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 5.3

Data Integrated from Existing Repositories V1

Start date of project:	01 December 2016	Duration:	60 months
Due date of deliverable:	30 November 2018	Actual submission date:	10 December 2018
Lead beneficiary for preparing the deliverable:		AWI	
Person-months used to produce deliverable:		3 pm	
Other contributing partners		ARMINES, DTU, FMI, GEUS, IMR, NERSC, NUIM, TERRADUE	

Authors: Ingo Schewe (AWI), Hervé Caumont (TDUE), Fabien Ors (ARMINES), Pedro Goncalves (TDUE), Torill Hamre (NERSC), Georg Heygster (UB), David Gustavsson (SMHI)
Reviewed by: Hanne Sagen (NERSC)

VERSION	DATE	CHANGE RECORDS	LEAD AUTHOR
0.1	16/03/2017	Template	Pedro Gonçalves
0.2	29/05/2017	Updates on ToC	Ingo Schewe
0.3	29/10/2018	1st Review	Ingo Schewe
0.3	21/11/2018	2nd Review	Ingo Schewe
1.0	26/11/2018	Final Version	Ingo Schewe
1.0	06/12/2018	Revised Version	H. Sagen / I. Schewe
1.1	10/12/2018	Submitted Version	Kjetil Lygre

Approval	Date:	Sign.
X	07 December 2018	Coordinator

USED PERSON-MONTHS FOR THIS DELIVERABLE					
No	***Beneficiary***	***PM***	***No***	***Beneficiary***	***PM***
1	NERSC	***x***	**24**	TDUE	***x***
2	UiB		**25**	GINR	
3	IMR	***x***	**26**	UNEXE	
4	MISU		**27**	NIVA	
5	AWI	**3**	**28**	CNRS	
6	IOPAN		**29**	U Helsinki	
7	DTU	***x***	**30**	GFZ	
8	AU		**31**	ARMINES	***x***
9	GEUS	***x***	**32**	IGPAN	
10	FMI	***x***	**33**	U SLASKI	
11	UNIS		**34**	BSC	
12	NORDECO		**35**	DNV GL	
13	SMHI	***x***	**36**	RIHMI-WDC	
14	USFD		**37**	NIERSC	
15	NUIM	***x***	**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	***x***	**44**	RADI	
22	UHAM		**45**	KOPRI	
23	NORUT		**46**	NIPR	
			47	PRIC	

x = contribution to the deliverable

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

This deliverable document contains a summary about the first 'expansion phase' of the 'Data Integration from Existing Repositories' into the iAOS. Basically the work for this deliverable would not be possible at all without a close entanglement of the duties in task 2.3 and task 5.3. Accordingly, there was close cooperation between WP2 and WP5 in the respective project phase. This manifests itself, for example, in the fact that an important basis for the first work steps in task 5.3 was very much directed by results of the assessment in WP2. The assessments of the technical maturity level were therefore an important driving factor for the selection of first 'show-cases' demonstrating the integration into the iAOS.

Against this background, this document is divided into a description which technical metadata is available after assessment and which general requirements for Data integration from existing repositories into the iAOS are necessary. After a further description which selection criteria were used to qualify for a show-case, an outlook follows which further steps will probably be necessary until the end of the project in order to be able to integrate as many data sources as possible into iAOS.

Basically this document is to be understood as a living document, which should be constantly updated in the course of the further project in order to document the corresponding steps of the integration of data sources into iAOS. This will then lead to the end of the project in version two of this deliverable 'Data Integrated from Existing Repositories V2', D5.9, M54.

Table of Contents

1. Introduction

Existing observing systems, data repositories and infrastructure available from partners and collaborators are the building blocks of the iAOS. The observing systems and data repositories were assessed in WP 2 (Exploitation of existing observing systems). New data and generated products from INTAROS will be stored in the optimal repositories based on the outcome of this assessment. Thus, INTAROS will not build up a new e-infrastructure for data storage and preservation, but instead capitalize on the many existing research data infrastructures in Europe, US, Canada and Asia, that hold environmental data for the Arctic.

ORDP: The European Commission (EC) is running a flexible pilot under Horizon 2020 called the Open Research Data Pilot (ORDP). This pilot is part of the Open Access to Scientific Publications and Research Data Program in H20201. The ORDP aims to improve and maximise access to and re-use of research data generated by Horizon 2020 projects and takes into account the need to balance openness and protection of scientific information, commercialisation and Intellectual Property Rights (IPR), privacy concerns, security as well as data management and preservation questions. In this context, this deliverable is to be understood as a report on the status of the implementation of data integration into iAOS in order to finally reach consistency with the requirements that characterize an ORDP.

1.1. linkage between WP2 'Exploitation of existing observing systems' and WP5 'Data integration and management'

In tasks 2.3 and 5.3, the narrow link between the two work packages manifests itself. In task 2.3, various partners have already described a large number of data sources in the proposal and promised to integrate them into iAOS. Many of these sources also use the repositories whose integration is mentioned in the description of task 5.3. The diagram below (fig. 1.1.1) illustrates how these links individually might look like.
Generally the assessment carried out in WP2 added a large number of potential candidates for integration. While the goal of WP2 was rather to gain an overview of which data sources are to be integrated at all and which technical framework conditions they bring along for integration and where adjustments have to be made. In the task documented here, the actual implementation of the integration into the iAOS takes place.

