

WP5 - DATA INTEGRATION AND MANAGEMENT

iAOS, data integration, Results and impact

INTAROS Final Meeting - Synthesis

January 20th-21st 2022

Hervé Caumont, Terradue Srl



Agenda

1. Objective and Tasks
2. Main Achievements
3. Expected Impact
4. Challenges
5. Recommendations

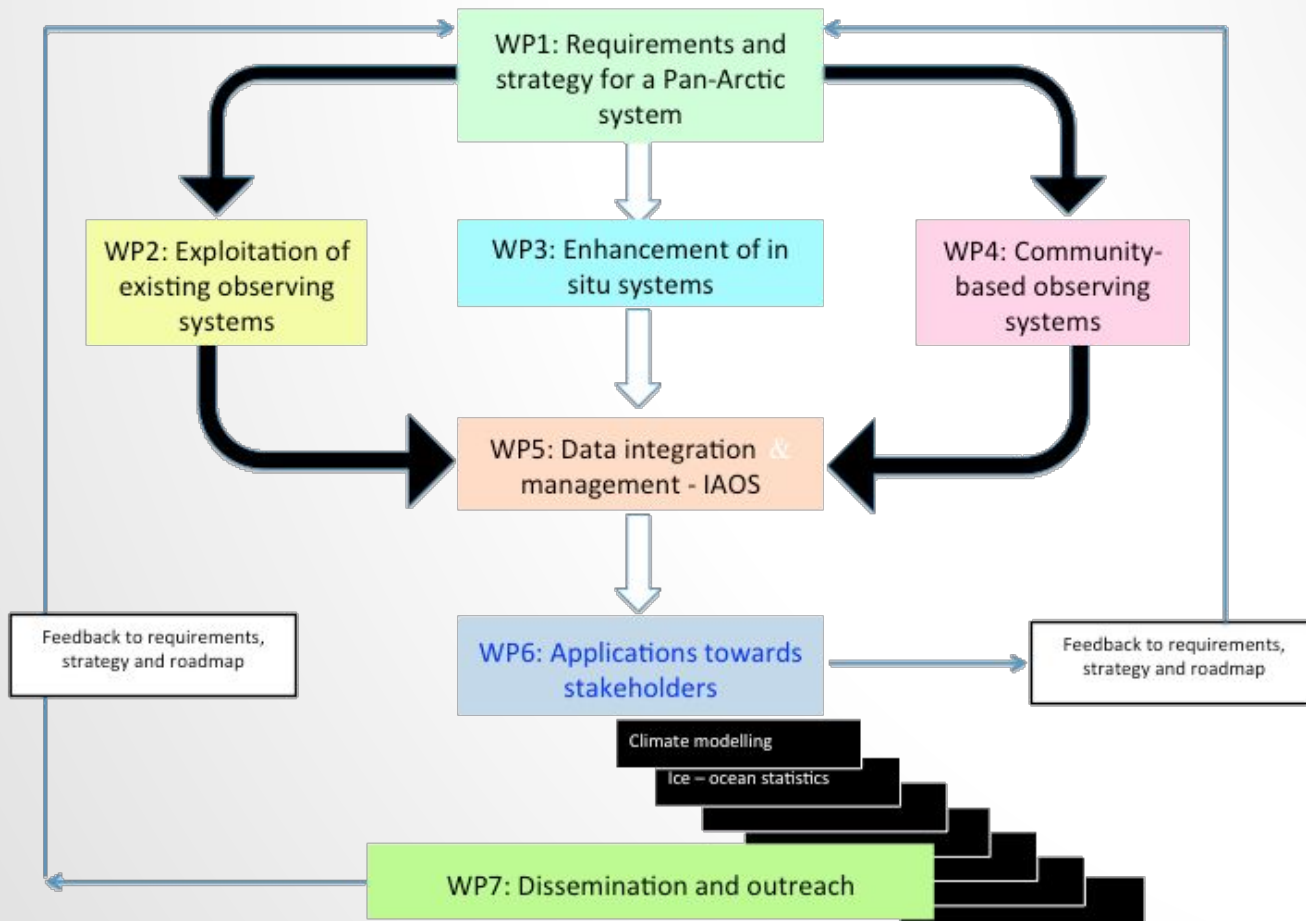
↳ **INTAROS Synthesis Report**
4 pages for WP5





1. WP5 Objective and Tasks





WP5 Objectives

- ❑ **Integrate multidisciplinary and distributed data repositories** into a scalable and resilient Pan-Arctic observing system (iAOS), which will offer seamless access to observations and derived parameters.
- ❑ iAOS will also provide a set of **tools for data analysis, transformation and visualization**.
- ❑ **Develop new geo-statistical methods** for interpolation of spatiotemporal datasets.
- ❑ **Process new observations** from WP2-4, and **store generated datasets** in an iAOS enabled repository.

WP5 Tasks

- **Task 5.0** Workpackage Coordination [**TERRADUE** & **NERSC**]
- **Task 5.1** System requirements and architecture [**TERRADUE**, **NERSC**, **AWI**]
- **Task 5.2** iAOS platform deployment and operation [**TERRADUE**]
- **Task 5.3** Integrate data from repositories [**AWI** & **TERRADUE**, **NERSC**, **FMI**, **UIB**, **GEUS**, **IMR**, **IFREMER**, **NUIM**, **NORDECO**, **CNRS-LOCEAN**]
- **Task 5.4** Geostatistical methods for data integration [**ARMINES**, **NERSC**, **DTU**]
- **Task 5.5** Integration of new processing services [**TERRADUE**, **NERSC**, **ARMINES**]
- **Task 5.6** iAOS portal development [**NERSC**, **TERRADUE**]
- **Task 5.7** Synthesis of iAOS infrastructure deployment and operation [**TERRADUE**, all partners]





2. WP5 Main Achievements



Deliverables

All WP5 deliverables submitted to EC

- D5.1 iAOS requirements and architectural design V1
- D5.2 iAOS platform and tools V1
- D5.3 Data integration from existing repositories V1
- D5.4 iAOS Portal with user manual V1
- D5.5 iAOS requirements and architectural design **V2** - revision 1.3
- D5.6 Geostatistical library V1 - revision 1.8
- D5.7 Processing services integration V1 - revision 1.1

