



# *LifeWatch*

**Jesús Marco de Lucas**  
**( marco [at] ifca.unican.es )**  
**Instituto de Física de Cantabria,**  
**CSIC-Universidad de Cantabria**  
Coordinador de “LifeWatch EGI-Engage  
Competence Center”



NUEVAS TECNOLOGÍAS APLICADAS A LA COMUNICACIÓN Y A LA CIENCIA CIUDADANA  
PROGRAMA DE FORMACIÓN AMBIENTAL DEL CENEAM OAPPNN-MAPAMA 2017

**CENEAM, Valsaín (Segovia) 24 mayo 2017**

# ¿LifeWatch?

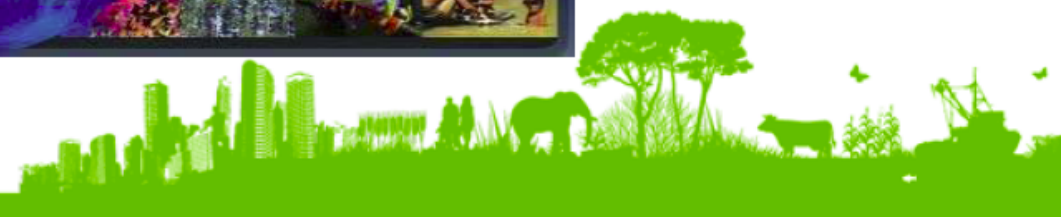
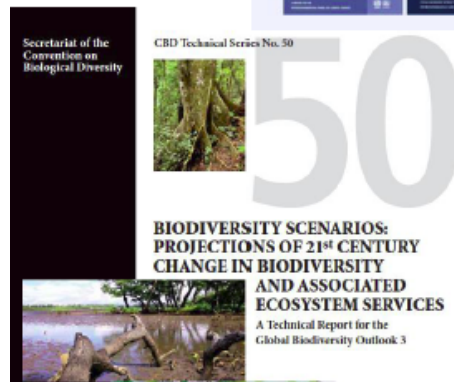
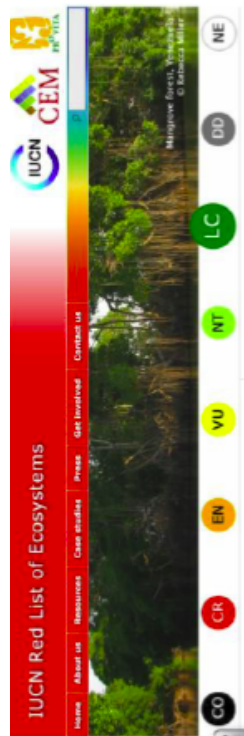
- Qué es la ESFRI LifeWatch y por qué es tan importante para la iniciativa LifeWatch participar hoy aquí
- Big Data? In the Cloud? Data Science?: trabajo en marcha
- Ejemplos prácticos “cercanos”
- ¿Cómo colaborar?





# 1. LIFEWATCH-ERIC: MISSION AND ORGANISATION

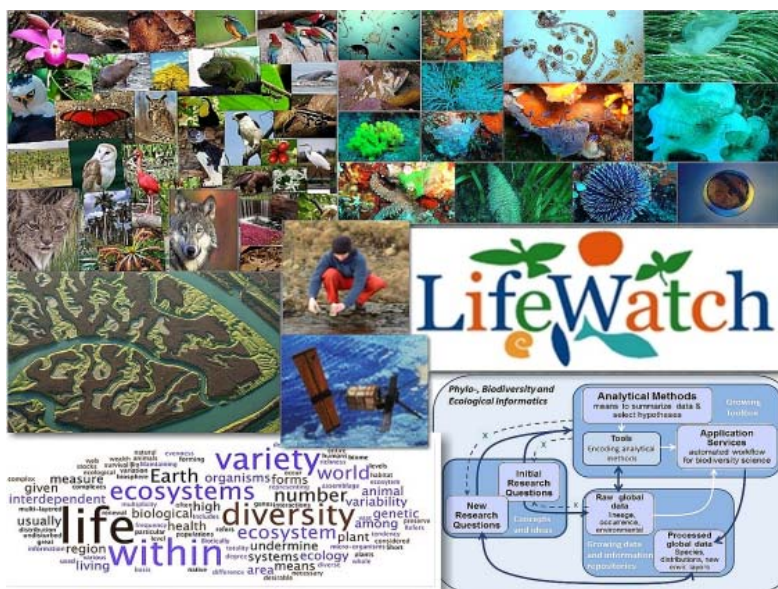
**LifeWatch** is the **European Infrastructure** supplying **e-Science** research facilities for **scientists** adding knowledge and deepening understanding on **Biodiversity** organisation **and Ecosystem** functions and services, with the goal of supporting **our societies** in addressing the key planetary challenges





# Presentando “la ESFRI” LifeWatch

**LifeWatch** es la infraestructura europea en **e-Ciencia** orientada a la investigación en Biodiversidad y Ecosistemas.



**TYPE:** distributed  
**COORDINATING COUNTRY:** ES  
**MEMBER COUNTRIES:** BE, EL, ES, IT, NL, PT, RO  
 March 7<sup>th</sup>, 2016 update  
**PARTICIPATING COUNTRIES:** FI, FR, HU, NO, SE, SI, SK

## TIMELINE

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

<http://www.lifewatch.eu>

LifeWatch está en el programa del Foro de Estrategia de Infraestructuras de Investigación Europeas (ESFRI) [http://ec.europa.eu/research/infrastructures/index\\_en.cfm?pg=esfri-roadmap](http://ec.europa.eu/research/infrastructures/index_en.cfm?pg=esfri-roadmap)