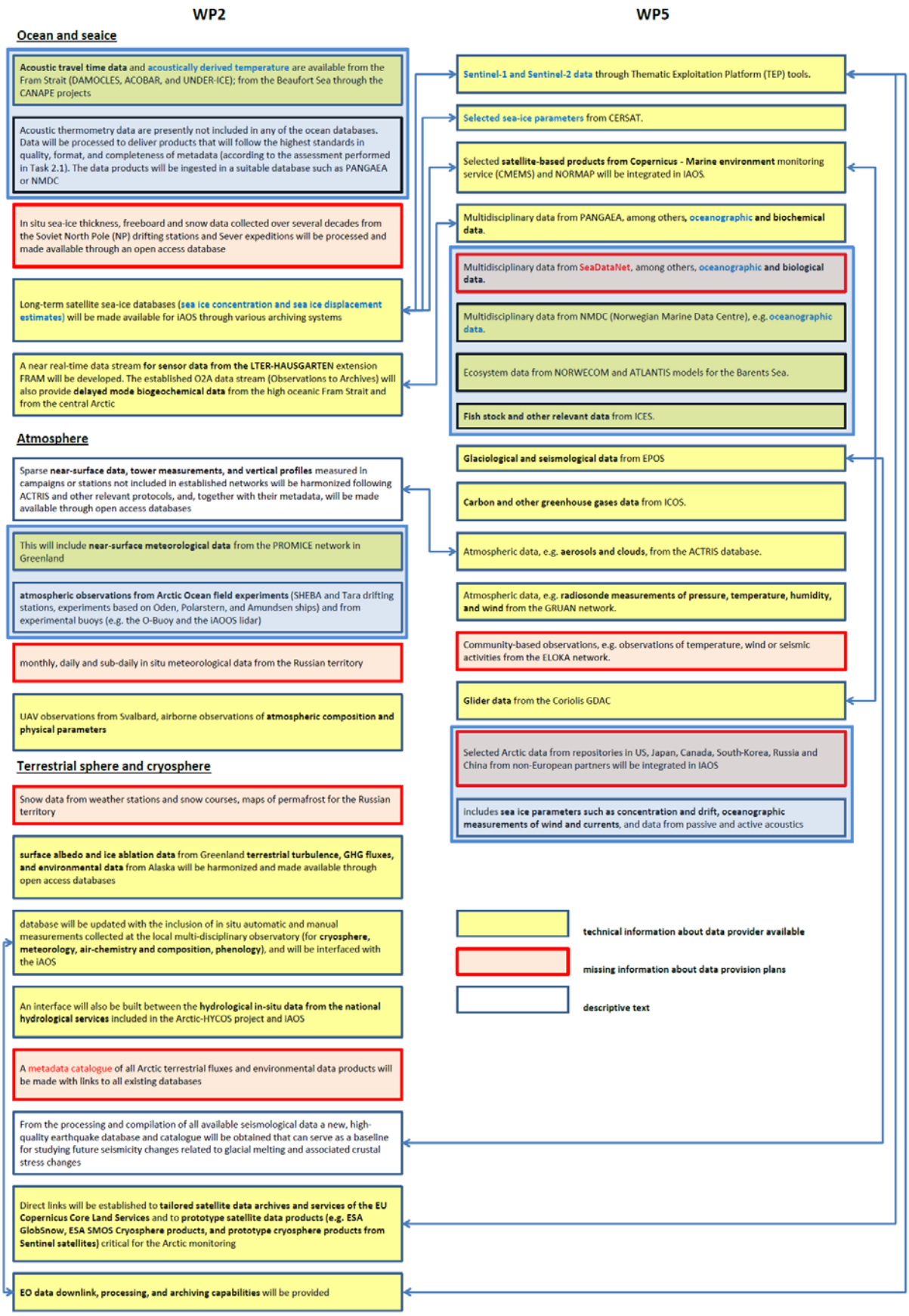

Figure 1.1.1. Overview of the connections between WP2 and WP5 and its data sources originally indexed for integration into the iAOS (link for original document)

2. Assessment of existing Arctic Observing Systems

The assessment was based on a survey, which covered part of the existing Arctic observational data that are anticipated to be most relevant for the INTAROS project (WP2 task 2.1). The survey initially was conducted among the INTAROS partners, in particular those who have promised to provide data to the integrated observing system. The survey addressed Arctic in-situ and satellite based observations of the ocean, atmosphere and terrestrial parameters retrieved through established networks/observing systems as well as individual measurement campaigns and projects.
In addition to the data collection component (infrastructure) itself, a significant focus was on the data management component (e-infrastructure) of the observing systems. The e-infrastructure usually includes hardware and software for data repository, the data processing, data discovery and visualization services. The management can be centralized in a single institution or distributed among several national institutions, which have agreed on common standards for the data and metadata formats, documentation and management.
For each network/observing system we assessed observed parameters in time and space, measurement accuracy and representativeness, data processing maturity, data delivery mode, quality control mechanism and information for current and historical time periods, and long-term sustainability (funding mechanism, technical readiness). These actions finally result in an overall picture which allows conclusions to be drawn regarding the maturity of a system or a collection with regard to its integrability into the iAOS. On this basis, those data collections from task 2.3 were selected whose maturity status was particularly high in order to be integrated already in an early implementation phase of the iAOS. A more detailed description of those 'show cases' is provided in chapter 5.

3. Compilation of data products from distributed databases and observatories

It is important to distinguish between the in task 2.3 generated INTAROS online data-catalogue and the 'working-catalogue' for task 5.3. The online catalogue provides to the end-user in a very condensed way most important information about data collections integrated into the iAOS and is presented by deliverables D2.3, D2.6 and D2.9.

3.1. INTAROS online data-catalogue

The INTAROS WP2 catalogue(s) was originally thought to be a static document, but in the recent months we have started to envision it as a web-based dynamic tool (which can be automatically updated), that will be used also in the INTAROS portal to enable partners (and all data users) to search for and (when possible) access the Arctic data.
The catalogue hosts the basic information needed to enable the users to understand the key data characteristics, and how to get the data. The INTAROS data catalogue will contain descriptions of and links to all datasets collected or generated through exploiting existing datasets and/or estimating new parameters within the project. In a first version of the catalogue, released at end November 2018, partners have registered the datasets that are resulting from their work in WP2 (Exploitation of existing observing systems) during the two first years of the INTAROS project. Each dataset is described by a set of metadata elements that capture key characteristics of the dataset.
The INTAROS data catalogue is online at https://catalog-intaros.nersc.no/.

3.2. 'working catalogues' of products and services based on data from different spheres:

The 'working-catalogue' presented here is a classical working database which summarizes all technical information relevant for integration into iAOS. Those information also originate from the results of all surveys, but are supplemented by information obtained e.g. through personal contact with data providers. It is a pure working catalogue not open to the public and its content is only used to have additional helpful information at hand when integrating data providers into the iAOS.

Table 3.2.1. working catalogue for the Ocean/ Sea-Ice Domain)

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	Wich discovery service is provided for the data product?	Which data access service is provided for the	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Fram Strait Multipurpose Acoustic System	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,	THREDDS Data Server	OpenDAP			acobar.nersc.no, under-ice.nersc.no	N/A
IMR SI_Arctic vessel mounted ADCP system	OCEANIC, PHYSICS: Subsurface currents, OCEANIC, PHYSICS: Surface currents						
IMR fixed hydrographic sections	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,						
IMR Barents Sea Opening mooring array	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,						
NorArgo	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,		FTP				
FRAM (FRontiers in Arctic marine Monitoring))	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,	OAI-PMH	DDI (Direct Download Interface; documentation: https://wiki.pangaea.de/wiki/DDI;			https://www.awi.de/en/science/special-groups/deep-sea-research/observatories/iter-observatory-hausgarten.html;	https://www.pangaea.de/?q=project%3AFRAM+OR+project%3AHAUSGARTEN
IMR-PINRO Ecosystem Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHRIC, SURFACE: Pressure,						
IMR Barents Sea Winter Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, OCEANIC, PHYSICS: Subsurface	None	None				
IMR Fixed hydrographic (near coastal) station network	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,	OAI-PMH	FTP				
Physical Oceanography Data of the AWI Fram Strait Mooring Array	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,	OAI-PMH	Retrieval of data sets is provided through a full text search engine (based on Apache Lucene / panFMP).			https://www.awi.de/en/expedition/observatories/ocean-fram.html	https://www.pangaea.de/
Polarstern ship borne CTD surveys	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,	OAI-PMH	Retrieval of data sets is provided through a full text search engine (based on Apache Lucene / panFMP).			https://www.awi.de/en/expedition/observatories/ocean-fram.html	https://www.pangaea.de/
AWI Polarstern VM ADCP measurements	OCEANIC, PHYSICS: Subsurface currents, OCEANIC, PHYSICS: Surface currents	OAI-PMH	THREDDS Data Server			https://www.awi.de/en/expedition/observatories/ocean-fram.html	https://www.pangaea.de/
Canada Basin Acoustic Propagation Experiment (CANAPE 2016)	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,	None	None			N/A	N/A
SIOS Airborne Infrastructure	OCEANIC, PHYSICS: Surface currents, OCEANIC, PHYSICS, Sea state (Wave height), OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice	Will be part of the SIOS data management system	Will be part of the SIOS data management system				
DAMOCLES-2008-2009-Depth-Range-Averaged-Ocean-Temperature	OCEANIC, PHYSICS: Subsurface temperature			netCDF	CF-compliant	N/A	
DAMOCLES-2008-2009-ambient noise	OCEANIC, PHYSICS: ambient noise			netCDF	CF-compliant	N/A	