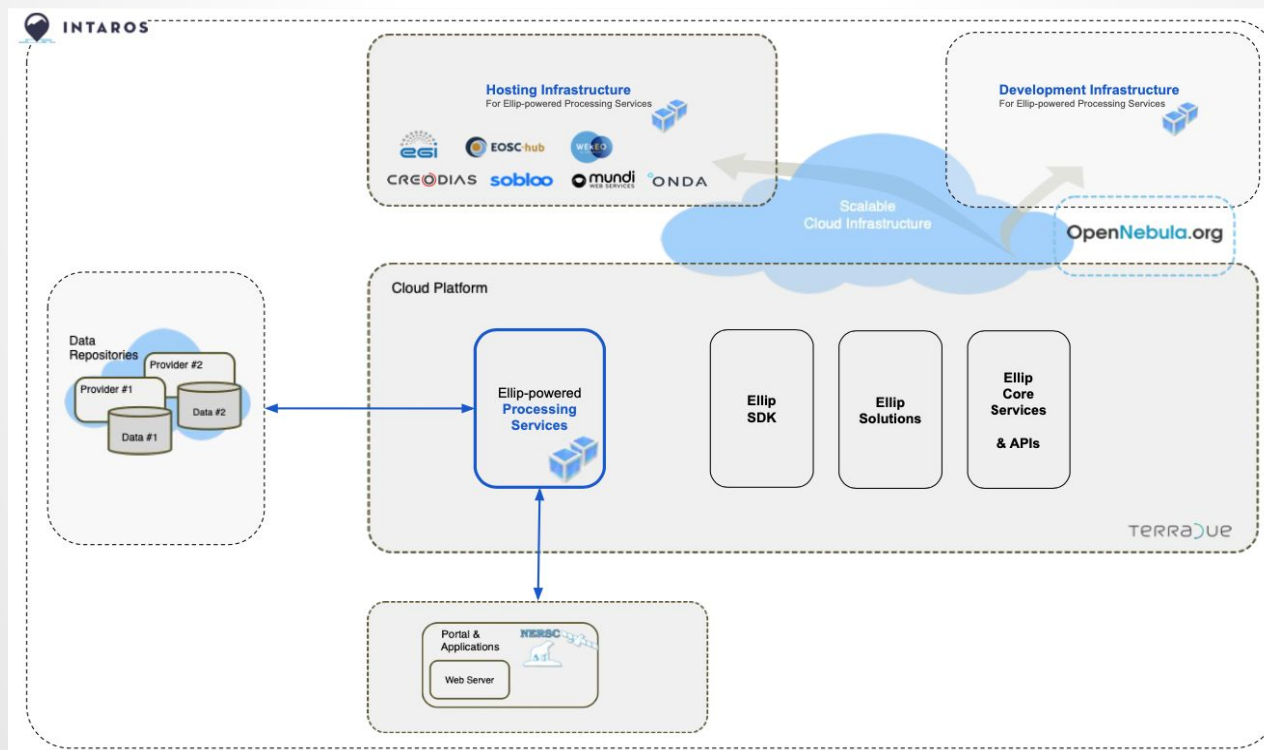
- D5.8 iAOS Platform and tools **V2** - revision 1.4
- D5.9 Data integration from existing repos **V2**
- D5.10 Geostatistical library **V2**
- D5.11 Processing services integration **V2**
- D5.12 iAOS Portal with user manual **V2**

- D5.13 Synthesis of the iAOS infrastructure



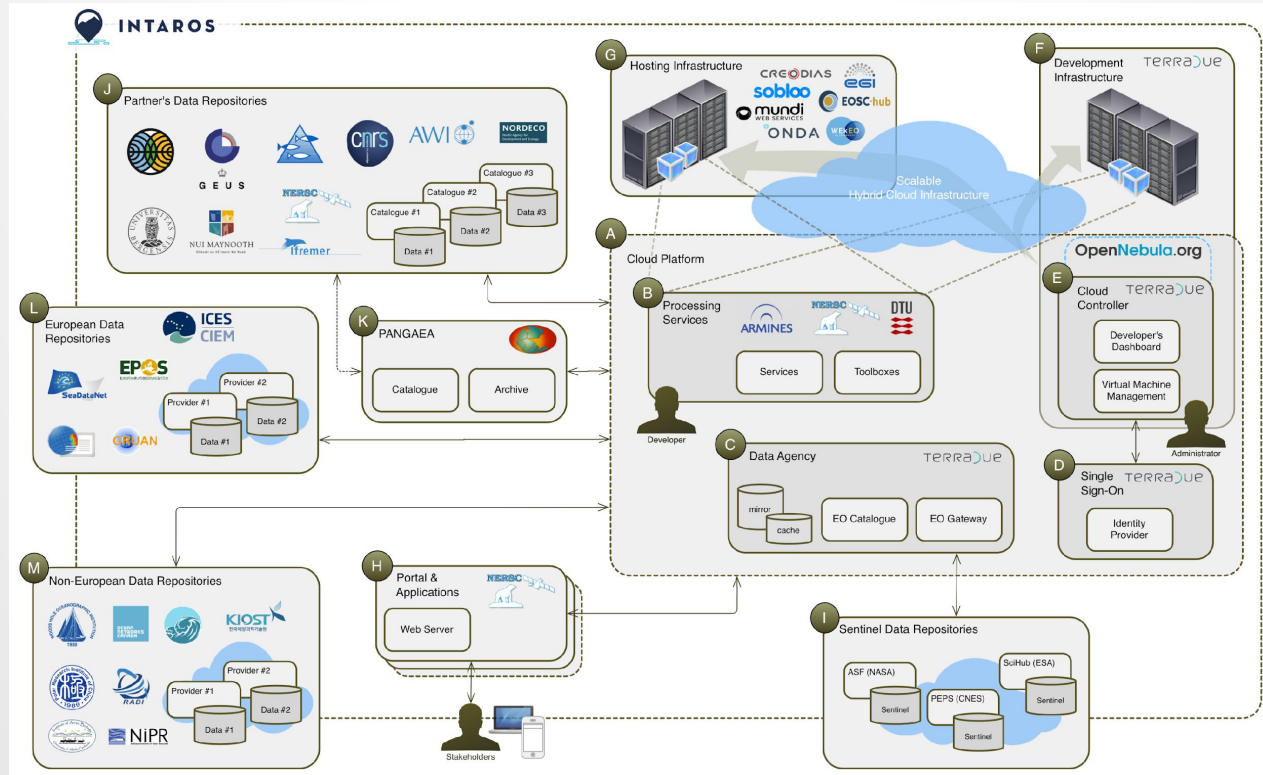
Synoptic view of the Cloud platform, linked to other components of the iAOS (e.g. data repositories, iAOS portal), and to development and deployment (hosting) infrastructures.

- As part of the INTAROS project, Terradue operates the **iAOS Cloud Platform**, a set of tools and services for online data access and data processing.



Architecture of the Information and Communication Technologies (ICT) components of the overall iAOS

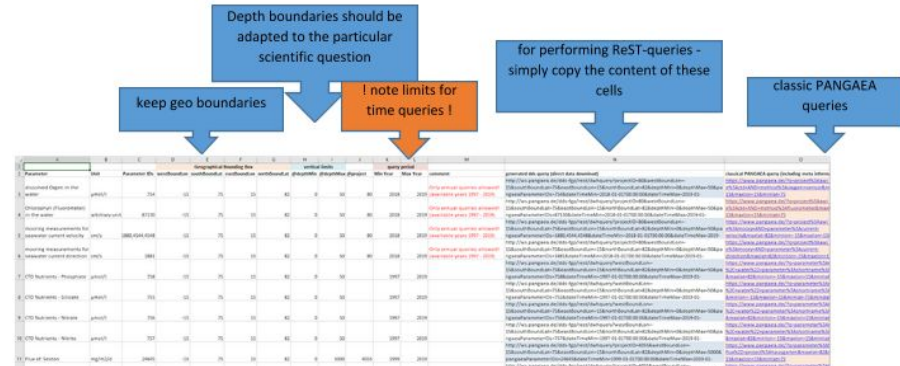
- One key achievement of INTAROS is to have extended, improved and unified existing observing systems in the different regions of the Arctic, as part of an integrated Arctic Observation System (iAOS)



Data Integration on PANGAEA

- Updates on AWI/PANGAEA/FRAM data-products
- Delivery of PANGAEA API for data-mining functionalities (data-warehouse), allowing to combine parameter records from different PANGAEA datasets in one file

cf. D5.9 Data integration from existing repos V2 (May 2021)



Base URL for all ReST-queries is <http://ws.pangaea.de/dds-fgp/rest/dwhquery?> followed by **mandatory** and **optional** search parameters.

