BioGeoSDI workshop

Using TDWG and OGC standards together

Javier de la Torre
jatorre@imaste-ips.com

Hi everybody, bla blah blah...
Workshop because of MoU

The Campinas workshop

BIOGEOSDI EXPERIMENT

Biodiversity Information Standards TDWG

- Tim Sutton (tim@linfiniti.com)
- Bart Meganck (bart.meganck@africamuseum.be)
- Dave Vieglais (vieglais@ku.edu)
- Aimee Stewart (astewart@ku.edu)
- Peter Brewer (p.w.brewer@reading.ac.uk)
- Renato di Giovanni
- Javier de la Torre (jatorre@imaste-ips.com)

http://wiki.tdwg.org/twiki/bin/view/Geospatial/GeoAppInter

-Took place in April.
-First reaction to the MoU paid by TDWG
-7 person
-Fromat of a Code fest, only programmers.

Objectives of the week

> *We aim to test and demonstration of the use of biodiversity informatics and geospatial standards, mainly TDWG and OGC, by creating and implementing a very simple use case that integrates various existing tools and standards within the TDWG universe*

We produced:

1. Prototype-proof of concept web application that uses several standards-technologies

   http://omtest.cria.org.br/biogeosdi/

2. Document-report explaining problems or issues setting up such a prototype

   http://omtest.cria.org.br/biogeosdi/doc/report/BioGeoSDIreport.html

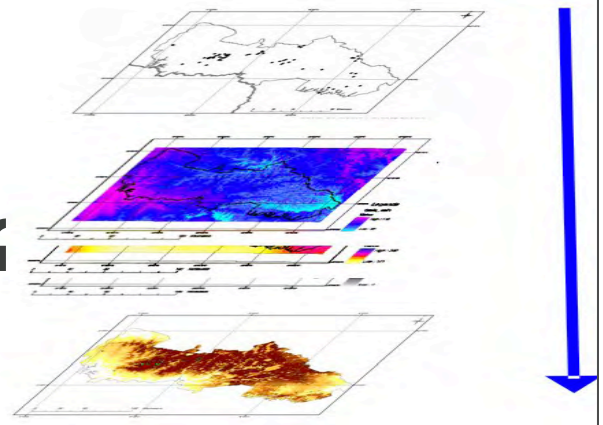
The objetive of the week was to test and demonstrate the use of TDWG and OGc standards in a simple use case implemented in a web application.

A simple use case for a week

BIOGEOSDI EXPERIMENT

Biodiversity Information Standards TDWG

> *Save Biodiversity using TDWG and OGC standards*



> *Online niche modelling experiments using web services*

So we wanted to be driven by a use case so we envision a simple one. save biodiversity using TDWG and OGC standards... simple one ... so we thought in implementing this in an online niche modelling experiment using web services.
Because not everybody in this room might know what niche modelling is I have taken some slides from other presentation first and i will give you a 5min class on niche modelling.

Here is the national natural history museum in madrid. Like in many natural history museums

you can find lots of curious animals, plants, etc in expositions... but what most of the people dont know is that in the back of the museum there are many many more specimens....

This are the so called biological collections... lot of specimens stored in shelves that had been collected for hundred of years. This is very useful for lot of research, but I will only explain you today the one i need for niche modelling.

if you look at one of the boxes in this shelve you can find two beatles with a label indicating where and when they collected.

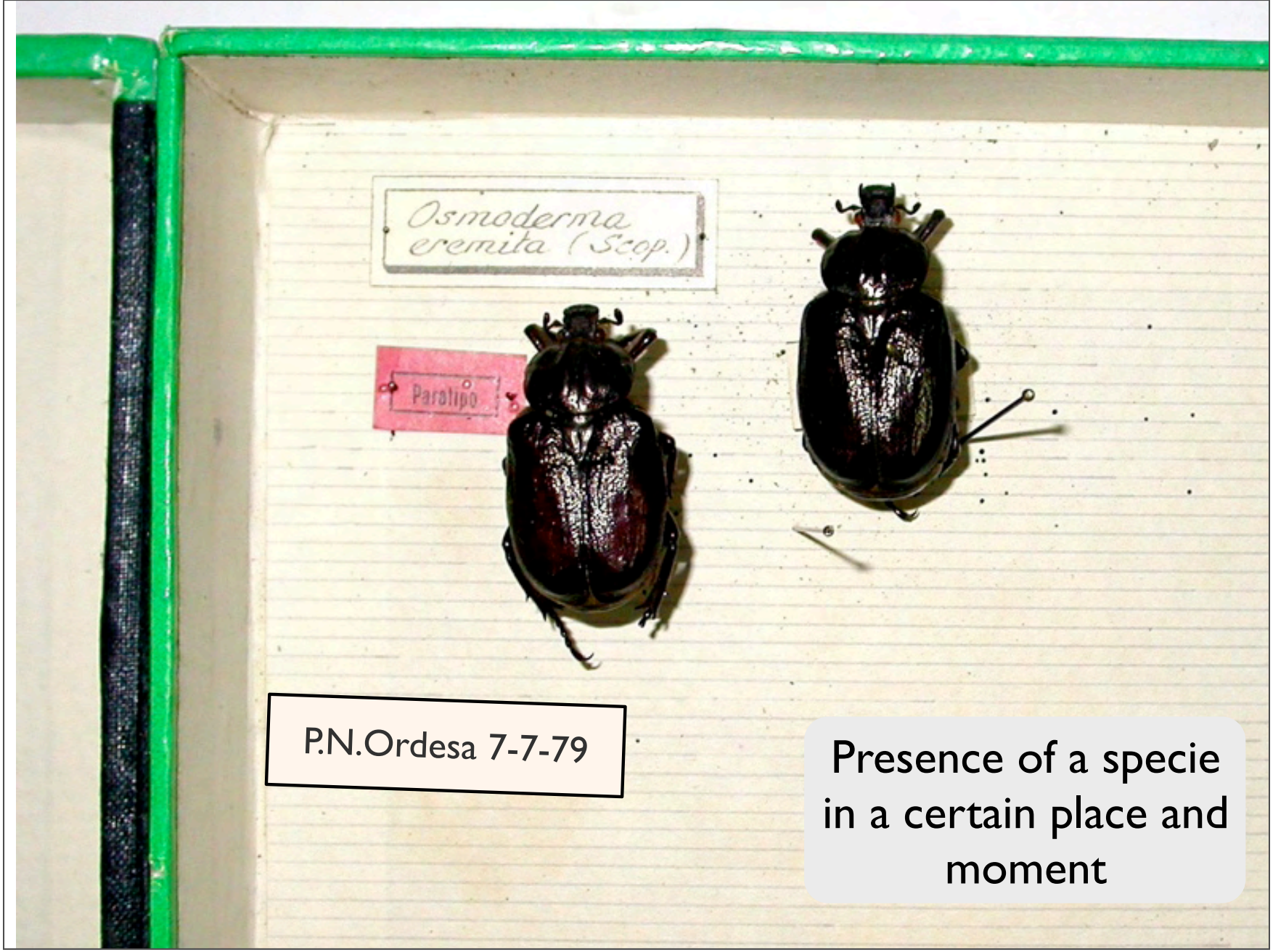
so, in the world of the 1979, the 7th July

