Integrated Arctic Observation System

Research and Innovation Action under EC Horizon2020
Grant Agreement no. 727890

Project coordinator:
Nansen Environmental and Remote Sensing Center, Norway

Deliverable 5.13

Synthesis of the iAOS cloud infrastructure

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	31 May 2021	Actual submission date:	3 December 2021
		Resubmission after review:	26 August 2022

Lead beneficiary for preparing the deliverable: TERRADUE

Other contributing partners

Person-months used to produce deliverable: 1 pm

Authors: Hervé Caumont (Terradue)

Version	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	06/01/2021	Initial draft - Table of content	Hervé Caumont
0.2	21/04/2021	Updates on ToC	Hervé Caumont
0.3	28/05/2021	First draft for internal review First version for internal project review (an update being scheduled to November 2021, in agreement with the project coordinator, in order to build upon all the "V2" deliverables from WP5)	Hervé Caumont
0.4	14/10/2021	Review	Torill Hamre
1.0	30/11/2021	Final for EC delivery	Hervé Caumont
1.1	03/12/2021	Final review and submission	Toril Hamre
1.2	26/08/2022	Updated according to external review	Hervé Caumont

Approval **x**	Date: 26/08/2022	Sign. Stein Sandven Stein Sandven

USED PERSON-MONTHS FOR WRITING THIS DELIVERABLE					
No	***Beneficiary***	***PM***	***No***	***Beneficiary***	***PM***
1	NERSC		24	TDUE	1
2	UiB		25	GINR	
3	IMR		26	UNEXE	
4	MISU		27	NIVA	
5	AWI		28	CNRS	
6	IOPAN		29	U Helsinki	
7	DTU		30	GFZ	
8	AU		31	ARMINES	
9	GEUS		32	IGPAN	
10	FMI		33	U SLASKI	
11	UNIS		34	BSC	
12	NORDECO		35	DNV GL	
13	SMHI		36	RIHMI-WDC	
14	USFD		37	NIERSC	
15	NUIM		38	WHOI	
16	IFREMER		39	SIO	
17	MPG		40	UAF	
18	EUROGOOS		41	U Laval	
19	EUROCEAN		42	ONC	
20	UPM		43	NMEFC	
21	UB		44	RADI	
22	UHAM		45	KOPRI	
23	NORUT		46	NIPR	
			47	PRIC	

DISSEMINATION LEVEL		
PU	Public, fully open	**X**
CO	Confidential, restricted under conditions set out in Model Grant Agreement	
CI	Classified, information as referred to in Commission Decision 2001/844/EC	

EXECUTIVE SUMMARY

WP5 (Data integration and management) is tasked with designing and implementing evolutions of the cloud platform and tools for the iAOS. This document gives a synthesis of the iAOS cloud infrastructure evolution during the project. The experience and challenges from the work include: (1) deploying and operating dedicated Cloud Platform infrastructure services, (2) integrating data repositories and data processing services based on these Cloud Platform infrastructure services, and (3) defining user-oriented applications powered by all these iAOS resources.

As a result, the iAOS cloud platform offers services to access observations and derived parameters, including new observations from WP2-3-4. In addition, it offers seamless integration with geo-statistical methods for interpolation of spatiotemporal datasets, and services to store the generated datasets in iAOS enabled repositories.

The report presents results on using the iAOS cloud platform to integrate multidisciplinary and distributed data repositories and to provide a set of tools for data analysis, transformation, and visualization. Selected applications established in WP6 have used the services and tools provided by the platform and have showcased the usefulness and functionality of the platform, as part of the benefit of enhanced integration of data from Arctic observing systems.

Table of Contents

Acronyms and abbreviations

API	Application Programming Interface
DAP	Data Access Protocol
DevOps	Development to Operations
DIAS	Data and Information Access Service
EGI	European Grid Infrastructure
EO	Earth Observation
FRAM	FRontiers in Arctic marine Monitoring project
HTTP	HyperText Transfer Protocol
IaaS	Infrastructure-as-a-Service
iAOS	integrated Arctic Observation System
ICT	Information and Communication Technologies
IDE	Integrated Development Environment
NCSS	NetCDF Subset Service
NERSC	Nansen Environmental and Remote Sensing Center
OGC	Open Geospatial Consortium
OpenDAP	Open Data Access Protocol
OSDD	OpenSearch Description Document
OWS	OGC Web Services
PaaS	Platform-as-a-Service
SDK	Software Development Kit
THREDDS	Thematic Real-time Environmental Distributed Data Services
VM	Virtual Machine
VPN	Virtual Private Network
WPS	Web Processing Service

Definitions

Application
A program or group of programs designed for end users. An application, application program or application software is a computer program designed to help people perform an activity.

Data Access Protocol (DAP)
The DAP2 and DAP4 are network protocols for structured data access, implemented as part of the open source initiative OpenDAP "Open-source Project for a Network Data Access Protocol" in particular supported by US agencies[1] (e.g. NASA and NOAA).

Infrastructure as a Service (IaaS)
IaaS is a standardized, highly automated offering, where compute resources, complemented by storage and networking capabilities are owned and hosted by a service provider and offered to customers on-demand. Customers are typically able to self-provision this infrastructure, using a Web-based graphical user interface that serves as an IT operations management console for the overall environment. Application Programming Interface (API) access to the infrastructure may also be offered as an option.

Platform as a Service (PaaS)
PaaS provides users with the capability to deploy user-created or acquired applications developed using programming languages and tools supported by the provider onto an infrastructure, typically cloud- based. The user does not manage or control the underlying infrastructure, including network, servers, operating systems, or storage, but has control over the deployed applications and possibly application hosting environment configurations.

Service
When an application software has been integrated and tested for cloud computing operations, it is deployed on production servers and made available as a "service": the application functionality can be delivered via a simple invocation interface, making the full application functionality available to others as a service of the iAOS.

Software as a Service (SaaS)
SaaS is characterizing application software that is owned, delivered and managed remotely by one or more providers. Typically the application provider delivers the software based on one set of common computing code and data definitions, that is consumed in a one-to-many model by all contracted customers, at any time, and on a pay-for-use basis or as a subscription based on the monitoring of service usage metrics.

Software Development Kit (SDK)
A set of software tools installable on the developer Integrated Development Environment (IDE) and usable as software dependencies in the developer's application code.

---

[1] OpenDAP Best practice resources can be found from these reference sites:
https://www.earthdata.nasa.gov/engage/open-data-services-and-software/api/opendap
https://psl.noaa.gov/data/gridded_help/using_dods.html
https://opendap.co-ops.nos.noaa.gov/erddap/index.html
https://wiki.esipfed.org/Making_Science_Data_Easier_to_Use_with_OPeNDAP
see also the INTAROS deliverable D5.11 "Integration of new processing services V2"

1. Introduction

This document provides return of experience from the integration and deployment of iAOS cloud infrastructure, leveraging the integration of multidisciplinary and distributed data repositories from different INTAROS partners and the cloud computing environment provided by Terradue. It describes in detail how iAOS stakeholders shall sustain their contribution to iAOS, in particular for the cloud platform and tools to build services using a federation of distributed compute and data resources.

1.1. Purpose and scope of the document

This document is directly attached to the WP5 task T5.7 Synthesis of IAOS infrastructure deployment and operation (led by Terradue), and covers the activities of the INTAROS Tasks:

- T5.1 System requirements and architecture consolidation (led by Terradue)
- T5.2 IAOS platform deployment and operation (led by Terradue)
- T5.3 Integrate data from existing repositories into iAOS (led by AWI & Terradue)
- T5.4 Development of geo-statistical methods for data integration (led by ARMINES)
- T5.5 Integration of new processing services (led by Terradue)
- T5.6 iAOS portal development (led by NERSC)

It summarizes the return of experience from deploying and operating the iAOS cloud infrastructure to establish a sustainable pan-Arctic iAOS cloud infrastructure, and reports the lessons learned from technical and non-technical challenges that occurred during the project in that matter.

1.2. Intended audience of this document

The main target audience of this document are the INTAROS Consortium Partners (especially the WP5 partners integrating processing services and the WP6 partners developing showcase applications) and European Commission services.

External stakeholders and collaborators such as organisations in charge of studies over the arctic region or other arctic-related projects, can benefit from this work by selecting iAOS resources presented in the report for their own usage (software, datasets) and thus taking part in developing the iAOS vision for data science and for services integration in a sustainable cloud infrastructure.

1.3. Document Structure

This section describes the document structure:

Section 1 is this introduction section.

Section 2 provides a review of experiences and challenges in deploying and operating the IAOS infrastructure, integrating data from existing data repositories and integrating EO data Processing services, as well as of the WP5 activities in support of selected WP6 iAOS showcase applications.

Section 3 describes identified actions for a sustainable pan-Arctic IAOS cloud infrastructure, in terms of cost, legal, trust, privacy, security and stability.

Section 4 is the report’s conclusions.

2. Experiences and challenges

INTAROS has addressed the development of community-based observing systems, where local knowledge is merged with scientific data, and the integrated Arctic Observation System is one aspect of it. It contributes in this way to enable better-informed decisions and better-documented processes within key sectors (e.g. local communities, shipping, tourism, fisheries), in order to strengthen the societal and economic role of the Arctic region and support the EU strategy for the Arctic and related maritime and environmental policies. The evolution of the iAOS into a sustainable contribution to an Arctic observing system required coordination, mobilization and cooperation between the existing European and international infrastructures (in-situ and remote, including space-based), the modeling communities and relevant stakeholder groups.

We'll present hereafter the key activities coordinated as part of WP5 in support of this objective. The following chapters discuss the tradeoff we had to face and the choices made to ensure of the good maturity of the selected technical solutions, in particular the solutions operated over cloud computing resources, and to ensure the cautious implementation of interoperability standards, related to computing system interfaces (storage and compute as-a-Service) as well as to data access and data processing (such as OGC Web Services standards and the de facto OPeNDAP standard).

These experiences and challenges are grouped in different sections depending on the expertise and teams required: from system infrastructure components and their federation in a working ensemble, through the validation of data access tools to feed data processing software in a flexible way that can be delivered on-demand to user communities, and to finally the development of applications serving some scientific questions selected as part of the INTAROS collaborations.

2.1. Deploying and operating the IAOS cloud infrastructure

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). 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.
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 (see section “Polarstern” further below) 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.
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 and we foresee a possible update in this matter by December 2021.
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.
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”.

The overall iAOS is made of observing infrastructures, data systems managing dataset resources, and software solutions (including Cloud-based Platforms) for the development, integration and deployment of data processing services. INTAROS supported notably the development of specific features for the iAOS Cloud Platform to search for and access data from distributed databases, providing a common entry point to data originating from a wide range of observation networks, scientific campaigns and satellite missions, as well as new data generated within the project.

The iAOS Cloud Platform leverages state-of-the-art cloud computing technologies to facilitate seamless access to multidisciplinary data, scalable allocation of data storage and computing power for big data processing, integration and analysis including geostatistical methods.

The synoptic view of the logical connections and integration of main iAOS data systems and computer services with the iAOS cloud platform has been described in the INTAROS deliverable D5.8 “iAOS platform and tools V2” (v1.4).

It showed how the iAOS cloud platform supports the development and maintenance of processing services using the cloud platform “Ellip Solutions” operated by Terradue to empower these processing services with:

- A service integration environment, where the processing services are under test and validation running on a “Development Infrastructure”;
- A service deployment environment, where the processing services are deployed in production running on a highly scalable “Hosting Infrastructure”;
- An embedded capacity for each Ellip-powered processing service to interact, whatever the infrastructure it runs on, with remote repositories and Web Portal(s).