Service call example:

<http://ws.pangaea.de/dds-fgp/rest/dwhquery?westBoundLon=-15&southBoundLat=75&eastBoundLon=15&northBoundLat=82&pangaeaParameterIDs=717&projectId=80&depthMin=0&depthMax=50&dateTimeMin=2018-01-01T00:00:00&dateTimeMax=2019-01-01T00:00:00>



OPeNDAP solutions for the iAOS

Integration of distributed data repositories

OPeNDAP Dataset Access Form

Action: Get ASCII Get Binary Show Help

Data URL: http://opendap1.nodc.no/terradue/phys/point/yearly/SSM_T70_2003.nc

Global Attributes:

```

version: 00000001: 1.2
platform: ender: SSM
platform_name: SSM
data_update: 2017-03-07T01:01:00Z
description: SSM/SSM-T70 Marine Research
  
```

Variables:

TIME: Array of 64 bit floats (TIME = 8.882)

```

TIME:
  platform: 999999.7
  data_type: float
  resolution: 1 km
  sector: days since 1850-01-01T00:00:00Z
  units: days
  
```

TIME_QC: Array of 8 bit bytes (TIME_QC = 8.882)

TIME_QC: False

OPeNDAP Contents of /data/

Name	Last Modified	Size	DAP Response	Links	Dataset Viewers
time2/	2020-07-13T13:11:40GMT	-	-	-	-
yrld_coordinates/	2020-07-13T13:14:55GMT	-	-	-	-
ase/	2020-07-13T13:17:32GMT	-	-	-	-
sample/	2020-07-13T13:18:01GMT	-	-	-	-
ssm_smap/	2020-07-13T13:18:37GMT	-	-	-	-
ts/	2020-07-13T13:18:37GMT	-	-	-	-

Open DAP Server

timeCOUT.nc

Information

Catalog: http://opendap1.nodc.no/terradue/phys/point/yearly/SSM_T70_2003.nc

FeatureType: Grid

Dataset: 6134675

URI: http://opendap1.nodc.no/terradue/phys/point/yearly/SSM_T70_2003.nc

Dates

Modified: 2020-05-27T08:43:12.280Z

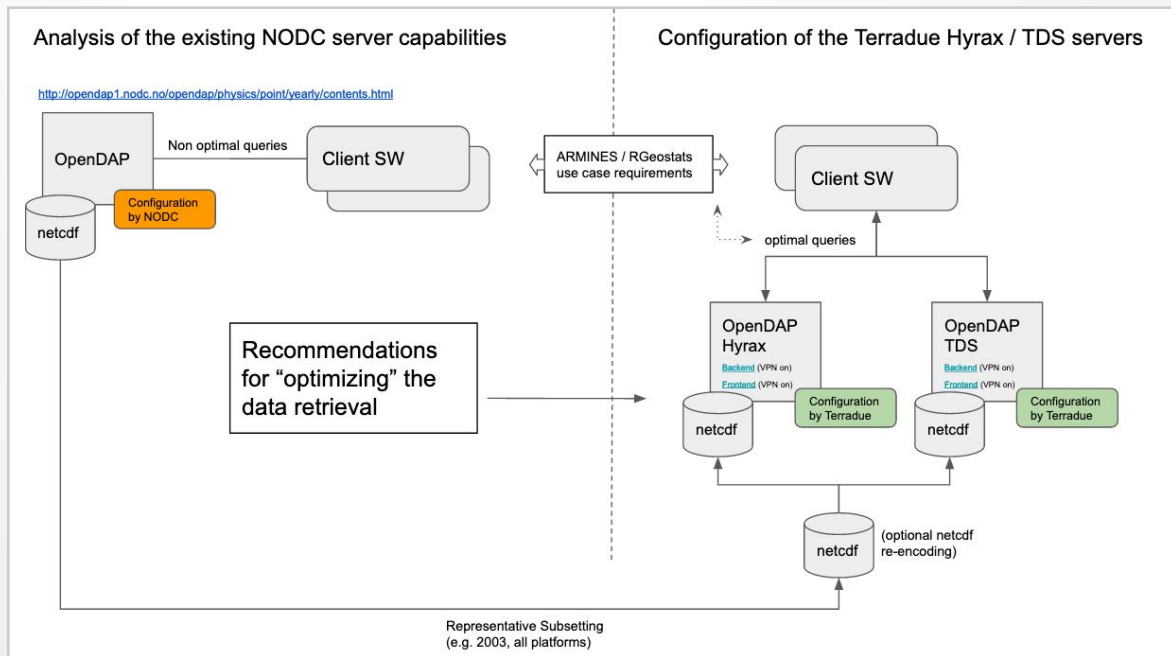
Access/Preview

OpenDAP

Type: Data Access

Description: Access dataset through OPeNDAP using the DAP2 protocol.

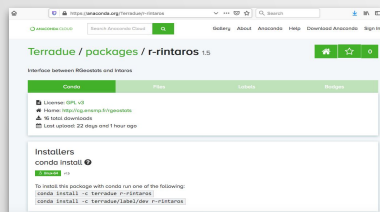
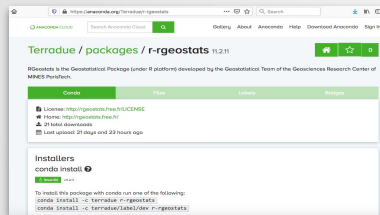
Supplier Notebook viewer



INTAROS

Geostatistics tools for the iAOS

Tools for data analysis, transformation and visualization



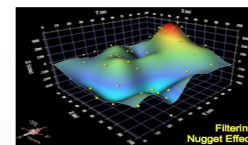
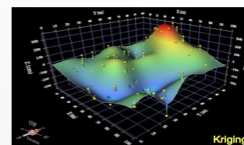
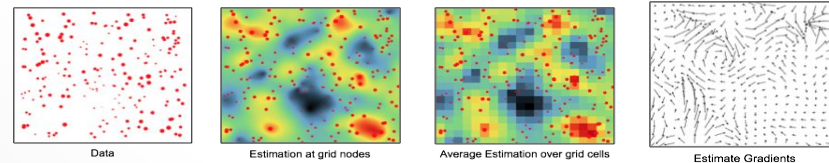
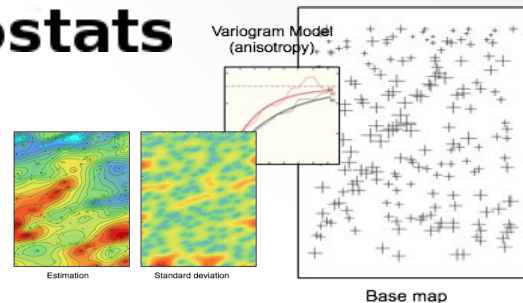
- Deployment of RGeostats package with anaconda for iAOS developers:

<https://anaconda.org/Terradue/r-rgeostats>

- Creation and deployment of RIntaros Geostatistical package for iAOS developers:

