

## THE ESA FELYX HIGH RESOLUTION DIAGNOSTIC DATA SET SYSTEM DESIGN AND IMPLEMENTATION

M. Taberner<sup>1</sup>, J. Shutler<sup>1</sup>, P. Walker<sup>1</sup>, D. Poulter<sup>2</sup>, J-F. Piolle<sup>3</sup>, C. Donlon<sup>4</sup>, and V. Guidetti<sup>4</sup>

1. Remote Sensing Group, Plymouth Marine Laboratory, Plymouth, UK. ([Malcolm.Taberner@pml.ac.uk](mailto:Malcolm.Taberner@pml.ac.uk))

2. Pelamis Scientific Software, Bath, U.K.

3. IFREMER, Brest, France.

4. European Space Agency.

**KEY WORDS:** Multitemporal data analysis, multisensor, insitu data, tracking, trend analysis

### ABSTRACT:

Felyx is currently under development and is the latest evolution of a generalised High Resolution Diagnostic Data Set system funded by ESA. It draws on previous prototype developments and experience in the GHRSSST, Medspiration, GlobColour and GlobWave projects. In this paper, we outline the design and implementation of the system, and illustrate using the Ocean Colour demonstration activities.

Felyx is fundamentally a tool to facilitate the analysis of EO data: it is being developed by IFREMER, PML and Pelamis. It will be free software written in python and javascript. The aim is to provide Earth Observation data producers and users with an open-source, flexible and reusable tool to allow the quality and performance of data streams from satellite, in situ and model sources to be easily monitored and studied. New to this project, is the ability to establish and incorporate multi-sensor match-up database capabilities. The systems will be deployable anywhere and even include interaction mechanisms between the deployed instances.

The primary concept of Felyx is to work as an extraction tool. It allows for the extraction of subsets of source data over predefined target areas(which can be static or moving). These data subsets, and associated metrics, can then be accessed by users or client applications either as raw files or through automatic alerts. These data can then be used to generate periodic reports or be used for statistical analysis and visualisation through a flexible web interface.

Felyx can be used for subsetting, the generation of statistics, the generation of reports or warnings/alerts, and in-depth analyses, to name a few.

There are many potential applications but important uses foreseen are :

- \* monitoring and assessing the quality of Earth observations (e.g. satellite products and time series) through statistical analysis and/or comparison with other data sources
- \* assessing and inter-comparing geophysical inversion algorithms
- \* observing a given phenomenon, collecting and cumulating various parameters over a defined area
- \* crossing different sources of data for synergy applications

The services provided by felyx will be generic, deployable at users own premises, and flexible allowing the integration and development of any kind of parameters. Users will be able to operate their own felyx instance at any location, on datasets and parameters of their own interest, and the various instances will be able to interact with each other, creating a web of felyx systems enabling aggregation and cross comparison of miniProds and metrics from multiple sources.

Initially two instances will be operated simultaneously during a 6 months demonstration phase, at IFREMER - on sea surface temperature and ocean waves datasets - and PML - on ocean colour.

**WEBSITE:** <http://hrdds.ifremer.fr/>

## 1. INTRODUCTION

### 1.1 Capability

felyx is, primarily, a set of data extraction tools used to subset source data over predefined target areas which can be static or moving. These subsets, and any associated metrics, are accessible to users or machines as raw files, automatic alerts and periodic reports.

felyx tools are open-source and provide both back-end and front-end software components to:

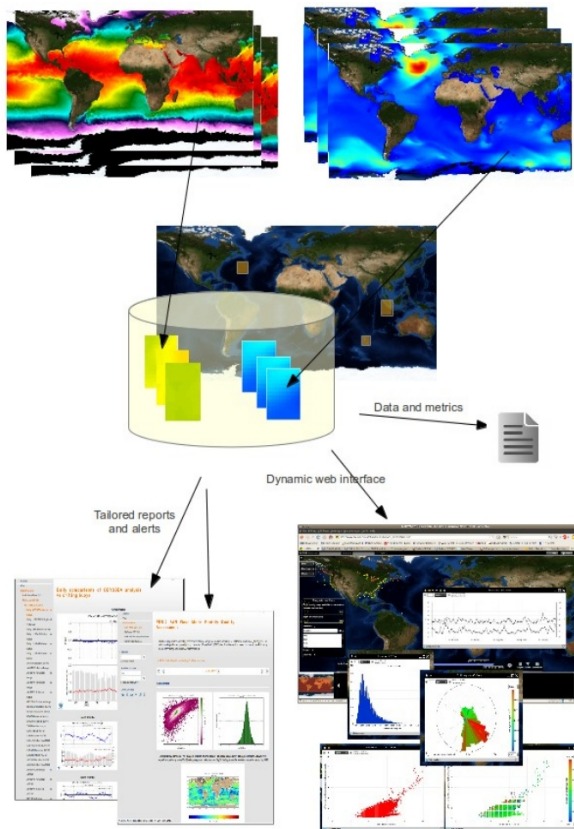
- subset large local or remote collections of Earth Observation data over predefined sites (geographical boxes) or moving targets (ship, buoy, hurricane), storing locally the extracted data (referred as miniProds). These data can be directly accessed by users, they constitute a much smaller representative subset of the original collection on which one can perform any kind of

processing or assessment without having to cope with heavy volumes of data.

- compute statistical (or any kind of) metrics over these extracted subset using for instance a set of classic statistical operators (mean, median, rms, ...) that is fully xtensible over some parameters of each dataset. These metrics are stored in a database and coupled with a fast search engine (ElasticSearch) from which they can be later queried by users or automated applications.

- provide periodic reports and raise alerts based on a user-defined set of inference rules through various media (email, twitter feed,...) and devices. The content and conditions on which this information is sent to the user is fully configurable through a web interface.

- analyse the content of the miniProds and metrics through a dedicated web interface allowing the information to be examined and useful knowledge to be extracted through multidimensional interactive display functions (time series, scatterplots, histograms, maps, etc.).



## 1.2 Applicability

Among several potential applications, users will be able to use felix for:

- monitoring and assessing the quality of Earth observations (e.g. satellite products and their time series) through statistical analysis and/or comparison with other data sources, in NRT or over longer periods
- assessing and inter-comparing geophysical inversion algorithms or different datasets
- alerting and reporting on performance degradation or specific conditions
- performing geophysical analysis (variability, trends, etc.)
- monitor a given phenomenon, collecting and accumulating various parameters over a defined area or time period.
- comparing and combining different sources of data for synergy applications

## 1.3 Utility

In this context, felix will serve different kind of users:

- instrument engineers
- calibration engineers
- quality control engineers
- project validation scientists
- external validation scientists
- algorithm developers
- scientific validation community
- ocean/atmosphere scientific community.

The target users and validators of the system will also be part of the project Reference User Group.

felix will provide services deployable at the users' own premises and adaptable enough to integrate potentially any kind of parameters. Users will be able to operate their own felix instance at any location, on datasets and parameters of their own interest. The different instances of felix will be able to interact with each other, creating a web of systems enabling aggregation and cross comparison of data subsets and metrics from multiple sources.

felix will base on standard and reusable technologies and components making it portable and allowing it to cross platforms. The client side front-end will run on any navigator and a large range of display devices (computer screen, tablet, smart phone).

felix will be demonstrated via operating 2 instances concurrently for 6 months: one will be at

IFREMER, operating on sea surface temperature and ocean waves datasets, and the other at PML based on ocean colour.

## 1.4 Summary

This paper gives an overview of the system and its potential use.

# 2. SYSTEM DESIGN

## 2.1 Requirements

**State-of-the-art:** Build on, and learn from, previous HR-DDS system pilot projects (GHRSS, GLOBWAVE, GLOBCOLOUR, etc.).

**Sustainability:** Developing for the sustained long-term – operation and maintenance.

**Universally applicable:** The HR-DDS system concept shall apply to any types of quantities (sea surface temperature, ocean colour, etc.).

**Reliable, quality, tool:** Need to support cal/val activities of upcoming missions (Sentinel)

**Prevent Redundancy:** Avoid multiple system implementation for such activities.

**High precision, high definition:** Concentrate on the small subset extraction principle, and do not duplicate existing global frameworks.

**Flexible:** It shall be easy for users to deploy their own HRDDS system instance (configure it on their local datasets). It shall be possible and easy to tailor the system to each user need - parameter, datasets, metrics, front-end.