INTAROS partners in WP5 have used the Cloud platform to develop:

- showcase applications e.g. Geostatistics in T5.4 (ARMINES),
- processing and analysis of acoustic data in T5.5 (NERSC),
- and other services related to WP6 partners activities (as listed in the table of content of this deliverable) that are related to:
    - Services for the use of Copernicus Sentinel products:
        - UAK research school (December 2018 at UNIS, Longyearbyen, Svalbard), Jupyter Notebook applications for:
            - AIS data to discover relevant Sentinel-1 products
            - Sentinel-1 for snow and ice classification
            - offset tracking techniques on Sentinel-1 Level-1 Ground Range Detected (GRD) products to derive glacier velocity maps
            - use of multi-year SAR data to study the seasonal dynamic of the snow melt patterns.
        - Sentinel-1 processing Sea ice classification products and Sea ice drift products generated and published in an open data repository by NERSC for exploitation in INTAROS.
    - Data integration work performed for:
        - OpenDAP services
        - PANGAEA data library services
    - Services for the use of openDAP served datasets:
        - 2.3.1.1. opendap-xarray-use-cases
        - 2.3.1.2. pydap-use-cases

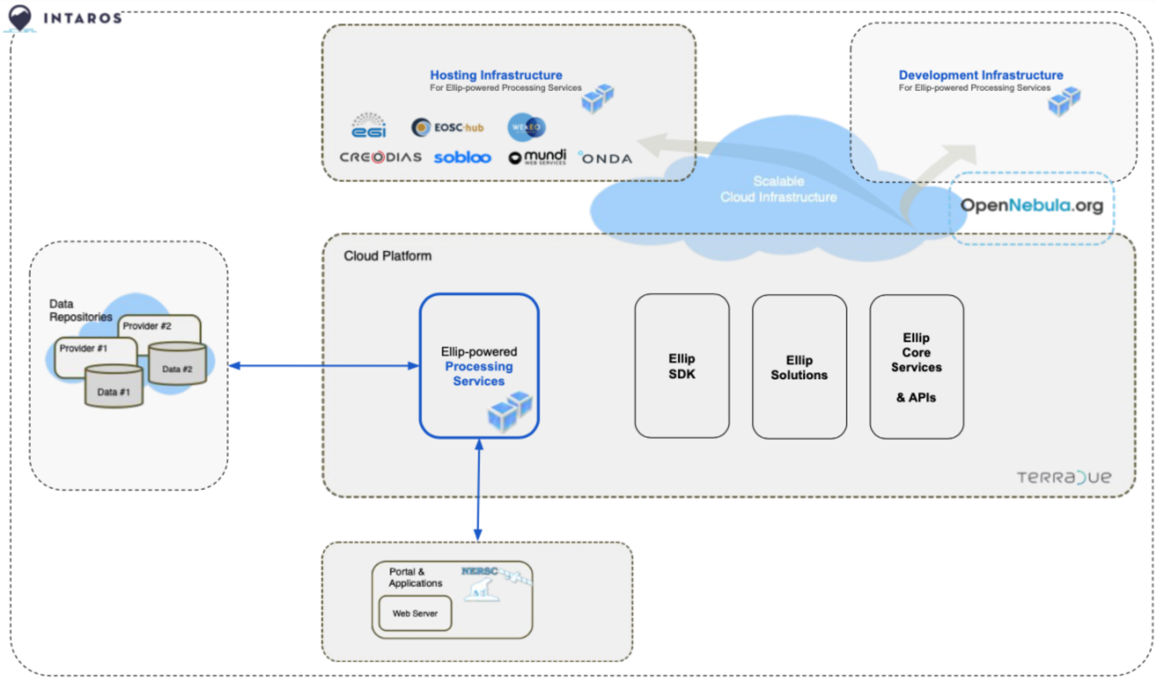

Figure 1. (from INTAROS D5.8) Synoptic view of how the Cloud platform is linked to other components of the iAOS e.g. data repositories, iAOS portal, as well as development and deployment (hosting) infrastructures.

A significant return of experience from the INTAROS activities has been collected over the project timeframe, with the following outcomes. We present these outcomes structured according to the D5.8 detailed view of the iAOS.

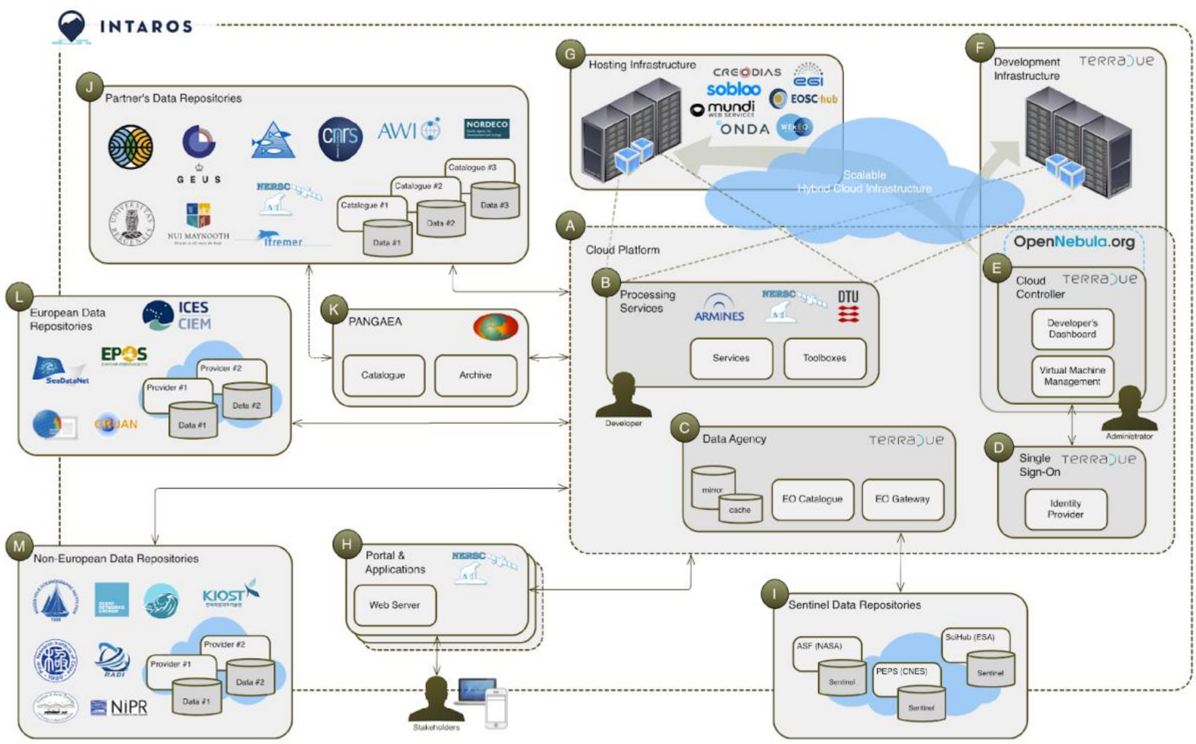

Figure 2. (from INTAROS D5.8) The architecture of the Information and Communication Technologies (ICT) components of the overall iAOS.

In **Figure 2** the ICT components of the iAOS are depicted in more details, where each of the elements are labeled A to M in the diagram, and for which we present hereafter the return of experience out of the INTAROS WP5 activities:

- **(A)** This set of iAOS Cloud Platform services, defined as a Collaborative workspace operated within a domain hosted on Terradue Cloud Platform, and accessed by the authorized partners via the use of the Ellip Solutions, was essentially improved as part of the INTAROS funding, on two aspects: integration of new software toolboxes for application developers (RGeostats, cf. point B. below), and addition of OPeNDAP server capabilities to the Data Agency (Hyrax server and Thredds Data Server, cf. point C. below). No particular evolution, as part of the INTAROS funding and WP5 activities, to be noted on the Identity Provider service side (cf. point D. below) and on the Cloud Controller technology side (cf. point E. below) of the Cloud Platform.

- **(B)** This specific part of the Ellip Solutions, related to software toolboxes and the processing services maintained on the Platform over their full lifecycle, from the initial versions for integration and testing to the released ones, deployed in operation on production servers, did not receive much traction as part of the INTAROS project, compared to other H2020 communities to which it was also delivered. It consumed a lot of software development efforts from ARMINES and support from Terradue to ARMINES to port the r-RGeostats and r-RIntaros software libraries onto the Cloud Platform environment as referenced packages (which is a success, still, and is a prerequisite for further integration steps) on one hand, and the efforts to ensure efficient CTD data access mechanisms (with efficient software driven data filtering and downloads) performed from within the Cloud-based software applications on the other hand (definition of reliable, query filters), did not allow the WP5 partners to bring up this specific part of the Ellip Solutions outside of the boundaries of WP5.

- **(C)** This set of Platform capabilities, related to Data Catalog services and Data Gateway services in charge of the data discovery and access, is used to feed processors with specific collections or product types, based on dataset files mirroring, caching and remote access strategies. It includes mechanisms for data registration. The Data Agency supports programmatic discovery and access to distributed EO data repositories for scalable data processing services. Processing Service developers can request data mirroring operations into their allocated Cloud storage on the Platform (data buckets), and exploit a standard interface to define data staging operations from the referenced repositories (mirrored or remote) to their Application processing nodes. as part of the INTAROS

- **(D)** No particular evolution, as part of the INTAROS funding and WP5 activities, to be noted on the Identity Provider service side, for the management of Platform user accounts, and the Single sign-on (SSO) service for authentication and authorization across all the Platform resources federated under the SSO. Some parts of the iAOS are currently able to operate from well established Identity providers related to research and open science (EGI Check-in, EduGain, …).

- **(E)**, **(F)** and **(G)** No particular evolution, as part of the INTAROS funding and WP5 activities, to be noted on the Cloud Controller technology side, for the management of Virtual Machines allocated to the iAOS platform domain and users conducting service integration activities, and for the management of Production Centers in which the integrated services are moved for their operational exploitation on a Cloud Computing infrastructure. Capacities on Terradue Cloud Platform currently allow organisations to tap into Cloud Providers from major Cloud Computing initiatives in Europe: Copernicus DIAS, EGI.eu, and EOSC. These providers are exploited for the 'background facilities' of the iAOS, for running the services through their lifecycle, from integration, testing and packaging (the main purpose of the ‘Development’ private Cloud environment) to Hosted Processing operations (the main purpose of the ‘Hosting” environment using public Cloud providers, commercial or institutional).

- **(H)** The iAOS Portal enables users to access federated data repositories, as well as providing links to selected Processing Services integrated and deployed on the Platform. This portal has been developed in WP5 task T5.6 iAOS portal development (led by NERSC), and is connected to the INTAROS Data Catalogue referencing the collection of INTAROS-related science, community-based and citizen science datasets.

- **(I)** The pool of Copernicus Sentinel distributed data repositories (e.g. Copernicus scientific hub, Copernicus collaborative ground segments) that are federated on Terradue Cloud Platform, thus available for access from the iAOS processing services, have significantly evolved over the past years, with in particular a change in the Sentinels rolling archive policies. Terradue for WP5 have carefully tracked these changes and implemented ad-hoc “data ordering” services to interact with the Data Hub Service capability of requesting products marked with an "offline" status, removed from the online archives, but available on the “cold” long term archive (tape storage and tape robot technology) .

