Work Package 5 - DATA INTEGRATION AND MANAGEMENT

WP Achievements

INTAROS M18 Review Meeting

Covent Garden Building,
Brussels, Belgium
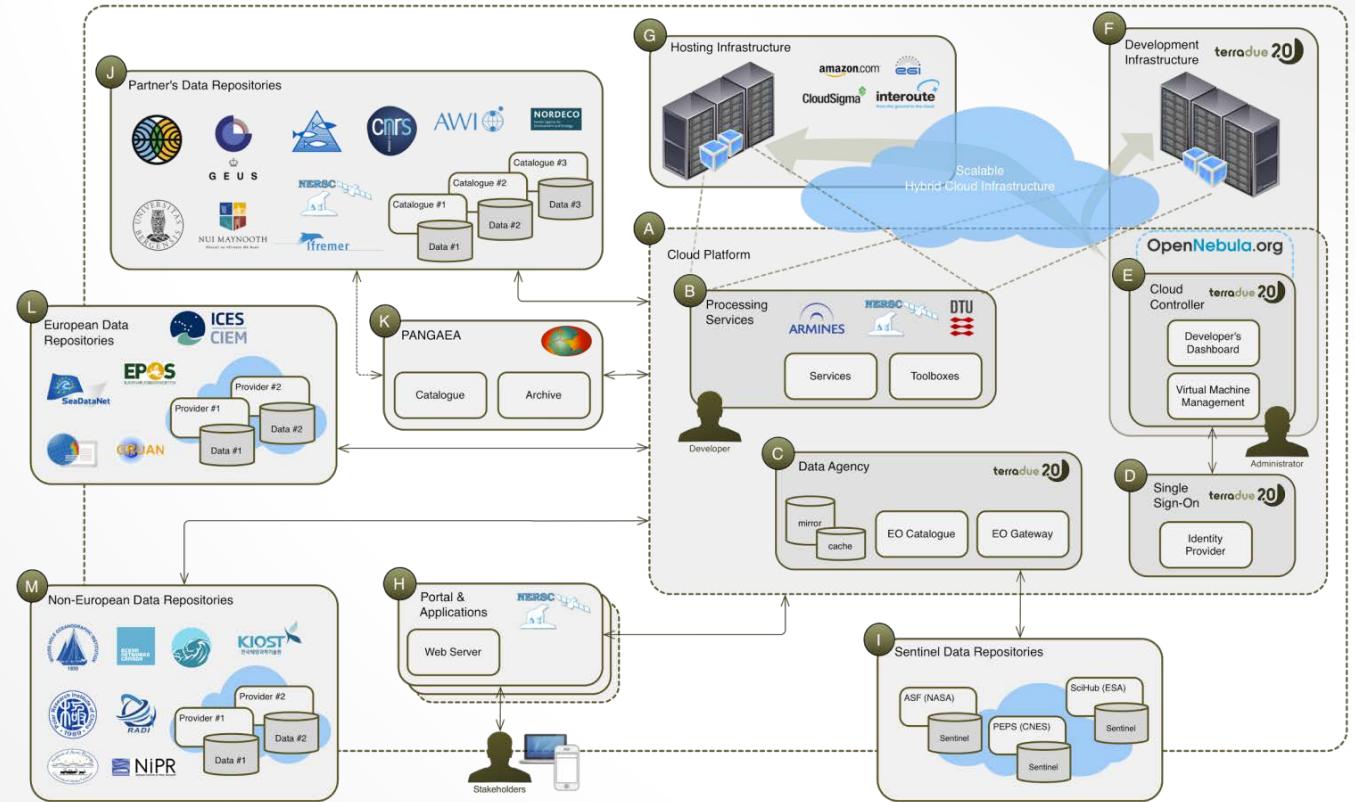
21st September 2018

Herve Caumont, Terradue Srl (lead)
Torill Hamre, NERSC (co-lead)

INTAROS

Agenda

- Objectives
- Overall achievements
- Tasks status report

WP5 DATA INTEGRATION & MANAGEMENT
OBJECTIVES

- **Integrate data repositories** (*multidisciplinary* and *distributed*) into a scalable and resilient integrated Arctic observing system (iAOS)

  - Connect to observations & derived parameters together with EO data services

- **Develop processing services** for sea ice statistics, for integrated acoustics-remote sensing data analysis, and other geostatistics

  - Integrate a set of tools for data analysis, transformation and visualization.
  - Support geostatistical methods for interpolation of spatiotemporal datasets.

- **Support processing campaigns** of new observations from WP2-4

  - Enable users to run processing "within iAOS" (using iAOS-funded Cloud resources)
  - Store generated datasets in an iAOS-enabled repository

INTAROS

WP5 DATA INTEGRATION & MANAGEMENT
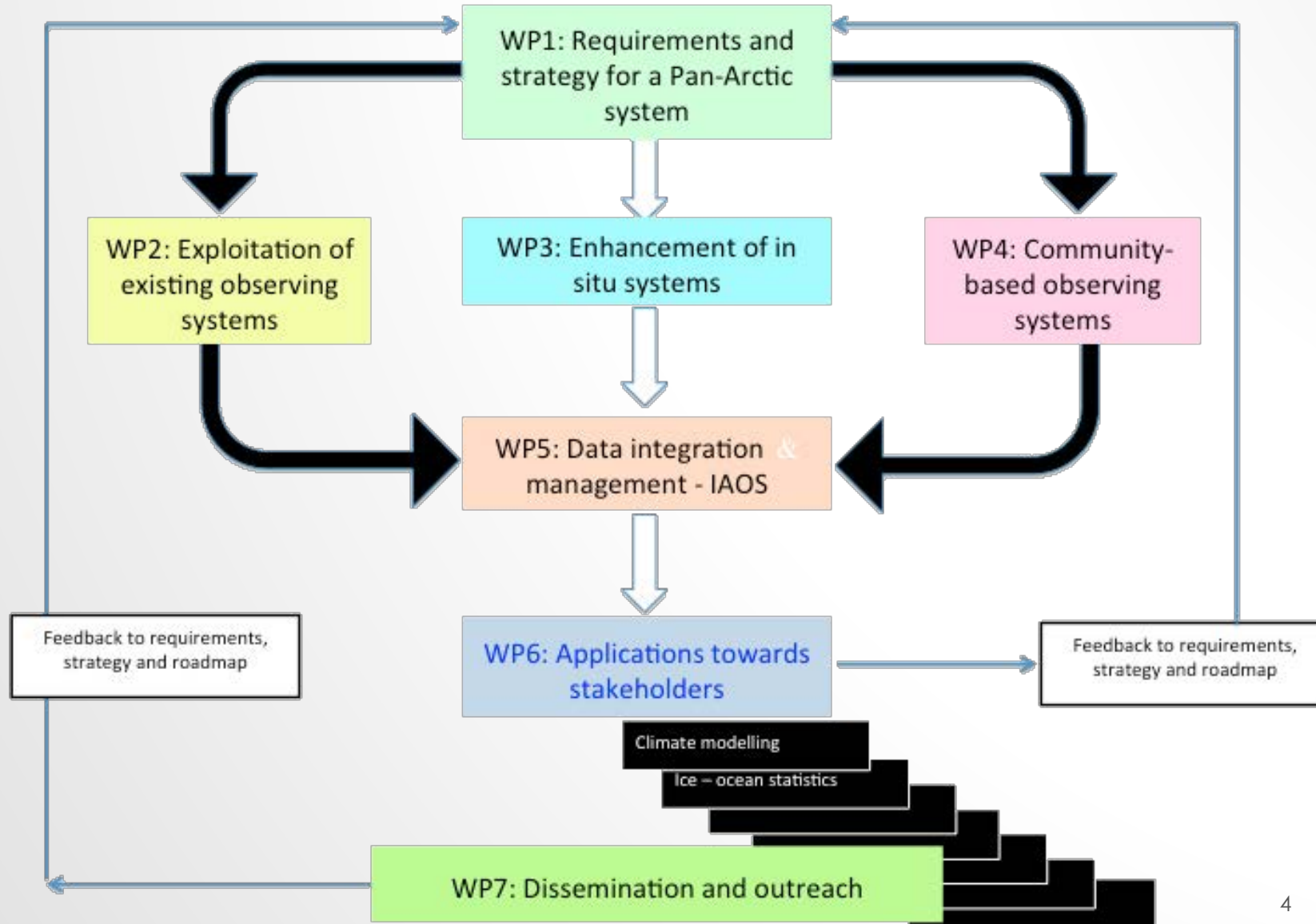
WP5 DATA INTEGRATION & MANAGEMENT ACHIEVEMENTS

- Monthly meetings with WP5 partners (actions and status review)

- Collaboration with WP2 for the Classification Parameters Document

- Outreach preparations for iAOS processing platform tools & services

- Design for the RGeostats toolbox integration as an iAOS Processing Service, and initial proof of concept (Sandbox service)

- Initial contacts for defining the iAOS Portal User Stories

- 1st release of the INTAROS Requirements and Architecture Design

- Prepared deliverables templates for D5.2, D5.3, D5.4, D5.5 and D5.6 in shared documents (Google Docs)

TASK 5.1 - SYSTEM REQUIREMENTS AND ARCHITECTURE CONSOLIDATION

Partners : Terradue, NERSC, AWI

- Analysis of the system requirements and architecture for the integration of multidisciplinary and distributed data repositories

  - Focused on data processing platform (T5.2), data discovery & access (T5.3), data analyses algorithms & toolkits (T5.4, T5.5) and user portal (T5.6)

- The first version of requirements and architecture was documented in the deliverable **D5.1 "IAOS requirements and architectural design"**, which was submitted in November 2017.

