WP5 - DATA INTEGRATION AND MANAGEMENT

Status and plan for the next period

INTAROS General Assembly

Finnish Meteorological Institute,
Erik Palmenin aukio 1, Helsinki
13.30-14.00
January 10th 2018

Pedro Gonçalves, Terradue Srl (lead)
Torill Hamre, NERSC (co-lead)

Agenda

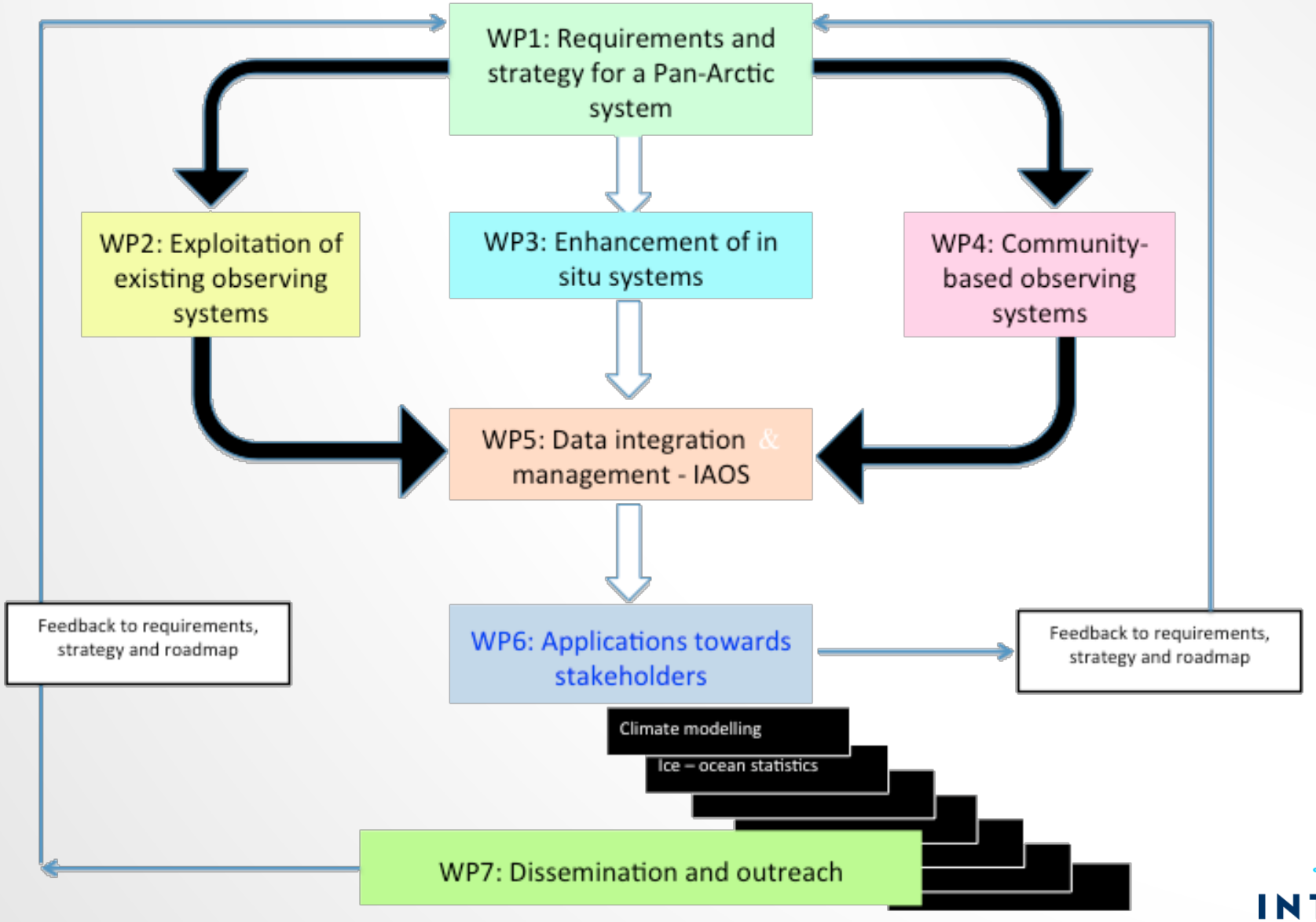
- Objectives
- Platform Usage Scenarios
- Technical Requirements
- Overview of Platform Architecture
- Status and Achievement
- Tasks Report

WP5 DATA INTEGRATION & MANAGEMENT OBJECTIVES

- Integrate multidisciplinary and distributed data repositories into a scalable and resilient integrated Arctic observing system (iAOS)
  - Provide seamless access to observations and derived parameters.
  - Integrate a set of tools for data analysis, transformation and visualization.
  - Support geo-statistical methods for interpolation of spatiotemporal datasets.
- Develop processing services for sea ice statistics, integrated acoustics-remote sensing data analysis
- Support the processing of new observations from WP2-4 and store generated datasets in an iAOS-enabled repository

INTAROS

WP5 DATA INTEGRATION & MANAGEMENT



WP5 DATA INTEGRATION & MANAGEMENT
ACHIEVEMENTS

- Monthly meeting 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)

INTAROS

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, the task activities are progressing according to the schedule.

