EGI-LifeWatch Competence Center Support to LW community

Research Infrastructures Roadmaps and Competence Centres

EGI 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

Agenda for the EGI-LW CC meeting

- Intro: Support to LifeWatch Community
- Big Data and Ecological Observatories: LW Marine
- Supporting Workflows & Virtual Labs in FedCloud
- Advanced Support to Citizen Science in Biodiversity

EGI- LifeWatch Competence Center

LifeWatch
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

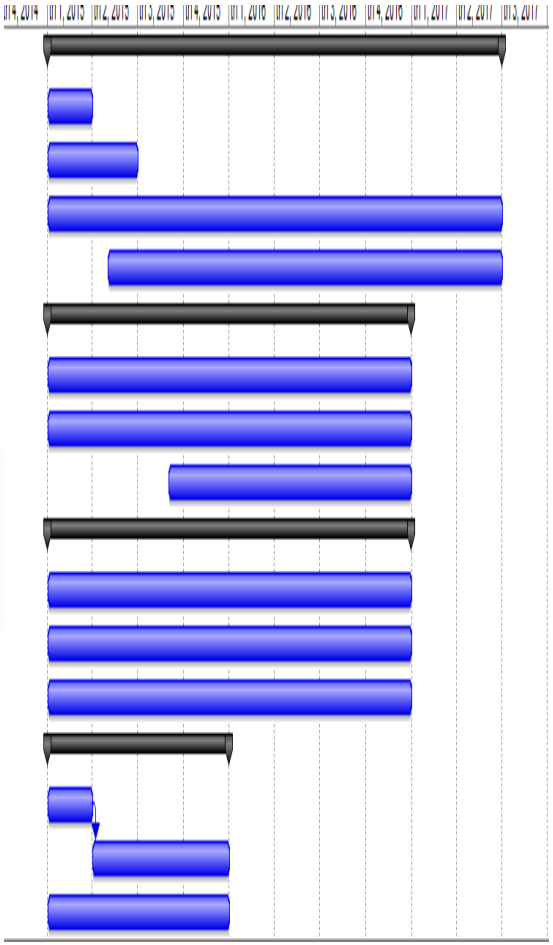
Proposal presented by
I.Blanquer & J.Marco

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.

Tasks/WP

TASK 1 Support to LifeWatch Community and Exploitation
T1.1 Connection of Project Management tools
T1.2 Implementation of common support tools.
T1.3 Deployment and operation of a support and training platform
T1.4 Definition of a dissemination and exploitation strategy
TASK 2 Big Data and Ecological Observatories
T2.1 Handling Data Streams from Ecological Observatories
T2.2 Supporting large software suites for Modelling Ecosystems
T2.3 Towards an integrated framework
TASK 3 Supporting Workflows & Virtual Labs in FedCloud for LifeWatch
T3.1 Integration of popular Bioinformatic interfaces
T3.2 An extensible framework for executing biodiversity pipelines
T3.3 Implementation of the Network of Life
TASK 4 Advanced Support to Citizen Science in Biodiversity
T4.1 Updated analysis of ongoing initiatives on nature observation
T4.2 Exploration of pattern recognition tools
T4.3 Engagement: outreach and inreach.



EGI-Engage

Participants

Participant No *	Participant organisation name	Role in the CC (user community/technology provider/service provider)
1 (Coordinator)	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	Flanders Marine Institute, VLIZ, Belgium	User Community
7	Research Centre on Biodiversity & Genetic Resources, CIBIO, Portugal	User Community

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 on-line specific courses*

Task 2: Big Data and Ecological Observatories

EGI-Engage

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.inaturalist.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.

LifeWatch Competence Center

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

LW-CC Scope

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

Advanced Support to Citizen Science in Biodiversity

F. Serrano (BIFI-Ibercivis)

With input from F. Sanz, BIFI-Ibercivis, J.Marco, CSIC-IFCA, F. Pando, CSIC-RJB

www.egi.eu

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

Citizen Science

- ***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.inaturalist.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.