In the
Pyrenees

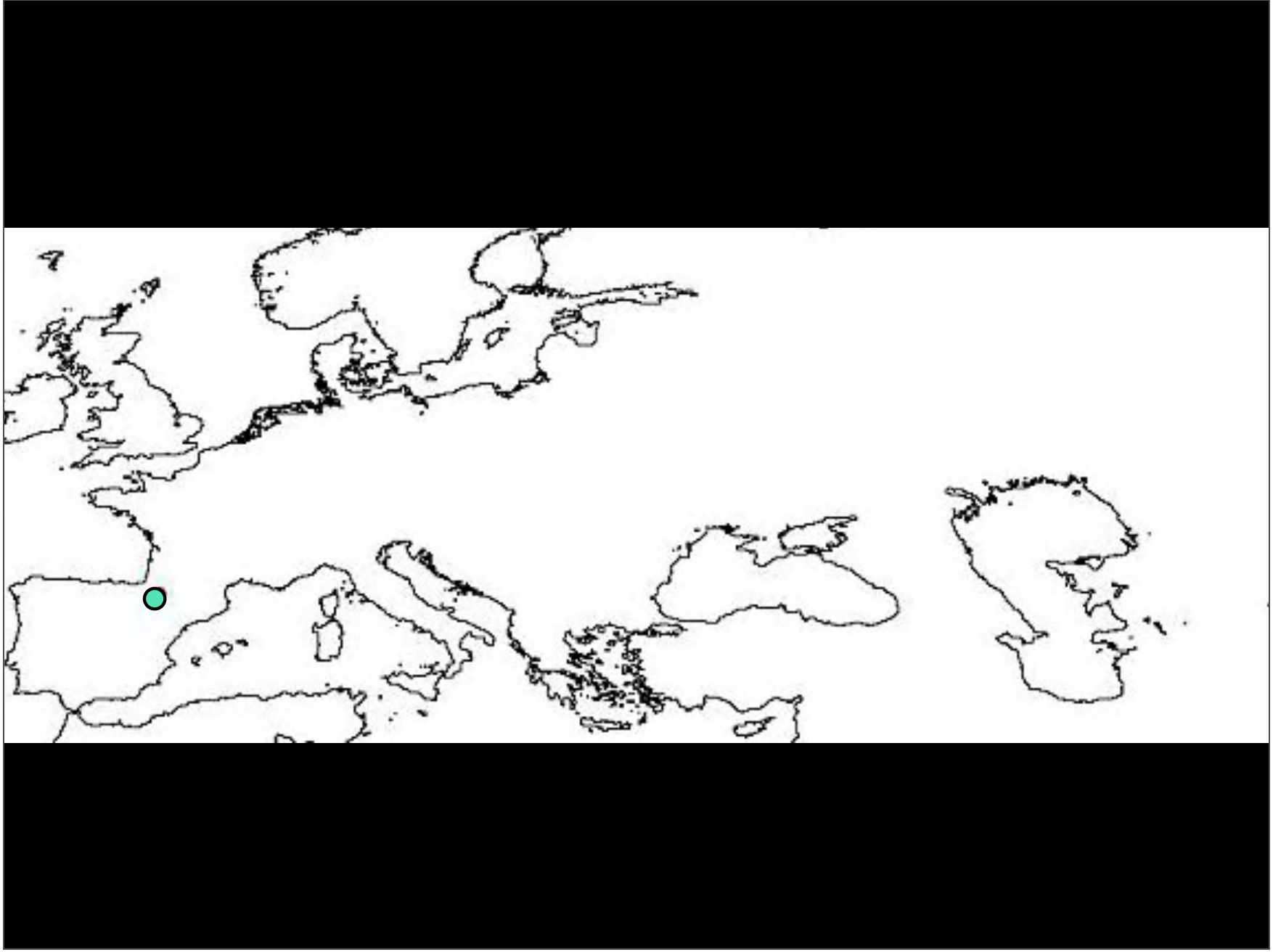
In a national
park

some biologist was
collecting ...