- **(J)** For the pool of INTAROS Partner’s data repositories that have been federated into the iAOS via interoperability arrangements (data and metadata formats, online access protocols), the main integration points are the iAOS Portal and the INTAROS Data Catalogue.

- **(K)** For the PANGAEA-federated data repositories, along with the PANGAEA catalog entries, that have been federated into the iAOS cloud platform via interoperability arrangements (data and metadata formats, online access protocols), progresses have been for the PANGAEA data download service, presented in the INTAROS deliverables D5.3 and D5.8, for query support to the PANGAEA data warehouse, and returning values in a tab delimited text file, based on search criteria that (1) specify a bounding box in time and space (latitude, longitude and water depth), and (2) specify a individual list of parameters.

- **(L)** For the pool of European data repositories with relevance for the Arctic areas that have been federated into the iAOS cloud platform via interoperability arrangements (data and metadata formats, online access protocols), applied data exploitation experiments have been conducted in the context of the iAOS Showcase applications, in particular for the use of river discharge data from the Arctic Hydrological Cycle Observing system (Arctic-HYCOS), HydroGFD v3 temperature and precipitation data, ECMWF deterministic medium range weather forecasts (with guidance from T6.1), as well as CTD datasets from nodc.no, ICES CTD and Bottle datasets from ocean.ices.dk, GINR trawl catches datasets, bathymetry datasets from gebco.net (with guidance from T6.2 and T6.8), and also NMI historical weather and climate data, NMI arome model data and Norwegian Polar Institute Svalbard Terrain Model (with guidance from T6.4).

- **(M)** for the pool of Non-European data repositories with relevance for the Arctic areas that have been federated into the iAOS cloud platform via interoperability arrangements (data and metadata formats, online access protocols) some experiments have been conducted in the context of the iAOS Showcase applications, in particular for the use of NASA OMG (Oceans Melting Greenland), NASA GTSPP (Global Temperature and Salinity Profile Programme) and NOAA NCEI WOD (World Ocean Database) datasets.

2.2. 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.

We present hereafter the key target data repositories integration goals defined for WP5 as part of the INTAROS project.

Notably, following discussions and findings during the initial project General Assembly, a dedicated Task Force within the project was led by TERRADUE and AWI to progress the exploitation of OPeNDAP servers from within client software applications. It focused on knowledge sharing between project partners on OPeNDAP and NetCDF through online sessions and email exchanges (up to 13 partner institutions took part and had an intensive exchange between May 2018 and June 2020). Moreover a workshop session at the Bremen General Assembly in January 2019 was focused on competence building for the setup of OPeNDAP servers by interested INTAROS partners (Terradue, Bremen University, IMR/NODC, and SMHI).

2.2.1. OPeNDAP servers - Operations

Specific challenges	Return of experience
Exploit operational, stable OPeNDAP endpoints to query and filter datasets accessible online based on software application data ingestion capabilities	Only a few operational OPeNDAP endpoints could be exploited during the INTAROS project for accessing Arctic-related datasets, namely: - SMHI (Arctic-Hype) - NODC (CTD) - NOAA NCEI (WOD)

	- NASA JPL PODAAC (AXCTD)

The following operational OPeNDAP servers have been investigated, with different levels of technical use cases, from simple to browser-based search and download to more elaborate machine-to-machine (M2M) queries based on the DAP2 or DAP4 protocol:

- SMHI Arctic-HYPE pan-Arctic hydro model
  - http://opendata-download.smhi.se/opendap/catalog/catalog.html
  - Well integrated effort under the lead of SMHI with user documentation embedded within the server endpoint.
- NODC CTD 1995-2016
  - http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html
  - http://opendap1-test.nodc.no/opendap/physics/point/yearly/contents.html
  - Some difficulties to fully implement M2M queries for the WP5 use case developed with ARMINES. The backend server technology (Hyrax) was not fully documented to allow autonomous work. Email and phone interactions with the NODC expert was needed but then this person left the organisation and activity has remained pending.
- NOAA NCEI
  - https://www.nodc.noaa.gov/OC5/WOA09/netcdf_data.html and https://www.ncei.noaa.gov/access
  - Manual, browser-based, data discovery and download only.
- NASA JPL PODAAC
  - https://podaac-opendap.jpl.nasa.gov/opendap/allData/omg/L2/CTD/AXCTD/
  - Manual, browser-based, data discovery and download only.

2.2.2. OPeNDAP servers - Developer support

Specific challenges	Return of experience
Support application developers to take full advantage of the OPeNDAP protocol for interoperability and application portability	The maturity of the DAP2 (circa 2007) and DAP4 (circa 2013) protocols was surprisingly difficult to assess, with too many roadblocks encountered for a non specialist due to lack of tutorials, practical guides and test servers.

Due to the difficult task to guide application developers on the expected behavior of operational OPeNDAP servers, and their query filter capabilities, a dedicated iAOS capability was set up (an OPeNDAP Hyrax server, and an OPeNDAP TDS server, both serving similar dataset samples). This allowed the cloud platform support team to conduct focused data discovery and retrieval experiments, and in particular to guide application developers to implement interoperable OPeNDAP query filters that can be tested against multiple OPeNDAP server implementations. Two of the leading OPeNDAP server software distributions have been selected and deployed, the Hyrax Server (maintained by the OPeNDAP group) and the Thredds Data Server (TDS, maintained by Unidata).

The following "developer support" OPeNDAP servers have been deployed on the iAOS, with a dedicated effort to document interoperable queries based on the DAP2 or DAP4 protocol:

Hyrax Server instance

- https://opendap.terradue.com/hyrax/
    - Access point for Ellip solutions registered users
    - VPN connexion required

Thredds Data Server instance

- https://opendap.terradue.com/thredds/
    - Access point for Ellip solutions registered users
    - VPN connexion required

The following work method is supported by this approach, illustrated for the case of software application interoperability (machine-to-machine) with the NODC OPenDAP server, in order to validate from the service integration environment the query filtering capabilities of the DAP2/DAP4 protocols against a particular dataset or group of datasets, as a representative subset of a bigger data archive:

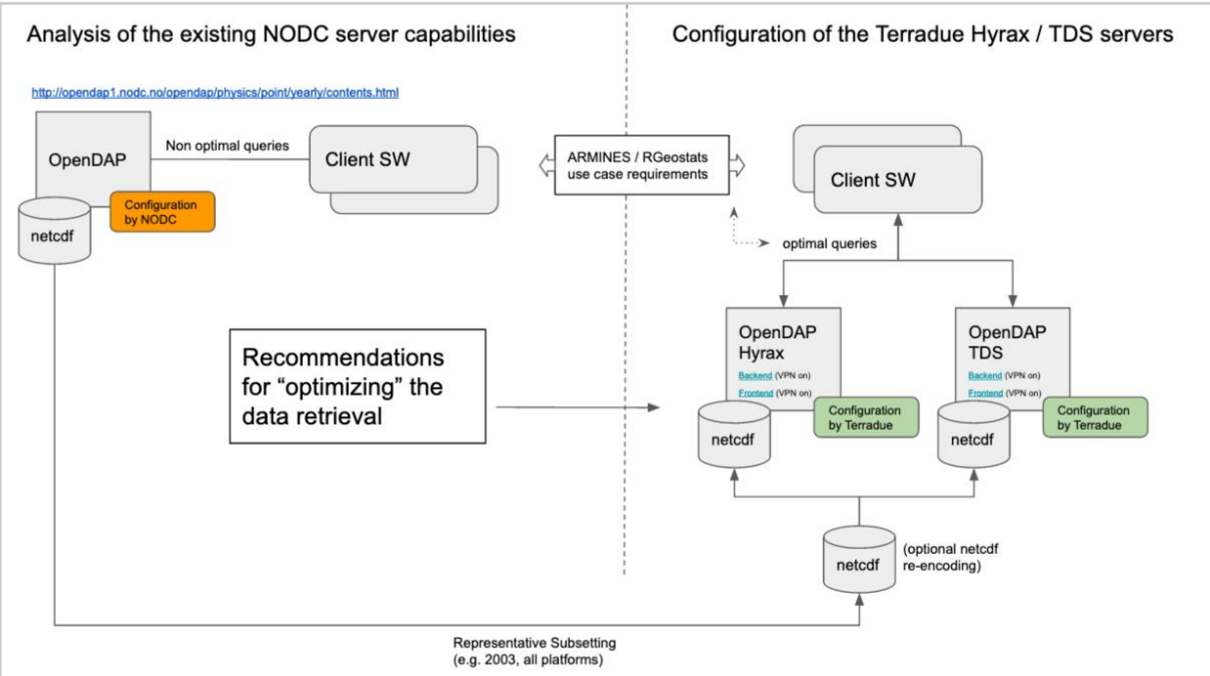

Figure 3. Knowledge building and test services delivery for the iAOS communities of data providers and software developers making use of the OPeNDAP protocols

As the query capabilities are dependent on the OPeNDAP server instance configuration (in particular as enabled by either stored functions or by NCSS settings), the description of the server configuration options are an essential part of this work, in order to formulate INTAROS-wide recommendations for optimizing the data retrieval capabilities and the interoperability limits via OPeNDAP.

2.2.3. Ellip Data Agency

Specific challenges	Return of experience

Keep up with the pace of Copernicus Sentinels acquisitions	The catalog technology within the Ellip Data Agency is mature and no specific issue was encountered for keeping data records synchronised with daily new acquisitions.
Provide support to science users for online data access experience including for data products transferred by the DataHub operators into Long Term Archives (tape robot technology)	Mechanisms for handling data requests towards the Copernicus SciHub endpoint for products moved offline are implemented on the Data Agency, along with a local cache providing longer persistence than the 24 hours advertised by the SciHub operator.

The Ellip Data Agency (cf. D5.8 iAOS platform and tools V2, section 2.2.1) was used during the INTAROS project for the programmatic discovery and access to Sentinel-1 EO data repositories, for scalable data processing services. It is made available to partners in WP6 for their applications making use of the Copernicus Sentinels data products.

2.2.4. PANGAEA data library

Specific challenges	Return of experience
Promote PANGAEA within INTAROS for the adoption of the PANGAEA API by application developers	To enable better integration of the PANGAEA API calls into existing workflows or code, various Python and R packages have been made available via GitHub (https://github.com/o2a-data/o2a-data-pangaea), for software developers to integrate data access from PANGAEA within their software applications (e.g. using Jupyter Notebook) and perform data analysis e.g. with one of the iAOS geostatistics packages. The goal to perform all these actions within the iAOS Cloud Platform could not be very well assessed for the case of PANGAEA , mainly because of not formally identified PANGAEA accesses from the selected iAOS Showcase applications.

PANGAEA is operated as an Open Access library aimed at archiving, publishing and distributing georeferenced data from earth system research. The system guarantees long-term availability of its content through a commitment of the hosting institutions.
Also each dataset can be identified, shared, published and cited by using a Digital Object Identifier (DOI Name).

PANGAEA's data-mining functionalities (data-warehouse) allow combining parameter records from different PANGAEA datasets in one file. A dedicated PANGAEA API provides a ReST-service which queries the data-warehouse and returns values in a tab delimited text file, based on search criteria that (1) specify a bounding cube in time and space (latitude, longitude and habitat depth (water or sediment)) and (2) specify an individual list of parameters. The use of this API has been documented in D5.5 iAOS platform requirements V2 and D5.8 iAOS platform and tools V2.

There was no particular measure of success (increase of data access calls on PANGAEA) performed as part of WP5 in this regard, so the main result is essentially the development of a capability and the communication about this capability being available to the iAOS community.

