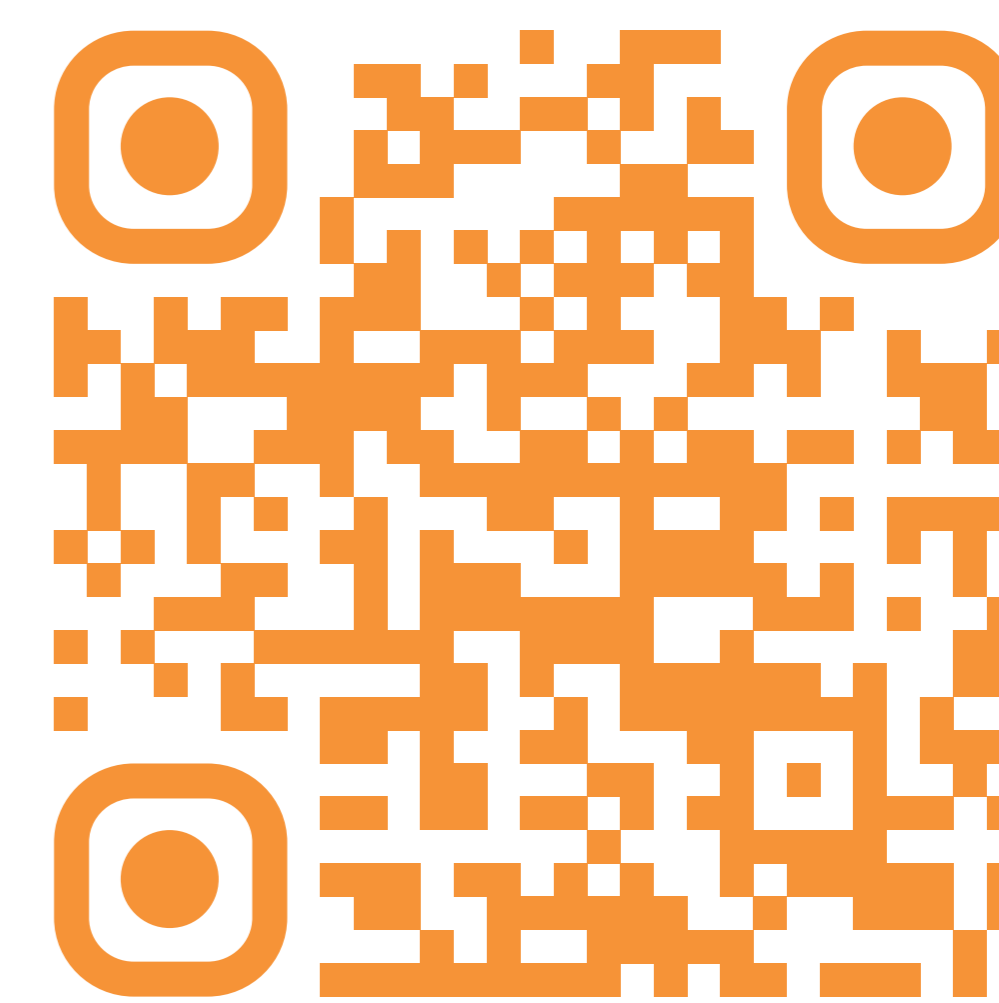


# Enabling access to ECV-related observation datasets from environmental Research Infrastructures

<sup>1</sup>Tjerk Krijger, <sup>2</sup>Peter Thijsse, <sup>3</sup>Marta Gutierrez, <sup>4</sup>Delphine Dobler, <sup>5</sup>Ulrich Bundke, <sup>6</sup>Alexandra Kokkinaki, <sup>7</sup>Gwenaëlle Moncoiffé, <sup>8</sup>MARIS, <sup>9</sup>EGL Foundation, <sup>10</sup>Euro-ARGO ERIC, <sup>11</sup>IAGOS/Forschungszentrum Jülich, <sup>12</sup>NOC-BODC



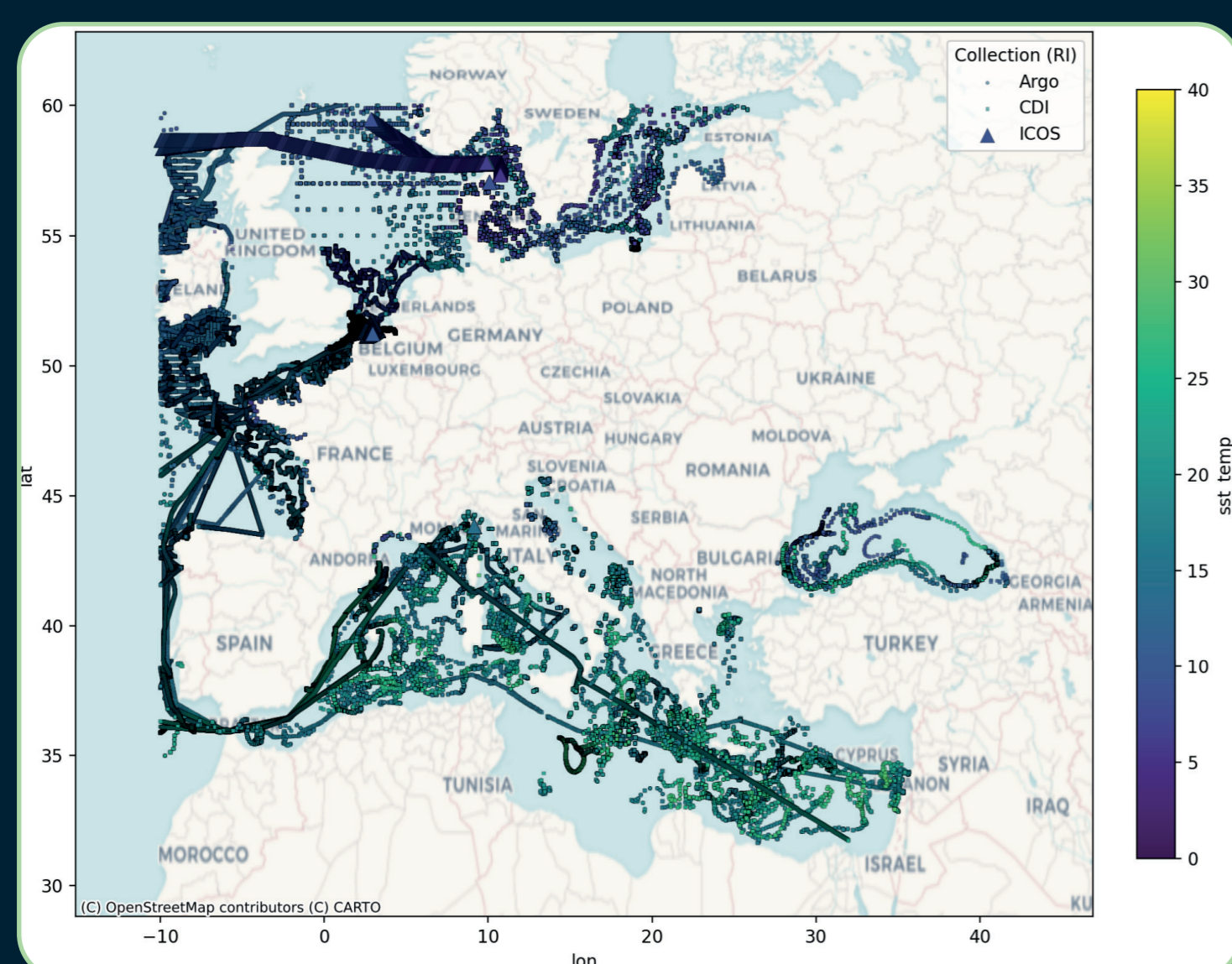
Test the ENVRI-Hub  
<https://envri-hub.envri.eu/>

## Abstract

ENVRI-Hub NEXT is improving access to Essential Climate Variable (ECV) data across environmental Research Infrastructures (RI). To support this, an ECV Working Group has been established to align the diverse vocabularies used by different RIs corresponding to selected ECVs. The NERC Vocabulary Server (NVS) and I-ADOPT were used to translate ECV requests into RI-specific observed parameters. Ongoing work addresses the heterogeneous data access services in use by the RIs by developing harmonised subsetting solutions delivering, query-specific single datasets-outputs, directly usable in Virtual Research Environments.

## Challenges

- **Heterogeneous machine-to-machine** data access services & lack of subsetting capabilities.
- **Inconsistent metadata** and vocabularies across RIs.
- **Limited alignment** between ECV concepts and RI-specific observed parameters.



## Our Proposed System

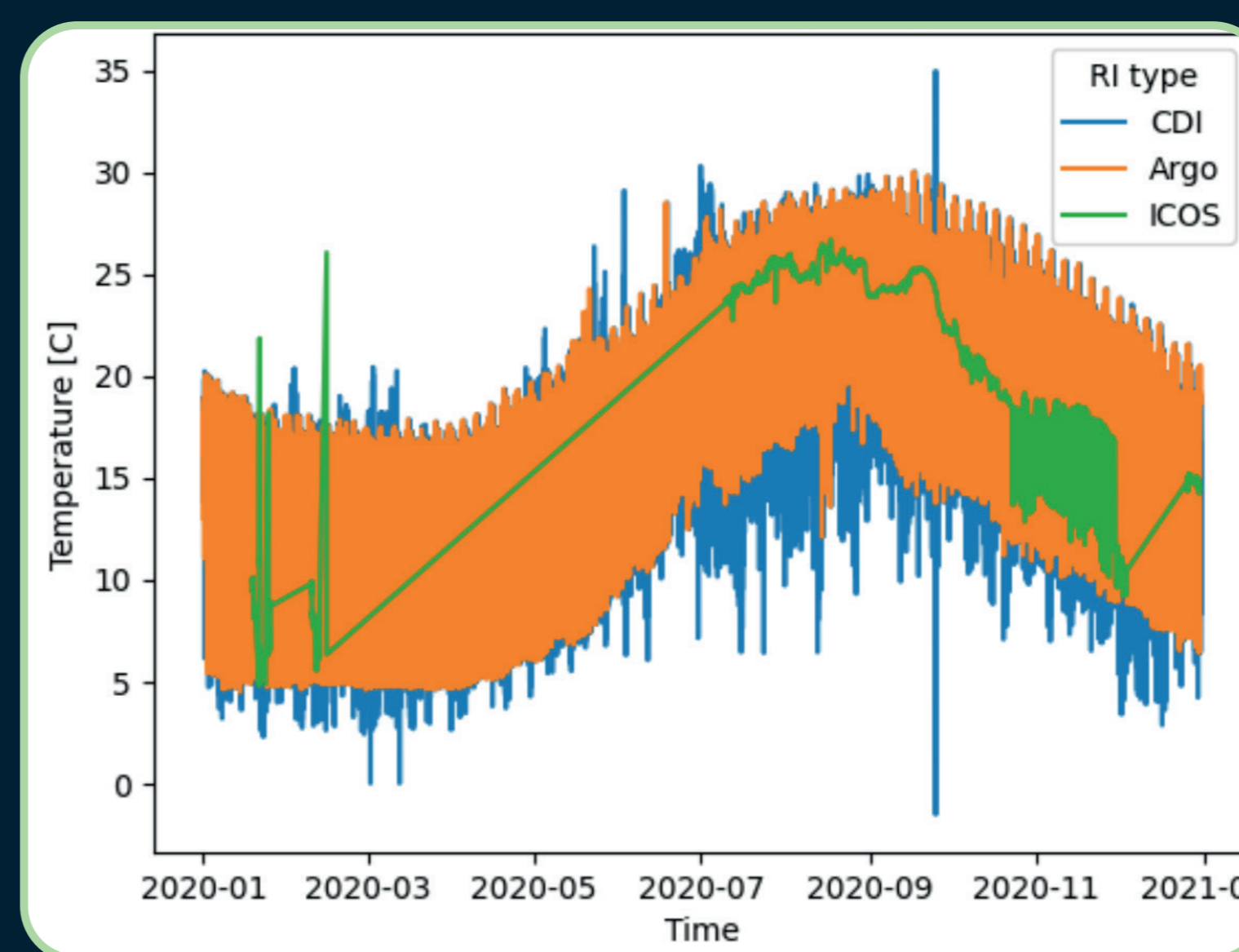
The Beacon system is developed as data lake software for subsetting large in situ data collections, used on the RI collections, and integrates the vocabulary mappings using I-ADOPT to dynamically map ECV concepts to RI-specific observed parameters.

This enables specific queries, such as "give me all data related to ECV-oxygen in this region, for this time period and this depth range", to **return one single file that matches exactly** this content from each RI.

## Next Steps

The current work demonstrates the workflows from ECV concepts to RI related observation datasets. Next steps will include integration into the ENVRI-Hub NEXT core components, including the Catalogue of Services.

Currently, we are focusing on a sub-set of ECV-related data from **IAGOS, Euro-Argo, ICOS, SeaDataNet and ACTRIS**.



## System Demonstration

The user selects which ECV they are interested in from the collection included in the NVS (<https://vocab.nerc.ac.uk/collection/EXV/current/>), along with filters on region, time and depth.

```
exv_variable = "EXV017"           # sea surface temperature
region = (-10, 45, 30, 60)        # lon_min, lon_max, lat_min, lat_max
time = ("2020-01-01", "2020-12-31") # date_min, date_max
depth = (0, 10)                   # depth below the sea level
```

The system returns one single file (csv/netcdf/parquet/..), in this case a Pandas dataframe, which contains only those measurements which exactly match the query filter.

The **RI data lake manager** decides the columns that are returned and what other filters are applied, described in pre-configured query tables.

For example, an RI can choose to include only those data that match certain quality flags. In this example for SeaDataNet, the EDMO code and local CDI id are returned, to be able to trace each individual measurement back to its original dataset in the CDI.

	LONGITUDE	LATITUDE	TIME	DEPTH	SDN_STATION	SDN_EDMO_CODE	SDN_LOCAL_CDI_ID	EXV017	RI
0	18.18594	42.08870	2020-12-28 01:14:00	2.820	6903250_176_D	486	ARGO_6903250_176_D	15.799	CDI
1	18.18594	42.08870	2020-12-28 01:14:00	2.920	6903250_176_D	486	ARGO_6903250_176_D	15.799	CDI
2	18.18594	42.08870	2020-12-28 01:14:00	3.020	6903250_176_D	486	ARGO_6903250_176_D	15.799	CDI
3	18.18594	42.08870	2020-12-28 01:14:00	3.140	6903250_176_D	486	ARGO_6903250_176_D	15.799	CDI
4	18.18594	42.08870	2020-12-28 01:14:00	3.230	6903250_176_D	486	ARGO_6903250_176_D	15.799	CDI
...	...	...	...	...	...	...	...	...	...
1459306	7.78618	43.40908	2020-09-29 08:38:08	4.110	7900563_9_A	486	ARGO_7900563_9_A	19.884	CDI
1459307	7.78618	43.40908	2020-09-29 08:38:08	4.445	7900563_9_A	486	ARGO_7900563_9_A	19.884	CDI
1459308	7.78618	43.40908	2020-09-29 08:38:08	6.130	7900563_9_A	486	ARGO_7900563_9_A	19.882	CDI
1459309	7.78618	43.40908	2020-09-29 08:38:08	7.480	7900563_9_A	486	ARGO_7900563_9_A	19.882	CDI
1459310	7.78618	43.40908	2020-09-29 08:38:08	8.160	7900563_9_A	486	ARGO_7900563_9_A	19.881	CDI

1459311 rows x 9 columns