EXTREME EARTH

ExtremeEarth

H2020 - 825258

Deliverable

D5.5

Implementation and evaluation of the Polar use case - version II

William Copeland, Alistair Everett, Nick Hughes

16.12.21

Status: Final
Scheduled Delivery Date: Dec2021

Executive Summary

The Polar use case aims at producing high resolution near real-time ice monitoring products from vast volumes of satellite data with a cloud-based approach. To this aim, we present in this deliverable an updated version of the D5.3 demonstrator, focussing on the integration of a deep learning schema with the data analytics and machine learning platform Hopsworks, and PolarTEP, the thematic exploration platform for polar uses. More specifically, we describe how a PolarTEP processor is prepared and how said processor is published, including a step by step user guide for using a published processor on the Polar TEP web interface. A reminder of the Polar Code Decision Support System concludes the demonstrator.

Document Information

Contract Number	H2020 - 825258	**Acronym**	ExtremeEarth
Full title	From Copernicus Big Data to Extreme Earth Analytics		
Project URL	http://earthanalytics.eu/		
EU Project officer	Riku Leppänen		

Deliverable	**Number**	D5.5	**Name**	Implementation and evaluation of the Polar use case - version II
Task	**Number**	T5.5	**Name**	Implementation and evaluation
Work package	**Number**	WP5		
Date of delivery	**Contractual**	M18	**Actual**	M18
Status	Draft☐ final ☒			
Nature	Prototype ☐ Report ☒			
Distribution Type	Public ☐ Restricted ☐ Consortium ☒			
Responsible Partner	METNO			
QA Partners	VISTA			
Contact Person	Nick Hughes			
	Email nick.hughes@met.no	**Phone**		**Fax**

Project Information

This document is part of a research project funded by the IST Programme of the Commission of the European Communities as project number H2020-825258. The beneficiaries in this project are the following:

Partner	Acronym	Contact
National and Kapodistrian University of Athens Department of Informatics and Telecommunications (Coordinator)	UoA	Prof. Manolis Koubarakis National and Kapodistrian University of Athens Dept. of Informatics and Telecommunications Panepistimiopolis, Ilissia, GR-15784 Athens, Greece Email: koubarak@di.uoa.gr
VISTA Geowissenschaftliche Fernerkundung GmbH	VISTA	Heike Bach Email: bach@vista-geo.de
The Arctic University of Norway Department of Physics and Technology	UiT	Torbjørn Eltoft Email: torbjorn.eltoft@uit.no
University of Trento Department of Information Engineering and Computer Science	UNITN	Lorenzo Bruzzone Email: lorenzo.bruzzone@unitn.it
Royal Institute of Technology	KTH	Vladimir Vlassov Email: vladv@kth.se

National Center for Scientific Research - Demokritos	NCSR-D	Vangelis Karkaletsis Email: vangelis@iit.demokritos.gr
Deutsches Zentrum für Luft-und Raumfahrt e. V.	DLR	Corneliu Octavian Dumitru Email: corneliu.dumitru@dlr.de
Polar View Earth Observation Ltd.	Polarview	David Arthurs Email: david.arthurs@polarview.org
METEOROLOGISK INSTITUTT	METNO	Nick Hughes Email: nick.hughes@met.no
Logical Clocks AB	LC	Jim Dowling Email: jim@logicalclocks.com
United Kingdom Research and Innovation – British Antarctic Survey	UKRI-BAS	Andrew Fleming Email: ahf@bas.ac.uk

Table of Contents

Table of Figures

1. Background and aims of the Polar Use Case

The polar regions play a critical role in regulating and driving the Earth's climate and ecological systems, and are currently experiencing significant change. New economic opportunities are resulting in increased attention and vessel traffic, leading to growing global interest both politically and economically. The anticipated economic developments in the Arctic, partially driven by reductions in sea ice cover, will see an increase in maritime shipping activity. High quality, timely and reliable information about sea ice and iceberg conditions is vital to ensure that vessels navigate efficiently and safely with minimal risk to the environment. This information is required by vessels in many sectors, including cargo transport, fisheries, tourism, research vessels, resource exploration and extraction, destination shipping and national coast guard vessels.

The new International Maritime Organisation (IMO) Polar Code now mandates that "ships shall have the ability to receive up-to-date information including ice information for safe navigation" and requires a risk assessment methodology (currently referred to as POLARIS) to determine the limitations for operating in ice conditions. Currently, sea ice information is provided to mariners in the form of ice charts issued by national ice agencies. Manual construction of ice charts results in time delays and in some cases out of date information. In addition to ice charts, many ships choose to access satellite imagery directly in order to obtain up-to-date information about sea ice conditions, however this requires a crew with relevant interpretation experience.

Collection of observational and in situ data is constrained by the remote and hostile environments of the Polar Regions. Satellite imagery provides the only source of consistent, repeatable, regional-scale, year round data covering the polar regions. Currently, parameters extracted from satellite imagery still lack the necessary veracity. High resolution ice operational ice charting relies upon synthetic aperture radar (SAR) data due to its independence from weather and light conditions, supported by information from other sources (optical, microwave, radiometer data, and sporadic in-situ). Copernicus satellites provide new remote sensing data for the basis of operational services, which greatly expands upon previous national Earth observation programmes.

The Polar use case aims at producing high resolution near real-time ice monitoring products from vast volumes of satellite data with a cloud-based approach. There is a need to develop new methods to automatically extract the information from satellite data in near real-time, so that it can be used to produce ice charts for maritime users with little delay. It is also necessary to provide an operational platform that provides the necessary tool box for researchers and casual users, who are not experienced with satellite or sea ice data, or have their own cloud/HPC capacity to experiment with bulk processing of satellite data. This aims to foster the development of new methodologies and techniques, particularly to address the temporal changes in sea ice surface characteristics, that lead to greater use of the Sentinels and future satellite missions.