2.2.5. Copernicus Marine environment monitoring service (CMEMS)

Specific challenges	Return of experience
Value the use of CMEMS data products, positioning iAOS as a set of downstream services making use of CMEMS as a reliable data source.	The return of experience was very positive, with several iAOS applications making use of CMEMS data products.

The iAOS can be seen as a set of downstream services making use of the Copernicus Marine Environment Monitoring Service (CMEMS) as a reliable data source. The Copernicus Marine Service provides key data products for the Polar Environment Monitoring sector to assess environmental impacts at both poles. The CMEMS data repository was used during the INTAROS project for access to, for data processing service with the iAOS Service “Time Series of Sea Ice Concentration (TS-SIC)“, allowing users to extract monthly time series of sea ice classification maps for a selected time period in the North of Svalbard region. The data in the time series can be used to generate monthly mean sea ice concentration (SIC) fields using the CMEMS daily SIC product for the Svalbard region. There was no particular hurdle in using the CMEMS data source for this iAOS application development.

Moreover, a specific iAOS data access scenario, involving the CTD data products coordinated under the Copernicus Marine Service, was developed as part of the iAOS Showcase applications led by T6.2 (Barents Sea Multi-depth Temperature & Salinity Maps) and T6.8 (Baffin Bay Bottom Temperature Map), developed in collaboration with WP5.
An important return of experience for the iAOS partners was to gain better knowledge of the Copernicus Marine Service, both as an organisation and as a service provider. We summarise hereafter some interesting finding related to the CMEMS Thematic Assembly Centres (TAC):

- The CMEMS in-situ TAC aggregates in-situ observations collected by 37000+ platforms (7000+ in near real time) from 300+ providers and applies systematic, consistent quality controls. Regularly enlarged to new data types, sources and platforms, the in-situ TAC products catalogue includes temperature and salinity, currents, sea level, waves, oxygen, chlorophyll a, nutrients and carbon.
- The CMEMS satellite TACs provide near-real time and reprocessed multi-mission satellite products of the wind stress, waves (significant wave-height and waves spectra), sea level and sea surface temperature. They deliver multi-missions, cross-calibrated global and regional products.
- A new multi-obs Thematic Assembly Centre started to be operated in 2018. It provides users with Global Ocean multi-observation products based on observations (satellite & in-situ) and data fusion techniques. The product catalogue delivered by the multi-obs TAC includes: sea surface salinity, density and current as well as 3D temperature, salinity, geostrophic velocities and mixed layer depth.

2.2.6. EMODnet web service

Specific challenges	Return of experience
Value the use of EMODNet data products, positioning iAOS as a set of downstream services making use of EMODNet as a reliable data source.	This goal could not be very well assessed for the case of EMODNet, mainly because of not formally identified EMODNEt accesses from the selected iAOS Showcase applications.

The iAOS can be seen as a set of downstream services making use of EMODNet as a reliable data source. EMODNet is developing and providing a harmonised Digital Terrain Model (DTM) for the European sea regions. The following information layers are made available to application developers:

- Mean depth full coverage on a grid of 1/16 * 1/16 arc minutes (ca 115 meter grid).
- Mean depth in multi colour style on a grid of 1/16 * 1/16 arc minutes (ca 115 meter grid) but no land.
- Mean depth in rainbow colour style on a grid of 1/16 * 1/16 arc minutes (ca 115 meter grid) but no land.
- Source references is a layer which allows to identify the prevailing data sources used for each DTM grid cell.
- Quality Index is an improved source reference layer with quality indication. Metadata has been expanded with characterization of the dataset by vertical, horizontal and temporal indicators and purpose of the survey.
- High Resolution bathymetry is a multi-resolution layer which allows to download a collection of higher resolution composite DTMs (HR-DTMs) for selected areas, with resolution of HR-DTMs varying between 1/32 and 1/512 arc minutes.
    - **Note**: GEBCO is one of these DTM providers, and has been directly used by the iAOS Showcase application led by T6.8 (Baffin Bay Temperature Maps)
- Survey tracks/polygons offer details on the conduct of bathymetric surveys.
- Land geography and topography derived from OpenStreetMap
- Geographic names on land is a customized layer of geographic names derived from OpenStreetMap.
- Coastline is also derived from OpenStreetMap. It is a highly detailed coastline and a topographic representation. The latest version covers the entire coastline of Europe.
- Underwater features is a layer provided by GEBCO (Undersea Features).
- Wrecks is provided by OceanWise (using the UKHO Wrecks database as source).
- EMODnet Bathymetry World Base Layer (EBWBL) is a gridded representation of worldwide bathymetric and topographic coverage adapted for a better representation of seabed morphological features.

2.3. Integrating data Processing services

Specific challenges	Return of experience
Support software developers to make use of iAOS resources for the development of new applications	iAOS support activities for software application development have been delivered either as tutorials or as reusable software libraries. Three types of applications have been developed during the INTAROS Project with that support: stand-

	alone Jupyter Notebook applications, stand-alone R software applications (dedicated to the use of the RGeostats library) and data processing services (online applications delivered as-a-Service)

We present hereafter the content of the software repositories populated over the course of the INTAROS project.

A summary of the developed iAOS services is provided on the INTAROS Portal "services" page https://portal-intaros.nersc.no/pages/services. This page is maintained and continuously evolved by NERSC.

First, three groups of software repositories, with all content made publicly available on the INTAROS GitHub community, consist of:

- tutorials and workshop material,
- data processing applications,
- and software libraries.

This INTAROS GitHub community is well referenced be web search engines and can be easily discovered by web users performing a search on the leading search engines.

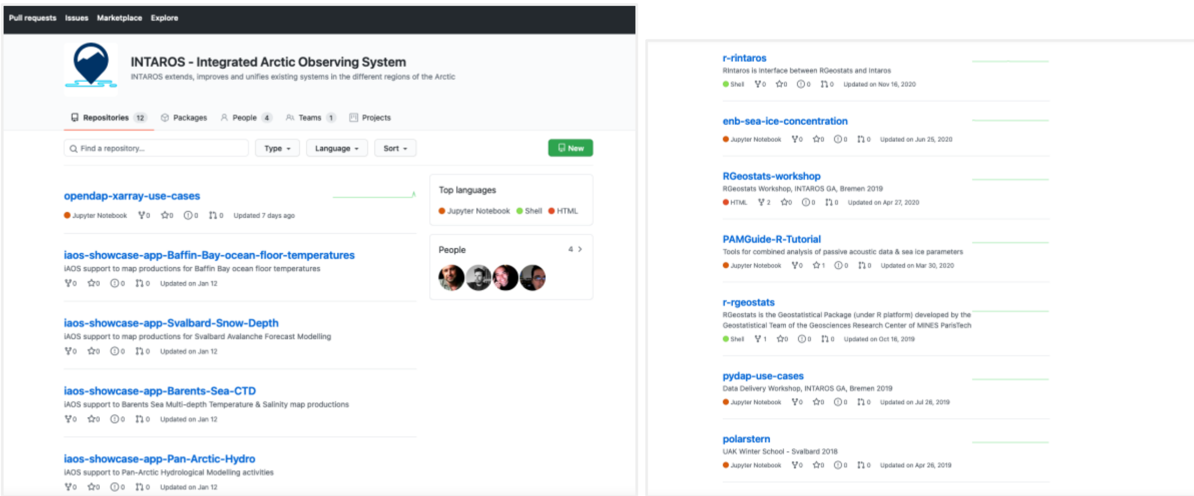

Figure 4. Overview of the INTAROS GitHub community contents

And then further below, a fourth group about applications made available online "as-a-Service", consisting in Ellip Workflows data processing services, ran onto Cloud Computing production clusters in order to deliver data products to the iAOS.

2.3.1. Tutorials and workshop material

Specific challenges	Return of experience
Train the iAOS community with guidelines and practical code examples making use of the iAOS data and processing services	Activities addressing this challenge have been conducted regularly over the project lifetime. The audience was mainly project partners, but a broader audience have also been engaged, especially for the UAK Winter School - Earth Observation module held

	on 02 – 07 December 2018 at UNIS, Longyearbyen, Svalbard
Consolidate the footprint of the iAOS architecture and technical choices in the Arctic science community by disseminating easy to use demo applications	Activities addressing this challenge have been conducted over the second half of the project lifecycle. The audience was mainly project partners but the provided resources being public, the generated results can support the iAOS sustainability goals and also, for the materials related to OPeNDAP, the stated goals for the partner Terradue in the INTAROS exploitation plan, i.e. in support of Terradue's business development activities towards organisations and scientists conducting applied research over the Arctic region.

In the following sections, we review the individual software repositories providing "Tutorials and workshop material" resources, shared via the INTAROS GitHub community during the INTAROS project. The topics are covering the use of OPeNDAP protocols from Python-based client applications, the use of the RGeostats library, the use of other R-based software libraries (PAMGuide), and the use of Satellite earth observations for Arctic-related software applications (snow-ice classification, glacier velocity, …).

2.3.1.1. opendap-xarray-use-cases

Specific challenges	Return of experience
Introduce concepts for the filtering and plotting of CTD data structures.	The learning curve was quite high in preparing this material, due to spreaded online documentation about CTD data structures, and the related concepts, e.g. as defined by the ship at sea acquisition modes (depth index, ...).
Present basic OPeNDAP functions and highlight efficient, lightweight queries to retrieve dataset descriptive information and perform initial data analysis to understand a dataset.	The learning curve was also quite high due to very spreaded online documentation about the OPeNDAP Data Model and the way it can be adapted to multiple types of datasets. Also the usage of the specific DDS (Dataset Descriptor Structure), while very powerful for an efficient pre-analysis of a dataset, is sparsely documented.
Illustrate the capabilities of Python Xarray data structures for in-memory data manipulation and plotting.	Very good and stable Python framework on Jupyter Notebook, and expressivity of xarray for data filtering and data structures manipulation, allowing to easily build a demonstration in a flexible way.
Provide a reference application (Jupyter Notebook) to help software developers to create a visual inspection of dataset contents based on different dataset variables and dimensions selection, in order to validate the definition of DAP queries that will perform similar data filtering, but server side.	Very good and stable Python framework on Jupyter Notebook, allowing to easily build a demonstration in a flexible way, with easy to use plotting functions operating on top of xarray structures.
Provide a demo application (Jupyter Notebook) to help software developers to create OPeNDAP query	The learning curve was quite high due to spreaded online documentation about the differences

filters that will ensure efficient data delivery from OPeNDAP servers to client applications, based on users interest and choices for specific dataset variables and dimensions.	between the DAP2 and the more recent DAP4 protocols, their authorised query strings, and their response behavior depending on the datasets being served.

Repository: https://github.com/ec-intaros/opendap-xarray-use-cases

This software repository provides an overview, as well as practical examples, to access and analyse a subset of NetCDF data served via the OPeNDAP protocol, and to manipulate the retrieved data as part of a powerful data structure "xarray".

A test dataset is exploited here. It was selected out of available CTD campaigns, involving five scientific ships (also called "Platforms"), and collected in the year 2003.

This data subset has been prepared and uploaded on the iAOS OPeNDAP servers (Hyrax server and Thredds Data Server) hosted on Terradue Cloud Platform, where it can be accessed directly by application developers, for them to test and validate their data access query filters, in particular when it comes to optimizing such filtering options with regard to a particular data processing service expected inputs.