<https://anaconda.org/Terradue/r-rintaros>

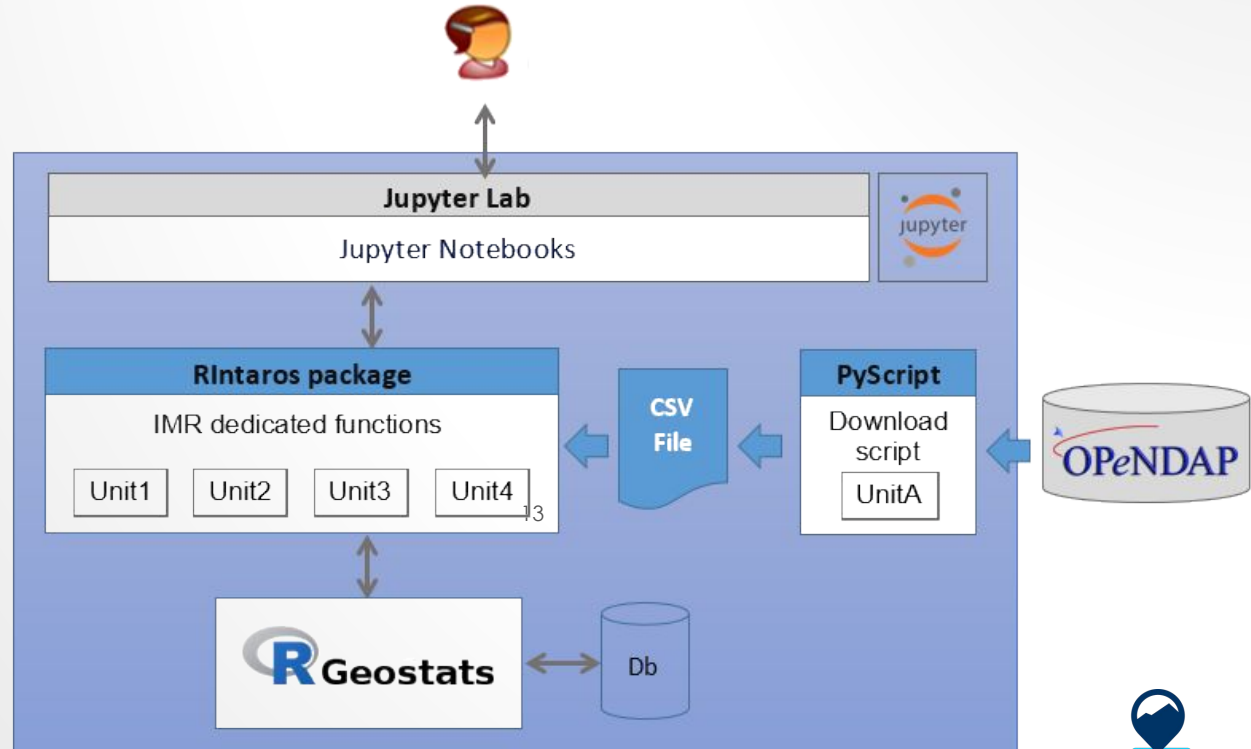
- Dissemination material produced to outreach the iAOS users community (Bremen Workshop, Terradue Seminar at Fontainebleau XIVème Journées de Géostatistique)



Geostatistics tools for the iAOS

Tools for data analysis, transformation and visualization

- Python scripts for downloading CTD data
- Unitary R scripts using RIntaros package for Geostatistical operations
- Jupyter Notebooks for download and ingestion of CTD data as part of Geostatistical models



Processing Services

Tools for data analysis, transformation and visualization

Data processing services integrated and ran by NERSC from collaboration with the EC H2020 project NextGEOSS

Run of complex, compute-intensive EO data processing chains (processing of copernicus Sentinel-1 observations) in order to generate value-added products supporting the INTAROS tasks

- Sea ice classification service
- Sea ice drift service

The services made use of Cloud Computing resources funded by the EC NextGEOSS project on the **EGI.eu Federated Cloud**, delivering data products for exploitation as part of INTAROS tasks.

For the **sea ice classification service**, two processing campaigns covering:

- a three week period in July-August 2018, coinciding with the INTAROS 2018 field experiment in the Fram Strait and north of Svalbard.
- a three week period in August-Sept. 2019, coinciding with the CAATEX/INTAROS 2019 field experiment in the Fram Strait – Eurasian Basin.

In total, over 500 Sentinel-1 SAR scenes were classified during these two campaigns.

For the **sea ice drift service**, two processing campaigns have been ran, for the same time periods as for the sea ice classification service:

- for the first period (2018), over 1500 pairs of Sentinel-1 images were processed,
- for the second period (2019) over 2000 pairs were processed.

The estimated ice drift vectors are grouped into daily datasets for up to 1-day, up to 2-day and up to 3-day time difference between the images in the pair. The ice drift vectors have been published back to the data server at NERSC for exploitation as part of INTAROS.



WP5 / WP6 coordination for the definition of iAOS Showcases

“WP6 will integrate **remote sensing data** and **in situ observations delivered through WP5**, from a variety of platforms and geographical scales and locations. *Incorporation of these data into analysis and modelling systems, including physical and **ecological process models, climate models and forecast methods**, will provide support for better products to key societal areas.*”

-- from WP6 Description of work

Helsinki, 2018 - Joint WP5-6 Workshop
(but WP6 starting later)

Bremen, 2019 - RGeostats Workshop

Sopot, 2020 - Interviews with WP6 task leaders for their work plan analysis, and identification of best 'showcase' opportunities to be supported by the iAOS (WP5)

- As return of experience (INTAROS internal) on how the WPs interact in order to illustrate the iAOS added value
- As a set of results-oriented data collections and services, which can support the INTAROS outreach activities in 2020-2021

Remote, 2020 - Intermediate results reviews and definition of final objectives for each Showcase



INTAROS


Contribution to the INTAROS Booklet

Geo-statistical methods


Application of RGeostats on oceanographical data


ARMINES and Terradue


- Analysis of oceanographical data (CTD)
- Presentation of spatial and temporal correlations
- Mapping multiple variables
- Combining different data sources

**INTAROS**

32

**Armines and Terradue**
Fabien Ors
fabien.ors@armines-terradue.fr



**The objective is to build and deploy a new geostatistical library as a service for scientists analyzing, interpolating and presenting oceanographic data.**



This work demonstrates a case study on:

- Analysis of oceanographical data from an IHR database
- Presentation of spatial and temporal correlations
- Mapping multiple variables
- Combining different data sources.

Powered by RGeostats



Application of RGeostats on oceanographical data

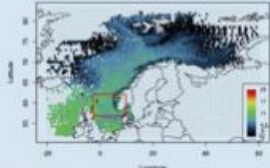



Fig. 11 Example of temperature map from an IHR database.

**Examples of applications of RGeostats:**

- Temperature interpolation map at a given depth and time interval (Figure)
- Estimation of salinity
- Probability of exceeding a given sea ice thickness
- Evolution of fish density in time
- Seasonal plankton concentration

Users include:

- Scientists, companies, NGOs, national and EU agencies working with climatology, meteorology, biology, oceanography, pollution, tourism and environmental management.

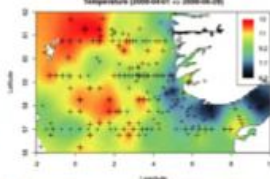






Fig. 12 Example of temperature interpolation map of water temperature at 25 m depth.