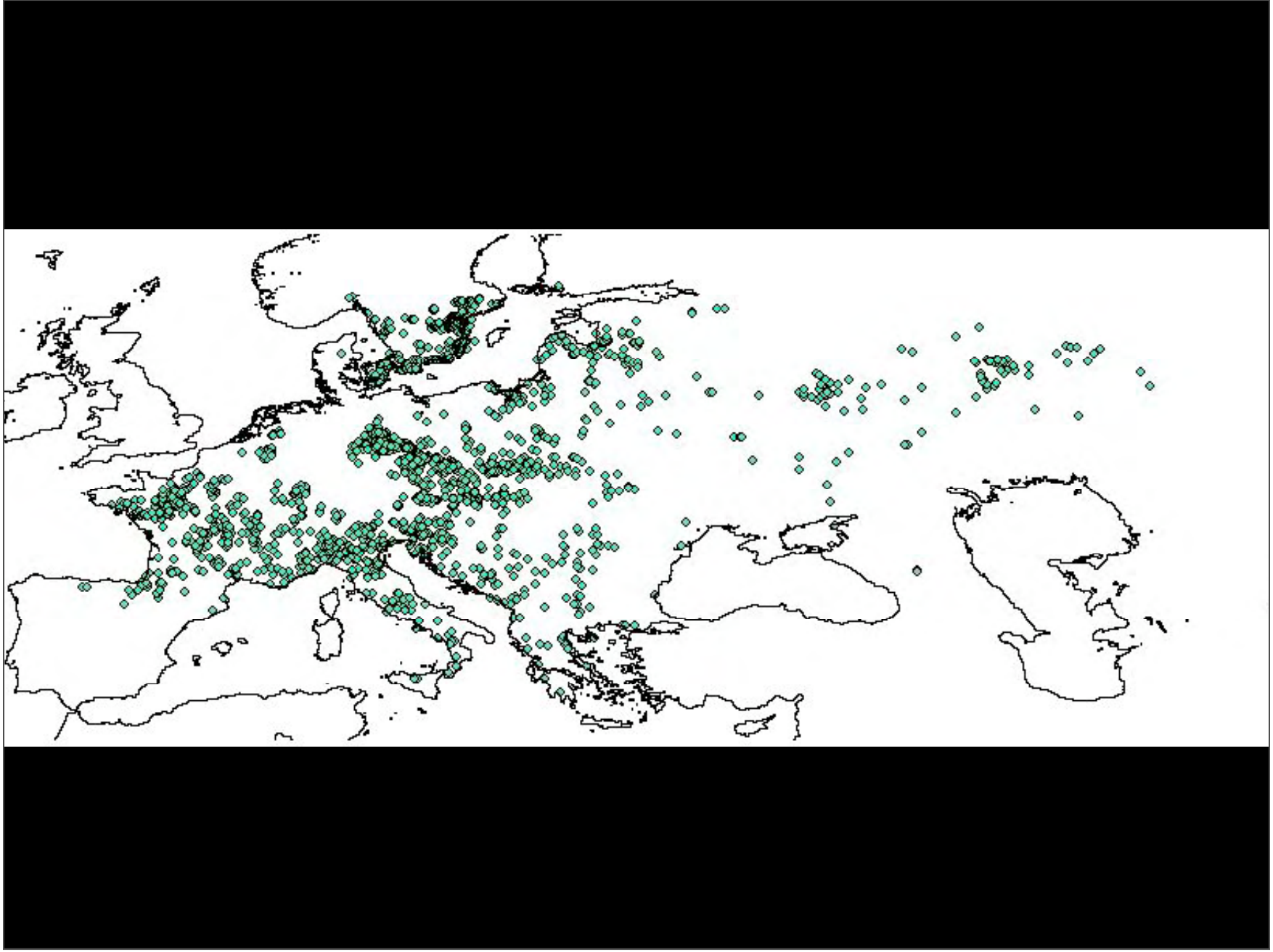
and found this beatle. Put it in a bottle and took it to the museum in madrid



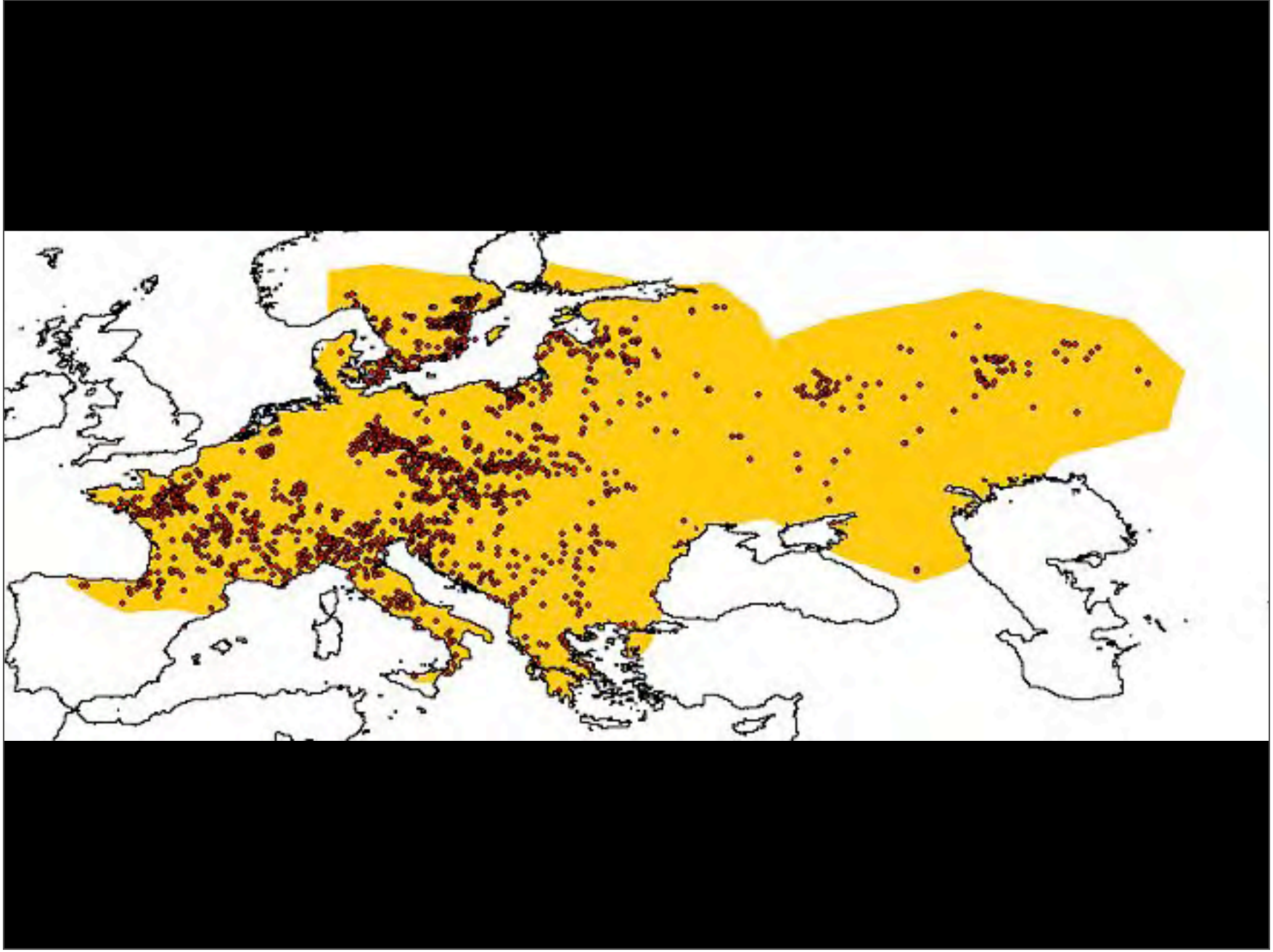
So this specimen at the museum is indicating that this specific specie was found in a certain place and moment.



Well, so if we can put in a map... but still is not very useful...