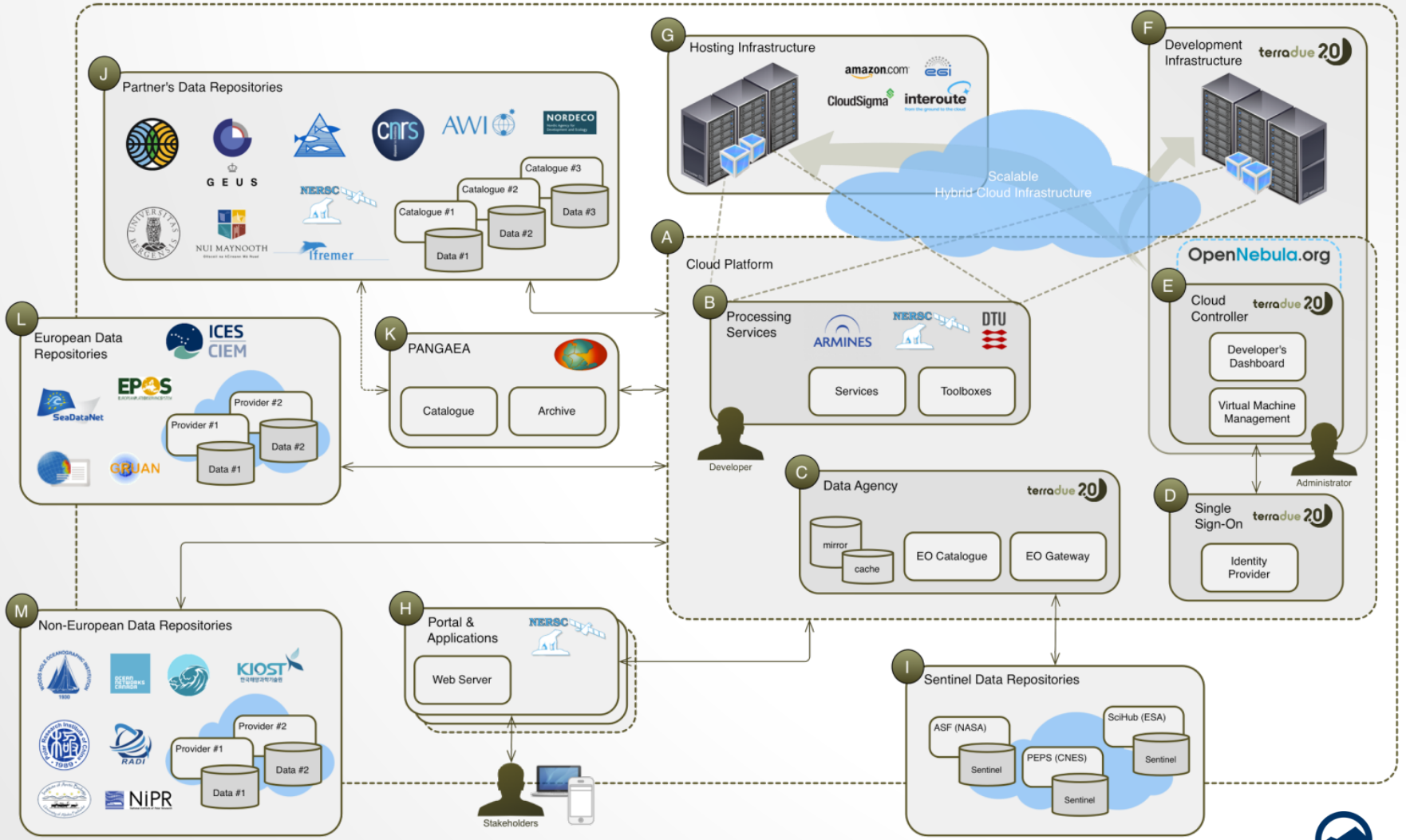
INTAROS

iAOS Platform
Operational scenarios

- Integration of Data Access facilities (Data)
- Design and integration of scalable processing applications (Cloud)
- Management of a Platform's resources for hosted data processing service (Cloud)
- Exploitation of data access services (Portal)
- Exploitation of data processing services (Portal)
- Administration of Cloud resources