In this deliverable, we present an updated version of the D5.3 demonstrator, focussing on the integration of a deep learning schema with the data analytics and machine learning platform Hopsworks, and PolarTEP, the thematic exploration platform for polar uses. More specifically, we present the process of PolarTEP processor creation, publishing a processor, and a step by step user guide for using a published processor on the Polar TEP web interface. This process of raw satellite data to classified image using a web based tool integrating machine and deep learning algorithms covers aims of the Polar Use Case outlined above. A reminder of the Polar Code Decision Support System concludes the demonstrator.

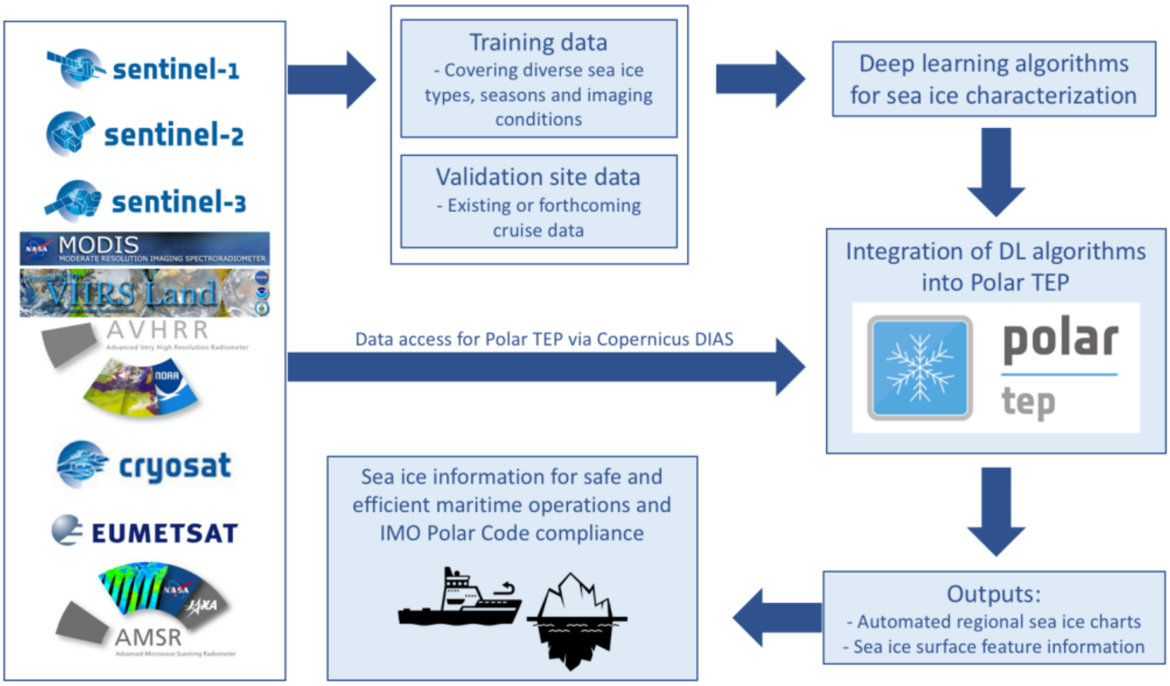

Figure 1: A schematic showing the aims of the Polar Use Case

2. AI techniques on sea ice imagery for the Extreme Earth Polar Use Case

For the purpose of this demonstrator, we will not go into detail for each individual algorithm developed for the Polar Use Case. Please see WP2, D5.3 section 2 and D2.5 - Evaluation report on Polar use case output products (M36) for full description and validation of these models which are summed up in figure 1.

2.1 Polar Use Case training data sets

MET Norway created a set of sea ice and iceberg analysis data for the ExtremeEarth project for training and validation of automated methods for processing satellite images and additional applications for ice analyst training (Hughes, N., 2020). An updated training set (2.0.0) was released April 13th 2021 using the satellite imagery from the same area (Danmarkshavn, Eastern Greenland) continuously monitored by key European Sentinel satellites during 2018. The updated version now contains the additional information on sea ice concentrations, ice type and form that was lacking in version 1.0.0.

Training data sets were also created by University of Tromsø (UiT Dataset V1.2) from north of Svalbard including information on open Water, leads with water, brash/ pancake ice, thin ice, thick ice-flat and thick ice ridged from Sentinel-1 images in 2018.

DLR developed two training datasets based on 24 Sentinel-1 images located in the Belgica Bank, an area of the Greenland Sea just to the north of the Danmarkshavn area. The first dataset (Dumitru, O. 2021) consists of ~62M patches of size 4x4 pixels where each patch is labeled with one of 11 topics using Latent Dirichlet Allocation. The second dataset consists of 153,600 patches of size 256x256 pixels classified into 8 classes using active learning based on support vector machines with relevance feedback (Karmakar, C. 2021).

2.2 Algorithms developed for the Polar Use Case

Three supervised deep learning methods developed by UiT were considered; an ad hoc Convolutional Neural Network (CNN) architecture, another CNN architecture named VGG-16 (Khaleghian et al, 2021, Simonyan and Zisserman, 2014), and a Fully Convolutional Network architecture named FCN-8. A semi-supervised label propagation architecture also developed at UiT (Khaleghian et al, 2021) and an unsupervised Latent Dirichlet Allocation (LDA) model developed by DLR (Dumitru et al, 2019) . In this demonstrator, we used UiT's VGG16 CNN architecture to present as a demonstration.