It shall be easy for users to build their own applications querying from existing Felyx instances (documented query API .t)

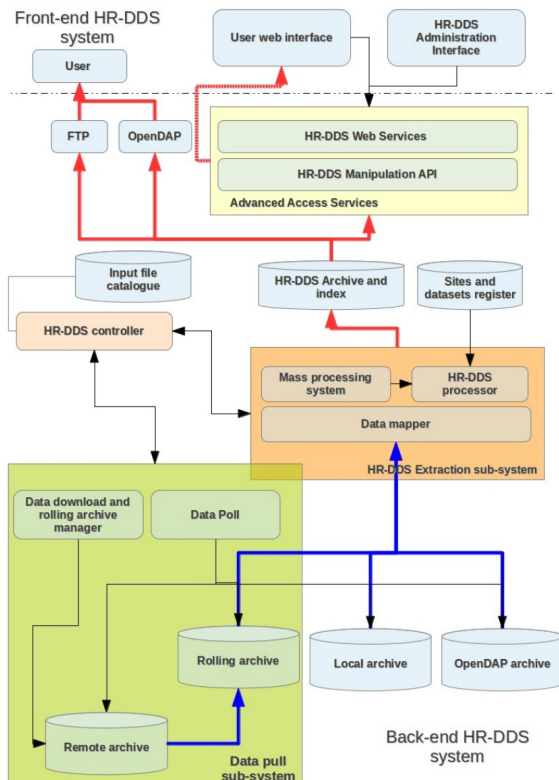
The various instances will be able to interact with each other, creating a web of felix systems enabling aggregation and cross comparison of miniProds and metrics from multiple sources.

**Community driven:** free and open source code. GPL v3 licensing (to be determined).

**Base code:** python and javascript – but not restrictive.

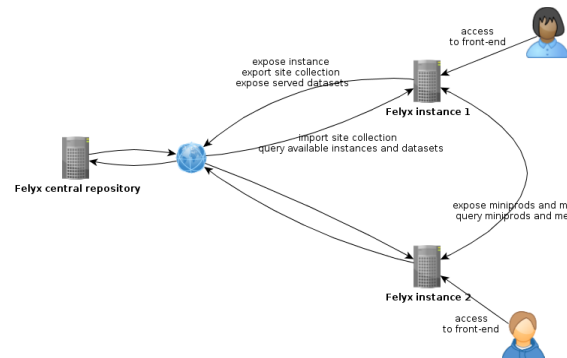
## 2.2 Engine

The following two illustrations show the design of the core HR-DDS system, and the user front-end.



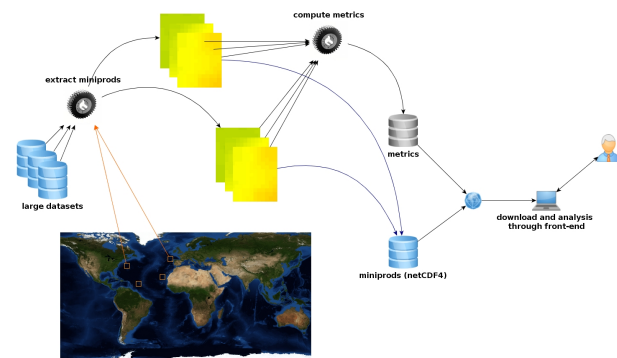
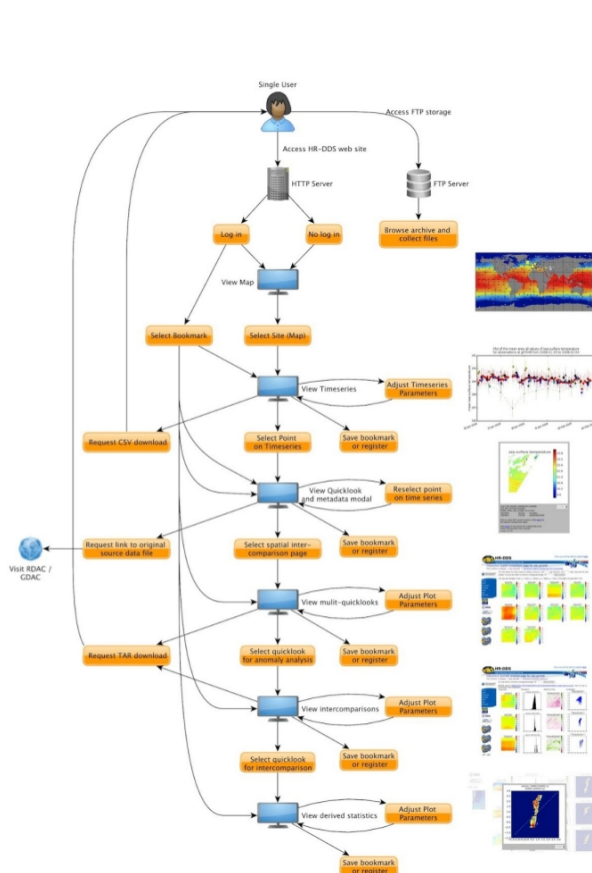
## 2.3 Connected, Distributed, Systems

