



EUBrazilOpenBio

EU-Brazil Open Data and Cloud Computing e-Infrastructure for Biodiversity

EUBrazilOpenBio - Open Data and Cloud Computing e-Infrastructure for Biodiversity (2011-2013) funded under the *Objective FP7-ICT-2011-EU-Brazil Research and Development cooperation* will deploy an e-Infrastructure of open access resources supporting the needs of the biodiversity scientific community.

Tackling the complexity of Biodiversity Science requires dealing with multiple multidisciplinary datasets spanning from climatology to earth sciences all of key importance to overcome the fragmentation and focus on uniting existing different European and Brazilian data sources to provide scientists with an even greater knowledge base, achieved through the integration and shared use of appropriate computing resources. In parallel EUBrazilOpenBio supports the Open Access Movement, promoting the concept of openness for scientific research, aligned with the OpenAIRE initiative launched in 2010 to establish an infrastructure for EC-funded researchers to publish their OA work. EUBrazilOpenBio supports these critical initial steps towards greater openness in the advancement of research and scholarship, through both a policy mandate for open access and a provision of infrastructure to support that policy.

The breadth and depth of the resulting data infrastructure and the openness of its resources will enable a large variety of new cost-effective, cross-disciplinary virtual research environment applications thus opening the way to its widespread adoption and exploitation by the biodiversity scientific community.

Objectives: EUBrazilOpenBio aims to ambitiously combine the two key themes above to deploy an e-Infrastructure of open access resources (data, tools and services) that will make significant strides towards fully supporting the needs and requirements of the biodiversity scientific community. This data e-Infrastructure will result from the federation and integration of existing EU and Brazilian developed infrastructures and resources, namely through Catalogue of Life, D4Science-II, openModeller and Venus-C.

Specifically EUBrazilOpenBio has three key objectives:

1. Drive forward the interoperation of existing Brazilian and European e-Infrastructures in the distributed computing, scientific data and portals & platform layers
2. Provide greater focus to the integration of data software platforms running through all of infrastructures
3. Identify further future EU-Brazil collaboration in support to the biodiversity area in all types of infrastructures

Description of work: The foundations for the biodiversity scientific research lie in a multitude of datasets (publications and observations' results but also datasets generated by powerful space, airborne and in situ sensors instrumentation) produced by a large number of worldwide disperse scientists by adopting very different methodologies, approaches and standards. A major effort performed by biodiversity scientists today is to produce, out of these huge amounts of datasets, consistent and complete reference resources that can be used by the entire scientific community. Indeed, building these resources is quite complex and expensive in terms of human and computing resources. Species niche modelling is another core activity performed by biodiversity scientists today that rely heavily on the availability of rich datasets and tools implementing algorithms dedicated to capture different biological aspects and predictions.

By integrating and making interoperable a large variety of open access datasets and collections (including publications), tools and computing resources maintained by EU and Brazilian initiatives, and by providing an environment where building and operation of modeling tools is simplified, the EUBrazilOpenBio project will facilitate these scientific activities in overcoming

Project acronym:
EUBrazilOpenBio

Contract n°: ICT-288754

Project type: Collaborative Project

Start date: 01/06/2011

Duration: 24 months

Total budget:

EUROPE

1.253.555 €

BRAZIL

R\$ 2.299.939,13

Funding from the EC:

1.049.737 €

Funding from CNPq:

R\$ 2.299.939,13

Total funded effort in

person-month:

288,5 PM

(EU 144,5 + BR 144)

Web site:

www.eubrazilopenbio.eu

Contact person:

EUROPE

Vassil Alexandrov

email: valexand@bsc.es

tel.: +34 93 413 7576

fax.: +34 93 413 7721

Donatella Castelli

Email:

Donatella.castelli@isti.cnr.it

BRAZIL

Vanderlei Cahnos

email: vcanhos@cria.org.br

tel.: +55 19 3288-0466

fax.: +55 19 3249-0960

Vinod Rebello

email: vinod@ic.uff.br

Project participants:

EUROPE

BSC E

CNR IT

TRUST-IT UK

UPVLC E

Species2000 UK

BRAZIL

CRIA SP

CESAR PE

UFF RJ

RNP RJ

some important limitations faced today by the biodiversity scientists. Furthermore, by laying the conditions for progressively and continuously expanding this set of shared resources, it will act as the springboard of a worldwide initiative which will accelerate comprehension of significant global biodiversity issues.