It becomes really interesting when you have a lot of points of where the specie was found



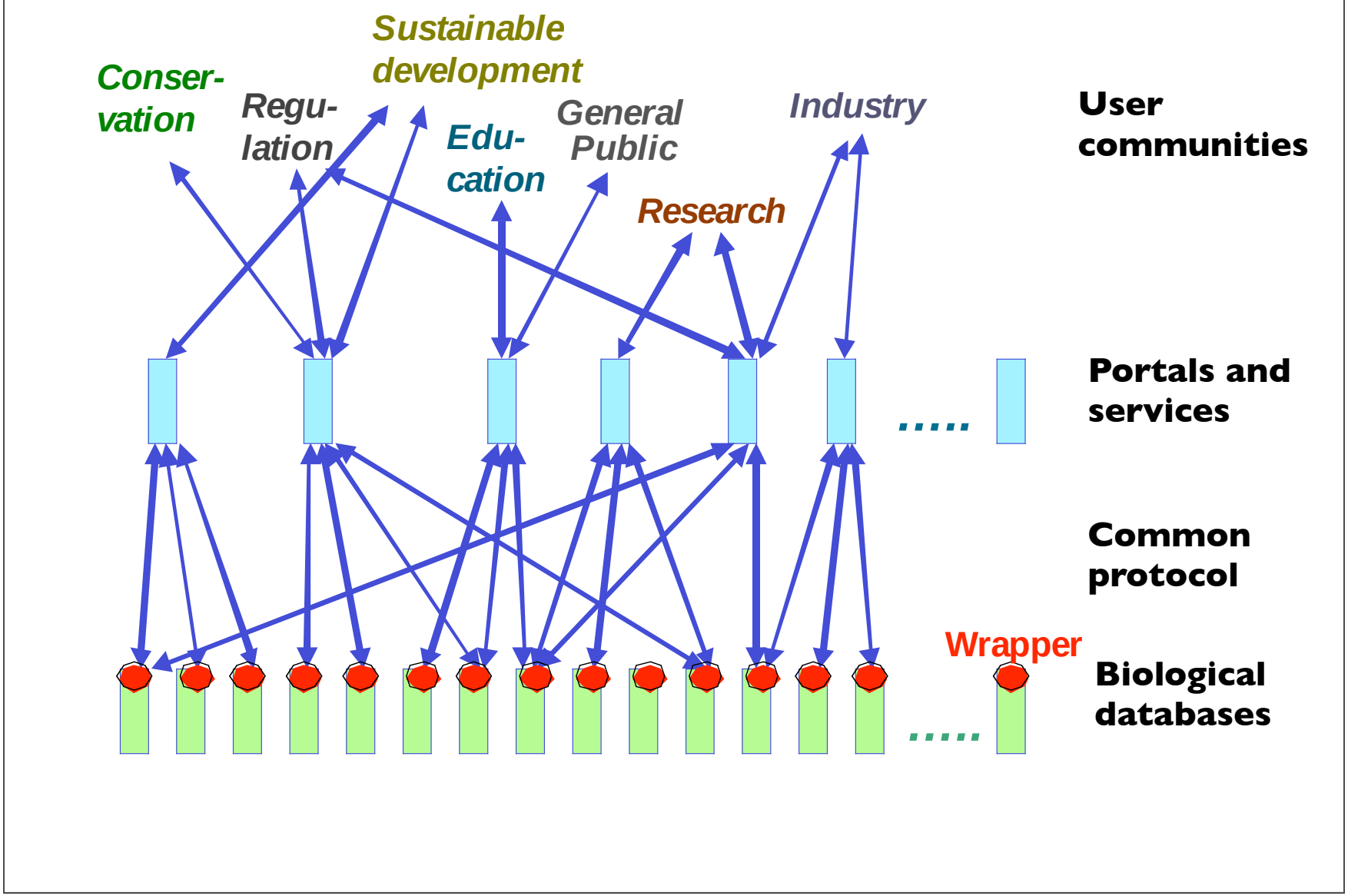
With this you can create the so called map of the distribution of the specie... something really useful to know where species occurred.