The Reference Notebook software provided to support developers is constructed as follows:

- Retrieval of DDS information: the OPeNDAP protocol defines a lightweight endpoint and payload, allowing an application developer to interact with the server only with the data dimensions information, that can be accessed through the Dataset Descriptor Structure (DDS). This allows us to understand the size of the datasets in an efficient way, without downloading all data into local memory. The dimensions under consideration for these NetCDF files are: 'TIME', 'LATITUDE', 'LONGITUDE', 'DEPTH', 'POSITION'. Extracted from an xarray data structure and projected into a 2-dimensional Dataframe structure, the LATITUDE and LONGITUDE (assembled per their index in the data structure) allow efficient and quick data visualisation on a 2D map projection plot of the full dataset ("Load and Plot Positions only"), and to apply some useful filtering for a data scientist to understand the dataset contents per index position, sampling time and coordinates and Platform (ship) identifier.
- Visual Analysis: Load and Plot Positions to visualise the geographic positions of the data points for each platform, and to perform some filtering operations based on locations and time queries. This is possible using only the necessary information retrieved from the DDS.
- Processing: Load and Plot all needed Data (Variables and their Attributes) and creation of reference plots based on a full dataset sample as a source stored in memory. The goal is to enable iAOS users to extract a time series of CTD measurements for a selected time period in the Arctic. The data in the time series can be used to generate different statistics for the water column or sea bottom, in particular when ingested as input to a RGeostats module.

The DAP Notebook software provided to support developers is constructed as follows:

- Retrieval of DDS information: the OPeNDAP protocol defines a lightweight endpoint and payload, allowing an application developer to interact with the server only with

the data dimensions information, that can be accessed through the Dataset Descriptor Structure (DDS).
- DAP Queries definition with Positions and Variables.
- Efficient Data load operations based on the defined DAP queries, populating xarray structures only with the necessary dataset information needed for a particular application (e.g. use geostatistics to interpolate temperature records for a specific area, period of the year and depth levels).

2.3.1.2. pydap-use-cases

Specific challenges	Return of experience
Introduce concepts for the filtering and plotting of Grid data structures.	The learning curve was quite high in preparing this material, due to very spreaded online documentation about PyDAP and ncML

Repository: https://github.com/ec-intaros/pydap-use-cases

This repository contains a set of simple Jupyter Notebooks showing examples of the usage of the "Pydap" Python library, in particular using its client functionalities, and providing an introduction to the use on ncML to support OPeNDAP dynamic aggregations of datasets.
- Setup of the Pydap Python library
- Accessing different dataset formats via the DAP protocol
- Using the DAP Constraint Expressions
- Using the Hyrax geogrid() Server Side Processing Function
- Hyrax Dynamic Aggregation function via NcML

2.3.1.3. RGeostats-workshop

Specific challenges	Return of experience
Introduce the concepts of Geostatistics and present user cases for implementation using the iAOS tools and services	There was a good attendance at the workshop, around 20 participants, and the course was a successful innovation as it introduced for the first time many iAOS features.

Repository: https://github.com/ec-intaros/RGeostats-workshop

This repository holds the material created for the RGeostats Workshop during the INTAROS GA, Bremen 2019. It provides Notebooks and slides, the workshop agenda being:
- Creating iAOS processing services
- Geostatistics and RGeostats
- Ellip Notebooks applications using RGeostats
- Ellip Workflows applications using RGeostats
- Case Study using IMR CTD data
- Geostatistics course and exercises

2.3.1.4. PAMGuide-R-Tutorial

Specific challenges	Return of experience
Reduce processing time and reduce the risk for loss of data due to file format conversion issues when accessing acoustic data records.	The generation of Noise level statistics and spectrograms was simplified for scientific users.

Repository: https://github.com/ec-intaros/PAMGuide-R-Tutorial

This repository presents a slightly modified version of the PAMGuide R software. A Jupyter Notebook is included to illustrate how to use the tool. It also shows how to access data available on a OPeNDAP server. It accompanies the work presented in the INTAROS deliverable D5.7 Processing services v1.1 for the iAOS Service "Characterization of Passive Acoustic Data (C-PAD), aiming at enabling users to analyse and present passive acoustic data, to support detection and classification of different sources of noise pollution in the ocean.

The primary modifications that have been made are 1) adding support for vector input, and 2) adding support for Jupyter Notebooks. The original version of PAMGuide only supports WAV-files (audio files) as input. However, passive acoustic data is often stored in NetCDF or other data files (e.g. MATLAB files). These sample values can now directly be processed, rather than converting this data into audio files. This reduces processing time and reduces the risk for loss of data.

2.3.1.5. polarstern

Specific challenges	Return of experience
Introduce concepts and tools for using satellite earth observation data as input to Arctic-related software applications	The winter school was well attended and collaborations have been established with some of the participants after the event.

Repository: https://github.com/ec-intaros/polarstern

This repository provides five Jupyter Notebook applications developed in support of the UAK research school, held on 02-07 December 2018 at UNIS, Longyearbyen, Svalbard. It is a research school on cross-disciplinary science in the Arctic and collaboration with local communities. It was organised by the Nansen Environmental and Remote Sensing Center under the project Useful Arctic Knowledge: partnership for research and education (UAK) funded by the INTPART programme 2018-2020.

The Jupyter Notebook applications are 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.

2.3.2. Data processing applications

2.3.2.1. enb-sea-ice-concentration

Specific challenges	Return of experience
Standardise the process of product ingestion and time series generation for selected regions and timeframes, and communicate the results in a simple to use interactive application.	The generation of time series out of sea ice concentration maps was eased for the scientific users, based on OPeNDAP and NetCDF standards.

Repository: https://github.com/ec-intaros/enb-sea-ice-concentration

This tool uses daily products from the Copernicus Marine Environment Monitoring Service (CMEMS) to generate monthly mean values of sea ice concentration for the Svalbard region. The daily products are accessed using the OPeNDAP protocol and the resulting sea ice statistics is stored in a NetCDF file. It accompanies the work presented in the INTAROS deliverable D5.7 Processing services v1.1 for the iAOS Service “Time Series of Sea Ice Concentration (TS-SIC)”, for users to extract a time series of sea ice concentration maps for a selected time period in the Fram Strait and North of Svalbard arctic regions.

2.3.2.2. iaos-showcase-apps

Specific challenges	Return of experience
Provide simple Jupyter Notebook applications demonstrating the use of remote sensing data and in situ observations delivered through the iAOS, from a variety of platforms and geographical scales and locations.	Incorporation of Geostatistics techniques was an important effort, due to difficulties in identifying the input datasets most relevant for a given showcase, and then in assessing their internal records structures and understanding their quality. The generation of synthesis Jupyter Notebooks communicating about the developed iAOS Showcase applications was therefore impacted by all this very demanding preparatory work.

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

Note: This repository features two Jupyter Notebook applications illustrating the data processing functions available for searching an OPeNDAP server hosting CTD datasets and applying Geostatistics to the referenced dataset (presented in this document as part of the section "T6.2 with IMR: Barents Sea Multi-depth Temperature & Salinity Maps").

This is a more focused and operational version of the Notebook file already presented in the section above "opendap-xarray-use-cases", as it refers to additional work performed by Terradue on the user of OPeNDAP servers, and the work performed by ARMINES presented in the INTAROS deliverable D5.6 "Geostatistical library for iAOS V1".

2.3.3. Software libraries

2.3.3.1. r-rintaros

Specific challenges	Return of experience
Improve the accessibility to geostatistical methods for the INTAROS community, and deliver a ready-to-be installed R software package.	The iAOS was used to provide Cloud resources to the project partner ARMINES, along with guidance from Terradue, to establish an automated process for the build of an anaconda software package (providing users with a simple software installation procedure from an online repository).

Repository: https://github.com/ec-intaros/r-rintaros

This repository provides the r-rintaros software library as an Anaconda package. RIntaros is an interface between RGeostats and the INTAROS-specific applications needs identified during the project, provided in order to improve the accessibility to geostatistical methods for the INTAROS community.
It incorporates:

- Utilities for processing IMR CTD data (reading according to specific formats, date conversion, selection based on time or depth intervals).
- Specific workflows for:
  - Basic statistics benefiting for spatial coarse gridding or analysis of time series;
  - Modeling the spatial structure of target variable(s);
  - Estimation in 2-D (by vertical layers for example) or in 3-D;
  - Blind test or Cross-validation.

The outcome is a standard online repository (on anaconda.org) holding the software package, for ease of installation on Cloud Computing environments such as defined for iAOS.

2.3.3.2. r-rgeostats

Specific challenges	Return of experience
Improve the accessibility to geostatistical methods for the INTAROS community, and deliver a ready-to-	The iAOS was used to provide dedicated Cloud resources to the project partner ARMINES, along

be installed R software package.	with guidance from Terradue, to establish an automated process for the build of an anaconda software package (providing users with a simple software installation procedure from an online repository).

Repository: https://github.com/ec-intaros/r-rgeostats

This repository provides the r-rgeostats software library as an Anaconda package. RGeostats is the Geostatistical Package (under the R platform) developed by the Geostatistical Team of the Geosciences Research Center of MINES ParisTech. More information at http://rgeostats.free.fr.

The outcome is a standard online repository (on anaconda.org) holding the software package, for ease of installation on Cloud Computing environments such as defined for iAOS.

2.3.4. Ellip Workflows data processing services

2.3.4.1. From collaboration with the EC H2020 project NextGEOSS

Specific challenges	Return of experience
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.	Two processing services were integrated by NERSC with the Ellip Workflows solution as part of the H2020 NextGEOSS project, deployed in production on Cloud Computing resources. The resulting data products were retrieved by NERSC and published on the NERSC TDS (THREDDS Data Server) where they are available for use in INTAROS as well as other research projects (featured on the INTAROS Portal home page and described in the services page https://portal-intaros.nersc.no/pages/services).

Two data processing services powered by Terradue's Ellip Solutions have been implemented by NERSC in collaboration with the H2020 NextGEOSS project:

- Sea ice classification service
- Sea ice drift service

These two services are making use of the Cloud Computing resources funded by the NextGEOSS project on the EGI.eu Federated Cloud, and delivering data products for exploitation as part of the INTAROS tasks.

NERSC staff conducted the service development, testing and packaging for deployment; Terradue staff deployed the service on the selected EGI.eu Cloud Provider, once the application was validated and packaged. Both processing services have transferred output files from the processing campaigns run on the NextGEOSS-funded Cloud Production environment, back to NERSC for publishing through the OPeNDAP data server at NERSC for exploitation as part of INTAROS.

For the sea ice classification service, two processing campaigns have been ran, 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-September 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.

2.3.4.2. From collaborations within INTAROS

Specific challenges	Return of experience
Run of complex, compute-intensive iAOS data processing chains (processing of CTD data) in order to generate value-added products supporting the INTAROS tasks.	A processing service was integrated by ARMINES with the Ellip Workflows solution, deployed in production on Cloud Computing resources. The resulting data products have not been fully satisfactory because of an identified issue with the initial geostatistical spatial analysis performed in 2018. The proposed follow-up is to build a unique spatio-temporal variogram model for Temperature (use of a single global variogram model for each computing node in charge of each kriging tile), taking into account co-variables like (salinity, bathymetry...), and potentially handling currents nonstationarities and distance to coastline.

A data processing service powered by Terradue's Ellip Solutions has been implemented by ARMINES as part of the INTAROS WP5 activities. It was established in collaboration with the project partner IMR, providing the iAOS with a first online version of an OPeNDAP data server and delivering the IMR CTD datasets (temperature, salinity and conductivity in the North Sea).