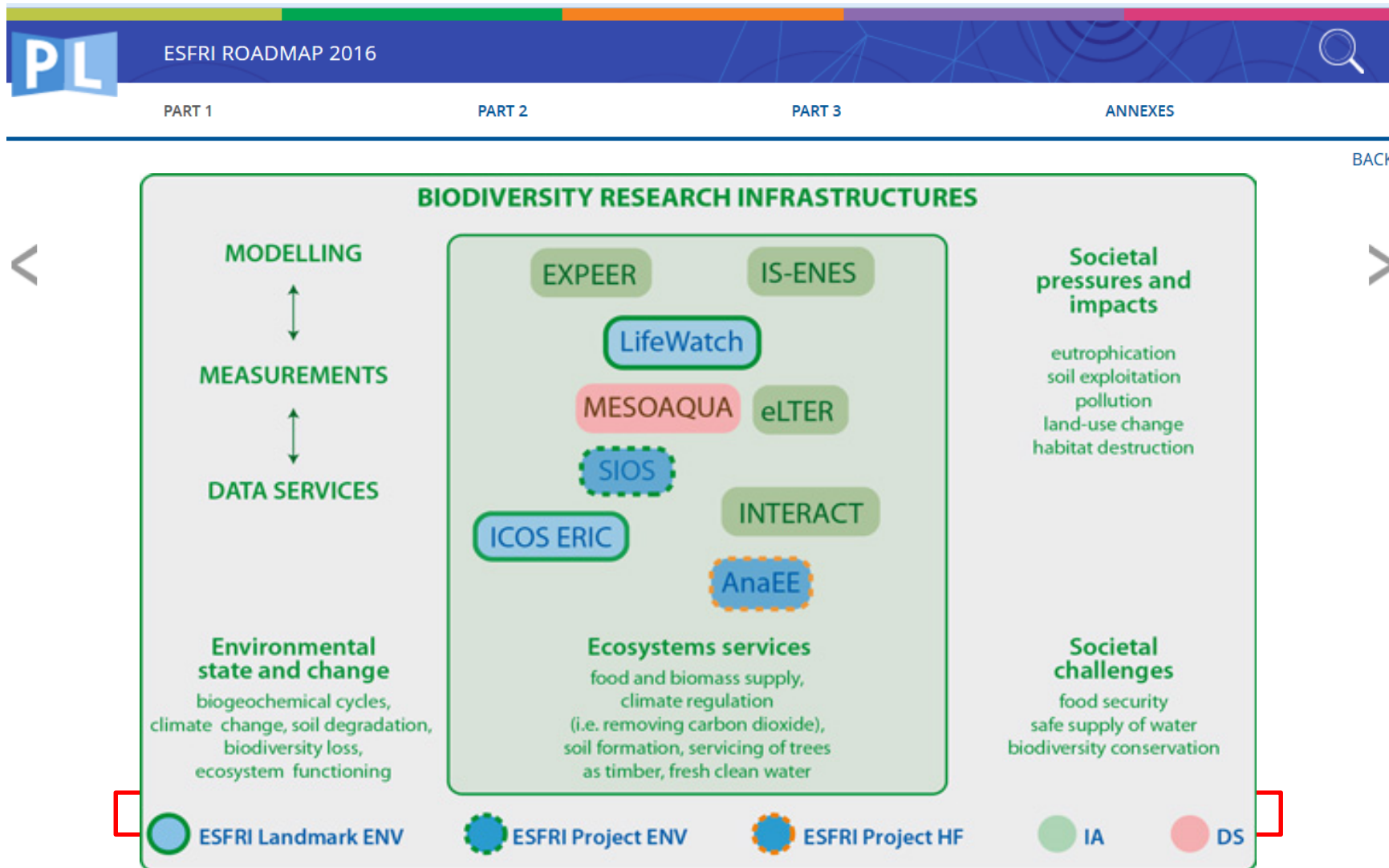
# *Objetivo de LifeWatch*

- ❁ Proporcionar la infraestructura y sistemas de información necesarios para contar con una plataforma analítica, que permita el modelado de datos, tanto existentes como nuevos, sobre biodiversidad, en un ecosistema, para mejorar nuestro conocimiento sobre su funcionamiento y su gestión.
- ❁ Ejemplos de casos de estudio relevantes:
  - ❁ Especies invasoras
  - ❁ Evolución de humedales
  - ❁ Evaluación de las condiciones ecológicas de habitats
- ❁ LifeWatch es una infraestructura orientada a dar soporte a la comunidad científica **y también a otros usuarios, especialmente a los gestores.**





# LifeWatch en el contexto ESFRI





**LifeWatch-ERIC** is a **distributed** infrastructure, operating through **Common Facilities** and **Thematic Centres**.



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

**PARTICIPATING COUNTRIES:** 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**

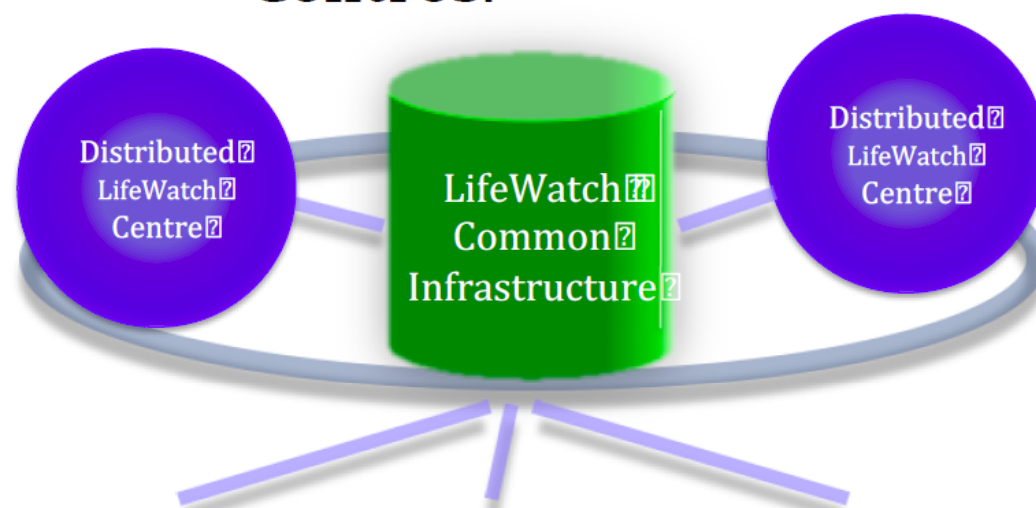
- Capital value: 66 M€
- Operation: 10 M€/year

#### **HEADQUARTERS**

Statutory Seat: Seville (ES)  
Common facilities: ES-IT-NL

#### **WEBSITE**

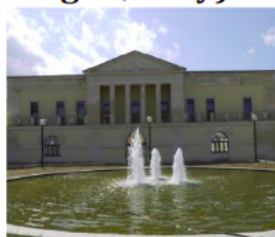
<http://www.lifewatch.eu>



**Virtual Labs and Innovations Centre**  
(Amsterdam, The Netherlands)



**Service Centre**  
(Lecce-Regione Puglia, Italy)



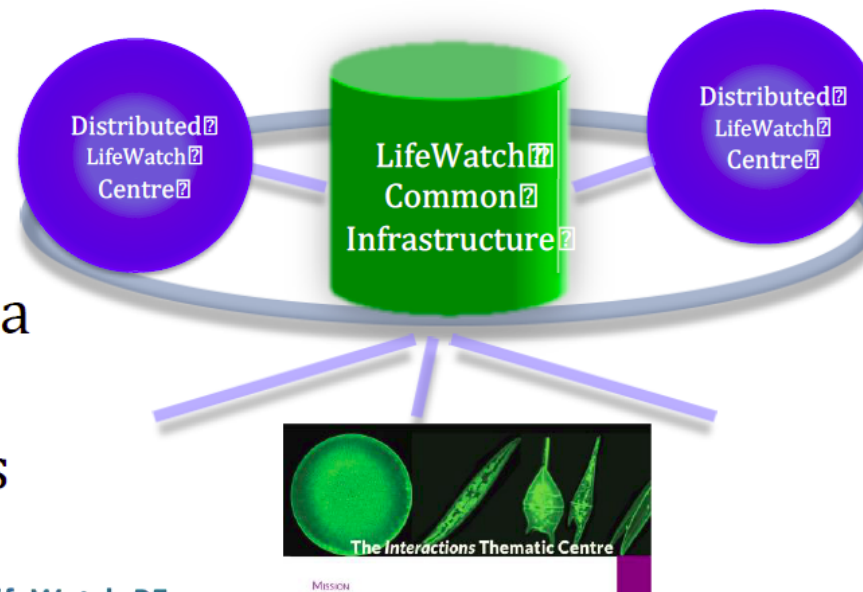
**Headquarters-Statutory Seat & ICT-Infra Office**  
(Andalusia, Spain)



