



VLabs for biodiversity and ecosystem research

Cloud PaaS - all user communities meeting
EGI Conference 2015 in Lisbon



INPUT FROM LIFEWATCH VLabs

Presented by
Jesús Marco de Lucas
IFCA-CSIC, Spain

www.egi.eu

EGI-Engage is co-funded by the Horizon 2020 Framework Programme
of the European Union under grant number 654142



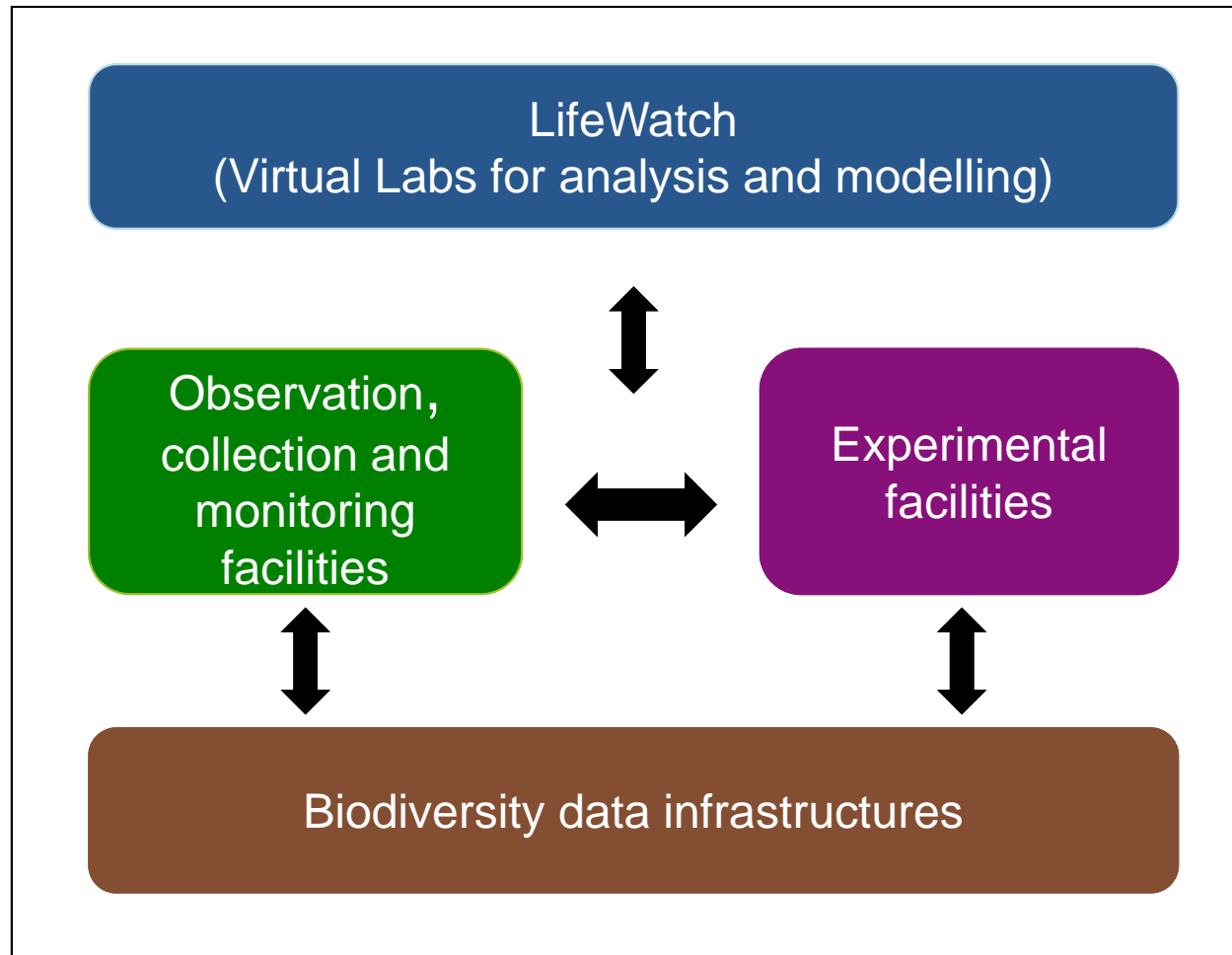
What is LifeWatch?