- Overall, task activities are progressing according to the schedule.

iAOS Platform Operational scenarios

- Integration of Data Access facilities (Data)

- Design and integration of scalable processing applications (Cloud Platform)

- Management of a Platform's resources for hosted data processing service (Cloud Platform)

- Exploitation of data access services (Portal)

- Exploitation of data processing services (Portal)

- Administration of Cloud Platform resources

# TASK 5.2 - IAOS PLATFORM DEPLOYMENT AND OPERATION

Partners: Terradue

- Provided Cloud Platform services and support to partners ARMINES and NERSC in their respective tasks (T5.4, T5.5)

  - Setup activities for user on-boarding (provision of VM and support)

- Defined Table of contents for **D5.2 "iAOS Platform and tools"** (due in M24): It will introduce Platform tools and services available for the integration of processing chains

- Overall, task activities are progressing according to the schedule

Integration of new Processing Services

- Exploit the data processing tools and geostatistical algorithms as Cloud processing services

- Support the full lifecycle of the integration of new processing services, offering simultaneous access to data, tools and Cloud resources

- Maintain and operate the supporting Platform-as-a-Service (PaaS) environment for the iAOS services implementation

- Demonstrate the iAOS capabilities through integration and deployment of selected data processing services and user Portal

Service Design

Service Implementation

Service Testing

Service Deployment

Service Validation

INTAROS
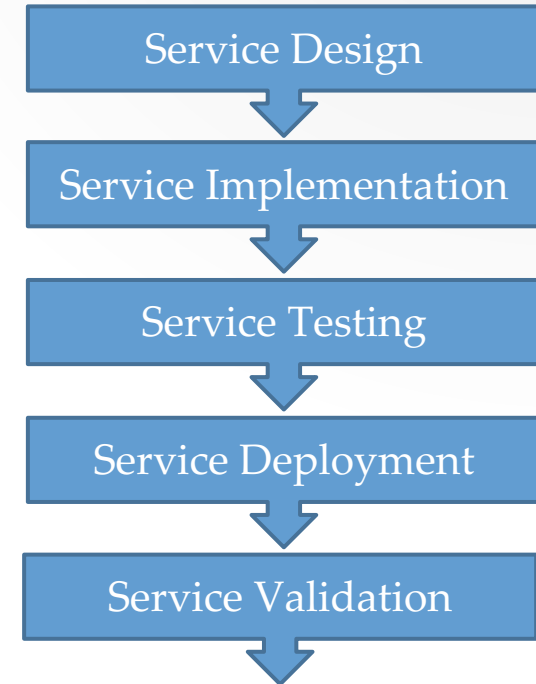
TASK 5.3 - INTEGRATE DATA FROM EXISTING REPOSITORIES INTO IAOS

Partners : AWI, Terradue, NERSC, IMR, AWI, DTU, GEUS, FMI, NUIM

- Established a solid link with WP2 from a very early stage.

- Defined initial task 5.3 targets:
  - Selection of suitable **show cases** for a first integration of datasets
  - Assess existing Arctic Observing Systems (**link to outcomes of WP2 task 2.1**)
  - Compile data products from distributed databases and observatories – linked from WP2 task 2.3 for data integration from existing repositories into iAOS

- Overall, task activities are progressing according to the schedule

INTAROS

Metadata Catalogue tool

- Catalogue records collected from WP2 surveys
  - Communities info + in-situ data + EO data
  - Input from three spheres (deliverables) into one metadata catalogue from WP2
- WP2/WP5 collaboration for their analysis
  - Must support machine-to-machine interface
  - Need additional (lots of!) information on how to access the data (AWI engaged with data managers), incl. updates
- Selected CKAN as iAOS Metadata Catalogue tool
  - Goal to feed the iAOS portal / website

INTAROS

Sea Ice Remote Sensing data

- Connect to iAOS the Sea Ice Remote Sensing data from University of Bremen
  - Analysed the current online repository (accessible via FTP and HTTP)
  - Analysed the products and metadata, to create collections foreseen as most attractive to share with a greater community
    - seaice thickness
    - ice concentration
  - Selected OpenDAP server solution
    - to be hosted on iAOS for initial experiment
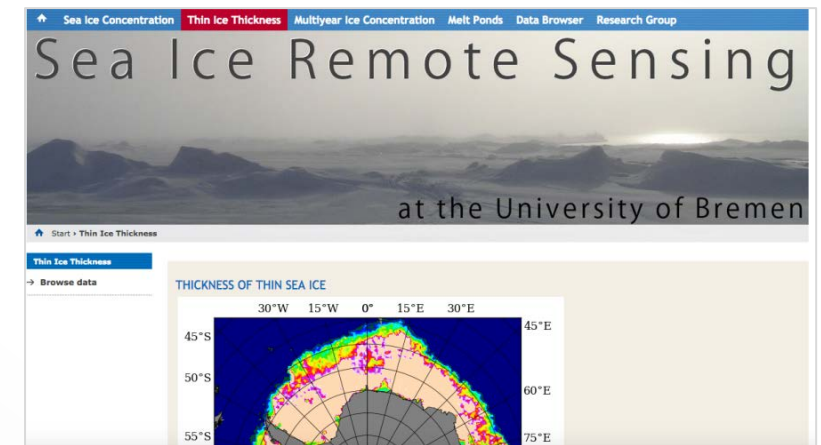
INTAROS

TASK 5.4 - DEVELOPMENT OF GEOSTATISTICAL METHODS FOR DATA INTEGRATION

Partners : ARMINES, NERSC, DTU

- Installation and deployment of the RGeostats package on the Cloud platform, now available to the INTAROS community.

- Development of a first application example with data relevant to the Arctic research community.

  - Dissemination material prepared to outreach the iAOS users community

  - Held workshop in Paris to prepare the January 2019 trainings (GA in Bremen)

- Overall, task activities are progressing according to the schedule.
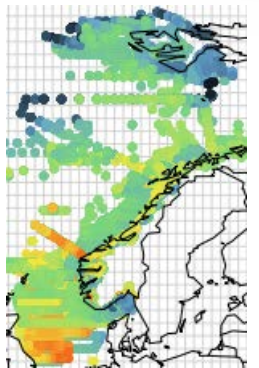
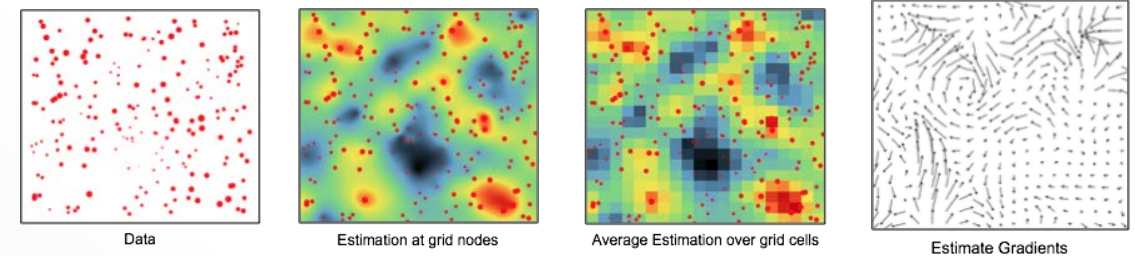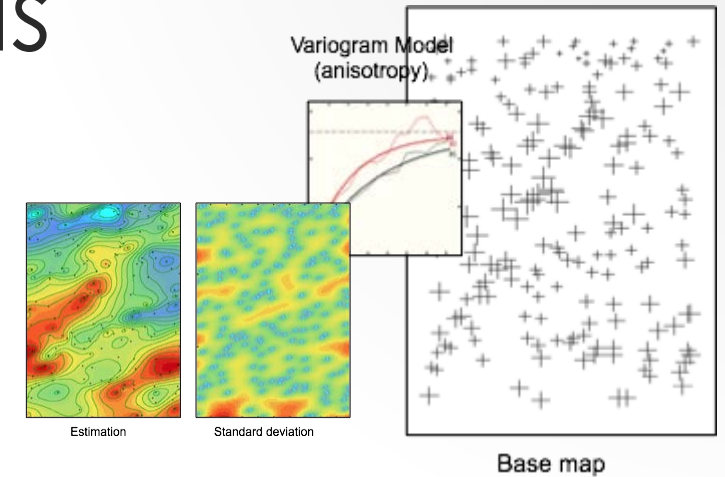# RGeostats Package Overview
Toolbox Capabilities

