Challenges in the EGI-LifeWatch Competence Center

Research Infrastructures Roadmaps and Competence Centers

EGI Conference 2015 in Lisbon

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
Outlook

- General Information on LifeWatch ESFRI
- LifeWatch distributed e-Infrastructure
- The EGI LifeWatch Competence Center
  - The Project
  - Task 1: Support to User Communities
  - Task 2: Big Data and Ecological Observatories
  - Task 3: Supporting Workflows & Virtual Labs in FedCloud for LifeWatch
  - Task 4: Advanced Support to Citizen Science in Biodiversity

- OUR REAL CHALLENGES
What is exactly 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
An example of requirements from a research-management project lead by an SME addressing biodiversity and environmental problems, integrating instrumentation, modeling, simulation and prediction across many different scales (courtesy of A. Monteoliva, ECOHYDROS CEO)

www.roemplus-life.eu
High Resolution AppROach for ManagEMent of Surface Water EutroPhication in Rural Areas of the DUero River Basin (LIFE11 ENV/ ES/ 590)

2nd Users Workshop
14 May 2015
ROEM+: SENSOR NETWORK
ROEM+: CORE STATION

Estación completa de medida (*) con posibilidad de efectuar medidas en toda la columna de agua y aplicación al análisis del estado ecológico de la masa de agua y diagnóstico precoz.

(*) Estación desarrollada por ECOHYDROS en colaboración con el IFCA (Cantabria) en el marco del proyecto europeo DORII

Parámetros: Temperatura ambiente, presión atmosférica y humedad relativa; Radiación; Profundidad, temperatura del agua y conductividad eléctrica; OD, pH y redox; Carbonatos, nitratos, nitritos, carbono orgánico total, sólidos en suspensión y demanda química de oxígeno; Biomasa de cianobacterias (concentración de ficocianinas); Biomasa de fitoplancton (concentración de clorofila a); Materia orgánica disuelta (CDOM); Irradiancia hiperespectral
Ubicación en ríos tributarios y sus vertientes. Controlan la carga actual de nutrientes que recibe el embalse y la respuesta a los tratamientos complementarios que se adopten en cada una de las subcuencas de drenaje al embalse para su aplicación a la *determinación cuantitativa de agentes causales y predicción*.

Parámetros: caudal, temperatura del agua, conductividad eléctrica, pH, oxígeno disuelto, nitratos, carbono orgánico total, sólidos en suspensión y demanda química de oxígeno.
ROEM+: Stations to detect Cyanobacteriae

Ubicación en puntos de control críticos de la masa de agua (zonas de baño, de cultivo, de abastecimiento y/o de control de modelos de simulación) para Indicación de Alerta de eutrofización en tiempo real.

Parámetros: sensores básicos de temperatura, profundidad y pH, y fluorómetros de clorofila (biomasa de algas, estado trófico), ficocianinas (biomasa de cianobacterias, riesgo de toxicidad) y un turbidímetro (transparencia del agua y corrección de influencia de sólidos no algales en las lecturas del fluorómetro de clorofila).
Simulation suite to model the physical, chemical, and biological evolution of water masses

**DELFT 3D**

- **FLOW** (Hydrodynamics)
- **WAVE** (Waves)
- **PART** (Particles)
- **SED** (Sediments)
- **WAQ** (Water Quality)
- **ECO** (Ecology)

**INPUTS**

- Tributaries flow
- Temperature
- Salinity
- Dam out-flows
- Meteo parameters (radiation, humidity, air temperature, wind speed and direction)

**Outputs**

- Hydrodynamic model results
- Parameters of tributaries: dissolved oxygen, phosphates, nitrates, etc. (daily resolution)
- Starting algae concentration

- Water reservoir model of temperature, salinity, water flow, water level
- Simulation of dissolved oxygen rate, algae growth, etc.
ROEM+: A COMPLEX MODEL

The cycle of Phosphorus
ROEM+: RUNNING PREDICTIVE MODELS

**ROEM+: RUNNING PREDICTIVE MODELS**

**Real Example**

Thermal stratification on the water column (2014, April to November)

---

Prediction of Cyanobacteria in the water column
ROEM+: HANDLING ALERTS: CYANOBACTERIAE EVOLUTION?

Changes in the phytoplankton community?

Evolution of cyano levels along last 5 years in the water reservoir:

2010 - 2011
*Dolichospermum planctonicum*
*Aphanizomenon flos-aquae*

2013
*Dolichospermum crassum*
*Woronichinia naegeliana*

2014
*Microcystis novacekii*
*Dolichospermum crassum*
LifeWatch, status and perspective

- LifeWatch is an ESFRI
  - **ERIC being signed this year, 2015**
  - Roadmap for next 5 years is prepared (large investment)
- **Countries participating with different implication levels:**
  - *Spain* (statutory seat, e-infra), *Italy* (service center), *Netherlands* (Virtual Labs)
  - *Belgium, Greece, Portugal*
  - Observing: Norway, Sweden, Finland, Hungary, Slovakia, Romania, Slovenia
  - Interested: France, Germany
- LifeWatch recent meetings
  - Feb 2015 in Malaga:
    - Strategic Plan Discussion
  - March 2015 in Amsterdam:
    - Devoted to Virtual Labs
Global Scheme

USER APPLICATIONS

- Portals Visualization
  - Liferay

- Workflows
  - Kepler,
  - Taverna

Software as a Service (SaaS)

Platform as a Service (PaaS)

Infrastructure as a Service (IaaS)

(Integeration in EGI FedCloud)

COMPOSITION

User Applications

Portals

Visualization

Workflows

Kepler,

Taverna

Software as a Service (SaaS)

Platform as a Service (PaaS)

Infrastructure as a Service (IaaS)

(Integration in EGI FedCloud)

RESOURCES

- "External" Data
- "Internal" Data

SITE

Network

Servers

Storage

SITE...

5th LifeWatch e-Infrastructure Construction Operational Meeting
**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 Lifera

COMPOSITION Software as a Service (SaaS)

Workflows Kepler, Taverna

Platform for Collaborative Framework

Open Data & Preservation Platform

Distributed Control Platform

SOA/ Cloud Computing

SOFTWARE AS A SERVICE (SaaS)

Infrastructure as a Service (IaaS)

PLATFORM AS A SERVICE (PaaS)

INFRASTRUCTURE

“External” Data

“Internal” Data

RESOURCES

SITE

Network

Storage

Servers

SITE

…
Example: Sensor Networks and Virtual Labs

• Sensor Networks deployed at Doñana's National Park

• Cloud-based sensor data storage, presentation and access

• Virtual Lab integration in the ICT-Core
  • Real Time monitoring
  • Historical data management
  • Assessing data quality

• Testing Lab development
  • Maintain, calibrate, and troubleshooting instrumentation
  • Install instruments using commonly-accepted best practices

DATA PROCESSING & DATA GATHERING & DATA VALIDATION

Testing Lab
Sensor Data Portal
Virtual Lab
WSN
Sensor Networks and Virtual Labs

• Biological Data Modelling for
  • Automated data from Sensor Networks
  • Manual tracking data

• Data Extractor tool
  • 7,036,418 observations since 2010
  • 89 sensors and 31 dataloggers
  • Historical and real time extractions

• Use Case: Eddy Covariance towers
  • Wetlands are monitored with high accuracy
  • 3 flux towers based on
    • CO$_2$/H$_2$O flux
    • Sonic anemometers

• Francisco M Sánchez is working at Estación Biológica de Doñana (Seville, Spain)
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
Objectives

- **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.
<table>
<thead>
<tr>
<th>TASK 1 Support to LifeWatch Community and Exploitation</th>
</tr>
</thead>
<tbody>
<tr>
<td>T1.1 Connection of Project Management tools</td>
</tr>
<tr>
<td>T1.2 Implementation of common support tools.</td>
</tr>
<tr>
<td>T1.3 Deployment and operation of a support and training platform</td>
</tr>
<tr>
<td>T1.4 Definition of a dissemination and exploitation strategy</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TASK 2 Big Data and Ecological Observatories</th>
</tr>
</thead>
<tbody>
<tr>
<td>T2.1 Handling Data Streams from Ecological Observatories</td>
</tr>
<tr>
<td>T2.2 Supporting large software suites for Modelling Ecosystems</td>
</tr>
<tr>
<td>T2.3 Towards an integrated framework</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TASK 3 Supporting Workflows &amp; Virtual Labs in FedCloud for LifeWatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>T3.1 Integration of popular Bioinformatic interfaces</td>
</tr>
<tr>
<td>T3.2 An extensible framework for executing biodiversity pipelines</td>
</tr>
<tr>
<td>T3.3 Implementation of the Network of Life</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TASK 4 Advanced Support to Citizen Science in Biodiversity</th>
</tr>
</thead>
<tbody>
<tr>
<td>T4.1 Updated analysis of ongoing initiatives on nature observation</td>
</tr>
<tr>
<td>T4.2 Exploration of pattern recognition tools</td>
</tr>
<tr>
<td>T4.3 Engagement: outreach and inreach.</td>
</tr>
<tr>
<td>Participant No *</td>
</tr>
<tr>
<td>-----------------</td>
</tr>
<tr>
<td>1 (Coordinator)</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>7</td>
</tr>
</tbody>
</table>
Task 1: Support to LifeWatch Community and Exploitation

- LifeWatch is implementing a comprehensive process to **support** its developers, operation and end-users
- **The LifeWatch support to end-users is handled through a Service Center being established in Lecce**
- The Lifewatch CC in EGI will connect a support team in EGI, operated by IBERGRID (NGI-ES and NGI-PT) and the core-ICT team in LifeWatch, with the communities of developers and end-users, in particular for the integration, operation and monitoring of new applications and services.
- This connection will be established at different levels:
  - **Full lifecycle support for application projects, including:** a register of projects, documentation, incidents and evolution management.
  - General forum for discussion of evolution, new ideas, and to gather feedback, implemented using communication tools and complemented with workshops.
  - **Training platform, including hands-on exercises, user guides, webinars and online specific courses**
### Task 2: Big Data and Ecological Observatories

**Description of work**

**Task 2.1 (UGr as JRU-LW-ES, VLZ as LW-BE, NGI-FR, CIBIO as LW-PT)** Handling Data Streams from Ecological Observatories: Flanders Marine Ship (BE), Mountain Observatory in Sierra Nevada (ES), Life under natural radiation (ZATU, FR), Lakes and Water Reservoirs (Sanabria Lake and CDP Water Reservoir, ES)

**Task 2.2 (CSIC as JRU-LW-ES)** Supporting large software suites for Modelling Ecosystems: Delft3D (on water quality and eutrophication), Community Land Model on Global Carbon.

**Task 2.3 (CSIC as JRU-LW-ES)** Towards an integrated framework/toolbox at international level including a catalogue of applications and final user interfaces based in R and Python.

### Deliverables/milestones (brief description and month of delivery)

- **D2.1** Proposal for a data flow handler to support integration of the information from Ecological Observatories. Type: Prototype. Due: M6
- **D2.2** Deployment of basic R tools to process data from Ecological Observatories using HTC/HPC infrastructure available in EGI. Type: Tools+Report Due: M12
- **D2.3** Support (installation, definition of images and context, connection to HTC/HPC/Data resources) to the execution of simulation packages Delft3D and CLM. Type: Report. Due: M12
- **D3.4** Report on the applications installed and usage record. Type: Report. Due: M24
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.
Task 4: Advanced Support to Citizen Science in Biodiversity

Task 4.1 (BIFI as NGI-ES + RJB-CSIC as JRU-LW-ES): Updated analysis of ongoing initiatives on nature observation and selection of an example of framework to be supported from the DCC.

There are several initiatives on nature observation that share some of the features we want to use about image/sounds uploading and analysis by the citizens, like for example http://www.naturalist.org/ or http://www.ebird.org/. This task will analyse the framework of some of these initiatives and the possibility to integrate them with our objectives. This has the double advantage of reducing the development costs and of using a platform already known by the potential collaborators.

Task 4.2 (BIFI+IFCA as NGI-ES): Exploration of pattern recognition tools that could benefit of EGI resources.

This task will address the technical point of exploring the integration and deployment of pattern recognition tools on EGI specific resources, including for example servers with GPUs or other relevant hardware for image/sound recognition.

Generic tools available in the market at different levels (like existing ones to identify grasshoppers, or bee identification from wing images) will be explored and considered, and an initial pack will be integrated and deployed. Tools considered will range from highly assisted, including support from experts or other citizen scientists, like in the inaturalist platform already cited, to fully automated. The results of the analysis will be taken into account to prepare future initiatives addressing the educational level.

Task 4.3 (BIFI + RJB-CSIC): Citizen engagement: outreach and inreach.

This task will deal with attracting and retaining people who would be willing to contribute with their skills, time and effort to the project. This task will rely for sustainability on the collaboration with existing associations with long tradition and experience in the field. Using social networking features, collecting experiences of the collaborators, approaching institutions or involving schools will be some of the instruments to be used, plus actions for further dissemination through workshops, press, etc.

The task will culminate both developments and general public engagement showing and evaluating the outcomes of the citizen science. A public participatory event oriented to bring tools, data and methods to the different stakeholders, in particular general public and younger students, is proposed as a demonstrator of the impact of these actions.
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 CC will
  – deploy cloud and GPGPU based e-Infrastructure services required to support data management, data processing and modelling for Ecological Observatories,
  – explore possibilities to increase the participation of citizens in data-intensive biodiversity research,
  – facilitate the adoption and exploitation of the EGI infrastructure by the LifeWatch user community.
LW-CC Deliverables & Milestones

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
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

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

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
Worries…or challenges

- To be successful on exploitation, how do we assemble and provide/propose/suggest a “complete” platform?

- How to interact with ongoing support projects to research communities with strong ICT flavour (example: ENVRI+..., next: VRE projects) and over different e-Infrastructures (EGI, PRACE, EUDAT...)
  - EGI Competence Centers
  - MoU with different initiatives? Concertation?

- Are we tracking what happens in the world?
  - Examples:
    - use of Amazon AWS,
    - XSEDE

- How do we really “connect” with the research teams and their needs and experience?
  - Person in the middle
  - Training? CRM?
The communication problem

As “ICT”, we have procedures to consider, support, track requirements using a project management tool (*from User Stories to Backlog items)*

But… our “final users” are not “ICT”, in fact, we work in a complex ecosystem (*More complex in reality than what we can try to model*)

- We may have a basic layer with ICT-infrastructure experts
  - They don’t know much about applications
- A middle layer with ICT data/modeling experts
  - They know about applications
  - They may do research, they usually do not write the papers
- A “final-user” layer with researchers
  - They want to use the applications, they provide the knowledge to write them, they usually write the papers

The three layers need to communicate/cooperate:
- The two first layers can talk “ICT” between them, the last two, not!
- A “mixed profile” expert is of course a very good solution.
- But, how can we “annotate” this communication, even if internal?
- This is a key question for building applications using a PaaS/SaaS system
- And the first key communication is on “requirements”
CHALLENGE 1: 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
CHALLENGE 2: DEFINITION AND SETUP OF THE e-INFRASTRUCTURE

• FedCloud framework, what do we need?
  – LW will go in production mode in 2016
  – Who will support LW VO?
  – Is FitSM a good idea? **we need SLA and CRM**
  – Additional components (Control Platform)

• Access to external data: GBIF, LTER, ESA, etc.

• Support to Open Data
  – The Complete Data Life Cycle
  – Preservation issues
CHALLENGE 3: ENGAGE THE COMMUNITY

• LW regional and national initiatives
• VRE `platforms: VRE marine LW,...
• Fragmentation of Biodiversity initiatives
  – Biodiversa, Natural Parks, LIFE...
  – Ecological Quality and “Management” projects
• Citizen Science
Thank you for your attention.

Questions?

You are invited to the LW CC meeting at 17h on Monday 18th and on Thursday 21st

www.egi.eu

This work by Parties of the EGI-Engage Consortium is licensed under a Creative Commons Attribution 4.0 International License.