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	Wich discovery service is provided for the data product?	Which data access service is provided for the	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
UDASH - Unified Database for Arctic and Subarctic Hydrography	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,	OAI-PMH	PANGAEA data download service (filter by geocodes and parameters; http://ws.pangaea.de/dds-	ascii txt	ISO 19115	https://doi.org/10.1594/PANGAEA.872931	www.Pangaea.de
UNDER-ICE-2014-2016-ambient noise	OCEANIC, PHYSICS: ambient noise			Not any standard	CF-compliant	N/A	
ACOBAR-2010-2012-ambient-noise	OCEANIC, PHYSICS: ambient noise			netCDF	CF-compliant	N/A	
IMR Fixed hydrographic (near coastal) station network	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,						
IMR SI_Arctic vessel mounted ADCP system	OCEANIC, PHYSICS: Subsurface currents, OCEANIC, PHYSICS: Surface currents						
IMR fixed hydrographic sections	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,			SEED			
IMR Barents Sea Opening mooring array	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,						
NorArgo	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,						
IMR Barents Sea Winter Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, OCEANIC, PHYSICS: Subsurface			Not any standard			
IMR-PINRO Ecosystem Survey Hydrography	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,						
IMR-PINRO Ecosystem Survey Nutrients	OCEANIC, BIOGEOCHEMISTRY: Nutrients (Interior ocean concentrations of silicate, phosphate, nitrate)						
IMR-PINRO Ecosystem Survey Fish	OCEANIC, BIOLOGY/ECOSYSTEMS: Fish abundance and distribution, OCEANIC, BIOLOGY/ECOSYSTEMS: Marine biodiversity			Not any standard			
Biogenic particle flux at the FRAM observatory from mooring sediment traps	OCEANIC, BIOGEOCHEMISTRY: Suspended particulates			CSV	ISO 19115	https://www.pangaea.de/?q=parameter%3Aflux&f.device%5B%5D=Mooring+(long+time)&f.location%5B%5D=Arct	
Inorganic nutrients measured on Fram-Strait water samples since 1997	OCEANIC, BIOGEOCHEMISTRY: Nutrients (Interior ocean concentrations of silicate, phosphate, nitrate)			CSV	ISO 19115	https://www.pangaea.de/?q=parameter%3Aname%3A%22depth%2C+water%22+parameter%3Aname%3Aphospha	
High resolution sea-bed photographs and footage from repeated long term surveys for fauna investigations	OCEANIC, BIOLOGY/ECOSYSTEMS: Marine biodiversity, OCEANIC, BIOLOGY/ECOSYSTEMS: Epibenthic megafauna abundance and distribution			jpg, mov	ISO 19115	https://www.pangaea.de/?q=project%3Aname%3Ahausgarten+project%3Aname%3Afram&f.device%5B%5D=Ocea	
Ship borne CTD surveys of temperature and salinity	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Sea surface temperature,			Tab delimited ASCII text files	ISO 19115	https://www.pangaea.de/?q=project%3Aawi_phyoce+AND+device%3Actd+AND+parameter%3Aname%3Atemperatu	

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	Wich discovery service is provided for the data product?	Which data access service is provided for the	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Ship borne CTD surveys of oxygen and chlorophyll	OCEANIC, BIOGEOCHEMISTRY: Oxygen (interior ocean oxygen concentration), OCEANIC, BIOGEOCHEMISTRY: Chlorophyll			Tab delimited ASCII text files	ISO 19115	https://www.pangaea.de/?q=project%3Aawi_phyoce+AND+device%3Actd+AND+method%3A%22fluorometer%22+O	
Biogeochemical parameters from deep-sea sediments taken at the long-term observatory AWI-HAUSGARTEN	OCEANIC, BIOGEOCHEMISTRY: benthic ecology			CSV	ISO 19115	https://www.pangaea.de/?q=project%3Ahausgarten+biochemical	
Physical Oceanography Mooring-Data of the AWI Fram Strait Mooring Array	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,			CSV	ISO 19115	https://www.pangaea.de/?q=device%3Amoory+project%3Aawi_phyoce&maxlat=82&minlon=-	
Benthic oxygen fluxes in the Arctic Fram Strait	OCEANIC, BIOGEOCHEMISTRY: Oxygen (interior ocean oxygen concentration)			CSV	ISO 19115	https://www.pangaea.de/?q=project%3Ahausgarten+parameter%3Aname%3A%22depth%2C+sediment*%22+para	
AWI Polarstern VM ADCP measurements	OCEANIC, PHYSICS: Subsurface currents, OCEANIC, PHYSICS: Surface currents			binary data	ISO 19115	https://www.pangaea.de/?q=method%3Atrdi&f.basis%5B%5D=Polarstern&minlat=75	
WIFAR/UNDER-ICE acoustic recording in the Marginal Ice Zone-2012	OCEANIC, PHYSICS: ambient noise		Download link is sent via e-mail	netCDF	CF-compliant	https://archive.norstore.no/pages/public/datasetDetail.jsf?id=10.11582/2017.00012	https://archive.norstore.no
Digital terrain model (DTM) of the central Fram Strait	OCEANIC, BATHYMETRY	OAI-PMH	DDI (Direct Download Interface; documentation: https://wiki.pangaea.de/wiki/DDI;	netCDF	ISO 19115	https://doi.pangaea.de/10.1594/PANGAEA.526589	www.pangaea.de
EGO gliders (European Gliding obervatories)	OCEANIC, PHYSICS: Subsurface temperature, OCEANIC, PHYSICS: Subsurface salinity, OCEANIC, PHYSICS: Subsurface currents,	?	FTP	netCDF	CF-compliant	http://www.coriolis.eu.org/Data-Products/Catalogue#/metadata/589bfa51-2219-4cc8-a19e-	http://www.coriolis.eu.org/
Arctic high resolution ice edge	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	OAI-PMH	THREDDS Data Server	netCDF			
ASI Sea ice concentration	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	interactive web site for browsing data; ftp	interactive web site for browsing data + ftp	hdf, GeoTIFF, png		https://www.seaice.uni-bremen.de	https://www.seaice.uni-bremen.de/
Multiyear sea ice concentration	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	interactive browsing and ftp	interactive web site for browsing data + ftp	netcdf, png		https://www.seaice.uni-bremen.de	https://www.seaice.uni-bremen.de/
Thickness of thin sea ice	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	interactive web site for browsing data; ftp	interactive web site for browsing data + ftp	hdf, geoTiff, png		https://www.seaice.uni-bremen.de	https://www.seaice.uni-bremen.de/
Ifremer/CERSAT Sea ice concentration	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	FTP	FTP	netCDF		ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-concentration/data/	http://cersat.ifremer.fr
Ifremer/CERSAT Arctic sea ice drift at large scale	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	FTP	FTP	netCDF	CF	ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/	ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/
Ifremer/CERSAT Arctic sea ice drift at medium resolution scale	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	FTP	FTP	netCDF	CF	ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/arctic/amsr2-merged/	ftp://ftp.ifremer.fr/ifremer/cersat/products/gridded/psi-drift/data/arctic/amsr2-merged/
Arctic Ocean - Sea Ice Concentration Charts - Svalbard	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	DIRECT GET FILE, FTP	DIRECT GET FILE, WMS, FTP	netCDF	Unidata Dataset Discovery v1.0	http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SE	http://marine.copernicus.eu/
Global Ocean Sea Ice Concentration Time Series REPROCESSED	OCEANIC, PHYSICS: Sea-ice (sea ice concentration, sea-ice extent/edge, sea ice thickness, sea-ice drift, snow thickness, albedo)	DIRECT GET FILE, FTP	DIRECT GET FILE, WMS, FTP	netCDF	Unidata Dataset Discovery v1.0	http://marine.copernicus.eu/services-portfolio/access-to-products/?option=com_csw&view=details&product_id=SE	ftp://osisaf.met.no/reprocessed/ice/conc-cont-reproc/v1p2

