



Progress with the **EGI-LifeWatch** Competence Center

WP6 Knowledge Commons EGI-Engage Review



www.egi.eu



Presented by
Jesús Marco de Lucas
IFCA-CSIC, Spain
for EGI LW CC team



unded by the Horizon 2020 Framework Programme
of the European Union under grant number 654142



What is LifeWatch?

- LifeWatch is an **e-science** and technology infrastructure for **biodiversity and ecosystem research** to support the scientific community **and other users**.
- It is putting in place the infrastructure and information systems necessary to provide an analytical platform for the **modeling and simulation** of both existing and new data on biodiversity to enhance the knowledge of biodiversity functioning and management
- Example of relevant case studies:
 - Invasive species
 - Evolution of wetlands
 - Evaluating the ecological quality of habitats

LifeWatch in the 2016 ESFRI Roadmap

ESFRI ROADMAP 2016			
PART 1	PART 2	PART 3	ANNEXES

BACK

http://www.esfri.eu/esfri_roadmap2016/roadmap-2016.php

ESFRI LANDMARKS								
NAME	FULL NAME	ROADMAP ENTRY (YEAR)	OPERATION (YEAR)	LEGAL STATUS (AS OF 10 MARCH 2016)	CAPITAL VALUE (M€)	OPERATIONAL ANNUAL BUDGET (M€/YEAR)		
JHR	Jules Horowitz Reactor	2006	2020*		1,000	NA	ENERGY	
EMSO	European Multidisciplinary Seafloor and water-column Observatory	2006	2016	ERIC under preparation	108	36		
EURO-ARGO ERIC	European contribution to the international Argo Programme	2006	2014	ERIC, 2014	10	8	ENVIRONMENT	
IAGOS	In-service Aircraft for a Global Observing System	2006	2014	AISBL, 2014	25	6		
ICOS ERIC	Integrated Carbon Observation System	2006	2016	ERIC, 2015	48	24-35		
LifeWatch	e-Infrastructure for Biodiversity and Ecosystem Research	2006	2016	ERIC under preparation	66	10		
BBMRI ERIC	Biobanking and BioMolecular resources Research	2006	2014	ERIC, 2013	170-200	3,5		



An e-Infrastructure to support research for the protection, management and sustainable use of biodiversity

TYPE: distributed
COORDINATING COUNTRY: ES
PROSPECTIVE MEMBER COUNTRIES:
BE, EL, ES, IT, NL, PT, RO

PARTICIPANTS: FI, FR, HU, NO, SE, SI, SK

TIMELINE

- ESFRI Roadmap entry: 2006
- Preparation phase: 2008-2011
- Construction phase: 2011-2016
- Operation start: 2016

ESTIMATED COSTS

- Capital value: 66 M€
- Operation: 10 M€/year

HEADQUARTERS

Statutory Seat: ES
Common facilities: ES-IT-NL

WEBSITE

<http://www.lifewatch.eu>



LifeWatch

e-infrastructure for Biodiversity and Ecosystem Research



Description

The e-infrastructure for Biodiversity and Ecosystem Research (LifeWatch) is a distributed RI to advance biodiversity research and to address the big environmental challenges and support knowledge-based strategic solutions to environmental preservation. This mission is achieved by providing access to a multitude of data sets, services and tools enabling the construction and operation of Virtual Research Environments.

Activity

LifeWatch is an e-Infrastructure of distributed nature, composed by Common Facilities and other Distributed LifeWatch Centres. Common Facilities are located in Spain (Statutory Seat and the ICT e-Infrastructure Technical Offices), Italy (Service Centre) and The Netherlands (Virtual Laboratories and Innovations Centre).

The Statutory Seat and the ICT e-Infrastructure Technical Offices will jointly assist to the coordination and management of the day-to-day institutional relationships, administrative, legal, and financial issues. Those include, among others, technology transfer, procurement and IPR matters, and the formal agreements with all the external data and e-Services suppliers, and the Service Legal Agreements (SLA) with local, regional, national and international entities, including decision makers and environmental managers. Also, they will coordinate and manage the ICT e-Infrastructure distributed construction, maintenance and deployment operations, including coordination of the design and implementation of e-Services demanded by the Service Centre, the Virtual Laboratories and Innovations Centre,



EGI- LifeWatch Competence Center



E-Science European Infrastructure for Biodiversity and Ecosystem Research

European Grid Infrastructure



EGI-LifeWatch Competence Centre

Call for Competence Centres
for inclusion in the EGI-Engage proposal, Call 3, EINFRA-1, Activity 6

Mail to: cc-call@mailman.egi.eu

Deadline for submission: 04 July, h 24:00 CEST

The proposal prepared in July 2014 included:

- A support task from NGIs (ES,PT,IT)
- Two lighthouse projects (24M):
 - Big Data and Ecological Observatories
 - **Supporting Workflows & Virtual Labs in FedCloud for LifeWatch**
- A path finding project (12M):
 - Advanced Support to Citizen Science in Biodiversity

#	Participant	Role in the CC
1	JRU-NGI-ES	Service Provider
2	JRU-LW-ES	Service Provider/User Community
3	NGI-PT (LIP)	Service Provider
4	NGI-FR (CNRS, INRA)	Service Provider/User community
5	NGI-IT (INFN)	Service Provider/User community
6	VLIZ, Belgium	User Community
7	CIBIO, Portugal	User Community

90 PM requested, EGI-Engage funds 59 PM

LIFE-WATCH related initiatives complement in what possible

EGI LifeWatch CC acting as a key technical collaboration forum!