We need data

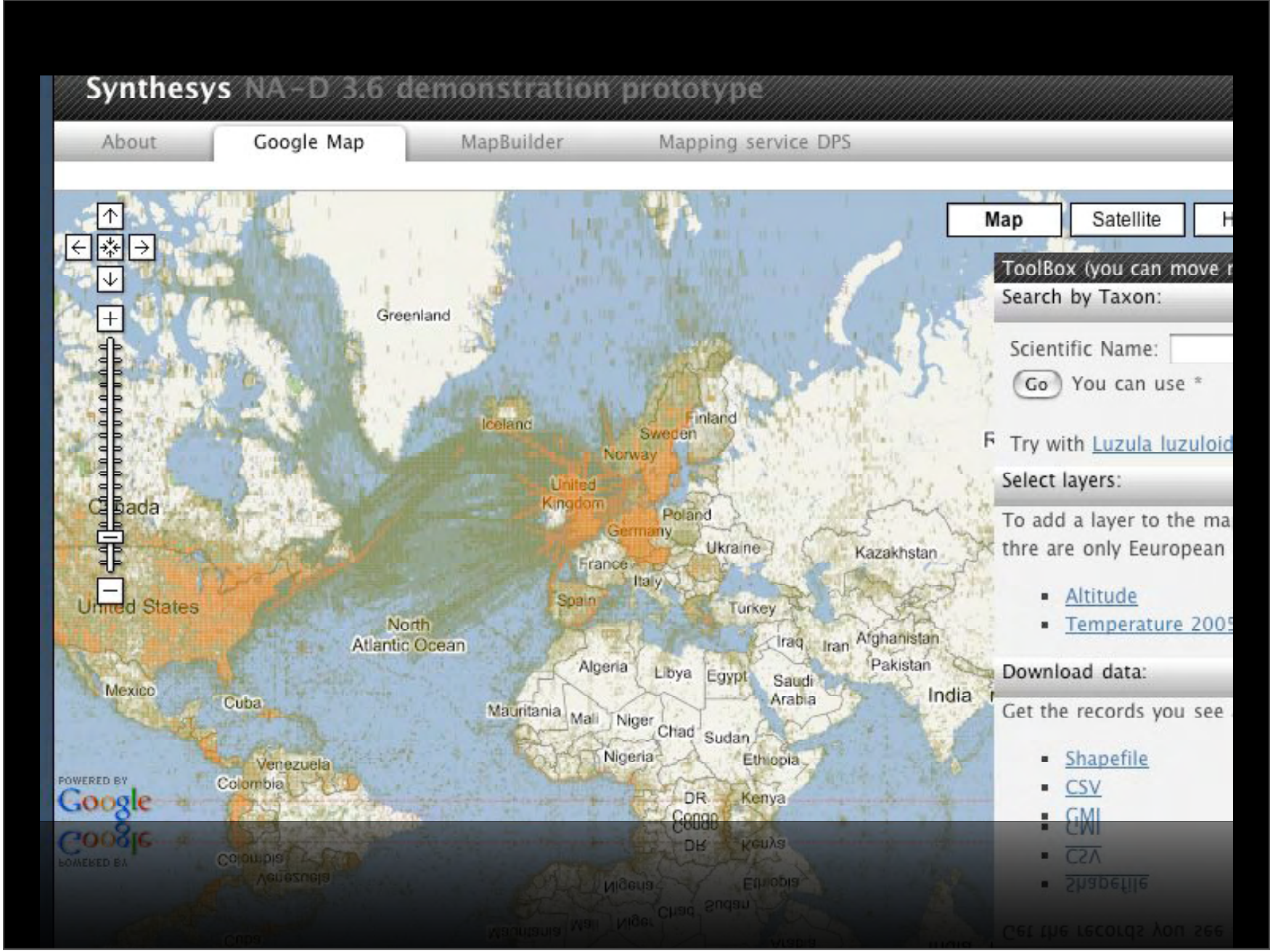
But to so we need lots of data

The most important iniciative right now on getting lots of this data is GBIF... more than 500 collections, like the one in madrid, are included. More than 180 million records... well Donald Hobern is the Deputy Director for Informatics and he is talking later... probably about TDWG and GBIF.

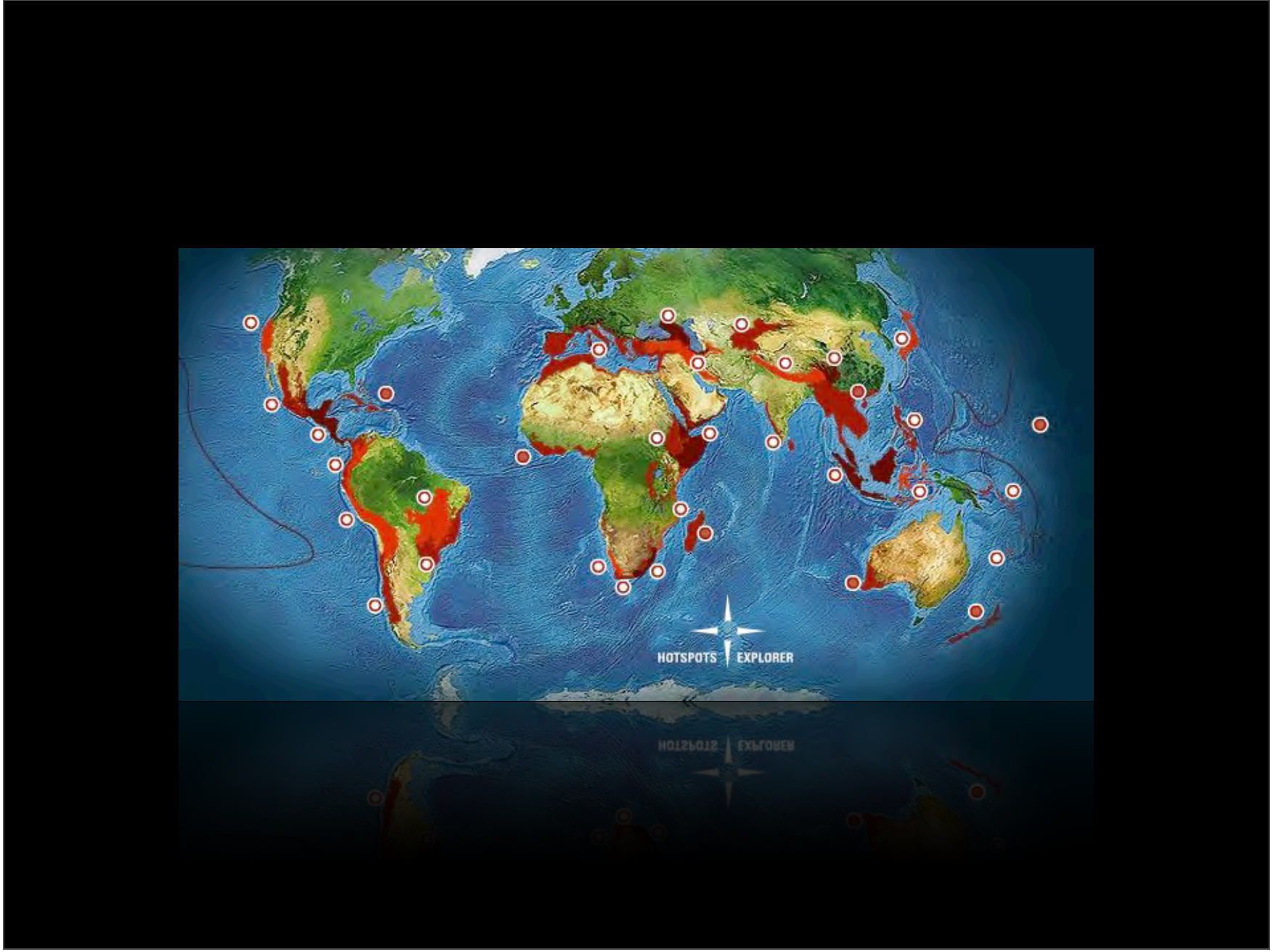
Well, just in simple, GBIF collects data from a distributed network of biodiversity databases from lots of collections...