WP5 TECHNICAL OVERVIEW

iAOS Platform architecture



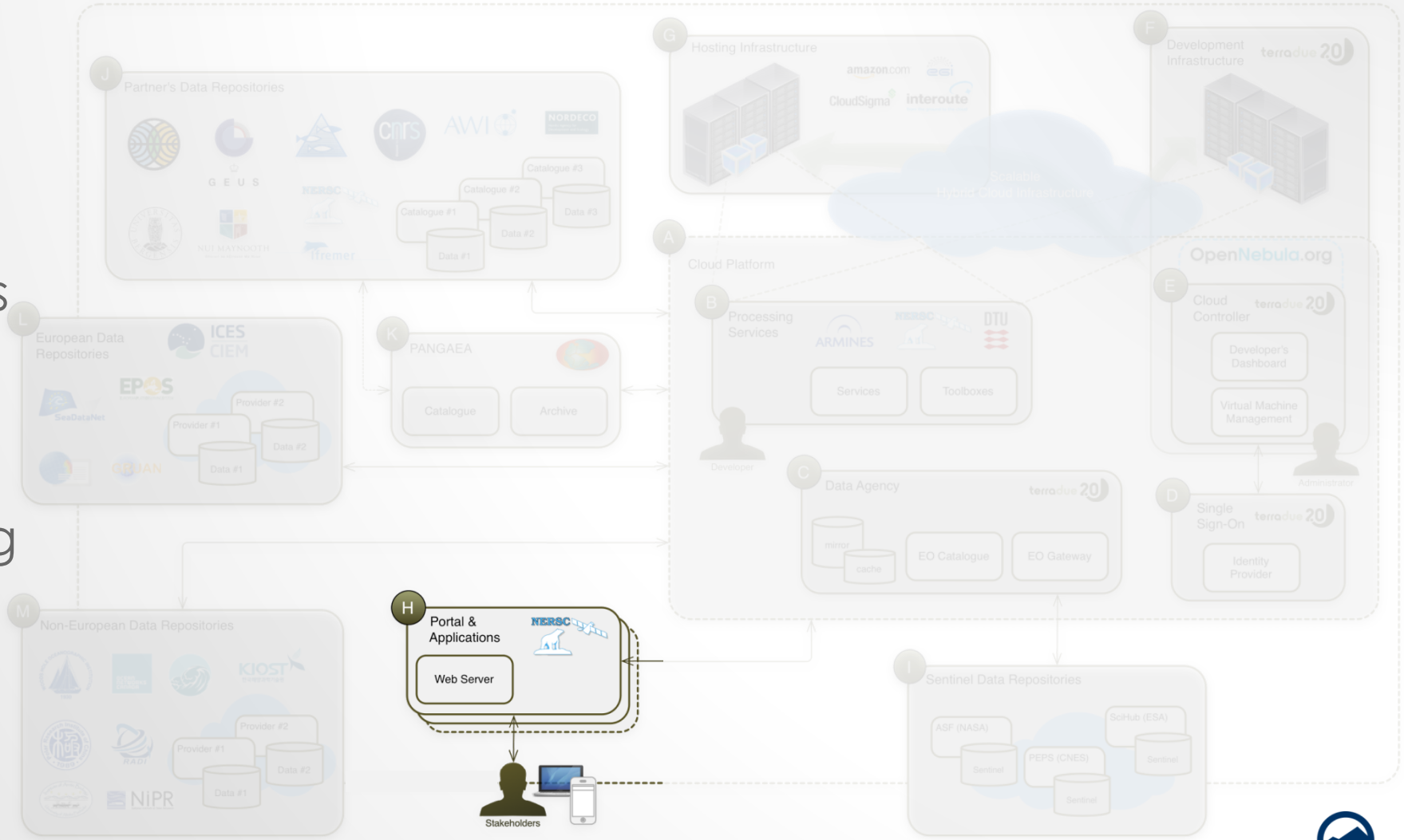
WP5 TECHNICAL OVERVIEW

iAOS Portal

End-user exploitation environment for Users

User stories defined

Development starting this year

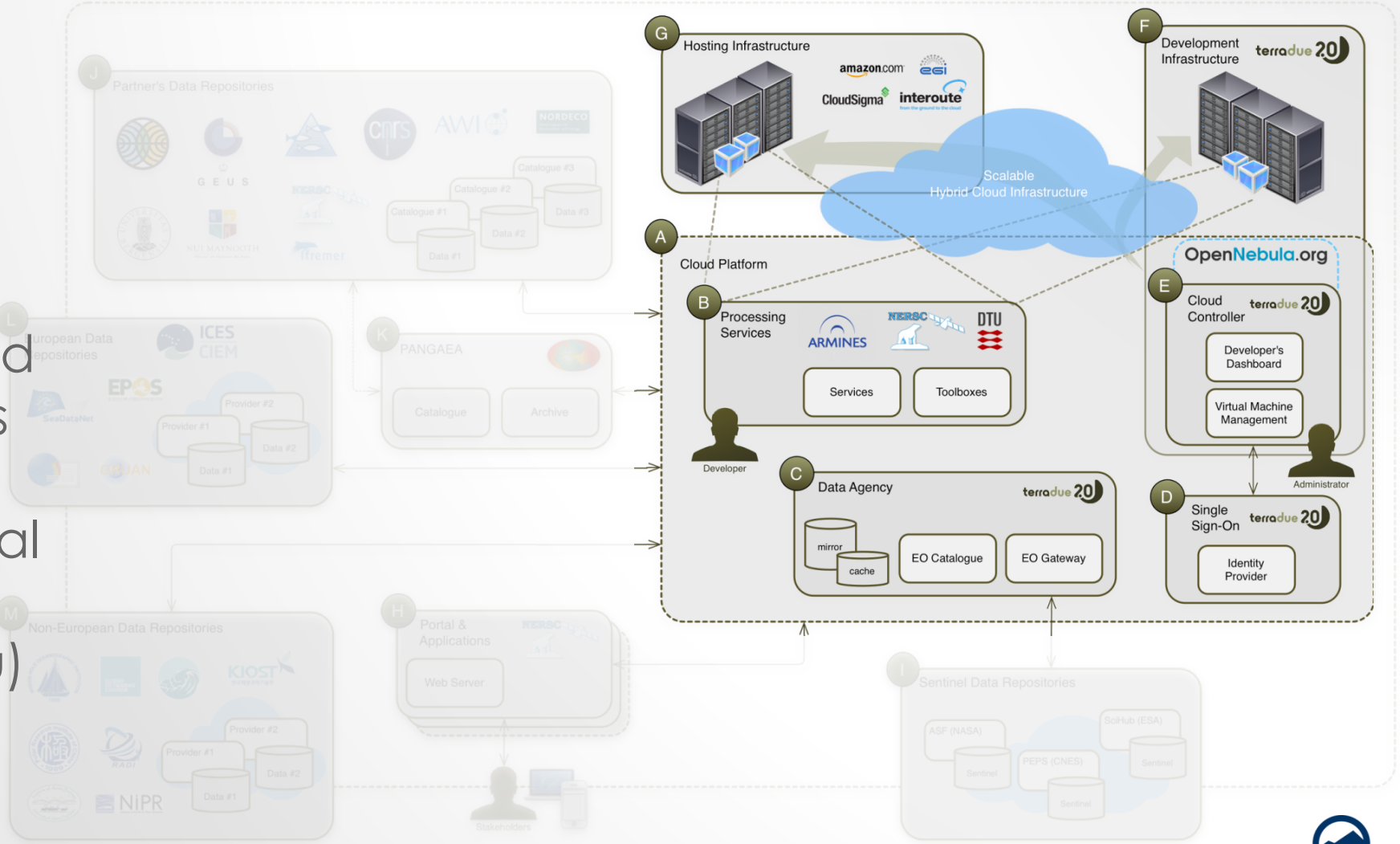


WP5 TECHNICAL OVERVIEW

Cloud Infrastructure

From development in private Cloud environment to Hosted Processing operations

Connections to several ICT providers already available (e.g. EGI.eu)