- The most complete free software for geostatistics
  - Package under R platform

- Main key features:
  - Data of any space dimension (space and time)
  - With any number of variables treated simultaneously
  - Possible extension for spatio-temporal models
  - Big number of data/targets (up to memory capacity for R)
  - Stochastic partial differential equation models (SPDE)

- Data organization:
  - Set of isolated points, Regular grids
  - Data (points) on profiles
  - Coordinates projections and spherical coordinates

Unleashing the Potential of Geostatistics for Data Analysis

- Mapping: Kriging provides optimal linear unbiased estimation

- With several types of estimations:
  - Punctual at grid nodes
  - Average over grid cells
  - Any linear quantity: gradients, convolution, …

- Kriging with nugget effect

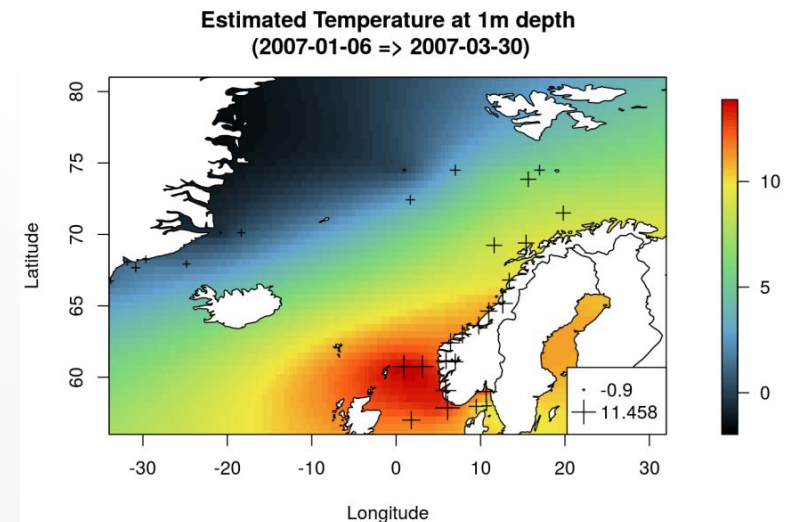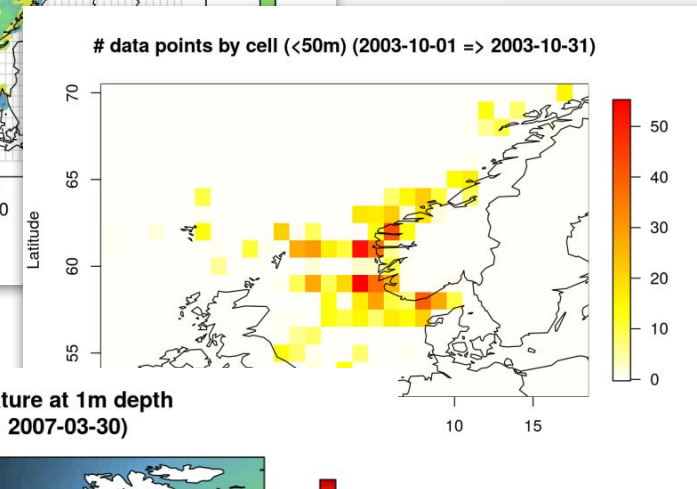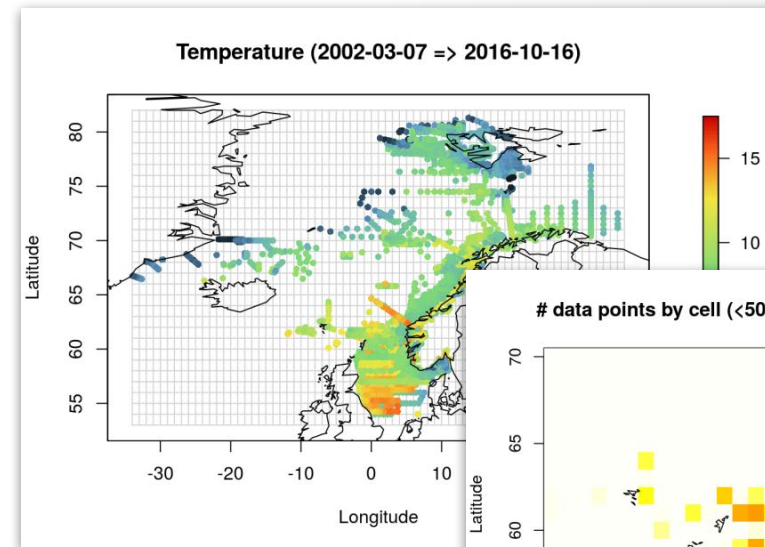- New methods to be developed or adapted for Arctics datasets specificites

Variogram Model (anisotropy)

Estimation   Standard deviation   Base map

Data   Estimation at grid nodes   Average Estimation over grid cells   Estimate Gradients

Kriging   Filtering Nugget Effect

INTAROS
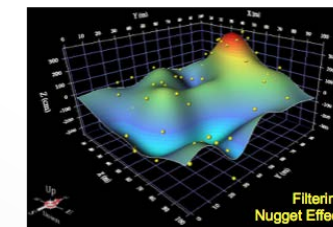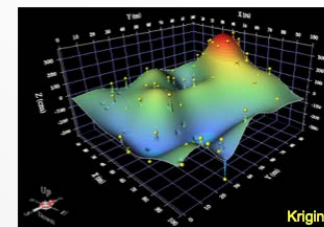
Test Case for Processing Services: Annual CTD Datasets

Data from R/V Håkon Mosby
(Norwegian research vessel)

Geostatistical analysis:
- Download from the IMR FTP site
- File structure analysis (NetCDF):
  - 2002-2016 period
  - Conductivity, Temperature, Salinity
- Spatial analysis of samples density per grid cell and per month
- Design and implementation of the **interpolation** service

http://rgeostats.free.fr/doc/Files/intaros7.html

Temperature (2002-03-07 => 2016-10-16)

data points by cell (<50m) (2003-10-01 => 2003-10-31)

Estimated Temperature at 1m depth
(2007-01-06 => 2007-03-30)

INTAROS

# TASK 5.5 - INTEGRATION OF NEW PROCESSING SERVICES

Partners : Terradue, NERSC, ARMINES

- Initial draft of D5.5 "Processing Services" with the "Service Description" form (input for partner service design)
  - Started defining a set of data processing services from NERSC (sea ice statistics, integrated acoustics-remote sensing data analysis, use of TRITON SW)
  - ARMINES ran the cloud platform online tutorials, and delivered an initial job design of the RGeostats service integration

- Overall, task activities are progressing according to the schedule

INTAROS

Integration of the RGeostats Toolbox Capabilities

- Job **design**: R Script (data wrappers) integration on a Cloud Sandbox instance