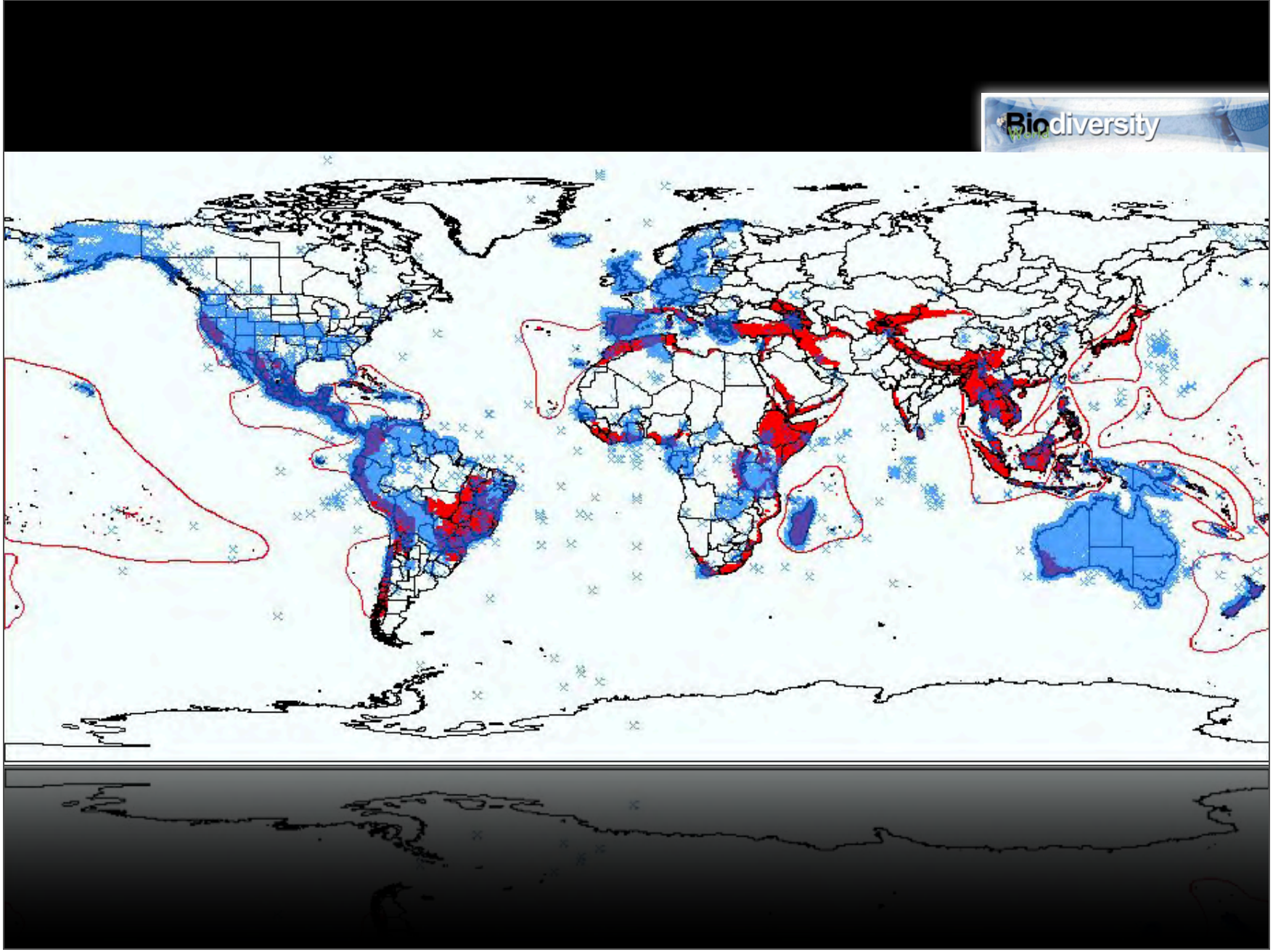
You have seen probably a diagram like this in any SDI presentation. The data is mantained by the source, the collections, and made avaialble for portals like GBIF to query them... not getting much in detail here



GBIF has a cache database to speed up analysis. And this is a map with data from GBIF. The color represent the amount of data... you can see how different the distribution of data is. Some country has lot of data some very little.



In fact, if we take a look at the supossely most important areas for biodiversity, the so called hot spots...

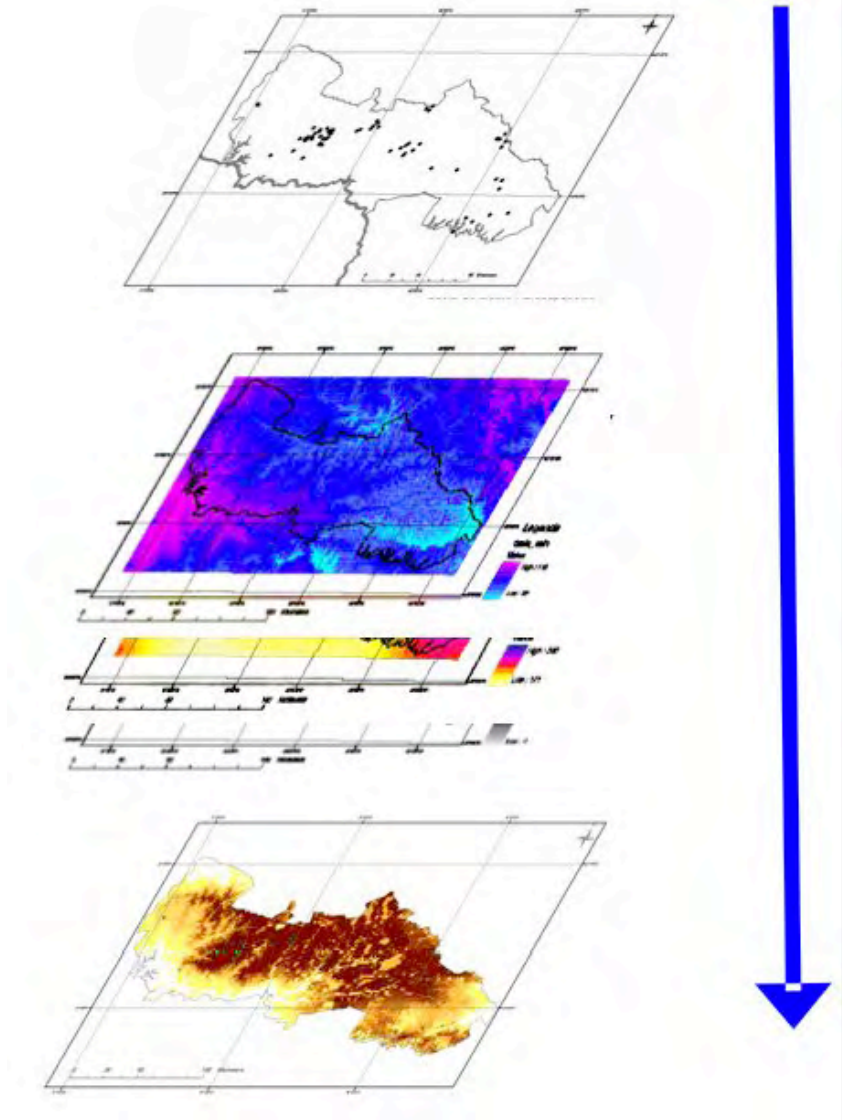


We can see that most data we have does not fall within them. Actually less than 10% of the data we have falls within a hotspot, and we know that at least 60% of world species occur there... so it is clear we dont have enough data actually to study.

