AnaEE
Analysis and Experimentation on Ecosystems
FROM EXPERIMENTATION TO GLOBAL PREDICTION
International conference Paris 2-3 March 2016

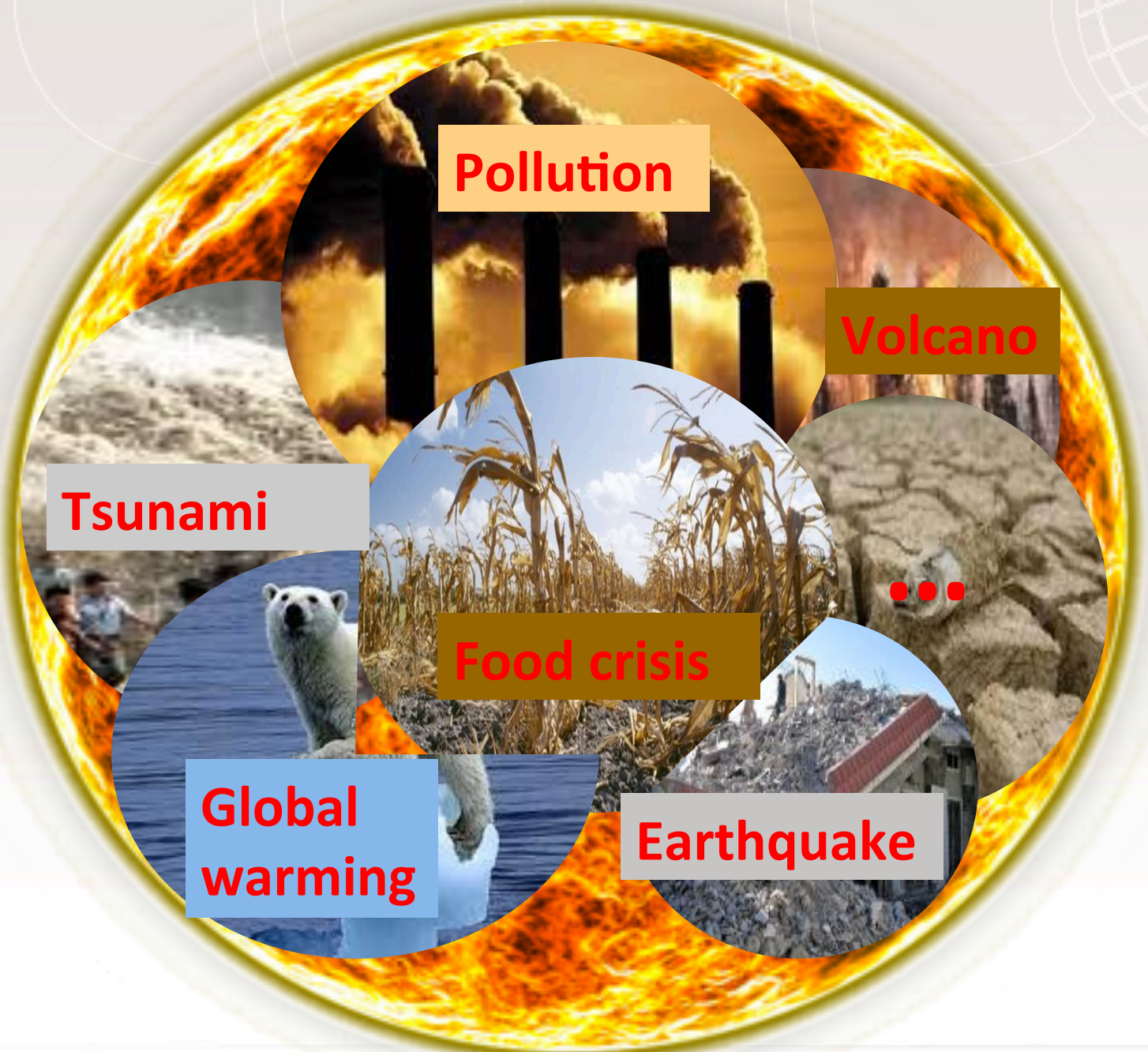
Interoperable research infrastructures for addressing societal challenges