The system will consist of a series of felix central repositories, sharing data, and a set of interconnected user oriented felix instances. The front-end machines will be able to communicate with the central repositories and with each other.

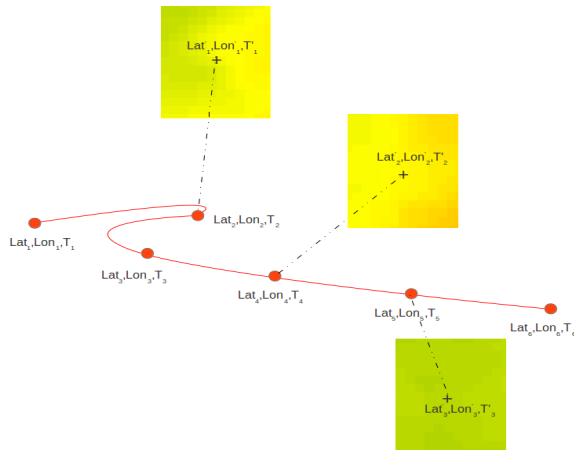


## 2.4 Data Flow and Processing

Input data can be of any form: satellite data in point source, swath or gridded format; model output; ancillary field information; climatologies, covering the atmosphere, ocean, and land.



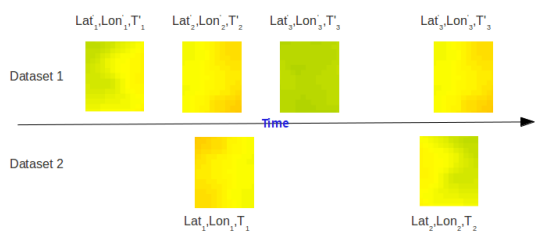
**Static Miniproducts:** Data processing is based on the idea of 'miniproducts' which are local areas or sites. Data from the core datasets are then extracted, processed, and aligned in these miniproducts. Static miniproducts are predefined areas of interest, allowing, for example, long-term trends to be examined.



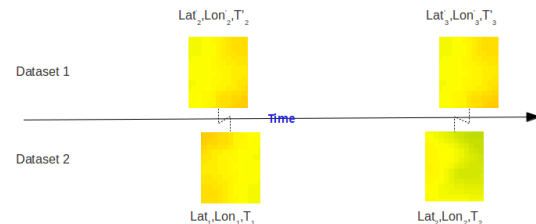
**Dynamic Miniprods:** The miniprods can be dynamic, allowing for the tracking of objects, phenomena or events, such as buoys, ships, hurricanes. In these cases, data is extracted based on moving space and time windows allowing data to be 'matched-up'.

**Multiple datasets:** are extracted over the miniprods. Additional filtering provides a multisensor matchup capability selecting miniprods close to each other. Rules of collocation take into account work done by CDR-TAG and CCI-SST. MMDB behaviour is through a special kind of query.

#### 1. HR-DDS Simple Access



#### 2. MMDB Access



**Metrics:** can be custom made by the user and are extensible. There are really too numerous to mention here completely but a core set includes:

Statistical operators (mean, median, rmse, etc.).

Bias indicators.

Cloud coverage.

Day/night tagging.

....

### 3. USER CASE STUDIES

#### 3.1 Case Studies

Case studies are being developed during the development of the system to test its functionality and performance. These will be

developed into demonstration activities to illustrate system capabilities and educate end-users. Several case studies are envisaged, a selected few follow.

