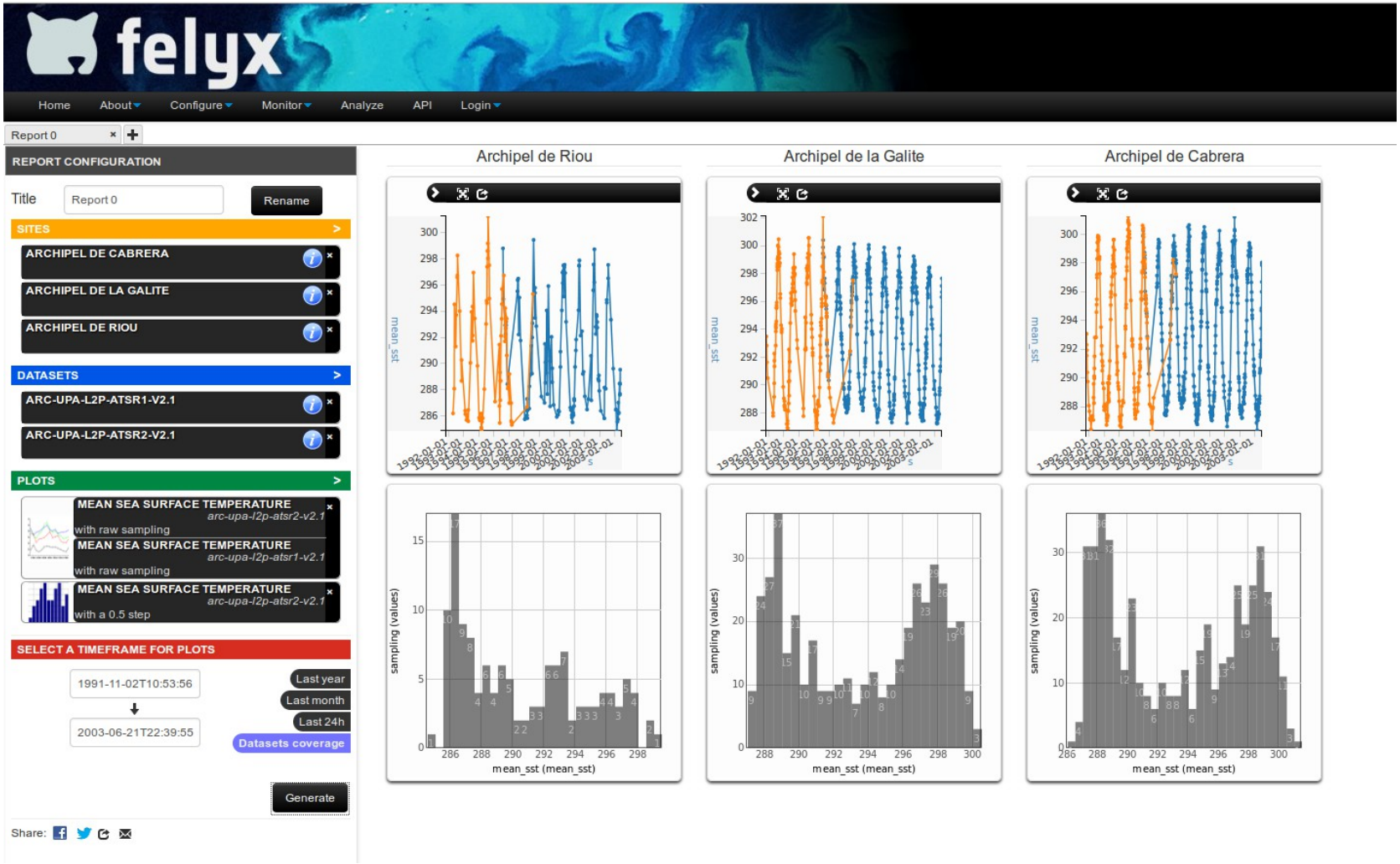




esa

felyx

display your report



esa

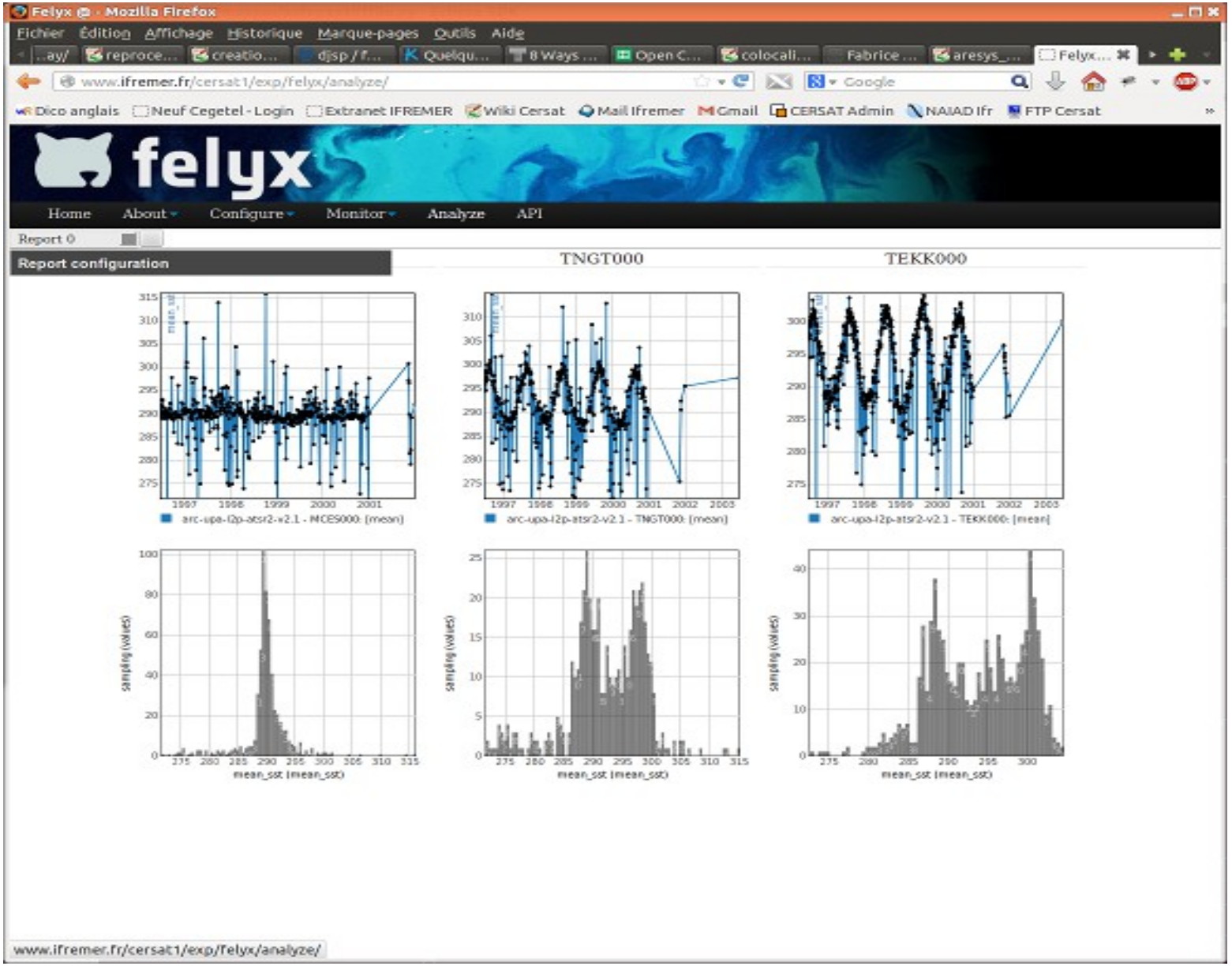
Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