**TERRADUE**

**ARMINES**

**PSL**

**INTAROS**

INTAROS has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 727995





INTAROS


Support to iAOS Showcase applications

Other collaborations

With SMHI

Climate Model initialisation

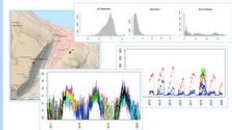
SMHI David Gustafsson David.Gustafsson@smhi.se	Pan-Arctic Hydrological Modelling
Objectives - Improve predictions of spring floods, river ice breakup and freshwater flow to the Arctic Ocean. Support the integration of river discharge into the ocean, providing data at user defined resolution	Climate model initialization Data resources for checking the modelling initial conditions, the model analysis of current conditions and the model forecast


<ul style="list-style-type: none">• Arctic-Hypos Observational data search and download from the iAOS Portal for daily pre-processing tasks:<ul style="list-style-type: none">◦ Archive of quality controlled data with 4months/years lag◦ Provisional datasets• Arctic-HYPE model results published as open data:<ul style="list-style-type: none">◦ Daily analyses of last 60 days◦ Medium range forecast of coming 10 days	 Figure 1. iAOS OpenDAP server (SMHI) for Arctic-HYPE model outputs
---	---

With FMI

Modelisation of spatial and temporal behavior for snow thickness

FMI Roberta Pirazzini Roberta.Pirazzini@fmi.fi	Maps for Svalbard Avalanche Forecast Modelling
Objectives - Derive geostatistical relationships between in-situ measurements and model data of snow and weather conditions to improve the snow accumulation forecast	Modelisation of spatial and temporal behavior for snow thickness Data resources for avalanche forecast modelling built from meteorological parameters and in-situ data recording

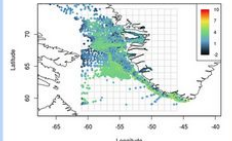
<ul style="list-style-type: none">• Analysis of snow stations discontinuous data records, arome model outputs and terrain model data• Modelisation of snow thickness through co-variables (temperature, wind speed by class of wind direction)• Use the Geostatistical Library (Rintaros / RGeostats) and build the R software for interpolating snow depth maps	 Figure 1. Longyearfjord valley utilized snow stations & automatic weather station. Temperature time series and Snow depth measurements over time.
--	--


Powered by **RGeostats** 

With Aarhus Univ.

Climate Change & Fish resources

Aarhus University Mikael Sejr mse@bios.au.dk	Baffin Bay Sea Bottom Temperature Maps
Objectives - Exploit temperature fields at bottom of the ocean for: <ul style="list-style-type: none">• Analysis of long term global warming influence• Analysis of the fish stock correlation to temperature	Climate Change & Fish resources Data resources for

<ul style="list-style-type: none">• Analysis of CTD, bottle and Trawl datasets (archives over 1960 to 2017)• Modelisation of spatial and temporal behavior for ocean floor temperature through Bathymetry co-variable• Use of the Geostatistical Library (Rintaros / RGeostats) to build the R software application for interpolating ocean floor temperature maps	 Figure 1. Maximum Depth Temperature (Overall)
--	--

Powered by **RGeostats** 



INTAROS

EO Data Discovery

Tools for data analysis, transformation and visualization

Polarstern enters uncharted Arctic waters in what once was thick perennial sea ice

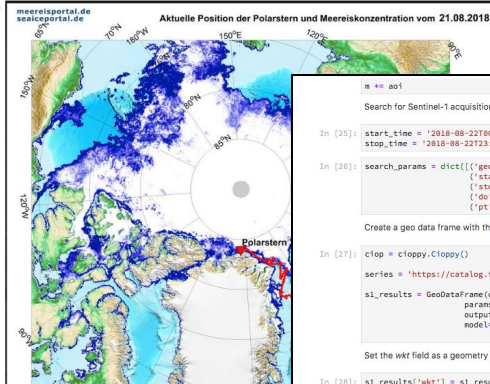
Mark Drinkwater's (@kryosat) tweet on August 22, 2018

<https://twitter.com/kryosat/status/1032172161635639296>

Polarstern enters uncharted Arctic waters in what once was thick perennial sea ice. Follow progress at <https://t.co/BxpVLWjD> @AWI_Media pic
— Mark Drinkwater (@kryosat) August 22, 2018

```
In [1]: from IPython.display import Image
        Image(filename='resources/kryosat_tweet.png')
```

```
Out[1]:
```



```
In [25]:
```

```
start_time = '2018-08-22T09:08:00Z'
stop_time = '2018-08-22T13:59:59Z'
```

```
In [26]: search_params = dict([('geom', 's1_track'),
                              ('start', start_time),
                              ('stop', stop_time),
                              ('iso', 'terradue'),
                              ('pt', 'GEO')])
```

Create a geo data frame with the Sentinel-1 search results (more info about GeoPandas and GeoDataFrames)

```
In [27]: c1op = c1oppy.C1oppy()
        series = 'https://catalog.terradue.com/sentinel1/search'
        s1_results = GeoDataFrame(c1op.search(end_point=series,
        params=search_params,
        output_fields='identifier,self,startdate,track,wkt,enclosure',
        mode='GEO'))
```

Set the wkt field as a geometry (it contains the Sentinel-1 footprints)

```
In [28]: s1_results['wkt'] = s1_results['wkt'].apply(shapecpy.wkt.loads)
```

```
In [29]: s1_results
```

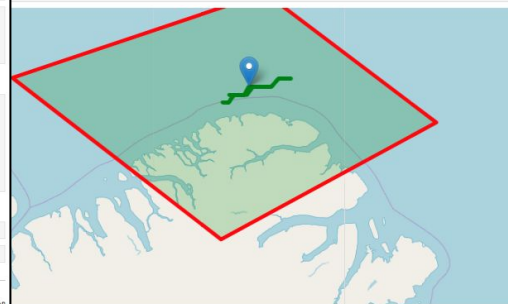
	enclosure	identifier	self	startdate	track
0	https://store.terradue.com/download/sentinel1/...	S1B_EW_GDRM_IDSH_20180822T181527.20180822T1815...	https://catalog.terradue.com/sentinel1/search?...	2018-08-22T18:14:27.738000Z	103 83.608
1	https://store.terradue.com/download/sentinel1/...	S1B_EW_GDRM_IDSH_20180822T163556.20180822T1636...	https://catalog.terradue.com/sentinel1/search?...	2018-08-22T16:35:56.146000Z	102 -1
2	https://store.terradue.com/download/sentinel1/...	S1B_EW_GDRM_IDSH_20180822T145749.20180822T1458...	https://catalog.terradue.com/sentinel1/search?...	2018-08-22T14:57:49.769180Z	101 -6
3	https://store.terradue.com/download/sentinel1/...	S1B_EW_GDRM_IDSH_20180822T131947.20180822T1320...	https://catalog.terradue.com/sentinel1/search?...	2018-08-22T13:19:47.427000Z	100 -75.4
4	https://store.terradue.com/download/sentinel1/...	S1B_EW_GDRM_IDSH_20180822T114138.20180822T1142...	https://catalog.terradue.com/sentinel1/search?...	2018-08-22T11:41:38.660000Z	99 -6

```
In [21]: wkt = to1['geometry'].values[0].wkt
        wkt
```

```
Out[21]: 'POINT (-33.8 84)'
```

Add map with boat position

```
In [22]: global m
        from IPyleaflet import Map, Polygon
        m = Map(center=(to1['geometry'].values[0].centroid.y,
        to1['geometry'].values[0].centroid.x), zoom=5)
        marker = Marker(location=(to1['geometry'].values[0].centroid.y,
        to1['geometry'].values[0].centroid.x), draggable=False)
```