#### 3.1 Case 1 : Data analysis and intercomparison, algorithm development

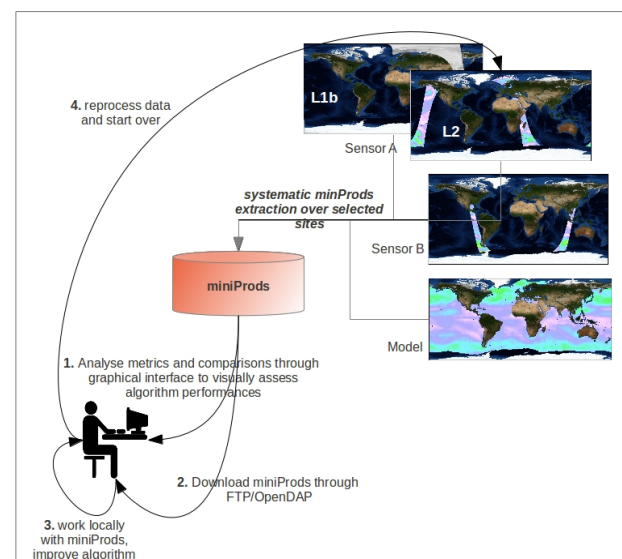
##### Application:

The use of felyx tools in:

- extracting data subsets (miniProds) from input (L1) and inversed (L2) data over predefined sites in a homogeneous format.
- pulling the extracted subsets from FTP or OpenDAP
- computing and displaying metrics (graphics or values) from the extracted miniProds (mean, median, standard deviation, rmse, ...)
- comparing through a web interface the metrics from various sources (similar products, model outputs, in-situ data) over different predefined sites :

- time series
- scatter plots
- histograms

##### Scenario



In this scenario, the felyx system can be used to extract minProds for several datasets (including L1b and one or more derived L2 products, other products for comparison) over a selection of geographical sites. These miniProds are stored in the system, together with metrics computed from their content, and can be queried by the user:

> either through the web interface for visual inspection, display of statistical plots (see example in outputs and next scenarii)

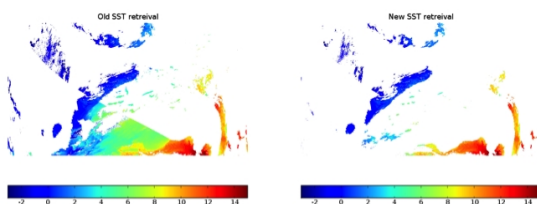
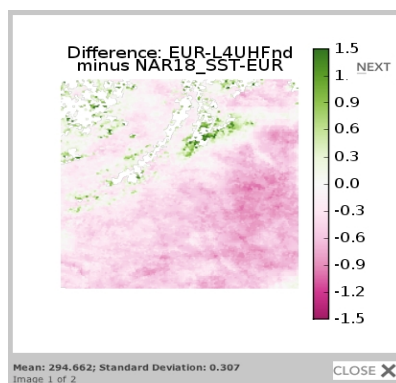
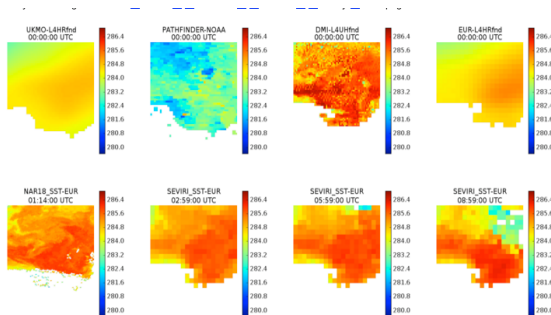
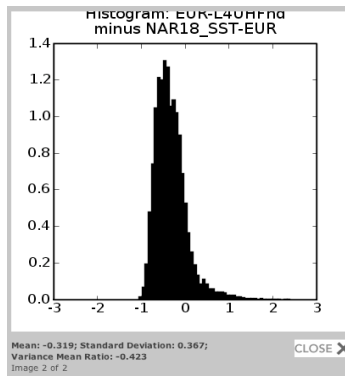
> or through FTP (in NetCDF4 format)/OpenDAP download : they can then be analysed or processed locally (for instance to test again various algorithm versions over these data subsets instead of complete orbit files)

This approach allows quick comparisons to be made of processing algorithms (L1, L2); the inter-comparison of parameter performance; and the detection and analysis of processing issues. Following these analyses, some datasets may be reprocessed and reanalysed in a feedback fashion allowing algorithms to be developed and parameters to be finely tuned.