Participation of more LifeWatch partners (not formal partners in EGI-Engage): LifeWatch Spain JRU, LifeWatch Greece team at HCMR, LifeWatch Italy team at UniSalento, LifeWatch Belgium at U.Lovaine, LifeWatch NL at UvA...

Assigned to SA2.7:

- ✓ **D6.1:** Assisted pattern recognition tools integrated with EGI for citizen science (OTHER, M09)
- ✓ **D6.6** Data flow handler and basic R tools to integrate and process data from Ecological Observatories on EGI (DEM, M12)
- D6.18** Report on the installed LifeWatch applications and their usage record (R, M24)

Related to SA2.1 Training

- M6.1** Joint training program for the first period is agreed M03
- M6.5** Joint training program for the sec. period is agreed M15

LifeWatch Competence Center Tasks

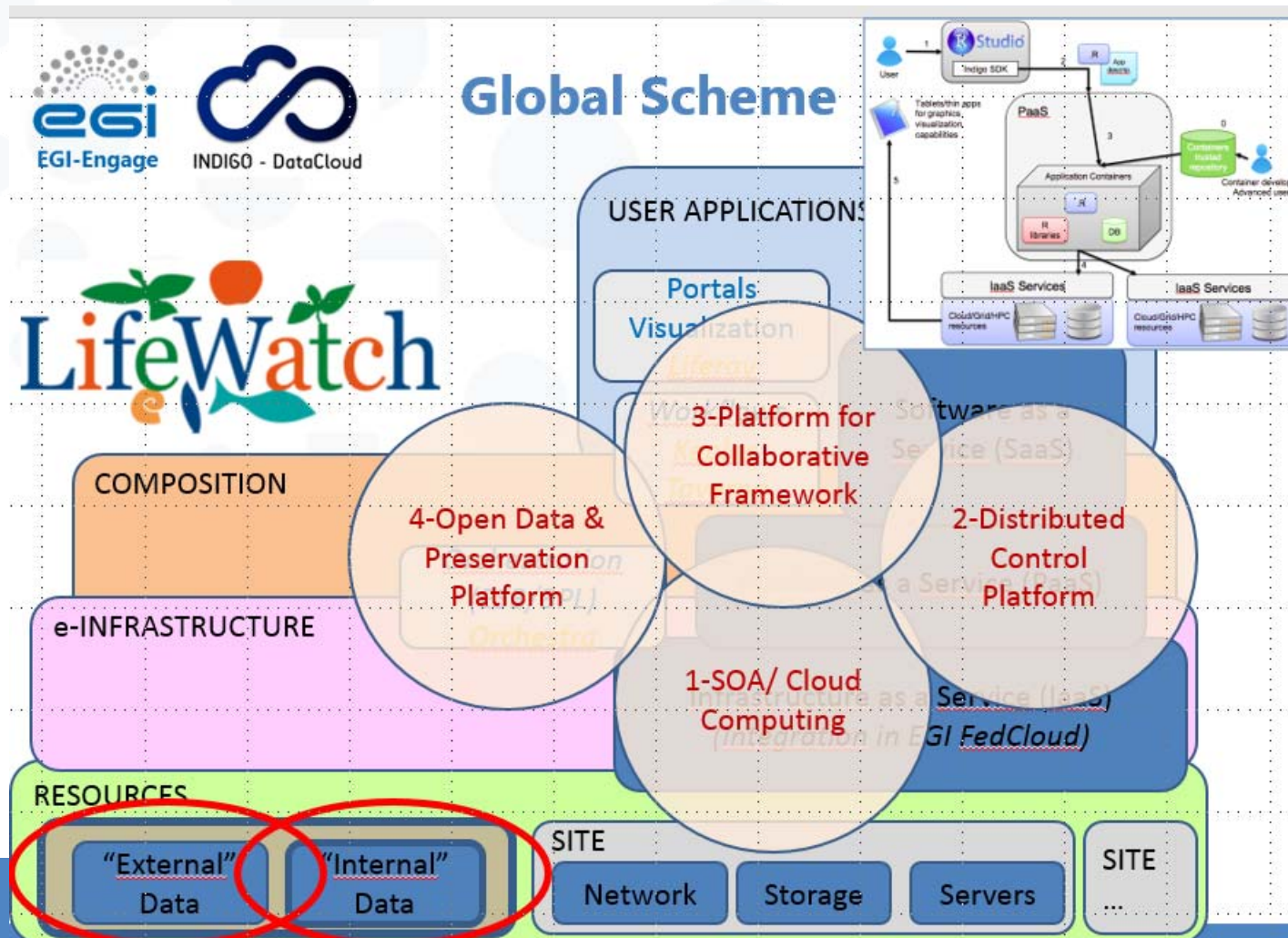
TASK SA2.7 LifeWatch (Lead partner: IFCA, M1 – M30)

- The goal of the LifeWatch EGI CC is to capture and address the requirements of Biodiversity and Ecosystems research communities.
- To achieve this, the Competence Center will
 - **deploy** cloud and gpgpu based **e-Infrastructure** services required to support data management, data processing and modelling for Ecological Observatories,
 - facilitate the adoption and exploitation of the EGI infrastructure by the LifeWatch user community.
 - explore possibilities to increase the participation of citizens in data-intensive biodiversity research,


Exploiting FedCloud


- **FedCloud adopted in LW CC as the basis (IaaS) for supporting the different services and applications**
- **Have we made the RIGHT SELECTION?**
- **We aim to integrate under an Open Science Cloud**
- **LifeWatch VO supported on FedCloud resources**
- **but... the cloud world is not “easy”**
 - **A LW PILOT PROJECT IN SEVILLE HAS SHOWN MANY OF THE POSSIBILITIES, BUT ALSO THE CHALLENGES!**
 - **We are collaborating directly with FedCloud team, and with Data Commons team within EGI-ENGAGE**
 - **We profit of the collaboration with INDIGO-DATACLOUD**
 - **With participation of several LW CC partners**