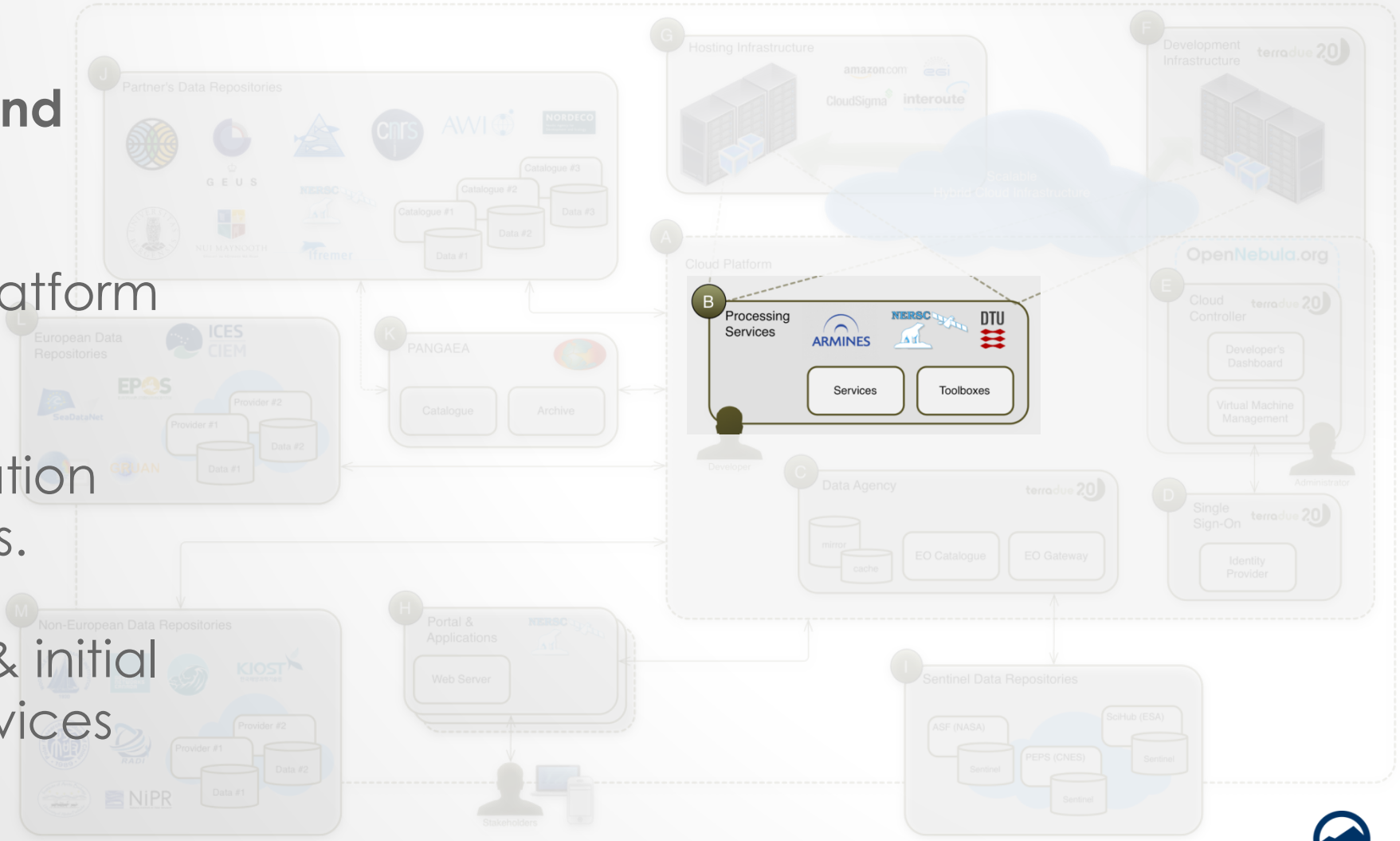
WP5 TECHNICAL OVERVIEW

Software toolboxes and processing services

Maintained on the Platform

From integration to deployment in operation on production servers.