## Output:

- miniProds, in netCDF4 format, through FTP and OpenDAP
- metrics and maps of miniProds, displayed through an interactive user friendly interface. Example of
- possible plots are provided hereafter. Please describe what additional plots or maps you would like to be displayed.



## 3.1 Use case 2 : Alerts and Alarms

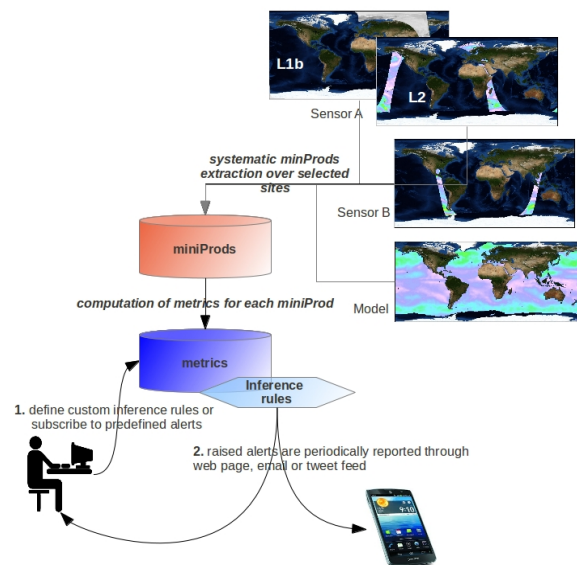
### Application:

Illustrate the use of felix tools to:

- compute metrics over each extracted miniProd (mean, standard deviation, anomaly,...).
- define rules for issuing alerts through a combination of operators over metrics (threshold, range over a selection of parameters, logical association of conditions,...).
- broadcast information automatically.

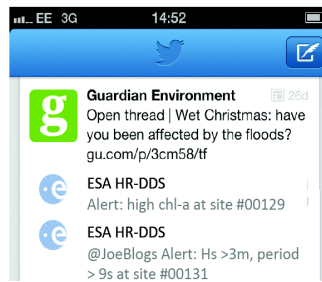
### Scenario:

In this scenario, the HRDDS system can be used to extract minProds for several datasets (waves, sst, ...) over a selection of geographical sites. Metrics are computed from their content. Users or system administrators can define inference rule to trigger alerts base on the range of the parameter(s) value over specific sites. and can be queried by the user. These inference rules are periodically run over the latest metrics. The alerts raised when applying these inference rules are broadcast to subscribed users.

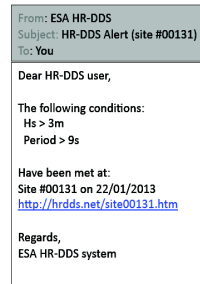


### Output

- Automatic detection of anomalies based on simple, user defined, indicators.
- Broadcast alarms and warnings as specified.



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



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

### 3.2 Case 3: Collaborative Science

#### Application:

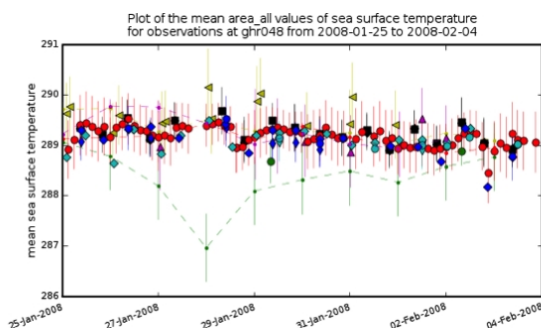
Illustrate the use of felix tools showing:

The social and collaborative features of the analysis.  
bookmarking of user commented analysis cases

#### Scenario:

In this scenario a particular observation appears to be erroneous. Using the social aspects of felix service, this analysis can be shared with other researchers using simple email services, social media such as Facebook or Twitter; or can be bookmarked and labeled for later analysis. Bookmarked cases can gather several plots selected as relevant by the user and rearranged within a single page that can be restored at any time.

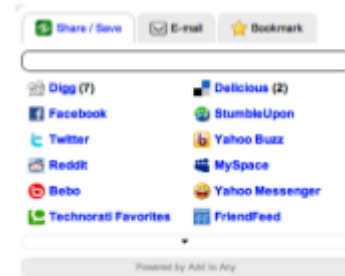
Users may categorize a particular view on the database as “<hot/cold> bias observed in <product\_name> at site <site\_name>”. Such entries could be searched to provide a list of already studied events, assisting new users and researchers.