felyx

display your report



esa

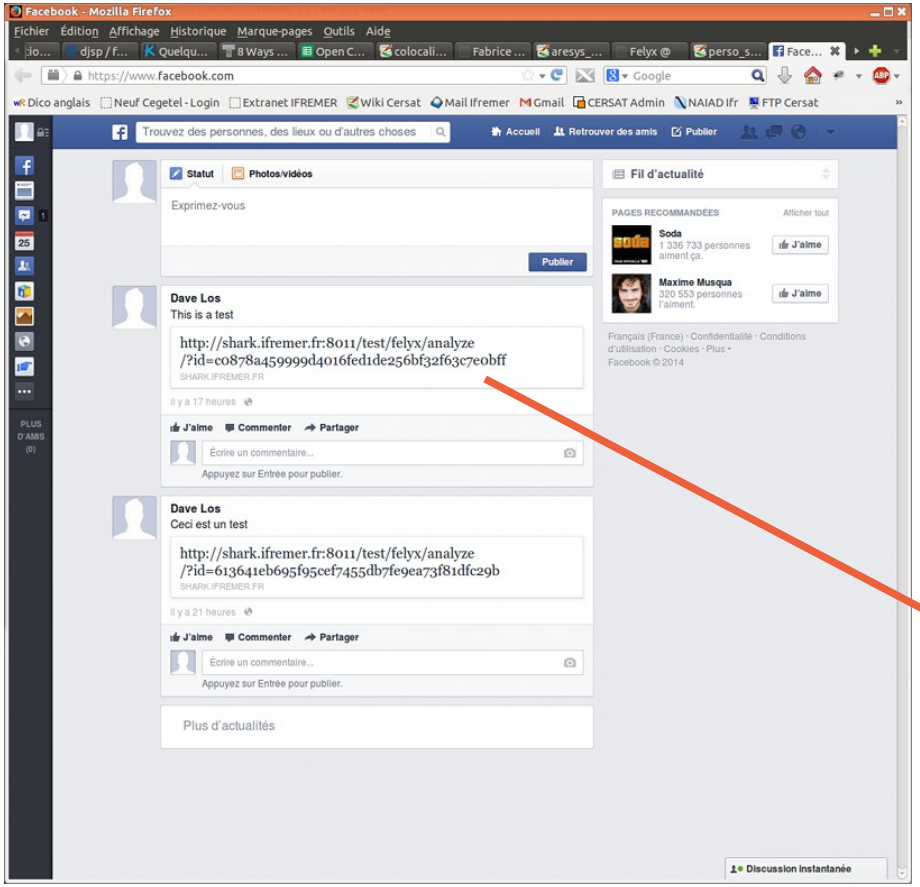
Ifremer

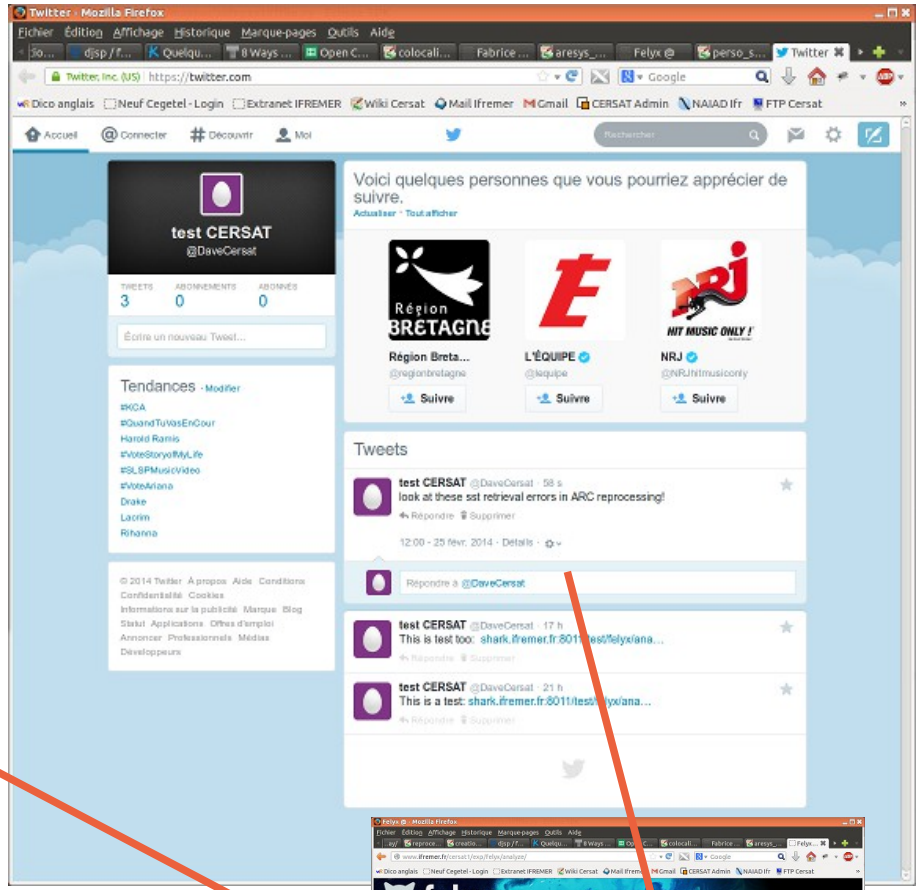
pelamis scientific software

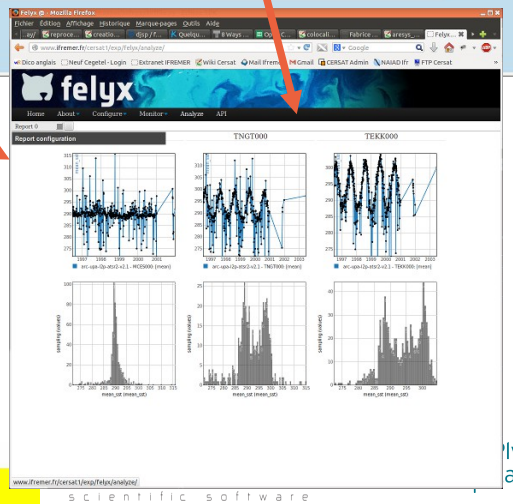
PML Plymouth Marine Laboratory

felyx

share your work







bookmark report