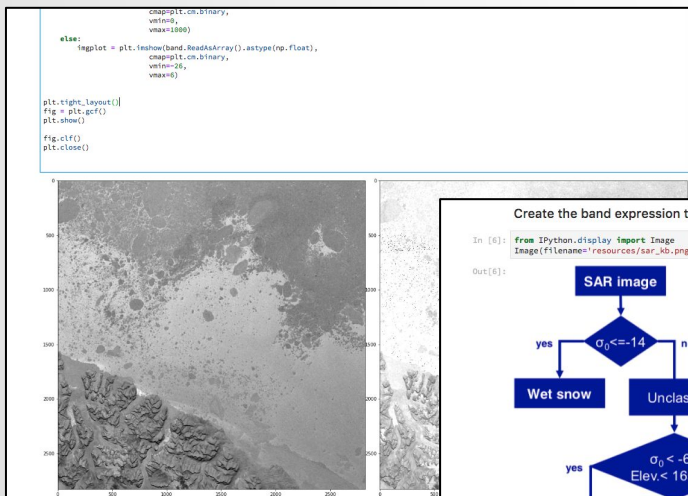
on 2018-08-22 as the geospatial filter for the discovery of Sentinel-1 data



INTAROS

EO Data Access and Processing

Tools for data analysis, transformation and visualization



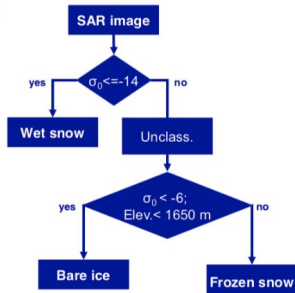
Create the band expression to classify the snow and ice

```

In [6]: from IPython.display import Image
Image(filename='resources/sar_kb.png')

```

Out[6]:



```

In [7]: band_expression = Template('if $den == -500 then 0 else if $sigma0 < -14 then $sigma0 - 1 else $sigma0 - 2 else 0 end if')
Use the SNAP BandMaths Operator to apply the expression for HH and HV sigma0 derived data:

```

```

In [8]: BandDescriptor = jpy.get_type('org.esa.snap.core.gpf.common.BandMathsOp$BandDescriptor')
targetBands = jpy.array('org.esa.snap.core.gpf.common.BandMathsOp$BandDescriptor', 2)
targetBand_HH = BandDescriptor()
targetBand_HH.expression = band_expression.substitute(sigma0=list(band_names)[0], den=list(band_names)[2])
targetBand_HH.name = 'snow_ice_classification_HH'
targetBand_HH.type = 'int8'
targetBands[0] = targetBand_HH

```

Plot the results

```

In [9]: fig = plt.figure(figsize=(20,20))
fig.suptitle('Snow and ice classification', fontsize=18, y=1.05)
for i in [0,1]:
    a=fig.add_subplot(2, 2, 1+i)
    data = snow_and_ice_product.getBand(snow_and_ice_product.getBandNames()[i])

    a.set_title(snow_and_ice_product.getBandNames()[i])
    w = data.getRasterWidth()
    h = data.getRasterHeight()

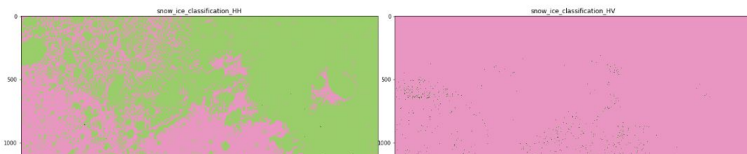
    band_data = np.zeros(w * h, np.float32)
    data.readPixels(0, 0, w, h, band_data)
    band_data.shape = h, w

    cmap = plt.cm.get_cmap('PiYG', 5)
    imgplot = plt.imshow(band_data,
                        cmap=cmap,
                        vmin=0,
                        vmax=30)

plt.tight_layout()
fig = plt.gcf()
plt.show()
fig.clf()
plt.close()

```

Snow and ice classification



INTAROS



Jupyter Notebooks for EO data processing

Tools for data analysis, transformation and visualization

Jupyter Notebook applications introducing EO data processing techniques for Arctic areas monitored using satellite earth observations:

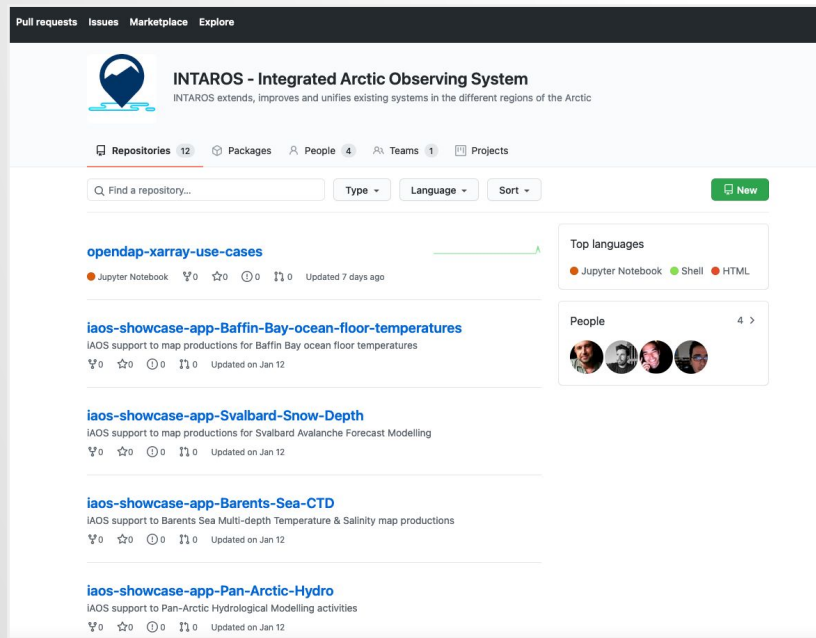
- **01-polarstern.ipynb**: get and clean the Polarstern AIS data, use the Polarstern position at 2018-08-22 03:00 to discover Sentinel-1 data, stage-in the discovered Sentinel-1 data, and plot a quicklook of the staged-in Sentinel-1 product
- **02-snap-intro.ipynb**: introduce the Sentinel Application Platform (SNAP) and create a data processing graph to extract the Sigma0 measure out of a Sentinel-1 product
- **03-snow-ice-classification.ipynb**: apply a simple snow and ice classification derived from a knowledge-based approach.
- **04-glacier-velocity.ipynb**: apply the offset tracking technique to derive the glacier velocity maps with Sentinel-1 Level-1 Ground Range Detected (GRD) products. Offset Tracking is a technique that measures feature motion between two images using patch intensity cross-correlation optimization.
- **05-multitemporal-rgb-2018.ipynb**: use of multi-year SAR data to study the seasonal dynamic of the snow melt patterns.



