



PHIDIAS

Prototype of HPC/Data Infrastructure for On-demand Services

PHIDIAS Scientific User Committee

Francois Bodin

University of Rennes



The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Cofinanced by the Connecting Europe

- ❖ The Committee have a general guidance role including
 - ❖ Provide guidance and opinion on project technical challenges, solutions and choices
 - ❖ Guide the project technical architecture in the best direction
 - ❖ Connect the project to world-wide groups and organisations

- ❖ It aims at helping better understanding how to mix HPC culture with Data culture

- ❖ And how to better use PHIDIAS advances?

 PHIDIAS coordinator

 Boris Dintrans

 External experts

Jean-Thomas ACQUAVIVA DDN Storage Evolve coordinator, data logistic	Michèle FICHAUT IFREMER SeaDataNet / SeaDataCloud	Maryvonne GERIN-LASLIER CNRS Astrophysics, research infrastructure	Sylvie JOUSSAUME IPSL Climate
Bruno RAFFIN INRIA Scientific workflows	Debora TESTI CINECA European Open Science Cloud	Jean-Pierre VILOTTE IPGP Seismology, INSU	

 SUC Coordinator & President

 François Bodin

 HPC, AQMO coordinator, former EXDCI2 scientific director

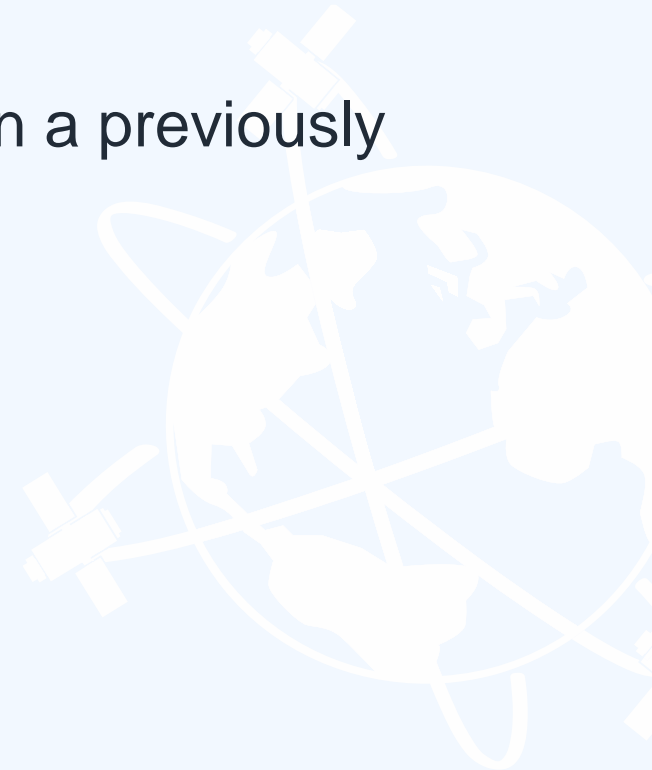
Committee Modus Operandi

❖ Committee meetings were generally split into two parts

❖ A short overview of last PHIDIAS results

❖ A debate between PHIDIAS and the committee members on a previously chosen topic

❖ An example to follow



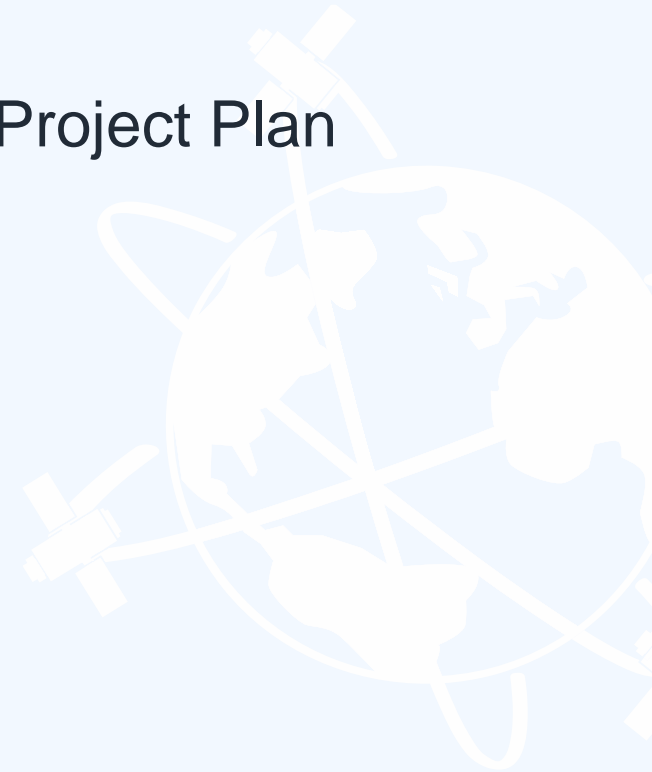
 The debate was

 Can PHIDIAS be the support for “a metadata centric approach”

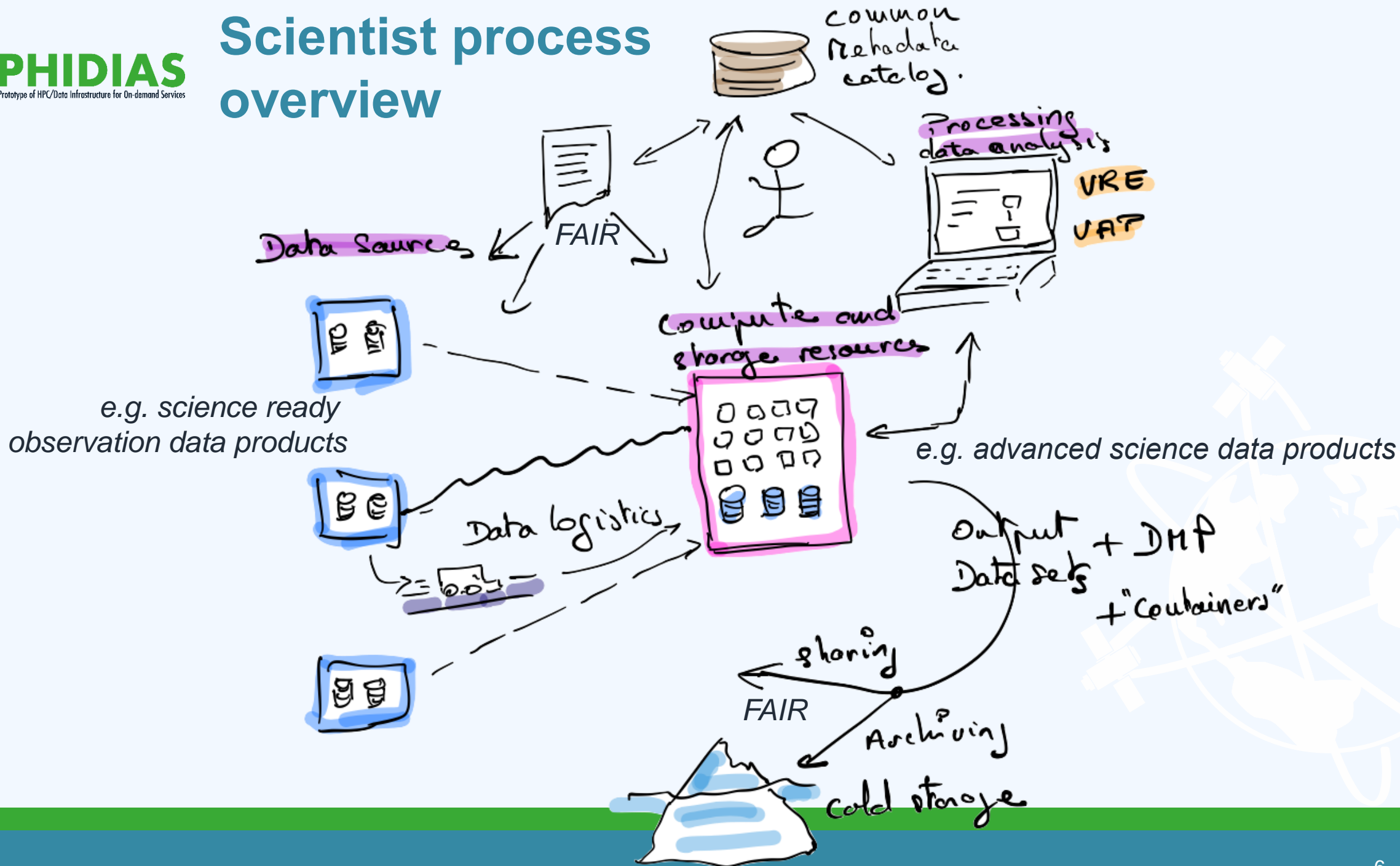
 This is of interest of some partners

 To explore the implementation of Machine Actionable Data/Project Plan

 To understand how to continue PHIDIAS work





Scientist process overview



Data Logistic & Compute Challenges

Distributed infrastructures

-  Volume of data available to a scientist is much larger than what can be hosted in a given data center

 -  HPC center is a data cache


 -  Protocol to update the cache is not the scientist problem

-  Large scientific instrument - important data sources

Exascale systems at least x petascale data size

-  Transferring a Petabyte at 10Gb/s \rightarrow 9 days full speed

Data resource essence differs from compute and storage

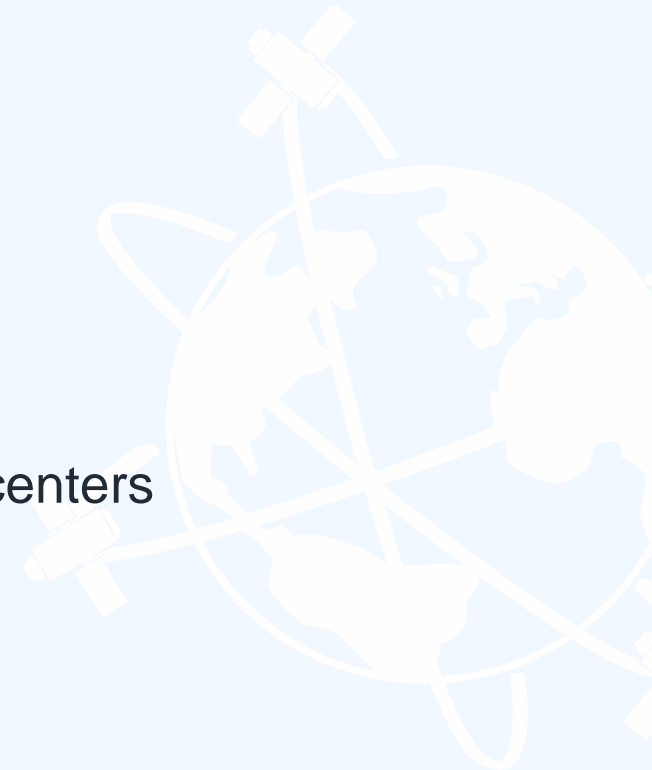
-  Not always local, not a commodity, no substitution

-  Volume of data available to scientist \gg storage capabilities of HPC centers

Data locality must drive compute resource allocation

 -  Most processing is performed remotely where the data is


 -  Efficiency must be evaluated at the global workflow level

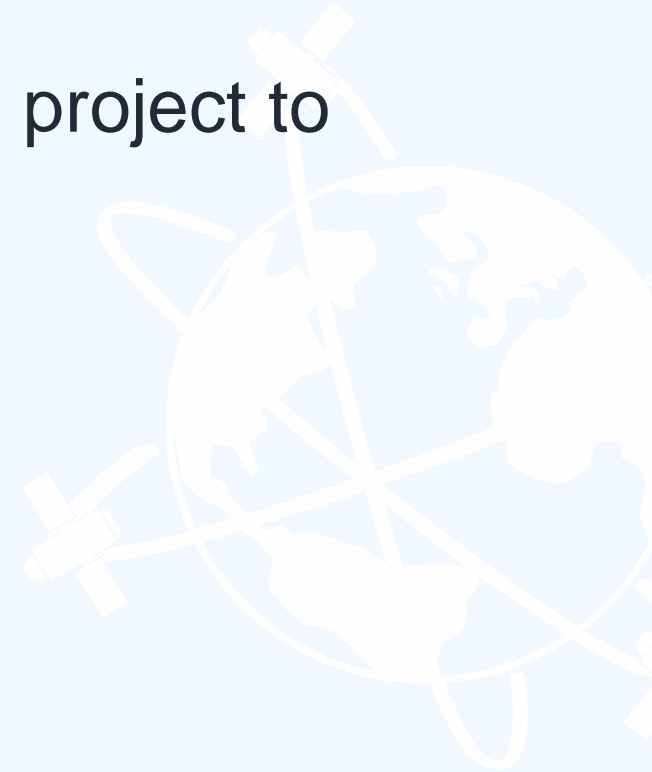


Overcoming Technical & Administrative Constraints

- ❖ ZRR
 - ❖ Cyber-security constraints
- ❖ Resources allocation
- ❖ FAIR requirements
- ❖ Reproducibility
- ❖ Offline capabilities
 - ❖ Oceanographic ships
- ❖ Online in flux processing
 - ❖ Pipelining data transfers and processing (in transit processing)
 - ❖ SKA for instance



 Hopefully the scientific user committee has help the project to make progresses 😊





PHIDIAS

Prototype of HPC/Data Infrastructure for On-demand Services

The End



The PHIDIAS project has received funding from the European Union's Connecting Europe Facility under grant agreement n° INEA/CEF/ICT/A2018/1810854.

Cofinanced by the Connecting Europe Facility