esa

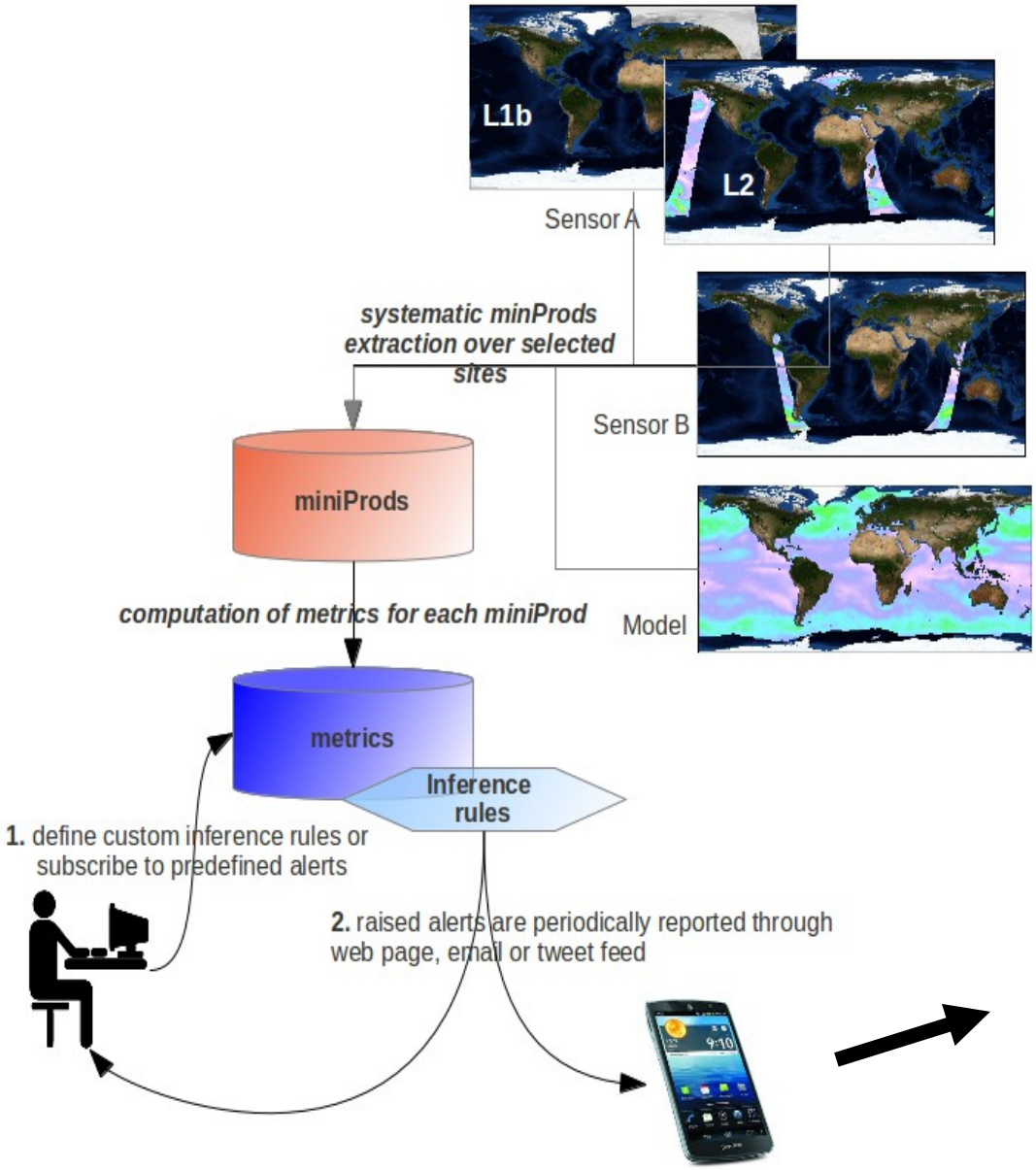
Ifremer

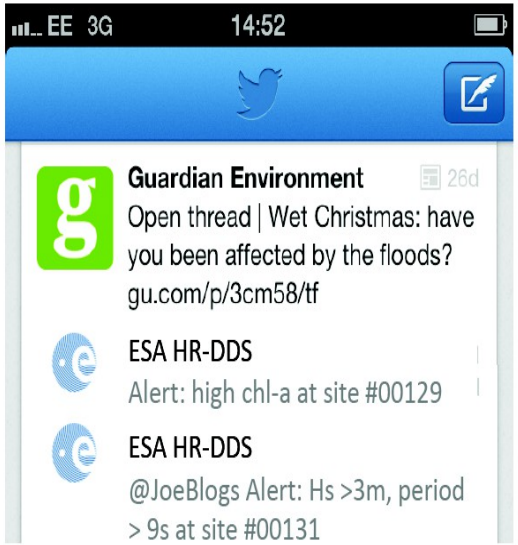
scientific software

Plymouth Marine Laboratory

felyx

anomaly detection and alert





Example of how HR-DDS alert tweets could appear on a phone twitter application.