- Architecture followed in the pilot project



- Open Science Framework in the pilot project

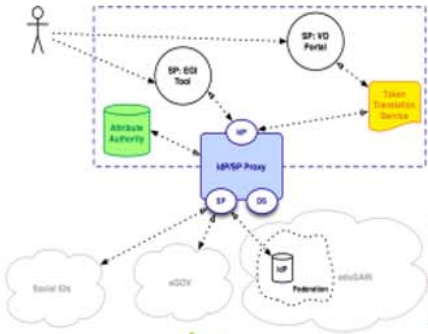

Open Science Framework

[SEARCH](#)
[COMMUNITIES](#)
[PROJECTS](#)
[DEPOSIT](#)
[ANALYZE](#)
[ADMIN](#)


New AAI services

From X.509 certificates to a multiple identity tokens

- Users will access EGI services with their Home Organisation credentials, which will be mapped to one persistent EGI unique identifier
- Different levels of Assurance
- Token Translation Services to convert users' credentials:
 - Online CA, PUSPs, etc.
- Pilot implementation ready by Q2 2016



ORCID

[Edit](#)

[Plan](#)

Plan

Create Data Management Plans for your data and upload the resulting document in the framework: describe the processes and resources for the entire data life-cycle.

DMP records

Data Management Plan for Cuerda del Pozo 2015-11-05 Data Management Plan for Cuerda del Pozo

PID [lifewatch.openscience/1](#) [open](#) [dmp](#) [archived](#)

[Create DMP](#)

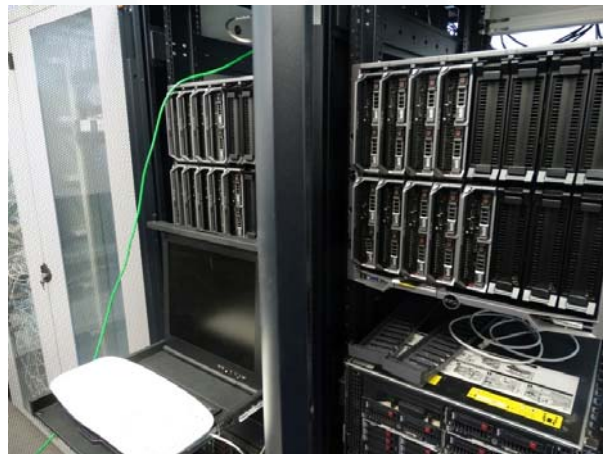
[Register DMP file](#)

[Publish](#)

[DATA LIFE CYCLE](#)

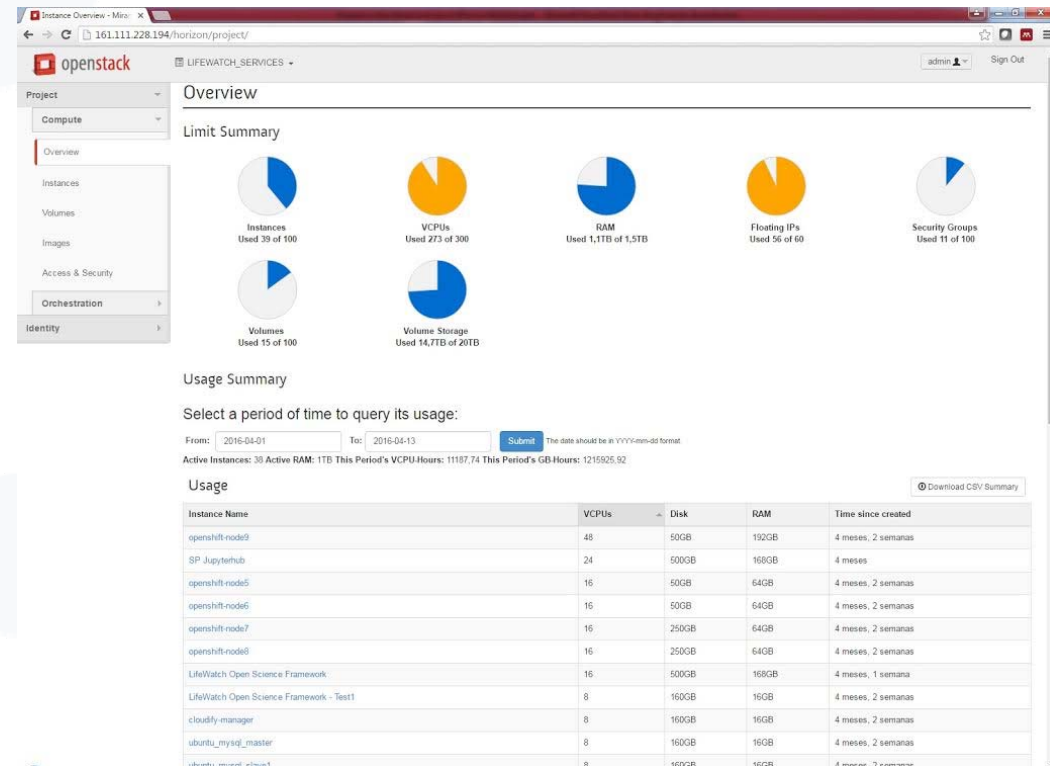
Exploiting FedCloud

- **Integrating resources in the dedicated pilot project**
 - New servers (around 1000 cores, including large RAM, GPU, etc.)
 - New storage (1 + 1 PB)
 - Dark fiber network connection
- **Working to define “service levels”**
 - FitSM
 - SLA