Table 3.2.2. working catalogue for the Atmosphere Domain)

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	discovery service	data access service	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Global-GAW	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water	More than one service, depends on data centre	More than one service, depends on data centre			https://public.wmo.int/en/programmes/global-atmosphere-watch-programme	https://www.wmo.int/pages/prog/arep/gaw/world_data_ctres.html
ICOS	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water					https://www.icos-ri.eu/	https://www.icos-cp.eu/
PROMICE automatic weather station network	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water	Data exploration possibilities provided through data portal				promice.org	promice.org
Airborne observations of surface-atmosphere fluxes	ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water vapor, ATMOSPHERIC, SURFACE: Surface	None	upon request				
GRUAN (GCOS Reference Upper Air Network)	ATMOSPHERIC, UPPER-AIR: Temperature (tropospheric temperature profile, stratospheric temperature profile, temperature of deep atmospheric layers),	FTP	FTP			www.gruan.org ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/RS92-GDP/version-002	www.gruan.org (provides a link to ftp area hosted by NOAA NCEI, soon to also be served via C3S CDS under
Arctic Summer Cloud Ocean Study (ASCOS); Arctic Clouds during Summer Experiment (ACSE)	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water	None					www.ascos.se ; http://bolin.su.se/data/
Polarstern Arctic field campaigns	ATMOSPHERIC, UPPER-AIR: Temperature (tropospheric temperature profile, stratospheric temperature profile, temperature of deep atmospheric layers),	IODP				https://www.pangaea.de	https://www.pangaea.de
Norwegian Young Sea Ice Cruise (N-ICE2015); Sea State 2015	ATMOSPHERIC, UPPER-AIR: Temperature (tropospheric temperature profile, stratospheric temperature profile, temperature of deep atmospheric layers),					N-ICE2015: https://data.npolar.no/dataset/216df9b3-e2bd-5111-9c02-fea848d76670	https://data.npolar.no/home/
IMR-PINRO Ecosystem Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHRIC, SURFACE: Pressure,						
IMR Barents Sea Winter Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, OCEANIC, PHYSICS: Subsurface	None	None				
Aerosol, Clouds, and Trace gases Research Infrastructure (ACTRIS))	ATMOSPHERIC COMPOSITION: Ozone and aerosol, supported by their precursors (Total column ozone, tropospheric ozone, ozone profile in					https://www.actris.eu	actris.nilu.no
AIRMETH_turbulent_fluxes_Polar5	ATMOSPHERIC, SURFACE: Latent and sensible heat fluxes, TERRESTRIAL, ECOSYSTEMS: greenhouse gas fluxes (CO2, CH4)			CSV			
AIRMETH_vertical_profiles_Polar 5	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Water vapor, ATMOSPHRIC, SURFACE: Pressure, Concentration of			CSV			
AIRMETH_vertical_profiles_Helipod	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Water vapor, ATMOSPHRIC, SURFACE: Pressure, Concentration of			CSV			
GRUAN	ATMOSPHERIC, UPPER-AIR: Temperature (tropospheric temperature profile, stratospheric temperature profile, temperature of deep atmospheric layers),	FTP	FTP ftp://ftp.ncdc.noaa.gov/pub/data/gruan/processing/level2/RS92-GDP/version-002	netCDF	Collection of rich metadata via RSlaunchclient. Extends beyond those stated here	www.gruan.org	
Long term monitoring at Polish Polar Station Hornsund	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHERIC, SURFACE: Water			CSV	WIGOS		

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	discovery service	data access service	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Surface net longwave radiation	ATMOSPHERIC, SURFACE: Surface radiation budget (surface longwave radiation budget, surface shortwave radiation budget)			Not any standard			
Surface net shortwave radiation	ATMOSPHERIC, SURFACE: Surface radiation budget (surface longwave radiation budget, surface shortwave radiation budget)			Not any standard			
Surface downwelling longwave radiation	ATMOSPHERIC, SURFACE: Surface radiation budget (surface longwave radiation budget, surface shortwave radiation budget)			Not any standard			
Surface downwelling shortwave radiation	ATMOSPHERIC, SURFACE: Surface radiation budget (surface longwave radiation budget, surface shortwave radiation budget)			Not any standard			
Cloud mask	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and			Not any standard			
Cloud liquid water path	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and			Not any standard	CF-compliant		
Cloud ice water content profiles	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and			Not any standard			
IMR Barents Sea Winter Survey	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, OCEANIC, PHYSICS: Subsurface			Not any standard			
GAW Aerosols	ATMOSPHERIC COMPOSITION: Ozone and aerosol, supported by their precursors (Total column ozone, tropospheric ozone, ozone profile in			Nasa-Ames	NASA-Ames	http://ebas.nilu.no	
GAW Aerosols	ATMOSPHERIC COMPOSITION: Ozone and aerosol, supported by their precursors (Total column ozone, tropospheric ozone, ozone profile in			NASA-Ames	NASA-Ames		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, SURFACE: Precipitation (amount of liquid precipitation, amount of solid precipitation)			CSV	WIGOS		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, SURFACE: Water vapor			CSV	WIGOS		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, SURFACE: Wind speed and direction			CSV	WIGOS		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHRIC, SURFACE: Pressure			CSV	WIGOS		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, SURFACE: Surface radiation budget (surface longwave radiation budget, surface shortwave radiation budget)			CSV	WIGOS		
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and			CSV	WIGOS		