From: ESA HR-DDS
Subject: HR-DDS Alert (site #00131)
To: You

Dear HR-DDS user,

The following conditions:
Hs > 3m
Period > 9s

Have been met at:
Site #00131 on 22/01/2013
http://hrdds.net/site00131.htm

Regards,
ESA HR-DDS system

Example of how an HR-DDS email alert might look.

esa

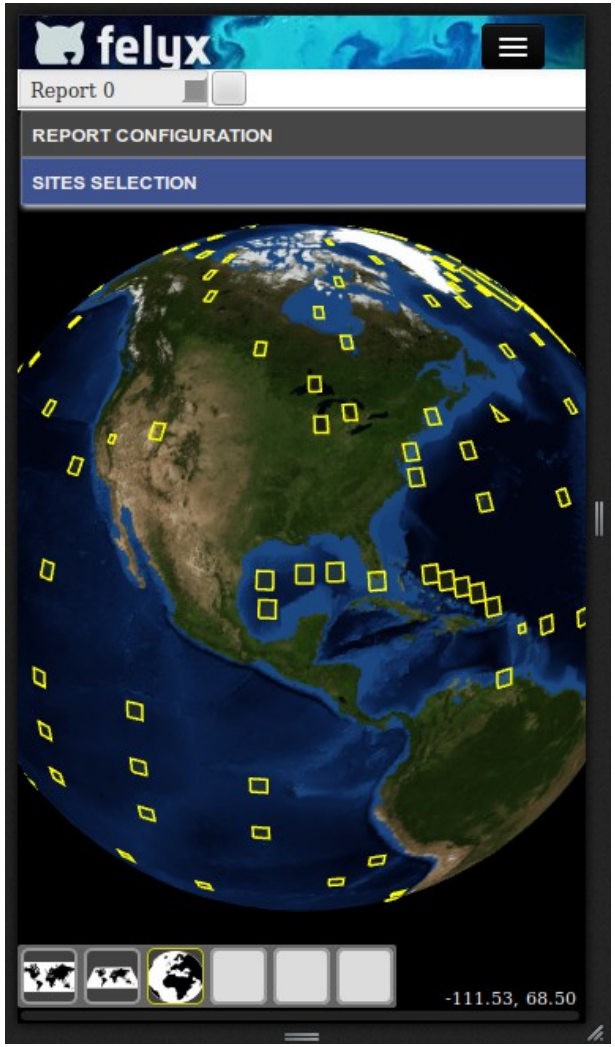
Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