- 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 case studies:
 - Invasive species
 - Evolution of wetlands
 - Evaluating the ecological quality of habitats



The research infrastructure landscape

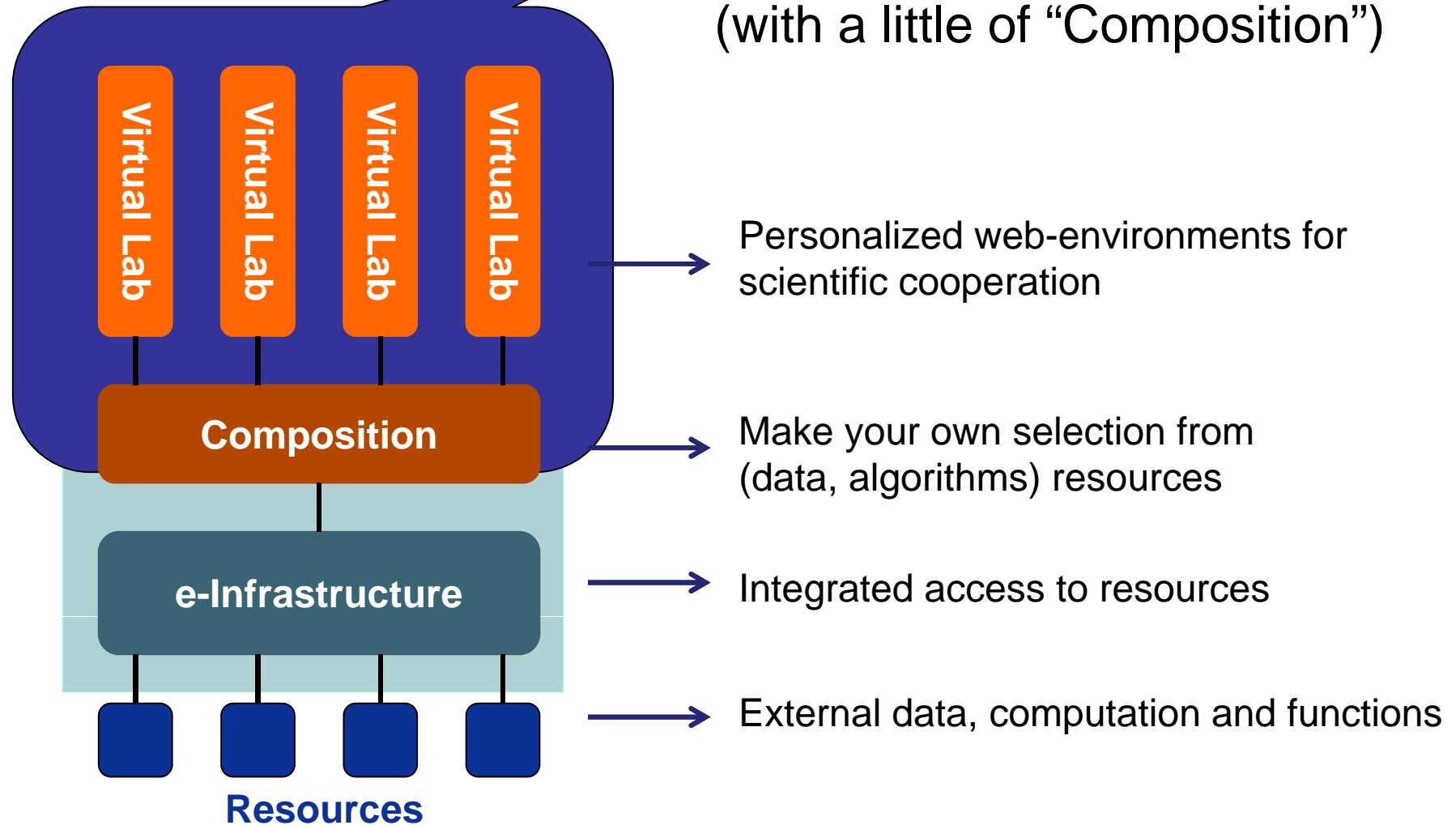
From W.Los, at LifeWatch Vlab meeting in Amsterdam, March 2015





From W.Los, at LifeWatch Vlab meeting in Amsterdam, March 2015

Meeting discussion is mainly on the layer of “Virtual Labs” (with a little of “Composition”)





Exposing VLabs in LifeWatch

- *Virtual Research Environment (VRE)*: brings together various Virtual Labs (and supporting Web services) within a research theme.
- *Virtual Lab*: remotely accessible virtual environments offering **interactive opportunities to collaborate with related web services**.
- *Web service*: access to web-based applications through a standardized interface allowing different applications from different sources to communicate with each other.
- *Workflow*: series of activities necessary to complete a task -> an orchestrated sequence of data transformations.

A LifeWatch VRE for marine research is already launched.

A next step should be a VRE for terrestrial / limnic research, bringing together the active community operating relevant Virtual Labs. Such a VRE will build upon many web services already linked within the marine VRE.



marine.lifewatch.eu VRE

LIFEWATCH
Marine VRE

Access Analyze Develop About

Access

Retrieve and access data resources holding marine biodiversity and ecosystem data. A range of data systems offering data on species names, traits, distribution and genes.

Analyze

Online tools that facilitate data analysis of marine biodiversity and ecosystem data. Analysis is performed on data from known data resources and/or data uploaded by the user.

Develop

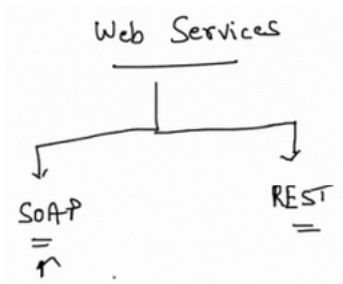
Build your own marine virtual lab making use of a range of available web services that access and process data. Service catalogues and 'how to' manuals help you to develop your own system.

LIFEWATCH Marine VRE

Access
 Analyze
 Develop
 About

Develop

What are web services?



Within the envisaged e-infrastructure of LifeWatch, data exchange and data analysis are largely based on the use of web services. Web services are systems that allow communication between two computers over the web, and allow the user to access the most recent and up-to-date information directly from within other applications.

Web services can roughly be divided into two categories: SOAP (Simple Object Access Protocol) and REST (Representational State Transfer). SOAP has the ability to discover and describe web services via the WSDL (Web Services Description Language) standard, but usually needs a platform dependent library to work. REST uses standard HTTP (and JSON) and is much simpler to use, but lacks a standard way of description (as is the case with SOAP WSDL).

Where to find the web services?

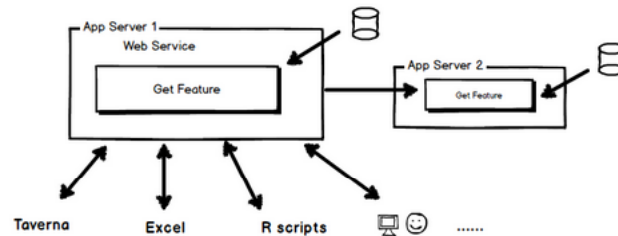
Several catalogues exist listing the available web services. The [BiodiversityCatalogue](#) (developed by BioVeL) is a curated catalogue of available web services that are specific to the interests of the biodiversity science community. This catalogue is related to the [BioCatalogue](#), which focus is on life science web services.

Both catalogues are community-oriented websites where service providers and community experts can register and curate services, and where users can discover them. Users can browse the catalogues and access full search options for services, checking their status and availability.

Geographic web services in general are listed in numerous spatial data catalogues, often with a

Catalogues

How to access the web services?

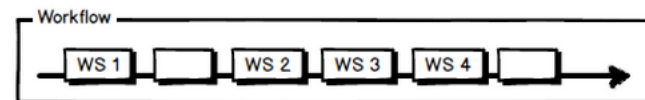


Web services can be accessed from within several platforms or software. Web services can for example be built into PHP web pages, service management tool, R scripts, and even spreadsheets software like MS Excel. Secondary application servers can use the web service to access data from the provider and combine this output with other local processes.

Some example implementations (specific for the WoRMS web services) can be found [here](#).

How to connect web services into workflows?

Web services can be executed consecutively, constituting so-called workflows, for all kind of analytical purposes. The existing BioVeL workflows can be downloaded from [myExperiment](#) and can be run in the [BioVeL Portal](#) and the [Taverna Workbench](#). The latter also enables you to edit workflows and to create your own workflows.

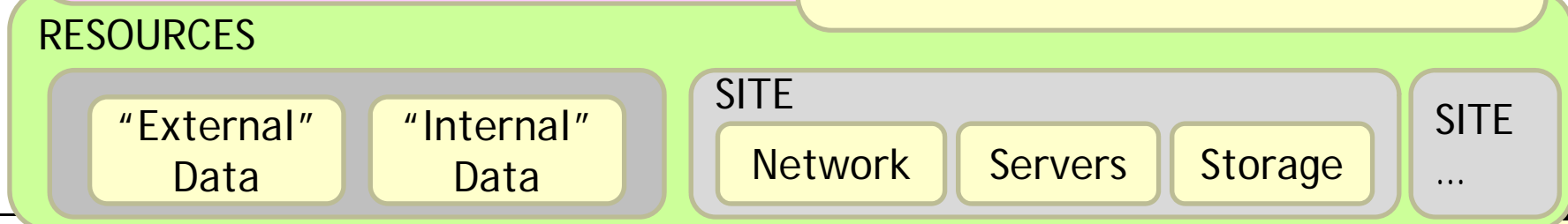
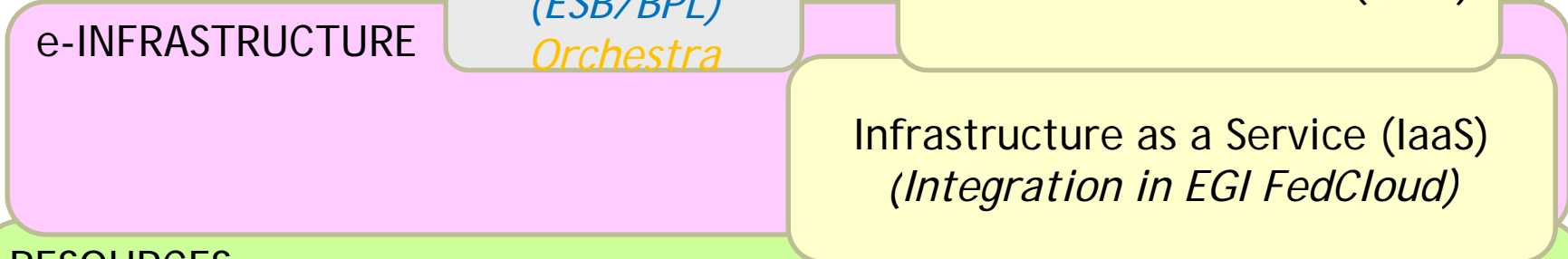
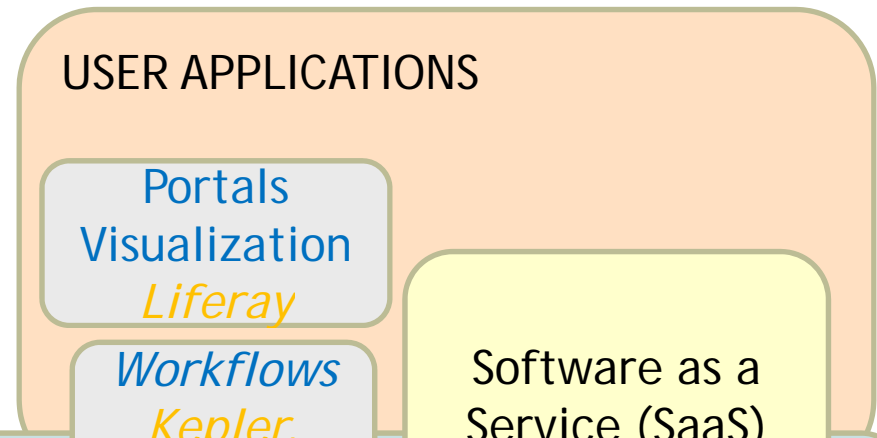
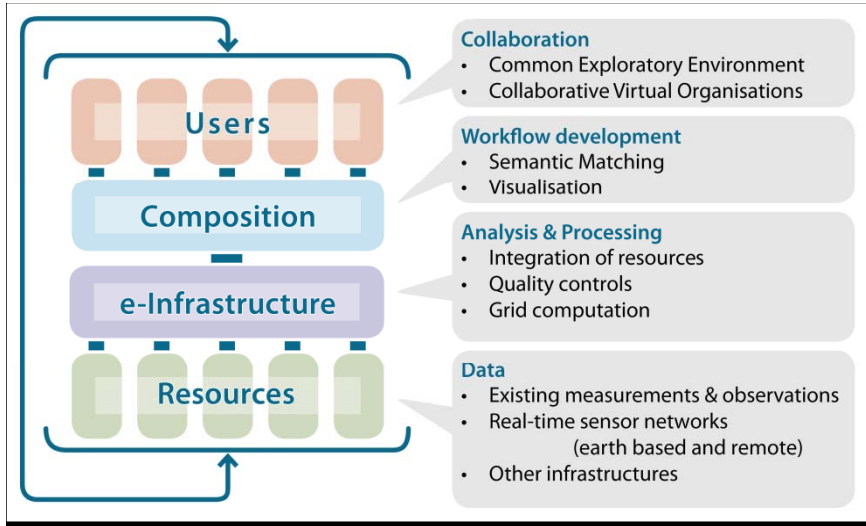




Overview of LW V Labs

- Main characteristics of developing Virtual Labs/services
 - Habitat/Taxon 10x
 - (marine, antarctic, estuaria, wetlands, urban, alien& invasive species, phytoplankton, Natura 2000 management, Sierra Nevada, tracking organism movements, water quality)
 - Specific Discipline/Method 8 x
 - (ecological niche modelling, phenology, synthesis of biodiversity data, taxonomic analysis, handling DNA throughput data, RNA-seq analysis)
 - Environmental Assessment 2 x
 - (assessment ecosystem services, essential biodiversity variables)
 - Data Processing 6 x
 - (R-stat processing, visualisation, 3D manipulation, sensor deployment, LiDAR data processing)
 - Processed data 6 x
 - (taxonomic backbone, Earth observation data, GIS data, plant traits, regional data)
 - Other Virtual Labs 6 x
 - (combining and chaining workflows, decision-support, species identification, digitisation support, biodiversity observation, user manuals)

Global Scheme



Progress with LW e-Infrastructure Implementation

- A pilot project to understand the global framework:
Adaptation and improvement of the e-Infrastructure ICTS-EBD

- Funded by MINECO (CSIC to be commissioned to execute it)
 - Setup an operational framework supporting from basic services to advanced data processing and collaborative work
 - Improve the sensor monitoring network at Doñana
 - MATCH & INTEGRATE ICT Services

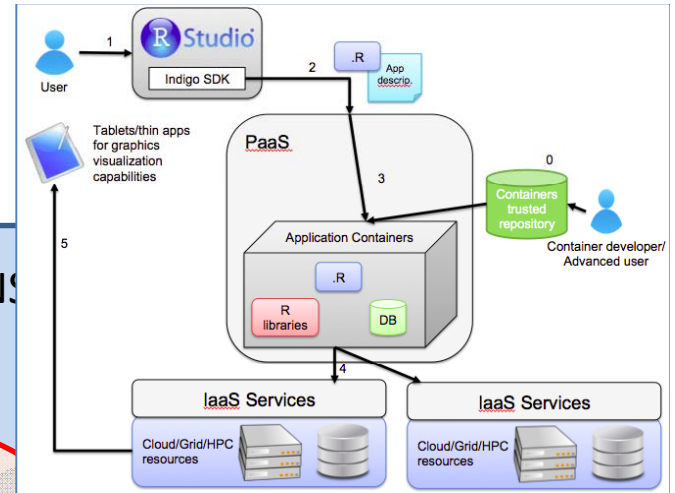
- Four actions:

- ICT e-Infrastructure
 - servers (cloud oriented + HTC/HPC), storage (>1PB), 10GB network
 - methodology for services design **completed in 2015**
 - Implementing services
- Integration of the network of sensors
- Conditioning for in-situ data process at RBD (Natural Reserve)
- New sensors and observational setups





Global Scheme



USER APPLICATIONS

Portals
Visualization
Liferay

Platform for Collaborative Framework

Distributed Control Platform

SOA/ Cloud Computing

COMPOSITION

Open Data & Preservation Platform (ODP)
Orchestra

e-INFRASTRUCTURE

RESOURCES

“External”
Data

“Internal”
Data

SITE

Network

Storage

Servers

SITE

...



EGI- LW 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

Proposal presented by

I.Blanquer & J.Marco

The proposal prepared in July 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 will fund 59 PM

LIFE-WATCH related initiatives will complement in what possible

**Spain 32 PM
 Portugal 9 PM
 Italy 3 PM,
 CIBIO 3 PM
 VLIZ 6 PM
 INRA 6 PM**

- Objective 1- **Adoption and exploitation of the EGI infrastructure by the LifeWatch user community**, reach users through dissemination of LifeWatch in EGI and assist them along the path of enrolment, learning and exploitation.
- Objective 2- Deploy the tools required to support **data management, data processing and modeling for Ecological Observatories** in the framework provided by EGI.eu.
- **Objective 3-** Integrate, and as necessary develop, on the EGI **FedCloud** framework, the services required **to support workflows** oriented to the deployment of **Virtual Labs for LifeWatch**.
- Objective 4- Support to the **direct participation of citizens** in LifeWatch contributing observation records, in particular those **including sounds or images uploading and processing**.



Task 3: Supporting Workflows & Virtual Labs in FedCloud for LifeWatch

Task 3.1 Integration of Bioinformatic interfaces and frameworks (Galaxy) on EGI FedCloud

- Adaptation of a Galaxy portal to run jobs on EGI FEdCloud
- Link the public part of INRA's numerical taxonomy database (R-Syst)
- Create a repository of configurations for addressing different Biocomputing problems

Task 3.2 An extensible framework for biodiversity pipelines on EGI Federated Cloud.

- Prototype available through the OpenModeller HTC service developed in EUBrazil OpenBio
- Niche Modelling Service is implemented through the COMPSs programming framework and available in the EGI AppDB.
- COMPSs will be adopted to develop the applications and to optimize their execution, through automatic parallelization techniques, on the EGI Federated Cloud.

Task 3 .3 Implementation of the Network of Life.

- After an analysis of the framework of different standards, protocols and tools available within GBIF, the needs of adaptation/expansion to support species relationship data will be defined.
- Storage and organization needs of geo-referenced information on species interactions, extracted from the primary literature, will be considered.
- The system implemented will be able to build networks of potential interactions, based on the species that have been reported in a given area. Social network algorithms will be used.

Thank you for your attention.

Questions?

You are invited to the EGI-LW Competence Center meeting at 17h on Thursday 21st



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/).