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	discovery service	data access service	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Long term monitoring at Polish Polar Station Hornsund on Spitsbergen	ATMOSPHERIC, SURFACE: Air temperature			CSV	WIGOS		
GAW-Ozone	ATMOSPHERIC COMPOSITION: Ozone and aerosol, supported by their precursors (Total column ozone, tropospheric ozone, ozone profile in			CSV	extCSV	http://www.woudc.org/data/	
CLARA-A2	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and		FTP	netCDF		https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002	https://doi.org/10.5676/EUM_SAF_CM/CLARA_AVHRR/V002
ESA Cloud_cci property datasets from passive satellite sensors: AVHRR-AM L3C/L3U cloud products	ATMOSPHERIC, UPPER-AIR: Cloud properties (cloud amount, cloud-top pressure, cloud-top temperature, cloud optical depth, cloud water path (liquid and		FTP	netCDF		http://www.esa-cloud-cci.org/	http://www.esa-cloud-cci.org/?q=data_download

Table 3.2.3. working catalogue for the Terrestial Domain)

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	Wich discovery service is provided for the data product?	Which data access service is provided for the	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
GLISN network Greenland	TERRESTRIAL, CRUST: Earthquakes	breq_fast	breq_fast			glisn.info	http://ds.iris.edu/SeismiQuery/virtual_net.htm
FMI Sodankylä	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, ATMOSPHRIC, SURFACE: Pressure,	OGC CSW	OGC WFS			http://fmiarc.fmi.fi/	https://en.ilmatieteenlaitos.fi/open-data, http://litdb.fmi.fi/
Fluxnet	TERRESTRIAL, HYDROLOGICAL: Soil moisture (Surface soil moisture content, Freeze/thaw status, surface inundation, vegetation optical depth, root-zone soil					http://ameriflux.lbl.gov/ https://fluxnet.ornl.gov	http://ameriflux.lbl.gov/ https://fluxnet.ornl.gov/
Norwegian National Seismic Network (NNSN)	TERRESTRIAL, CRUST: Earthquakes	web search	web portal			skjelv.no	ftp://ftp.geo.uib.no/pub/seismo/DATA/
Automated Weather and Snow Measuring System	ATMOSPHERIC, SURFACE: Air temperature, ATMOSPHERIC, SURFACE: Wind speed and direction, TERRESTRIAL, CRYOSPHERE: Snow					http://158.39.149.181/SASM/index.html	http://158.39.149.181/SASM/index.html
GLISN	TERRESTRIAL, CRUST: Earthquakes				SEED	http://ds.iris.edu/SeismiQuery/virtual_net.htm	
eddy flux data (CO2 and CH4)	TERRESTRIAL, ECOSYSTEMS: greenhouse gas fluxes (CO2, CH4)			CSV		http://ameriflux.lbl.gov/ http://fluxnet.fluxdata.org/	
Snow cover - Hornsund glaciers	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	None	None	Not any standard			
Front velocity of tidewater glaciers	TERRESTRIAL, CRYOSPHERE: Glaciers (Area, elevation change, glacier mass change, glacier topography)	None	None	Not any standard			
Soil frost/snow stations	TERRESTRIAL, HYDROLOGICAL: Soil moisture (Surface soil moisture content, Freeze/thaw status, surface inundation, vegetation optical depth, root-zone soil			CSV		http://litdb.fmi.fi/distributed_stations.php	
Snow depth stations	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow			CSV		http://litdb.fmi.fi/suo0003_data.php, http://litdb.fmi.fi/ioa0003_data.php	
Snow scale SWE	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow			CSV		http://litdb.fmi.fi/ioa0011_data.php	
Manual SYNOP observations	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow			CSV		http://litdb.fmi.fi/luo0016_data.php	
ERA-CLIM2 in situ SWE	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	None	None	Matlab, txt		http://litdb.fmi.fi/eraclim2.php	http://litdb.fmi.fi/eraclim2.php
ArcticHYCOS_GRDC_archive	TERRESTRIAL, HYDROLOGICAL: River discharge (River discharge, water level, flow velocity, cross-section)			GRDC CSV format	WIGOS	http://www.bafg.de/GRDC/EN/04_spcldtbss/41_ARDB/arcticHycos.html	
Mass balance of Hans Glacier at Svalbard	TERRESTRIAL, CRYOSPHERE: Glaciers (Area, elevation change, glacier mass change, glacier topography)			CSV		http://wgms.ch/products_gmbb/	

Name of the observatory/ data collection/ sattelite product	Relevant variables observed by the system	Wich discovery service is provided for the data product?	Which data access service is provided for the	International/ community standard format of the data	metadata format	URL othe observatory/ data collection/ sattelite product	URL of the data repository(s) (if it exists)
Norwegian National Seismic Network	TERRESTRIAL, CRUST: Earthquakes			miniSEED	DCAT/CERIF/QuakeML/SEED	ftp://ftp.geo.uib.no/pub/seismo/DATA/	
NNSN catalog	TERRESTRIAL, CRUST: Earthquakes			Nordic format, QuakeML	DCAT/CERIF/QuakeML/SEED	ftp://ftp.geo.uib.no/pub/seismo/DATA/	
AWS	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow			CSV		http://litdb.fmi.fi/luo0015_data.php	
Front positions of tidewater glaciers in Hornsund (S Svalbard)	TERRESTRIAL, CRYOSPHERE: Glaciers (Area, elevation change, glacier mass change, glacier topography)	None	None	Not any standard			
Soil frost/snow stations	TERRESTRIAL, HYDROLOGICAL: Soil moisture (Surface soil moisture content, Freeze/thaw status, surface inundation, vegetation optical depth, root-zone soil			CSV		http://litdb.fmi.fi/distributed_stations.php	
Soil frost/snow stations	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow			CSV		http://litdb.fmi.fi/distributed_stations.php	
ESA DUE GlobSnow v2.0 SWE	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	FTP	FTP	netCDF		www.globsnow.info	http://www.globsnow.info/se/
SMOS soil frost	TERRESTRIAL, HYDROLOGICAL: Soil moisture (Surface soil moisture content, Freeze/thaw status, surface inundation, vegetation optical depth, root-zone soil	None	None	Not any standard			
GlobSnow snow extent	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	None	http	netCDF		http://www.globsnow.info/index.php?page=Snow_Extent	http://www.globsnow.info/index.php?page=SE
ERA-CLIM2 Northern Hemisphere snow water equivalent	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	None	None	HDF		http://litdb.fmi.fi/eraclim2.php	http://www.globsnow.info/swe/archive_v2.1_Eraclim/
IMS Daily Northern Hemisphere Snow and Ice Analysis at 1 km, 4 km, and 24 km Resolutions, Version 1	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	FTP	FTP	GeoTIFF	OAI 2.0	http://nsidc.org/data/G02156	
MODIS/Aqua Snow Cover Daily L3 Global 500m Grid, Version 6	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	THREDDS Data Server	FTP	HDF	OAI 2.0	http://nsidc.org/data/MYD10A1	http://nsidc.org/data/MYD10A1
AMSR-E/Aqua Daily L3 Global Snow Water Equivalent EASE-Grids, Version 2	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	https	https	HDF	OAI 2.0	http://nsidc.org/data/AE_DySno	
SMAP L3 Radiometer Northern Hemisphere Daily EASE-Grid Freeze/Thaw State, Version 1	TERRESTRIAL, HYDROLOGICAL: Soil moisture (Surface soil moisture content, Freeze/thaw status, surface inundation, vegetation optical depth, root-zone soil	FTP	FTP	HDF	OAI 2.0	http://nsidc.org/data/SPL3FTP http://nsidc.org/data/SPL3FTP_E	
JASMES snow depth	TERRESTRIAL, CRYOSPHERE: Snow (Spatial extent of snow, fractional snow cover (viewable and canopy adjusted), snow depth, snow water equivalent, snow	None	None	HDF		http://kuroshio.eorc.jaxa.jp/JASMES/WC.html	

4. General requirements for Data integration from existing repositories into the iAOS

Federated datastores allow linking of distributed data collections. The iAOS allows for federation of instances by linking them through APIs, controlled by the iAOS admins. The end result is that users can access data from multiple datastores from a single portal or API while the data remains within the control of the experts (data centres). A federated system should always serve the latest version of the data, thus solving the 'multiple copies' issues found in a traditional distributed system. While federation between the same software (e.g. OpenDAP to OpenDAP) isn't complex, federating between different systems/software is more complex and relies on good coordination between the repository and the iAOS.

