



## EGI Services for LifeWatch

EGI towards Horizon 2020 Workshop Session: User requirements for data services



Input from J.M. González Aranda, W.Los, A.J.Saenz, presented by Jesus Marco (marco@ifca.unican.es) Instituto de Física de Cantabria (IFCA) CSIC, NATIONAL RESEARCH COUNCIL SANTANDER, SPAIN



## What is LifeWatch???

### LifeWatch is an ESFRI initiative in biodiversity and ecosystem research

An *exploratory research environment* that allows scientists to find data, combine resources, compose workflows, run analyses, develop models and visualize predictions.

User-friendly services to combined data, using established standards and unique identifiers are developed that can be further combined into work flows for different types of questions.



### http://lifewatch.eu/

Flexibility to be offered through grid- or cloud-based computational services. While the facility takes care of access and matching of resources 'under the hood', scientists and other users can focus on the analyses.







## Scientific challenges





From Wouter Los presentation in Madrid

### AN EXAMPLE SELECTED ON PURPOSE:

## **Monitoring Cyanobacterial Blooms**

ECOHYDROS SL for Confederación Hidrográfica del Duero (CHD)

DORII (FP7 project with IFCA-CSIC), ROEM & ROEM+ (Avanza & LIFE projects with ITG)



#### **Advanced water quality measurement**

#### **Relevant environmental measurements**

Atmospheric-meteorological variables Hyperspectral solar radiation in water Nutrients

**High sampling frequency** (minutes/hours) + data transmission to cloud services (replicated database, on-line monitoring system)

Mobility (GPS integrated) plus **Profiling** (wincher programmed/remote down to 50m)





(Coterillo, 2012).

#### **ONLINE ACCESS (data, profiles, images) USING CLOUD RESOURCES AT IFCA**

←



#### **THE CHALLENGE:**

## **INTEGRATION**

## of information and models

### using e-infrastructures

### for advance water reservoir management

#### **INFORMATION:**

**-DMP (Data Management Plan)** integrating internal and external information, key for preservation (PID).

-GIS system

#### **MODELS:**

-DELFT3D (open model) Water reservoir modeling *Running: Cloud + Supercomputer* -physical (abiotic) VALIDATED ! -water quality (biotic) Working on it DECISION/MANAGEMENT TOOLS

**NEXT STEP!** 



# Our scheme must take into account our stakeholders: researchers, citizens & managers



# *My naive view of the process to publish a research paper or complete a report*

NASA, ESA

- Data Taking A: Biodiversity specimens observation/collection
  - Manual or Automated (Instrumentation?)
  - Professional or Amateur/Citizens for Museums/Research centers/Organizations
  - Occasional/Campaign or Systematic/Monitoring
  - Integration of annotations/previous references
- Data Taking A': Genomic information
- Data Taking B: Environment
  - From external remote monitoring (satellites, radar, LIDAR, etc.)

**GBIF** 

- From in-place monitoring (basic to to proper sensors and probes, cameras, spectrometers, etc.)
- Data Integration
  - Collections, Papers
  - Databases, Maps
- Data Curation: Identification/Classification, Taxonomy, Integration in GEObase
- Model
  - Specimens evolution, niches, interaction, etc.
  - Impact of changes (eg. Environmental, human activities)
- Validation, Publication/Report and Design of new experiments

## Better global image (Vince Smith @ BIH2013)

### An informaticians view of biodiversity



- Many audiences (policy makers, scientists, amateurs, citizen scientists)
- Many scales (global solutions to local problems)

Figure adapted from Peterson et al 2010

## Indeed, a long list of initiatives

- LlfeWatch & National LW Initiatives
- LTER-Europe, LTSER (supported by ALTER-Net)
- GBIF, TDWG
- 🕸 RDA
- IPBES (Intergovernmental Platform on Biodiversity & Ecosystem Services)
- FAO interest for fisheries and agriculture, AG-Infra, i-Marine, FLOD,
- GEO Ecosystems
- GEOBON genomic layer
- Biosos Earth Observation / EBONE; NATURA2000 sites
- General Ecosystem Models (Predicts, BioVel)
- Ecological Observatories & Genomic Observatories
- Biocode / BiSciCol: VertNet/Genbank
- Microbiome project
- Local Ecological Footprint Tool, Connectivity:www.groms.de
- Ecological Index, BICT, Vibrant
- Catalogue of Life
- Traits: integration of pheno and genotypic data; Phenotype Ontology Research Coordination Network
- BiodiversityDataJournal
- Integrating information using OCR / Vibrant

- Service Networks, Service Sets (deployed on e-Infra) and Biodiversity Catalogue: Integrated Virtual Environment (IVE) for Biodiversity Science (Creative-B)
- Workflows and provenance (Wf4ever, SCAPE)
- Virtual Research Environments (i-Marine, D4Science, gCube)
- Scratchpads (websites for taxonomists)
- OpenAgrid / Agrovoc; data.fao.org
- EnvEurope (semantics and data)
- COMPSs: programming framework for distributed infrastructure
- EUBrazilOpenBio Ecological Niche Modeling Service
- EUBrazilCloudConnect
- New tools for environmental monitoring (Acoustic, Trackers...)
- AAA solutions
- Long Term Preservation (Rebind)
- Ocean Sampling Day
- GeoBroker & A Broker Framework for Next Generation Geoscience (BCube)
- FreshWaterBiodiversity (Mobilizing data and constructing data networks)
- pro-iBiosphere
- PESI
- EUBON
- GN (Global Names)



## Focus on a challenge

- Grand Challenge: Predictive Modeling of Biosphere
  - Global Carbon cycle
  - Essential Biodiversity Variables (EBV) for IPBES
    - IPBES=Intergovernmental Platform on Biodiversity & Ecosystem Services (cf. IPCC)
    - EBV= a measurement required for study, reporting, and management of biodiversity change.
  - Examples of candidate EBV:
    - Species populations: Abundances and distributions (inc. invasive alien)

		LAAMI LLS OF CAN		SSEIVITAL DIODIVERSITI VI	AKIADLES	
EBV class	EBV examples	Measurement and scalability	Temporal sensitivity	Feasibility	Relevance for CBD targets and indicators (1,9)	
Genetic composition	Allelic diversity	Genotypes of selected species (e.g., endangered, domesticated) at representative locations.	Generation time	Data available for many species and for several locations, but little global systematic sampling.	Targets: 12, 13. Indicators: Trends in genetic diversity of selected species and of domesticated animals and cultivated plants; RLI.	Pereira et al., Science 2013
Species populations	Abundances and distributions	Counts or presence surveys for groups of species easy to monitor or important for ES, over an extensive network of sites, complemented with incidental data.	1 to >10 years	Standardized counts under way for some taxa but geographically restricted. Presence data collected for more taxa. Ongoing data integration efforts (Global Biodiversity Information Facility, Map of Life).	Targets: 4, 5, 6, 7, 8, 9, 10, 11, 12, 14, 15. Indicators: LPI; WBI; RLI; population and extinction risk trends of target species, forest specialists in forests under restoration, and species that provide ES; trends in invasive alien species; trends in climatic impacts on populations.	
Species traits	Phenology	Timing of leaf coloration by RS, with in situ validation.	<mark>1</mark> year	Several ongoing initiatives (Phenological Eyes Network, PhenoCam, etc.)	Targets: 10, 15. Indicators: Trends in extent and rate of shifts of boundaries of vulnerable ecosystems.	
Community	Taxonomic	Consistent multitaxa surveys and	5 to >10	Ongoing at intensive monitoring sites	Targets: 8, 10, 14.	

Examples of candidate Essential Biodiversity Variable

## A Reference Model?



## "Simple" IT Reference Model



#### Collaboration

- Common Exploratory Environment
- Collaborative Virtual Organisations

#### Workflow development

- Semantic Matching
- Visualisation

#### **Analysis & Processing**

- Integration of resources
- Quality controls
- Grid computation

#### Data

- Existing measurements & observations
- Real-time sensor networks (earth based and remote)
- Other infrastructures

## Core ICT (e-)Infrastructure

- Essential 'central' components
  - Single portal access for all users
  - Datasets & services / tools catalogues
  - Access to computational resources
  - Security (AAA)
  - Provenance and citation tracking
  - Annotations
  - Virtual Collaborative Environments / VO / BTCN
  - Workflow composition, execution and management
- Data & tool resources
  - New data resources to be 'admitted'
  - Statistical, analytical & modelling tools
- Innovation Lab
- Intellectual property management

www.lifewatch.eu

### Solution for HETEROGENEITY: An SOA approach







A.J.Saenz, Rome EUDAT meeting

## EGI services for LW?

- MODEL: LW brings users & resources together!
  - **LW core-ICT (Spain) will operate an e-infrastructure in 2014** 
    - LW core-ICT could/will integrate grid/cloud infrastructure in EGI
    - LW VOMS will be supported by LW core-ICT
    - LW core-ICT will rely on IberGrid for this integration in EGI
  - **LW** national initiatives will be integrated
    - LW core-ICT will support integration at different levels (NGI role?)
  - **LW** will explore successful examples in EGI FedCloud:
    - EUBrazilOpenBio Ecological Niche Modeling Service
    - EUBrazilCloudConnect
    - New challenge for phenology with LTER/Univ.Granada
- So, LW will use existing EGI services
  - But LW needs additional services...



## Additional services

- Additional services (related to EGI) are being studied:
  - Considering output of ongoing projects
    - EUDAT, ENVRI, BIOVEL, COOPEUS, IMARINE, CREATIVE-B
  - Some of them may be implemented with EGI support/collaboration
    - Identity federation for researchers, educators and students
    - Digital Identifier e-Infrastructure for digital objects (and PID issues)
    - Simple Storage/File System + Medium/Large DBMS cloud/grid instances
    - Large, persistent DBMS, GIS systems in cloud/grid framework
    - Parallel (multithread?) datamining ( in phytom OPR) cloud/grid instance
    - Systems to handle & process real time reams
    - Access to large databases/directives common to other research areas
    - Workflows connecting to HPC resources (o(10<sup>2</sup>-10<sup>3</sup>) processes, 1-100 TB)
    - Support to virtual eLaboratory
    - Data discovery and access

Along 2014 we will work to complete a VRE proposal



## THANKS!