- D4.1 **Updated analysis of ongoing initiatives on citizen science** for Biodiversity and Ecosystems Research. Type: Report. **Due M3**
- D4.2 Initial list of actions oriented to citizen science using LW-DCC Type: Report, Due: M6
- OTHER (UNFUNDED) CONTRIBUTIONS FROM
  - **LifeWatch Greece (see talk on Tuesday)**
  - **JRU-LW-ES: Phenology project**
  - **LifeWatch Portugal / IBERLIFE : CIBIO (see talk on Thursday)**

Citizen Science

- D4.1 (review of ongoing initiatives, M3=end may, internal)

Best choice (to integrate with our objectives): iNaturalist

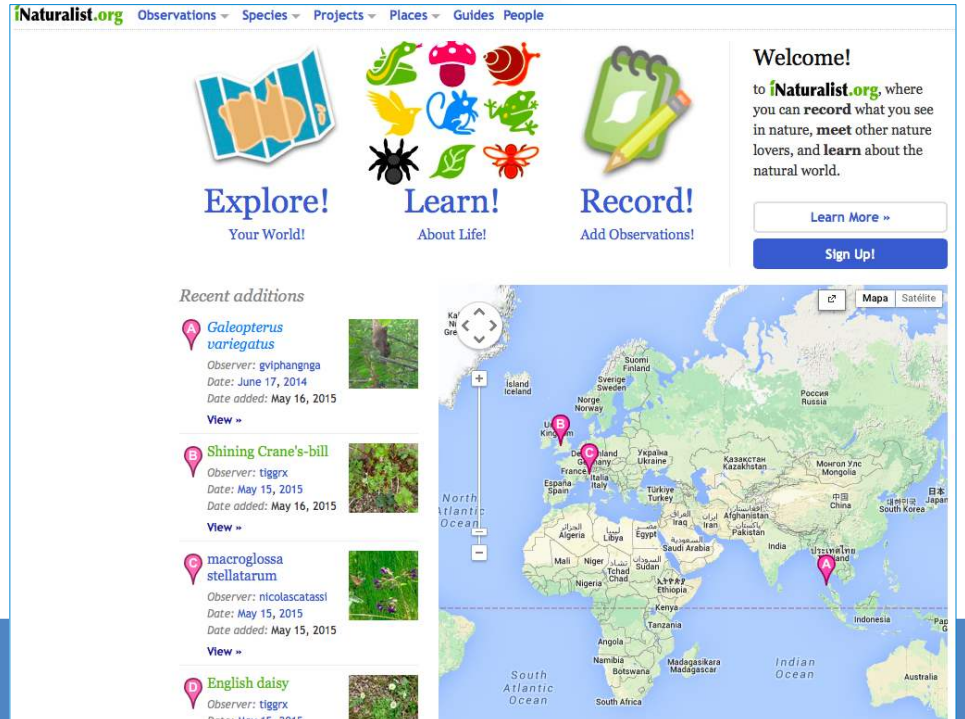
contact: RJB (F.Pando)

Framework for biodiversity data [gathering/management/visualization] + social network

Open & accesible & well documented (both sources, API) → web&mobile

Widely used (e.g. gbif, NationalGeographic), aggretation of data, community of developers

Developed on Ruby on rails



Citizen Science

EGI-Engage

LifeWatch
E-Science European Infrastructure for Biodiversity and Ecosystem Research

inaturalist — Ruby ★ 69 ⑂ 54
The Rails app behind iNaturalist.org
Updated 2 days ago

INaturalistIOS — Objective-C ★ 16 ⑂ 17
iNaturalist iOS app
Updated 2 days ago

iNaturalistAndroid — Java ★ 11 ⑂ 15
Android app for iNaturalist.org
Updated 4 days ago

inaturalist-cookbook — Ruby ★ 0 ⑂ 0
Chef recipes needed to build development and production machines for iNaturalist
Updated 7 days ago

Windshaft-inaturalist — JavaScript ★ 3 ⑂ 3
Wrapper around Windshaft to provide a map tiler for iNaturalist.
Updated on Mar 20

inaturalist.github.io — HTML ★ 0 ⑂ 0
iNat gh-pages repo, mostly for fun

https://github.com/inaturalist

/observations.json

Response


```
[{
    "user_login": "greatjon",
    "place_guess": "San Francisco, San Francisco",
    "location_is_exact": false,
    "quality_grade": "casual",
    "latitude": 37.713539,
    "created_at": "2012-04-10T21:48:25-07:00",
    "timeframe": null,
    "species_guess": "gray wolf",
    "observed_on": "2012-04-10",
    "num_identification_disagreements": 0,
    "delta": true,
    "updated_at": "2012-04-10T21:49:50-07:00",
    "num_identification_agreements": 0,
    "license": null,
    "geoprivacy": null,
    "positional_accuracy": 354,
    "coordinates_obscured": false,
    "taxon_id": 42048,
    "id_please": false,
    "id": 2281,
    "iconic_taxon": {
        "name": "Mammalia",
        "ancestry": "48460/1/2",
        "rank": "class",
        "id": 40151,
        "rank_level": 50,
        "iconic_taxon_name": "Mammalia"
    },
    "time_observed_at_utc": "2012-04-11T04:47:55Z",
    "user_id": 53,
    "time_observed_at": "2012-04-10T21:47:55-07:00",
    "observed_on_string": "Tue Apr 10 2012 21:47:55 GMT-0700 (PDT)",
    "short_description": "",
    "time_zone": "Pacific Time (US & Canada)",
    "out_of_range": null,
    "longitude": -122.340054,
    "description": "",
    "user": {
        "login": "greatjon"
    },
    "positioning_method": null,
    "map_scale": null,
    "photos": [],
    "iconic_taxon_name": "Mammalia",
    "positioning_device": null,
    "iconic_taxon_id": 40151
}]
```


iNaturalist

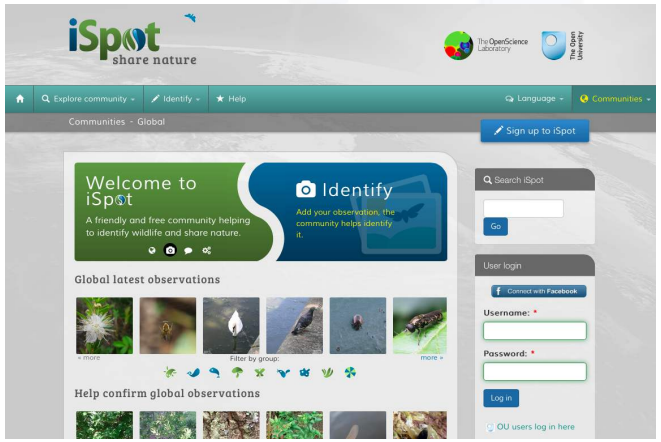
https://www.inaturalist.org/pages/api+reference

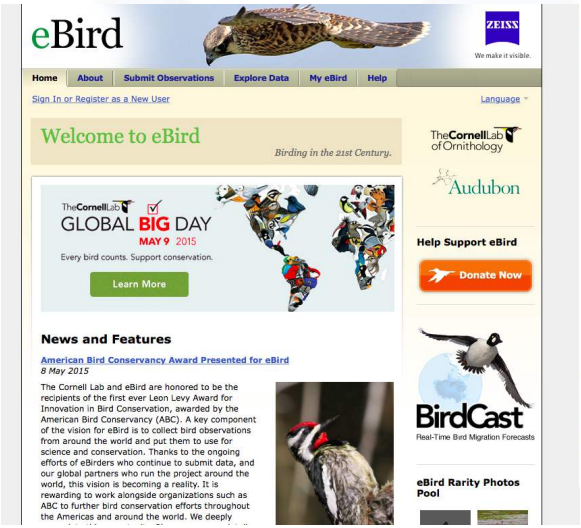
Citizen Science

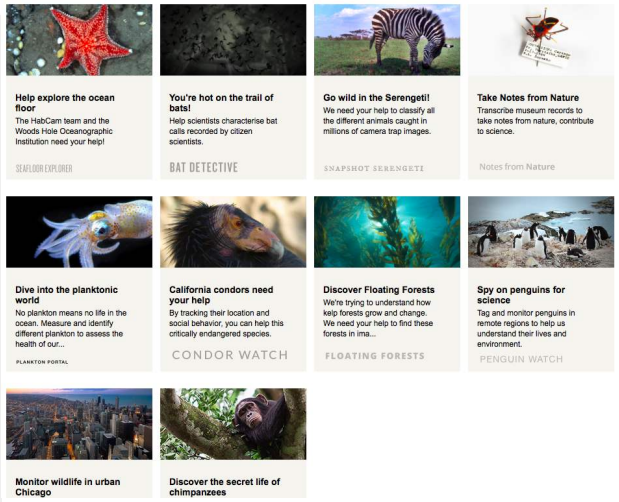
- D4.1 (review of ongoing initiatives, M3=end may, internal)

Other reviewed options alternative to iNaturalist:

- Ebird (Cornell Lab, USA) Api well documented, widely used
- iSpotNature (Open University, EU)
- EpiCollect (EU) Source code accesible, we have experience
- EpiWork (EU), Source code accesible, we have experience
- NatureWatch (USA), based on iNaturalist
- Zooniverse (EU), widely used



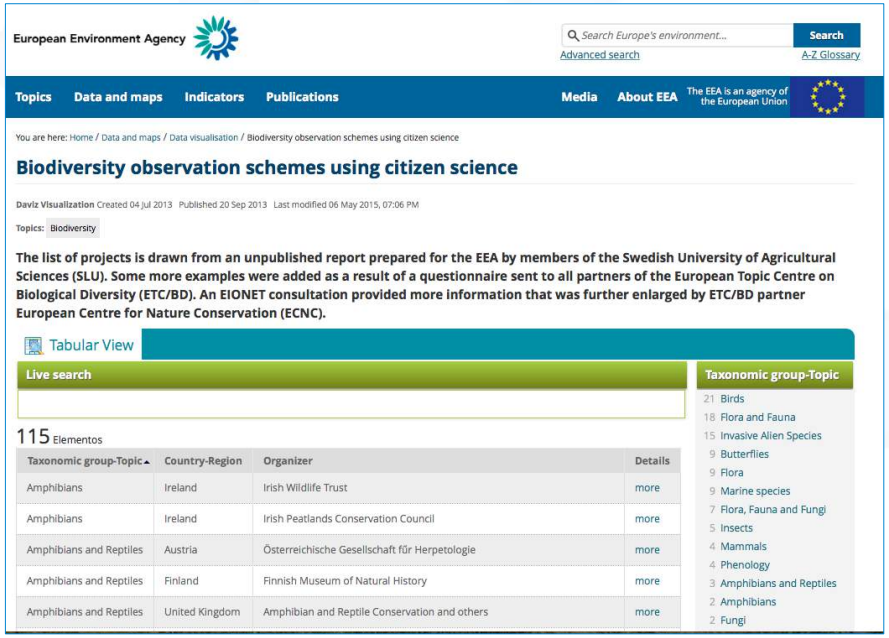




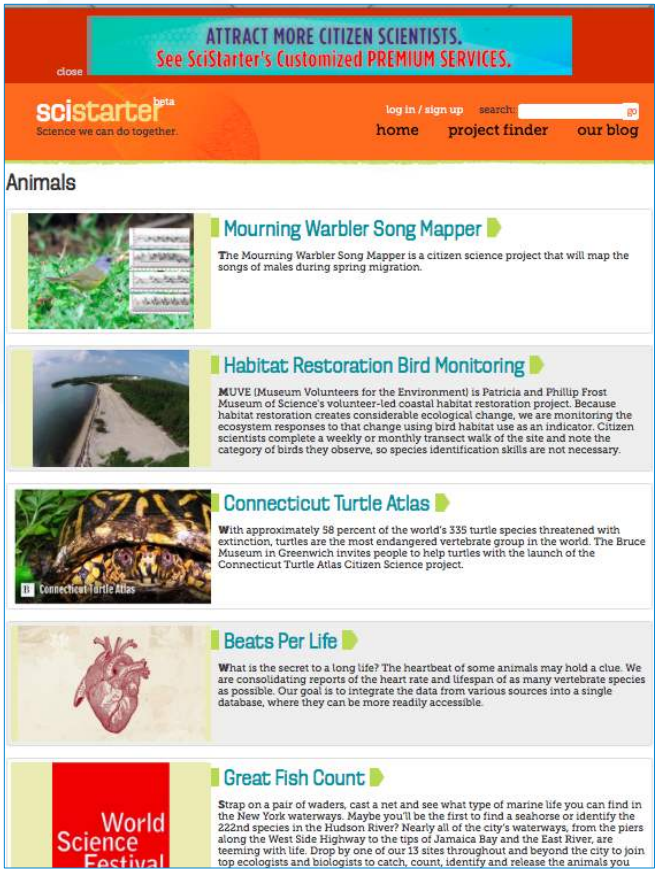
Citizen Science

- D4.1 (review of ongoing initiatives, M3=end may, internal)
  Hundreds of ongoing initiatives / phenology + observation projects:

www.scistarter.com

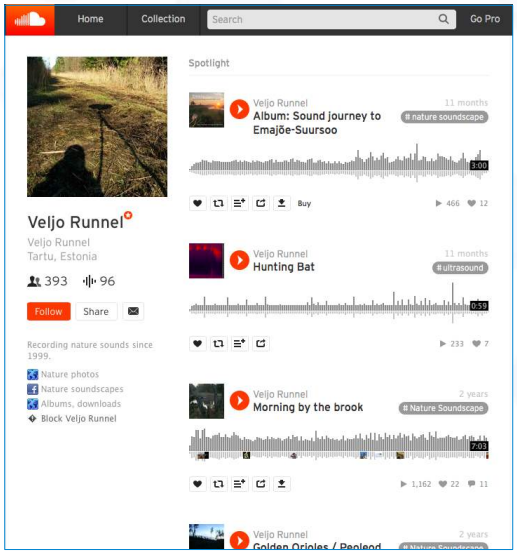


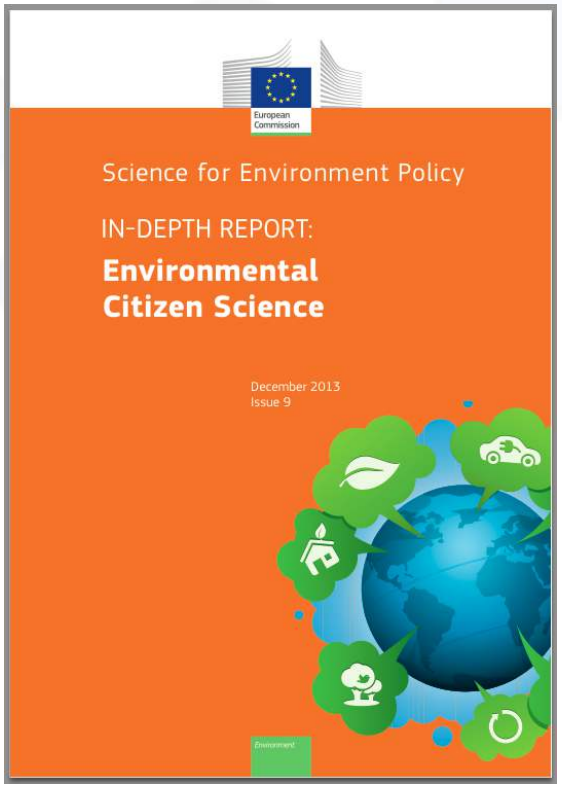
http://www.eea.europa.eu

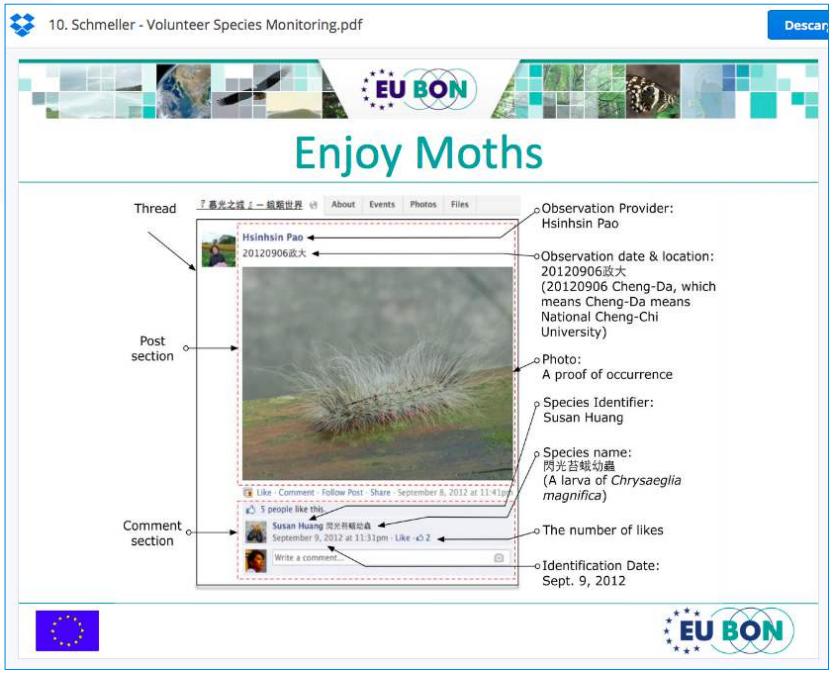


Citizen Science

- D4.1 (review of ongoing initiatives, M3=end may, internal)
  Hundreds of ongoing initiatives / phenology + observation projects:







Citizen Science

- ***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.

D4.3 Integration of assisted pattern recognition tools. Type: prototype. Due: M9

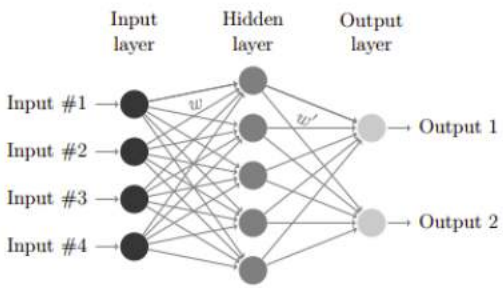
Example of assisted pattern recognition tool

Identification of birds from their songs

Breaking down the problem

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \\ x_d \end{bmatrix}$$

Feature vector



Data Reduction → Feature Engineering → Classification

Automatic Segmentation	Averaged MFCCs estimators	Neural Network (MLP)

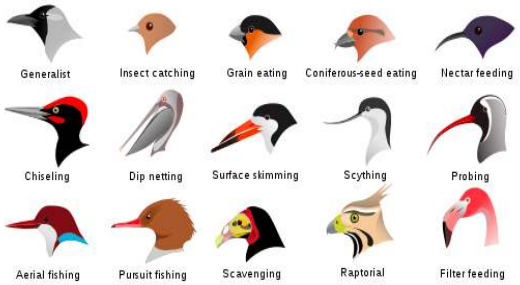
More details in Federated Accelerated Computing session, tomorrow at 14h

Other options for automatic pattern recognition (CPU → GPU)

http://vision.ucsd.edu

Visipedia Project

Visipedia is a joint project between Pietro Perona's Vision Group at Caltech and Serge Belongie's Vision Group at UCSD. Visipedia, short for "Visual Encyclopedia," is an augmented version of Wikipedia, where pictures are first-class citizens alongside text. Goals of Visipedia include creation of hyperlinked, interactive images embedded in Wikipedia articles, scalable representations of visual knowledge, largescale machine vision datasets, and visual search capabilities. Toward achieving these goals, Visipedia advocates interaction and collaboration between machine vision and human users and experts.



Sub-Projects

Visual Recognition With Humans in the Loop

We present an interactive, hybrid human-computer method for object classification. The method applies to classes of problems that are difficult for most people, but are recognizable by people with the appropriate expertise (*e.g.*, animal species or airplane model recognition). The classification method can be seen as a visual version of the *20 questions game*, where questions based on simple visual attributes are posed interactively. The goal is to identify the true class while minimizing the number of questions asked, using the visual content of the image. Incorporating user input drives up recognition accuracy to levels that are good enough for practical applications; at the same time, computer vision reduces the amount of human interaction required. The resulting hybrid system is able to handle difficult, large multi-class problems with tightly-related categories.

Citizen Science

- ***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.

  Expected date: Oct-Nov 2015 (with the support of Ibercivis Foundation)

- D4.4 Report on the success of citizen oriented initiatives and proposal for future work. Type: report. Due:M12
- D4.5 Report of the public participatory event day. Type: report. Due: M12

Join us tomorrow for further discussions

Session on Platforms for Citizen Science

- Tuesday 19 May 15:30-17:00
- Location B-104

Thank you for your attention.

Questions?

www.egi.eu

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