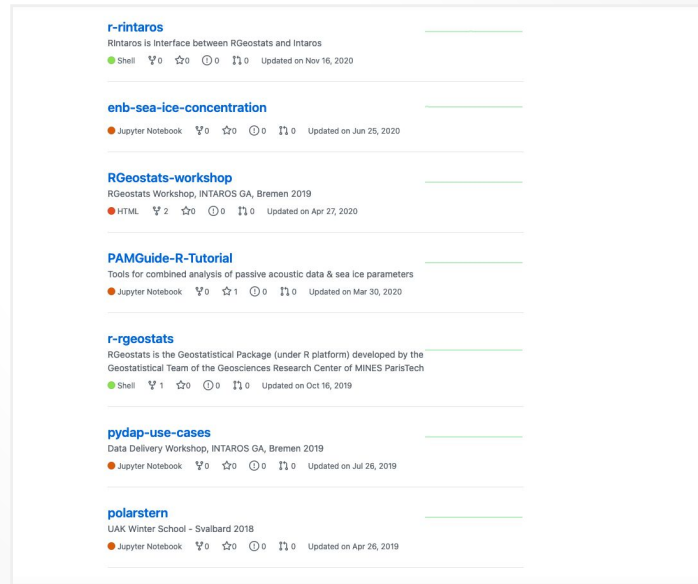
Software Repositories

Overview of the INTAROS GitHub community contents

<https://github.com/ec-intaros>



The screenshot shows the main page of the INTAROS GitHub repository. At the top, there's a navigation bar with links for Pull requests, Issues, Marketplace, and Explore. Below this is the repository header for "INTAROS - Integrated Arctic Observing System", which includes a description: "INTAROS extends, improves and unifies existing systems in the different regions of the Arctic". The main content area displays a list of repositories, each with its name, a brief description, and statistics like stars, forks, and updates. The repositories listed are: opendap-xarray-use-cases, iaos-showcase-app-Baffin-Bay-ocean-floor-temperatures, iaos-showcase-app-Svalbard-Snow-Depth, iaos-showcase-app-Barents-Sea-CTD, and iaos-showcase-app-Pan-Arctic-Hydro. On the right side of the repository list, there are sections for "Top languages" (Jupyter Notebook, Shell, HTML) and "People" (a group of four profile pictures).



This screenshot shows a list of repositories from the INTAROS GitHub community. Each repository entry includes the repository name, a brief description, and statistics such as stars, forks, and updates. The repositories listed are: r-rintaros (Rintaros is interface between RGeostats and Intaros), enb-sea-ice-concentration (Jupyter Notebook), RGeostats-workshop (RGeostats Workshop, INTAROS GA, Bremen 2019), PAMGuide-R-Tutorial (Tools for combined analysis of passive acoustic data & sea ice parameters), r-rgeostats (RGeostats is the Geostatistical Package (under R platform) developed by the Geostatistical Team of the Geosciences Research Center of MINES ParisTech), pydap-use-cases (Data Delivery Workshop, INTAROS GA, Bremen 2019), and polarstern (UAK Winter School - Svalbard 2018).





3. WP5 Expected Impact



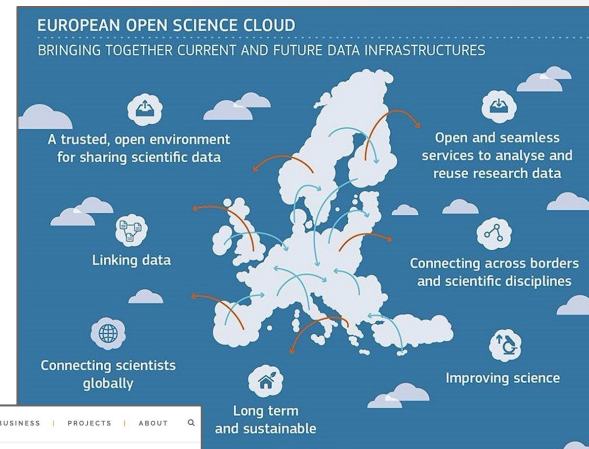
Key factors for iAOS to have an impact

Cost, legal, trust, privacy, security, usability

With the achieved maturity level of the iAOS cloud infrastructure, having a sustained impact after the completion of the project implies to consider **funding sources** (e.g. other projects), and stakeholder activities as part of a broad picture (e.g. **EOSC**), having in scope transformative changes and socio-economic impacts.

See also D5.13 - Synthesis of the iAOS infrastructure

Terradue is involved in EOSC-support actions with EGI.eu and ESA, building further Cloud Platform capacity of interest for the iAOS



The screenshot shows the ESA website page titled "About ESA" and "The challenge". The page is part of a project titled "EGI-ACE" and includes a navigation bar with links to SERVICES, FEDERATION, USE CASES, BUSINESS, PROJECTS, and ABOUT. The main content area is divided into two columns. The left column contains the following sections:

- ESA**
European Space Agency
- About ESA**
The **European Space Agency (ESA)** is Europe's gateway to space. Its mission is to shape the development of Europe's space capability and ensure that investment continues to deliver benefits to the citizens of Europe and the world. ESA is an international organisation with 22 Member States.
- The challenge**
Terradue is a company specialised in delivering e-infrastructures for earth sciences and was tasked by ESA to lead the development of a cloud infrastructure to support the Geohazards and hydrology thematic exploitation platforms. Terradue needed cloud resources to make this possible and to be able to handle massive and complex data streams.
- EGI-ESA collaboration**
The two thematic exploitation platforms were integrated within the **EGI Federated Cloud** to guarantee enough computational power for their use cases. This task was successfully completed by developing an interface between the Terradue cloud framework and the standard OGC interface of the EGI Federated Cloud. Seven EGI cloud providers from Italy, UK, Greece, Germany, Poland, Belgium and Spain committed the cloud resources necessary for the project through Service and Operational Level Agreements.

The right column contains the following sections:

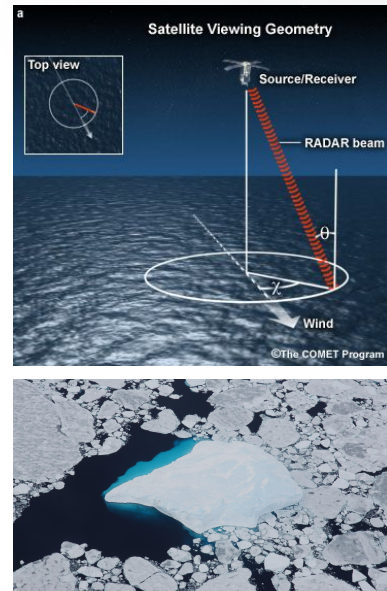
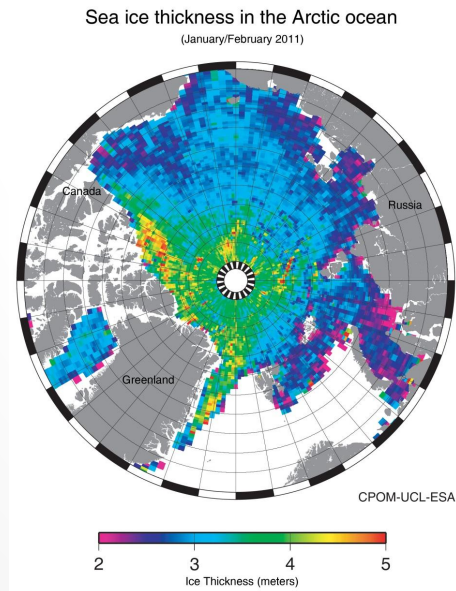
- Services used by ESA**
Cloud Compute
- Partners involved**
ESA: Terradue
EGI: EGI Foundation, 100%IT, HG-09-Okeanos-Cloud, GoeGrid, CESSGA, RECAS-BAR, CYFRONET-CLOUD, BEGrid-BELNET
- More information**
ESA website
News item: EGI and Terradue: a better cloud service for science

Exploitation plan - Terradue

Cloud processing services using EO data

- Earth observation data, in particular the Copernicus Sentinel products, have proven a good level of relevance and applicability in the iAOS cloud infrastructure context
- Tools and platform services successfully demonstrated with the iAOS applications:
 - UAK research school
 - INTAROS collaboration with the EC NextGEOSS project

This target impact was defined from the start of the WP5 activities in INTAROS, and a number of tutorials and training assets have been produced consequently. They are publicly shared on the INTAROS community on GitHub.