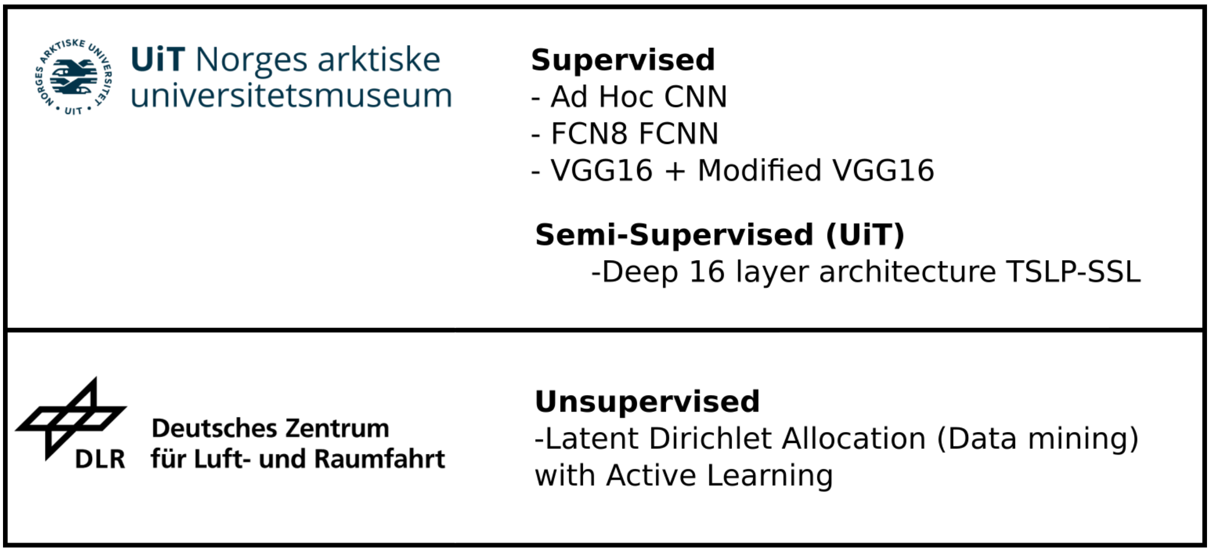

Figure 2: A list of the algorithms developed for the Extreme Earth Polar Use Case.

3. Second version of Polar use case running on Polar TEP and DIAS

The Polar TEP's main purpose is to address the growing complexity and volume of Copernicus big data for the polar regions. Allowing industry representatives, scientists, operational service providers, regional authorities and policy makers to access and exploit the data on an all-in-one web platform. The Polar TEP provides a complete remote working environment for the user, in which access to algorithms, with corresponding computing resources and tools, avoids the need for individual users to download and locally manage large volumes of data. The platform also opens up an arena for polar researchers from all fields to collaborate and facilitate the development of new algorithms, data products and processors, thus encouraging multifaceted exploitation of Earth observation data. Users can visualize their own data processors and products, which can be added to the shared environment, allowing for cross-disciplinary usage of resulting products. Anticipated Sentinel expansion missions, along with development of new polar-focused instrumentation, will only increase the demand of functionality that Polar TEP can deliver.

Now installed on the CREODIAS infrastructure as part of the Extreme Earth project, Polar TEP has access to a data repository that includes Sentinel-1,2, 3, and 5-P, Landsat-5, 7, 8, Envisat, and many other Copernicus services in a scalable framework. CREODIAS is a cloud infrastructure that is specially adapted to process big earth observation data, supporting operational, research and commercial applications. By combining the large data repository with customer accessible big processing power, CREODIAS allows for a cluster of sustainable commercial services to be developed. In terms of the Polar TEP, the CREODIAS infrastructure is the primary source of all raw Sentinel-1 satellite data, and users of Hopsworks (also on CREODIAS) can directly access this raw data from their services and applications that run on the Hopsworks platform. This object store is primarily used in the Polar use case for storing training data which is fed to the deep learning pipelines.

The ExtremeEarth infrastructure has been architectured around the open source platform Hopsworks. This scalable platform allows for development and operation of end-to-end machine learning (ML) and deep learning (DL) models on both an on premises platform (open-source or enterprise version), and as a managed platform on AWS. Since D5.3, Hopsworks has been migrated to the CREODIAS infrastructure, reducing the significant network overheads at the prediction stage.

3.1 Introduction to Hopsworks

The Hospworks platform offers:

- End to End ML & DL pipeline - Feature store to training to serving;
- Management of ML assets: Features, Experiments Model Repository/Monitoring;
- Project based multi-tenancy - collaborate with sensitive data in a shared cluster;
- Enterprise integrations with active directory, LDAP, OAuth2 and Kubernetes;
- Full Governance and provenance for ML assets, with GDPR compliance;
- Open-source ML with Spark, TensorFlow, PyTorch, Scikit-Learn;
- Solving the hardest scaling problems: training, hyperparameter tuning, feature engineering.

Hopsworks is always involved in the training phase within the framework of the ExtremeEarth project. In D5.3, it was decided that the most reasonable implementation for the Polar use case was to train the model on Hopsworks and export the model to disk. The model can then be embedded into a processor that can run on the Polar TEP, completely independent of Hopsworks. Even with Hopsworks now on the CREODIAS infrastructure, model serving via a REST API (Case 1 - D5.3) would not be as robust compared to in-memory processing.

3.2 Preparing a Polar TEP processor part 2

The Polar TEP supports two types of users; (i) the (potentially) non-technical end user who consumes information by requesting that a satellite image is to be processed using a predefined processor, and (ii) the technical user or algorithm developer who may want to make such a processor available to end users. Please refer to sections 1.1.1 and 1.1.2 in D5.3 where the process of making ExtremeEarth algorithms available to the platform is described in great detail. A summation of the process can be seen in figure 2.

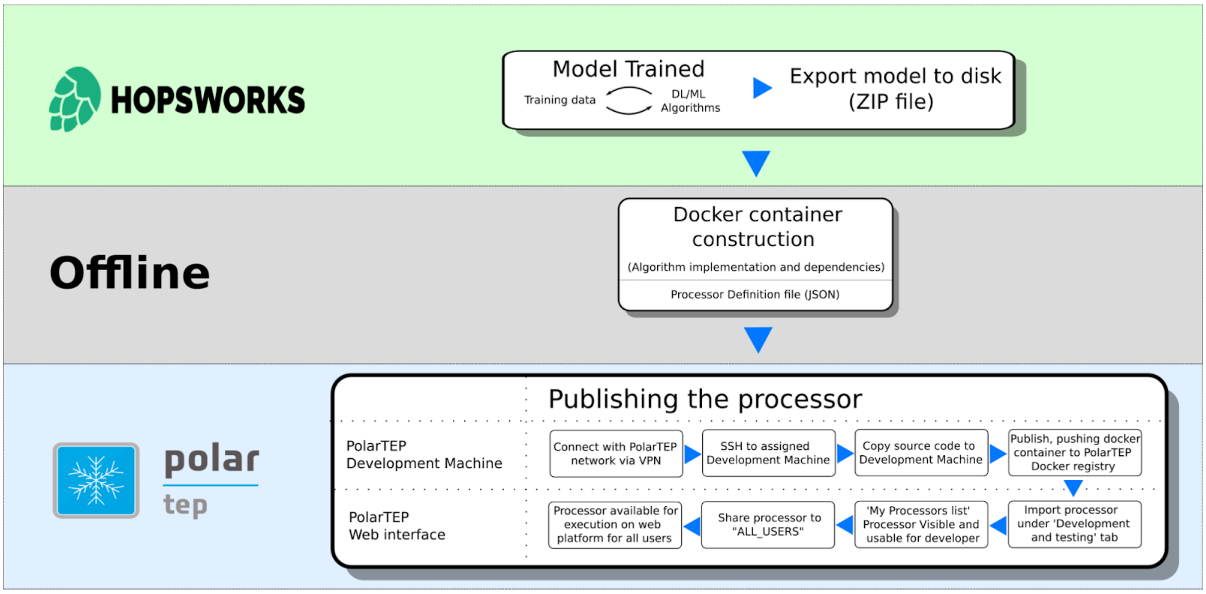

Figure 3: Workflow from training the model in Hopsworks to publishing the processor to the PolarTEP

The Docker container can be developed on any machine, for example one's own laptop, making it a dynamic tool for researchers globally. In this demonstrator, all pre-processing, post-processing, and import/export is handled using the ESA Sentinel-1 Toolbox and Python running within the Docker container on the Polar TEP platform.

Figure 3 shows the basic pipeline of how an end user would interact with the processor on the Polar TEP platform, and the flow of events that take place once a processor operation is executed.

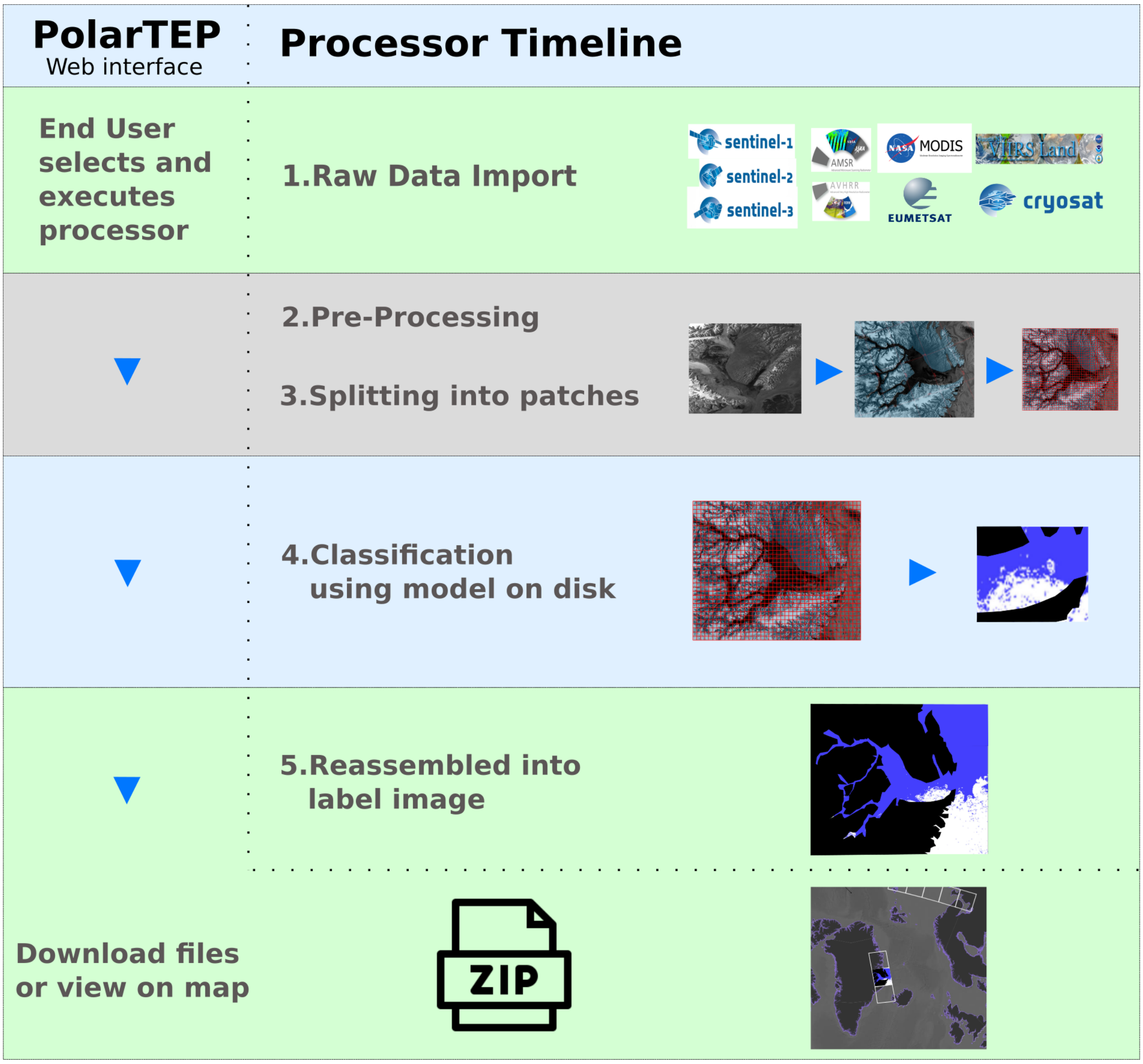

Figure 4: User interaction within the Polar TEP platform and processor workflow. On the web interface itself, it is simply a matter of executing a processor and waiting for the processor to classify the image (s) before downloading/viewing the result. The processor timeline all takes place hidden from view, but what is important to note is that the classification takes place using the model on disk.

3.3 Notes about the implementation

Training the model on Hopsworks and exporting the model to disk would be the process used at MET Norway. The same docker container developed for use on Polar TEP can also be run on MET's post processing infrastructure PPI with singularity.

The current demonstrations are utilizing CPUs for the prediction step, but assuming that GPUs are available and set up correctly in the Polar TEP execution environment, enabling GPU support should only be a matter of building the Docker container with GPU support in the TensorFlow Python library.

3.4 Using the published processor in PolarTEP

In this section we present a step by step demonstration of how an end user would interact with the Polar TEP web interface. Before using the platform, the user must register for an account with Polar TEP, which is verified by email and soon activated by a moderator. Once signed into the portal, the user selects '*Data Search and Processing'* located in the header, and is presented with a map of the Arctic region. In figure 4 we see the platform interface and the first three steps taken in order to access the satellite imagery the end user wishes to process.

Firstly, data provenance is chosen by selecting, for example, Sentinel-1 Ground Range detected, Medium Resolution (GRDM) imagery. The user then sets the date range and area of interest (AOI). For date range the user can use '*absolute'* which allows date constraints to be set in calendar format. While under '*relative',* minutes, hours, days or weeks previous can be chosen as date constraints. On pressing enter, polygon boxes will appear on the map showing the location of imagery in the selected date range. The associated granules will also appear in the '*Granules*' section on the bottom right of the page. However, many end-users will want to specify the region in which to conduct the search. This is done by selecting the drawing tool under '*Bounding Box',* which allows the drawing of a polygon on the map specific to the interest area. Advanced users can choose their location by inputting a WKT under '*Complex/WKT'.* Now when pressing enter, the granules that fully or partially cover the selected region are shown on the map.

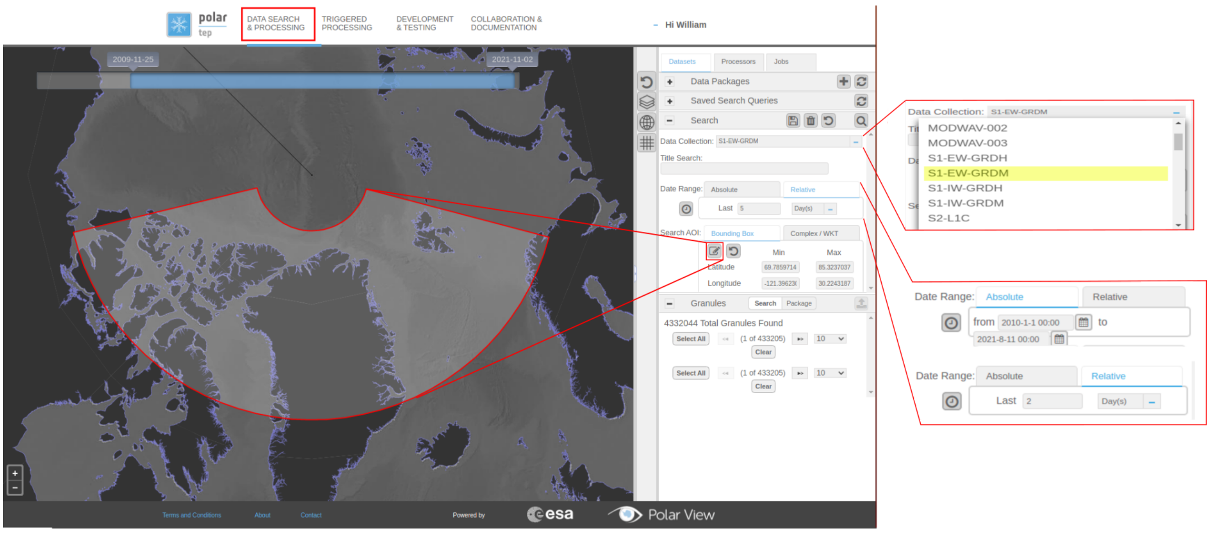

Figure 5: The first two steps taken by the end user which includes selecting *'Data collection'* , '*Date range'* and drawing a boundary box over their preferred area or entering the *WKT*.

Once a region has been selected, the user can look through the granules that cover the chosen region, and select the ones they wish to send to the processor. If using only one image, they can skip creation of a data package which is represented on the right hand side of figure 5. However, the user may wish to create a package of more than one granule, when dealing with multiple, overlapping sets of images. To do this, the user highlights the granules they wish to assimilate into a data package and selects the right hand side + symbol on the *'Data packages'* header. This will bring up a box where they can fill in a *'user key* (title)*'* and the *'info* (Basic description)*'*. After pressing *'save'*, the package is created and stored under *'Data packages'* for later use.

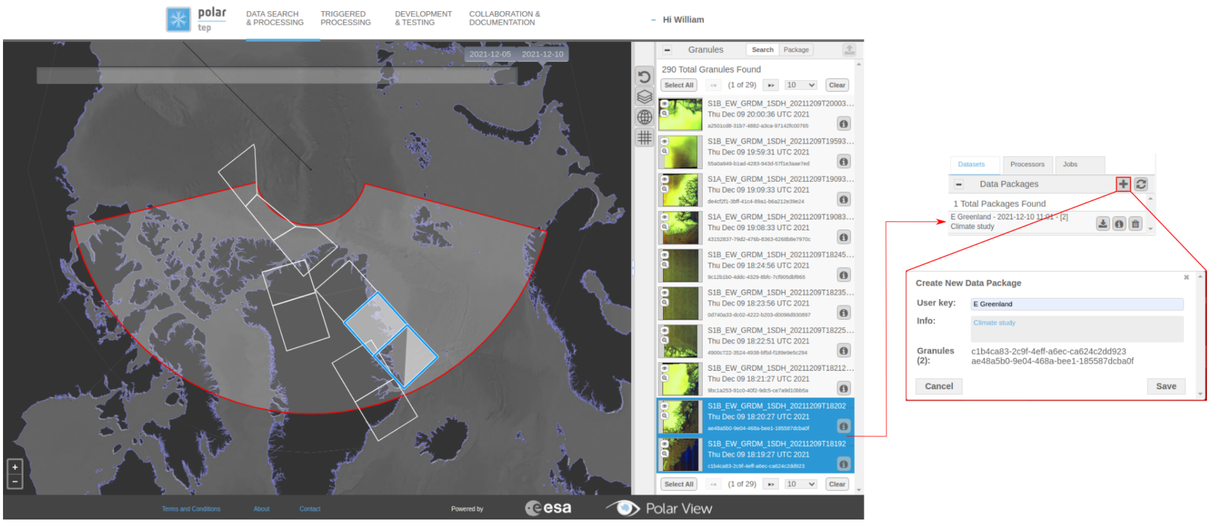

Figure 6: Selecting granules (images) for creation of a data package which can later be used as the '*input file*' for the processor.

In figure 6, the *'processors'* tab located next to *'datasets'* on the top right header box is selected. This opens up a new side window with listings of all available processors. Advanced users who are using their own imported algorithms can find these under the header *'My Processors'.* Users can select the processor of which they would like to use by simply clicking on it, which produces the *'Processing Parameters'* box to the bottom right of the screen.. Under *'Processing Parameters'*, the data package created earlier, or single files from the granules menu, are dragged into the empty *'infile'* box. Making sure that the *'is URL?'* is selected. If dealing with more than one input file, users select *'is Batch Parameter'* and then *'continue on error'* in the '*Batch error behavior'* section. This allows the user to be able to view the processing stages of each image being processed, and to stop delays caused by single file processing failures in the queue. Parameters can be changed according to the needs of the user.

Figure 7: Selection of a processor and input file within processing parameters. More parameters can be chosen by the docker container developer if desired.

Once the input file and parameters are set, the user can select to execute the processor as seen in the bottom right window in figure 6. As the processor starts to run, the user selects the *'Jobs'* tab in the top header menu, showing the current and previously run jobs, along with the processor output (figure 7). A notification email is sent to the user's email account as to whether the job was successful or a failure. It is then possible to download the result files or inspect logs from the standard output and error streams of the docker container.

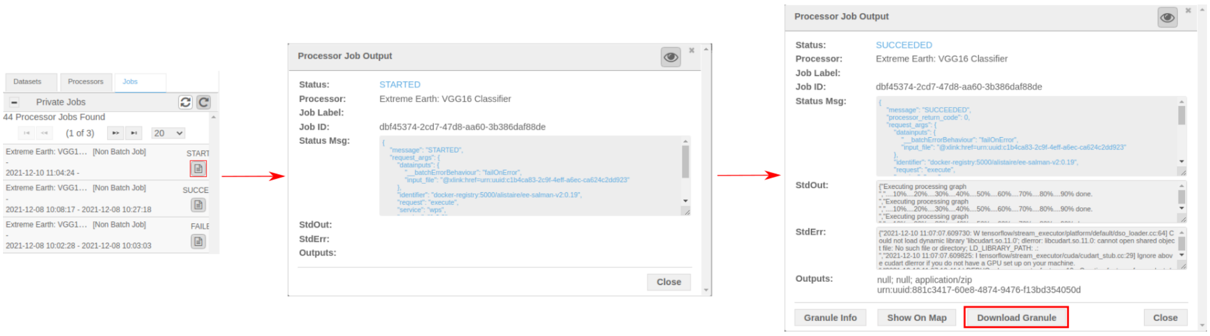

Figure 8: Under the *'jobs'* tab, the user can keep track of the progress of the processor. When successful, options are given to view the output, and for this deliverable we choose to download the granule.

In this example, the result is a zip file containing two 32 bit and 8 bit GeoTIFFs, each containing the HH band from the Sentinel-1 image and the classifier output where 0=no data, 1=water, 2=Sea ice, and 3=Land. The user can import the extracted zip file into any GIS software. In this example we use QGIS as is outlined in Figure 8.

Once imported into QGIS, layer properties can be changed to better view the classified image in isolation. Figure 8b shows the process of selecting Band 2 (classified image), changing the render type, selecting minimum and max values, choosing a preferred colour range and selecting the *'equal interval mode'* with 4 classes. You can then *'classify'*, *'apply'* and press *'ok'* to view the classified image (8c). In the classified image the end user can see the distinct boundaries between land, sea ice and water.

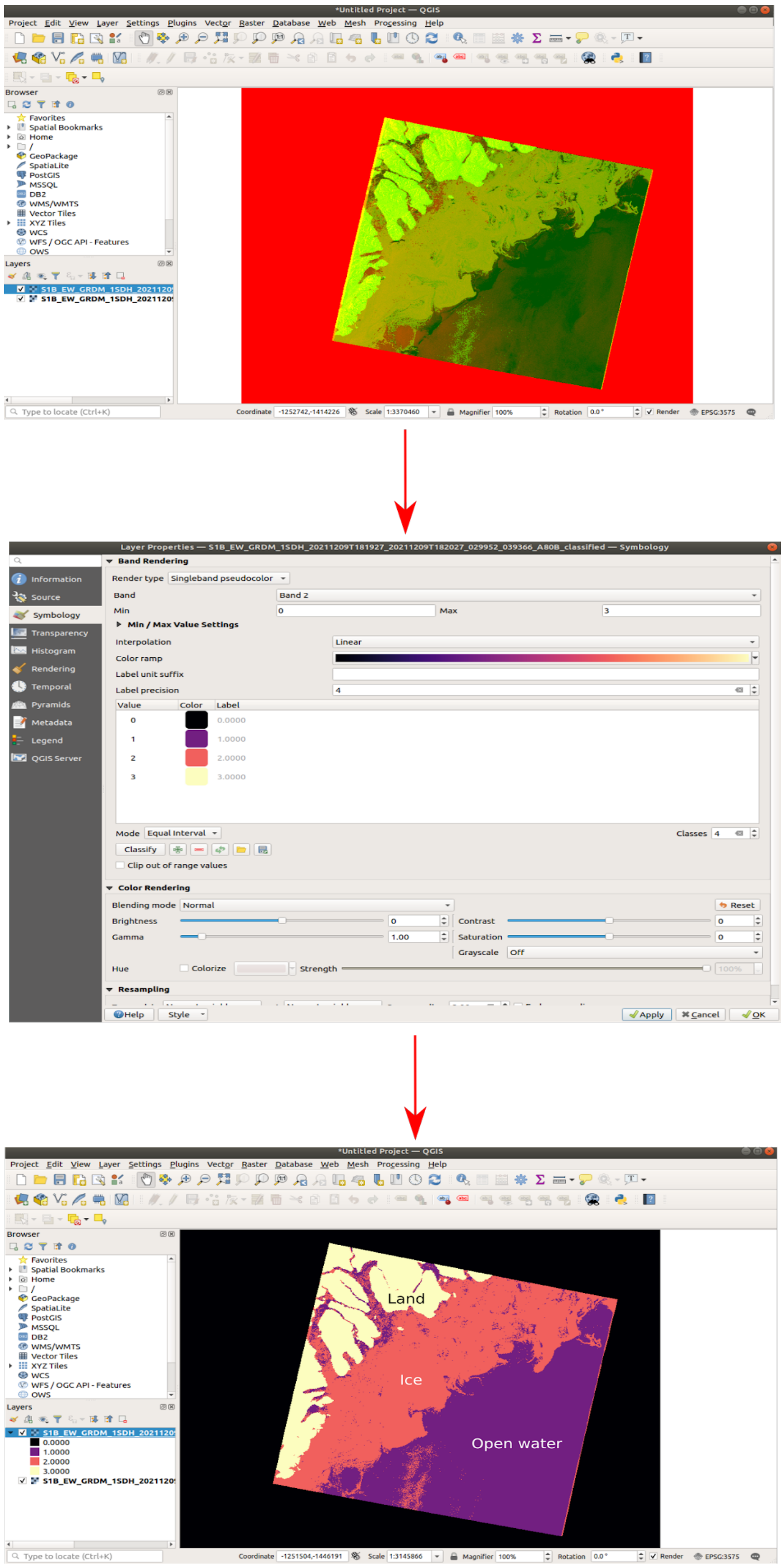

Figure 9: Importing the zip file into QGIS and editing layer properties to view the classified image.

3.5 Plans for operational use of the implemented techniques

The options for implementing these techniques in MET Norway's operational systems are outlined in Deliverable 5.2, Section 4.2.2. Downloading and preprocessing of satellite imagery would be done in-house, which is already carried out by ice analysts who currently produce ice charts. Images would be downloaded, pre-processed and classified on the same infrastructure operated by MET Norway and the docker containers created for the Polar TEP would also be able to run on the internal infrastructure at MET with singularity.

4. Linked data integration

An important objective of ExtremeEarth is to extend the capabilities for Earth Observation (EO) access and data discovery with semantic catalogue services that scale to the big data, knowledge and information of Copernicus data. This allows us to interlink these resources with other linked geospatial data and produce thematic maps for the Polar TEP user community.

The Polar Ontology. In deliverable *D1.3 Semantic catalogue design and implementation - version I* the design and implementation of the first version of the polar semantic catalogue was presented. This ontology covers the Copernicus and other data exploited by the polar use case. Please see D5.3 section 1.6 for a full background and aims for linked data integration in the polar use case. New additions to the ontology include ice observation classes corresponding to concentration ranges (fast ice, very close drift ice, open drift ice, very open drift ice, open water and ice free) and an interlink with the Global Administrative Dataset (GADM).

The end-to-end workflow using the newly distributed version of the Semantic Web tools that are implemented in ExtremeEarth are presented in figure 8. Firstly a spatiotemporal join between the two data sources using JedAI is set up. Then, conversion of interlinked items into RDF triples using GeoTriples and the polar ontology takes place. These triples are finally stored in the spatially enabled triple store Strabon and are visualized with Sextant using GeoSPARQL queries, to produce a thematic map[1] for the Polar TEP users.

[1] http://test.strabon.di.uoa.gr/SextantOL3/?mapid=mkn7oa88hi0s4pea_

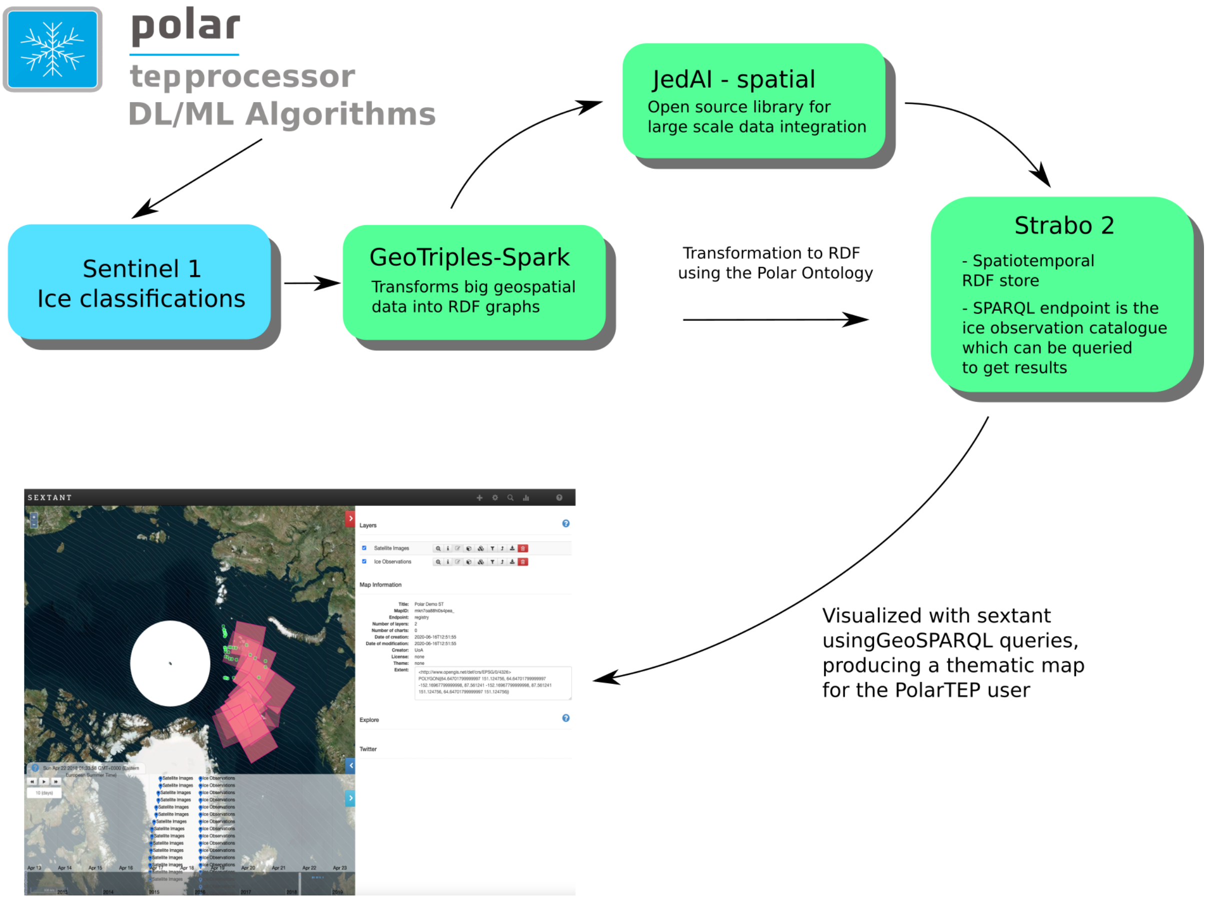

Figure 10: Diagramatic view of the end-to-end linked data pipeline employed in the Extreme Earth Polar Use Case.

5. Conclusions of the demonstrator

The process we lay out in Polar TEP is largly undocumented, and there would need to be greater investment in developmental time to make executing a processor smoother. However, the platform has succeeded in providing an arena for scientists, developers and end users to share and exploit Copernicus big data. The platform has also offered a bridge between operations and research. These are two distinctively different groups with a history of un-synchronized co-operation in terms of using the same data format specifications and standards. Greater interlinkage of data and products will allow for the research community to take advantage of the extensive ice service knowledge, while the ice service can benefit from the scientific achievements of the research community. The Polar TEP interface is one that will be expanded upon in Polar TEP 2, an ecosystem that will incorporate Hopsworks and Strabo 2 as a result of the Extreme Earth Project.

In terms of operational use, processing times must be as efficient as possible to ensure fluidity in product delivery to end users. The current automated products laid out in this deliverable are deemed currently too immature at this time. However, early results show promise for at least seasonal / regional implementation. At this juncture, we must weigh up whether these products can efficiently cut the manual time spent drawing ice charts, and at the same time keep quality and consistency. Full automation would require input of multiple environmental variables, including gradual local knowledge and ice analyst input as the product develops over time. There is little ground to support rapid implementation of fully unsupervised and unchecked automation at the operations level at this current time. However, it is perfectly acceptable to say that due to the ExtremeEarth project we have come significantly closer to developing operational techniques which could be vetted by our ice analysts.

6. Polar Code Decision Support System

The Polar View 'Polar Code Decision Support System' (PCDSS) enables vessels to meet the information requirements of the IMO Polar Code by assembling the information in one place and integrating that information:

- Across information types
- Across information providers
- Across geographic regions

PCDSS provides:

- Coverage of all polar regions (Arctic and Antarctic)
- Aggregated historical and contextual information concerning weather, ice conditions, and other auxiliary data;
- Near real-time information obtained from earth observation satellites concerning weather and ice conditions;
- Forecasted information obtained from models concerning imminent weather and ice conditions;
- Risk analysis tools; and
- Delivery of the information in intermittent and low-bandwidth communications conditions.

PCDSS will be used to deliver the ExtremeEarth Polar Use Case results to ships in the polar regions.

PCDSS normally obtains ice charts from the various national ice services either directly by FTP or via the Ice Logistics Portal (http://www.bsis-ice.de/IcePortal/). Once MET Norway incorporates the ExtremeEarth results into their operational ice charts, they will become part of the PCDSS normal data processing chain. Until then, PCDSS will obtain the ExtremeEarth results directly from Polar TEP.

7. References

C. Karmakar et al (2021), Sea Ice Data Content Representation Based on Latent Dirichlet Allocation for Belgica Bank in Greenland [Dataset], Zenodo https://doi.org/10.5281/zenodo.5075861

C. Karmakar, C.O. Dumitru, M. Datcu, "Feature-free Explainable Data Mining in SAR Images Using Latent Dirichlet Allocation", JSTARS, 2020, submitted for review.

C.O. Dumitru, V. Andrei, G. Schwarz, M. Datcu, „Machine Learning for Sea Ice Monitoring from Satellites", the International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Germany, XLII-2/W16, 2019, pp. 83-89.

N. Hughes and F. Amdal (2021) Extreme Earth Polar Use Case Training Data (Version 2.0.0) [Data set] Zenodo https://doi.org/10.5281/zenodo.4683174

N. Hughes, (2020). ExtremeEarth Polar Use Case Training Data (Version 1.0.0) [Data set]. Zenodo. http://doi.org/10.5281/zenodo.3695276

O. Dumitru et al (2021), Semantic Sea-Ice Classification for Belgica Bank in Greenland [Data set], Zenodo https://doi.org/10.5281/zenodo.5075448

S Khaleghian, H Ullah, T Krmer, T Eltoft, A Marinoni, "Deep Semi-Supervised Teacher-Student Model based on Label Propagation for Sea Ice Classification", IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, 2021

S. Khaleghian et al (2020), Synthetic-Aperture Radar (SAR) based Ice types/Ice edge dataset for deep learning analysis", https://doi.org/10.18710/QAYI4O, DataverseNO, V1

Salman Khaleghian, Habib Ullah, Thomas Kræmer, Nick Hughes, Torbjørn Eltoft, Andrea Marinoni, "Sea Ice Classification of SAR Imagery Based on Convolution Neural Networks", Remote Sensing, 2021, 13(9).

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.