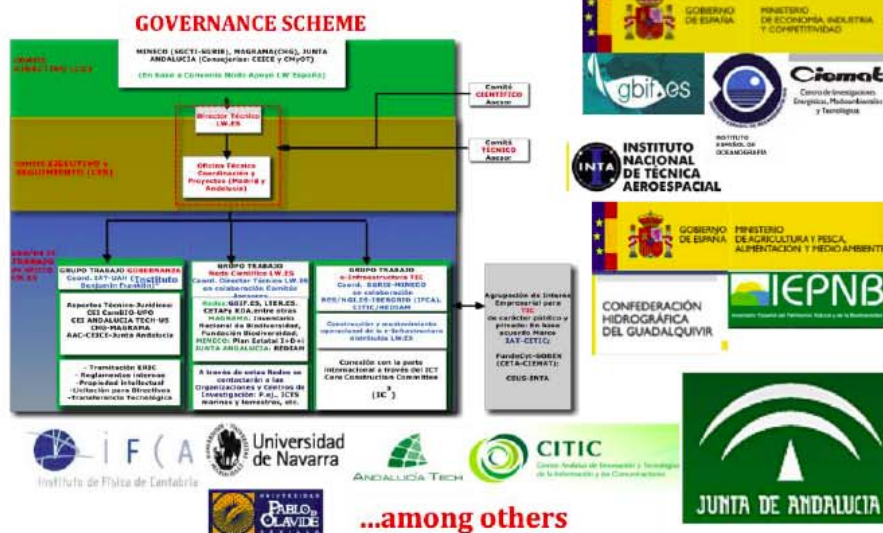
**Thematic Centres** are in member countries (and REGIONS) and develop components of the **e-Science** facilities



**LifeWatch-ERIC national networks**  
(BE, EL, ES, IT, NL, PT, RO, SI) include a large and representative component of academic and research institutions and national authorities.



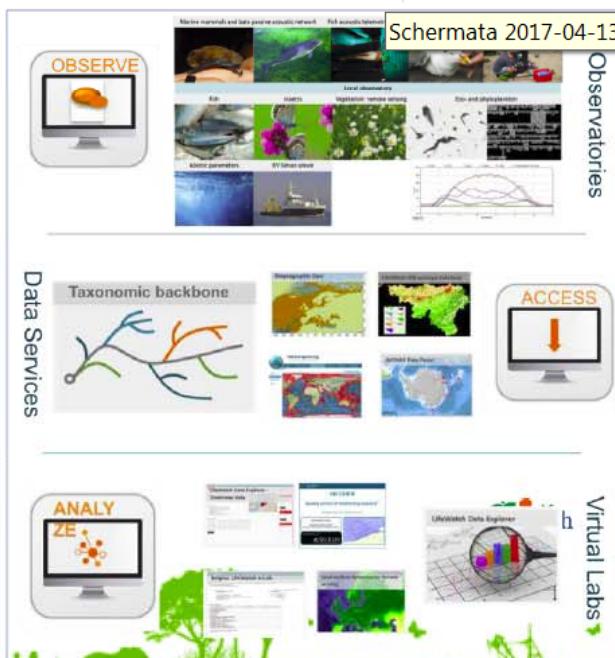
**LIFEWATCH SPAIN Community** is structured by its Joint Research Unit (acronym **JRU LW.ES**)  
Composed of **29 Organizations & Institutions** (April 11, 2017)



**LifeWatch.BE**



Schermata 2017-04-13

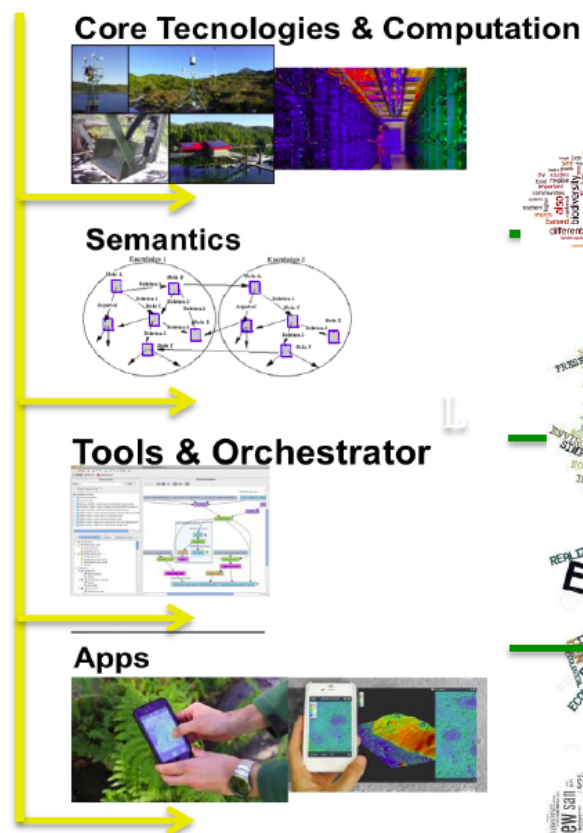




# 2. LW-ERIC: MAIN SCIENTIFIC DIRECTIONS

## COMPUTER AND BIOLOGICAL SCIENCES, DATA ACCESSIBILITY AND THEIR INTEGRATION

### Computer Science



### Data



### Biological Science



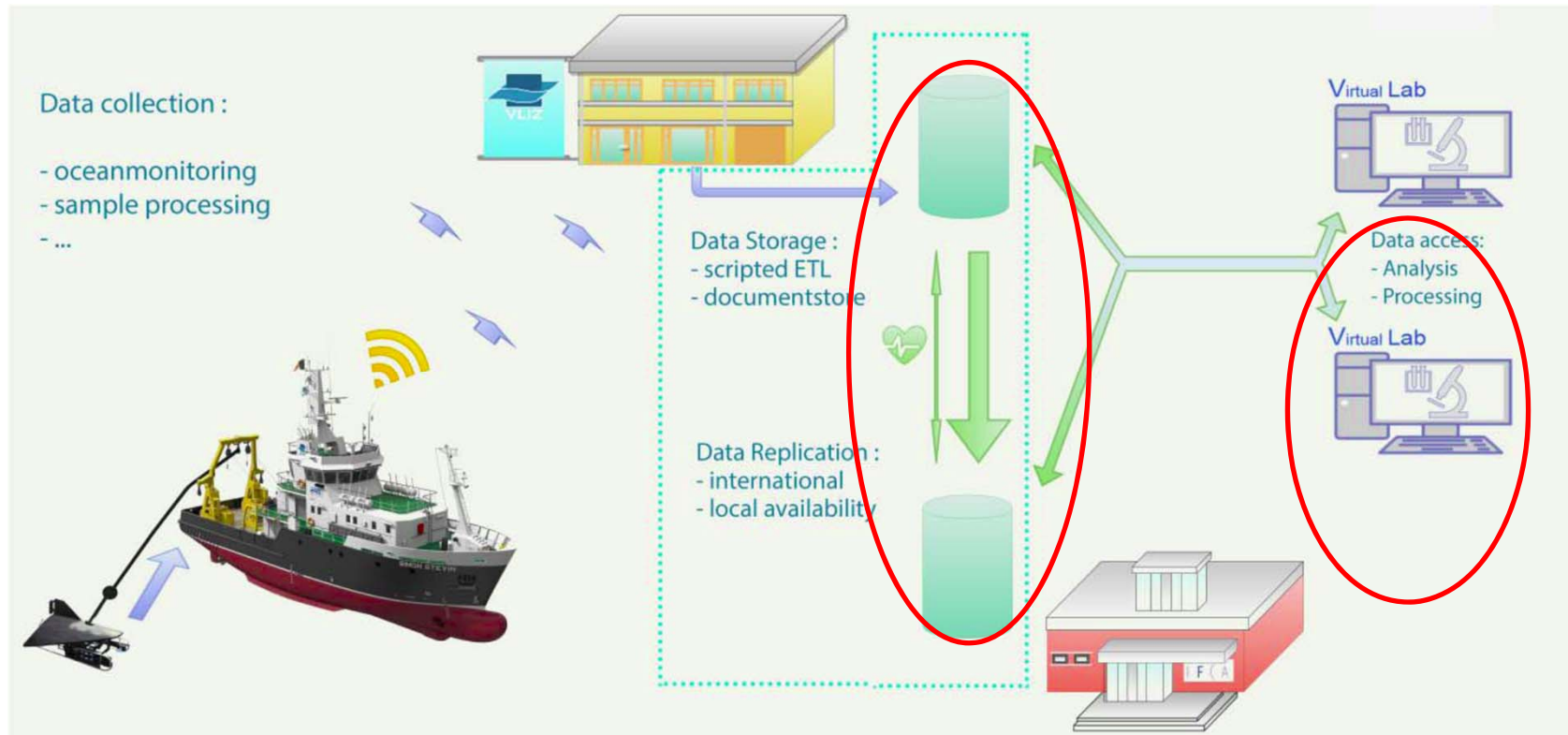
- Providing data  
Validating facilities
- Running experiments  
Building new knowledge
- Deepen understanding  
Supporting policy