felyx

use it wherever



smartphone

tablet

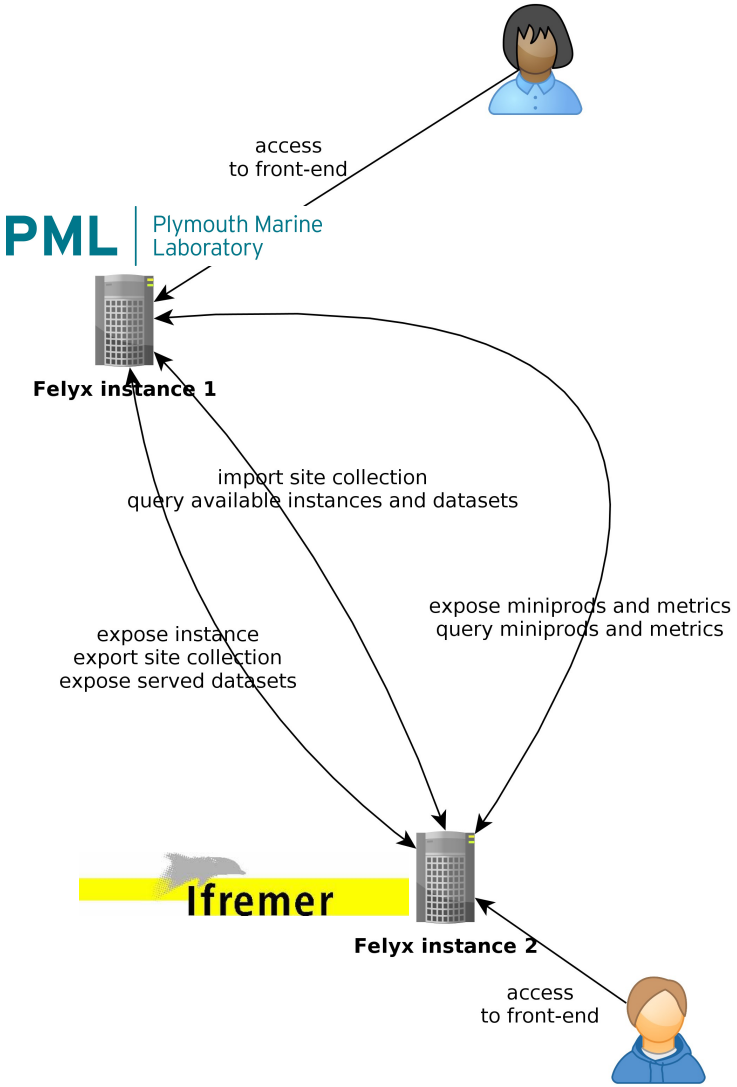
esa

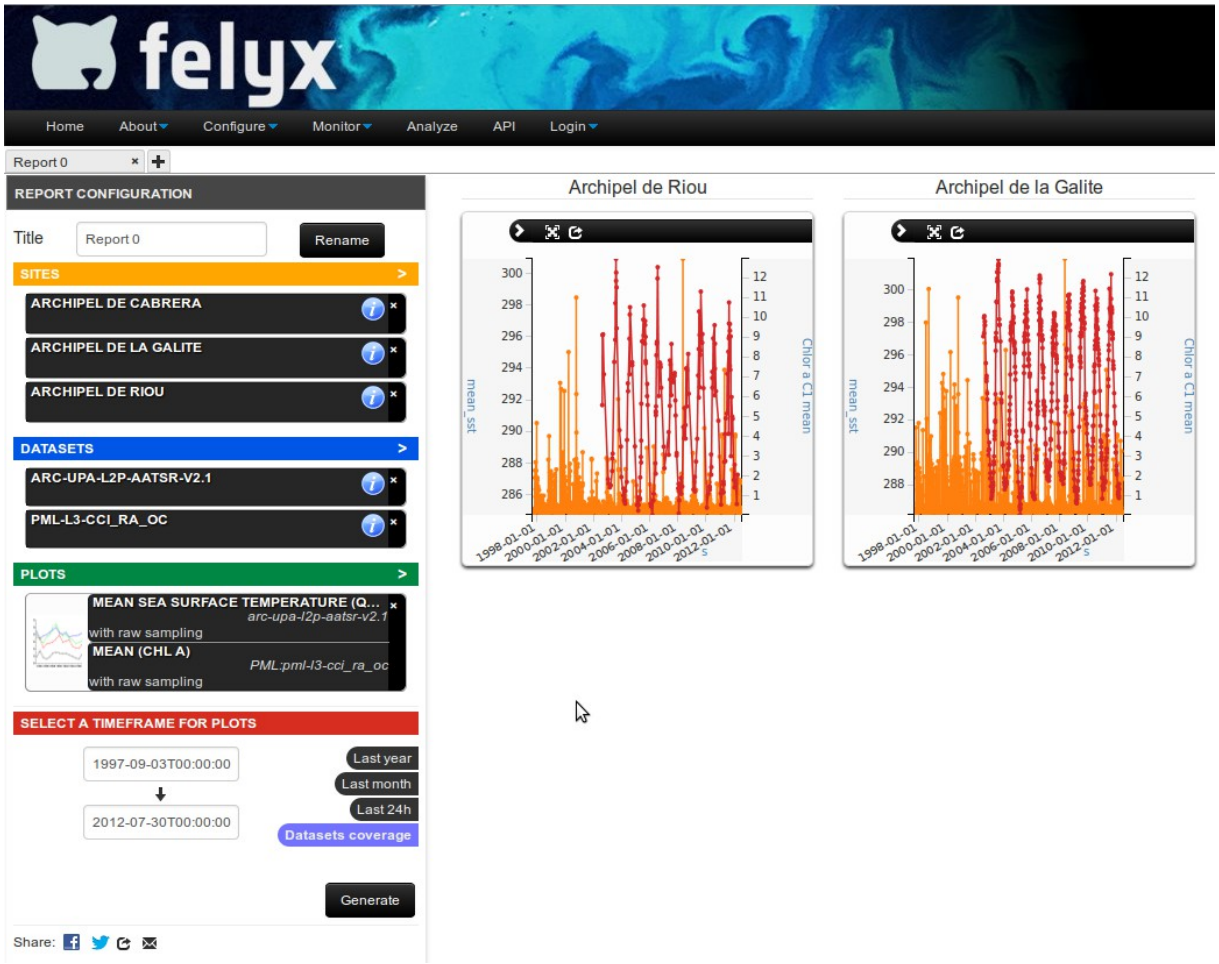
Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