- Job **deployment** (upcoming upon user requests): service scaling on a Production environment
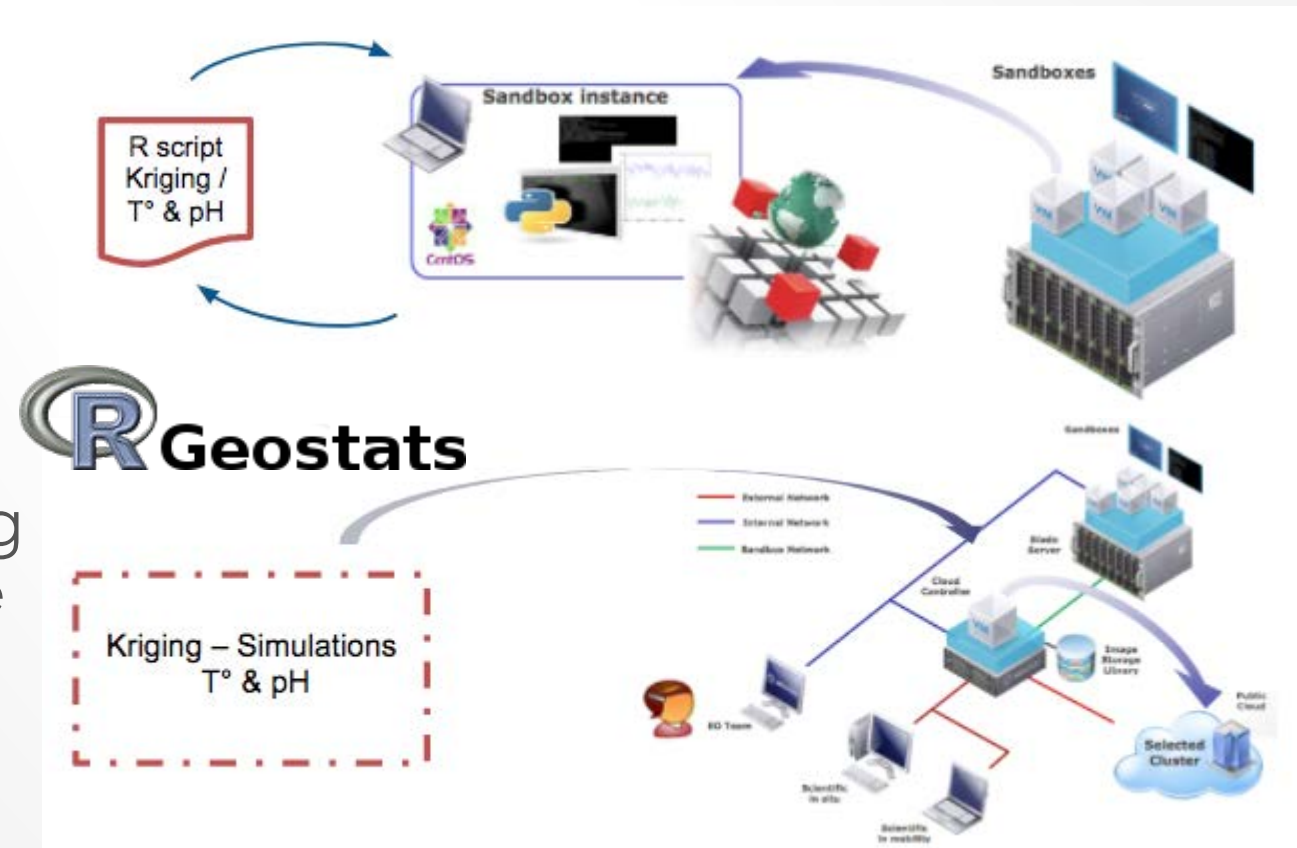
TASK 5.6 - IAOS PORTAL DEVELOPMENT

Partners: NERSC, Terradue

- Completed design and development plan for the web-GIS portal component of iAOS

- Engagement with other tasks and initiated discussion with WP6

  - Focus on requirements gathering for an initial definition of the iAOS portal user stories, including the description of data services (access/storage) and processing services (analysis, interpolation).

- Overall, task activities are progressing according to the schedule

INTAROS

iAOS Web Portal

- Provide an intuitive user interface to the search, the data access and the processing services in iAOS

- Provide an **entry point to the federated remote data repositories** and the developed data processing services

- **Visualize** retrieved multi-source data in a common map projection with basic GIS operations

- Enable the **execution of processing services** and the retrieval of data processing results

INTAROS

iAOS Web Portal

Design of processing services (tools)

- Sea ice statistics based on satellite data (NERSC)
- Assessing noise pollution in the ocean using passive acoustics data (NERSC)
- Generating SST fields for validation of climate models (Armines)

Web mapping component for survey data

- Linked to database with metadata (from questionnaire A,B,C)
- Planning started for data ingestion in DB and updating mapping

Collaboration on Metadata web tool for eddy covariance flux sites

- Desktop GIS for mapping (QGIS) → QGIS 2 Web Client

WP5 CURRENT ACTIONS ON UPCOMING DELIVERABLES

- D5.2 iAOS platform and tools (M24)
  - Provides secure work environment for VM access and support to application integration

- D5.3 Data integrated from existing repositories into iAOS (M24)
  - Selection of suitable show cases for a first integration

- D5.4 iAOS portal with user manual (M24)
  - Definition of User stories

- D5.5 iAOS requirements and architectural design v2 (M36)
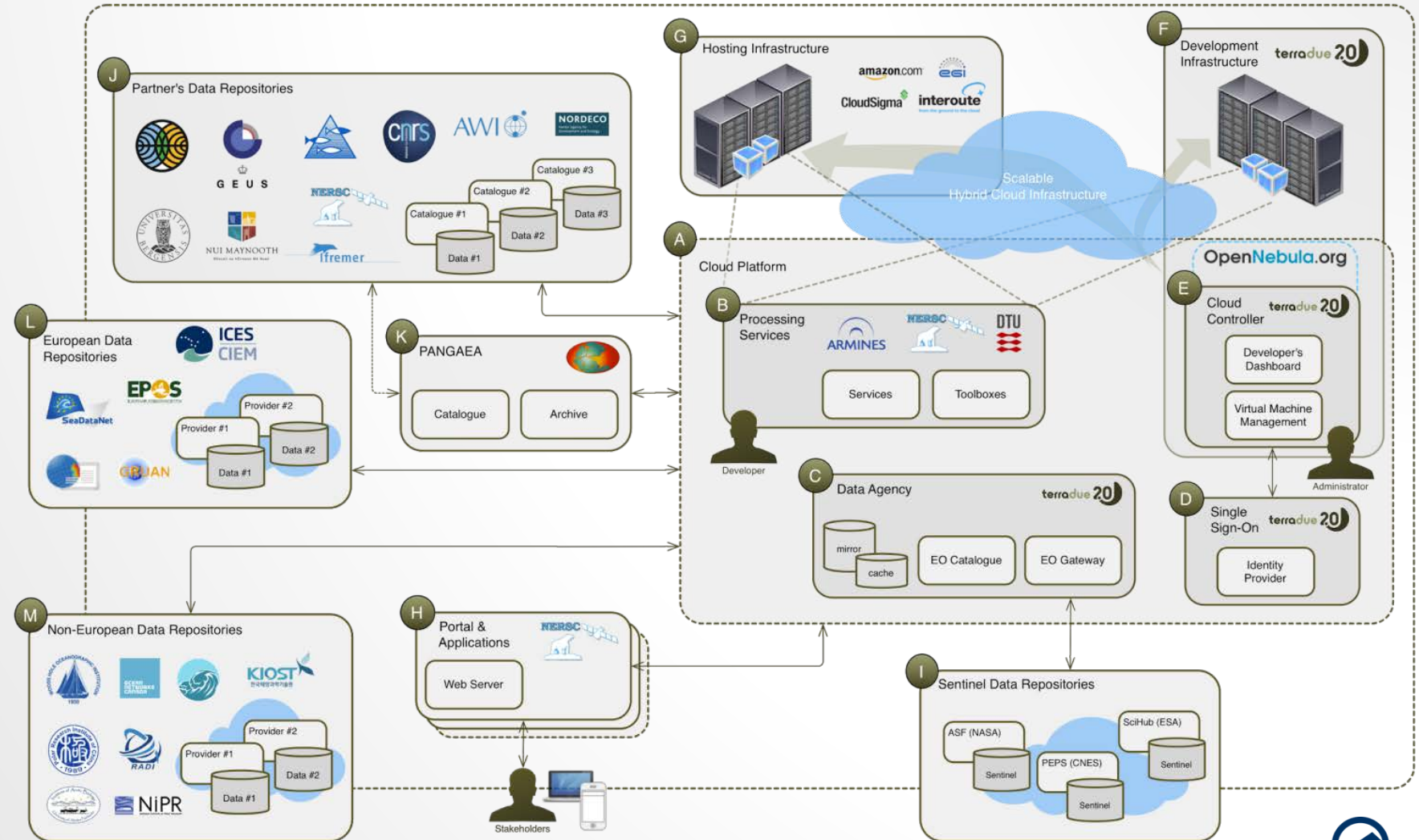 - Evolve architecture to meet the main challenges of the observing system