LifeWatch community

LifeWatch e-infrastructure



# *Ejemplo: Observatorios Marinos*



**Flujo de datos en el caso del Observatorio Marino gestionado por el centro VLIZ (Ostende, Bélgica)**

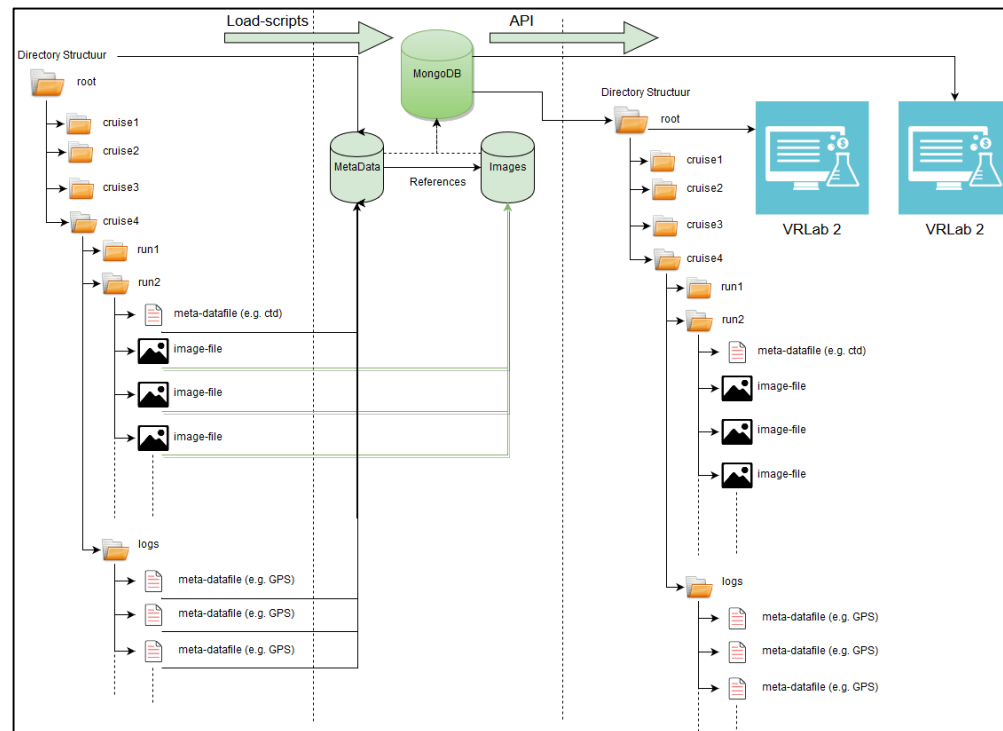
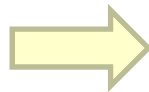
# Ejemplo: Observatorios Marinos

## Sincronización y accesibilidad a los datos

La base de datos en MongoDB de VLIZ e IFCA son accesibles desde los “Virtual Labs” basados en Rshiny/Rstudio: explorador de datos de LifeWatch data: <http://rshiny.lifewatch.be/ZooScan>

## Acceso a los ficheros en MongoDB

El laboratorio virtual debe dar acceso también a los ficheros individuales generados por los diferentes instrumentos/sensores



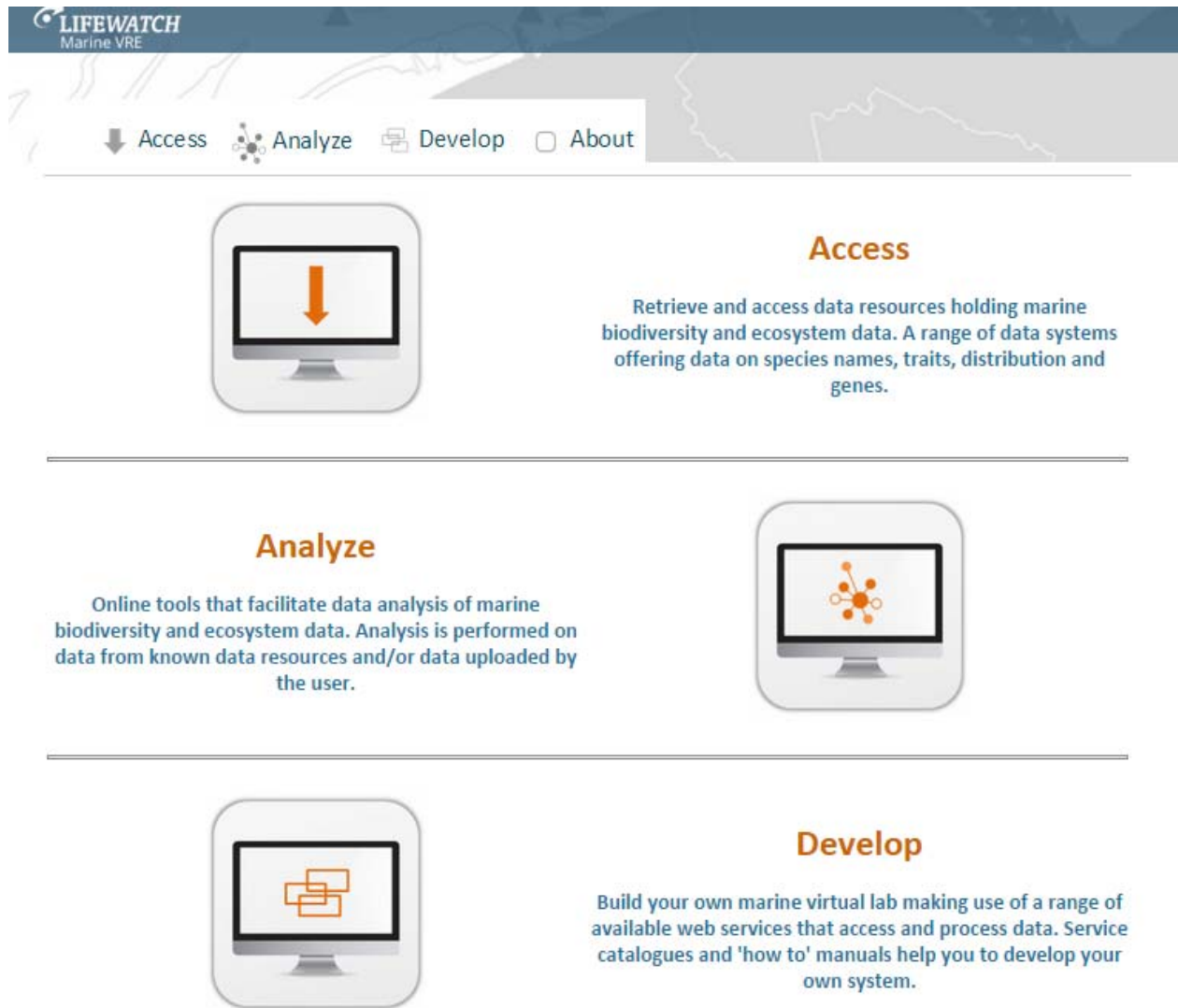
Acceso a los datos en SQL: <http://rshiny.lifewatch.be/>

Acceso a los datos a través de servicios web via Geoserver

Usando **Geoserver-cluster** se puede acelerar la respuesta para los usuarios!



# *Ejemplo: Observatorios Marinos*



Acceso a los datos

Análisis de los datos

Desarrollo de  
“Laboratorios Virtuales”

### 3. LW-ERIC: PRODUCTS AND SERVICES

**DATA RESOURCES:** MoUs directly established with few international initiatives, e.g., GBIF, ELTER, & through national nodes



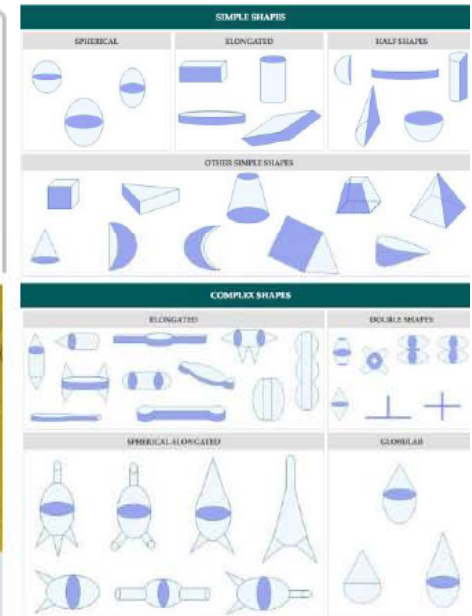
WoRMS Currently holds:  
239,117 accepted species, of which  
93% checked [more stats]



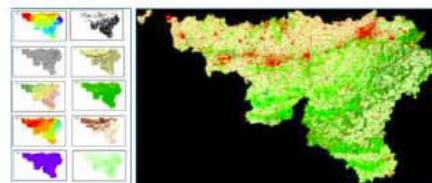
Catalogue of Life (CoL)



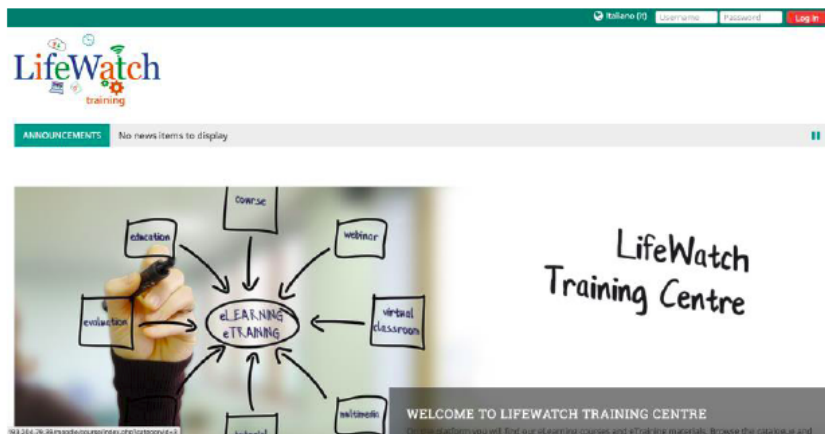
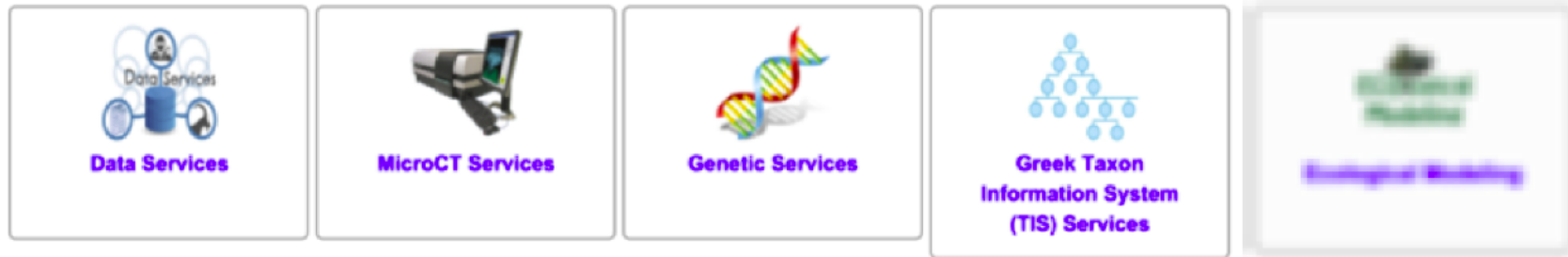
FishBase



LifeWatch-WB ecotope database



# 3. LW-ERIC: PRODUCTS AND SERVICES



## BioMaS

Bioinformatic analysis of Metagenomic AmpliconS



:GI-LIFEWATCH / INDIGO-WP2 Case Study

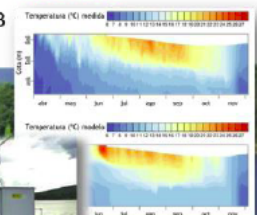
## Monitoring & Modeling ALGAE BLOOM in a Water Reservoir

LIFE+ Project lead by a SME, collecting monitoring data (environmental station+ water quality and chloro-cyano profiler), and modeling hydro+bio

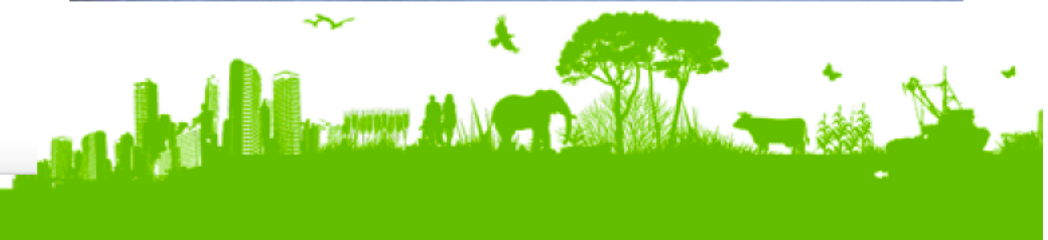
- **INTEGRATE EXISTING OPEN RESEARCH DATA**
  - USE **METEO**, **TERRAIN**, **BATHIMETRY**, **LAND USE**
  - **HYDROLOGICAL INPUT**
- **PRESERVE NEW OPEN RESEARCH DATA & ANALYSIS**
  - REMOTELY COLLECTED DATA INTO REPLICATED DB
  - COMPLEX MODEL OUTPUTS
  - MULTIPARAMETRIC ANALYSIS

**STORAGE+COMPUTING  
NEED HPC FOR DELFT-3D**

**Model already running in FedCloud  
Adapt Multiparametric scan  
Preserve Thermoclines analysis (in R)**



An Innovative Fresh Water VRE for I





# ¿Cómo desarrollar las soluciones TIC?

## ✚ **Requerimientos reales** de aplicaciones **realistas!**

- ✚ Que cubran tanto investigación básica como gestión
- ✚ Diferentes ámbitos (marino, fluvial, terrestre...)
- ✚ Inter-disciplinario, a diferentes escalas
- ✚ **Basándose en un catalogo de soluciones en abierto**
- ✚ **Beneficiándose de la e-infraestructura de LifeWatch**
  - Miles de procesadores, cientos de terabytes

## ✚ “People in the middle”: expertos BIO + expertos TIC

- ✚ ¿Es sostenible?

## ✚ Soluciones “amigables”

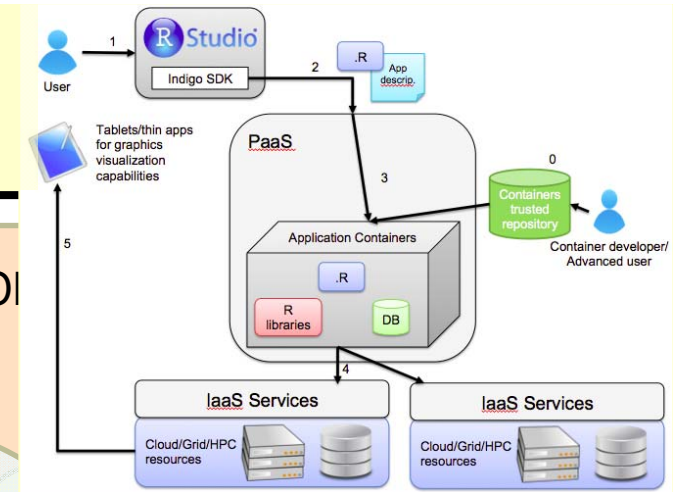
- ✚ Desde la **“identificación”**... hasta la visualización

## ✚ Workflows? VRE (Virtual Research Environments)? V-Labs?

## ✚ En colaboración con otras iniciativas...



# Piloto CLOUD



COMPOSITION

e-INFRASTRUCTURE

RESOURCES

"External"  
Data

"Internal"  
Data

SITE

Network

Storage

Servers

SITE  
...

4-Plataforma  
de publicación  
y preservación  
de datos en  
abierto

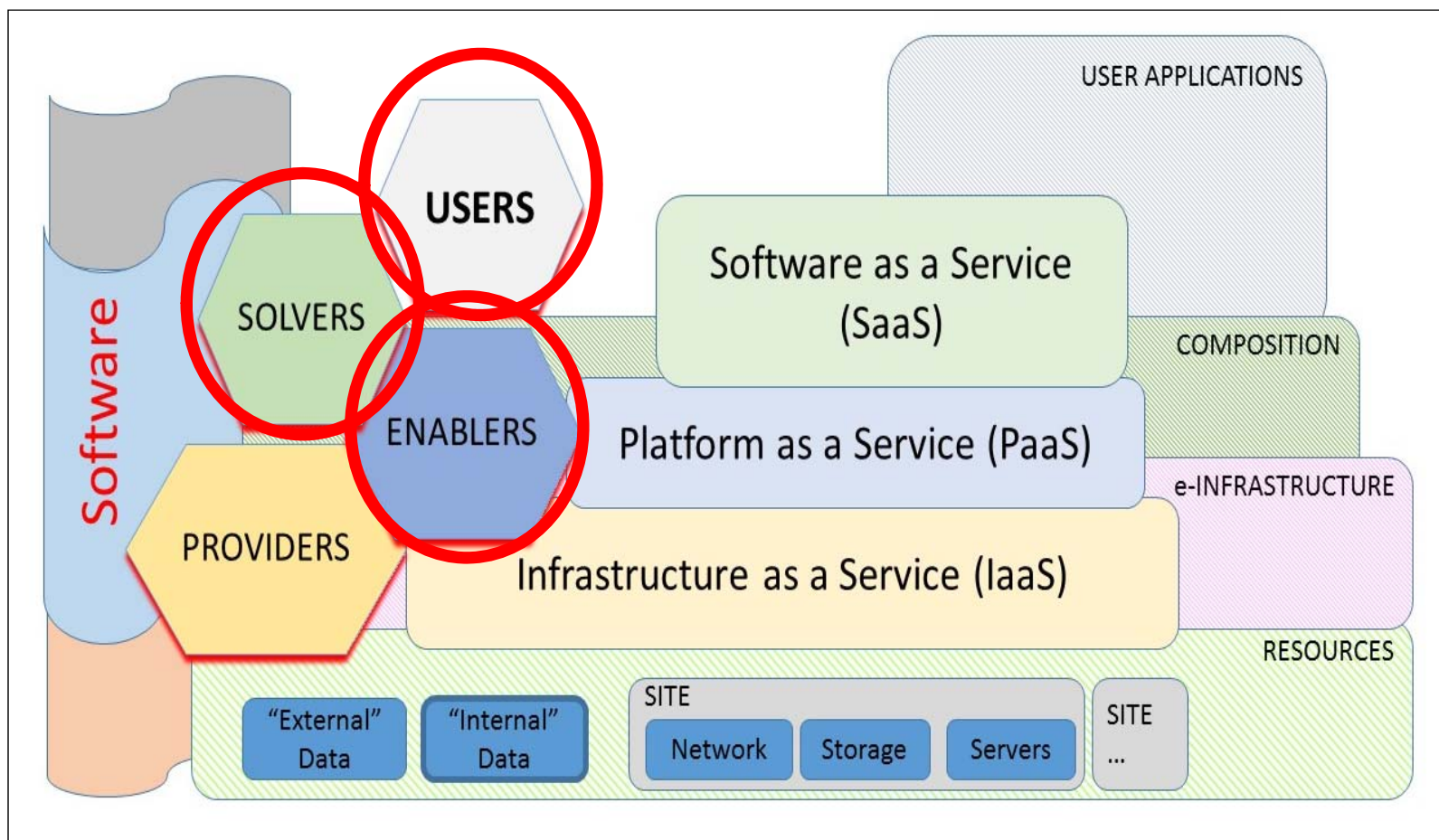
1-SOA/ Cloud  
Computing

3-Plataforma  
de trabajo  
colaborativo

2-Plataforma  
de control  
distribuido



EOSC = EUROPEAN OPEN SCIENCE CLOUD





## TRUFA (<http://trufa.ifca.es/>)



A web-based RNA-seq application

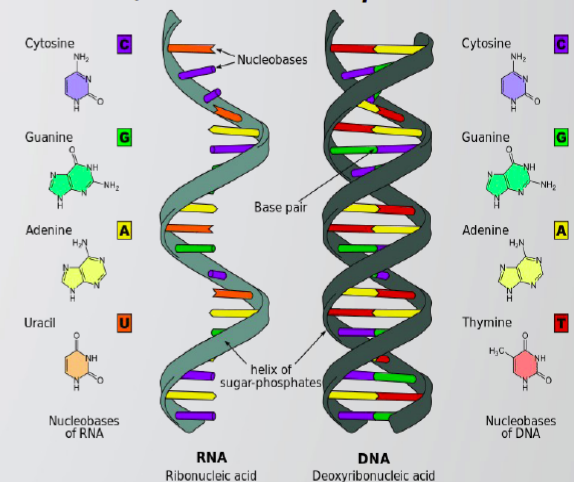
**RNA-seq is used to analyze the continually changing cellular transcriptome**

→ *RNA comparison* between healthy and unhealthy organisms can lead to *better understanding* diseases such as *cancer, diabetes, alzheimer...*

→ Also useful for *comprehending several biological mechanisms* (i.e venom production, silk production...)

**Application of next-generation sequencing (NGS) methods for transcriptome analysis (RNA-seq) has become increasingly accessible** in recent years and are of great interest to many biological disciplines including, *evolutionary biology, ecology, biomedicine, and computational biology.*

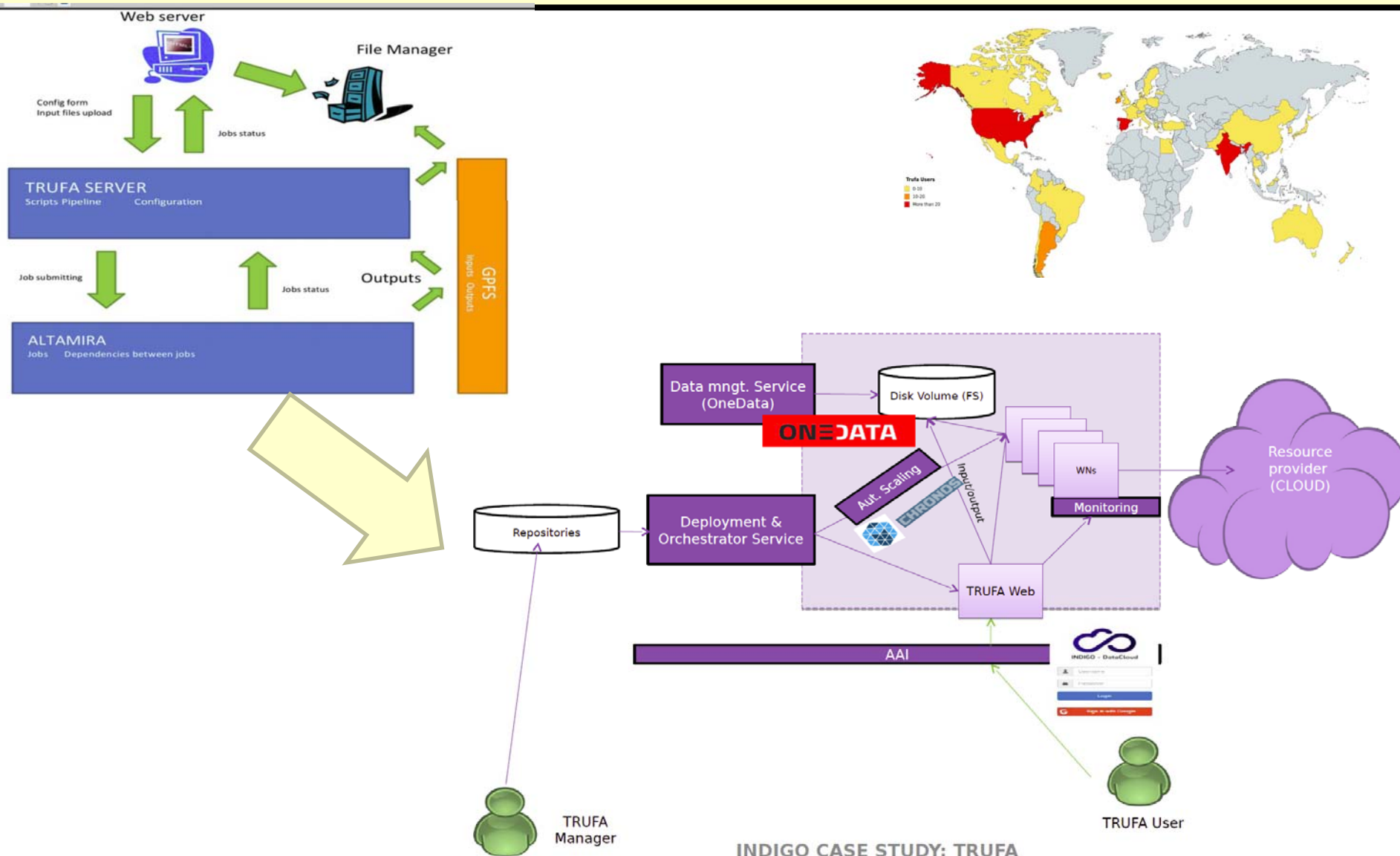
Although virtually any research group can now obtain RNA-seq data, **only a few have the bioinformatics knowledge and computation facilities required for transcriptome analysis.**



INDIGO CASE STUDY: TRUFA



# Adaptando TRUFA al EOSC



INDIGO CASE STUDY: TRUFA



# GBIF.es & GBIF.pt

**-Esfuerzo conjunto de**

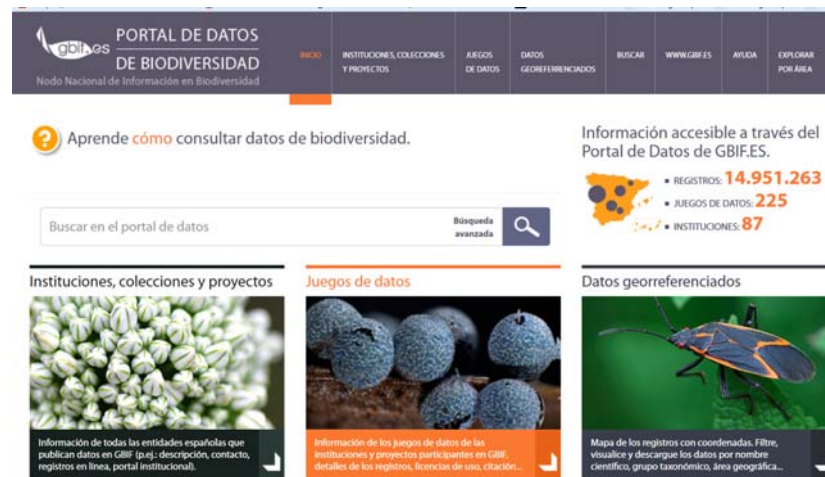
- GBIF.es (RJB), IFCA (ES)
- GBIF.pt, LIP(PT)

**-Implementación de componentes del “Atlas of Living Australia”**

-más de 15M de registros específicos

-una comunidad de más de 10.000 investigadores

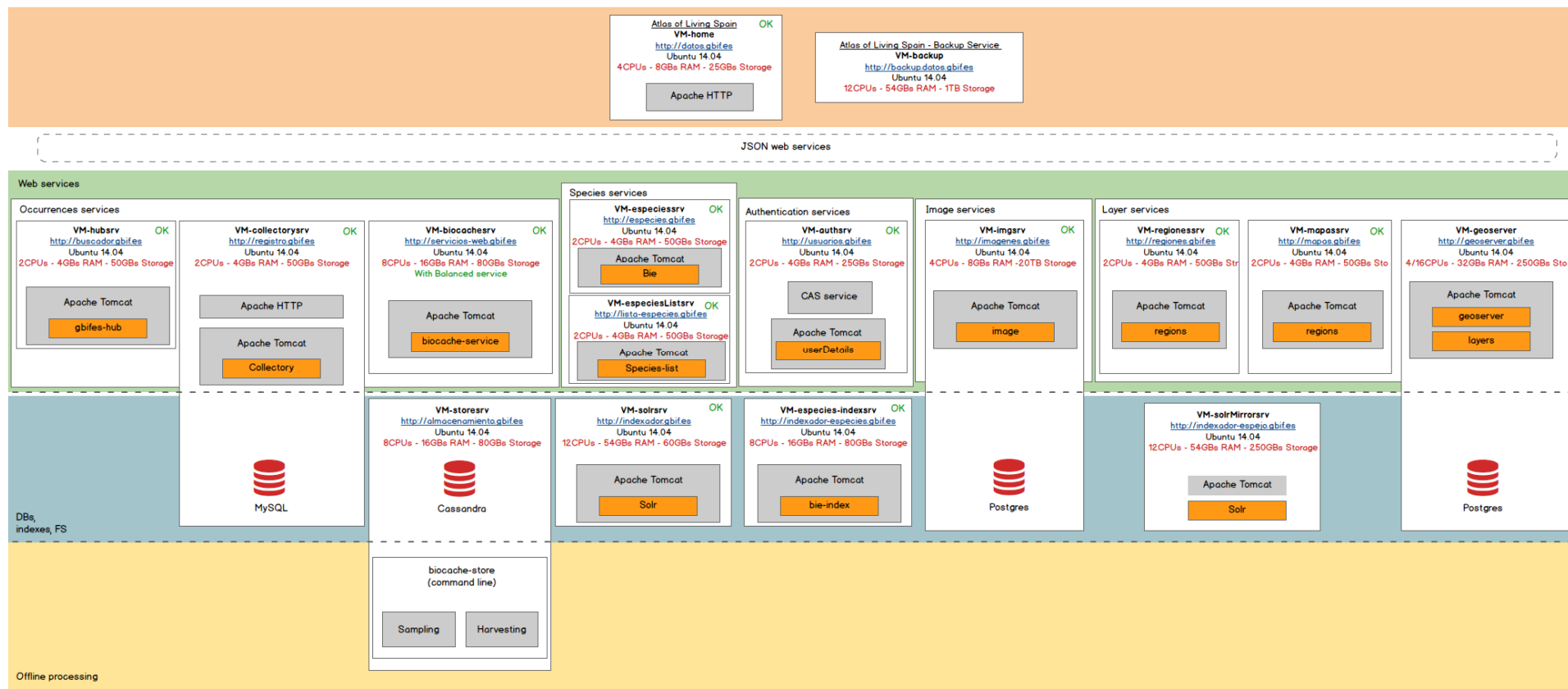
**-Uso creciente de recursos LifeWatch en el Cloud Federado (en el IFCA)**







# Despliegue de GBIF.es



# Sobre Ciencia Ciudadana

*The best data for the best answers*  
(slide courtesy of F.Pando, Royal Botanic Garden Madrid)

## 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



*offers a tremendous increase in data available!*

- Aplicación “open source”
- Colaboración CREAM, RJB, IFCA (CSIC)  
(idea impulsada desde GBIF.es)
- extiende la funcionalidad de iNaturalist
  - Proyectos didácticos específicos
  - Comunidad hispano hablante
- más de 1300 usuarios
  - muchos estudiantes
  - también aficionados expertos!
- sobre recursos de Fedcloud
- en uso:
  - escuelas e institutos
  - agencias medio ambiente
- pronto (o ya) : Parques Nacionales!





## 🌀 DEEP LEARNING:

- ❏ Entrenar una red neuronal convolucional (CNN) sobre 200.000 imágenes de plantas identificadas (PlantNet) para ser capaz de reconocer con una foto de que planta se trata
- ❏ Requiere hardware especializado (GPUs)
- ❏ Trabajo presentado en el reciente congreso ACM (Siena 2017)
- ❏ PERO LO MEJOR ES PROBARLO: <http://deep.ifca.es>

## 🌀 Combinar Deep Learning con aplicaciones:

- ❏ Zooscan de Marine VRE: reconocer plankton
- ❏ Natusfera: ayudar a identificar plantas
- ❏ ... aves, etc.



# Una idea sencilla...

## DEEP Learning + Ciencia ciudadana versus Invasoras

- El ciudadano saca una foto con el móvil
- La aplicación contrasta con la red (entrenada en invasoras)
- Identifica las 5 primeras opciones,
- e indica si alguna es invasora...
- Reporta al “proyecto”

## Prueba inicial con la red actual:

|                  | Num especies tot | Num especies in our DB | Num test images | Top1 (%) | Top5 (%) |
|------------------|------------------|------------------------|-----------------|----------|----------|
| Picos de Europa  | 12               | 9                      | 2802            | 30       | 48       |
| Garajonay        | 30               | 25                     | 7714            | 41       | 57       |
| EEI establecidas | 173              | 116                    | 30602           | 39       | 57       |
| EEI lista negra  | 128              | 64                     | 16391           | 35       | 50       |
| EEI urgentes     | 35               | 31                     | 7587            | 42       | 57       |

### Mejor identificadas

Vinca difformis [49 images in DB]  
Lantana camara [320 images in DB]  
Datura stramonium [301 images in DB]  
Agave americana [147 images in DB]  
Tropaeolum majus [147 images in DB]  
Phytolacca americana [220 images in DB]  
Amaranthus retroflexus [41 images in DB]  
Asclepias curassavica [74 images in DB]  
Carpobrotus edulis [106 images in DB]  
Eriobotrya japonica [368 images in DB]

- ✿ Lifewatch en España ha constituido una “red” (JRU)
- ✿ Se colabora entre diferentes instituciones
  - ✦ Proyectos específicos
    - Adquisición de datos, modelado, simulación...
  - ✦ Nuevos proyectos internacionales
  - ✦ Integración de datos en abierto (Data life cycle)
  - ✦ Integración de “servicios”
  - ✦ Apoyo en e-infraestructura
    - Almacenamiento
    - Computación
    - Hardware específico
- ✿ Formación (en ambas direcciones)





# Master en Data Science



## 1.- FUNDAMENTOS

(OCTUBRE-ENERO)

Módulo obligatorio, cinco materias

| Panorama de Data Science                      | Métodos de Data Science                      | Data Management                                                                                          |
|-----------------------------------------------|----------------------------------------------|----------------------------------------------------------------------------------------------------------|
| <b>Introducción a Big Data y Open Science</b> | Estadística para Data Science<br>Data Mining | Modelos de datos y sistemas de información<br>Ciclo de vida de los datos: de adquisición a presentación. |

## 2.- MÓDULO DE ESPECIALIZACIÓN

(ENERO-ABRIL)

Un área de especialización a elegir:

| Data Science Analytics                   | Data Science Engineering              | Open Data Management                     |
|------------------------------------------|---------------------------------------|------------------------------------------|
| Machine Learning I                       | Sistemas de Computación para Big Data | Acceso a los Datos: Portales y Servicios |
| Machine Learning II                      | Cloud para Data Science               | Preservación de Datos                    |
| Semantics, Linked Data, Text Data Mining | Desarrollo de Proyectos               | Repositorios de Datos                    |

## 3.- MÓDULO PROFESIONAL

(ABRIL-JUNIO)

Este módulo obligatorio incluye:

**Seguridad, Privacidad y Aspectos Legales**

**Nuevos Desarrollos en Data Science (seminarios)**

## 4.- MÓDULO DE ORIENTACIÓN PROFESIONAL

(MAYO-SEPTIEMBRE)

De acuerdo a su interés profesional y cualificación, el/la estudiante podrá optar por realizar prácticas externas y/o "Data Labs" en diferentes áreas

| Data Labs                                                  |                              |                     |
|------------------------------------------------------------|------------------------------|---------------------|
| Biomedicina                                                | Medio Ambiente, Meteorología | Física y Astronomía |
| Economía y Finanzas                                        | Internet of Things           | Ciencias Sociales   |
| + Prácticas Externas en empresas y grupos de investigación |                              |                     |

## 5.- TESIS DE MÁSTER

(Comienzo en MAYO, presentación en SEPTIEMBRE)

Un trabajo avanzado desarrollado de forma autónoma por el estudiante bajo la supervisión de un profesor del Master. La temática y orientación dependerán de la especialidad elegida. Supondrá un trabajo de iniciación al contexto profesional que permita unirse en el futuro a una empresa o a un grupo de investigación.

**Se promoverá que la tesis se realice asociada a un contrato externo remunerado en las empresas colaboradoras o grupos de investigación.**