Exploiting FedCloud

- **Openstack view in LW pilot in Seville**
- **Other FedCloud resources used (supporting LW VO):**
 - IFCA, CESGA, BIFI
 - LIP
 - IISAS
 - INFN
- **Applications in production deployed on FedCloud resources:**
 - VLIZ/Marine VRE
 - GBIF Spain
- **OpenProject (development)**
 - 7 Working groups
 - >30 Registered Developers



- **Data flow from observatories**
 - **Marine Observatories**
 - **Water Reservoir (contribution to LIFE+ project ROEM+)**
- **Data processing and workflows:**
 - **R and python**
 - **Galaxy (elastic clusters) and TRUFA (pipeline) in the FedCloud**
 - **Python based workflows**
- **Support to Citizen Science:**
 - **Support to Natusfera**
 - **Deep NN using GPUs and assisted image recognition (Bari' demo)**
 - **Outreach events!**
- **Integration of Preservation framework under Data Commons**
- **Lessons learnt on Requirements, OpenProject, and Working Groups**

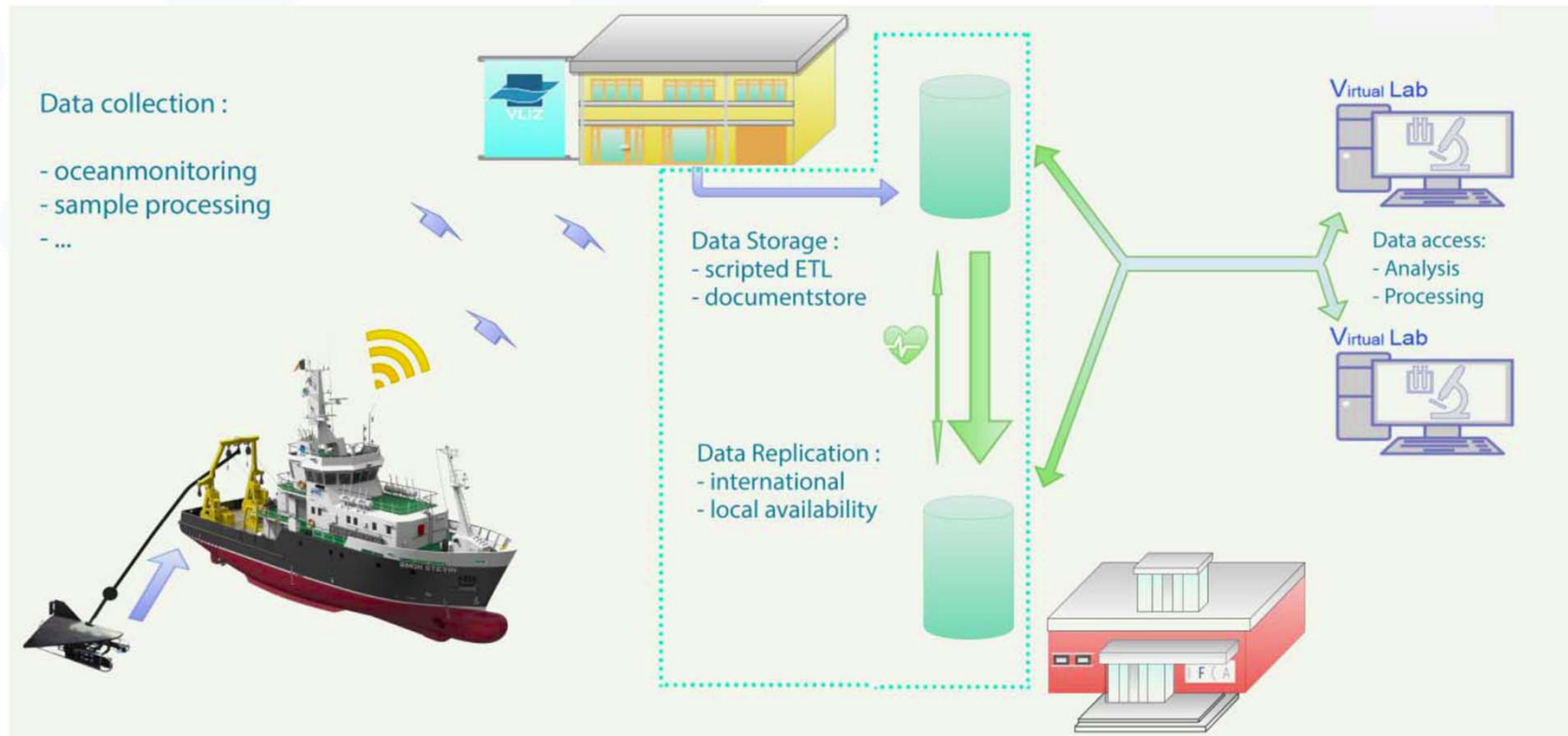
Ecological Observatories and R tools

Two Ecological Observatories provide data into FedCloud via LW-EGI-CC:

- Flanders Marine Institute (VLIZ) has installed a number **of biosensors on board of the Research Vessel Simon Stevin**, as part of the Flanders Marine LifeWatch Observatory, providing a data flow that will reach about 50Tb of data per year, **mainly video and images, collected by the vessel in quasi real time** and requiring a substantial computational power, to incorporate a framework based in R for the final researcher.
- IFCA and a Spanish SME (Ecohydros SL) have been operating for the last five years **an advanced monitoring platform in a water reservoir to detect cyano-algae blooms, that is providing a continuous data flow** and requires also the integration of external data into EGI FedCloud, used by the SME researchers **to contrast the modelling tools**. R is used systematically to provide to the online monitoring with the computation of relevant quantities like the vertical temperature profile parameters evolution (epilimnion /hypolimnion parameters among many others)

Ecological Observatories and R tools

Marine Data Stream



Data Flow in the Case Study of the marine observatory managed by VLIZ center

Ecological Observatories and R tools

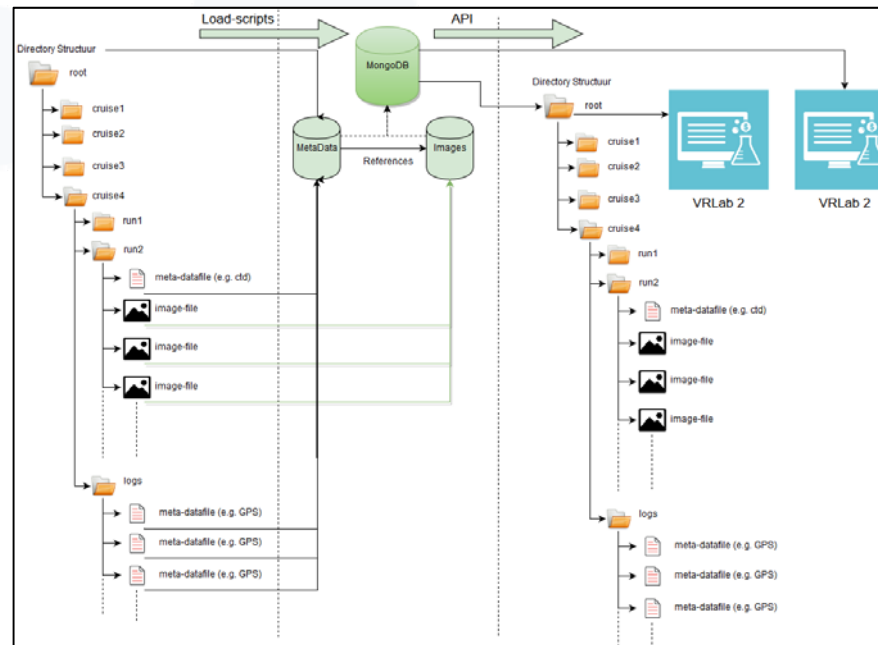
Data synchronization + Data accessibility

The MongoDB databases in VLIZ and IFCA are accessible from the Rshiny/ Rstudio based virtual lab running at VLIZ: the LifeWatch data explorer.

Demonstration website accessing server at VLIZ: <http://rshiny.lifewatch.be/ZooScan%20data/>

Access to files through MongoDB

The virtual labs should also have access to the individual files generated by the different biosensor instruments:



Access to data in SQL databases: <http://rshiny.lifewatch.be/>

Access to data through Geoserver webservices

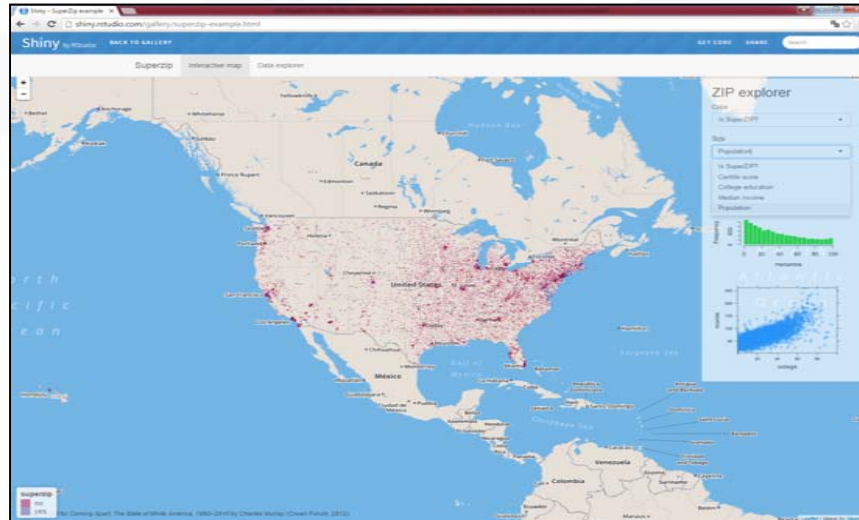
Using Geoserver **clusters** could boost the speed of accessing data.

This is ongoing work within the Geoserver working group in LW-EGI CC.

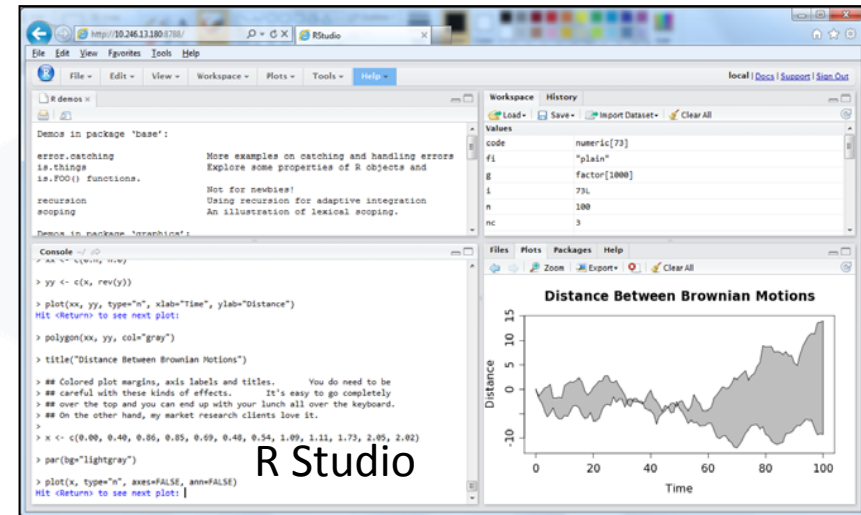
Ecological Observatories and R tools

- A detailed analysis of the possibilities to implement and deploy services oriented to support the use of R is presented in EGI-Engage D6.6
 - starting from the previous experience in the Grid framework (processing data from the LTER Observatory of Sierra Nevada in Spain)
 - describing the implementation in HPC systems, in clusters in other LifeWatch centers (HCMR, VLIZ, IFCA)
 - also starting the discussion on how to compare the performance in order to improve it combining the experience and different approaches of the different teams in the LW-EGI-CC.

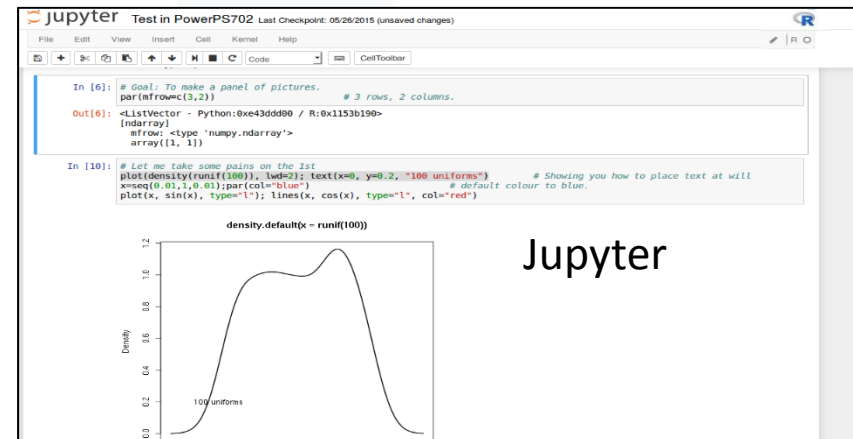
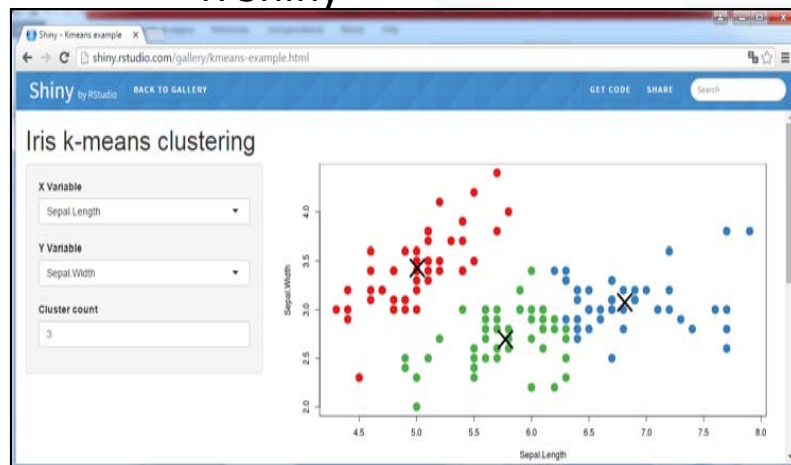
R tools / frameworks



R Shiny



R Studio



Jupyter

Implementations within LifeWatch

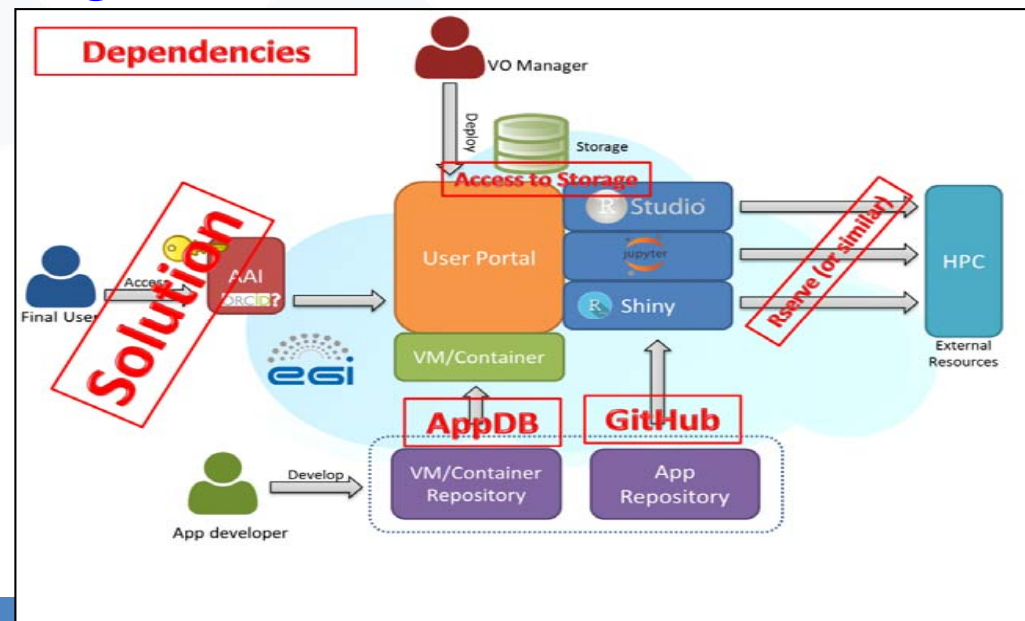
- LW-Be

<http://rshiny.lifewatch.be/> and <http://rstudio.lifewatch.be/>

- LW-Gr (@HCMR)

<https://rvlab.portal.lifewatchgreece.eu/>

- Draft proposal for implementation as service:



CHALLENGES FOR Y2: A REALISTIC FRAMEWORK

- We need **Real** Requirements from **Real** Applications
 - Covering both basic research and management
 - Different scope (Marine, Fluvial, Terrestrial...)
 - Cross-disciplinary, cross-scales
 - Need a **catalog of Open Source solutions**
 - **Benefiting from LW e-Infrastructure**
- Human in the middle?
 - Sustainable?
- User friendly
 - Starting from **Authentication**... to Visualization
- Workflows?
 - easy or sophisticated?
- In collaboration with other H2020 initiatives

CHALLENGES FOR Y2: DEFINITION AND SETUP OF THE e-INFRASTRUCTURE

- FedCloud framework, what else do we need?
 - **LifeWatch is entering in production mode in 2016**
 - Additional support to LW VO?
 - Is FitSM a good idea? **we need SLA and CRM**
 - Additional components (Control Platform)
- Access to external data: GBIF, LTER, ESA
- Support to Open Data
 - The Complete Data Life Cycle
 - Preservation issues

CHALLENGES FOR Y2: **ENGAGE THE COMMUNITY**

- Engage LW regional and national initiatives
 - 7 technical working groups
- Support VRE platforms:
 - **VRE marine LW** (<https://marine.lifewatch.eu>)
 - VRE terrestrial LW
- Target fragmentation of Biodiversity initiatives
 - Biodiversa, Natural Parks, LIFE...
 - Ecological Quality and “Management” projects
- **Citizen Science**

The best data for the best answers

(slide courtesy of F.Pando, Royal Botanic Garden Madrid, GBIF)

- Biodiversity data suitable for analysis

From Science

- very formal
- very reliable
- standardized
- highly integrable (\pm)
- accessible, \pm available
- has representativity issues (emphasis in discovery)
- usually not up-to-date (slow publication)
- very technical (hard to use by non-scientists)



from Public Administrations

- highly representative
- \pm Current
- focused on specific territories
- focused on selected of species
- \pm accesible, \pm available
- hard to integrate

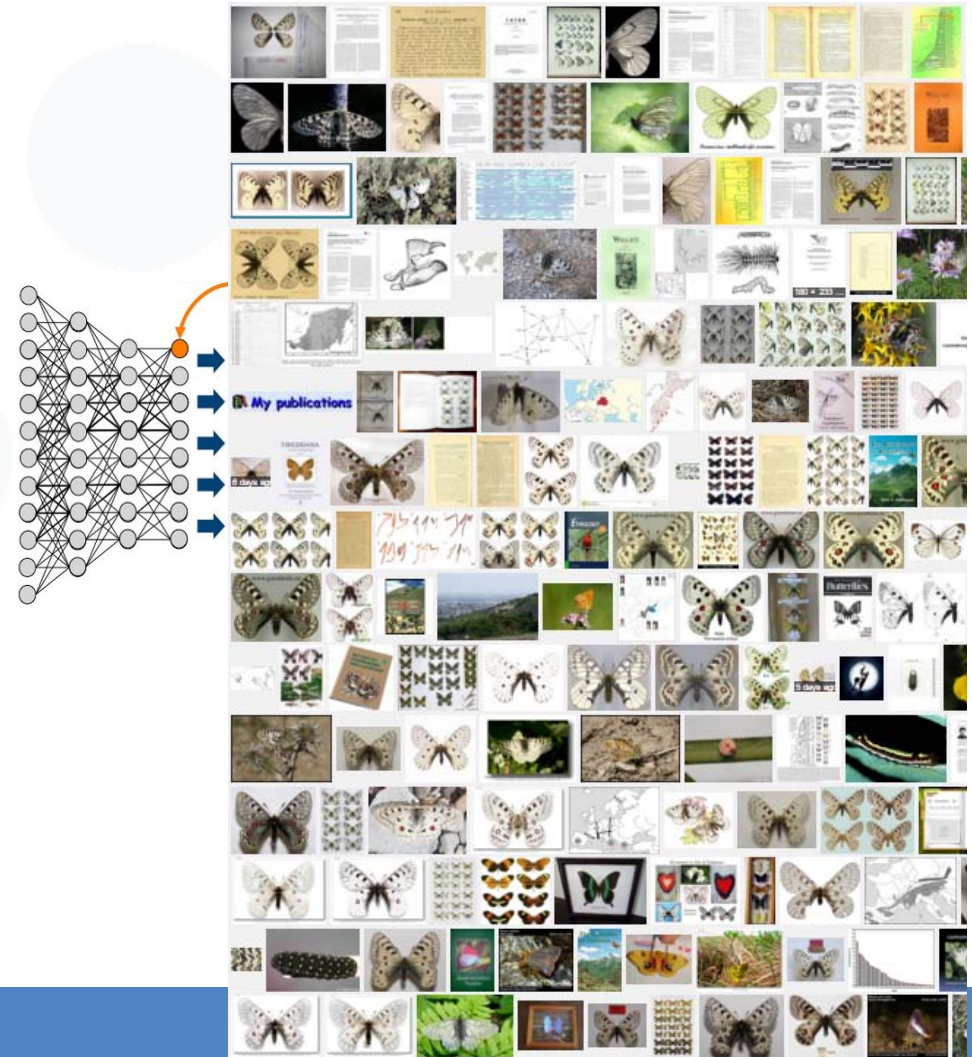
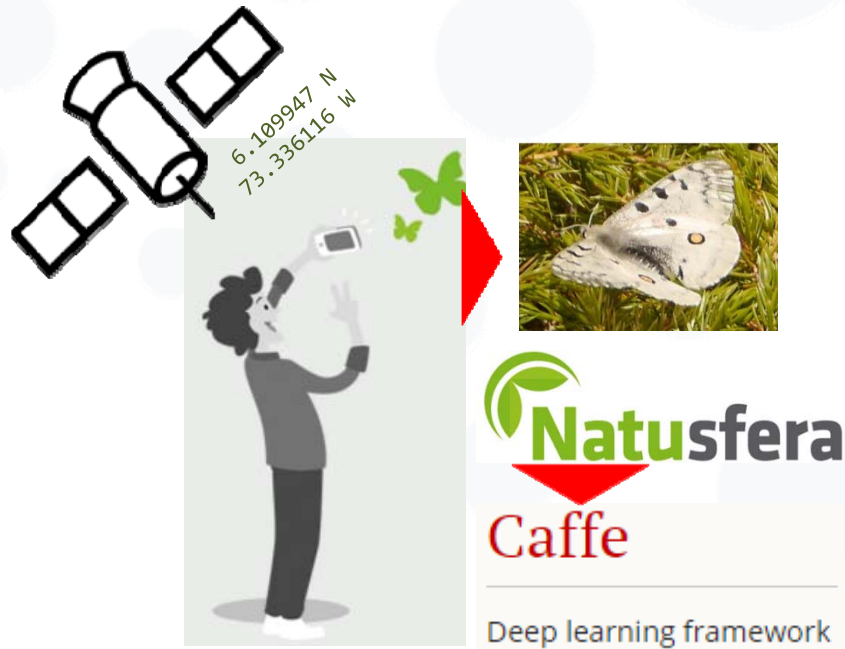
From citizens

- current
- expanding
- recognised
- several bias
- reliability under question
- low accesibility, low availability



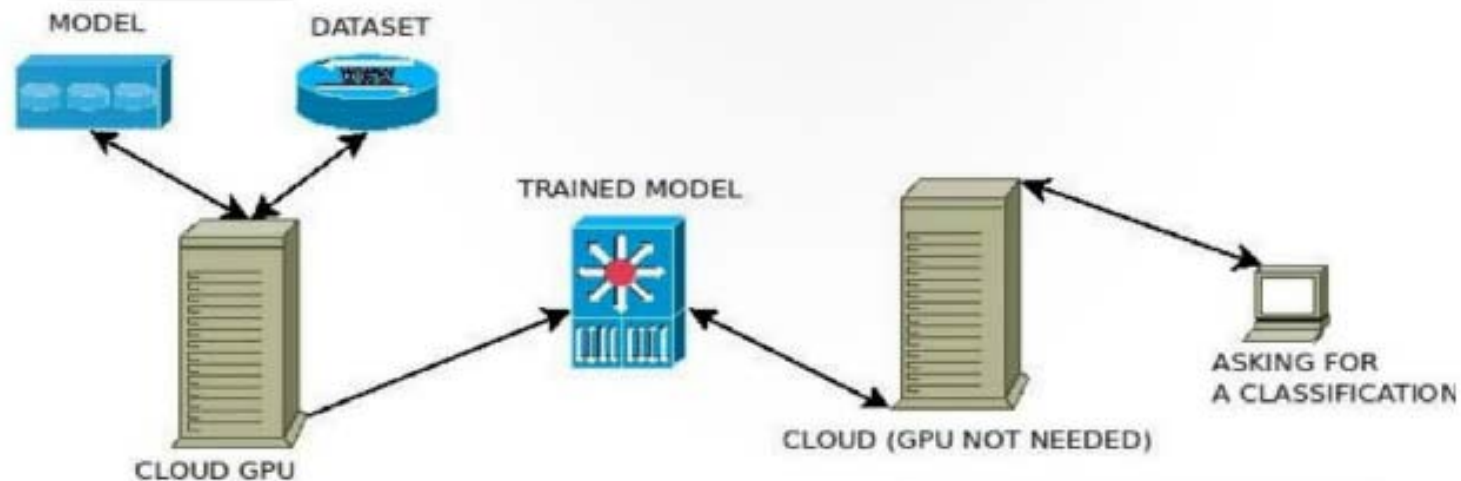
offers a tremendous increase
in data available!

Linking CS Biodiversity observation platform(s) and Species identification based on automatic image analysis



Observer/SC gets species ID
Instant satisfaction ↑
Species distribution DB gets ID
DB quality ↑

Task 4.2: Exploration of pattern recognition tools that could benefit of EGI resources



The Institute for Biocomputation and Physics of Complex Systems (BIFI). Zaragoza. Spain

Concluding remarks

- EGI LW Competence Center has been **instrumental** for us to progress !
- We have adopted the FedCloud basis, and are exploring the components (PaaS, SaaS levels), in collaboration with other projects (like INDIGO)
- A lot of effort put!
-Many thanks to all people and teams!!!
- We start to have an idea of how this framework can be sustained, what tools do we need and what do we miss (in particular more resources)
- **Because the challenge in front of us is very large**
- A final reflection towards our integration into Open Science (Cloud) , motivated by *"Creating a Learning Society: A new approach to Growth"* (Stiglitz&Greenwald): **WE CAN HAVE A LARGE (SOCIETAL) IMPACT IF**
 - We are able to reach an adequate scale
 - We realize that our experience should be exploited to provide learning support!

Thank you for your attention.

Questions?

You are cordially
invited to follow us in
www.lifewatch.eu



www.egi.eu



elroto.elpais@gmail.com

This work by Parties of the EGI-Engage Consortium is licensed under a
[Creative Commons Attribution 4.0 International License](https://creativecommons.org/licenses/by/4.0/).