The service is implementing an RGeostats-based data interpolation application (interpolating scattered in situ observations of ocean temperature and salinity to gridded fields), based on the iAOS cloud platform APIs which empowers the application with interoperable data access mechanism (based on OpenDAP), scalable data processing capabilities (based on Hadoop MapReduce) and standard processing invocation and results retrieval (based on OGC WPS specifications), for integration with the iAOS Portal of other Geobrowser applications.

While the technical result was satisfying in terms of demonstration of Ellip-powered applications, the RGeostats capabilities involved have been limited by the difficulties to retrieve and analyse data from the provided OPeNDAP server. Several improvements could be identified and a tentative roadmap established on this basis where Terradue would hand-over part of the technical work, releasing ARMINES from that aspect, so they can focus on the RGeostats modelling effort.

2.4. iAOS support to selected WP6 showcase applications

2.4.1. Overview

Specific challenges	Return of experience
Integrate remote sensing data and in situ observations delivered through the iAOS, from a variety of platforms and geographical scales and locations. Incorporate data into analysis and modelling systems, including physical and ecological process models, climate models and forecast methods, providing support for better products to key societal areas.	A dedicated WP5 and WP6 collaboration was established, through a selection of iAOS Showcase applications, which could make use of the iAOS tools and services for data access, data processing and data products dissemination.

WP6 is integrating 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, is providing support for better products to key societal areas. A dedicated WP5 and WP6 collaboration was established in this matter, through a selection of iAOS Showcase applications that we present hereafter from the perspective of the delivered input data and of the gathering of the data products generated from these applications.

For each showcase, the WP5 coordination and support actions are covering the following aspects, that have a direct benefit for the selected tasks within WP6:

- Definition of the overall need for data integration and management, as well as the technical use cases and resources to be involved;
- Integration of data sources in preparation of the iAOS Showcase applications integration, and description of the data products generated from the showcases;
- Use of Geostatistics by exploiting specific RGeostats / RIntaros software developments, and by developing methods and related ad-hoc (if/when needed) RGeostats software libraries, in support of the iAOS Showcase applications data exploitation goals;
- Use of iAOS platform services for the iAOS Showcase applications integration, in particular reusable Jupyter Notebooks in order to interactively run geostatistical tools on the iAOS data sources needed by the iAOS Showcase applications.

The collaboration took place over the project phases, with milestone events as follows:

- **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.
- **Remote, 2021** - Peer-to-Peer workshops presenting the iAOS support to actual and potential users of each selected Application.

From the perspective of Data integration from existing repositories, the WP5 support to the showcases has been focused onto the following objectives:

- Guarantee a good data availability
- Leverage the iAOS services (one-stop shop portal & data catalogue) to guide on the selection of most relevant datasets for the setup and elaboration of scientific applications over the Arctic
- Show the data flow from a source to the application developer and its end users
- Help to implement a smooth data flow between systems.

We present hereafter the selected data sources, data exploitation tasks, delivered results and remaining improvements identified by the end of the INTAROS project, for selected WP6 Showcase applications carried out in collaboration with WP5 by the project tasks T6.1 (with SMHI) for Pan-Arctic Hydrological Modelling, T6.2 (with IMR) for Barents Sea Multi-depth Temperature & Salinity Maps, T6.4 (with FMI) Maps for Svalbard Avalanche Forecast Modelling and T6.8 (with Aarhus University) for Baffin Bay Bottom Temperature Maps.

2.4.2. For T6.1 with SMHI: Pan-Arctic Hydrological Modelling

Specific challenges	Return of experience
Improve predictions of spring floods, river ice breakup and freshwater flow to the Arctic Ocean. And support the integration of river discharge into the ocean.	The coordination with WP6 partners on the input data selection was smooth and efficient.. The main effort was put by T6.1 on producing the expected data resources for climate model initialisation: for checking the modelling initial conditions, for the model analysis of current conditions and the model forecast.

Objectives

- Have the “observational” data available for search and download from the iAOS Portal
- Have the Arctic-HYPE produced at SMHI and provide the data as open data from SMHI repositories:
  - Daily analyses of last 60 days
  - Medium range forecast of coming 10 days
- Improve predictions of spring floods, river ice breakup and freshwater flow to Arctic Ocean, cf.INTAROS D6.1 Climate model initialization v1.4

Data sources

- River discharge data from the Arctic Hydrological Cycle Observing system (Arctic-HYCOS) - assessed and enhanced in INTAROS WP2 https://catalog-intaros.nersc.no/dataset/arctic-hycos-hydrological-data/
- HydroGFD v3 temperature and precipitation data
- ECMWF deterministic medium range weather forecasts

Data exploitation

- Implement HYCOS pre-processing (both archive of quality controlled data with 4months/2years lag, or provisional datasets)
    - In-house server at SMHI
    - Cloud-based, using Ellip, to compare
- Schedule HYCOS pre-processing operations to be made daily at a certain time
- Setup OpenDAP server for publishing the Arctic-HYPE model results

Delivered results

- OpenDAP server publishing Pan-arctic hydrological model Arctic-HYPE results provided by SMHI http://opendata-download.smhi.se/opendap/catalog/catalog.html

iAOS Showcase (demo)	Use of Arctic-HYCOS sources (multi-provider)	Use of Arctic-HYPE (SMHI) outputs (per station ID, with 1 dimension time)
Flood forecasting use cases Select a point, or some AOI (in Russia MPI in Yakutsk - L. Lebedeva)	For checking the initial conditions	For checking the model analysis of current condition and the model forecast
Entire freshwater inflow to the ocean (in collaboration with IMR - R.Hordoir)	To integrate discharge into the ocean from observations (only represent 60% draining land, see WP2 deliverables).	To integrate model analysis/predictions on river discharge into the ocean on user defined resolution

Foreseen improvements

Results:

- The approach was successful with the new model running for the Yakutian case study: https://hype-eras.org/forecasts/

2.4.3. For T6.2 with IMR: Barents Sea Multi-depth Temperature & Salinity Maps

Specific challenges	Return of experience

Generate temperature and salinity fields for the modelling of Arctic Ocean biogeochemistry and the validation of climate model projections (NorCMP).	The main effort was put on understanding the input datasets for Geostatistical modelling. Errors and gaps in the data have been identified at multiple points in time. Some communication issues also occurred, slowing the process and somehow the dynamic of the activity. The main effort had to be put by T5.4 on the Geostatistics modelisation of spatial behavior for the Temperature and Salinity.

Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating maps from CTD datasets
- Generate temperature and salinity fields for:
    - modelling of Arctic Ocean biogeochemistry
    - validation of climate model projections (NorCMP)
- Build a Web Processing Service for iAOS

Data sources

- Institute of Marine Research (IMR) center of Norway, 7 research vessels
- collecting conductivity, temperature, depth (CTD) data in the North Sea.
    - Acquisitions between 7th January 1995 and 29th November 2016.
    - 5.5 billion of samples measured over 63 500 positions (vertical profiles).
- All files are freely available on an OPeNDAP server (operational and test instances):
  http://opendap1.nodc.no/opendap/physics/point/yearly/contents.html
  http://opendap1-test.nodc.no/opendap/physics/point/yearly/contents.html
- NetCDF files (one file by year and per vessel). Coordinates are in degrees (Long/Lat) and the timestamp is the number of minutes since the 1st January 1950. The whole dataset volume is 880 GB.

Data exploitation

- Use of the iAOS OpenDAP server at NODC
- Exploratory data analysis and variography
- Modelisation of spatial behavior for Temperature and Salinity

Delivered results

Standalone solution (R software)
- Map productions per run
    - Base maps
    - Average per cell
    - Cross-validation (blind test) maps
    - Estimation (Temperature / Salinity) and corresponding uncertainty maps

Solution as-a-Service (Cloud software)

- Map productions per WPS run
  - On-demand, self performed by each user from the Portal
  - Split tiles for large areas

Foreseen improvements

Approach:

- Improve the initial geostatistical spatial analysis performed in 2018
  - Build a unique spatio-temporal variogram model for Temperature
  - Take into account co-variables like (salinity, bathymetry...)
  - Handle currents nonstationarities and distance to coastline

Standalone solution (R software)

- Promote the Jupyter Notebook and R Markdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

Solution as-a-Service (Cloud software)

- Promote the use of the Ellip Solutions for parallel processing https://gitlab.com/ec-intaros/dcs-imr-estim
- Consider the use of a single global variogram model for each computing node in charge of each kriging tile

2.4.4. For T6.4 with FMI: Maps for Svalbard Avalanche Forecast Modelling

Specific challenges	Return of experience
Generate snow depth maps at regular time intervals as input for avalanche forecast model.	Input data selection was efficient, but still had to face a team communication issue, resolved early 2021, on the availability of wind speed within stations data. This has slowed the progress of the Showcase for a few months. The main effort had to be put by T5.4 on the Geostatistics modelisation of spatial and temporal behavior for snow depth through co-variables (temperature, wind speed by class of wind direction).

Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating maps from snow stations, arome model output and terrain model
- Generate snow depth maps at regular time intervals as input for avalanche forecast model

Data sources

- NMI Frost API historical weather and climate data stations (selected files on shared Drive)

- NMI arome model: https://thredds.met.no/thredds/catalog/aromearcticarchive/catalog.html
- Norwegian Polar Institute Svalbard Terrain Model: https://doi.org/10.21334/npolar.2014.dce53a47

Data exploitation

- Exploratory data analysis and variography
- Handle the different spatial distributions and resolutions of the data (“support”)
- Modelisation of spatial and temporal behavior for snow depth through co-variables (temperature, wind speed by class of wind direction)

Delivered results

Pre-analysis of the data:

- Few stations: temporal series of snow thickness measured at short time steps
- Arome models: various maps covering the whole area, every 6 hours, on a large scale grid (incl. snow thickness derived from model)
- Several co-variables

Processing:

- Regularization of the station time series by averaging over 6 hours
- Correlation (space-time) of snow depth variable with arome model output
- Estimation using both information sources (with relevant co-variables) over a small scale grid, at regular 6 hours intervals

Foreseen improvements

Approach:

- Finalize the showcase with a proof-of-concept
    - Create snow depth variogram models (with relevant co-variables) for each class of wind orientation
    - Generate snow depth map
        - for a given date and time interval (6 hours)
        - accounting for a global wind orientation

Standalone solution (R software):

- Promote the Jupyter Notebook and R Markdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

2.4.5. For T6.8 with Aarhus: Baffin Bay Bottom Temperature Maps

Specific challenges	Return of experience
Generate temperature fields at the bottom of the ocean in support of the analysis of long term global	The main effort was put on understanding the input datasets for Geostatistical modelling. Errors and gaps

warming influence, as well as potentially the analysis of the fish stock correlation to bottom temperature.	in the data have been identified at multiple points in time. Some communication issues also occurred, slowing the process and somehow the dynamic of the activity. The main effort had to be put by T5.4 on the modelisation of spatial and temporal behavior for ocean floor temperature through Bathymetry co-variable.