Further to the use and importance of standards, the standardisation of metadata and data are crucial if data are to be readily usable by a machine or the dataset aggregated. Former activities like Ocean Data View and SeaDataNet have introduced a standard ASCII representation of data. For multidimensional and larger datasets where binary formats are used key advances have included the introduction of the CF NetCDF standards and the Attribute Convention for Dataset Discovery (ACDD). Elements of cf-netCDF and ACDD have been used in the NetCDF formats produced by the community observing programmes (like Argo, OceanSITES, Ocean Glider network etc). In a separate development stream, the Open Geospatial Consortium (OGC) developed standards including SensorML for sensor metadata and Observations and Measurements (O&M) for sensor data. These are XML based representations but are readily converted to other formats such as JSON.

Technologies like OpenSearch can be used to complementary serve as the query aspect of individual data access protocols built on core protocols like HTTP and commonly accepted methodologies like REST that can be handled by a large set of client tools as simple as common web browsers, download-managers or computer programs. Those provide a way to access identified or located results and download them and allow publishing of search results in a standard and accessible format. OpenSearch itself is RESTful technology and can be used to complementary serve as a query aspect, which provides a way to access identified or located results and download them.

All these described standards and services were used by us to assess the 'cloud readiness' of an individual. Which simply means the technical capability to perform machine to machine queries and retrieve references to dataset files that can be fetched by a data processing program.

In addition to the purely technical requirements of a system and its maturity level, an important requirement is also the relevance of the data collections hosted by such a system. Here the importance for the partners working within INTAROS, especially those from WP6, is very important. In order to find out which data collections are of particular importance for the partners from WP6, we invited the interested persons to a workshop at the general assembly 2018 in Helsinki. The results of the discussions at this workshop provided us with important information for selecting suitable 'show cases' for early integration into the iAOS.

5. Show cases: first integration of most mature data collections into the iAOS

Since the development of the iAOS still is in an early and preliminary phase WP5 decided to initially focus on the integration of 'show case' data-products and data systems in order to demonstrate pathways for data integration, but also to demonstrate the capabilities and potentials of fully integrated data collections. The aim is to show the full iAOS data streams from the data 'ingestion', combined with search and retrieval functionalities of the task 2.5 'Data Catalogue', towards newly integrated data processing services presented within deliverable 5.7.

The main selection criteria for becoming a potential show-case were that the data-product/ data-provider:

1. have a general impact on tasks/ research-foci in INTAROS
2. are useful to demonstrate further functionalities of the iAOS and its tools
3. have a good level of maturity with regard of integration into the iAOS (proper APIs, search and retrieval functionality, machine to machine communication)

The following chapters describe those data providers that have been selected as 'show cases' for the first implementation phase.

The description essentially contains a documentation of technical features for the data integration of the respective systems.

5.1 Multidisciplinary data from NMDC (Norwegian Marine Data Centre)

The Norwegian Marine Data Centre provides a huge amount of oceanographic data collected during Norwegian expeditions. The OPeNDAP server at IMR is now operational. Through a 'Terradue Ellip' solution, ARMINES can browse the server and download files but ARMINES cannot yet send queries to the server in order to filter the data at the server side. Here still some work is in progress.

Once the case study script will be enhanced by using OPeNDAP access, ARMINES will deploy it as a service through Ellip and iAOS will be able to call it.

ARMINES already is able download data files (CTD by year and by vessel), filter the data and work on statistics and interpolation locally.

To demonstrate the dataflow and general capabilities of 'Geostats' solutions demonstration study has been implemented which has been documented on a 'Jupyter Notebook' (http://rgeostats.free.fr/doc/Files/intaros7.html). The study is meant to demonstrate how to use a simple Kriging interpolation from the RGeostats package applied to the Annual CTD datasets from R/V Håkon Mosby (Norwegian research vessel).

5.2 SeaIce.Uni-Bremen.de

The Sea-Ice Portal of partners from University Bremen (UB) provides high-quality data of seaice-thickness and ice-concentration, which has high relevance of many research tasks within INTAROS. All data can be found in the data archive, which is accessible via FTP and HTTP:

- For FTP access, connect to "seaice.uni-bremen.de" (use a ftp client, browser will not work). The products are available in the given directories, e.g., "amsr2" for the AMSR2 data. Please set your ftp client to "active mode". Otherwise no connection is possible. The user name is "anonymous" and no password is needed.
- For HTTP access, visit https://seaice.uni-bremen.de/data, the structure is the same as for FTP.
- To quickly browse the data, please have a look at the Databrowser.

Because neither FTP nor HTTP access-options provide sufficient APIs for search and retrieval functionalities UB partners were looking for an adequate solution to implement an openDAP service.

Due to the fact, that Uni-Bremen would only have the funds for the purchase of an openDAP server available at the end of the year it was discussed if Terradue could provide a 'remote openDAP' solution in order to in order to gain first experiences with the implementation. After ensuring that the UB data center had no security concerns about such a solution if this solution affects certain security issues Terradue has set up an OpenDAP service on a remote server, to be hosted on an

external cloud provider (Hetzner). This action has been advertised as a contributed resource to the iAOS (integrated Arctic Observation System), using INTAROS WP5 budget to track this spending.
Apart from seaice-thickness, ice-concentration data-sets seem to be most attractive to share with a greater community and thus are selected for first tests. First test for data integration from FTP servers will be undertaken by TERRADUE with GEO-tif files while metadata seem to be hosted within the GEO-tif files.

5.3 PANGAEA/ FRAM data-products

PANGAEA provides a service which queries the PANGAEA data warehouse and returns 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.
The returned text file contains following columns: Date/Time, Latitude (north), Longitude (east), Depth water [m], Parameter 1, Parameter 2, ..., Data source (DOI)
Base URL for all queries is: http://ws.pangaea.de/dds-fgp/
For FRAM data-products within the iAOS individual direct download queries were prepared and provided for each data collection (table 5.3.1). The generated tab-delimited text-files of a single query contain:

- Date/Time, Latitude, Longitude, Waterdepth, Parameter, Origin of Values (DOI which leads to the original full PANGAEA dataset with full metadata information)
- Queries just provide data to the iAOS within a given geographical bounding box. Queries for smaller geographical areas within that area are of course allowed.
- The same for time and depth intervals. Due to dataset sizes for some parameters single queries are only permitted for yearly queries (otherwise system overflow).
- With each dds-query also a 'classical' PANGAEA-query is provided, which leads to a list of relevant data-sets with full meta-information.