End of presentation

The iAOS Platform Architecture
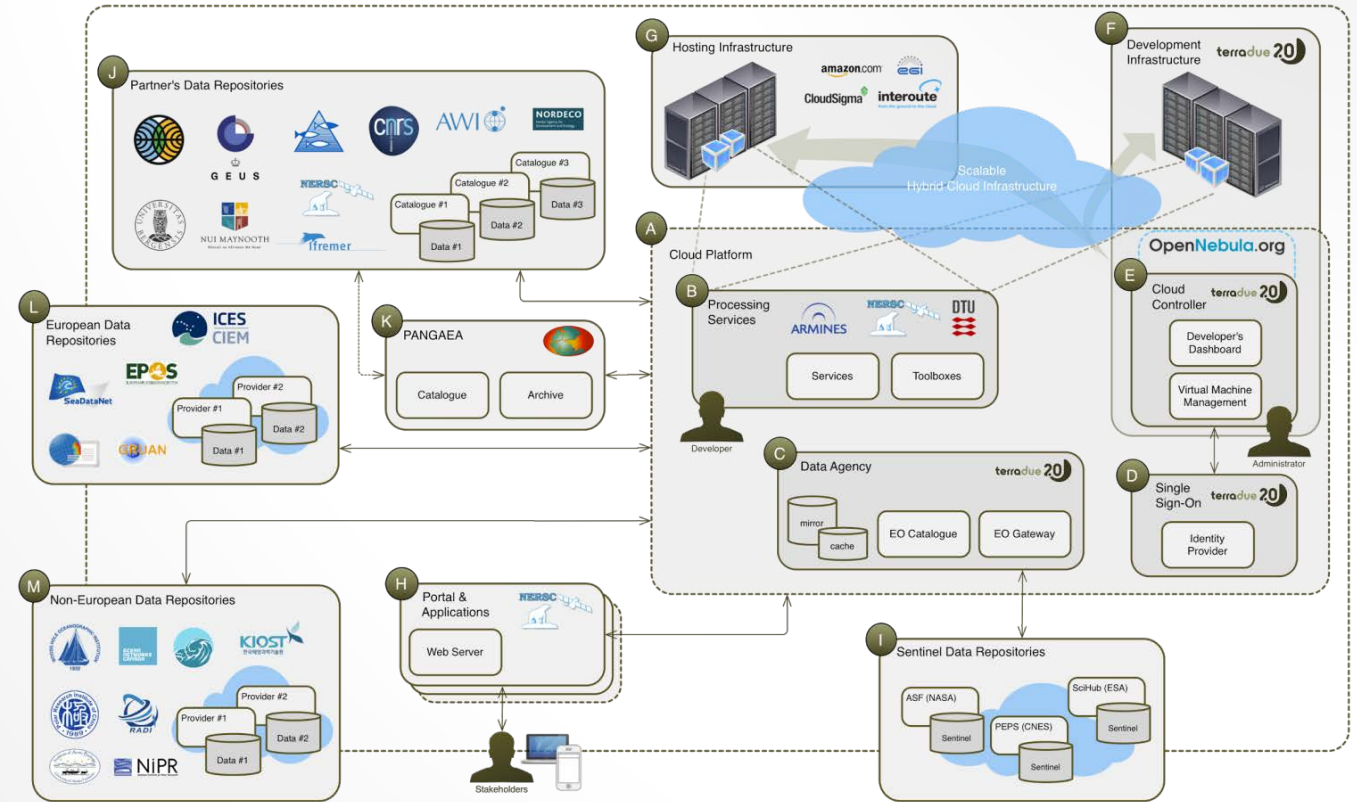
iAOS Platform Architecture

WP5 TECHNICAL OVERVIEW

iAOS Portal

End-user exploitation environment for Users

User stories defined

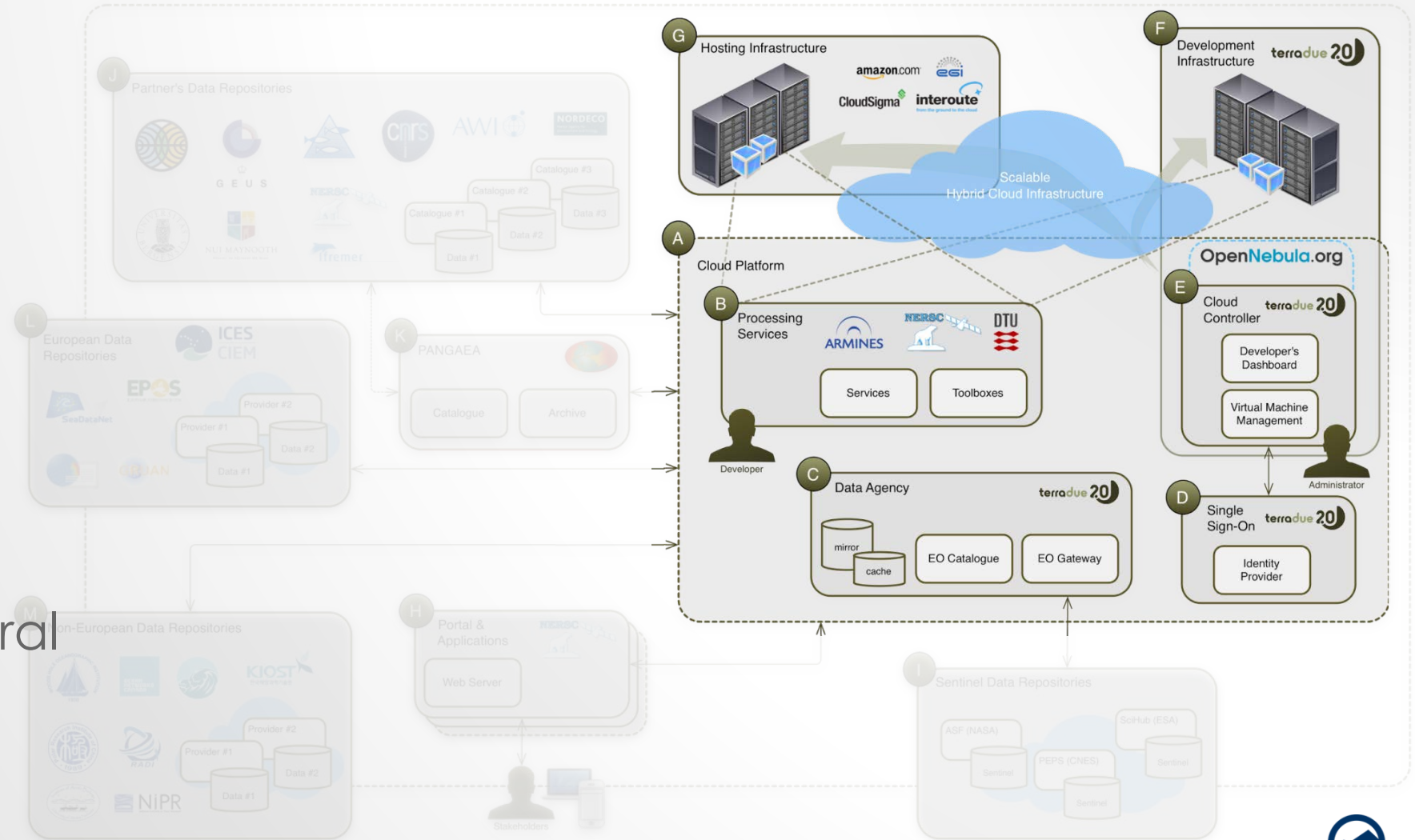Development starting this year
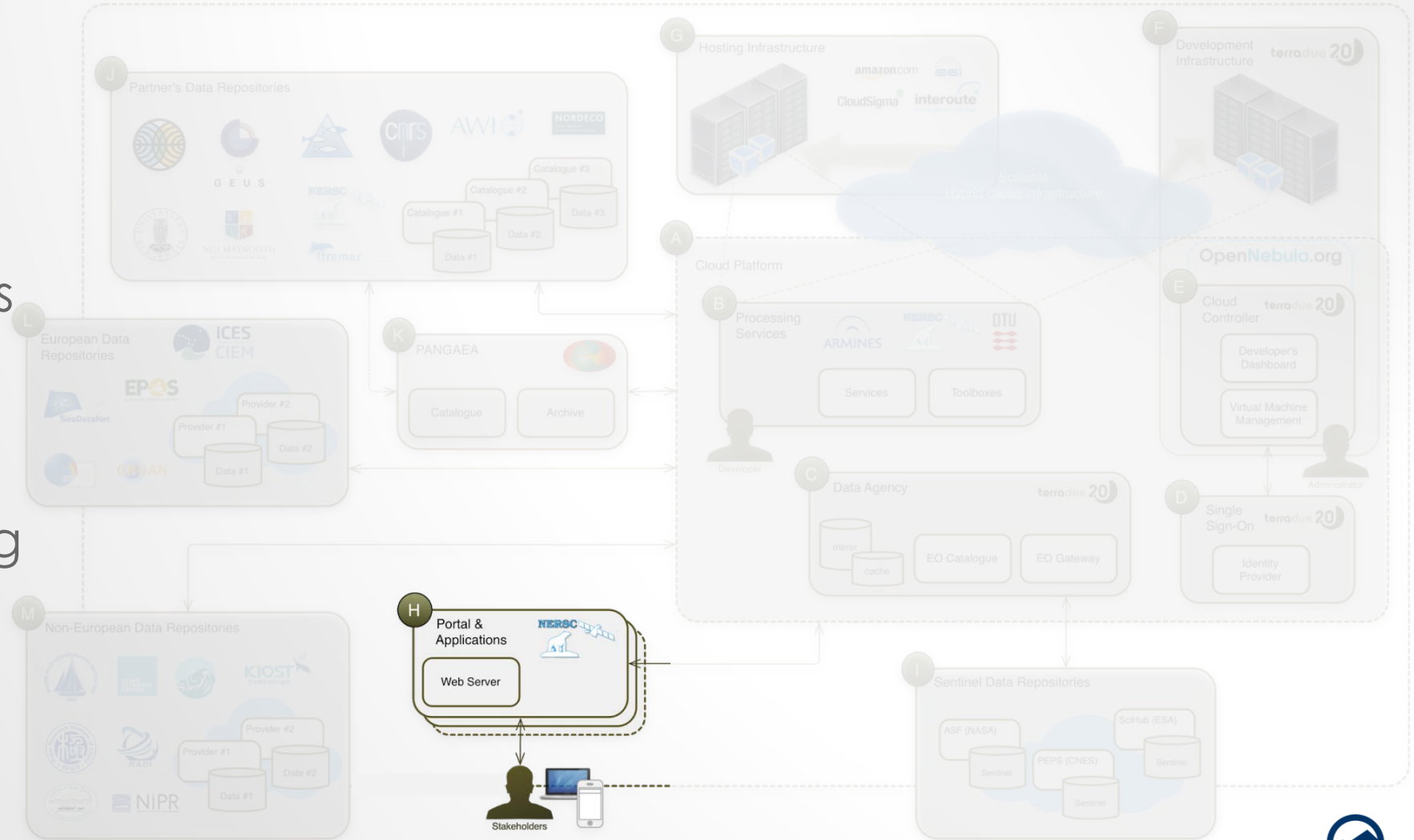
WP5 TECHNICAL OVERVIEW