networking felyx





orange / red curves correspond to metrics fetched from Ifremer and PML instances

Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

felyx

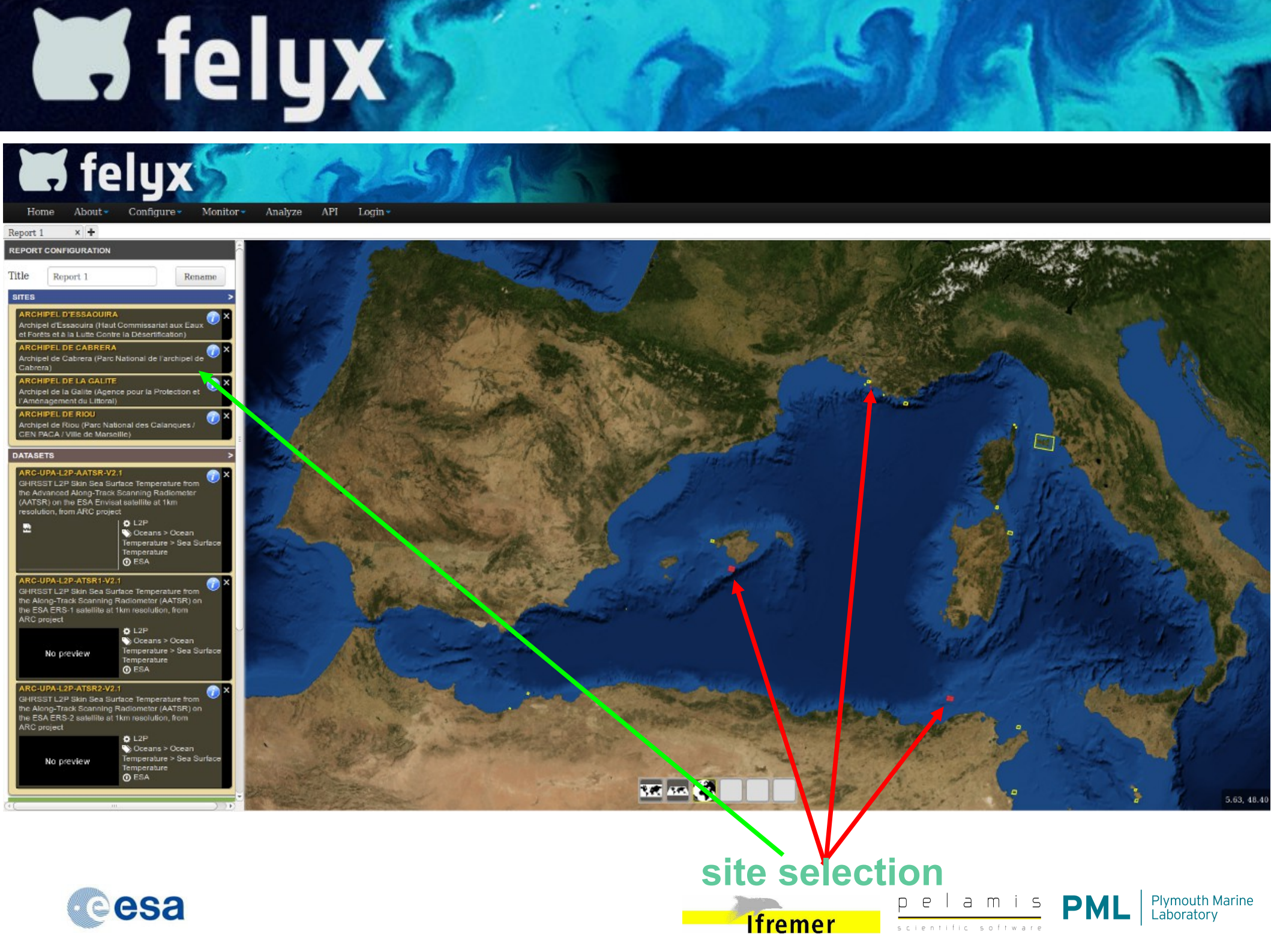
sample use cases

MED SEA

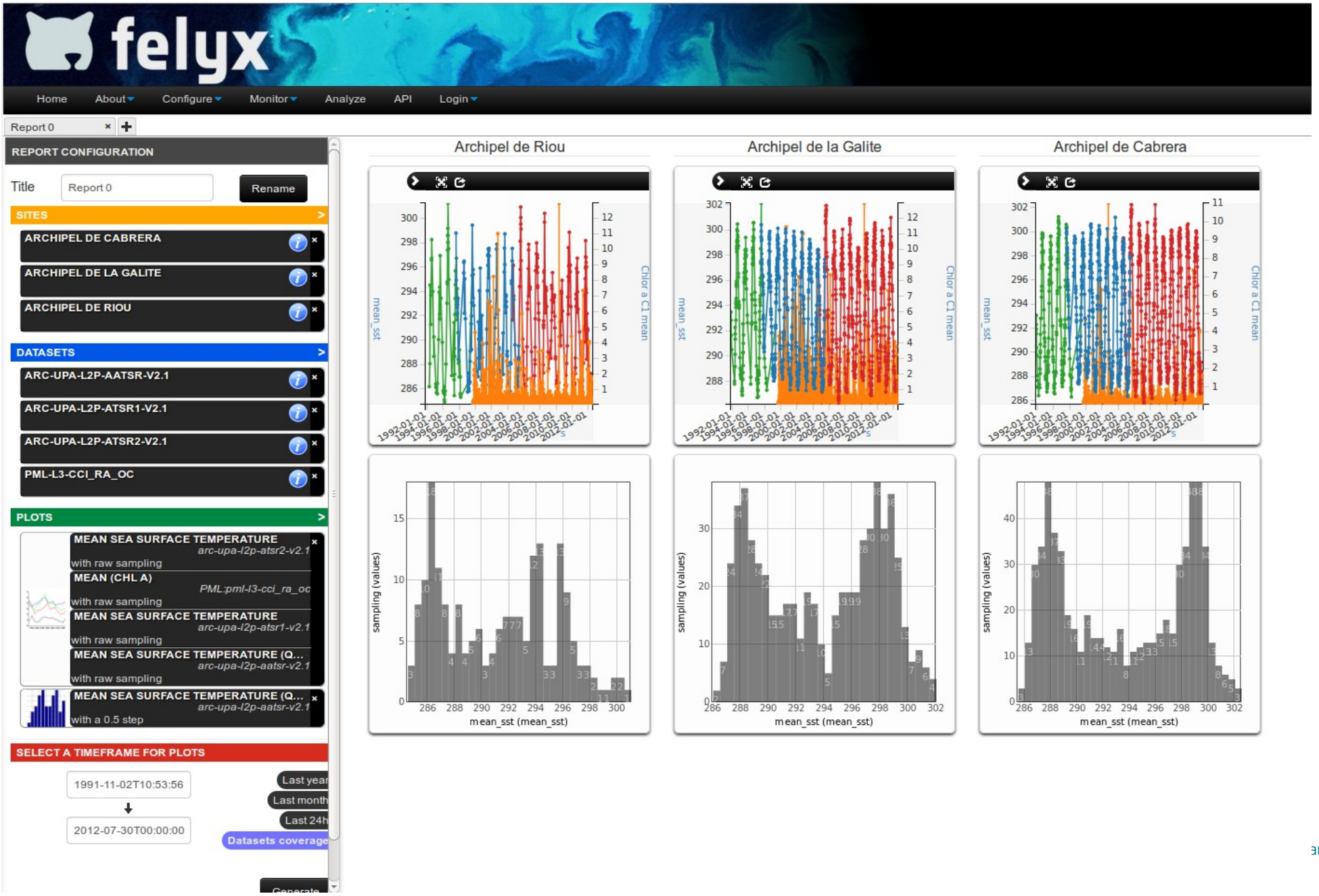
- assess climate variability and change
- process, display long-time series of climate variables, SST (ARC + Pathfinder reprocessing), OC, ALT
- cross instance communication

GHRSST match-up

- cross sensor inter-comparison in GHRSST
- NRT datasets METOP, VIIRS, AMSR2, SEVIRI, AVHRR
- SAT-INSITU matchups
- dynamic sites use (iQUAM drifters + argo floats)

