POLARIN

POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

Virtual Access in POLARIN

Daan Kivits (SIOS-KC) on behalf of the POLARIN
WP5 – Provision of Virtual Access

SIOS ett UNIVERSITY OF OULU Consiglio Nazionale delle Ricerche inkode

FUNDED BY THE EUROPEAN UNION

What is Virtual Access (VA)?

in the POLARIN context

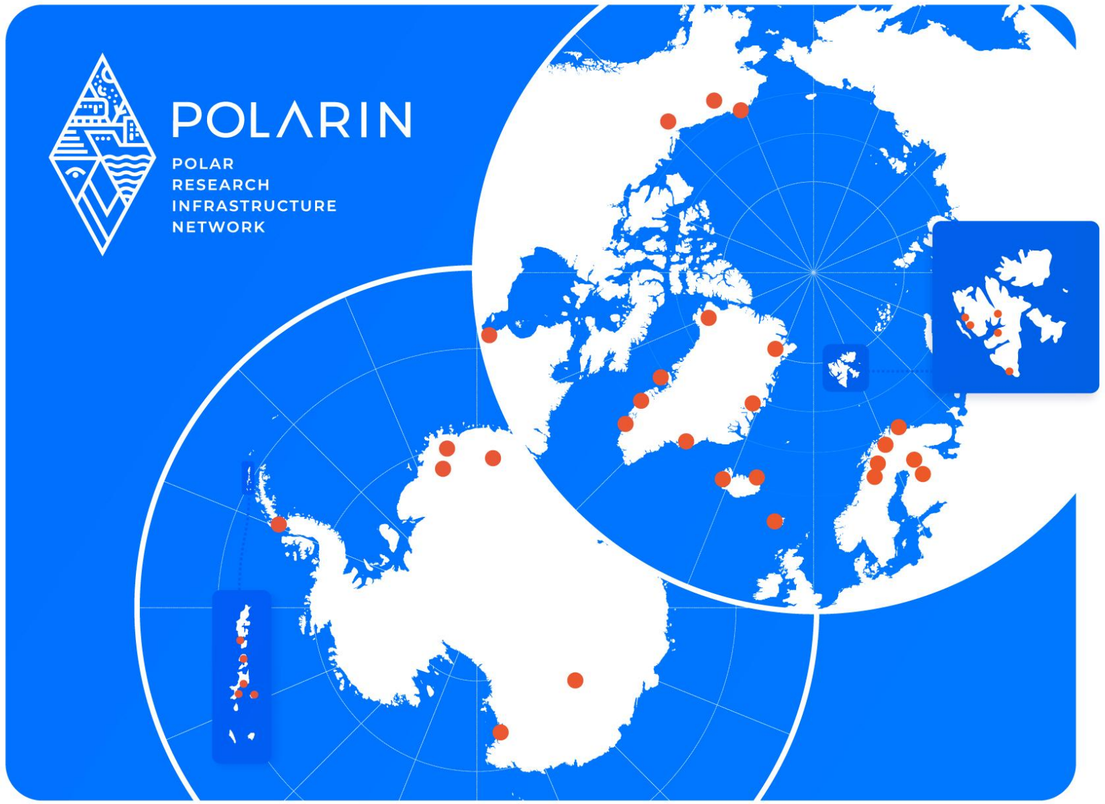
POLARIN VA

“ In the context of POLARIN, Virtual Access (VA) refers to the provision of user-friendly, free, and open online access to scientific data, metadata, and data services from select polar research infrastructures (RIs). ”

POLARIN VA

“ *In the context of POLARIN, Virtual Access (VA) refers to the provision of user-friendly, free, and open online access to scientific data, metadata, and data services from select polar research infrastructures (RIs).* ”

… and you are these *select polar RIs!*



Observation Networks

- ARC-MO: Multi-purpose mobile observatory
- GEM: Greenland Ecosystem Monitoring database

Data Infrastructure

- Arctic Biodiversity Data Service
- ARICE Data Infrastructure
- Italian Arctic Data Center
- INTERACT data portal
- Italian National Antarctic Data Center
- POSEDA
- SIOS Data Management System

Research Stations & Observatories

- Abisko Scientific Research Station
- Whapmagoostui-Kuujjuarapik Research Complex
- Dirigibile Italia Station
- Kevo Subarctic Research Institute
- Kilpisjärvi Biological Station
- Koltur Station
- Oulanka research station
- Pallas-Sodankylä Atmosphere-Ecosystem Supersite
- Tarfala Research Station
- Western Arctic Research Centre
- Concordia Station
- Mario Zucchelli Station

Objectives

POLARIN VA will:

- publish currently non-public data at your RI available in open access repositories to expose it to a broader audience
- improve the FAIR-compliance of data already existing at your RI
- aggregate all these data and serve these data in a standardized and unified manner

Why should you care?

Making your data compliant with POLARIN VA will:

- accelerate research by providing open access to historically unavailable datasets, and by making cross-institutional and cross-discipline collaborations and meta-analyses more feasible
- allow your data to be used to benefit society by automatic aggregation into operational decision-support systems, weather/climate models, and aggregated data products
- reduce the data management burden on researchers using your RI, increasing their efficiency
- improve impact of your RI by allowing for easy data citations (pro-actively preparing for envisioned ‘data index’ as alternative for ‘h-score’ index)
- make your RI more competitive for future funding rounds, as funders increasingly require open, FAIR-compliant data as a condition of grants

By being part of POLARIN VA your RI is legally obliged/committed to offer your data in a manner that is compliant with POLARIN VA requirements. You risk losing the POLARIN VA funding if not fulfilling these requirements.

Why should you care?

Making your data compliant with POLARIN VA will:

- **accelerate research by providing open access to historically unavailable datasets, and by making cross-institutional and cross-discipline collaborations and meta-analyses more feasible**
- allow your data to be used to benefit society by automatic aggregation into operational decision-support systems, weather/climate models, and aggregated data products
- reduce the data management burden on researchers using your RI, increasing their efficiency
- improve impact of your RI by allowing for easy data citations (pro-actively preparing for envisioned ‘data index’ as alternative for ‘h-score’ index)
- make your RI more competitive for future funding rounds, as funders increasingly require open, FAIR-compliant data as a condition of grants

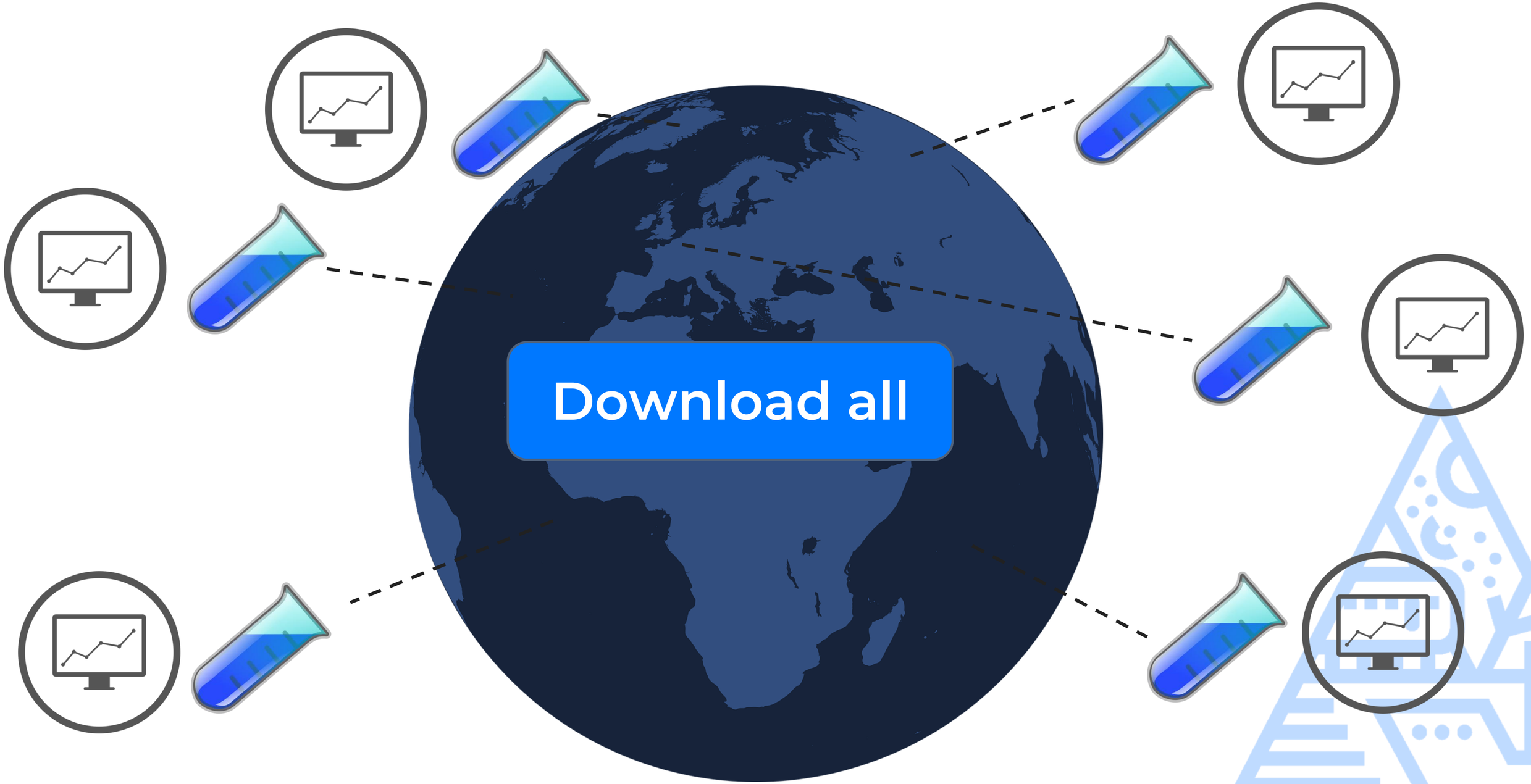
By being part of POLARIN VA your RI is legally obliged/committed to offer your data in a manner that is compliant with POLARIN VA requirements. You risk losing the POLARIN VA funding if not fulfilling these requirements.

Data availa

Data will be m request.

Data Availability Statement

The original contributions presented in t the article/Supplementary Material, further inquiries can be directed to the corresponding a


Download all

Why should you care?

Making your data compliant with POLARIN VA will:

- accelerate research by providing open access to historically unavailable datasets, and by making cross-institutional and cross-discipline collaborations and meta-analyses more feasible
- **allow your data to be used to benefit society by automatic aggregation into operational decision-support systems, weather/climate models, and aggregated data products**
- reduce the data management burden on researchers using your RI, increasing their efficiency
- improve impact of your RI by allowing for easy data citations (pro-actively preparing for envisioned ‘data index’ as alternative for ‘h-score’ index)
- make your RI more competitive for future funding rounds, as funders increasingly require open, FAIR-compliant data as a condition of grants

By being part of POLARIN VA your RI is legally obliged/committed to offer your data in a manner that is compliant with POLARIN VA requirements. You risk losing the POLARIN VA funding if not fulfilling these requirements.

Occurrence search

gbif.org/occurrence/map?has_coordinate=true&has_geospatial_issue=false&taxon_key=4261391

Get data | How-to | Tools | Community | About | Login

Occurrences

SEARCH OCCURRENCES | 3,038 WITH COORDINATES

TABLE | GALLERY | MAP | TAXONOMY | METRICS | DOWNLOAD

Licence

Scientific name

- [x] Grewia occidentalis L.

Basis of record

Year

Month

Location

- [x] Including coordinates

Administrative areas (gadm.org)

Country or area

Continent

Dataset

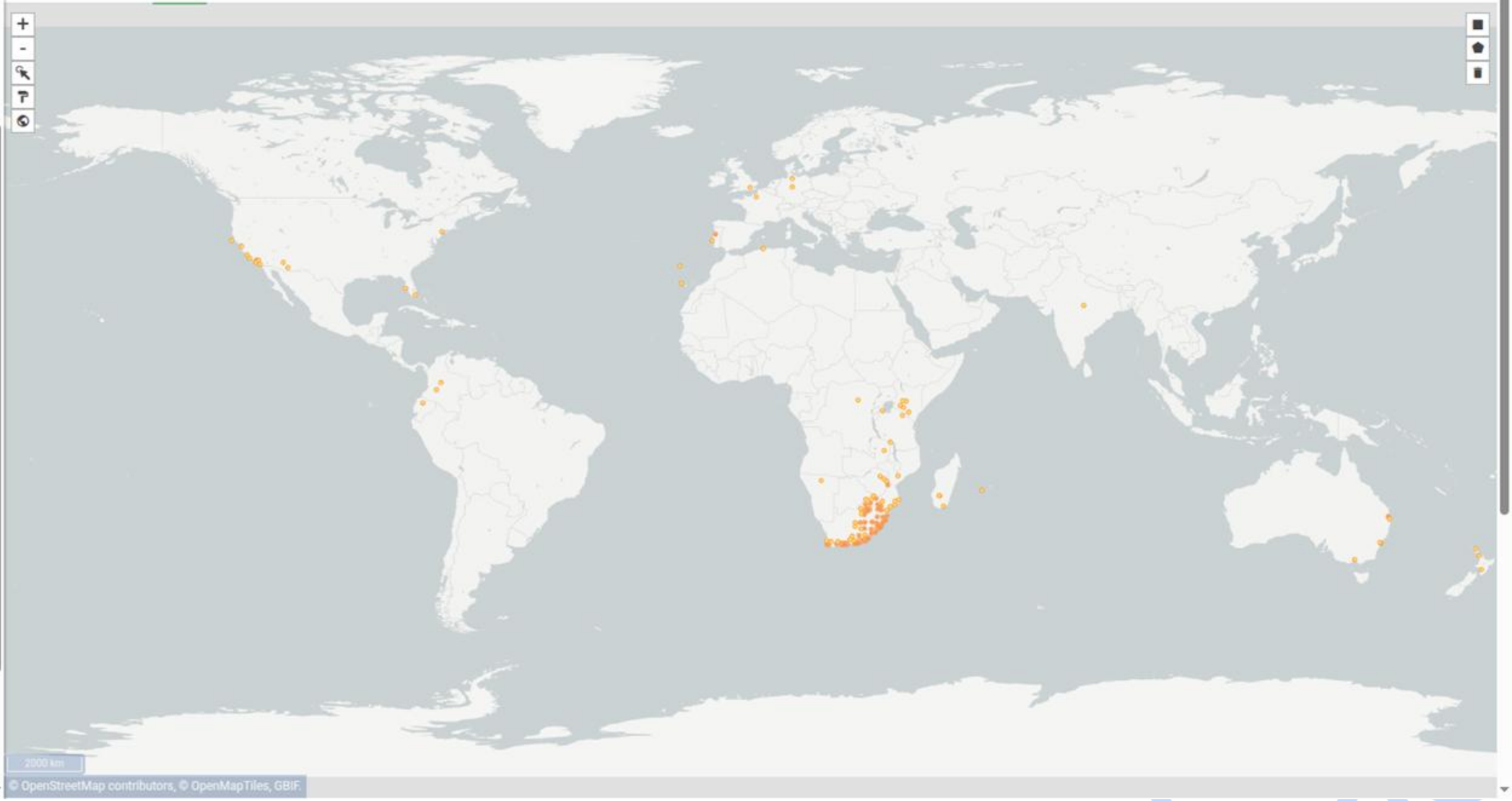
Search

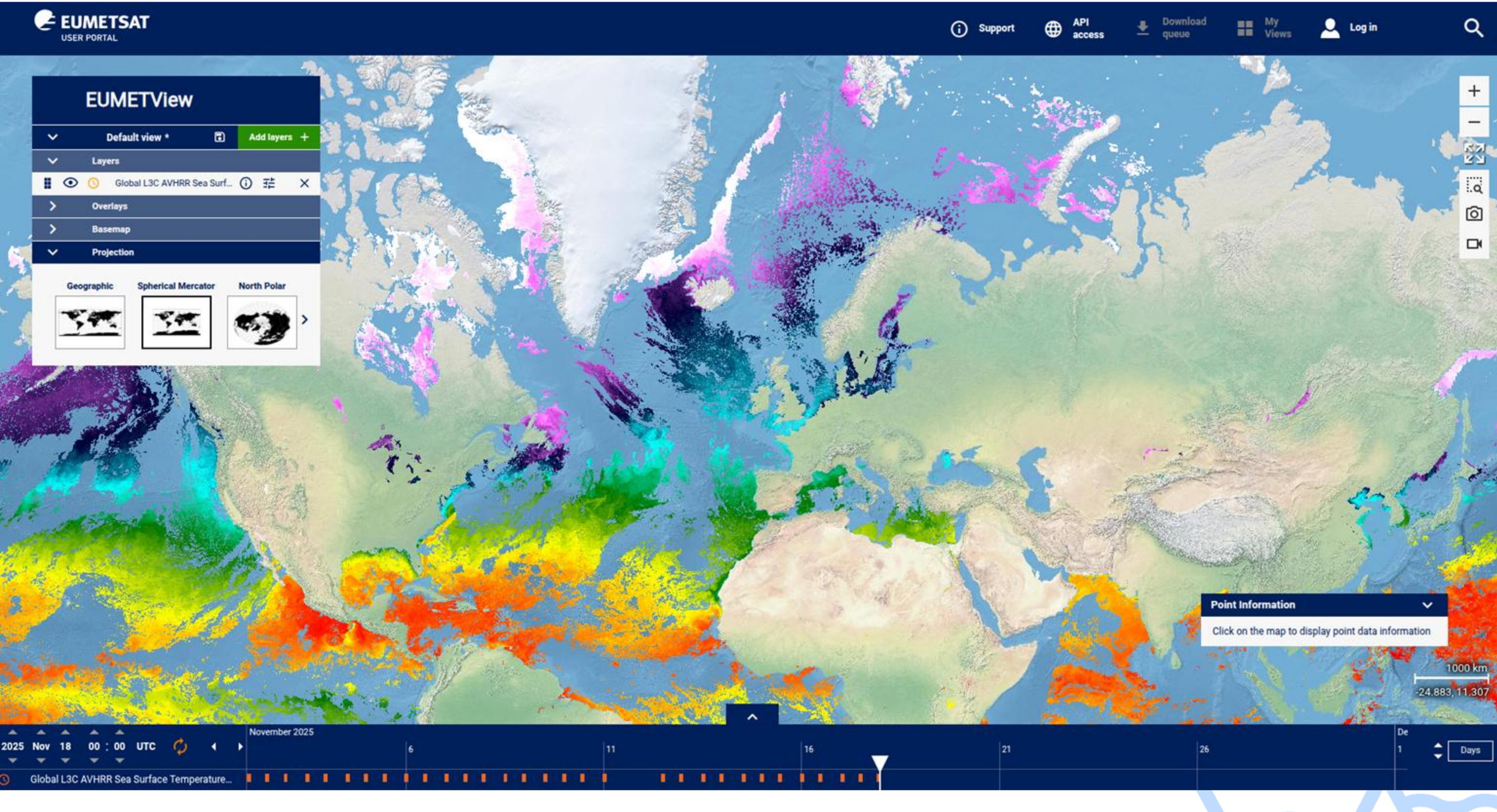
- [ ] iNaturalist Research-grade Observations 1,985
- [ ] Botanical Database of Southern Africa (BODA... 766
- [ ] Tree planting occurrences 71
- [ ] occurence_data_for_wild_food_species_in_zi... 31
- [ ] O armário: Fruiting phenology data for 4,462 pl... 21
- [ ] Pl@ntNet automatically identified occurrences 18
- [ ] National Herbarium of Victoria (MEL) AVH data 17
- [ ] IICT Herbário LISC 14
- [ ] African Plants - a photo guide 11
- [ ] Pl@ntNet observations 8

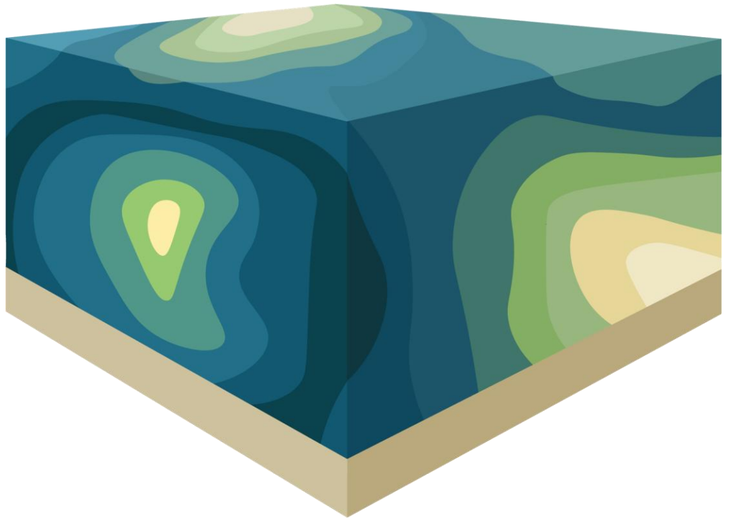
Publisher

IUCN Global Red List Category

Issues and flags




EUMETSAT
USER PORTAL
Support
API access
Download queue
My Views
Log in
EUMETView
Default view *
Add layers
Layers
Global L3C AVHRR Sea Surf...
Overlays
Basemap
Projection
Geographic
Spherical Mercator
North Polar
Point Information
Click on the map to display point data information
1000 km
-24.883, 11.307
2025 Nov 18 00 : 00 UTC
November 2025
6
11
16
21
26
1
Days
Global L3C AVHRR Sea Surface Temperature...



3D visualisation of the data

Extending existing datasets or collections

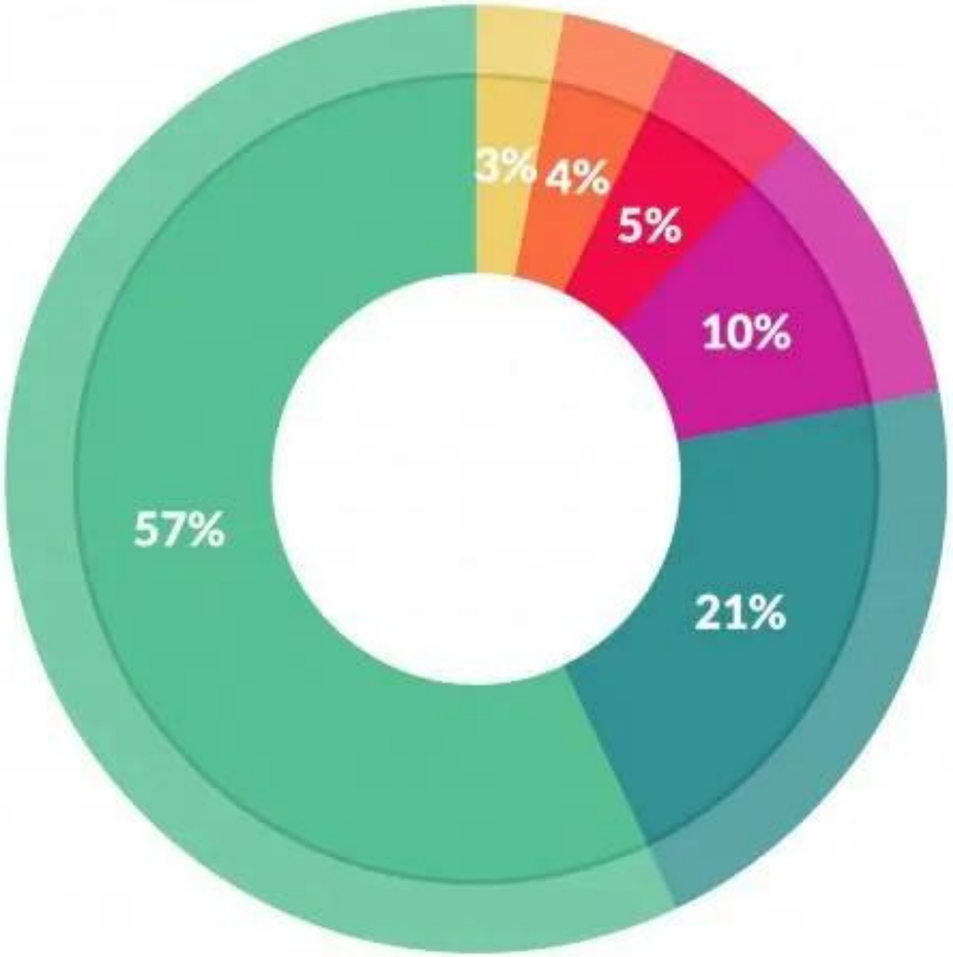
Illustrations by Frida Cnossen, Nansen Legacy

Why should you care?

Making your data compliant with POLARIN VA will:

- accelerate research by providing open access to historically unavailable datasets, and by making cross-institutional and cross-discipline collaborations and meta-analyses more feasible
- allow your data to be used to benefit society by automatic aggregation into operational decision-support systems, weather/climate models, and aggregated data products
- **reduce the data management burden on researchers using your RI, increasing their efficiency**
- improve impact of your RI by allowing for easy data citations (pro-actively preparing for envisioned ‘data index’ as alternative for ‘h-score’ index)
- make your RI more competitive for future funding rounds, as funders increasingly require open, FAIR-compliant data as a condition of grants

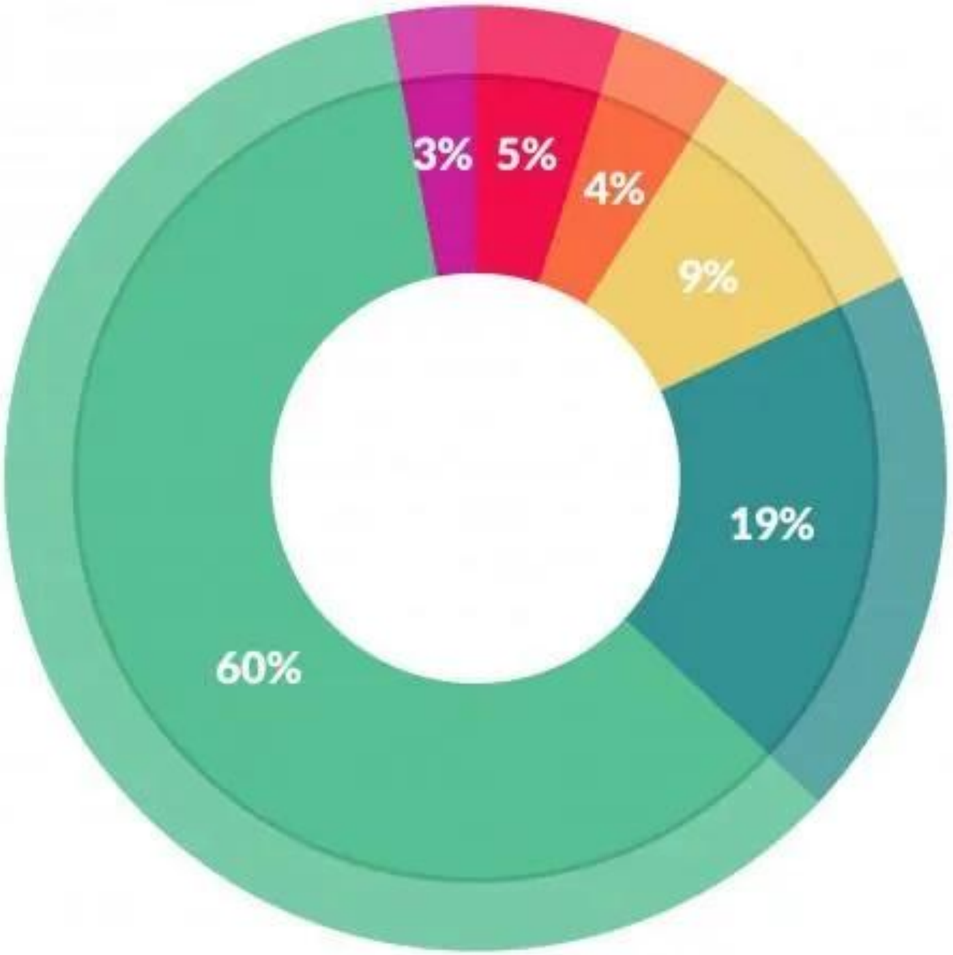
By being part of POLARIN VA your RI is legally obliged/committed to offer your data in a manner that is compliant with POLARIN VA requirements. You risk losing the POLARIN VA funding if not fulfilling these requirements.



What's the least enjoyable part of data science?

- *Building training sets: 10%*
- *Cleaning and organizing data: 57%*
- *Collecting data sets: 21%*
- *Mining data for patterns: 3%*
- *Refining algorithms: 4%*
- *Other: 5%*

G. Press, Cleaning Big Data: Most time-consuming, least enjoyable data science task, survey says, Forbes [Internet], 2016 Mar 23. Available at http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#3643728a7f75.



What data scientists spend the most time doing

- *Building training sets: 3%*
- *Cleaning and organizing data: 60%*
- *Collecting data sets; 19%*
- *Mining data for patterns: 9%*
- *Refining algorithms: 4%*
- *Other: 5%*

G. Press, Cleaning Big Data: Most time-consuming, least enjoyable data science task, survey says, Forbes [Internet], 2016 Mar 23. Available at http://www.forbes.com/sites/gilpress/2016/03/23/data-preparation-most-time-consuming-least-enjoyable-data-science-task-survey-says/#3643728a7f75.

Why should you care?

Making your data compliant with POLARIN VA will:

- accelerate research by providing open access to historically unavailable datasets, and by making cross-institutional and cross-discipline collaborations and meta-analyses more feasible
- allow your data to be used to benefit society by automatic aggregation into operational decision-support systems, weather/climate models, and aggregated data products
- reduce the data management burden on researchers using your RI, increasing their efficiency
- **improve impact of your RI by allowing for easy data citations (pro-actively preparing for envisioned ‘data index’ as alternative for ‘h-score’ index)**
- **make your RI more competitive for future funding rounds, as funders increasingly require open, FAIR-compliant data as a condition of grants**

By being part of POLARIN VA your RI is legally obliged/committed to offer your data in a manner that is compliant with POLARIN VA requirements. You risk losing the POLARIN VA funding if not fulfilling these requirements.

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

H2020 Programme

Guidelines to the Rules on

Open Access to Scientific Publications
and
Open Access to Research Data
in Horizon 2020

EUROPEAN COMMISSION
Directorate-General for Research & Innovation

Version 3.2
21 March 2017

H2020 Programme

Guidelines on

FAIR Data Management in Horizon 2020

Version 3.0
26 July 2016

Ecology and Evolution

Forward Series

ACADEMIC PRACTICE IN ECOLOGY AND EVOLUTION | Open Access

The data-index: An author-level metric that values impactful data and incentivizes data sharing

Amelia S. C. Hood, William J. Sutherland

First published: 13 October 2021 | https://doi.org/10.1002/ece3.8126 | VIEW METRICS

top: Hood, A. S., & Sutherland, W. J. (2021). The data-index: An author-level metric that values impactful data and incentivizes data sharing. Ecology and Evolution, 11(21), 14344-14350. Available at https://doi.org/10.1002/ece3.8126

left 1: European Commission. (2024). H2020 Guidelines to the Rules on Open Access to Scientific Publications and Open Access to Research Data in Horizon 2020 [PDF]. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-pilot-guide_en.pdf

left 2: European Commission. (2024). H2020 Guidelines on FAIR Data Management in Horizon 2020 [PDF]. https://ec.europa.eu/research/participants/data/ref/h2020/grants_manual/hi/oa_pilot/h2020-hi-oa-data-mgt_en.pdf

‘But AI will easily fix this for us!’

AI can not make up metadata. If there is ambiguity in the data description, AI (or a person) will never be able to reliably and correctly interpret the meaning of the data.

POLARIN

POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

Differences between POLARIN Data Hub and VA

Rachele Bordoni (ETT) on behalf of the POLARIN WP4 – Improvement of data services and customised data products

SIOS ETT UNIVERSITY OF OULU Consiglio Nazionale delle Ricerche inkode

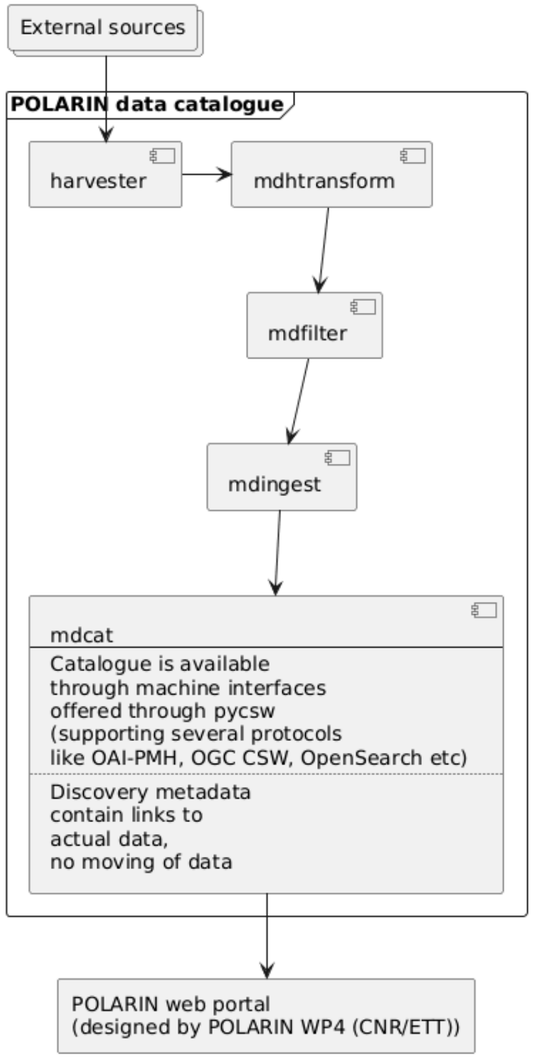
FUNDED BY THE EUROPEAN UNION

POLARIN
POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

How is POLARIN VA set up?

And you can offer your RI data in a POLARIN VA-compliant manner

The POLARIN metadata catalogue for VA



A **machine interoperable catalogue** that

... works as a broker: compatible with several mechanisms for exchanging discovery metadata

... harvests information from Ris providing VA without moving data

... transforms harvested information into an internal data model and enriches information in the process where possible

Minimum viable generic data model

POLARIN
POLAR RESEARCH INFRASTRUCTURE NETWORK

Element	Purpose
Metadata identifier	A unique identifier for the dataset. This is used to avoid duplicate records in aggregator catalogues. Utilisation of UUID with a namespace prefix is recommended.
Last update of metadata	An ISO8601 datetime for the last update of the metadata record.
Title	To provide a brief explanatory title for the dataset
Abstract	A short summary of the dataset, its purpose and how it was generated.
Temporal Extent	The temporal spanning of the dataset. This can be a multi temporal dataset.
Geographical Extent	The geographical location of the dataset. This can be a point, a bounding box, trajectory or a polygon.
Keywords	Keywords describing the dataset. Ideally this comes from controlled vocabularies and describes the variables of the dataset.
Personnel	Identification of all people that have contributed to the dataset. This requires full name, email, affiliation and a role description in the dataset. The latter has to come from a controlled vocabulary.
Publisher	This is identification of the data centre publishing the data. It contains a long and short name for the data centre and URL to the landing page.
Use constraint	This is a license for the data. This is a URL to the license text and an identifier. Utilisation of SPDX is recommended.
Data Access	This provides direct access to the dataset for download etc. It is not a landing page, but a direct link to the data and indication using a controlled vocabulary of the access mechanism (ranging from direct download to OPeNDAP and OGC WMS).
Project	A list of projects that has contributed to the creation of the dataset. Polarin has to be one of the projects.

Data model: a blueprint that defines how data should be structured and outlines relationships between data elements

Minimum viable: what should at least be communicated about datasets in the ***discovery metadata*** to be fully compatible with POLARIN VA

Generic: should work for ***all*** scientific domains covered in POLARIN

Mapping metadata elements to standards

POLARIN
POLAR RESEARCH INFRASTRUCTURE NETWORK

Metadata field (Table 1)	ISO-19115 (ISO-19139, ISO-19115)	DIF 10.2 (DIF-10.2)	Schema.org (Schema.org)	DCAT (DCAT-2.0)	ACDD (ACDD)
Metadata identifier	gmd:fileIdentifier	Entry_ID	identifier	dct:identifier	id
Last update of metadata	gmd:dateStamp	Metadata_Dates/Metadata_Last_Revision	dateModified	dct:modified	date_metadata_modified
Title	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:citation/gmd:CI_Citation/gmd:title	Entry_Title	name	dct:title	title
Abstract	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:abstract	Summary/Abstract	description	dct:description	summary
Temporal Extent	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:temporalElement/gmd:EX_TemporalExtent/gmd:extent/gml:TimePeriod/gml:beginPosition (gml:endPosition)	Temporal_Coverage/Range_DateTime/Beginning_Date_Time (Ending_Date_Time)	temporalCoverage	dct:temporal	time_coverage_start and time_coverage_end
Geographical Extent	gmd:identificationInfo/gmd:MD_DataIdentification/gmd:extent/gmd:EX_Extent/gmd:geographicElement/gmd:EX_GeographicBoundingBox/gmd:westBoundLongitude (gmd:eastBoundLongitude, gmd:southBoundLatitude, gmd:northBoundLatitude)	Spatial_Coverage/Geometry/Bounding_Rectangle/Westernmost_Longitude (Northernmost_Latitude, Southernmost_Latitude, Easternmost_Longitude)	spatialCoverage	dct:spatial	geospatial_lat_min,geospatial_lat_max, geospatial_lon_min,geospatial_lon_max

The Architecture Document on GitHub has mappings for relevant metadata standards in use:

https://github.com/polarin-he/va_add/blob/main/doc/polarin_add.pdf

Critical elements for discovery metadata

If this is missing the information harvested will not be ingested!

All datasets **need** to be described with
… **unique identifier** (DOI or a UUID with namespace)
… **title**
… **abstract**
… **spatiotemporal location** (time, lat, lon, z, sometimes even more dimensions)
… **link to data access**
… **project** (in this case ‘POLARIN’)

That way we can differentiate between unique datasets and avoid storing duplicate datasets in the database

However, including the other elements is highly beneficial to provide proper credit to PIs and data repositories serving data!

How to make your database compatible with POLARIN VA

POLARIN
POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

Set up a machine-readable interface that serves your data in a standardized manner, more commonly known as an **Application Programming Interface (API)**!

Do implement:

- Widely-adopted, standardised information (metadata) exchange protocols. POLARIN currently supports: **OAI-PMH**, **OGC CSW**, **OpenSearch**, and **schema.org**.
- Widely-adopted information (metadata) documentation standards. POLARIN currently supports: **GCMD DIF** (multiple flavours), **ISO19115** (multiple flavours), **DCAT**, **schema.org**, **ACDD, CF**
- Widely-adopted, proprietary, standardized file formats: **CF-NetCDF**, **Darwin Core Archives**
- Widely-adopted, open-source, and standards-compliant data server solutions: **THREDDS** or **Hyrax** (OPeNDAP compatible)
- Sitemaps (increased efficiency)
- High dataset granularity (avoid single, massive, *monolithic* files)

Do not implement:

- Proprietary, non-standardised APIs
- Proprietary, non-standardized, domain-specific file formats
- Lazy-loading of webpages (introduces unpredicatability)

There are ongoing efforts to accommodate more protocols!

POLARIN
POLAR
RESEARCH
INFRASTRUCTURE
NETWORK

Take-home message

You don’t need to overhaul everything at once. Focus on one dataset, apply 2-3 easy wins, and scale from there. The goal is to reduce friction for users while making data self-serviceable and interoperable!

Contact Us

Questions related to Virtual Access

SIOS (polarin@sios-svalbard.org)

Questions related to the POLARIN Data Hub

Rachele Bordoni rachelebordoni@dedagroup.it)

Antonio Novellino (antonio.novellino@dedagroup.it

POLARIN

POLAR
RESEARCH
INFRASTRUCTURE
NETWORK