Table 5.3.1. provided PANGAEA parameters with DDS-queries and classical queries.
(Link to usable table: https://goo.gl/2kf3zt)

Parameter	Unit	Min Year	Max Year	comment	dds query (link)	classical PANGAEA query (including meta information)
Water-Temperature	°C	1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Salinity		1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
dissolved Oxgen in the water	µmol/l	1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Chlorophyll (Fluorometer) in the water	arbitrary units	1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
mooring measurements for seawater current velocity	cm/s	1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
mooring measurements for seawater current direction	cm/s	1997	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
AWI Polarstern VM ADCP measurements	binary	1997	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/?...
CTD Nutrients - Phosphate	µmol/l	1997	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
CTD Nutrients - Silicate	µmol/l	1997	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
CTD Nutrients - Nitrate	µmol/l	1997	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
CTD Nutrients - Nitrite	µmol/l	1997	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Flux of: Seston	mg/m2/d	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Flux of: CaCO3	mg/m2/d	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Flux of: POC	mg/m2/d	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Flux of: PON	mg/m2/d	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
OFOS Sea-Floor Images	tif, jpg	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: Sediment porosity	% vol	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: sediment bound Chlorophyll a	µg/cm3	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: sediment bound Phaeopigment	µg/cm3	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: Esterase activity per sediment volume	nmol/ml/h	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: Phospholipid content	nmol/ml	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
HAUSGARTEN biogeochemistry: sediment bound Protein	mg/cm3	1999	2019		http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Benthic oxygen fluxes in the Arctic Fram Strait	µmol/l	2004	2019	single query period limited to one year max!	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...
Digital terrain model (DTM) of the central Fram Strait	binary			Attention!- Diverging dds-query scheme	http://ws.pangaea.de/dds-fdp/rest/panquery...	https://www.pangaea.de/...

5.4 Copernicus Marine environment monitoring service (CEMS)

Many partner services are already adapted to the Copernicus data-model. Therefore it is obvious to integrate this into iAOS at an early stage. CEMS provides a **SUBSETTER download mechanism**. User can retrieve a subset of gridded datasets through http and https protocol. The main idea is: download exactly what user needs. User can retrieve the needed variable, the geospatial coverage, the temporal coverage. There are two ways to access this service: via GUI (through the web portal), or **via a machine to machine interface (script)**. The advantages of this service are:

- User minimizes the volume of data transiting on the network, and the volume of data to be stored on his computer.
- User selects the different parameters of his request in a way “easy to understand”: variable can be selected via their names or standard names (from CF convention standard name table), date format is harmonized at CMEMS level (ex: 2013-02-22).
- Information on dataset (notably the updated temporal coverage) and on the request (volume of the request), can be provided to user before running the request (via GUI or via script).
- Security of the authentication (https, CAS authentication with a ticket).

How to write and run the script to download CMEMS products through Subset or Direct download mechanisms?

The CMEMS web portal allows you generating a template command line using the python script for downloading product you are interested in. This command line integrates the extraction parameters. In order to first create your command line, you have to navigate as if you wanted to download the data through the web portal.

=> Technical Info (http://marine.copernicus.eu/faq/how-to-write-and-run-the-script-to-download-cmems-products-through-subset-or-direct-download-mechanisms/?idpage=169)

Examples of command lines using the motu-client to download CMEMS products

If you choose the SUBSETTING or DIRECT GET FILE download mechanism you can build command lines using the motu-client to interact with **MOTU REST API** .

The client is a python script used to connect to **Motu HTTP server** in order to:

- extract the data of a dataset, with geospatial, temporal and variable criterias (default option)
- get the size of an extraction with geospatial, temporal and variable criterias
- get information about a dataset

and it can be launched under different environments in order to be integrated into a processing chain (usual aim -> automate the downloading of products):

- Windows
- Linux
- MacOS.

=> Technical Info: http://marine.copernicus.eu/faq/can-you-give-a-few-examples-of-command-lines-to-download/?idpage=169

Table 5.4.1. Overview of available Arctic Ocean Variables within CEMS

	Generic variable name	acronyms	Specific variable name
PHY	Temperature	T	Temperature
		SST	Sea surface temperature
		bottomT	Bottom temperature
		SSD	Sea surface density
	Salinity	S	Salinity
		SSD	Sea surface density
	Sea surface height	SSH	Sea surface height
	Current velocity	UV	Current velocity
			Geostrophic current velocity
			Barostropic current velocity
	Mixed layer thickness	MLD	Mixed layer thickness
	Sea Ice	SIC	Sea ice concentration
		SIE	Sea ice edge
		SIT	Sea ice thickness
		SIUV	Sea ice velocity
			Sea ice drift
		SNOW	Snow
		ICBG	Iceberg
		SIAGE	Sea ice age
		SIALB	Sea ice albedo
		WIND	Wind
	Wind	WIND	Stress
		SWH	wave significant height
WAV	Wave	MWP	wave mean period
		VMDR	wave mean direction
		VSDXY	Stokes drift
		WW	wind wave (period, height, direction)
		SW1	primary swell wave (period, height, direction)
		SW2	secondary swell wave (period, height, direction)

	Generic variable name	acronyms	Specific variable name
BIO	Plankton	CHL	Chlorophyll-a
		PHYC	Phytoplankton
		ZOOC	Zooplankton
	Oxygen	O2	Dissolved oxygen
	Nutrients	NO3	Nitrate
		PO4	Phosphate
		SI	Silicate
		FE	Iron
		NH4	Ammonium
	Primary production	PP	Primary production
	Reflectance	RRS	Reflectance
	Transparency	CDM	Absorption coefficient
		ATOT	
		APHY	
		BBP	Back scattering coefficient
		KD	Light attenuation
		ZSD	Secchi depth
		ZEU	Euphotic zone depth
	Turbidity	SPM	Suspended matter
		CDM	Absorption coefficient
		ATOT	
		APHY	
		BBP	Back scattering coefficient

5.5 Descriptions for Arctic-HYCOS river discharge data integration into the iAOS

Background information

The Arctic-HYCOS observing system provides daily and monthly gauged river discharge data from a selection of stations operated by the national hydrological services (NHS) in the Arctic Council member states (Canada, Denmark, Finland, Iceland, Norway, Russian Federation, Sweden and United States of America). The observation system is established by the Arctic-HYCOS project (https://hydrohub.wmo.int/en/projects/Arctic-HYCOS). The set of stations have been selected to provide a basis for monitoring fresh water flow to the Arctic Ocean and for monitoring changes in the hydrological regime. The current list includes 423 stations of which 72 are listed as flow-to-ocean stations, representing the most reliable downstream station in the river basins.

The Global Runoff Data Centre (GRDC, http://www.bafg.de/GRDC) serves as a focal point redistributing historical data and station metadata, whereas provisional and (when available) near-real-time data should be provided directly by the NHS.