Cloud Infrastructure

Collaborative workspace already available

Hosted on Terradue Cloud Platform
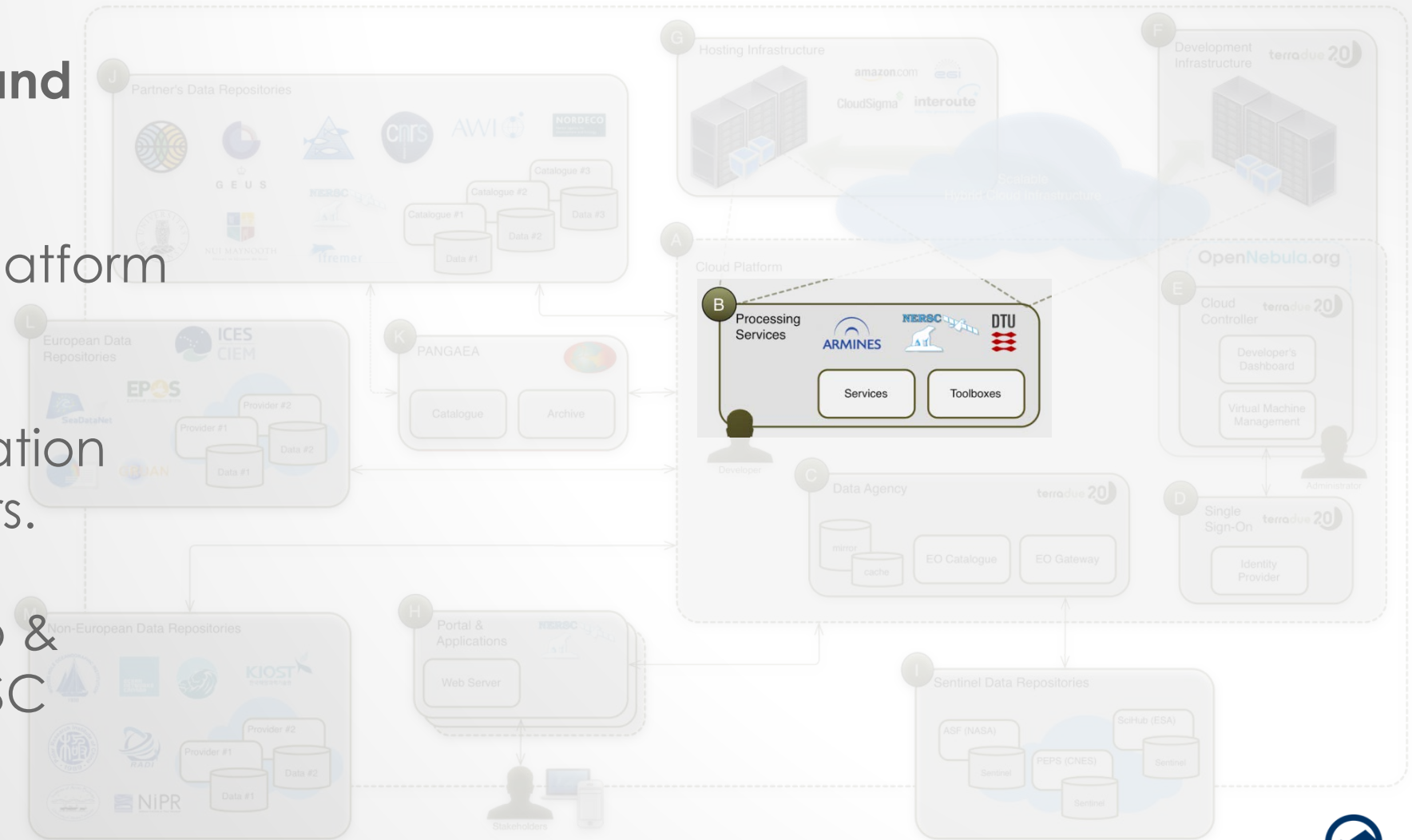
Connections to several ICT providers

WP5 TECHNICAL OVERVIEW

Software toolboxes and processing services

Maintained on the Platform

From integration to deployment in operation on production servers.

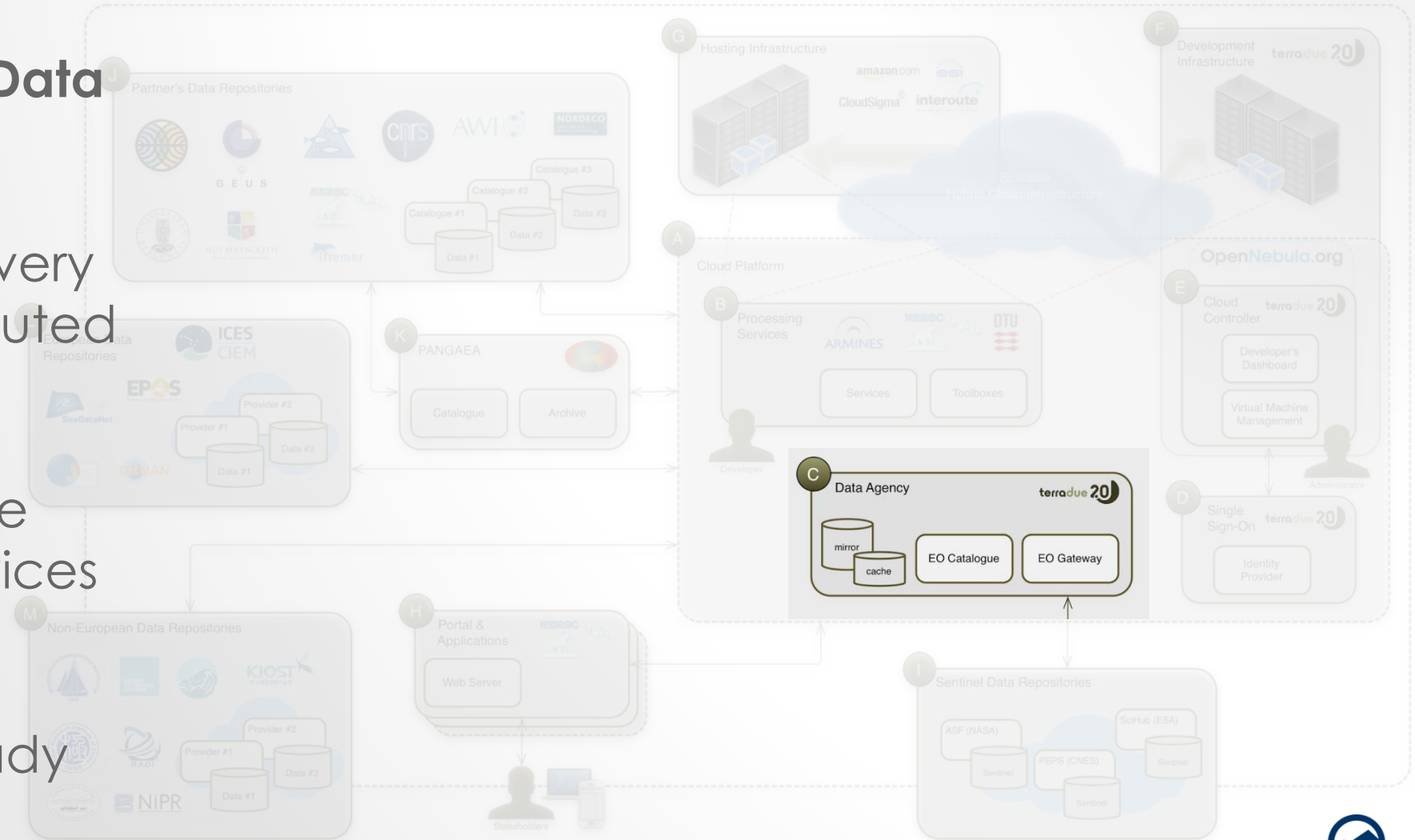First RGeostats demo & initial design for NERSC Services