The project plans to initiate this exploitation path by deploying and operating in the EUBrazilOpenBio data e-Infrastructure a number of Virtual Research Environments (VREs). Two use cases will be implemented by these VREs introducing novelty with respect to the state-of-the-art biodiversity scientific environments.

Use Case I – Integration between Regional & Global Taxonomies. Taxonomical intelligent mappings are often algorithmically complex and demanding in terms of computational resources, data required and exchange between widely dispersed scientists. The EUBrazilOpenBio e-Infrastructure offers a large variety of capabilities for supporting such kind of mappings. In order to fully exploit this solution, the pilot study will analyse and cross-map all or part of the regional plant catalogue of the Brazilian Catalogue of Life (CoL) Regional Hub served by CRIA (over 30,000 species of flowering plants) with the global plant catalogue used in the CoL and served by Species2000 (with up to 150,000 species of plants). This case study provides a component in the 4D4Life Global Multi-Hub Network, and will use test versions of two tools under development in the i4Life project – the CoL Cross-mapping Tool, and the CoL Piping Tool.

Use Case II- Data usability and the use of ecological niche modeling. Ecological Niche Modelling (ENM) is a widely used approach to understand distribution of species. Practical problems are associated with intensive computational requirements in an analysis involving a large number of species, and the complexity of some ENM algorithms. Frequently the ENM process requires a long tedious process of data cleaning, and harmonisation of scientific names for synonyms, misspellings, etc. This use case will exploit existing data resources (the speciesLink network, the List of Species of the Brazilian Flora and the results of the Brazilian National Institute of Science & Technology – Virtual Herbarium of Flora and Fungi) and ecological niche modeling algorithms to produce potential distribution models for specific species to orient new surveys in Brazil.

Progress beyond the state-of-the-art: Over the last decade the effort to capture biodiversity data in the traditional field of life sciences has resulted in the development of public archives that actively store and curate data. Today, the biodiversity domain can leverage network access to a substantial amount of relevant on-line data sources, information services and tools. A large fraction of these resources are already open access and many more are going to move in this direction in both Europe and Brazil. EUBrazilOpenBio will develop a biodiversity dedicated data e-Infrastructure that, by connecting major EU and Brazilian computing and data e-Infrastructures, will make available an extremely rich set of open access resources to biodiversity scientists.

In terms of software platforms and services, the Distributed Computing Infrastructures technologies are progressively moving towards the adoption of a mixed approach that brings together Grid and Cloud infrastructures, a distinguished example in this context being the gCube Software Platform. By integrating openModeller, the software framework offering a rich library of niche modeling facilities, with the gCube Framework, the EUBrazilOpenBio Software Platform will provide the support for building applications and workflows to integrate, cross-reference, postprocess, analyse and more generally synthesise new knowledge from the available information. It will become possible to achieve the automation, transparency, and timeliness that are required to turn an e-infrastructure into a credible biodiversity and life science platform. It will be equally possible to meet the goals of cost-effectiveness, sustainability, performance, and scale by exploiting the capabilities of resource sharing within the infrastructure that is not limited to data. Rather, it will offer the capabilities to access domain-specific applications, services, computing resources, and storage resources acquired on-demand and used as a commodity for the time needed to carry out the scientific workflows.



Keywords:

Biodiversity, cloud computing, Europe - Brazil collaboration, open access, open data, einfrastructures, niche modeling, taxonomy, virtual research environments

Collaboration with other EC funded projects:

4D4Life
D4Science -II
eScienceTalk
GRDI2020
iMarine
OpenAIRE
VENUS-C

Collaboration with other CNPq Funded projects:

INCT - HVFFINCT-
INCT-Reflora
SICol
SisBiota
speciesLink