First rGeoStat demo & initial design for NERSC Services



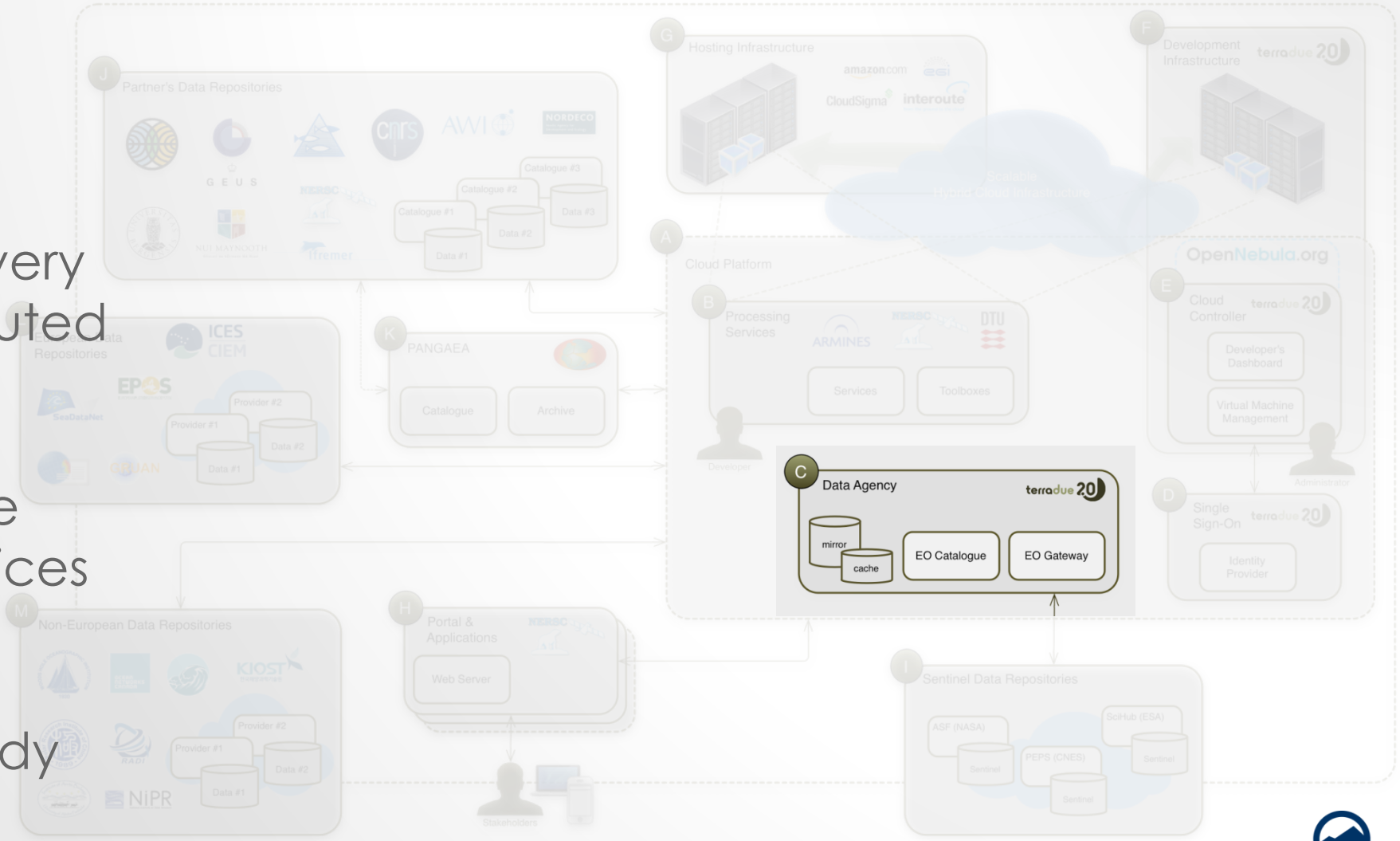
WP5 TECHNICAL OVERVIEW

Catalog and Data Gateway services

Programmatic discovery and access to distributed EO data repositories

Designed for scalable data processing services

Connections to Copernicus Hubs ready

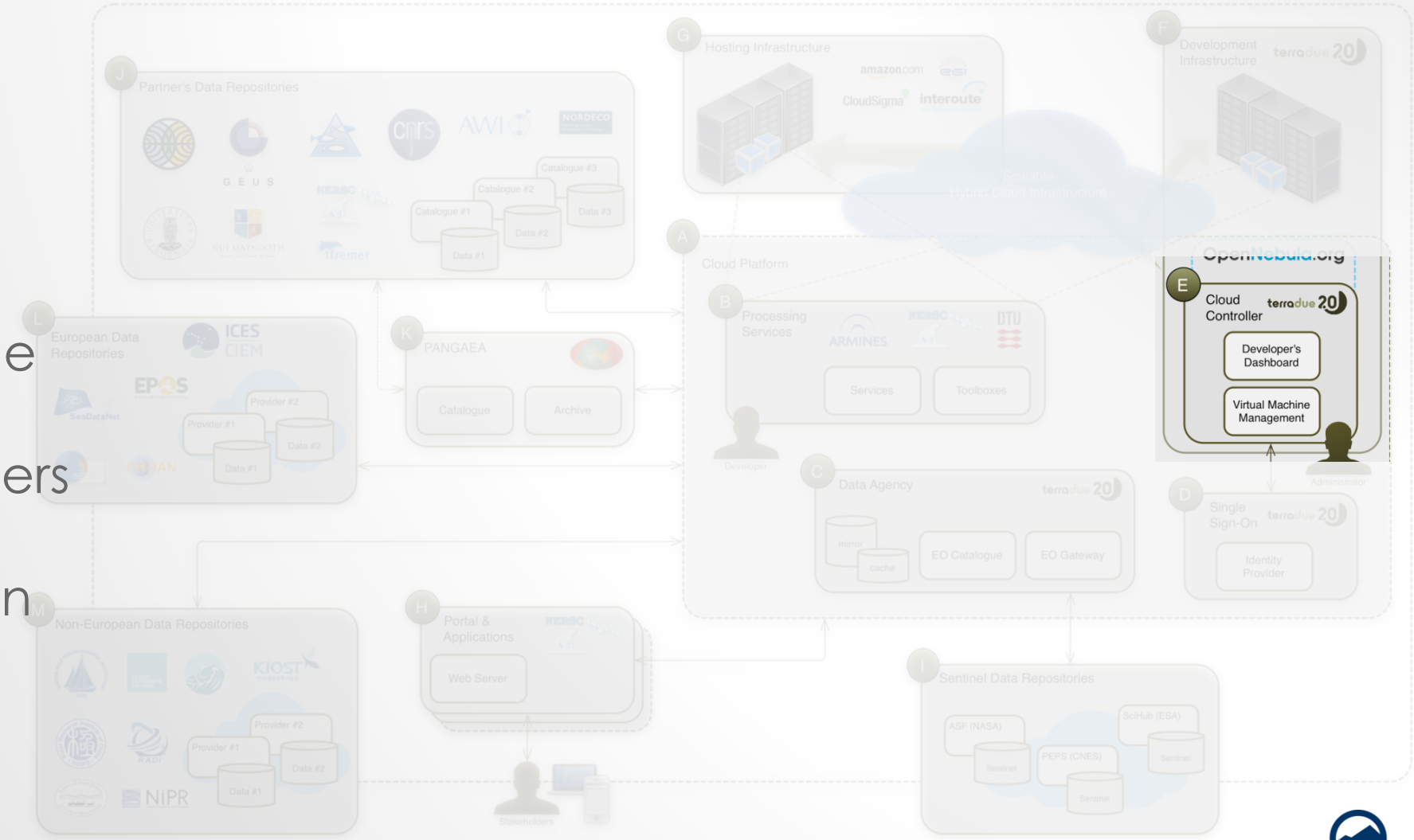


WP5 TECHNICAL OVERVIEW

Cloud Controller Service

Cloud Dashboard interfaces to manage Virtual Machines allocated to iAOS users

Operates Application Integration and Production Centers



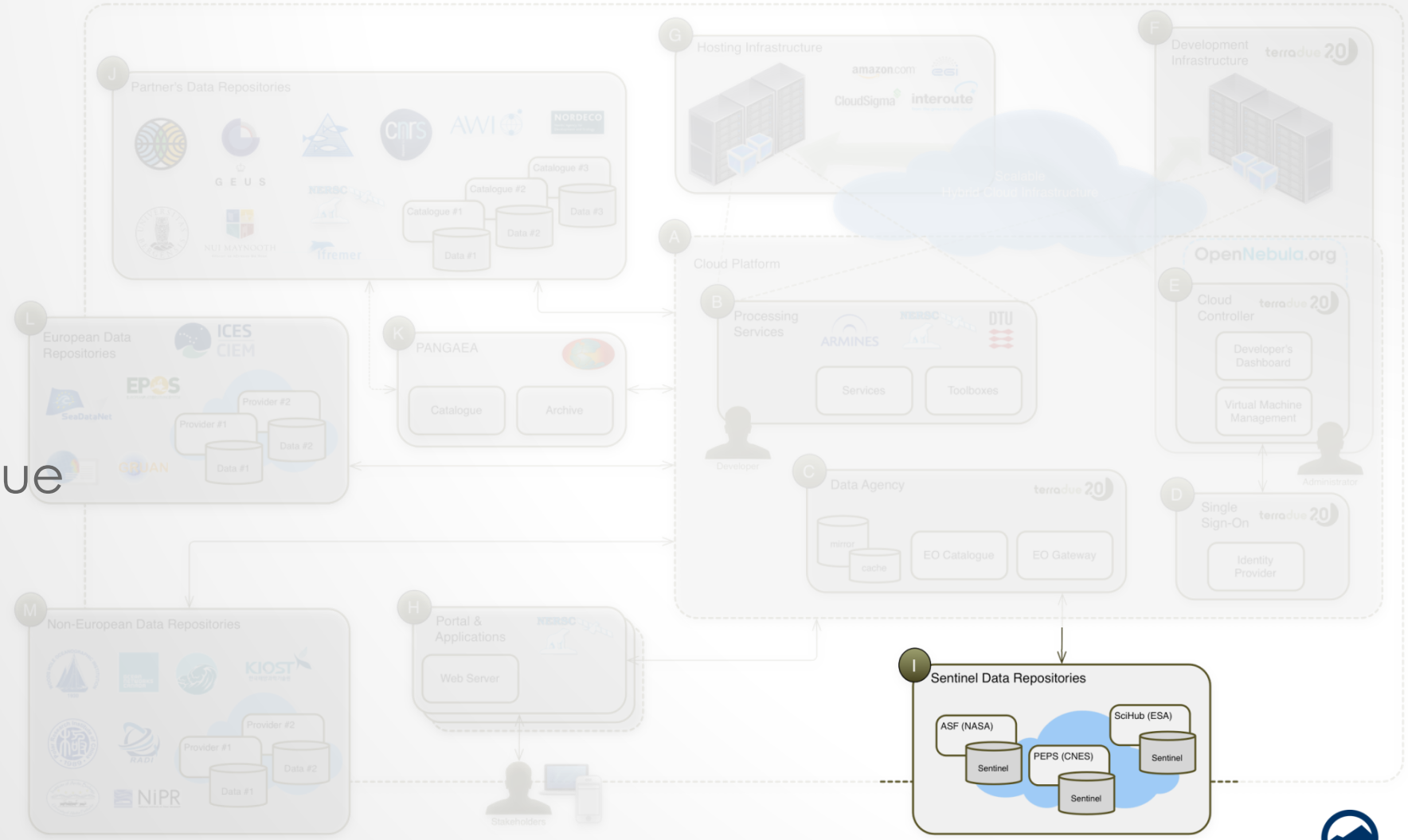
WP5 TECHNICAL OVERVIEW

Copernicus Sentinel data repositories

Pool of Copernicus data repositories

Federated on Terradue Cloud Platform

Available from the processing services.



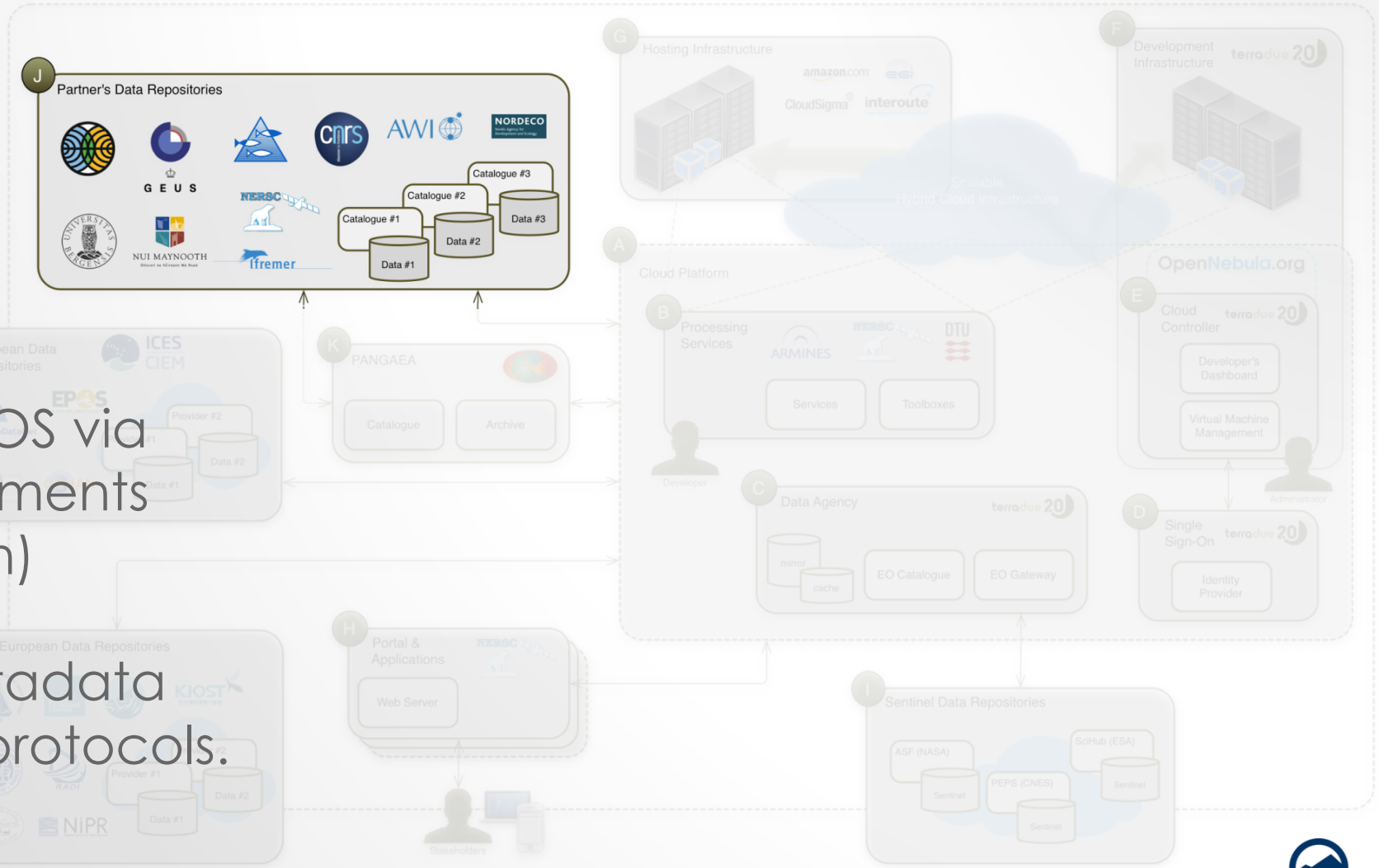
WP5 TECHNICAL OVERVIEW

Data Repositories

INTAROS Partner's data repositories

To be federated into iAOS via interoperability arrangements (WP2/WP5 collaboration)

Covering data and metadata formats, online access protocols.



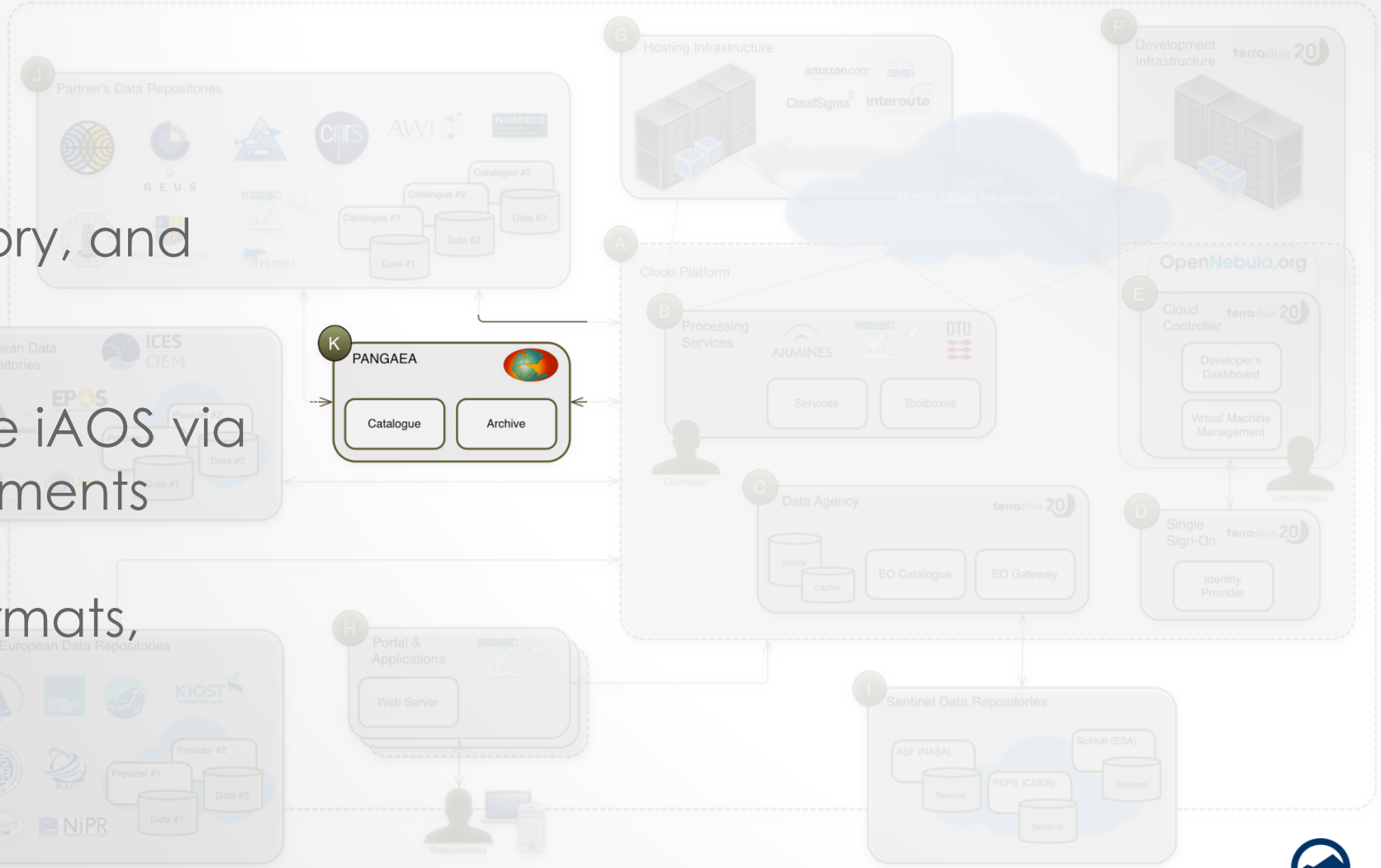
WP5 TECHNICAL OVERVIEW

PANGAEA Repository

PANGAEA data repository, and catalog entries

To be federated into the iAOS via interoperability arrangements

Data and metadata formats, online access protocols

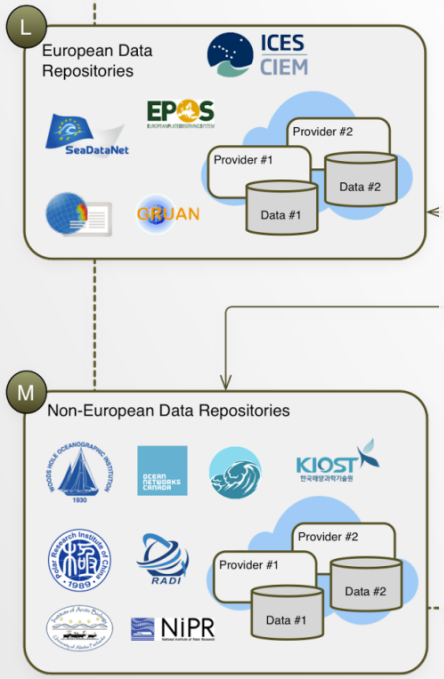


WP5 TECHNICAL OVERVIEW

Federated Data Repositories

Relevant for the Arctic areas management

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



WP5 TECHNICAL OVERVIEW

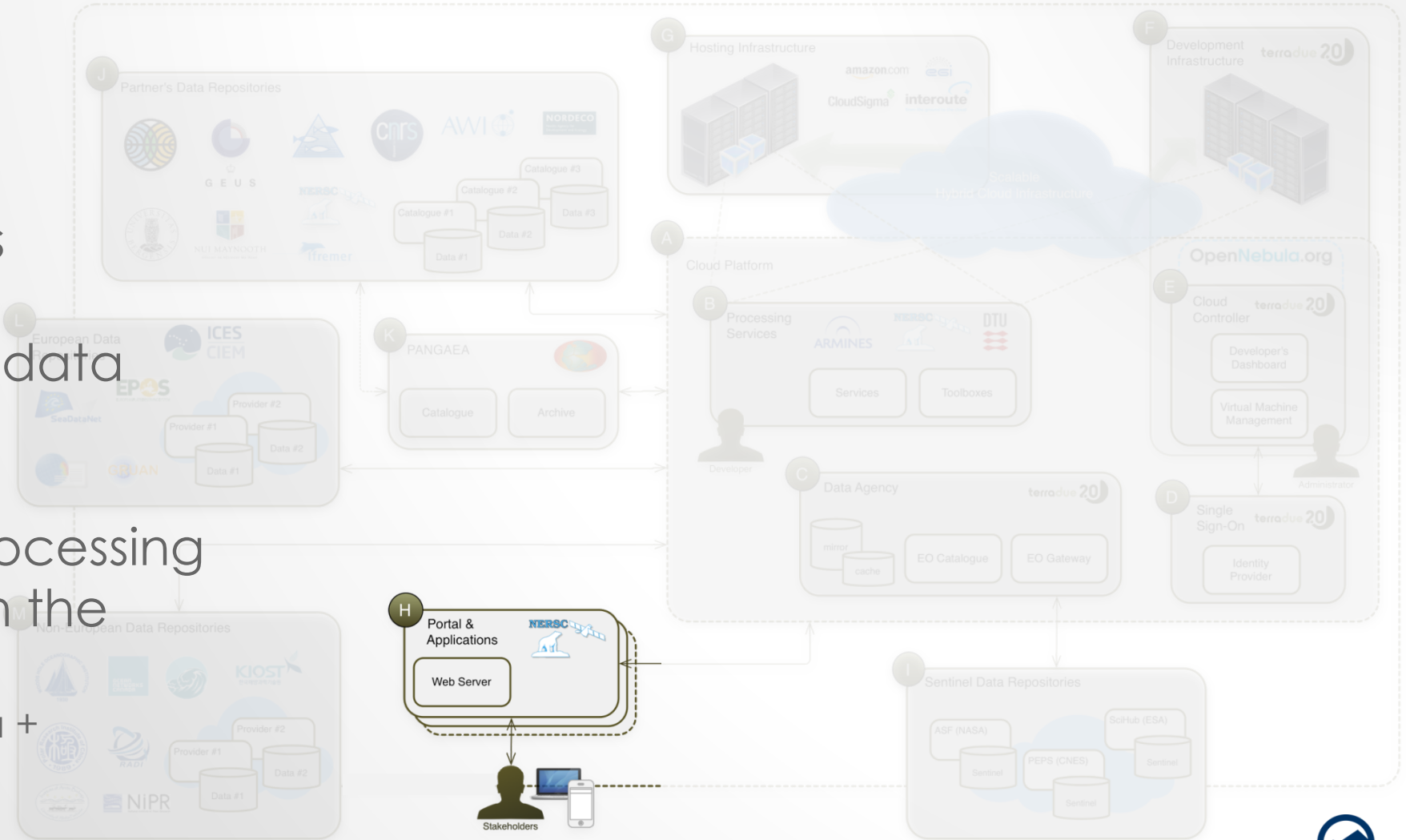
iAOS Portal

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



TASK 5.2 - IAOS PLATFORM DEPLOYMENT AND OPERATION

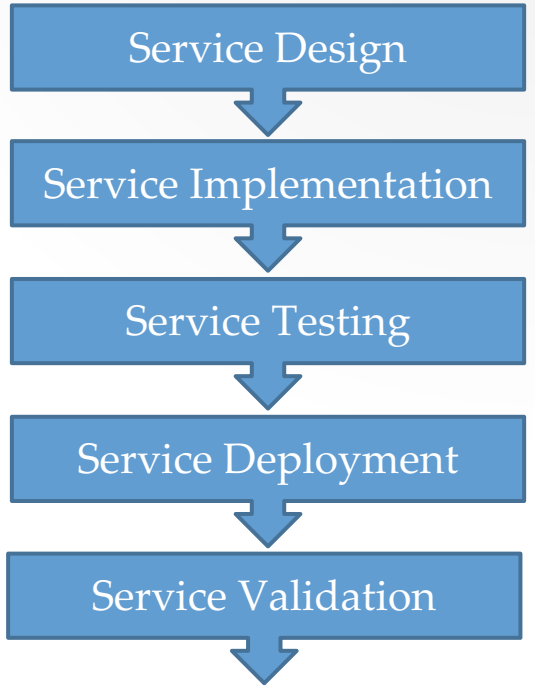
Partners: Terradue

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

INTAROS

Integration of new Processing Services

- Exploit the data processing tools and geo-statistical 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



TASK 5.3 - INTEGRATE DATA FROM EXISTING REPOSITORIES INTO IAOS

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

- Established a solid link with WP2 from an very early stage.
- Defined initial task 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 the data integration from existing repositories into iAOS
- Overall, the task activities are progressing according to the schedule.

INTAROS

TASK 5.4 - DEVELOPMENT OF GEO-STATISTICAL METHODS FOR DATA INTEGRATION

Partners : ARMINES, NERSC, DTU

- Installation and deployment of the RGeostats package on the Cloud platform readily available to the INTAROS community.
- Development of a first application example with data similar to that used by the Arctic research community.
  - Dissemination material prepared to outreach the iAOS users community.
- Overall, the task activities are progressing according to the schedule.

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)
  - ARMINES ran the cloud platform online tutorials, and delivered an initial job design of the RGeostats service integration
- Overall, the task activities are progressing according to the schedule.

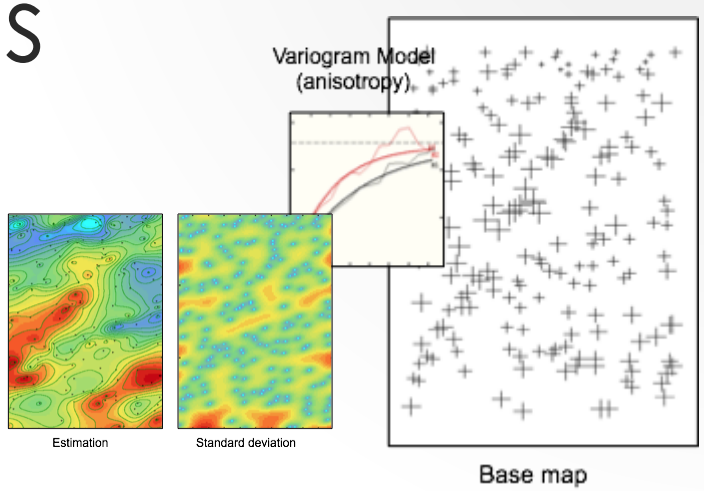
INTAROS

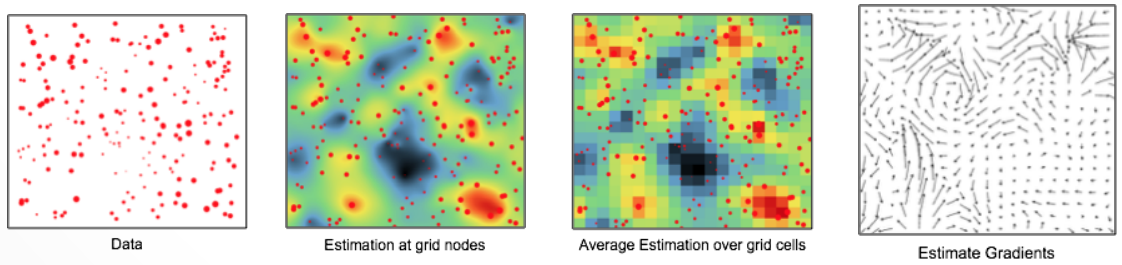
Integration of the RGeostats Toolbox Capabilities

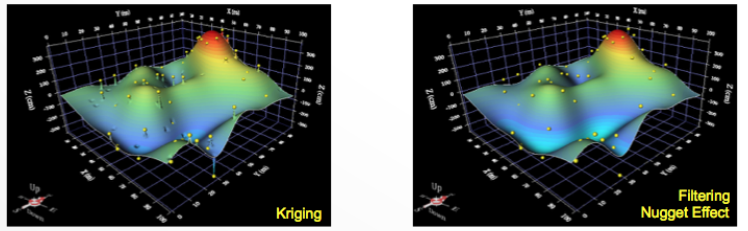
- The most complete free software for geostatistics
  - Packaged under R platform
- Main key features:
  - Data of any space dimension (space and time)
  - With any number of variables treated simultaneously
  - Possible extension for spatio-temporal models
  - Big number of data/targets (up to memory capacity for R)
  - Stochastic partial differential equation models
- 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







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

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

INTAROS

iAOS Web Portal

- Provide an intuitive user interface to the search, access and processing services in iAOS
- Provide an entry point to the integrated data repositories and the developed 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

WP5 CURRENT ACTIONS ON UPCOMING DELIVERABLES

- D5.2 iAOS platform and tools (M24)
  - Provide 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
- D 5.4 iAOS portal with user manual (M24)
  - Definition of User stories
- D5.1 iAOS requirements and architectural design (M36)
  - Evolve architecture to meet the main challenges of the observing system

INTAROS

End of presentation