WP5 TECHNICAL OVERVIEW

EO Catalog and EO Data Gateway services

Programmatic discovery and access to distributed EO data repositories

Designed for scalable data processing services
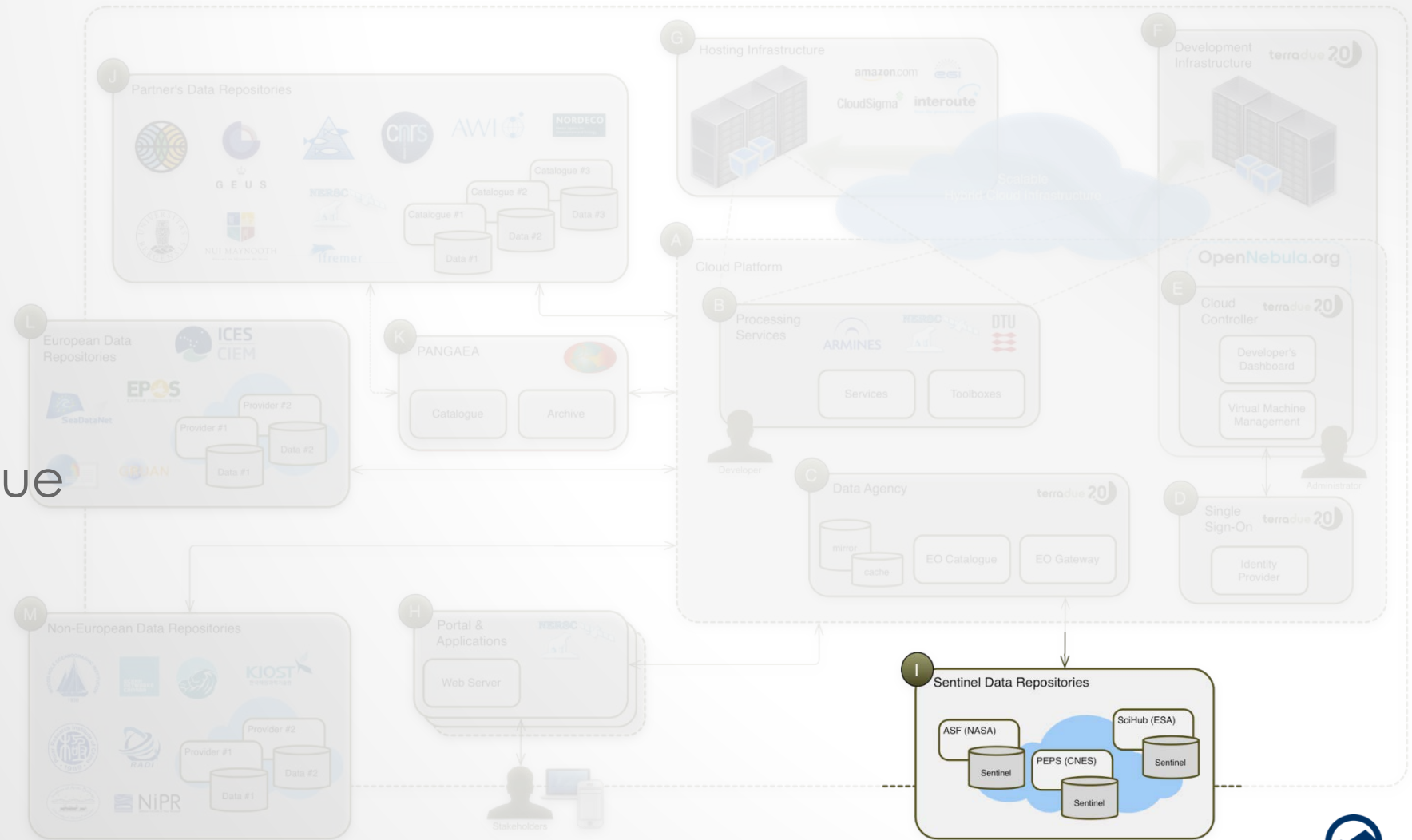
Connections to Copernicus Hubs ready

INTAROS

WP5 TECHNICAL OVERVIEW

Copernicus Sentinel data repositories

Pool of Copernicus data repositories

Federated on Terradue Cloud Platform
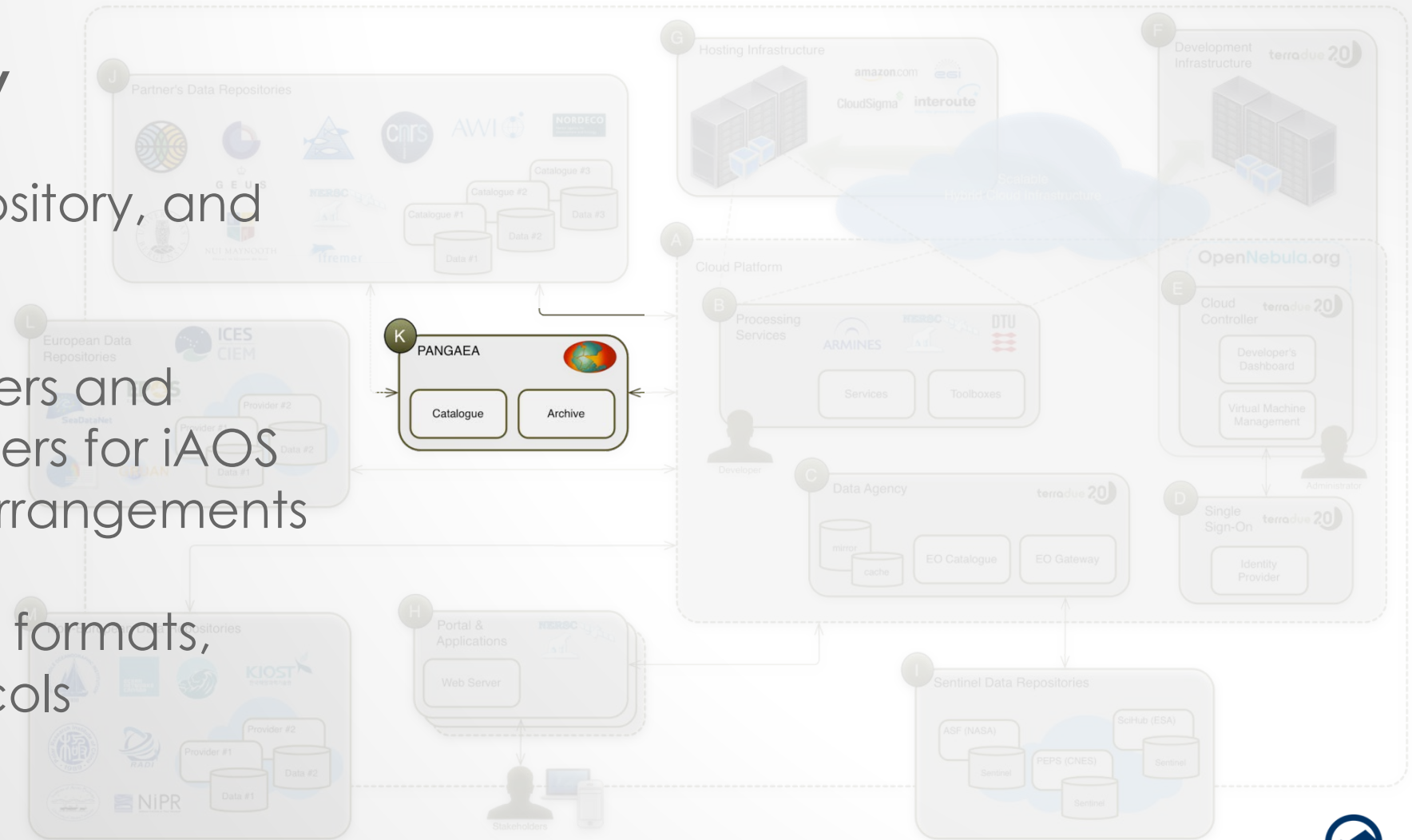
Available from the processing services.