Exploitation plan - Terradue

Cloud processing services using EO data

Providing data services/processing services

Improved “Ellip Solutions” EO data processing services and documentation, which are used both to generate new EO products (level 3-4) and for enabling scientists to prototype new algorithms and validate services.

Portal:

<https://www.terradue.com/portal/ellip>

Documentation:

<https://docs.terradue.com/ellip>

Dashboard:

<https://ellip.terradue.com>

(private access, Ellip users)

Providing improved access to data from repositories

New capacity to test and validate Ellip-powered applications over OpenDAP standard endpoints (Hyrax/Thredds) configured over dataset samples, thus releasing operational servers from testing-only workloads.

TDS OpenDAP Server for Ellip users:

<https://opendap.terradue.com/thredds/>

(private access, Ellip users)

Hyrax OpenDAP server for Ellip users:

<https://opendap.terradue.com/hyrax/>

(private access, Ellip users)

Demonstrating useful applications towards stakeholders

Jupyter Notebook files executable on a Jupyter Lab software environment (<https://jupyter.org/>), with access to Open Science software repositories, stored by the EC-INTAROS GitHub organisation.

UAK Winter School - Svalbard 2018

<https://github.com/ec-intaros/polarstern>

RGeostats Workshop - Bremen 2019

<https://github.com/ec-intaros/RGeostats-workshop>

Barents Sea Multi-depth Temperature & Salinity Maps

<https://github.com/ec-intaros/iaos-showcase-app-Barents-Sea-CTD>

Data extraction from OPeNDAP server at NMDC

<https://github.com/ec-intaros/iaos-CTD-extract-from-opendap>





4. WP5 Challenges



Deploying and operating the IAOS cloud infrastructure (1/3)

Specific challenges	Return of experience
Federate access to distributed data repositories from multiple stakeholders and with disparate technical maturity levels.	Only Pangaea and OPeNDAP-based servers provide structured support (software tools for developers, online documentation) but still the maturity level is low, estimated at TRL6 “prototype demonstration in a relevant environment”, compared to TRL8 “System complete and qualified” and TRL9 “Actual system proven in operational environment”.
Federate access to Cloud providers from major initiatives (DIAS, EGI.eu, EOSC).	The capacity within iAOS to tap into Cloud Computing resources was delivered to few use cases in INTAROS (support to ARMINES and NERSC). While Terradue Cloud Platform provides the capability to connect to the Cloud Providers from major initiatives in Europe (DIAS, EGI.eu, EOSC), the INTAROS partners with high compute load needs made use of their pre-established corporate access to HPC resources.



Deploying and operating the IAOS cloud infrastructure (2/3)

Specific challenges	Return of experience
Validate and brand a Software Development Toolkit (SDK) encapsulating all the key developer functions for API-based functions	Cross-projects coordination was difficult to handle in this particular scope. The Ellip Software Development Kit (SDK) was presented in the INTAROS deliverable D5.8 iAOS Platform and Tools V2 (revised 10 June 2020). The goal to repackage the Ellip SDK tools as presented was partially attained within INTAROS , since a technical trend emerging from the ESA and OGC communities (EOEPCA), while anticipated and contributed to by Terradue, is going to change some of the orientations initially described in D5.8. Terradue is still actively contributing to EOEPCA (ESA/OGC press release upcoming end January 2022)
Build a data catalogue federating Arctic-related data sources	This objective was successfully attained by NERSC as part of the task T5.6. Cf. updates on the iAOS Portal and the INTAROS catalogue in D5.12 “iAOS Portal with user manual V2”.



Deploying and operating the IAOS cloud infrastructure (3/3)

Specific challenges	Return of experience
On-board and support a developer user community	Due to technical efforts allocated by WP5 on building and consolidating the iAOS cloud service features, it was not possible to put much extra efforts on developing the user community of the iAOS beyond the WP5 partners and the WP6 partners in charge of a selected iAOS showcase application . A notable success in this matter was the “polarstern” Earth Observation module and training delivered during the UAK Winter School held on 02-07 December 2018 at UNIS, Longyearbyen, Svalbard, that brought together leading researchers, educators and young scientists from Norway, USA and Canada, and working on Arctic science topics.
Deliver data processing applications as online services accessed from a user Portal	It was not possible to address this objective, simply due to the fact that no specific INTAROS-specific web processing service has reached a maturity level allowing this type of system integration . The orientation was put halfway through the INTAROS project to focus on delivering Jupyter Notebooks instead. Latest developments (December 2021) on CTD extraction tool might address further this challenge (under evaluation between Terradue and NERSC).



Integrating data from existing data repositories

Specific challenges	Return of experience
Find and assess datasets accessible online , based on their potential relevance for a given use case or a set of initial requirements	It remains difficult for persons that were not previously exposed (meetings, telcos, reports) to a specific dataset to rely on current search engine technologies (web search engine or portal search engine) in order to get a clear overview of the initiatives, data producers and online repositories able to deliver on the use case expectations. The data discovery processes involved in support of the definition of the iAOS showcases were largely dependent on human expertise and advice , and therefore not straightforward from the start.





5. Recommendations from WP5



Establishing a sustainable pan-Arctic iAOS cloud infrastructure

Address long-term infrastructure properties from the start, along the lines of the technical work performed during the INTAROS project timeframe:

- infrastructure cost assessment,
- legal dimensions,
- trust building,
- privacy policies,
- security policies
- and usability criteria.

See D5.13 - Synthesis of the iAOS infrastructure



Focused work to be pushed forward

Implementation of **Data standards**, while being well established in the software systems involved within iAOS, are still lacking support to software application developers, and it is still needed to:

- Assess the **learning curve** and improve **user experience** related to all the aspects of an infrastructure such as iAOS cloud platform, for an application integrator.
- Improve **online support resources related to interoperability protocols** (e.g. DAP2, DAP4), for software application developers.

This challenge appeared quite from the start of the WP5 activities in INTAROS, and a number of tutorials and training assets have been produced consequently.

These Data standards related resources are publicly shared on the INTAROS community on GitHub, and shall be pushed forward.

Tooling support for the **volume, variety and variability of data sources**. The set of iAOS showcase applications demonstrated this remains a challenge, and it is still needed to:

- Optimise tools for **data preparation** effort, in particular to assess the data quality, as platform-based automations to support this effort-consuming step.
- Optimise tools for **data analysis** effort of poly-structured data sources, in order to perform exploratory analysis and come up with work plan hypotheses.

The WP5-WP6 collaboration on the iAOS showcases helped to pinpoint the high interest of scientific users for improved, platform-based, data science capabilities.

The overall data value that is represented here is very important for research and development and shall be pushed forward.



WP5 - DATA INTEGRATION AND MANAGEMENT

Results and impact

INTAROS Final Meeting - Synthesis

Looking forward
hearing from you !

January 20th-21st 2022

Hervé Caumont, Terradue Srl