Title	Description	Site code	Delete
<a href="#">Time series example of gross METOP errors</a>	There are extremely large deviations in SST over the course of 12 hours, up to 10 K	ghr020	<a href="#">Delete this MyDataSet</a>
<a href="#">Example very cold METOP</a>	METOP 6AM is 10K colder than METOP 6PM. AVHRR17 and AVHRR18 agree with the warmer METOP	ghr020	<a href="#">Delete this MyDataSet</a>
<a href="#">Very cold METOP</a>	Look at the last 4 histograms. They show the distribution of SST from the warmer METOP.	ghr020	<a href="#">Delete this MyDataSet</a>
<a href="#">High Coverage Morning 6AM</a>	Really good example	ghr020	<a href="#">Delete this MyDataSet</a>
<a href="#">Atlantic day night issue</a>	Same as pacific, but opposite times	ghr063	<a href="#">Delete this MyDataSet</a>
<a href="#">Atlantic example evening cooling</a>	This one is made spurious by gridding issues, but still provides some insight.	ghr063	<a href="#">Delete this MyDataSet</a>
<a href="#">Example north of Russia</a>	Cold evening METOP observations, not as pronounced by this example has NAR, NAVO and AATSR obs.	ghr066	<a href="#">Delete this MyDataSet</a>
<a href="#">Example o everything working just right</a>		ghr121	<a href="#">Delete this MyDataSet</a>
<a href="#">This one is just plain weird!</a>	Might already be marked, not sure.	ghr020	<a href="#">Delete this MyDataSet</a>
<a href="#">AATSR cold compared to METOP</a>	AATSR one degree colder compared to METOP in Antarctic, and swath times only 10 minutes apart.	nmi070	<a href="#">Delete this MyDataSet</a>

#### Output:

Users could request that they are sent all bookmarked examples of a particular phenomenon. For example the provider of the ocean colour product 'X' could be sent an email whenever a user bookmarks an example of error in that product.



### 3.3 Case 4: long-term data analysis, climate analysis

#### Application:

Illustrate the use of felix tools to:

- define custom sites of interest (such as islands, estuaries,...), over a geographical square box or any polygonal shape.
- extract data subsets (miniProds) from long data collections over predefined sites in a homogeneous format.
- compute metrics from each miniProd (mean, median, rms, ...)
- detect and highlight values outside a given range or threshold (e.g. anomalies)
- query and download metrics or miniProd time series for further analysis (trend calculation,...)

This case showss the use of felix for surveys and climate analysis and highlights the trend estimation, cross-correlation, and consistency capabilities.

#### Scenario:



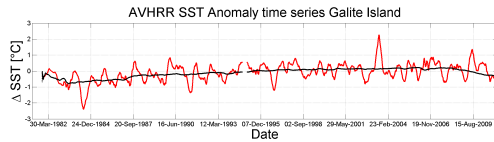
For example, a study on the small Mediterranean islands to assess the climate change in the Mediterranean basin, using a synergy of various parameters (SST, winds, SAR-derived radial current,...). Each island could be defined as a custom site

The user would start by loading the default set of sites defined by the system. Alternatively, with the proper permissions (granted by the felix administrator), the user could define their own custom set of sites. These sites can be defined by:

- uploading a CSV-type file in a comprehensible format for felix.
- defining the sites (square box or polygon) interactively on a map with high-resolution zoom capability.

Custom sites defined by users implies the development of new miniprods so the user has the capability to process the associated miniProd and extract metrics for these sites, for a selection of datasets and parameters of their choice.

Results can then be analysed through time series for instance :



Climatologies can be processed as any dataset by felyx system, making possible to compare any dataset to a reference.

#### Output:

As in the previous use cases, the miniProds and metrics can be visualised as graphics. These graphics can be saved or emailed individually (in SVG or PNG format for a publication quality plot) or together as reportpage. They can also be download by the user in netCDF4 (miniProds and metrics) or CSV (metrics only) format for further analysis (trend calculation, EOF, ...) with users own tools (matlab, IDL, etc...).

## 4. OUTPUTS

### 4.1 System Outputs:

Transfer: FTP, OpenDap, Web query (tar file, values)  
Format: Netcdf/csv/json  
Plots: Displayed, saved, emailed, and exported in reports.  
Reports: Automatic generation for period/latest data.  
Alerts: SMS etc.