WP5 TECHNICAL OVERVIEW

PANGAEA Repository

PANGAEA data repository, and catalog entries

Also federates partners and relevant data providers for iAOS via interoperability arrangements

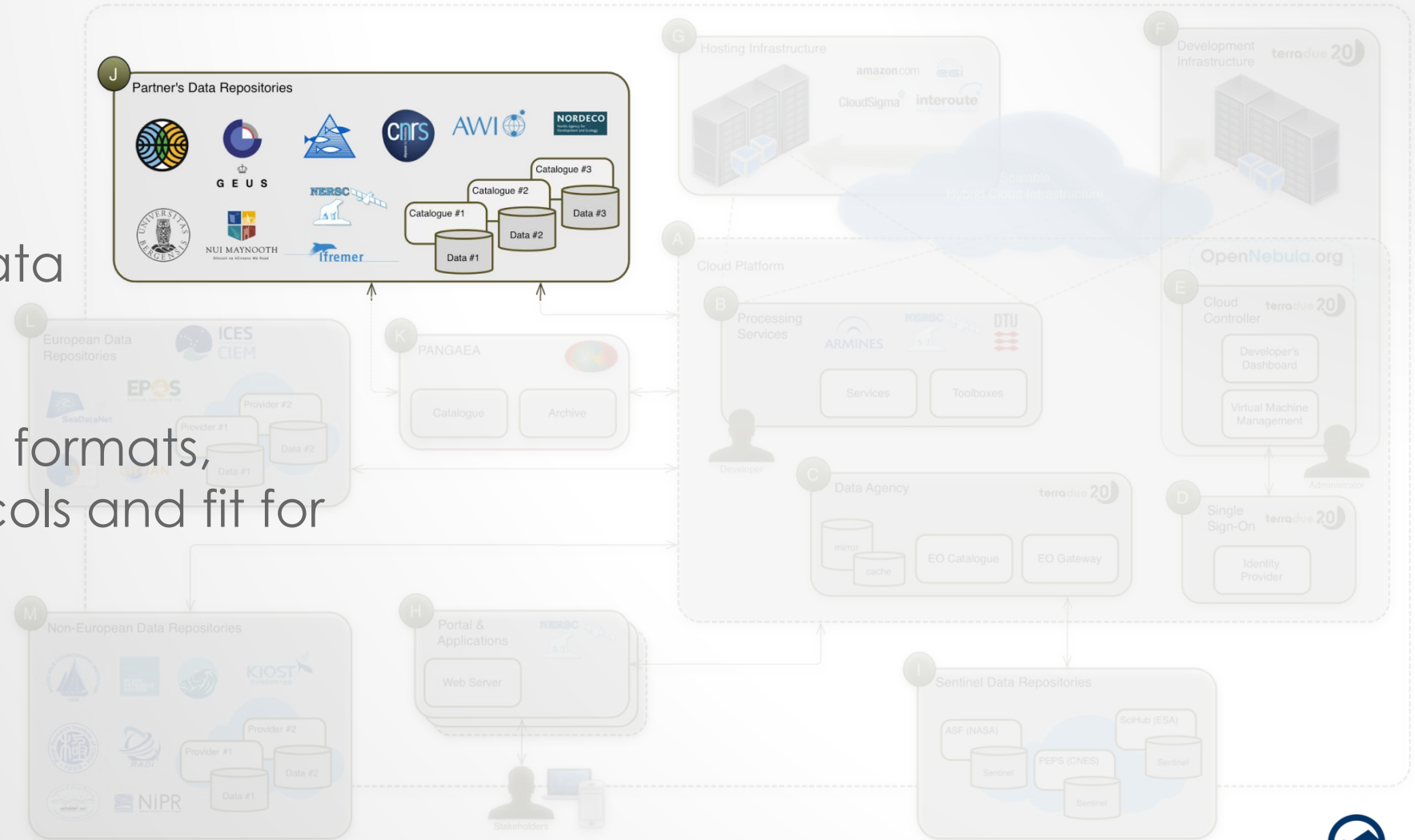Data and metadata formats, online access protocols

WP5 TECHNICAL OVERVIEW

Data Repositories

INTAROS Partner's data repositories

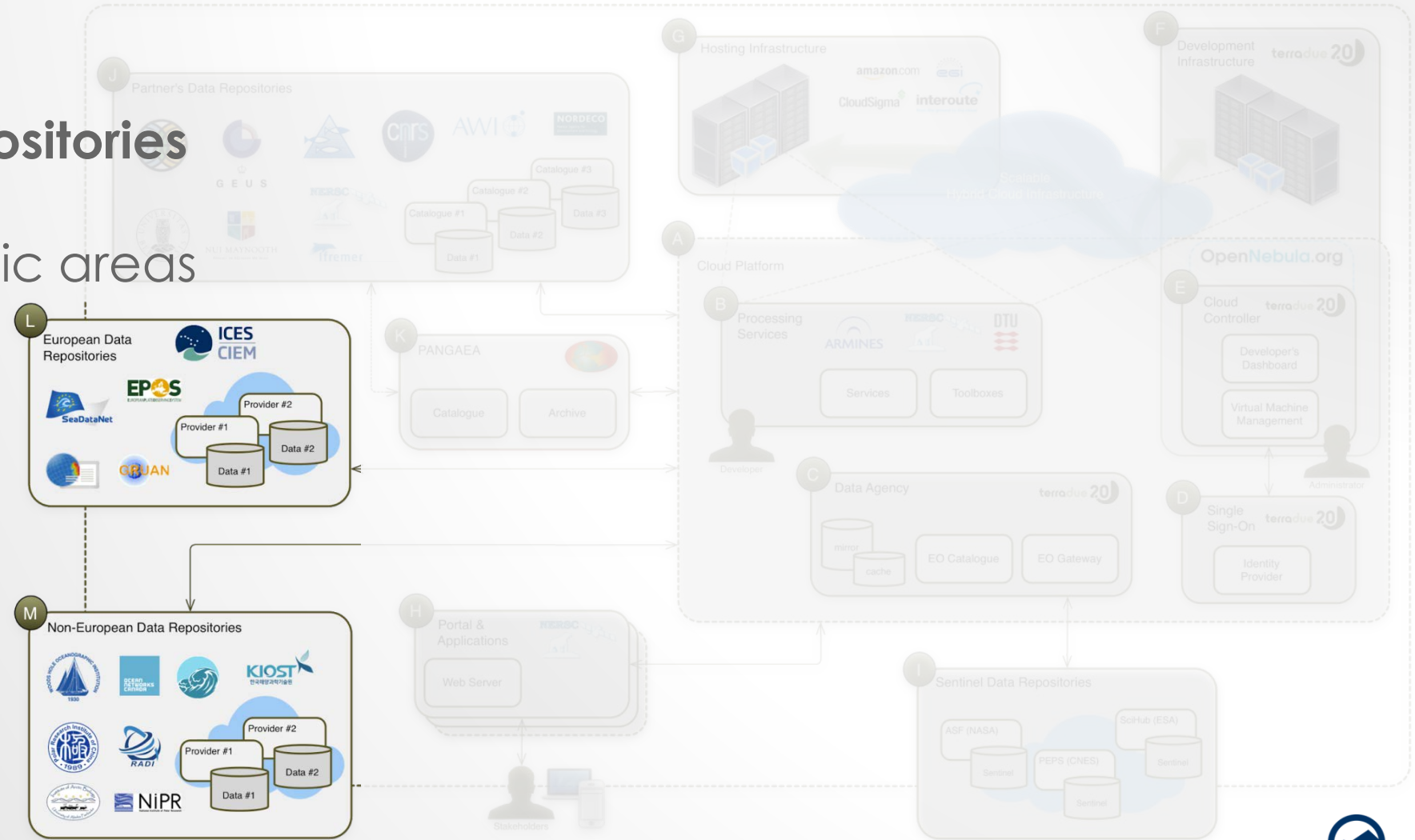Data and metadata formats, online access protocols and fit for purpose

WP5 TECHNICAL OVERVIEW

Federated Data Repositories

Relevant for the Arctic areas management

To be Federated in iAOS (WP2/WP5 collaboration) by interoperability arrangements

WP5 TECHNICAL OVERVIEW

iAOS Portal

End-user exploitation environment for Users

Access to federated data repositories

User access to the Processing Services deployed on the Platform.

Sea Ice + Acoustic Data + Geostatistics