DR. ZHIMING ZHAO

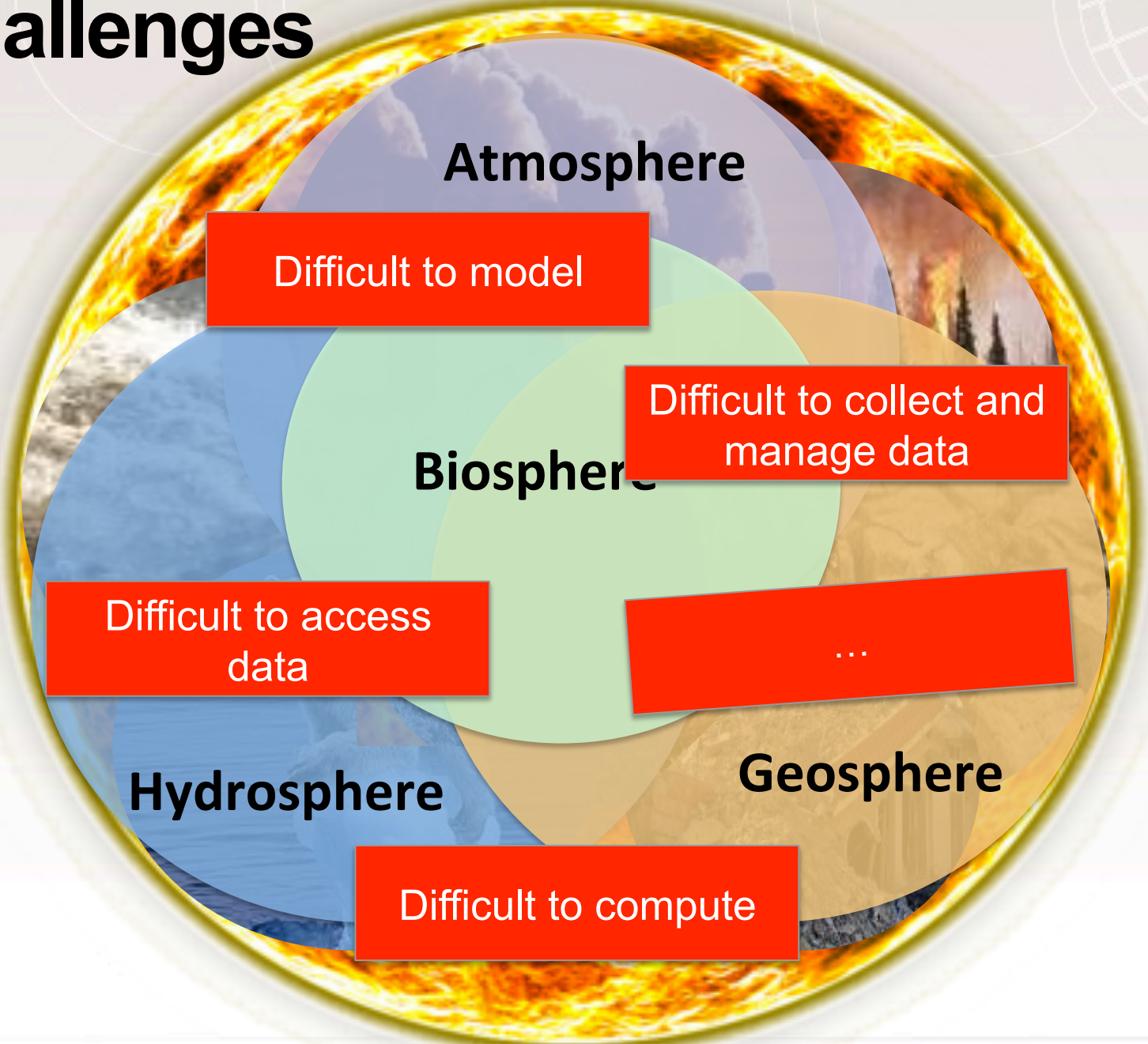
UvA Universiteit van Amsterdam

H2020 Project
SWITCH
plus ENVRI
Environmental Research Infrastructures Providing Shared Solutions for Science and Society
VRE4EIC

Environmental challenges



To understand those environmental challenges

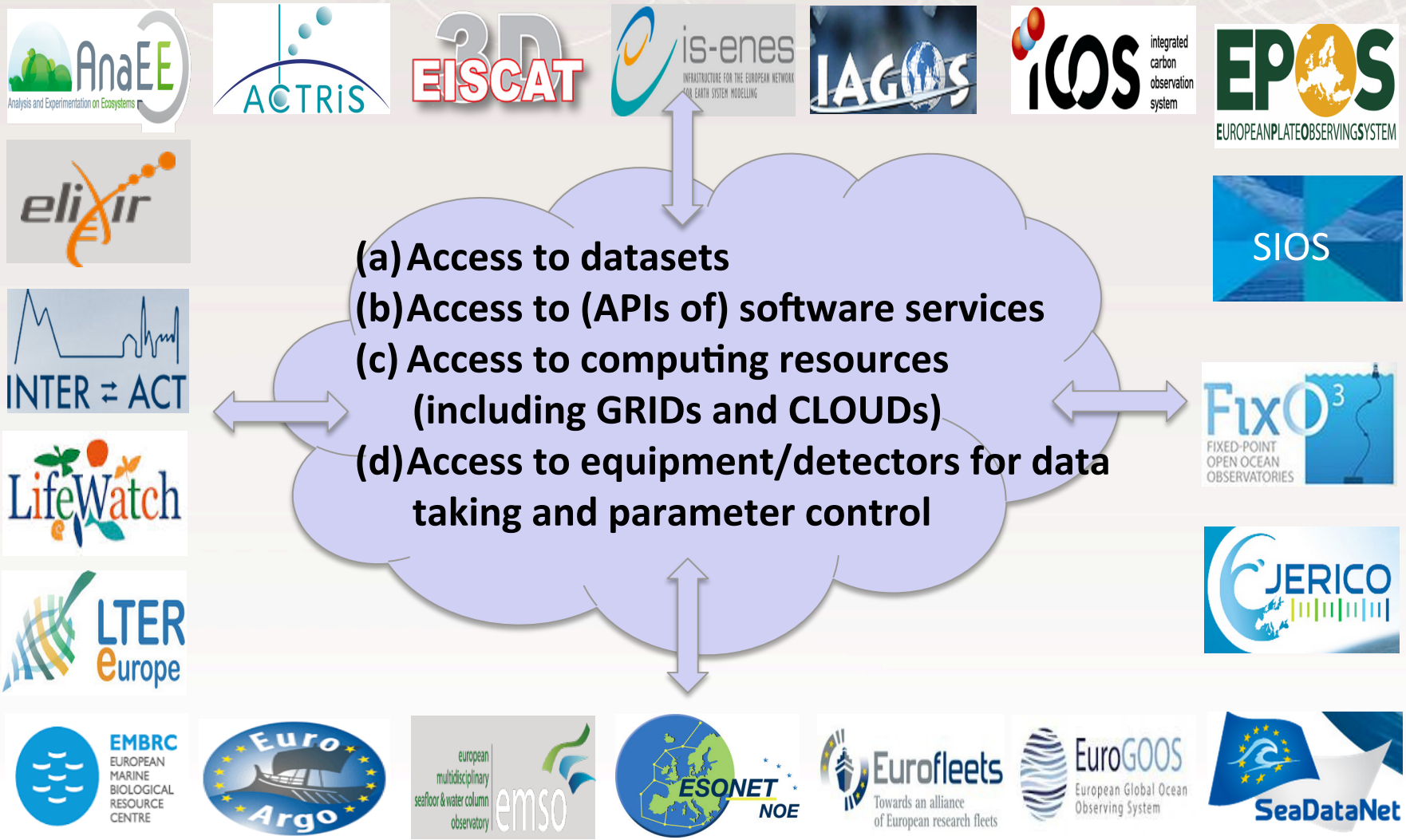


Outline

- Interoperability requirements and reality
- Reference architecture based approach
- Current status
- Summary

Research Infrastructures, I3, and ESFRIs in environmental Sciences

Research Infrastructures, I3, and ESFRIs in environmental Sciences



The reality of RIs

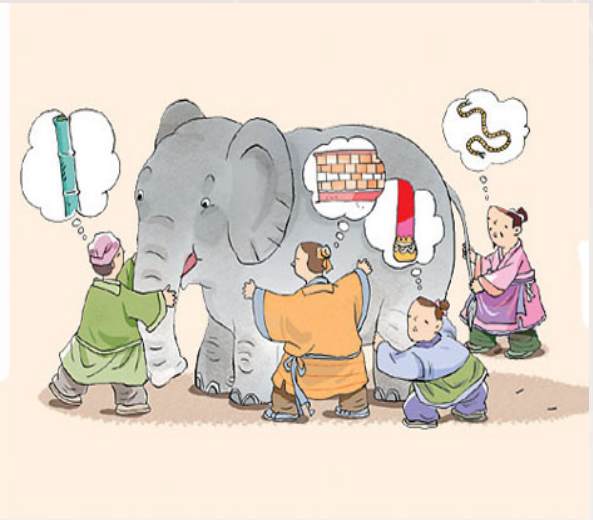
- RIs usually provide access to datasets
  - Many: simple catalogue access with (maybe selection) download to user's home computing environment
  - Few (any?) providing homogeneous view over heterogeneous datasets (none?)
- RIs sometimes provide access to software services/components
  - For download/use in user's home computing environment
- RIs sometimes provide (manual?) workflow construction
- RIs rarely (if ever?) provide access to computing resources
  - With movement of datasets and software components/services and workflow composition
- RIs sometimes provide access to equipment/detectors for data taking and parameter control
- RIs rarely provide integrated e-communication facilities for researchers
  - Scholarly publications, grey literature, blogs, wikis, e-conferencing, social media
- RIs rarely (if ever) provide assistance with research management tasks
  - Proposal preparation, finding related work/researchers, finding specialist equipment, reporting to funders…