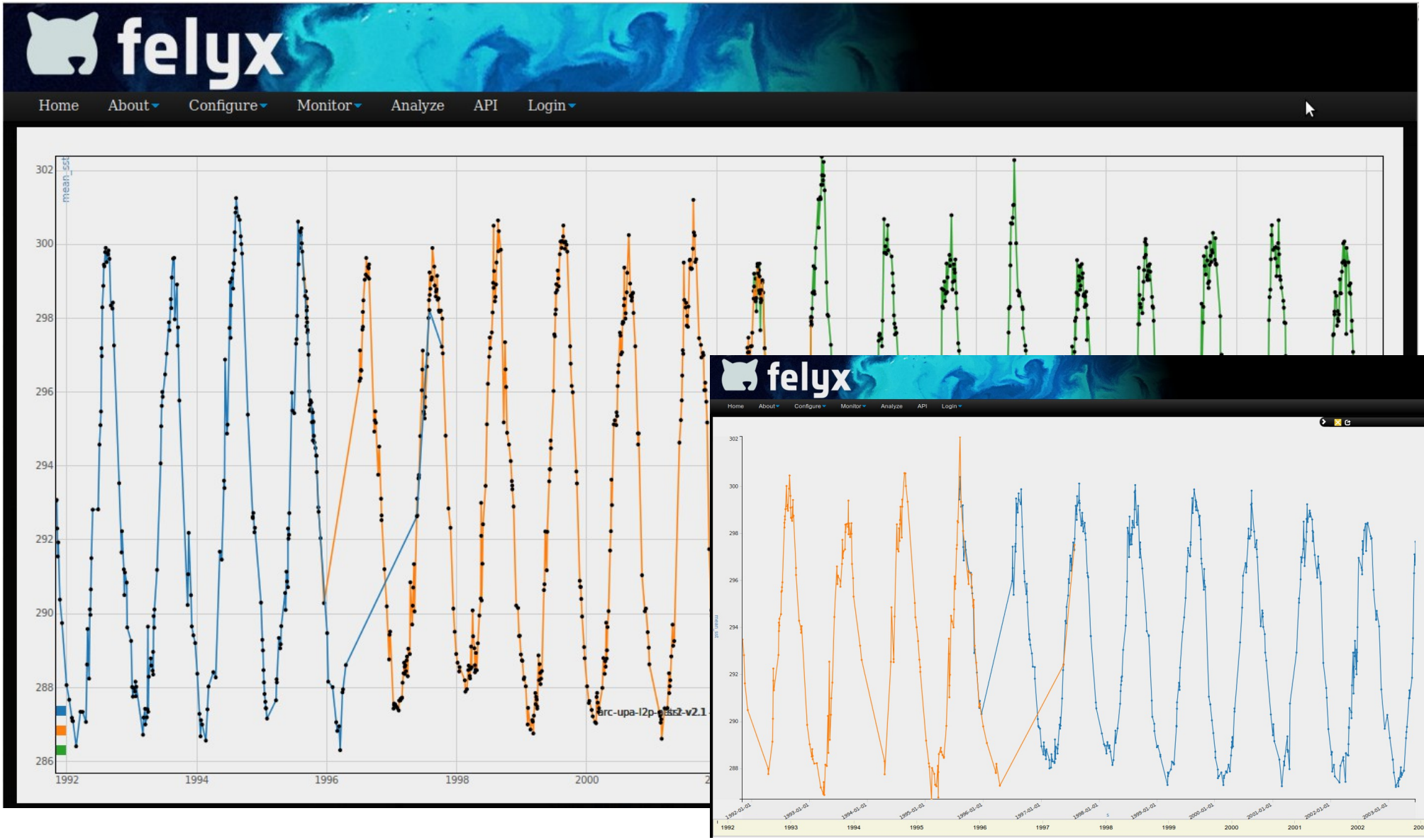

felyx
felyx
Home
About
Configure
Monitor
Analyze
API
Login
Report 1
REPORT CONFIGURATION
Title
Report 1
Rename
SITES
ARCHIPEL D'ESSAOUIRA
Archipel d'Essaouira (Haut Commissariat aux Eaux et Forêts et à la Lutte Contre la Désertification)
ARCHIPEL DE CABRERA
Archipel de Cabrera (Parc National de l'archipel de Cabrera)
ARCHIPEL DE LA GALITE
Archipel de la Galite (Agence pour la Protection et l'Aménagement du Littoral)
ARCHIPEL DE RIOU
Archipel de Riou (Parc National des Calanques / CEN PACA / Ville de Marseille)
DATASETS
ARC-UPA-L2P-AATSR-V2.1
GHRSST L2P Skin Sea Surface Temperature from the Advanced Along-Track Scanning Radiometer (AATSR) on the ESA Envisat satellite at 1km resolution, from ARC project
L2P
Oceans > Ocean Temperature > Sea Surface Temperature
ESA
ARC-UPA-L2P-ATSR1-V2.1
GHRSST L2P Skin Sea Surface Temperature from the Along-Track Scanning Radiometer (AATSR) on the ESA ERS-1 satellite at 1km resolution, from ARC project
No preview
L2P
Oceans > Ocean Temperature > Sea Surface Temperature
ESA
ARC-UPA-L2P-ATSR2-V2.1
GHRSST L2P Skin Sea Surface Temperature from the Along-Track Scanning Radiometer (AATSR) on the ESA ERS-2 satellite at 1km resolution, from ARC project
No preview
L2P
Oceans > Ocean Temperature > Sea Surface Temperature
ESA
5.63, 48.40
site selection
esa
Ifremer
pelamis
scientific software
PML
Plymouth Marine Laboratory

felyx


felyx
Home
About
Configure
Monitor
Analyze
API
Login
Report 0
REPORT CONFIGURATION
Title
Report 0
Rename
SITES
ARCHIPEL DE CABRERA
ARCHIPEL DE LA GALITE
ARCHIPEL DE RIOU
DATASETS
ARC-UPA-L2P-AATSR-V2.1
ARC-UPA-L2P-ATSR1-V2.1
ARC-UPA-L2P-ATSR2-V2.1
PML-L3-CCI_RA_OC
PLOTS
MEAN SEA SURFACE TEMPERATURE
arc-upa-l2p-atsr2-v2.1
with raw sampling
MEAN (CHL A)
PML:pml-l3-cci_ra_oc
with raw sampling
MEAN SEA SURFACE TEMPERATURE
arc-upa-l2p-atsr1-v2.1
with raw sampling
MEAN SEA SURFACE TEMPERATURE (Q...
arc-upa-l2p-aatsr-v2.1
with raw sampling
MEAN SEA SURFACE TEMPERATURE (Q...
arc-upa-l2p-aatsr-v2.1
with a 0.5 step
SELECT A TIMEFRAME FOR PLOTS
1991-11-02T10:53:56
2012-07-30T00:00:00
Last year
Last month
Last 24h
Datasets coverage
Generate
Archipel de Riou
Archipel de la Galite
Archipel de Cabrera
mean_sst
Chlor a C1 mean
sampling (values)
mean_sst (mean_sst)

arine

felyx



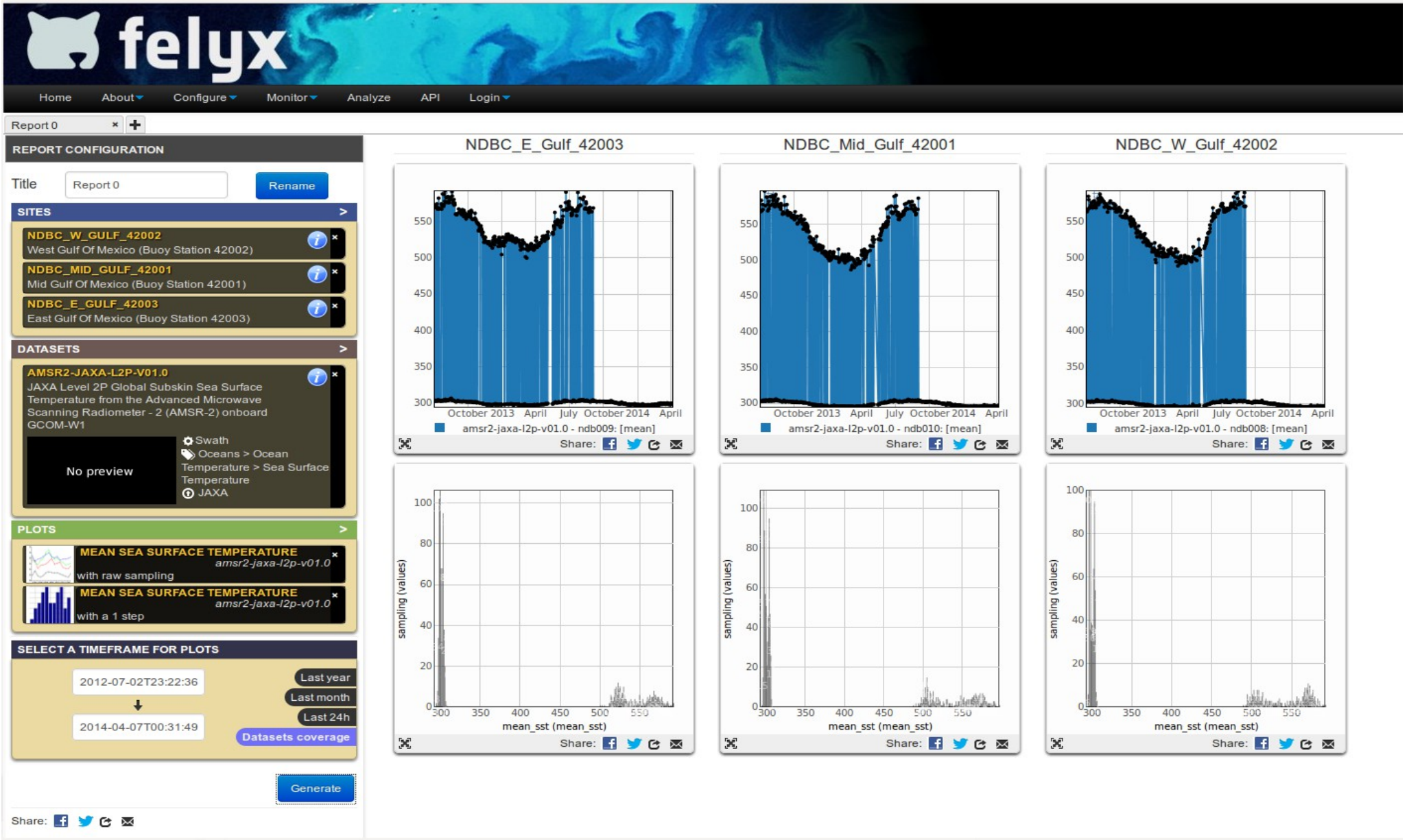
esa

Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

felyx



esa

Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

Felyx is agnostic and extensible

- Can handle any type of acquisition pattern :
  - Swath, grid, trajectories or along-track, …
- Can handle any kind of geophysical parameter
- Can handle any kind of domain : ocean, land, atmosphere, cryosphere, …
- Can handle new data format through an extensible plugin mechanism that can be tailored by any administrator of felyx system
- Metrics processing can be extended too through plugin mechanism

Current datasets

SST

ERS-1 ARC ATSR L2P
ERS-2 ARC ATSR L2P
ENVISAT ARC AATSR L2P
JAXA AMSR2 L2P
O&SI SAF METOP L2P
O&SI SAF METOP L3
O&SI SAF SEVIRI
O&SI SAF GOES13
O&SI SAF NAR L2P
JPL MODIS L2P
NAVO LAC & GAC L2P
RSS AMSRE L2P
RSS TMI L2P
Medspiration L4
MyOcean L4 (NWE,Baltic,Arctic)
Odyssea Global L4
AVHRR Pathfinder

WAVES

All GlobWave L2P products
JASON-1 GDR & NRT
JASON-2 GDR & NRT
AltiKa IGDR
CryoSat-2 IGDR
Envisat GDR & NRT
Topex GDR
ERS-2 GDR
ERS-1 GDR
GFO
GeoSat

OCEAN COLOUR

NASA SeaWIFS L1A,L2
NASA MODIS L1A,L2

Installation and deployment

Objectives

To be easy for users to build their own applications querying from existing Felyx instances

- » Instances share a common, base RESTfull API.
- »Users are encouraged to build their own applications, and to host their own instance with their own data, configured to their needs.!

To be easy for users to deploy their own instance

- »from source code
- »providing linux container (http://en.wikipedia.org/wiki/LXC)

To be possible and easy to tailor the system to each user need : parameter, datasets, metrics, front-end, ….

- »Free open source (GPL v3), fully documented, for uptake by the user community. Users can modify source code, and resubmit modifications to GitHub.
- »Felyx is fully written in python (back-end) and javascript (front-end), using third party components compatible with above licensing
- »Users can make commercial usage of the system.
- »Providing source code modifications is not mandatory (but encouraged)

Yes....but is it working ?

Where do we stand ?

Back-end is almost implemented, advanced services and interfaces yet under development

More bug correction and performance improvement is required

Improvement of administration, activity monitoring, overall API and system documentation

So when ?

Starting with full scale use cases to assess performances and scalability of the system

Completing implementation

Expected release is September with 6 month full scale demonstration phase

felyx

Felyx brings distributed diagnostic datasets and multi-sensor match-up database capabilities and analysis tools to many, virtually all environmental data. Beginning September 2014.

Project web site : http://www.felyx.org

Contact us: jfpiolle@ifremer.fr

esa

Ifremer

pelamis scientific software

PML Plymouth Marine Laboratory

Indications on data volumes

Dataset type	Volume on static sites	Volumes for Argo and drifters
SST L4 (grid) product, 10km resolution, daily	4.2 GB / year	
high resolution SST L2 (swath) product, 1km resolution (typical LEO radiometer for SST or Ocean Colour)	~ 80 GB / year	
AMSR-2	16 GB / year	60 GB / year

current metrics storage (in Elasticsearch search engine) is : 3.5 GB.

large storage (several TB) when you want to handle several collections of data.

Optimization of output format still to be done