This version of the description (Version 1, 2018-11-03) provides information for integrating the station metadata and the historical discharge data provided through GRDC. Description for integration of provisional data directly from the NHS will be provided in the next version of the document.

Integration of the Arctic-HYCOS station metadata and archive data-collections from the Global Runoff Data Centre (GRDC) repository

GRDC provides metadata as well as historical data in the form of compressed (zip) archives through an open ftp access (see URL further below). This means that the compressed data need to be downloaded and further processed in order to retrieve/query for station metadata and river discharge data.

To simplify the integration of the Arctic-HYCOS data in the iAOS, SMHI has developed a set of scripted functions using the R programming language (https://www.r-project.org/) to download, extract, search, reformat, and export the data provided by GRDC. Depending on the implementation of the iAOS, these functions could be provided as a stand-alone or integrated application in the iAOS as an interface to the Arctic-HYCOS data. The R script functions and example uses are given in

Appendix to this document.

The R-script functions can be used to export the Arctic-HYCOS station metadata to the following formats:

- Tab-separated ascii text files
- ESRI shapefile (point layer)
- KML file (point layer)

The river discharge data can be exported into the following formats:

- The original text-format provided by GRDC
- A comma-separated format (csv) developed in the Hydrology-TEP project (by Terradue, SMHI, and others), including for each observation record: timestamp, latitude, longitude, unit of measure, value, and additional metadata fields as defined by the file header.
- A text-format used by the hydrological model HYPE developed by SMHI.

Further details of the text-formats exported by the R-script functions is given in the R-script itself. See Appendix 1.
The original data provided by GRDC is further described in the following section:
Station metadata is distributed by GRDC as a compressed MS Excel file on the URL: ftp://ftp.bafg.de/pub/REFERATE/GRDC/catalogue/grdc_arctichycos_stations.zip

The station metadata contains a table with the following information:

Table 5.5.1 Column key to the Arctic-HYCOS station metadata provided by GRDC

Column	Name	Description
1	grdc_no	GRDC station number
2	wmo_reg	WMO region
3	sub_reg	WMO subregion
5	nat_id	national station ID
6	river	river name
7	station	station name
8	country_code	country code (ISO 3166)
9	lat	latitude, decimal degree
10	long	longitude, decimal degree
11	area	catchment size, km2
12	altitude	height of gauge zero, m above sea level
13	ds_stat_no	GRDC station number of next downstream GRDC station

15	d_start	daily data available from year
16	d_end	daily data available until year
17	d_yrs	length of time series, daily data
18	d_miss	percentage of missing values (daily data)
19	m_start	monthly data available from
20	m_end	monthly data available until
21	m_yrs	length of time series, monthly data
22	m_miss	percentage of missing values (monthly data)
23	t_start	earliest data available
24	t_end	latest data available
25	t_yrs	maximum length of time series, daily and monthly data
26	lta_discharge	mean annual streamflow, m3/s
27	r_vol_yr	mean annual volume, km3
28	r_height_yr	mean annual runoff depth, mm

The GRDC station number (grdc_no) and the national station id (nat_id) provides the link to the historical time-series provided by GRDC and the provisional data provided by the NHS, respectively. Historical daily mean discharge and monthly mean discharges are distributed by GRDC on the following URLs:

- ftp://ftp.bafg.de/pub/REFERATE/GRDC/ARC_HYCOS/arc_hycos_day.zip (daily data)
- ftp://ftp.bafg.de/pub/REFERATE/GRDC/ARC_HYCOS/arc_hycos_mon.zip (monthly data)

The compressed archives contain sub-folders with comma-separated (;) text-files for daily and monthly mean river discharge data, respectively. Data are provided in one file per station identified by the GRDC station number on the file name, for example:

- /archycos_day/2903087_Q_Day.Cmd.txt (daily data for GRDC station 2903087)
- /archycos_month/2903087_Q_Month.txt (monthly data for GRDC station 2903087)

The original files for daily data provided by GRDC the following structure:

- Metadata and additional remarks are given in initial lines commented with a leading #.
- The column header is given in the first data line after the comment line:
    - (daily data) YYYY-MM-DD;hh:mm; Value
- The daily mean discharge data is given in the column Value with unit m3/s.

The original files for monthly data provided by GRDC the following structure:

- Metadata and additional remarks are given in initial lines commented with a leading #.
- The column header is given in the first data line after the comment line:
    - (monthly data) YYYY-MM-DD;hh:mm; Original; Calculated; Flag
- The column Original contains the monthly mean discharge data (unit m3/s) provided to GRDC by the original data providers (NHS).

- The column Calculated contains monthly mean discharge (unit m3/s) calculated from daily mean discharge data if available.
- The column Flag contains the percentage of days with daily data used for the calculated monthly mean value.

Appendix 1 – R script functions to simplify the integration of Arctic-HYCOS river discharge data in the iAOS (https://github.com/dgustafsson/arctic-hycos-iaos-utils)

6. Further steps for Data integration from existing repositories into the iAOS

In order to further advance the integration of the data providers listed in Figure 1.1.1 and in Chapter 3.2, it will be necessary to take various further steps. On a short term basis - with regard of termination of WP2 obligations until end of May 2019 - we aim to focus on wishes and problems for data integration of WP2 partners. Also on this short term basis we aim to finalize the integration of described 'show cases' into the iAOS.
Starting from the first implementation phase, it should then be somewhat easier to carry out further integrations into the iAOS. On a long term perspective two factors will be key drivers for selecting systems and data-collections for integration during the second implementation phase. First, if a system has already been listed in the description of work for integration. Secondly, as already described, whether a data collection or a data provider is highly relevant for the tasks in WP6.

6.1 partner related actions

In order to promote the transfer of know-how between the partners, we have already set up a Partner Task Force, which will initially focus on the exchange of experience on openDAP and NetCDF. In general, this task force aims on making the knowledge of experienced partners available to less experienced partners. In a first web session in mid-November, 13 partner institutions took part and had an intensive exchange. In a further step, a Data Management Training Workshop is planned for the coming General Assembly in January 2020, based on this web session.
As described in chapter 3 the quality of technical information provided by partners has very heterogeneous quality. For this reason it will be necessary to continue the lengthy task of single consultation and communication in order to complement technical information gaps and also to provide individual support to partners.

6.2 system related actions

One essential factor will be to provide the end user with a system that is as easy to use as possible. One very important aspect is the already realized implementation of the INTAROS online-catalogue which should make it as easy as possible for the end user to find out which data collections will be available for use in iAOS. Through a deeper integration into the iAOS, the system-specific discovery (i.e. search) and retrieval (i.e. access) framework will be directly supported. In the end the aim is to provide tools which support data access to the relevant data-infrastructures.
In this context, it will be essential to develop and provide of tools for multi-dimensional data sub-setting and basic processing (e.g. statistics extraction, spatial/temporal rescaling, temporal merging, gap filling, formatting, compositing, etc.). Some of these tools are already available in the first iAOS developments phase and can be tried out in connection with the show-cases.

6.2.1 necessary adaptations of the iAOS itself – findings from the assessment and from interaction with partners

(This chapter will be filled with corresponding findings in the further course of the project.)

---------- 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: