EUROPEAN OPEN SCIENCE CLOUD

Implementing an inclusive European Open Science Cloud

EOSC SYMPOSIUM 2021

A Common Approach to Data Ecosystems and Data Spaces: Implications for EOSC Architecture

M. Dietrich, S. Fiore, F. Pando, S. Mieruch

mark.dietrich@egi.eu; sandro.fiore@unitn.it; pando@gbif.es; sebastian.mieruch@awi.de

17/06/2021

Swirling Terms and Technologies

EOSC SYMPOSIUM 2021

Can we make sense of them all?

data sharing space

Data Sharing Domain

INTERNATIONAL DATA SPACES ASSOCIATION

IHAN

EUROPEAN OPEN SCIENCE CLOUD

Open Data Ecosystem

X-ROAD

gaia-x

DAWEX

FIWARE

Data Exchange Approach (DEA)

A Unifying Concept

Data Exchange Approach: defined as

- a set of (1) organizational policies and roles and
- (2) technical specifications,
- required to enable the trustworthy exchange of data between two parties.

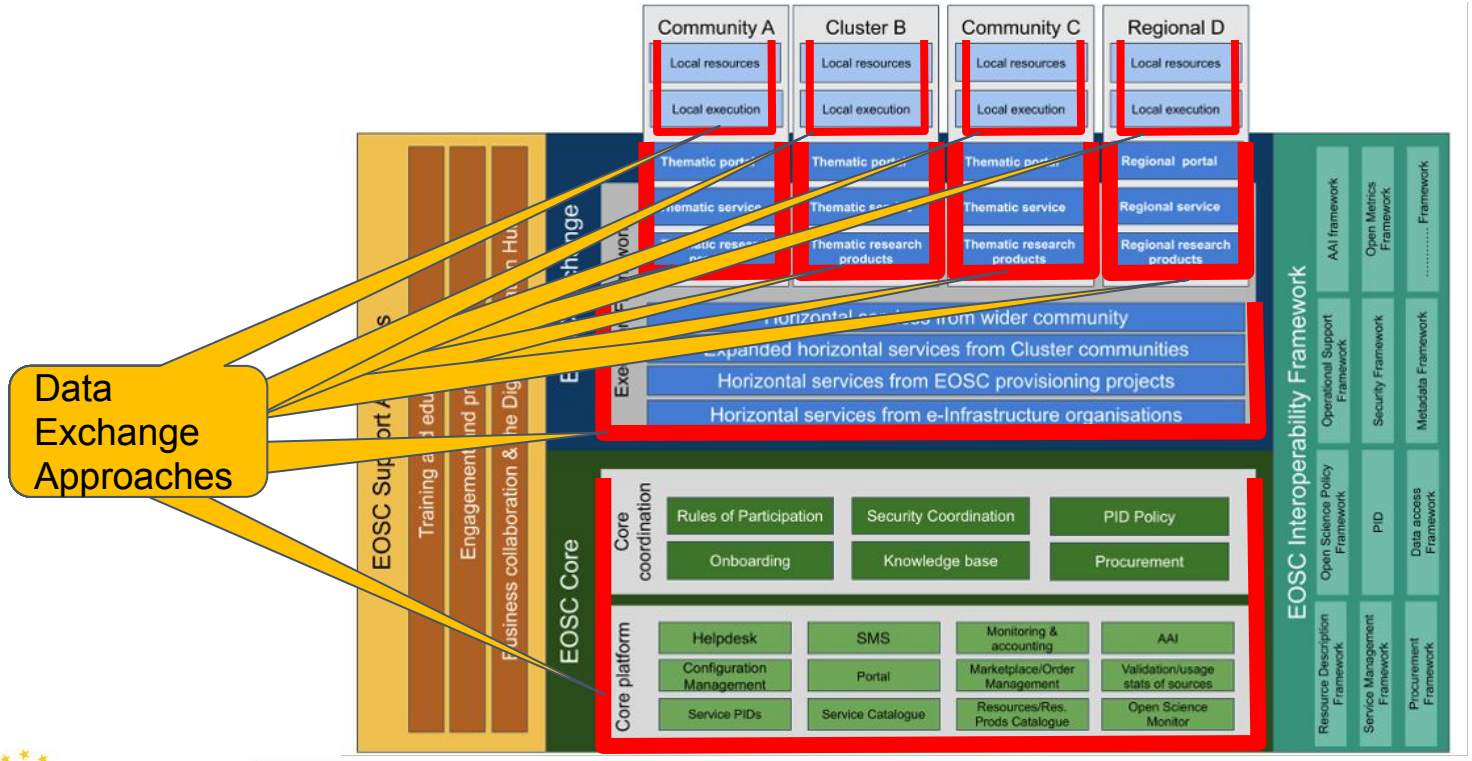
Data Ecosystems and Data Spaces are types of DEA.

DEAs can flexibly describe different characteristics of exchange structures.

- E.g. the **visibility** of data, which can be limited or controlled:
  - exposure to only one other party
  - combined exposure of metadata and data to limited groups
  - public exposure of both data and metadata

How does this concept map to EOSC?

We Need To Respect Distinct Data Exchange Approaches



Data Ecosystems, Data Space are types of DEA

EOSC SYMPOSIUM 2021

Data Ecosystem: *a purposeful collaboration or partnership consuming, producing and providing interoperable data and related resources*.

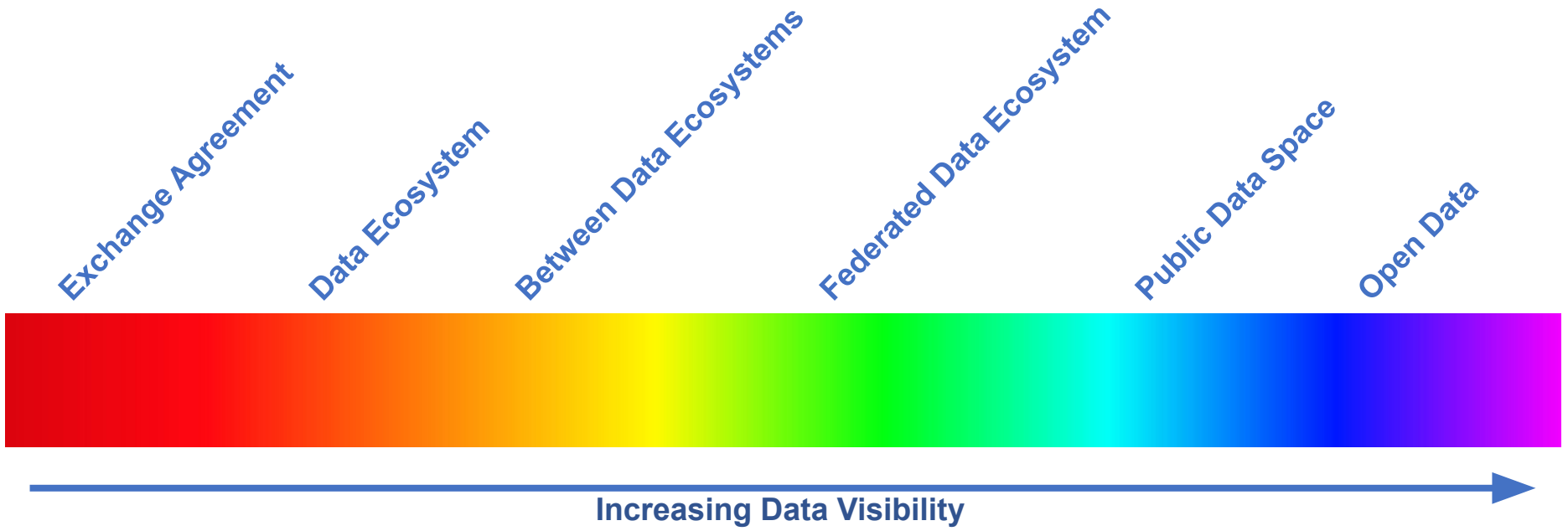
- Multiple data ecosystems possible -- they already exist:
  - E.g. in agriculture: Djust Connect, API-Agro, DKE Agrirouter, JoinData, Agrimetrics, Aladin.farm, DataConnect
  - "Science communities" in different areas of research

Data Space: *a collection of FAIR*, quality data and related resources consumed, produced and provided by identified participants, each respecting societal values and operating within an explicit framework of trust and governance*.

- EC refers to **Public** Data Space
  - Does not identify any collaborating actors or a community
  - Is not "purposeful"
- EC aspires to a "single common European data space"
- Domain Data Spaces, collecting data and data-services relevant to a domain or sector.

Data Ecosystem = Private Data Space + Community + Purpose

Spectrum of DEAs



Research → Already Provides Public Data Spaces

many examples in both research and industry

IS-ENES / ESGF

GBIF

SeaDataNet

EUROPEAN OPEN SCIENCE CLOUD

Applying DEAs to 3 Use Cases: GBIF, IS-ENES, SeaDataNet

EOSCsecretariat.eu has received funding from the European Union's Horizon Programme call H2020-INFRAEOSC-05-2018-2019, Grant Agreement number 831644.

Global Biodiversity Information Facility

GBIF

EOSC SYMPOSIUM 2021

Free and open access to biodiversity data: numbers as of 31/03/2021

Species occurrence records

1,667,617,812

Datasets

57,432

Country Participants **61**

Organizational Participants **39**

Data-publishing institutions

1,659

Peer-review papers using data

5,658

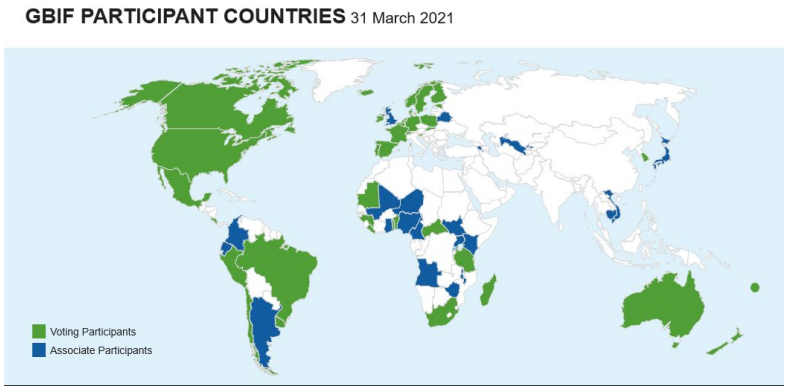
Avg records downloaded per month (2021)

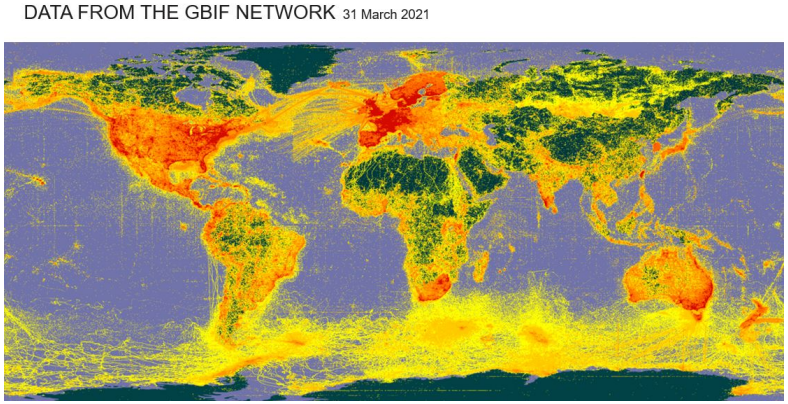
85.6 billion

FAIR, Free and Open Access to Biodiversity Data

Data: biological species occurrences, species check-lists, sample based datasets

- Findable?
  - Indexed: data & metadata
  - Standardized: data & metadata
- Interoperable?
  - Inside network: full
  - Outside network: limited
  - SPARQL endpoint: no
- Accessible?
  - Cached: yes
  - Downloadable: yes
  - Visualizations: yes
- Reusable?
  - Downloadable: yes
  - Licenses schema: yes
  - DOIs for datasets & downloads
  - UUIDs for records: uneven implementation

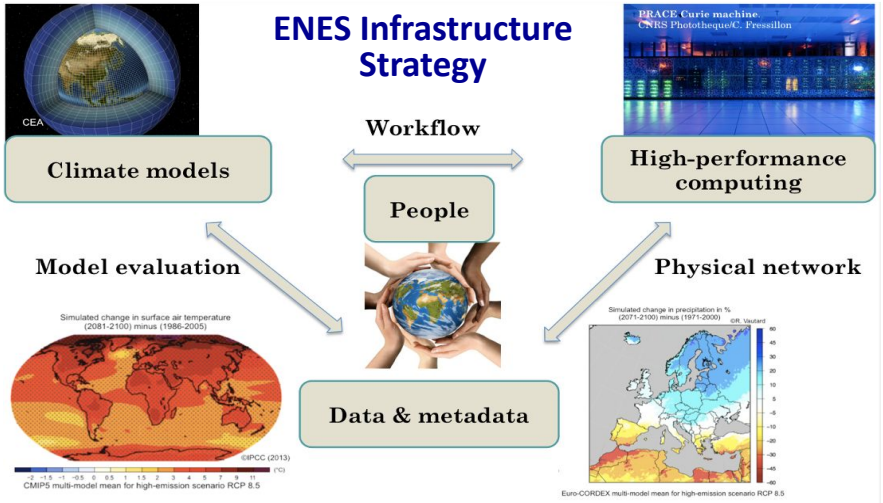




Federated data infrastructure for climate model data

is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

EOSC SYMPOSIUM 2021



Services

Data access and publication
Citation & PID, Statistics
Long-term archival and replica
Climate4impact portal
Evaluation and diagnostics
Compute services
Service on data & metadata standards, CF convention & Data Request, Model documentation

Download Statistics

ESGF: 8 M datasets
23,4 PB (w/o replica 12,7)
CMIP6: 7 M datasets
16,1 PB (w/o replica 9,3)
CMIP5: 5,3 PB (1,5)
ca 15 000 registered users

Earth System Grid Federation ESGF

is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

EGI-ACE

EOSC-hub

EUROPEAN OPEN SCIENCE CLOUD

Climate Change Service

opernicus Europe's eyes on Earth

Easy access to climate model (open) data for analysis needs

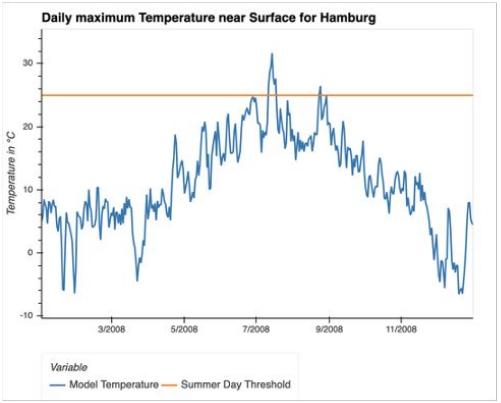
is-enes INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

EOSC SYMPOSIUM 2021

ESGF Data download and publication metrics

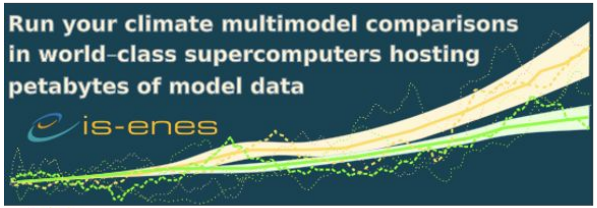
31	162	24	11,929,956	560,381,850	22,76
Projects	Countries	Data nodes	published datasets	files downloaded	PB downloaded

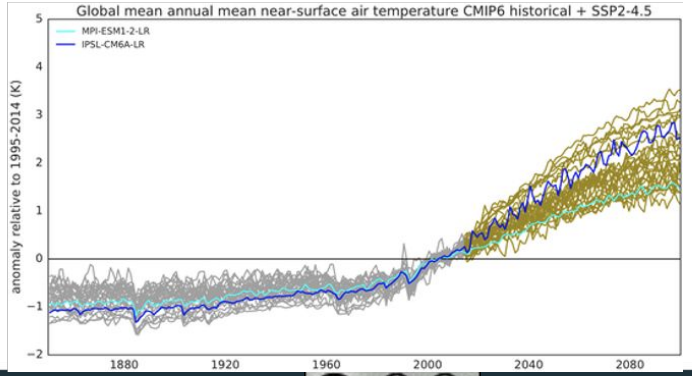
Climate data analysis use cases



IS-ENES Compute services

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

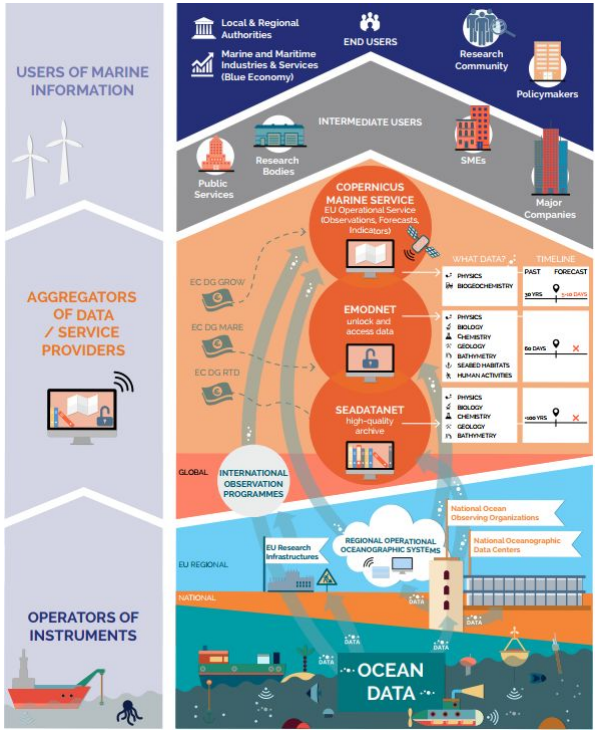




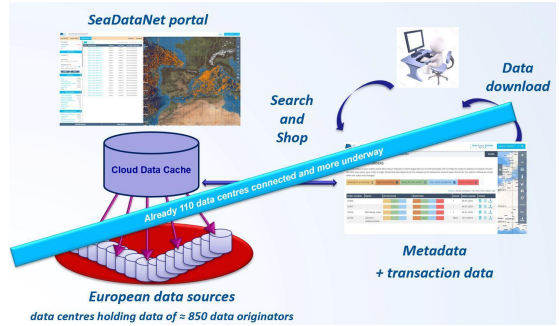
Major European Aggregator/Manager of Marine Data SeaDataNet

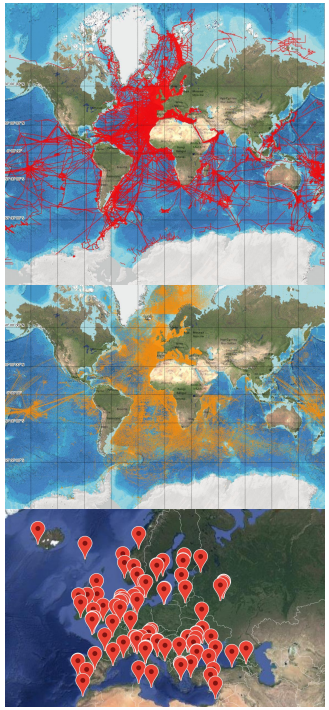
EOSC SYMPOSIUM 2021

SeaDataNet - pan-European network of NODCs



- Network of marine data centres for curation and long term archiving
- Developing and promoting standards, tools & services for ocean and marine data management
- Controlled vocabularies and European directories
- Giving access to data sets from 110 data centres and more than 850 data originators