Objectives

- Use the Geostatistical Library (RIntaros / RGeostats) and build the R software for interpolating ocean floor temperature maps from CTD and Trawl datasets
- Generate temperature fields at bottom of the ocean in support of:
    - Analysis of long term global warming influence
    - Analysis of the fish stock correlation to bottom temperature

Data sources

CTDs (ICES) ● 1977 to 2017 ● 3700 vertical profiles ● 1.34M samples ● 1 CSV file (95Mo)	Trawls (GINR) ● 1988 to 2016 ● Catches near seafloor only ● 51K samples ● 1 CSV file (11Mo)
Bottles (ICES) ● 1960 to 2017 ● 7800 vertical profiles ● 167K samples ● 1 CSV file (11Mo)	Bathymetry (Gebco) ● Grid lag = 1/250 degree ● Grid size = 10320x4560 nodes ● 47M samples ● 1 NetCDF file (94Mo)

Additional data sources provided by Aarhus / Mikael Sejr (Nov. 2020):

- Better quality of arctic bathymetry
  https://www.nature.com/articles/s41597-020-0520-9
- World ocean database (Temp fields)
  https://www.nodc.noaa.gov/OC5/WOD/pr_wod.html
- Global temperature and salinity profile program (data?)
  https://www.nodc.noaa.gov/GTSPP
- NASA project in Greenland (new CTD)
  https://omg.jpl.nasa.gov/portal/browse/OMGEV-AXCTD/

Data exploitation

- Explanatory data analysis and variography
- Local and global temperature evolution analysis
- Modelisation of spatial and temporal behavior for ocean floor temperature through Bathymetry co-variable

Delivered results

- Temporal evolution (global and local)
    - Between 1960 and 2015, T°C has gained 1°C (around 1995)
    - Lower T°C values around 2008 have been recorded
- Map productions
    - Basemap of data
    - Bottom temperature estimation by year and its standard deviation
- Time series of average temperature by region (Lat/Depth)

Foreseen improvements

Approach:
- Improve first kriging estimation
    - Reduce the estimation error (currently around 0.7°C)
    - Reduce estimation smoothing
- Next improvements
    - Multi-directional variography and zonal anisotropies
    - Local cross-validation with additional abyssal (or new) data
    - Bathymetry as external drift using non linear regression
    - Salinity as co-variable to be studied

Standalone solution (R software):
- Promote the Jupyter Notebook and R Markdown scripts on iAOS and their documentation on GitHub
- Build a flyer for addressing the reusability of the solution

2.4.6. Status summary on the iAOS Showcase applications

Different maturity levels have been reached by the selected iAOS Showcase applications, in demonstrating the integration of remote sensing data and in situ observations, and the use of such datasets into analysis and modelling systems.

For **T6.1 with SMHI**, Pan-Arctic Hydrological Modelling, the coordination with WP6 partners on the input data selection was smooth and efficient, with the goal to improve predictions of spring floods, river ice breakup and freshwater flow to the Arctic Ocean, as well as for supporting the integration of river discharge into the ocean.
The main effort was put by T6.1 on producing the expected data resources for climate model initialisation: for checking the modelling initial conditions, for the model analysis of current conditions and the model forecast.

For **T6.4 with FMI**, Maps for Svalbard Avalanche Forecast Modelling, the coordination with the WP6 partners on input data selection was efficient (NMI Frost API historical weather and climate data stations, NMI arome model, Norwegian Polar Institute Svalbard Terrain Model), but still the team had to face a communication issue, resolved early 2021, on the availability of wind speed within stations data, which has slowed the progress of the activity for a few months. The main effort had to be put by T5.4 on the Geostatistics modelisation of spatial and temporal behavior for snow depth through co-variables (temperature, wind speed by class of wind direction).

For **T6.2 with IMR**, Barents Sea Multi-depth Temperature & Salinity Maps and **T6.8 with Aarhus**, Baffin Bay Bottom Temperature Maps, the main effort was put on understanding the input datasets for Geostatistical modelling. Errors and gaps in the data have been identified at multiple points in time. Some communication issues also occurred, slowing the process and somehow the dynamic of the activity. The main effort had to be put by T5.4 on the Geostatistics modelisation of spatial behavior for the Temperature and Salinity (for T6.2) and on the modelisation of spatial and temporal behavior for ocean floor temperature through Bathymetry co-variable (for T6.8).

3. Establishing a sustainable pan-Arctic iAOS cloud infrastructure

We discuss hereafter work items for the definition of a sustainable pan-Arctic iAOS cloud infrastructure, that are derived from lessons learnt out of the WP5 activities presented above in this report, and that are formulated as extended recommendations for addressing long-term infrastructure properties going beyond the sole technical work performed during the INTAROS project timeframe. These recommendations are structured within strategic and organisational issue categories, related to the infrastructure cost assessment, legal dimensions, trust building, privacy policies, security policies and usability criteria. Each set of recommendations is referencing practical iAOS activities conducted for the INTAROS project by the WP5 team.

3.1. Cost

The overall cost of maintenance and operation, along with opportunity costs for evolutions, of a distributed system of systems such as the iAOS, is nearly impossible to assess without a directed, independent effort based on financial audit approaches.

A generic approach to identify costs at a finer grain level, and communicate these to third-parties or stakeholders, can be nevertheless discussed as follows:

- resources exposed as-a-service (meaning, with a relatively high level of operational maturity) have by definition their own business model and sustainability plan, and the cost estimate at the user entry level for a given subscriber can be based on the public pricing policies exposed by each of these well matured online services. This is typically the model developed by Cloud Computing operators, with a significant range of offers developed by this industry, amongst which the most established ones are Infrastructure-as-a-Service (IaaS), Platform-as-a-Service (PaaS), and Software-as-a-Service (SaaS).
- resources exposed as open source software for use by software application developers (software libraries, software frameworks and solutions, …) come at a cost of exploitation also depending on their maturity level (quality of the documentation, tutorials, community support on discussion forums, …) but not only, as other important factors will come into play such as: operational stability and robustness, community adoption (that can incur a cost of switching to another solution in case of lack of evolutions), and the intrinsic learning curve (including the potential need to hire specialists).
- resources exposed as internet-based free resources (especially for data repositories

that are not attached to particular terms of service) represent both an immediate cost reduction opportunity and a potential risk of extra cost on the data consuming applications side, in case of unilateral decisions involving changes in the provided service.

Let's also emphasize that the economic gains in terms of capital expenses (CAPEX) and operating/operational expenses (OPEX) for a customer organisation of cloud computing services can be huge, as large parts of corporate expertise and assets can be delegated to a service provider. Moreover, the scalability of cloud computing resources (for upscaling or downscaling depending on contextual needs) also represents a cost reduction factor for a customer organisation of cloud computing services, freed from a burden that is in this context delegated to the market (as a customer, flexibility to move to another Cloud Computing provider with more interesting conditions and capacity) and/or delegated to the service providers know-how to optimise and rationalise their service offer to users (as a cloud provider, capacity to optimise the match, and therefore the pricing, between their customers requirements and their virtualized cloud computing resources).

With the current maturity level of the iAOS cloud infrastructure, sustainability considerations with regards to costs have to be considered as part of a broader picture, not reduced to a static cost–benefit analysis effort, and having in scope transformative changes and socio-economic impacts. To the request of the European Commission, economists and specialists have shaped a framework to guide the future European Union Framework Programme for Research and Innovation, in which the definition of clear targeted missions, supported by a portfolio of projects and bottom-up experimentations, help address grand challenges pertaining to political agenda (Mission-Oriented Research & Innovation in the European Union - A problem-solving approach to fuel innovation-led growth, by Mariana MAZZUCATO). One important idea in this approach is that "Missions should not be achievable by a single development path, or by a single technology. They must be open to being addressed by different types of solutions''. The mission-based approach is defining the expected outcome, while the path to reach the outcome is to be 'discovered' on a bottom-up approach of multiple solutions assessed and adjusted along the way.

In this view, the iAOS cloud infrastructure has developed some valuable features over the course of the INTAROS project. In terms of a distributed system, it is freed from sectorial lock into a single technology or innovation path. It therefore provides a capacity to deploy and operate on-demand data processing services, powered by standard APIs, for a defined duration and a defined set of cloud computing providers. Overall, this guarantees organisations an easier path to experiment with new solutions, under the known cost conditions advertised by such Cloud computing providers.

3.2. Legal

The adoption and sustainability over time of a cloud infrastructure capable of federating a distributed system of systems of cloud providers, such as the iAOS cloud platform can heavily depend on the way it addresses some legal issues and/or opportunities:

- Open Data Sharing licenses: the adequacy of the infrastructure services to enable international collaborations and to support open data sharing or information sharing agreements (e.g. as a place supporting the access to research data underlying open

access scientific publications);
- Preparedness for investors, business partners: for example related to commercial exploitation of the by-products of research and development efforts, by providing support for embedded software licensing or data licensing, such as:
    - automation mechanisms to support network-compatible, digital license agreements, as done by creative commons,
    - automation mechanisms to support author attribution and allocation of unique identifiers such as Digital Object Identifiers (DOI),
    - automation mechanisms to support user subscriptions to online services, with well documented service access conditions, terms of use, demo tutorials, … that are easy for the users to self-assess and test (free trials, ...) and easy to revoke.

Cloud computing, both as a technology and as a business, presents specific challenges to the traditional bylaw protections to ensure the protection of corporate assets. The data centers or server farms delivering cloud computing capabilities provide a network-based model, with remote software and remote data access from the users software systems, which is a paradigm shift compared to on-site computing resources potentially connected only to user applications operated from a same private network.

Contracts, Service Level Agreements (SLA), Operational Level Agreements (OLA), terms of service and licensing agreements in this context have to provision adequate legal recourse and remedies, especially with regards to domestic and international laws. Intellectual property and GDPR are of course coming into play (see also next section) in relation to any user organisation concerned about privacy and data protection.

The readiness for such legal aspects related to the access to scientific inputs and outputs (data, software) under open science guidelines is an opportunity for research organisations and their partners to:
- Promote more research activities from a same data collection or software development;
- Reduce duplication of effort and optimise the costs of creating, transferring and exploiting knowledge;
- Multiply opportunities to develop synergies;
- Improve their own effectiveness and productivity.

The iAOS cloud infrastructure demonstrated already some readiness levels in this matter (user access to Platform-as-a-Service resources, consolidation of scientific data under trustable data repositories exposing machine-to-machine interfaces) and these efforts should be strengthened, as the organisations contributing to such operating modes certainly still lack from a coordinated capacity development (clearly shared guidelines, under a shared framework), in order to be ready for strategic developments around the new legal landscape attached to Cloud Computing and open science. This is one of the key objectives of the European Open Science Cloud that we discuss hereafter in the context of Trust aspects.

3.3. Trust

Trust about a distributed cloud infrastructure will be generally defined by the perception and/or the reputation of a few representative service providers, which is typically the case for cloud computing.

There are some best practices and mechanisms for establishing trust in the context of Cloud Computing, either based on evidence, certification, or self-service assessment offered to users. Trust is also dependent on key operational capabilities in terms of risk mitigation and recovery mechanisms in case of outages.

For example, the NIST Cloud Computing Reference Architecture identifies Cloud Computing technical use cases in terms of “Centralized vs. Distributed” deployment models, and “Within vs. Crossing” Trust Boundaries characteristics of these deployments, and it references cloud brokers and cloud auditors as entities entitled to conduct assessment of cloud computing services with regards to trust issues. It also discusses the Cloud Computing Standards for Interoperability and in particular their importance for data portability and workload (software runtime) portability across Cloud providers.

On this side, while the Copernicus programme is addressing, in a large trust-building effort, the challenge of cloud computing in order to support the valorisation of investments done into earth observation systems, in particular via the funding of the Data and Information Access Services (DIAS), this effort remains part of a multitude of solutions, not yet harmonized or interoperable, that the European Open Science Cloud will seek to optimise, especially with regards to standardized application program interfaces (API) having the potential to reduce risks of vendor lock-in. Therefore, on the sole Cloud Computing front, much is still to be built in terms of trust.

