

TC_2 EuroArgo Data Subscription Service

1. Background

1.1 Short description

The objective is to provide a data subscription service to scientific users. The scientific user provides his data selection criteria. Regularly, a dataset of his selected data will be extracted from Research Infrastructures cloud and delivered in his personal cloud account.

1.2 Contact

Background	Contact Person	Organization	Contact email
RI-ICT (Use Case proposer, Agile Group leader)	Thierry Carval	Ifremer, Euro-Argo	Thierry.Carval@ifremer.fr
RI-ICT	Robert Huber	University of Bremen, EMSO	rhuber@uni-bremen.de
RI-ICT	Benjamin Pfeil	University of Bergen, ICOS	benjamin.pfeil@gfi.uib.no
e-Infrastructure	Yin Chen	EGI	yin.chen@egi.eu
e-Infrastructure	Leonardo Candela	CNR	leonardo.candela@isti.cnr.it
e-Infrastructure	Gergely Sipos	EGI	gergely.sipos@egi.eu
e-Infrastructure	Daan Broeder	EUdat	daan.broeder@meertens.knaw.nl
RI-ICT	Jerome Detoc	Ifremer, Euro-Argo	Jerome.Detoc@ifremer.fr
RI-ICT	Antoine Queric	Ifremer, Euro-Argo	Antoine.Queric@ifremer.fr
Task 5.3/7.2 leader	Zhiming Zhao	University of Amsterdam (UvA)	z.zhao@uva.nl
	Paul Martin		P.W.Martin@uva.nl

1.3 Use case type

This use case is an implementation case

Euro-Argo, EMSO and ICOS Research Infrastructures will push a series of data and metadata on a so-called ENVRIPLUS cloud.

A data subscription service will be developed for scientific users.

Data provided for the cloud service:

- All Argo observations, daily updated
- Copernicus in situ observations
- A selection of EMSO observatories data
- ICOS-SOCAT carbon data observed from voluntary observation ships

1.4 Scientific domain and communities

Scientific domain

Atmosphere, hydrosphere, *geosphere*?

Community

Data service provision, data usage.

Behavior

2. Detailed description

Objective and Impact

The objective is to provide a regular data flow to scientists, from different Research Infrastructures.

Challenges

- Provide data and metadata from complementary Research Infrastructure (Ocean, Atmosphere, space)
- Aggregate and distribute billions of observations
- Link between RI and E-infrastructures

Detailed scenario

The data subscription service to scientific users:

- The user provides his criteria
 - time, spatial, parameter, data type
 - update period for delivery (daily, monthly, yearly, on the spot)
- The relevant data are extracted on ENVRI cloud
- Data may be converted/transformed on ENVRI grid
- The user's cloud account is updated regularly with the new data provided above
- An accounting of data delivery is performed (MDC ?)
 - A citation scheme is attached to the delivered data (DOI)
 - bibliographic surveys can track the use of these data in publications
 - reproducibility is possible
- A users identification scheme is implemented (MarineID, OpenID, Shibboleth ?)

The cloud content:

- Euro-Argo and Copernicus datasets
 - 4 billion ocean observations
 - 300 parameters
 - 15 000 observing platforms
 - from 1900 to today.
- A selection of EMSO observatories data
- ICOS-SOCAT carbon data observed from voluntary observation ships

This "cloud" of observations is pushed and continuously updated on ENVRIPLUS cloud (EGI, EUDAT ...)
Copies are replicated in different places, close to users location (EU, US, Australia, Japan)

The cloud data model

- Observation data model : a flat table of 4 billion records
- ID DOI platformCode dataType x y z t parameter value dateUpdate
- The observations table is hosted in a workplace such as
 - NoSQL Elasticsearch
 - The use of in-memory features would provide the best reactivity (instant answers)
- The workplace is activated in a virtual server, replicated on the cloud

Data services to be developed around Euro-Argo cloud

- Ocean observations API : scoop
- Metadata visualisation : map wms services
- Data visualisation : graphics
- Data products such as mixed-layer depths maps

Agile and incremental implementation

- Step 1 : Euro-Argo data file on the cloud (1to file daily updated)
- Step 2 : VM for indexation of data file (Elasticsearch)
- Step 3 : data file generation service (scoop Java API)
- Step 4 : data subscription/distribution service to OwnCloud accounts
- Step 5 : replicate data and VM in mirror sites (EU, US, AU, JP)
- Next steps : promote the development of various services around an ENVRIPLUS cloud

Technical status and requirements

This use case involves EGI and EUDat

Implementation plan and timetable

Timetable

- Step 1: 2016 Q1 Euro-Argo RI, EGI, EUDAT
- Step 2: 2016 Q2 Euro-Argo RI, EGI, EUDAT
- Step 3: 2016 Q3 Euro-Argo RI, EGI, EUDAT, EMSO RI, ECOS-SOCAT RI
- Step 4: 2016 Q4 Euro-Argo RI, EGI, EUDAT, EMSO RI, ECOS-SOCAT RI
- Step 5 : 2017 Euro-Argo RI, EGI, EUDAT, EMSO RI, ECOS-SOCAT RI

Allocation of resources

- Euro-Argo RI: 10 man months, 20 000€ (2 months of subcontracting)
- EGI
- EUDAR
- EMSO RI
- ECOS-SOCAT RI

Expected output and evaluation of output

In 2017, the first data subscribers and the number of downloaded files will be counted.
The experience will be a success if regular users receive the data files they need.

External Links:

1. TC_2 notebook: <https://envriplus.manageprojects.com/projects/wp9-service-validation-and-deployment-1/notebooks/630/pages/331+>