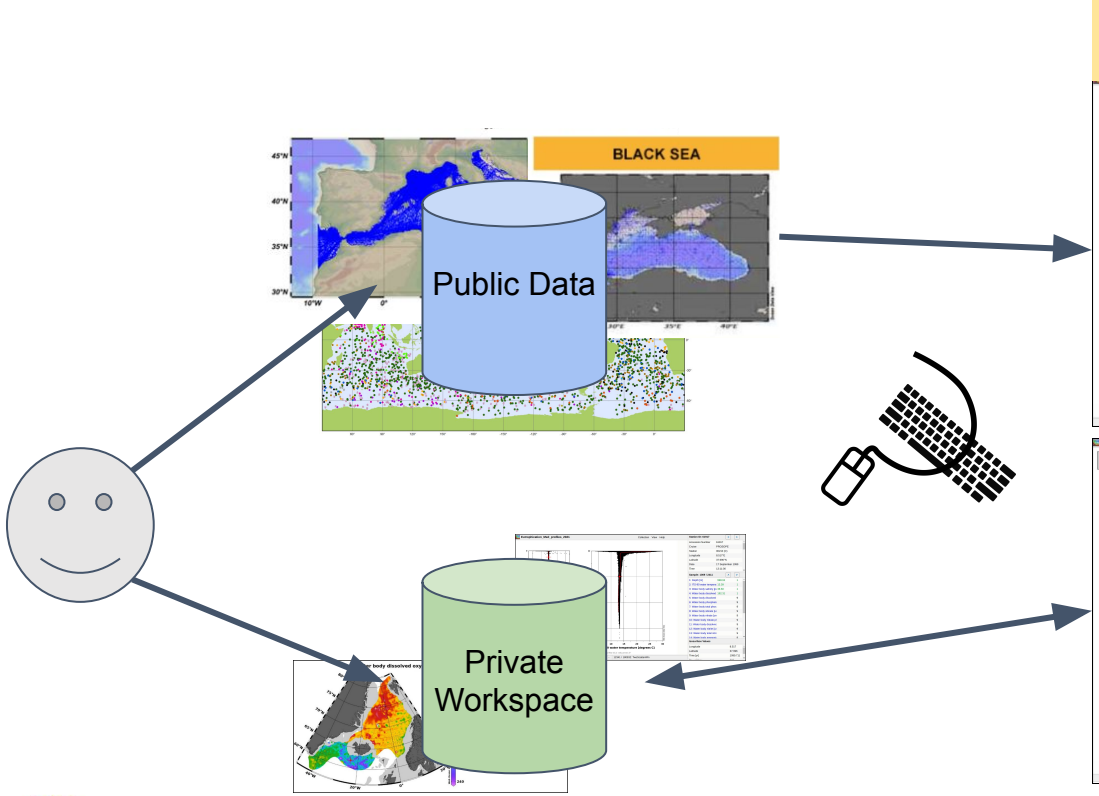


webODV: Web Ocean Data View

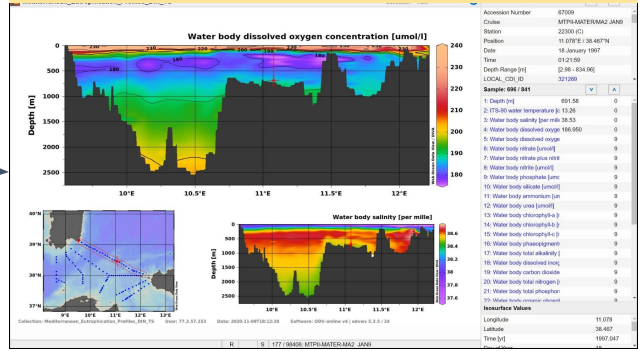
SeaDataNet

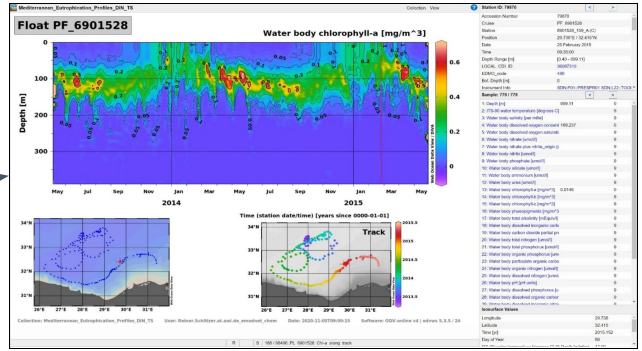
EOSC SYMPOSIUM 2021

Interactive online analysis of marine data from 110 data centers and 34 countries



Explore, Extract, Analyze, Visualize





EUROPEAN OPEN SCIENCE CLOUD

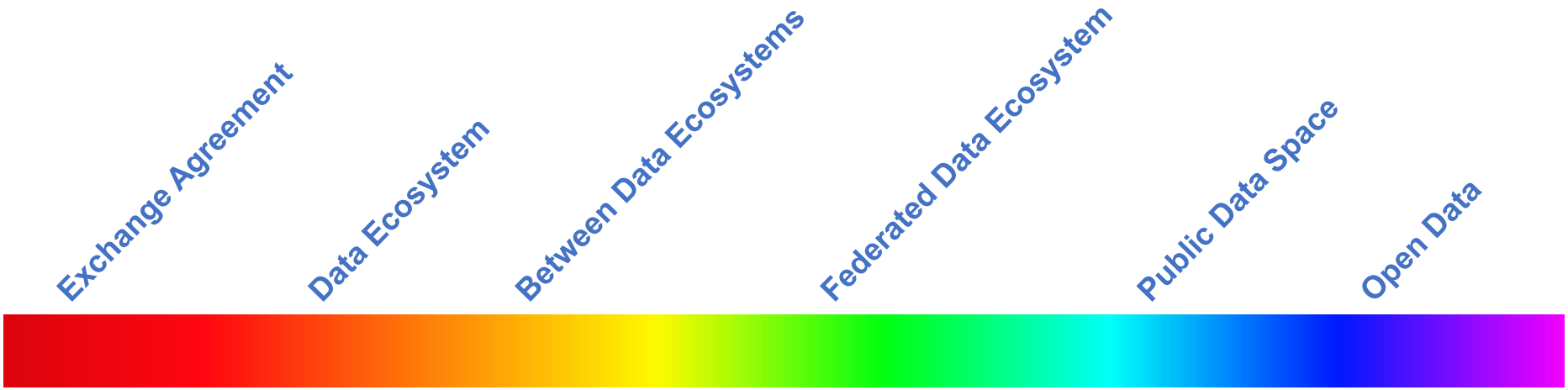
H-CLOUD

EGI-ACE

is-enes
INFRASTRUCTURE FOR THE EUROPEAN NETWORK FOR EARTH SYSTEM MODELLING

This work was supported by the European Commission H2020 grant number 101017567 EGI-ACE (https://www.egi.eu/projects/egi-ace/), the H2020 grant number 871920 H-CLOUD (https://www.h-cloud.eu/), the H2020 grant number 824084 IS-ENES3 (https://is.enes.org/)

Specific DEAs along the Spectrum



IHAN

X-ROAD

many examples in both research and industry

EUROPEAN OPEN SCIENCE CLOUD

gaia-x

European Commission

data.europa.eu