The iAOS efforts, as defined as part of the INTAROS project, have been fully aligned from the start with the objective to deliver trustable Cloud Compute services (cf. section 2.1 “Deploying and operating the IAOS infrastructure” of this report, as well as section 1.4. “What is the iAOS?” and section 2. “Processing Service Integration & Deployment Guide” of the INTAROS deliverable D5.8 iAOS Platform and Tools V2, Revised 10 June 2020). For example, considering the use of Terradue’s Ellip Solutions as part of INTAROS, application portability from one Cloud provider to another was done by trained users (Operations Support team at Terradue) having a technical profile. This is not yet a high level function that can be operated from a dashboard. Work is still ongoing by Terradue to deliver it as part of the Ellip v2 solutions. Establishing a sustainable pan-Arctic iAOS cloud infrastructure shall build upon this INTAROS achievement, and make extensive use of it in order to promote both the iAOS vision for system of systems, and the iAOS cloud infrastructure readiness for capacity building in the scientific community and beyond, to other communities (government bodies & policy makers, commercial sector, …).

3.4. Privacy

The EU General Data Protection Regulation is about data privacy rights and principles for “natural persons” (data subjects). Under GDPR regulation translated into law, organizations collecting personal information (data controllers) as part of their businesses are held responsible for facilitating these “rights and principles”.

This is strongly impacting all software systems and infrastructures that operate user registration services, user authentication services and user communication services. It can have even more complex organisational and potentially legal impacts on such systems when the databases supporting these services are distributed and rely on interoperability protocols, such as single sign-on (SSO) where personal information initially registered on one system is accessed by third-party systems (in principle under the user explicit approval, and under the GDPR legal terms since May 25th, 2018 when the GDPR regulation was put into effect).

Transparency and communication: organisations collecting personal data (e.g. from users interacting with their services), they have to communicate specific information to these persons. These organisations have to explain how they process personal data in “a concise, transparent, intelligible and easily accessible form, using clear and plain language”, have to make it easy for people to make requests, and have to respond to those requests quickly and adequately.

Right of access and modification: data subjects have the right to know about the processing activities of an organisation controlling their personal data records. This includes clarity on the source of their personal data (user registration form, …), the purpose of processing such personal data, and the duration the data will be held. Moreover, individuals have a right to correct inaccurate or incomplete records of their personal data that an organisation is processing.

Right to restrict processing and right to object: data subjects have the right to request an organisation to temporarily change the way their personal data is processed (such as removing it temporarily from a website) if they believe the information is inaccurate, is being used illegally, or is no longer needed by the controller organisation for the claimed purpose.

Right to erasure: data subjects have the right to request that a controller organisation deletes any information about them.

Data portability: data subjects have the right to control who can access their personal information. To this end, personal data must be stored in a format that can be easily understood by data subjects, and therefore must have the potential to be easily shared across systems by these data subjects.

Establishing a sustainable pan-Arctic iAOS cloud platform that fully adheres to the state-of-the-art practices and regulations related to Privacy issues should not be seen as a remote objective. The main effort has generally to be allocated to procedures. For INTAROS, the collaborative setting between WP5 and WP6 , and in particular through the showcases, have shown the criticality of such procedures for any future iAOS operations involving co-development of services based on multi-provider datasets.

Hence, translating the world-leading GDPR regulation into practice is mainly a matter of organisational decisions and procedures implementation to ensure that a data controller organisation, or a pool of organisations contributing to a federated system, has the capability to react properly to individuals’ requests about the management of their user accounts (delete my account and account data), of their transactions on the system (remove me from that newsletter subscription), and of their data portability requests.

3.5. Security

Beyond the obvious security aspects related to cyberattacks and crises, tackled by the Cybersecurity Act (EU Agency for cybersecurity, ENISA) by establishing a cybersecurity certification framework for products and services, and which concerns primarily communication network and data center operators, the other key security concerns for an infrastructure such as the iAOS cloud infrastructure are to ensure the confidentiality, integrity, and availability of digital assets managed within information systems, and can be related to the prevention or mitigation of accidents, natural disasters and external loss of service, of unauthorized access to information through compromised user account authentication or authorization mechanisms, ...

In this view, security aspects for the iAOS cloud infrastructure addressed during the INTAROS project were related to the:

- Protection of user personal data from unauthorized access or monitoring. This includes identity management mechanisms such as the ones based on OpenID, for Single Sign On (SSO) capabilities from trusted Identity Provider services. This includes the ability for a user to make access to its personal data selectively available to others.
- Protection from unauthorized access to cloud computing infrastructure resources, using security domains that provide logical separations of user workloads running on the same physical server in a multi-tenant Cloud Computing environment.
- Note: there was no possible allocation of project resources onto related issues such as monitoring services of Web Portal traffic.

These are mainly security aspects related to the nature of the iAOS cloud infrastructure as a “federated services” infrastructure, where each provider is responsible for the core ICT security questions such as the ones related to cyberattacks, but where interoperability between these services can introduce vulnerabilities that have to be taken care off.

3.6. Usability

The usability of the iAOS cloud infrastructure covers a large set of information system features, and can involve a tremendous amount of tools and techniques to ensure and improve the usability of these information system features:

- Properties of the User interfaces, for example as defined in the ISO IEC 9126 standard “Software engineering — Product quality”, in terms of Understandability, Learnability, Operability, or Error handling design and implementation choices for the software applications running on the iAOS cloud infrastructure.
- Methods and tools to ensure satisfying levels of usability of the software systems:
    - Early focus on users requirements, activity analysis and user task modeling;
    - Iterative design involving application mockups, application prototypes and evaluation & usability inspections methods involving user panels;
    - Software validation procedures based on expert quantitative measurements of the software properties (e.g. checking software conformance against implementation standards such as User Interface design guidelines).

While it is a commonplace to consider that involving the foreseen system users themselves (e.g. scientists, decision-makers, operators) is part of the overall approach, it is sometimes overlooked that involving other levels of stakeholders (e.g. management teams in environmental agencies, local and national authorities, pan-Arctic organizations) is also important when addressing usability concerns. Indeed, these actors are also interested in the direct impacts of usability properties, potentially delivering organisational gains such as decreased training and support costs, increased productivity. Moreover, these actors can be driving forces to position and guide usability-related activities in a context of organisational changes (and change management) that will impact work procedures, and ultimately make the user tasks evolve.

This overall approach was addressed as part of INTAROS WP5 activities via the initial definition of tow evolutionary cycles (iAOS V1 and V2) each supported by the deliverables D5.1 - iAOS requirements and architectural design V1 and D5.5 - iAOS requirements and architectural design V2, each followed by implementation work (Platform tools and libraries, data integration / interoperability with remote data repositories and Portal services).

The iAOS cloud platform is certainly bringing some potential changes in the way research and development activities would be conducted in a distributed information system work environment, and the usability topic has to be approached mainly from that angle. The governance of the iAOS cloud platform has to be formulated in a transparent and well communicated set of policies, considering this is closely related to usability matters: governance is a necessary framework for establishing the guiding principles and goals, upon which usability-related activities can build, in order to prioritise activity analysis and user task modeling, to prioritise the subsequent specification of the expected user experience and to plan the iterative design and testing of the software features addressing these requirements.

4. Conclusion

We have presented in this report a synthesis of the cloud platform evolutions during the INTAROS project for improved iAOS cloud infrastructure capabilities. The synthesis is presented in terms of return of experience and challenges related to: (1) deploying and operating a distributed system of systems; (2) integrating remote data repositories and data processing services based on these Cloud Platform infrastructure services, and (3) defining user-oriented "iAOS showcase" applications powered by all these infrastructure resources. This was done to demonstrate the usefulness and functionality of the platform, and the benefit of enhanced integration of data from Arctic observing systems. As a result, the cloud platform offers services to access observations and derived parameters, including new observations from WP2-3-4, and it offers a new capacity of integration with geo-statistical methods for interpolation of spatiotemporal datasets, as well as services to store the generated datasets in iAOS enabled repositories.

For each group of iAOS cloud based assets generated, we have summarized the specific challenge that had to be addressed, and the lessons learnt from the generated results. These summaries indicate that earth observation data, in particular the Copernicus Sentinel products, have proven a good level of relevance and applicability in the iAOS cloud infrastructure context, with tools and platform services successfully demonstrated with the iAOS applications. At the UAK research school (December 2018 at UNIS, Longyearbyen, Svalbard), Jupyter Notebook applications were used to introduce EO data processing techniques. The students were introduced to applications using satellite earth observations and other data: 1) AIS data to discover relevant Sentinel-1 products, 2) Sentinel-1 for snow and ice classification, 3) offset tracking techniques on Sentinel-1 Level-1 Ground Range Detected (GRD) products to derive glacier velocity maps, 4) use of multi-year SAR data to study the seasonal dynamic of the snow melt patterns. From the INTAROS collaboration with the NextGEOSS project, where Sentinel-1 processing Sea ice classification products and Sea ice drift products have been generated and published in an open data repository by NERSC for exploitation in INTAROS.

Moreover, data standards, while being well established in the software systems involved within iAOS, are still lacking support to implementation, as some interoperability protocols (e.g. DAP2, DAP4) do still lack effective online support for software application developers. The learning curve is still high for an application integrator who would have to leverage all the aspects of an infrastructure such as the iAOS cloud infrastructure for efficient data access and reliable data quality assessments, high performance data processing, and easy results sharing. This challenge appeared quite from the start of the WP5 activities in INTAROS. Consequently, a number of tutorials and training assets, also reviewed in this report, have been produced. They are publicly shared on the INTAROS community on GitHub.

The objectives of the iAOS cloud infrastructure were to improve the integration and interoperability of existing observation systems to support the science community with enhanced capability to assess and predict Arctic environmental change. The approach was to demonstrate integration of several key variables of Arctic meteorology, climatology, oceanography, ecosystems and pollution at various scales in selected showcases. This requires co-operation between existing European and international infrastructures (in-situ and remote including space-based).

Working on a set of iAOS showcase applications demonstrated that the volume, variety, and variability of data sources are still a challenge for the scientific community. The WP5-WP6 collaboration on the iAOS showcases helped to pinpoint the high interest of scientific users for improved, platform-based, data science capabilities (descriptive, diagnostic, predictive, and prescriptive). In this context, data preparation for inclusion into the showcase applications is the main effort, in particular to assess the data quality.

There are still only few platforms supporting this effort-consuming step, and WP5 focused on an OPeNDAP-related capacity building effort, along with a consolidation work on the Pangaea-related tools. Data analysis of poly-structured data sources is another time-consuming effort. To perform exploratory analysis and come up with work plan hypotheses, WP5 focused on capacity building for using Geostatistics. The orchestration of efficient data processing campaigns, once the previous steps have been addressed, is also a critical aspect. This requires platform-based tools for running and delivering results in a well controlled fashion. WP5 focused on the use of Jupyter Notebooks with ad-hoc programming environments and execution kernels solving common computing environment issues.

Finally, we have addressed the question of the iAOS cloud infrastructure sustainability, through a discussion of the transverse aspects: cost, legal, trust, privacy, security, usability. WP 5 has delivered results on each of these aspects, and general recommendations for the future have been formulated (Section 3). These transverse aspects are identified as means and conditions, besides the technology and technical resources, to establish long-term community support. They are perceived as strategic differentiators to provide the iAOS cloud platform with a better capacity to evolve and to be maintained over time.

----------- END of DOCUMENT-----------

INTAROS

This report is made under the project
Integrated Arctic Observation System (INTAROS)
funded by the European Commission Horizon 2020 program
Grant Agreement no. 727890.

Project partners: