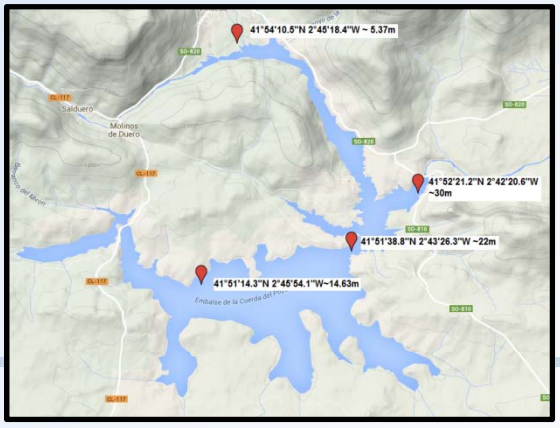
An Integrating Fresh Water VRE for LifeWatch

(ongoing work)

Showcasing tools and services from Research Infrastructures at EGI CF 2015 in Bari



Jesus MARCO DE LUCAS . Fernando AGUILAR , Daniel GARCÍA, Guadalupe CAÑAS (IFCA)

Agustin MONTEOLIVA, Tamara SANTIAGO , Jose Augusto MONTEOLIVA (ECOHYDROS SL)

Ana Yaiza RODRIGUEZ (AEONIUM)

Juan Miguel GONZALEZ ARANDA (MINECO)
AJ SAENZ (CITIC-CENTA)

JRU LW ES

www.egi.eu

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

SCOPE

An Integrating Fresh Water VRE for LifeWatch

- IFCA Experience on Fresh Water systems
- The current framework in LW: VRE, Cloud, existing examples
- An Integrating VRE following the Data Life-Cycle:

  Building an Open Science Framework
- The “project” orientation
- The hard question: integrating Digital Knowledge

EVOLUTION

Our interest on Fresh Water systems

- **2005: Start of IFCA-ECOHYDROS (SME) collaboration**
  - **Modeling for Itoiz water reservoir,** *CWM models (ELCOM, DYRESM+ CAEDYM)*

  Modelling of A Watershed: A Distributed Parallel Application in a Grid Framework
- **2008: DORII FP7 project**
  - **Monitoring platform @CdP, Instrumentation, Labview, Web Services, Grid Instrument Element**
- **2012: ROEM+ (LIFE+ project)**
  - **Extension of monitoring system**
  - **Start of use of Delft3D for hydrological and water quality (algae bloom)**
- **2013: Connecting with LifeWatch**
  - **Presentation to EGI**
  - **SCARCE Int.Conference, NETLAKE**
- **2014: Extension of monitoring network**
  - **Cogotas water reservoir: new monitoring station connected via 3G**
- **2015: Integration in the LW framework**
  - **Observatory in EGI-Engage LW CC**
  - **CdP & SWIRL presented at Ecology at Interface (Rome)**
  - **Case Study for Structural Funds project**

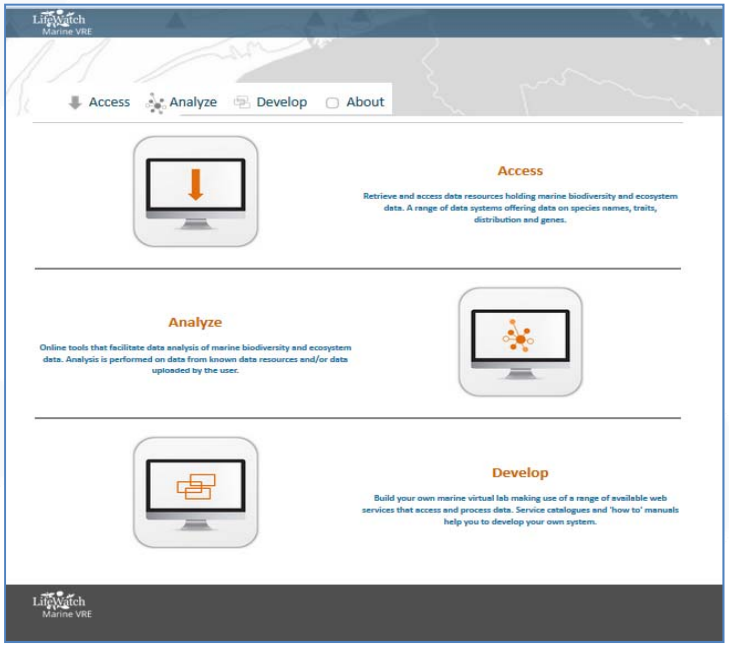
FRAMEWORK

- **LifeWatch (lifewatch.eu) is an ESFRI** *(EU Research Infrastructure)*
  - **Addressing Biodiversity & Ecosystems**
  - **An e-Infrastructure to build Virtual Research Environments (VRE)**
  - **Integrating OPEN DATA information**
    - **GBIF, LTER, GENBANK, SATELLITE IMAGES, TERRESTRIAL MAPS…**
- **EGI-LifeWatch Competence Center**
  - **Framework: EGI FedCloud**
  - **Dedicated Resources (~5000 cores + PB,** new node in Seville, Andalusia, SPAIN **)**
- **Support LW VRE**
  - **Marine VRE (marine.lifewatch.eu)**
  - **Terrestrial + FreshWater VRE**
- **Pilot projects**
  - **Ecological Observatories Data Flow and "Big Data" analysis**
  - **Workflows: Galaxy and TRUFA; Network of Life**
  - **Citizen Science: Assisted Pattern Recognition**

Exploration of VREs in LW

- **LifeWatch Marine VRE**

See LW EGI CC session @ Lisbon

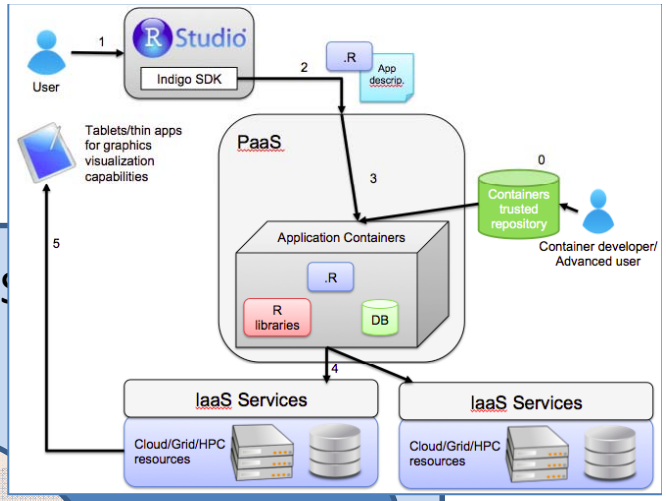


- **Terrestrial-Fresh Water VRE**
  - Both SWIRL and CdP are included

See LW EGI CC session @ BARI (TOMORROW!)

- **Invasive Species VRE**

Global Scheme



USER APPLICATIONS

Portals

Visualization

COMPOSITION

e-INFRASTRUCTURE

Platform for Collaborative Framework

Open Data & Preservation Platform

Distributed Control Platform

SOA/ Cloud Computing

RESOURCES

"External" Data

"Internal" Data

SITE

Network

Storage

Servers

SITE

...

- Human impact in Spanish basins
  - Dams, reservoirs, etc.
- Changes in the environment due to new uses of the ground and built of infrastructures
- Need of ensuring certain water quality and flow levels.
- Swirl combines 2000 different data sources (flow, purity, pluviometry rate, etc.)

- Open(*) Resources and Products
- LifeWatch Standard Architecture
- 1700 Web Services in Catalogue
- Focused on Science-based Exploitable Scenarios
- Iterative-Agile Integrations, Deployments and Results
- Integrating Researchers, Environmental Managers, Enterprises and Citizens.

open science

EGI-LIFEWATCH / INDIGO-WP2 Case Study

Monitoring & Modeling ALGAE BLOOM in a Water Reservoir

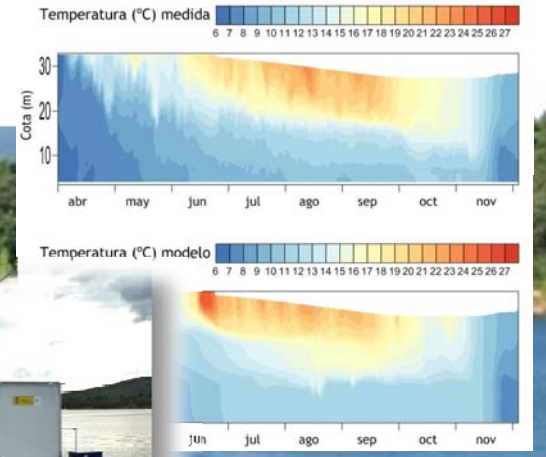
LIFE+ Project lead by a SME, collecting monitoring data (environmental station+ water quality and chloro-cyano profiler), and modeling hydro+bio

- **INTEGRATE EXISTING OPEN RESEARCH DATA**
  - **USE METEO, TERRAIN, BATHIMETRY, LAND USE**
  - **HYDROLOGICAL INPUT**

STORAGE+COMPUTING NEED HPC FOR DELFT-3D

- **PRESERVE NEW OPEN RESEARCH DATA & ANALYSIS**
  - REMOTELY COLLECTED DATA INTO REPLICATED DB
  - COMPLEX MODEL OUTPUTS
  - MULTIPARAMETRIC ANALYSIS

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



SOLUTIONS EXPLORED

See OPEN DATA Commons session @ RDA

- Support external resources (data, tools): **VRE**
- Enable a "/lifewatch/home" for each researcher/each community, accessible with ID via a **preservation portal**
- Users will define the "openness" of their
  - DATA (private/**embargo**/open/published-DOI)
  - ANALYSIS (R/python, via github)
  - WORKFLOWS at SaaS level (R,python)
- Support it with a global (federated) distributed storage
  - OneData (Data Commons basic component)
- Integrated also with FedCloud computing resources
  - We will rely on INDIGO project developments to optimize!
- Enforce DMP (Data Management Plan)

If it needs to be preserved => DMP & OPEN (after embargo)

AN INTEGRATING VRE FRAMEWORK

- Based on the idea of a Digital Knowledge Platform as a Framework for Open Science
  - Start from Data Life-Cycle
  - Dual View: Data & Software
  - Incorporate Digital Knowledge from start
  - Integration of external resources
  - Include Acquisition
  - Incorporate Analysis
  - &… Simplify Workflows
  - Enable Publication of Datasets
- Developed initially as a Open Data Preservation Portal
  - Using Structural Funds Project for LW in Spain
  - In collaboration with
    - Aeonium
    - Viavansis
    - Adevice
    - Telefonica

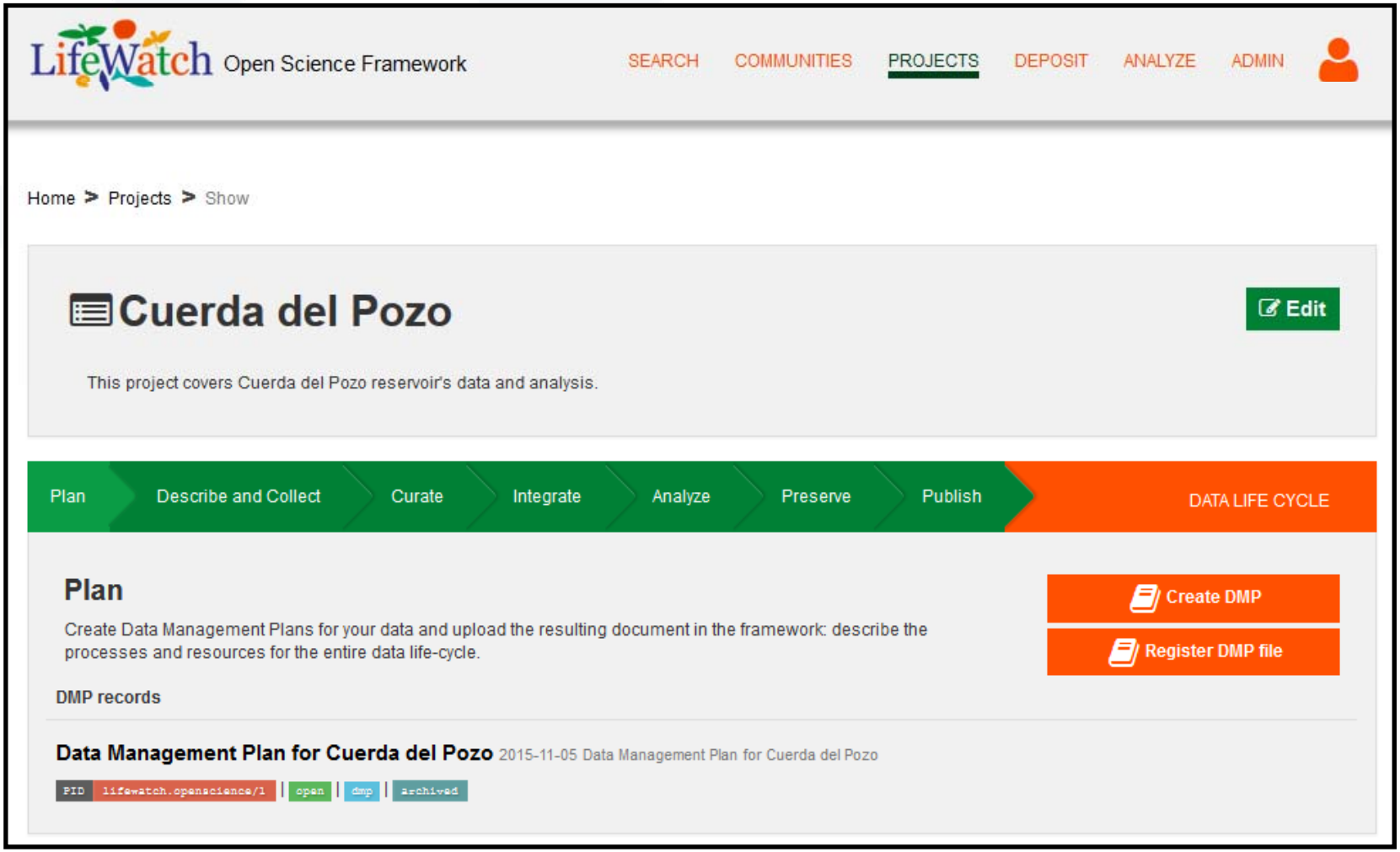
See DKP talk tomorrow in "Data without boundaries" session (16h20)

See "Setting up a new FedCloud site in collaboration with the industry" talk Thursday in "Community Cloud" session (15h35)

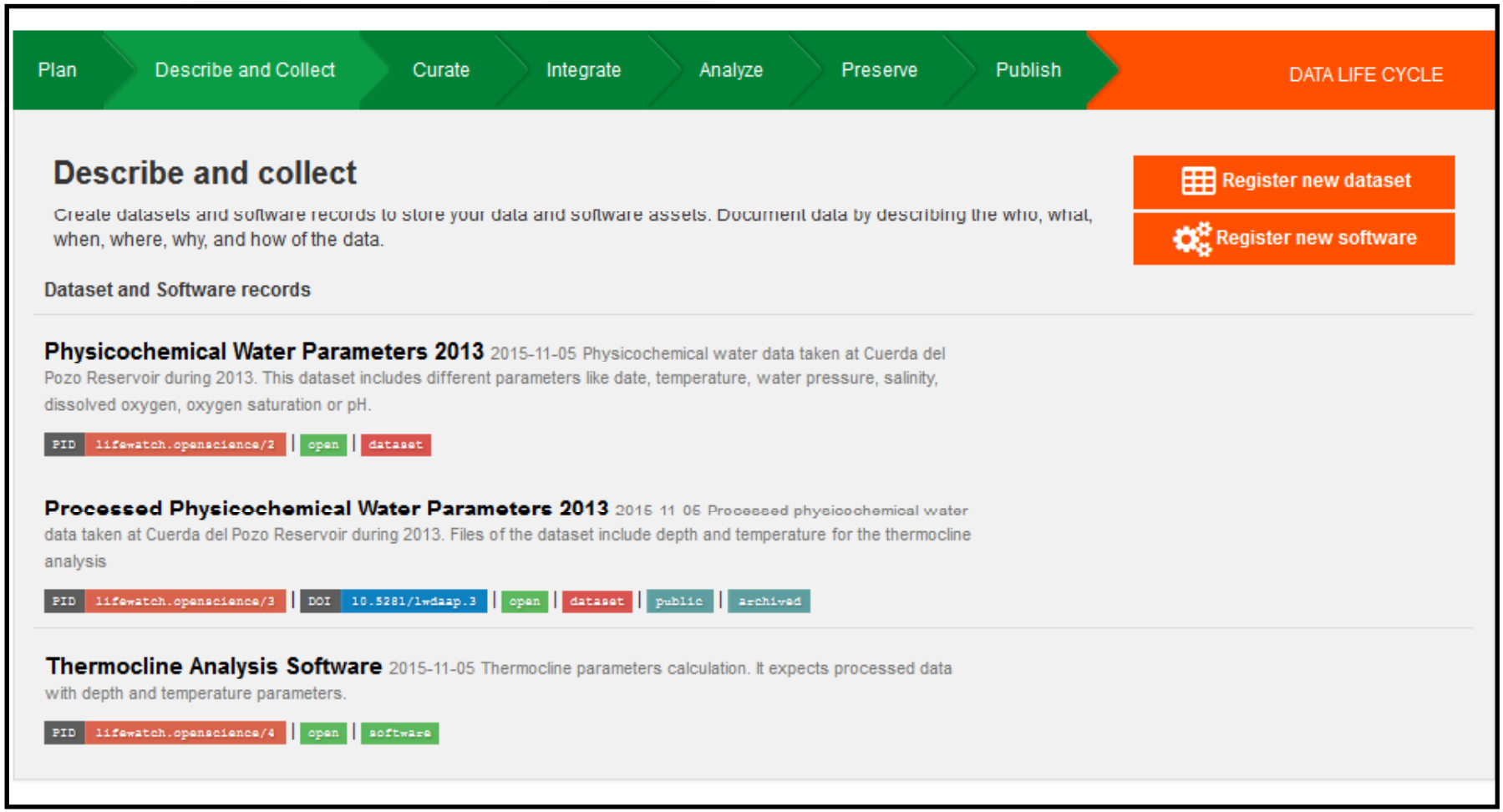
APPLIED TO CdP WATER RESERVOIR

- Start from Data Life-Cycle
  - DMP for CdP (already in INDIGO), DMP Online Tool
- Incorporate Digital Knowledge from start
  - Understand the use of “ontologies”
- Dual View: **Data & Software**
- Integration of external resources
  - Other measurements, maps & GIS…
  - Models Output (example: Delft3D, ABM based)
- Include Acquisition
  - Ex: replace LabView web services using pyVISA from Jupyter
  - Directly store into cloud enabled workspace (OneData)
- Incorporate Analysis
- &… Simplify Workflows
  - Jupyter/Python notebook Thermoclinae example
- Enable Publication of Datasets
  - Mint DoI

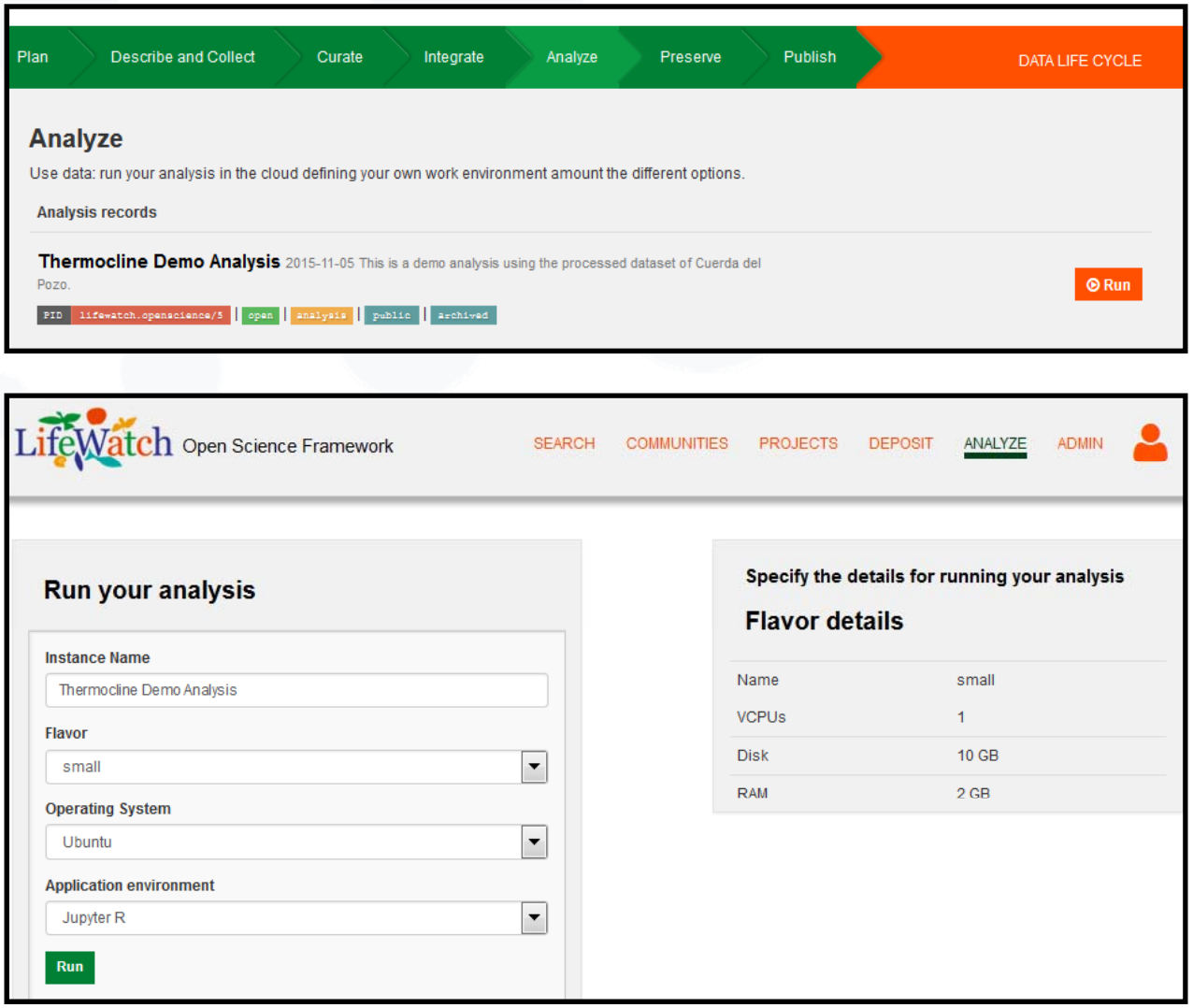
Projects follow the Data Life-Cycle: Plan



Projects follow the Data Life-Cycle: Describe



Projects follow the Data Life-Cycle: Analyze



Projects follow the Data Life-Cycle: Analyze

Your instances

Run an analysis

Instance Name	OS	Size	Env.	Status	Created	
Thermocline Demo Analysis	Ubuntu	small: 1 VCPU, 10 GB Disk, 2 GB RAM	jupyter-r	BUILD	11 seconds from now	Connect Terminate
Thermocline Demo Analysis	Ubuntu	small: 1 VCPU, 10 GB Disk, 2 GB RAM	jupyter-r	ACTIVE	5 days ago	Connect Terminate

File Edit View Insert Cell Kernel Help

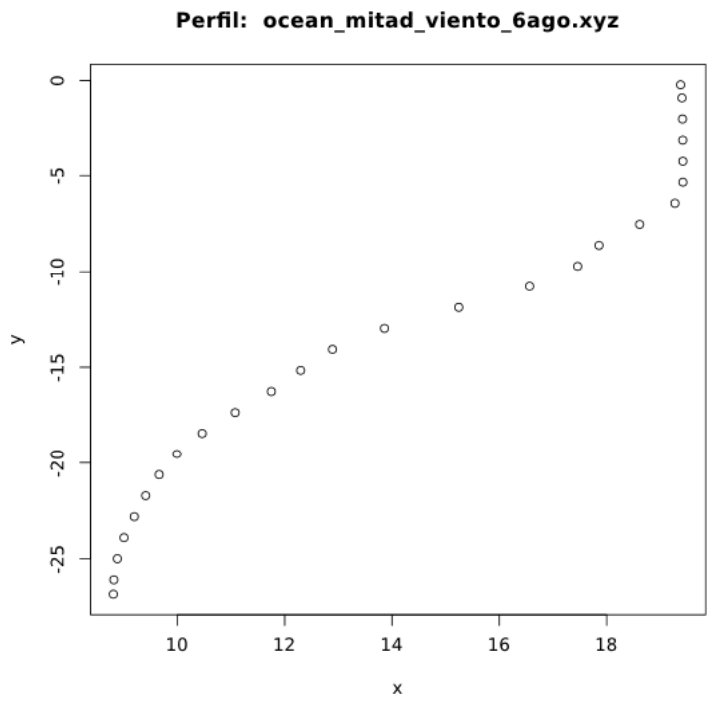
Code Cell Toolbar: None

In [6]:

```
getTe <- function(data){

    media = data$x[1];
    numElem = 1;

    for(i in 1:length(data$x))
    {
        if (data$y[i] > -1)
        {
            media = media + data$x[i]
            numElem = numElem + 1
        }
    }

    return (media/numElem)
}

min.RSS <- function(data, par) {
    r <- with(data, sum((x-(Th+((Te-Th)/((1+((par[1]*(-y))^par[2]))^(1-1/par[2])))))^2))
    return (r)
}

mod.RSS <- function(data, par) {
    with(data, (Th+((Te-Th)/((1+((par[1]*(-y))^par[2]))^(1-1/par[2])))))
}


paramTermoclina <- function(file){

#Esto es una ñapa que hacia falta solo en las salidas de los modelos, porque el nivel de agua es relativo,
#pero con los valores reales no es necesario porque la profundidad que está almacenada en la base de datos
    water_level = 0
    if (grepl("murakami",file))
    {
        if(grepl("29jun",file))
        {
            water_level = 1.02
        }
        if(grepl("22jul",file))
```

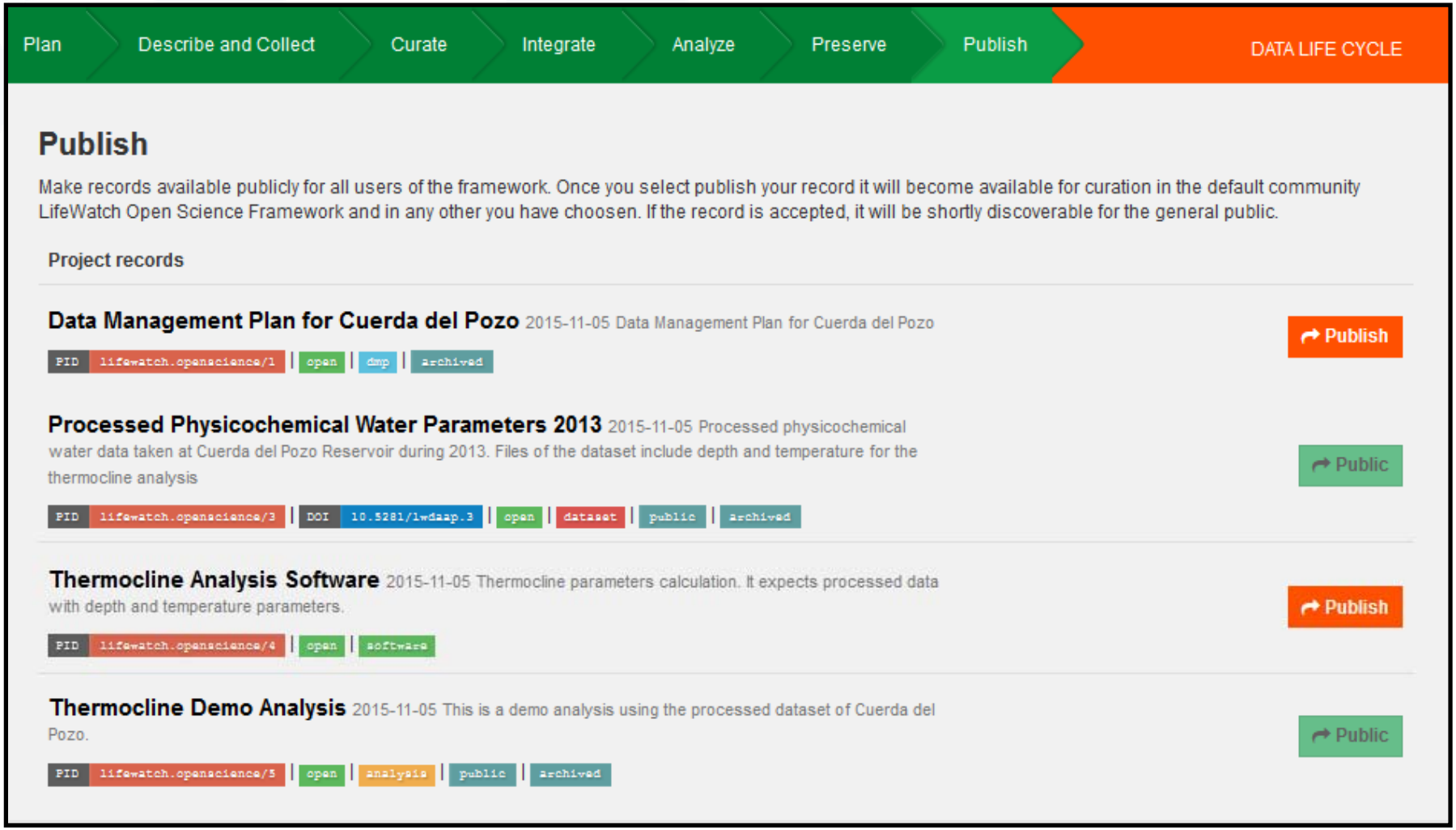
File Edit View Insert Cell Kernel Help

Code Cell Toolbar: None

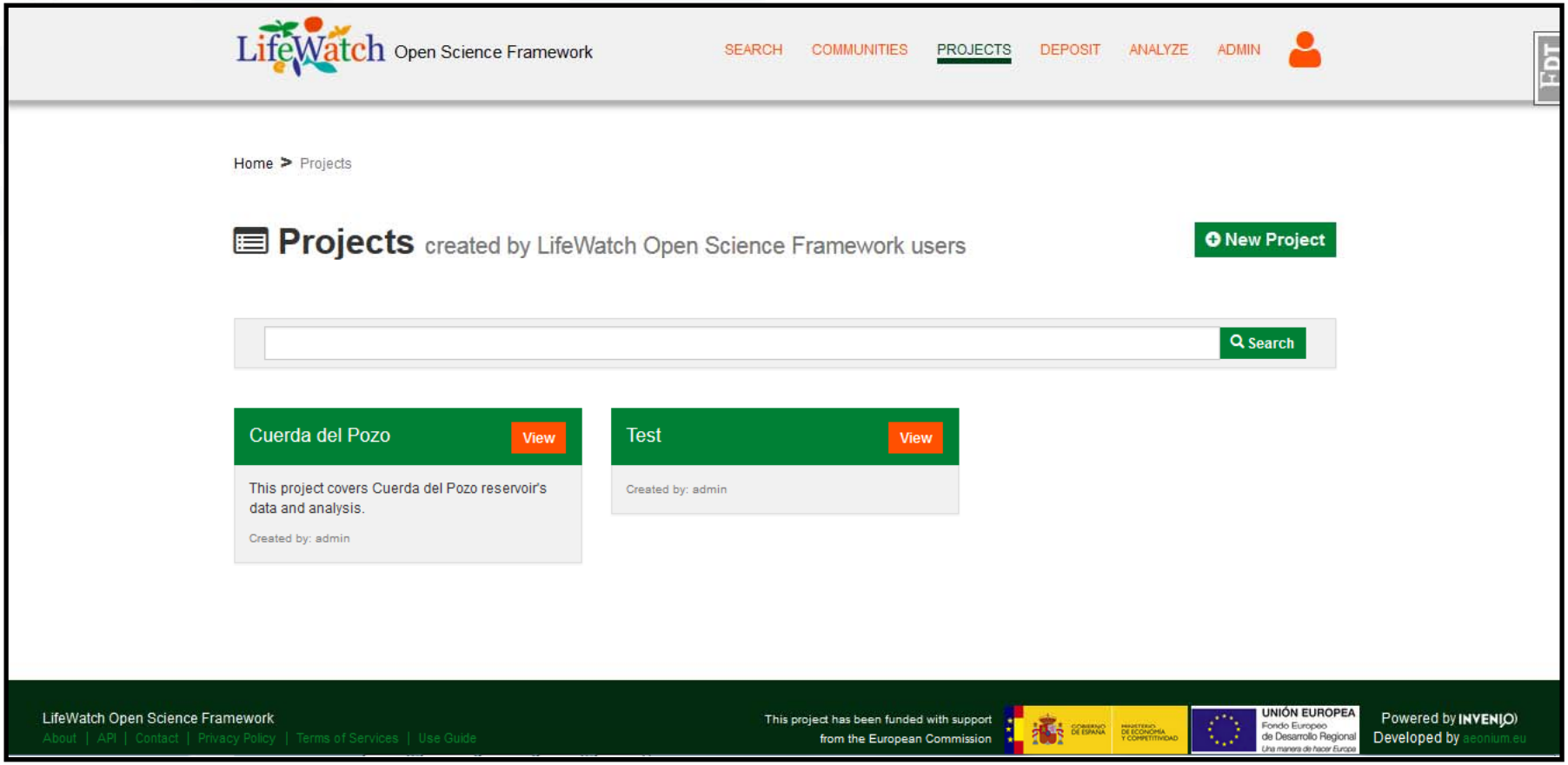


Projects follow the Data Life-Cycle: Preserve

Projects follow the Data Life-Cycle: Publish



An integrating VRE oriented to Projects



Technical question: connect with OpenProject?

-code: github

-resources: VO policies, groups , VM…

Key question: Incorporate Digital Knowledge

- Software (VM) + Data preservation is not enough
  - Ideas explored under CMS preservation
    - Validation
    - Analysis Description
  - CHEP 2014 discussion → Knowledge Preservation
- Incorporate Digital Knowledge from start
  - Understand the use of “ontologies” / “semantics”
    - Ontologies are not taxonomies
    - Ontologies are not metadata
    - Ontologies are not (restricted) vocabularies
- *Under analysis for Fresh Water VRE:*
  - *Ontological Framework explored: SWEET*
  - *Ontologies: EML and WaterEML*
  - *Consider report at RDA 2015 (Paris) devoted session*
  - *What about INSPIRE directive?*
- *First try:*
  - *Build on SWEET*
  - *Start from vocabularies used in CdP*
- Yet, how to integrate into DMP? Ideal (unify) is obvious…

Incorporating Digital Knowledge…

CuerdaDelPozo (http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/10/CuerdaDelPozo) : [/Users/guadalupecanasherrera/CuerdaDelPozo

CuerdaDelPozo (http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/10/CuerdaDelPozo)

Active Ontology | Entities | Object Properties | Data Properties | Individuals by class | OntoGraf | SPARQL Query

SPARQL query:

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX owl: <http://www.w3.org/2002/07/owl#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>
PREFIX onto: <http://www.semanticweb.org/guadalupecanasherrera/ontologies/2015/10/CuerdaDelPozo#>

SELECT ?instruments  ?magnitudes
        WHERE {?instruments rdf:type  onto:ServoirPlatform .
                ?instruments onto:is_type onto:Water .
                ?instruments onto:measures_magnitude ?magnitudes
                }
```

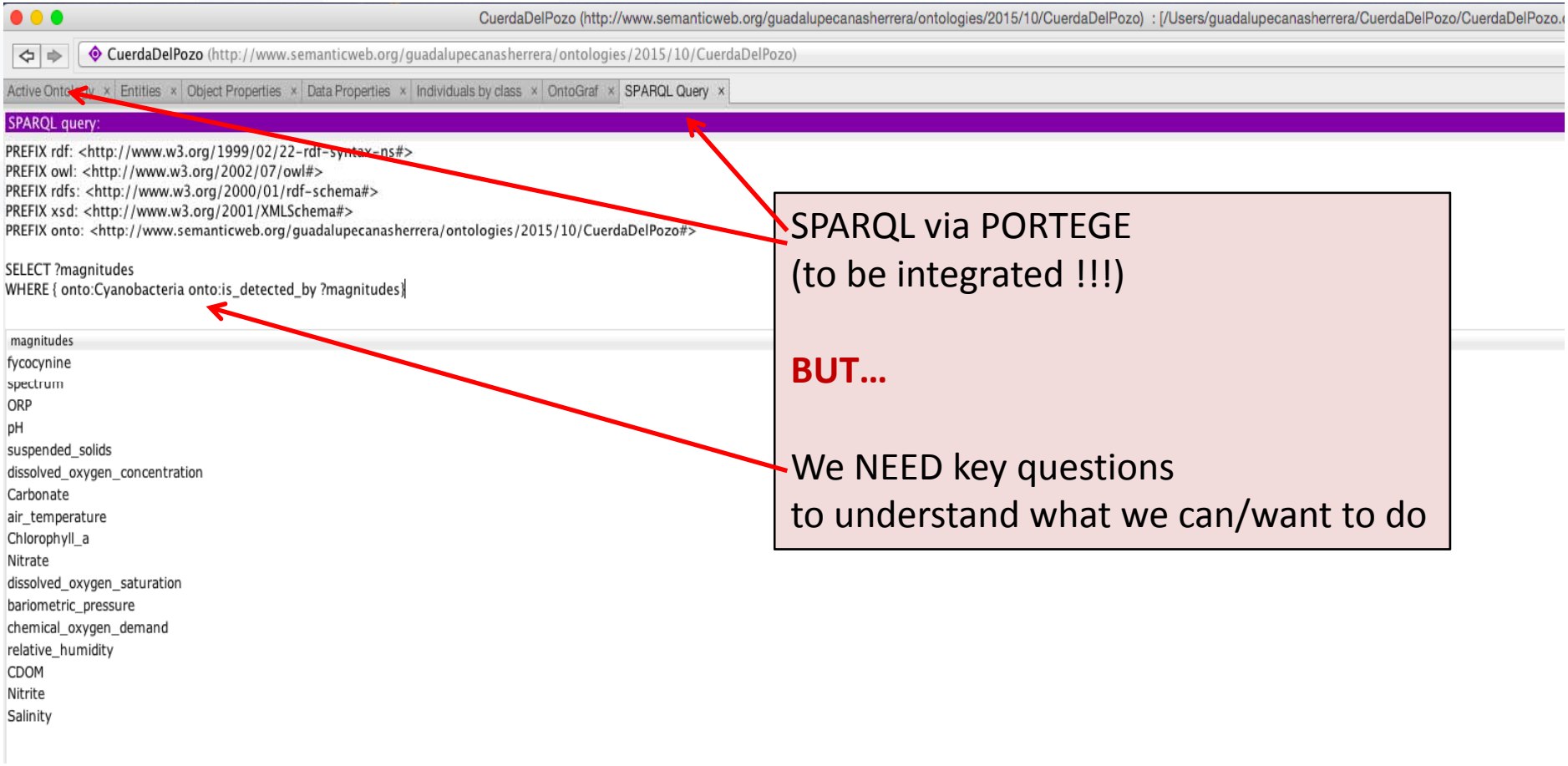
OWL via PORTEGE
(to be integrated !!!)

BUT…

We NEED key questions
to understand what we can/want to do

instruments	magnitudes
radiometer	incident_upcoming_irradiance
radiometer	incident_downcoming_irradiance
fluorometer	CDOM
fluorometer	fycocynine
fluorometer	Chlorophyll_a
UV_Spectrometer	Nitrate
UV_Spectrometer	chemical_oxygen_demand
UV_Spectrometer	suspended_solids
UV_Spectrometer	Nitrite
UV_Spectrometer	Carbonate
UV_Spectrometer	spectrum
multiparametric_probe	depth_at_measurement
multiparametric_probe	water_temperature
multiparametric_probe	ORP
multiparametric_probe	Salinity
multiparametric_probe	dissolved_oxygen_concentration

Incorporating Digital Knowledge…



SPARQL via PORTEGE
(to be integrated !!!)

BUT…

We NEED key questions
to understand what we can/want to do

Thank you for your attention.

Questions?

Join EGI-LW CC meeting in
EGI Conference in BARI
Tomorrow 11/11 @ 9h

Tourists in the ARTIC seeing how glaciers melt

elroto.elpais@gmail.com

www.egi.eu

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