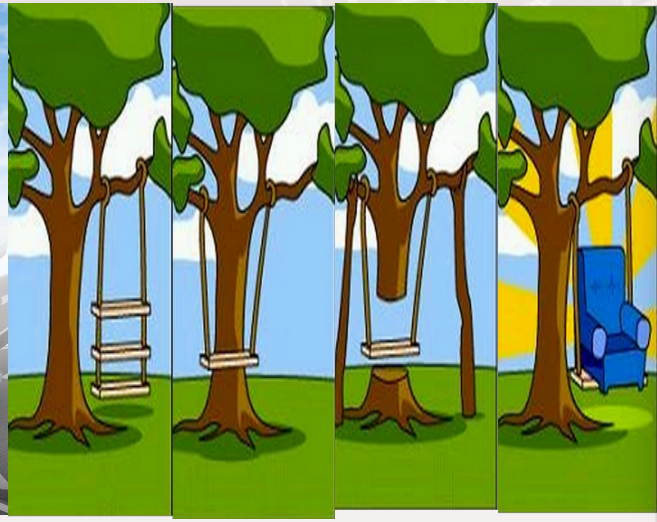
The key is

- RIs usually provide access to datasets
  - Many: simple catalog access with (maybe selection) download to user's home computing environment
  - Few (any?) providing homogeneous view over heterogeneous datasets (none?)
- RIs sometimes provide access to software services/components
  - For download/use in user's home computing environment
- RIs sometimes provide (manual?) workflow construction
- RIs rarely (if ever?) provide access to computing resources
  - With movement of datasets and software components/services and workflow composition
- RIs sometimes provide access to equipment/detectors for data taking and parameter control
- RIs rarely provide integrated e-communication facilities for researchers
  - Scholarly publications, grey literature, blogs, wikis, e-conferencing, social media
- RIs rarely (if ever) provide assistance with research management tasks
  - Proposal preparation, finding related work/researchers, finding specialist equipment, reporting to funders…

RI catalogs have very variable richness

Interoperability challenges



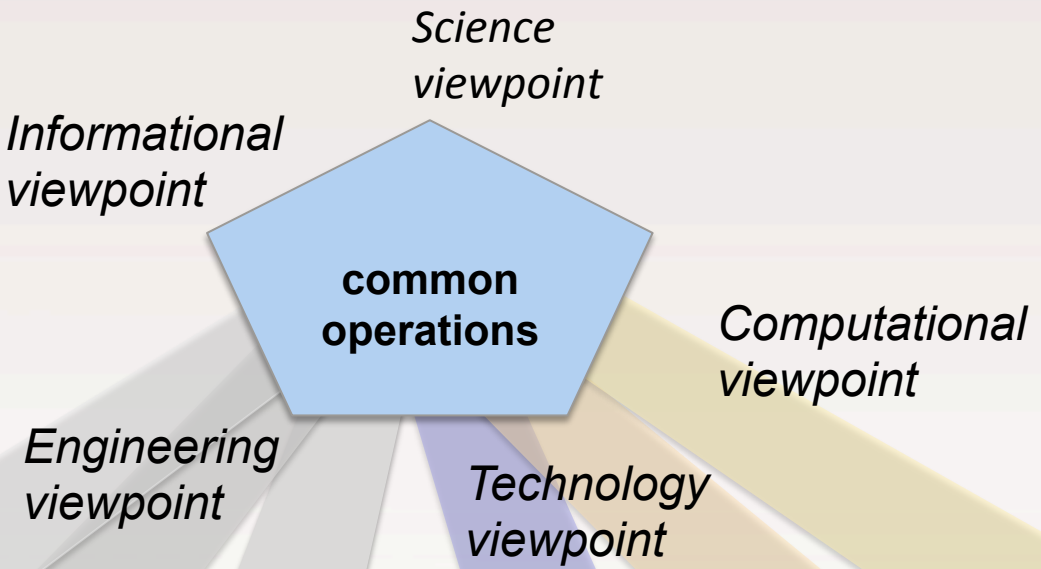


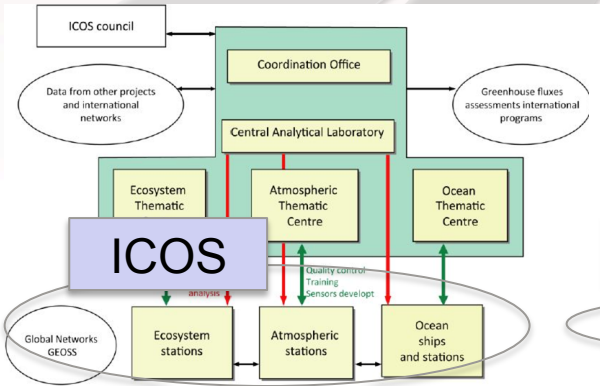
- **Diverse visions, different development agenda, and the lack of common vocabulary** make the data and functional components difficult to be interoperable.
- Approaches in EU FP 7 **ENVRI**, H2020 **ENVRI^{PLUS}**, **VRE4EIC** and **SWITCH**
  1. **Reference model**
  2. **Reference model guided common solution engineering**
  3. **Virtual Research Environment**
  4. **Virtualized infrastructure**

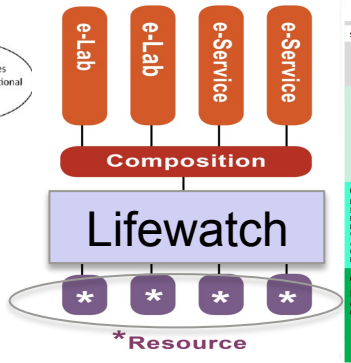
ENVRI Common Operations of Environmental Research Infrastructures

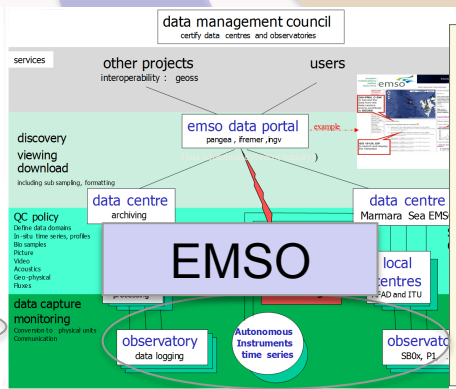
EU FP7 ENVRI project

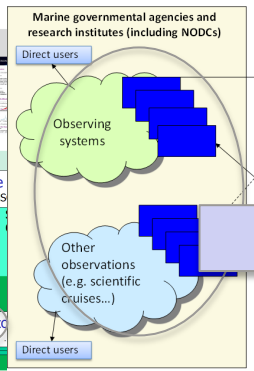
1. Reference model: abstract common characteristics from RIs

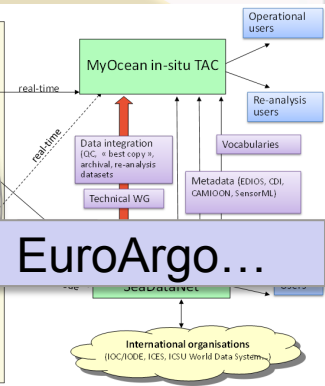












Establish a common set of vocabulary

User community

Environmental RI

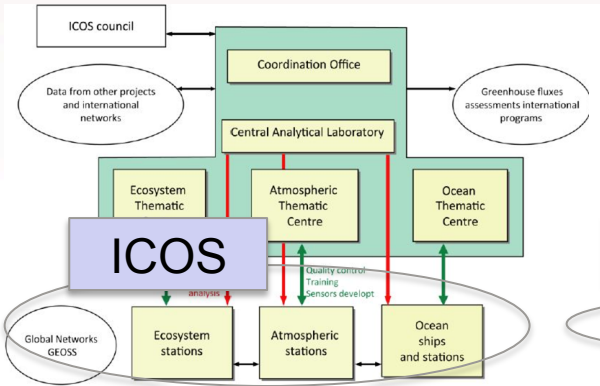
Community support

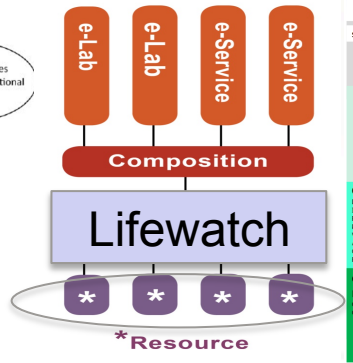
Data acquisition

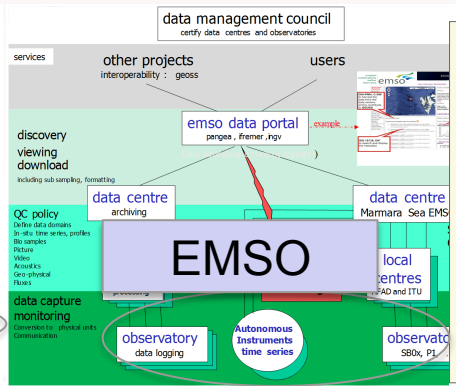
Data Curation

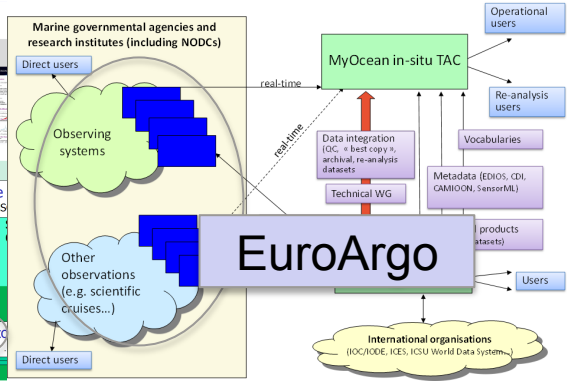
Data access

Data processing









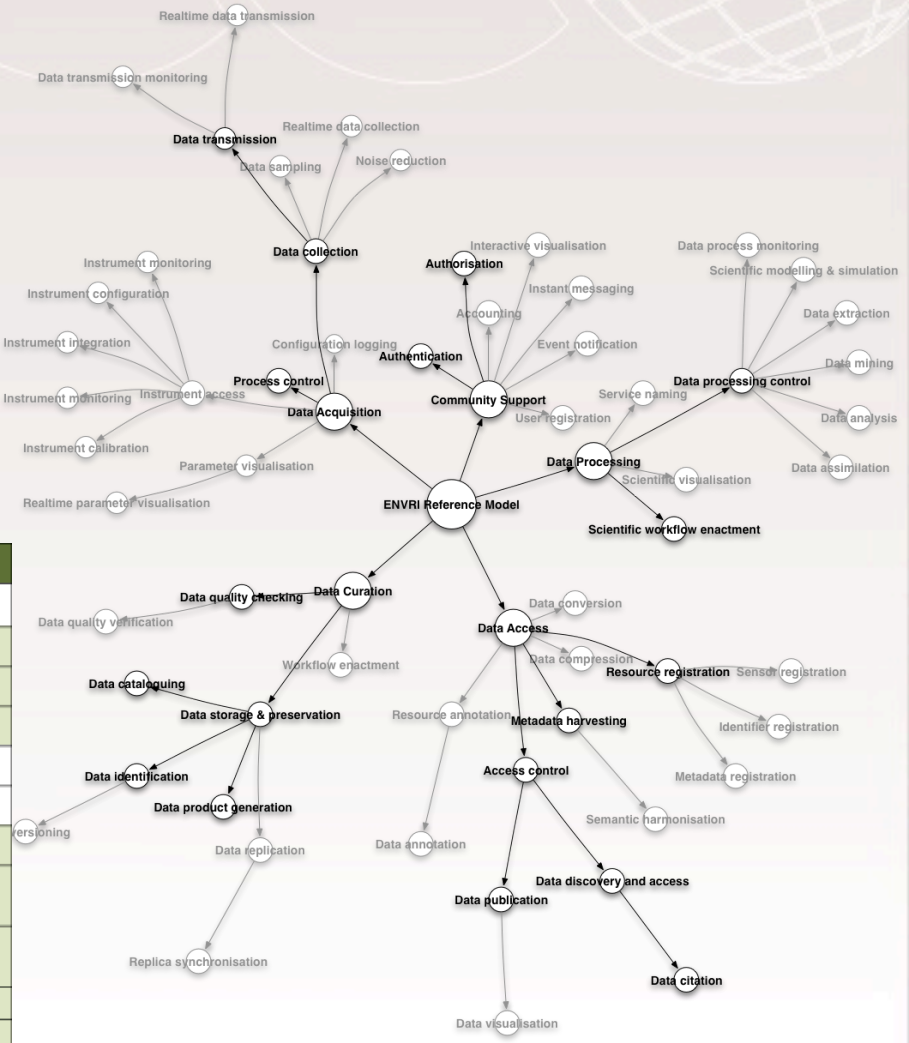
Reference model

ENVRI Common Operations of Environmental Research Infrastructures

Functions/Embedded Services	ICOS	EPOS	EMSO	EISCAT-3D	LifeWatch	Euro-Argo
Access Control	Unknown	Yes	Unknown	Yes	Unknown	Unknown
Data Conversion	Yes	Yes	Yes	Yes	Yes	Yes
Data Compression	No	No	No	No	Yes	No
Data Visualisation	Yes	Yes	Yes	Yes	Yes	Yes
Data Publication	Yes	Unknown	Yes	Unknown	Yes	Yes
Data Citation	No	Unknown	Yes	No	Unknown	No

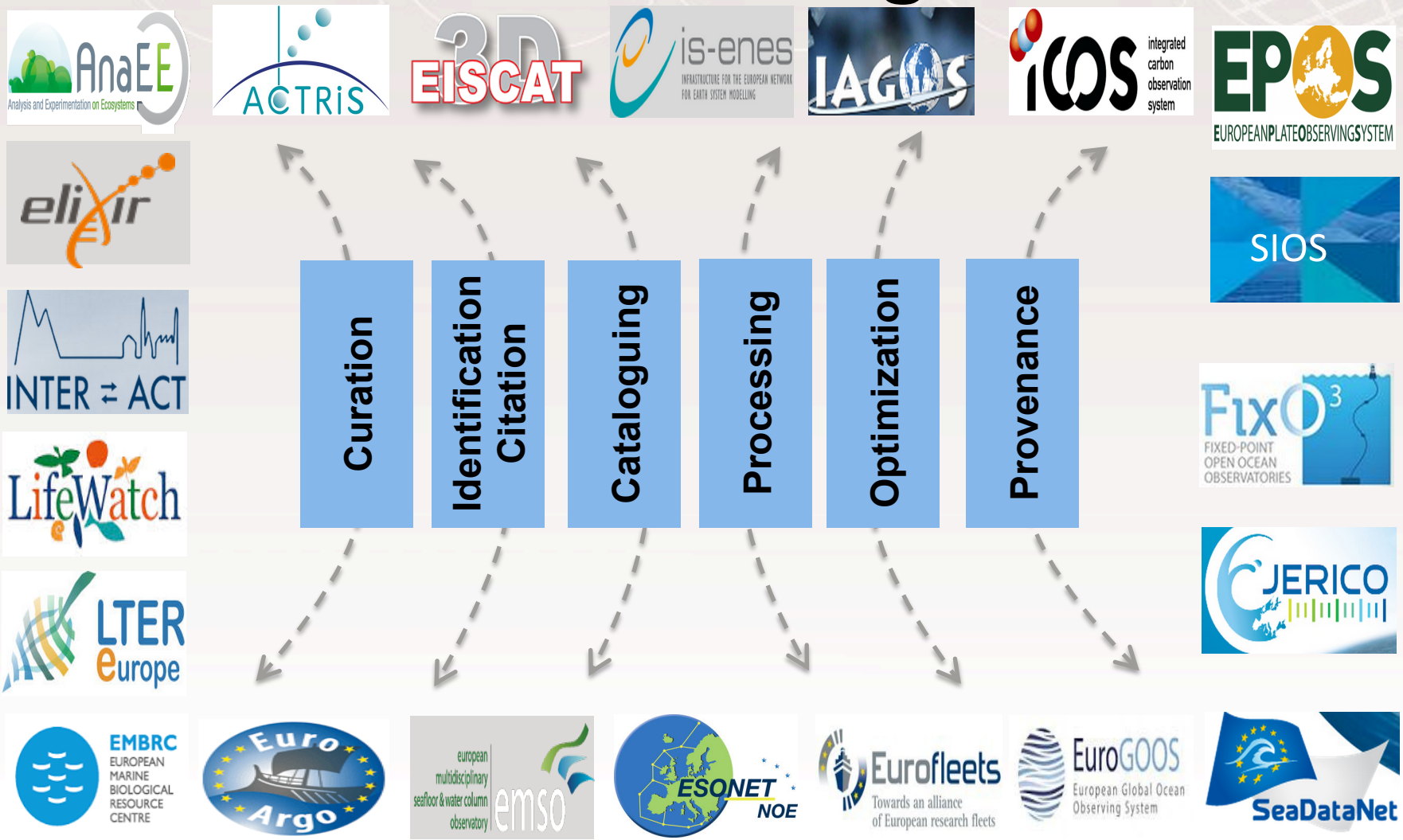
Functions/Embedded Services	ICOS	EPOS	EMSO	EISCAT-3D	LifeWatch	EURO-Argo
Data Quality Checking	Yes	Yes	Unknown	Yes	Not Applicable	Yes
Data Quality Verification	Yes	Unknown	Unknown	Unknown	Not Applicable	Yes
Data Identification	Yes	Yes	Yes	Unknown	Not Applicable	Unknown
Data Cataloguing	Unknown	Yes	Yes	Unknown	Not Applicable	Unknown
Data Product Generation	Yes	Yes	Yes	Yes	Not Applicable	Yes
Data Versioning	Yes	Unknown	Unknown	Unknown	Not Applicable	Unknown

Functions/Embedded Services	ICOS	EPOS	EMSO	EISCAT-3D	LifeWatch	Euro-Argo
Access Control	Unknown	Yes	Unknown	Yes	Unknown	Unknown
Data Conversion	Yes	Yes	Yes	Yes	Yes	Yes
Data Compression	No	No	No	No	Yes	No
Data Visualisation	Yes	Yes	Yes	Yes	Yes	Yes
Data Publication	Yes	Unknown	Yes	Unknown	Yes	Yes
Data Citation	No	Unknown	Yes	No	Unknown	No
(Resources/Data) Annotation	Yes	Yes	Yes	No	Yes	Yes
Metadata Harvesting	Unknown	Unknown	Yes	No	Unknown	No
Resource Registration	Unknown	Yes	Yes	No	Yes	No
Semantic Harmonisation	No	Yes	Yes	No	Yes	No
Data Discovery and Access	Yes	Yes	Yes	Yes	Yes	Unknown



- A full function list is on ENVRI wiki: http://envri.eu/rm

2. Common data challenges



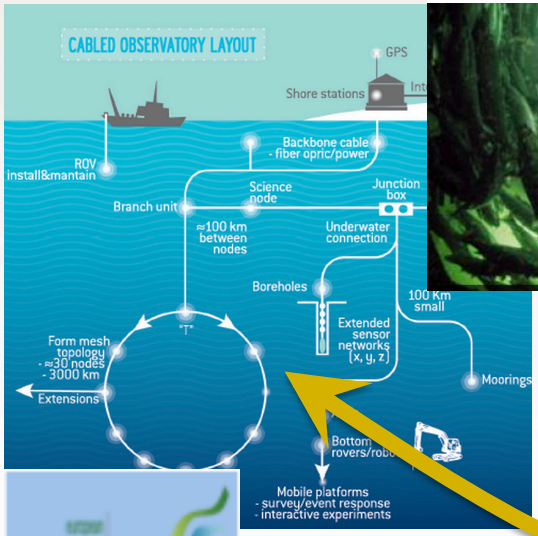
Data identification and citation

DOI

Marine biology

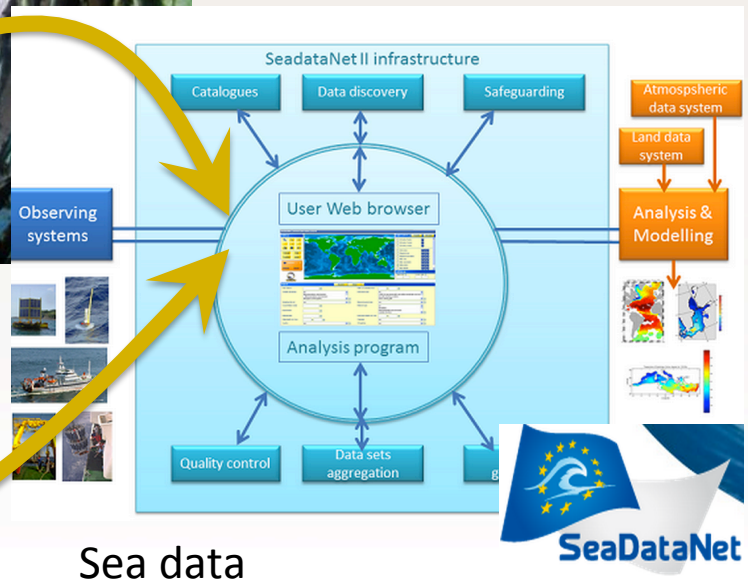
Bio-diversity

DOI



Marine and Seafloor data

DOI

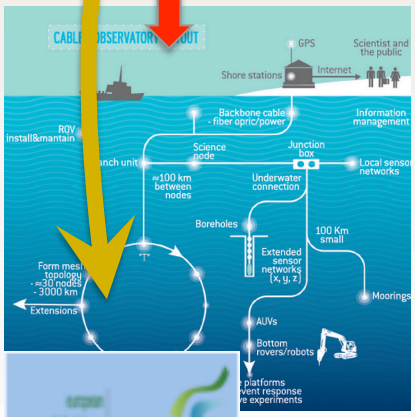


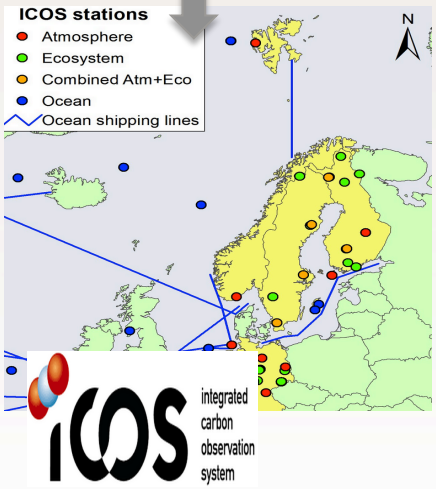
Sea data

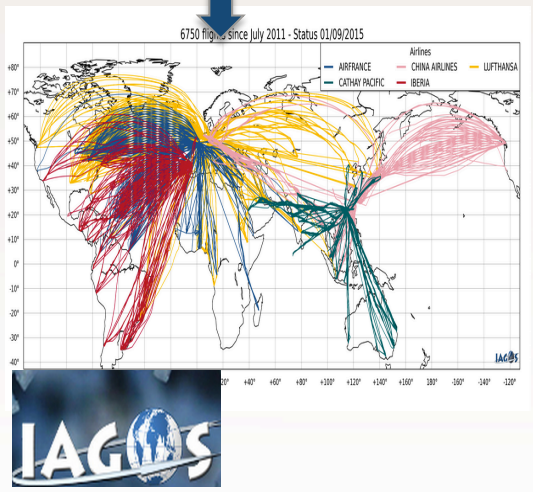
DOI

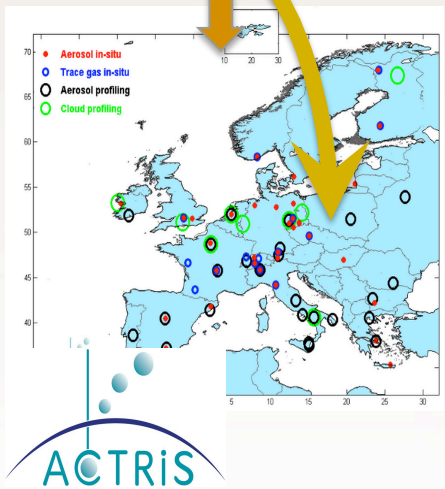
Data cataloguing

Interoperable cataloguing?









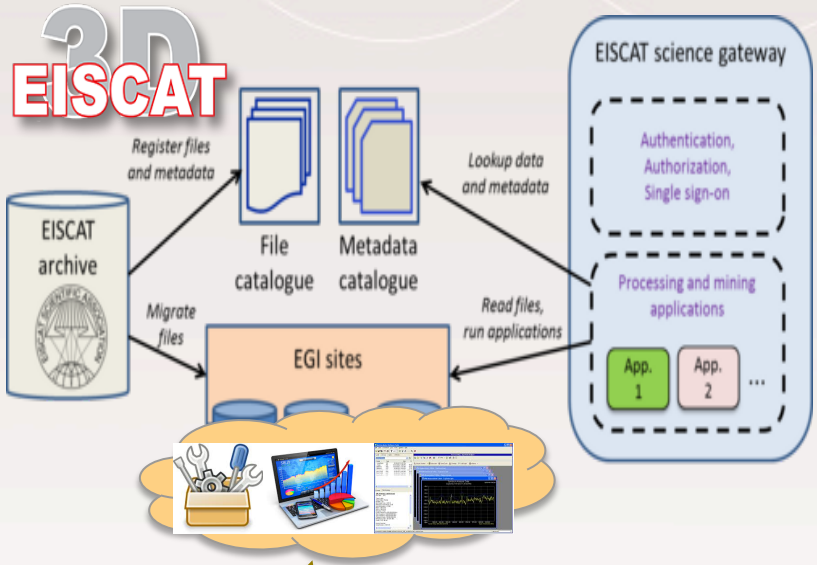
Marine and Seafloor data

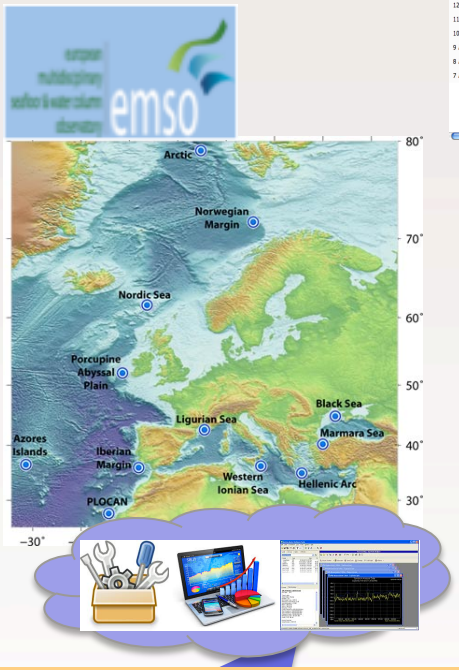
Carbon, ocean, eco-sys

Cloud particles, atmosphere composition

Gas-sphere spices, etc.

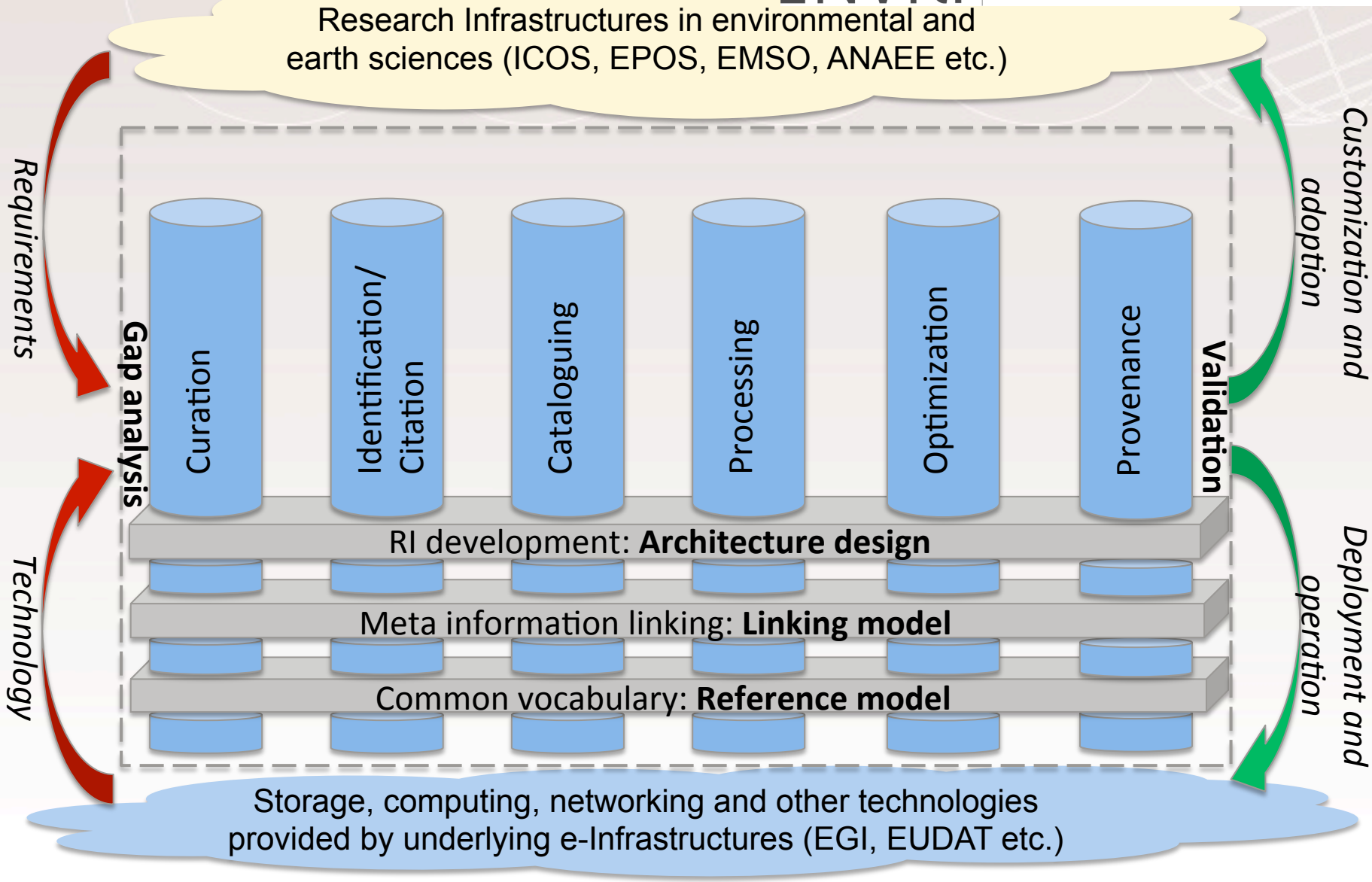
Processing, monitoring, diagnosis and optimization





How to optimize: data discovery, access, delivery and processing.

Tools: processing, monitoring, and diagnosis, virtual research environments



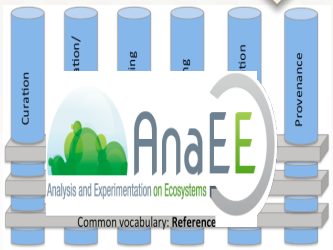
Zhao, Z., et al. (2015) Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures, 1st IT4RIs in IEEE eScience

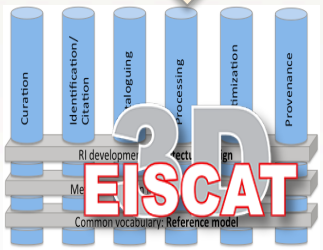
3. From research infrastructures to Virtual Research Environment

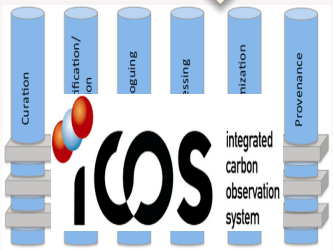
Virtual Research Environment

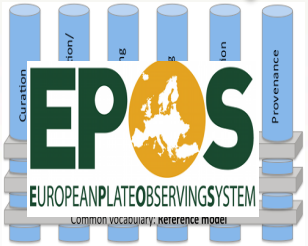
- AAAI for managing access
- Cataloguing, metadata for inter-RI resource integration
- Workflow management for managing experiments

Interface









...

Storage, computing, networking and other technologies provided by underlying e-Infrastructures (EGI, EUDAT etc.)

From research infrastructure to Virtual Research Environment

(a) AAAI systems
(b) Catalog of users, datasets, software services, resources (computing, equipment/detectors), services with associated information on rights, costs

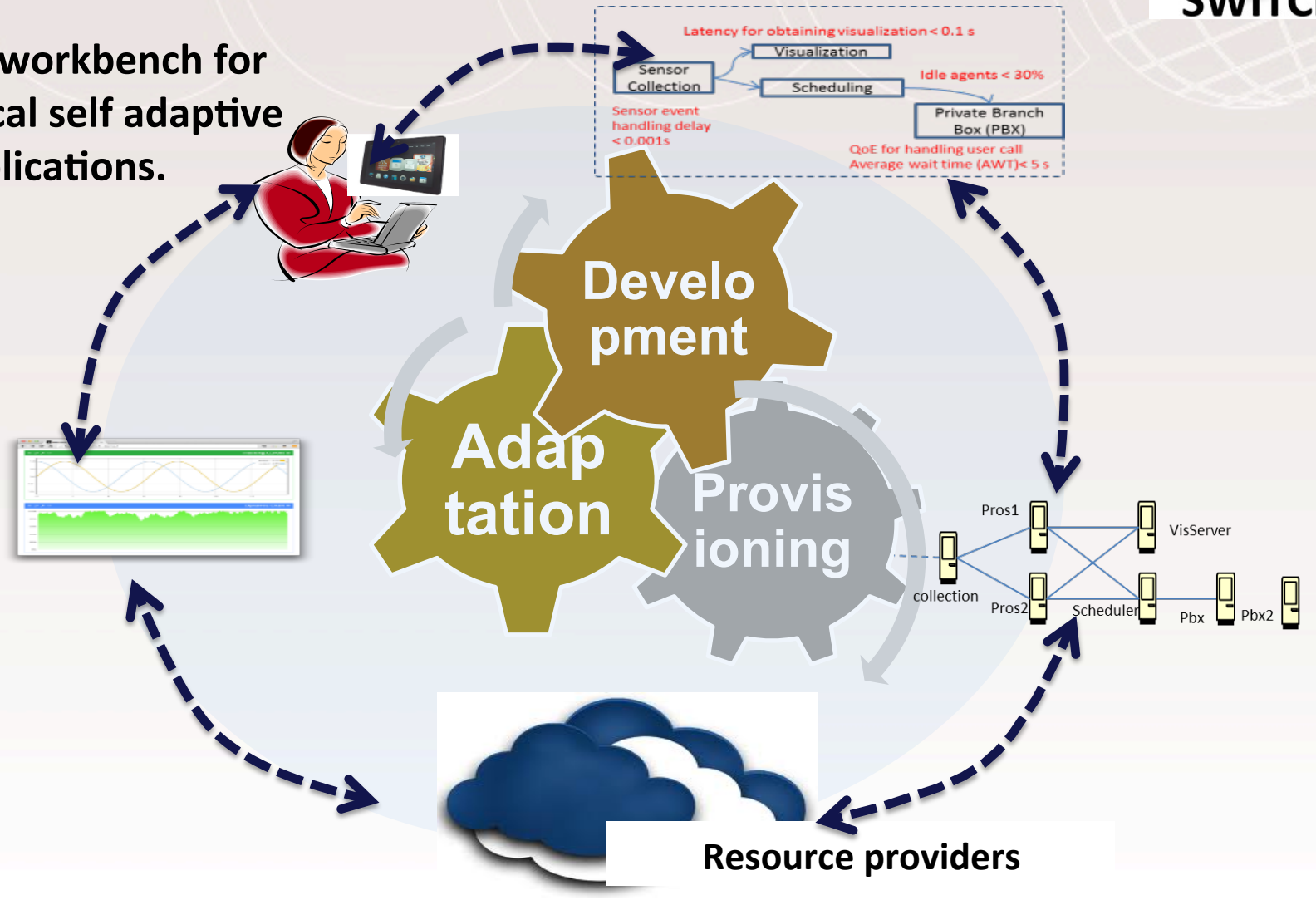
And from individual RI……

(a) Access to datasets
(b) Acces to (APIs of) software services
(c) Access to computing resources (including GRIDs and CLOUDs)
(d) Access to equipment/detectors for data taking and parameter control

Storage, computing, networking and other technologies provided by underlying e-Infrastructures (EGI, EUDAT etc.)

4. Quality critical data processing on virtualized infrastructures

Software workbench for time critical self adaptive cloud applications.



Zhao, Z., et al.,(2015), A software workbench for interactive, time critical and highly self-adaptive cloud applications (SWITCH), Cluster, Cloud and Grid Computing)

Use cases: implementation/test/ business cases

Implementation cases

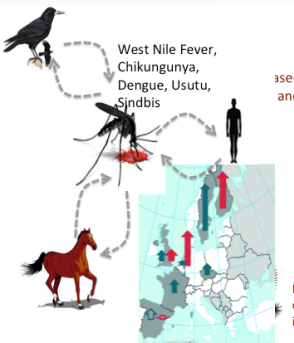
- Small but concrete
- Clear objectives on functional components from specific sub-system
- Fixed time line

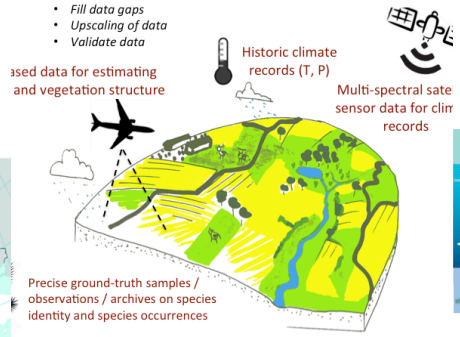
Test cases

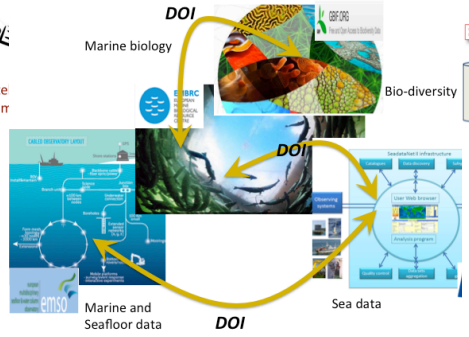
- Extend implementation cases with cross sub-system functionality
- Demonstrate SWITCH functionality of all sub systems based on SME cases

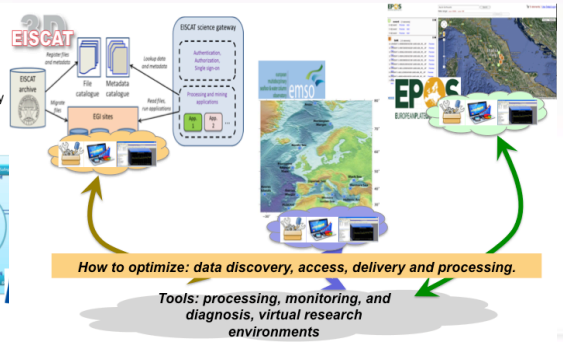
Business cases

- Business stories based on the extensions of test cases
- Stories of challenges, market potential and business models of time critical cloud applications









...

Use case and topics

	RM	Semantic Linking	Ident./ Citation	Curation	Catalog.	Pocessing	Optim.	Provenance	Others
SC-3		Y	Y			Y	Y		Y/ trainning
TC-2					Y	Y	Y		
TC-4				Y	Y				
TC-16	Y	Y							
IC-2		Y						Y	
IC-3						Y	Y		
IC-9		Y						Y	
IC-10		Y							
IC-11		Y							
IC-12	Y	Y							

Current status

- ENVRI Reference model
  - Available online www.envri.eu/rm
  - Ontology: OIL-e
- Common operation development
  - www.envriplus.eu
  - 2015-2019
- VRE4EIC
  - www.vre4eic.eu
  - Initial prototype in 2018
- SWITCH workbench
  - www.switchproject.eu
  - 2015-2018
  - Version 1, 2016

Summary

- Discussed challenges and approaches in interoperable infrastructures
  - Reference architecture
  - Shared solutions to common problems
  - Virtual Research Environment
  - Virtualized infrastructure
- Share solutions to common problems increate the reusability of the software components and increase the system interoperability
- A proper interface between Research Infrastructures and **Virtual Research Environments** promotes the resource access and workflow execution across different RI boundaries
- Virtualized infrastructures provide elastic resources for quality critical applications

References

- EU FP7 ENVRI www.envri.eu
- EU H2020 SWITCH www.switchproject.eu
- EU H2020 ENVRI^{PLUS} www.envriplus.eu
- EU H2020 VRE4EIC www.vre4eic.eu
- Martin P., et al, Zhao Z. (2016) Information modelling and semantic linking for a software workbench for interactive, time critical and self-adaptive cloud applications, (CCPI-2016)
- Mork, R., Martin, P., Zhao, Z. (2015) Contemporary Challenges for Data-intensive Scientific Workflow Management Systems, WORKS 2015
- Zhao, Z., et al., (2015), Developing and operating time critical applications in clouds: the state of the art and the SWITCH approach, Cloud Forward 2015
- Jeferry, K., et al., Zhao, Z. (2015) Challenges emerging from future cloud application scenarios, CloudForward 2015
- Zhao, Z., et. al., (2015) Reference Model Guided System Design and Implementation for Interoperable Environmental Research Infrastructures, (IT4RIs 2015)
- Martin, P., et al., Zhao, Z. (2015) Open Information Linking for Environmental Research Infrastructures, (IT4RIs 2015)