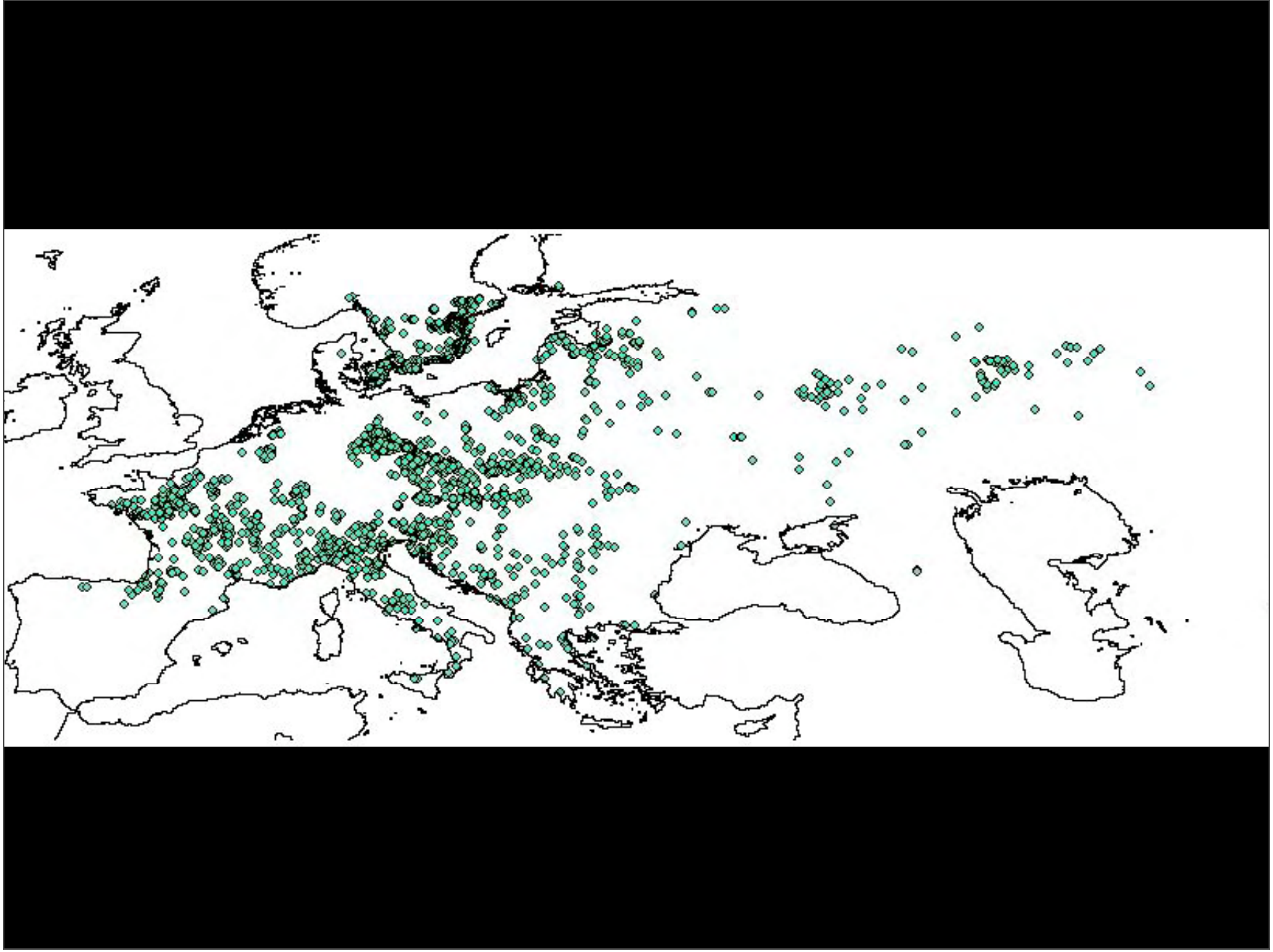
We need solutions to lack of data

So we dont have much data about where species occur... how are we gonna protect biodiversity if we dont know where it happens? well, we have a problem and we need a solution.

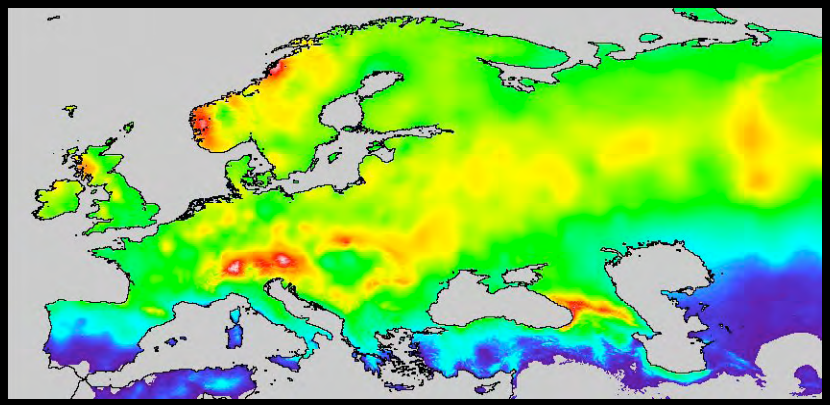
Niche Modelling



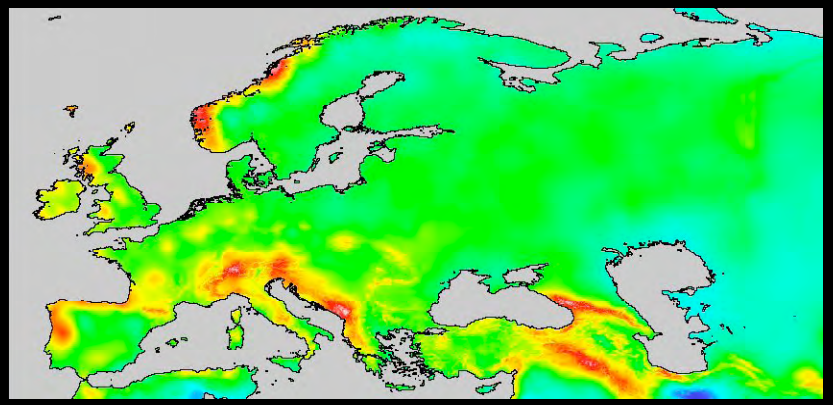
Niche modelling is one possible short term solution to the lack of data... lets take a look to it.



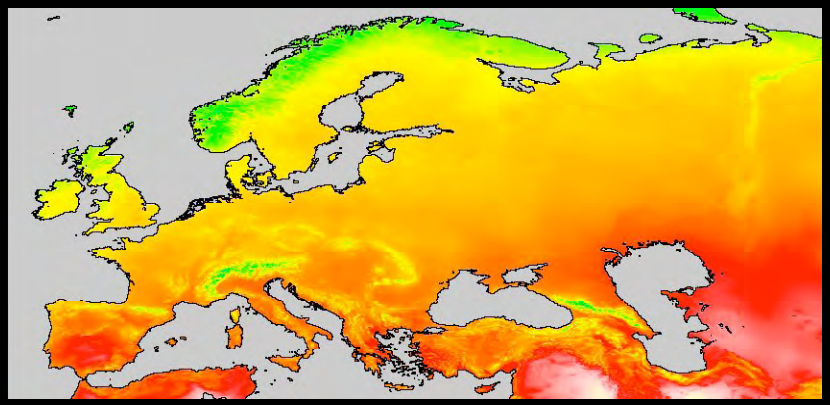
Using the incomplete data we already have...



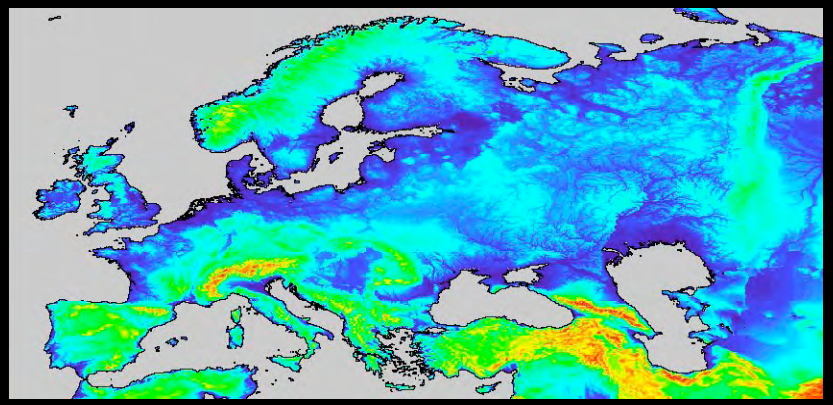
Summer precipitation



Winter precipitation

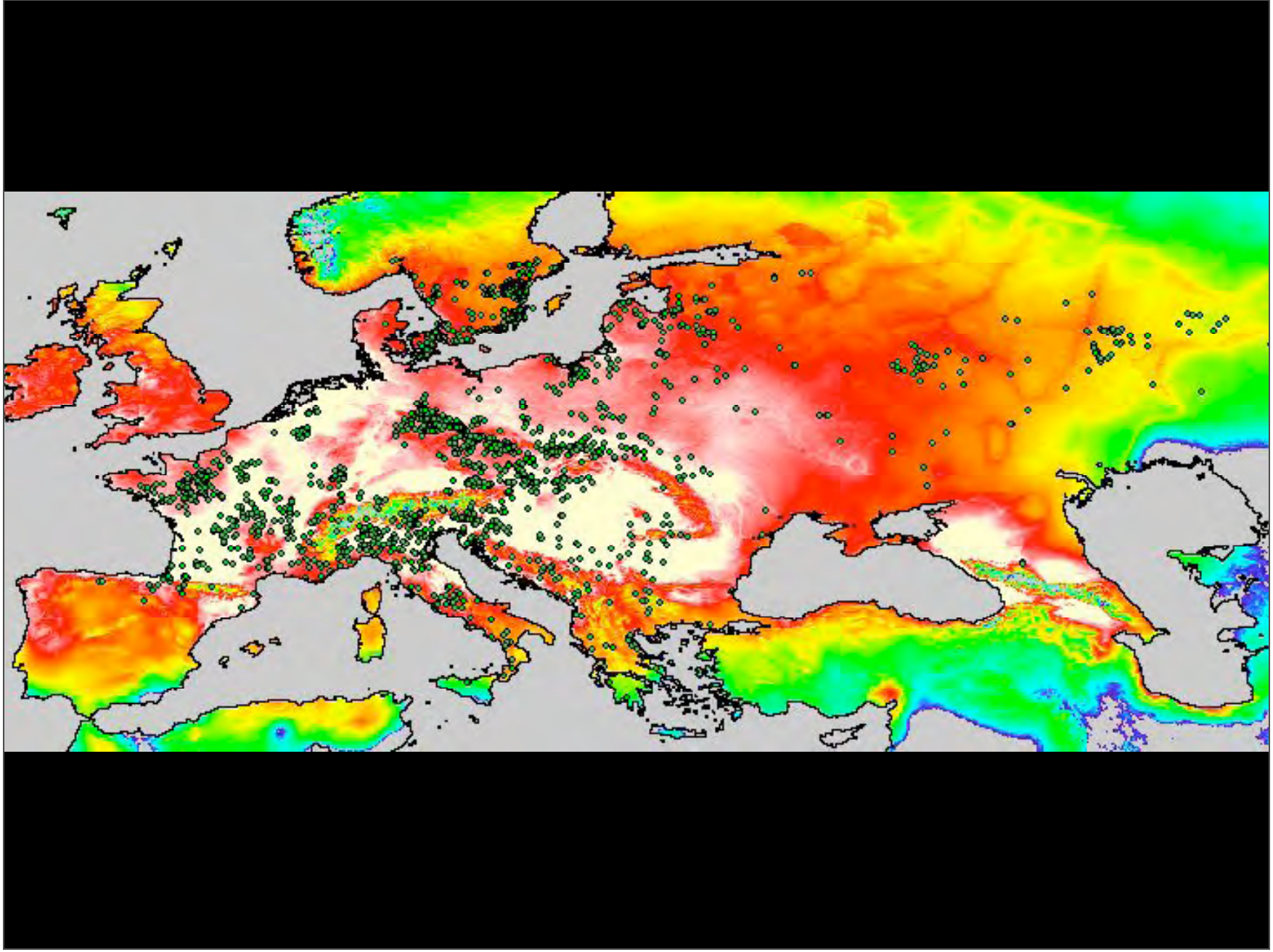


Maximum temp. summer

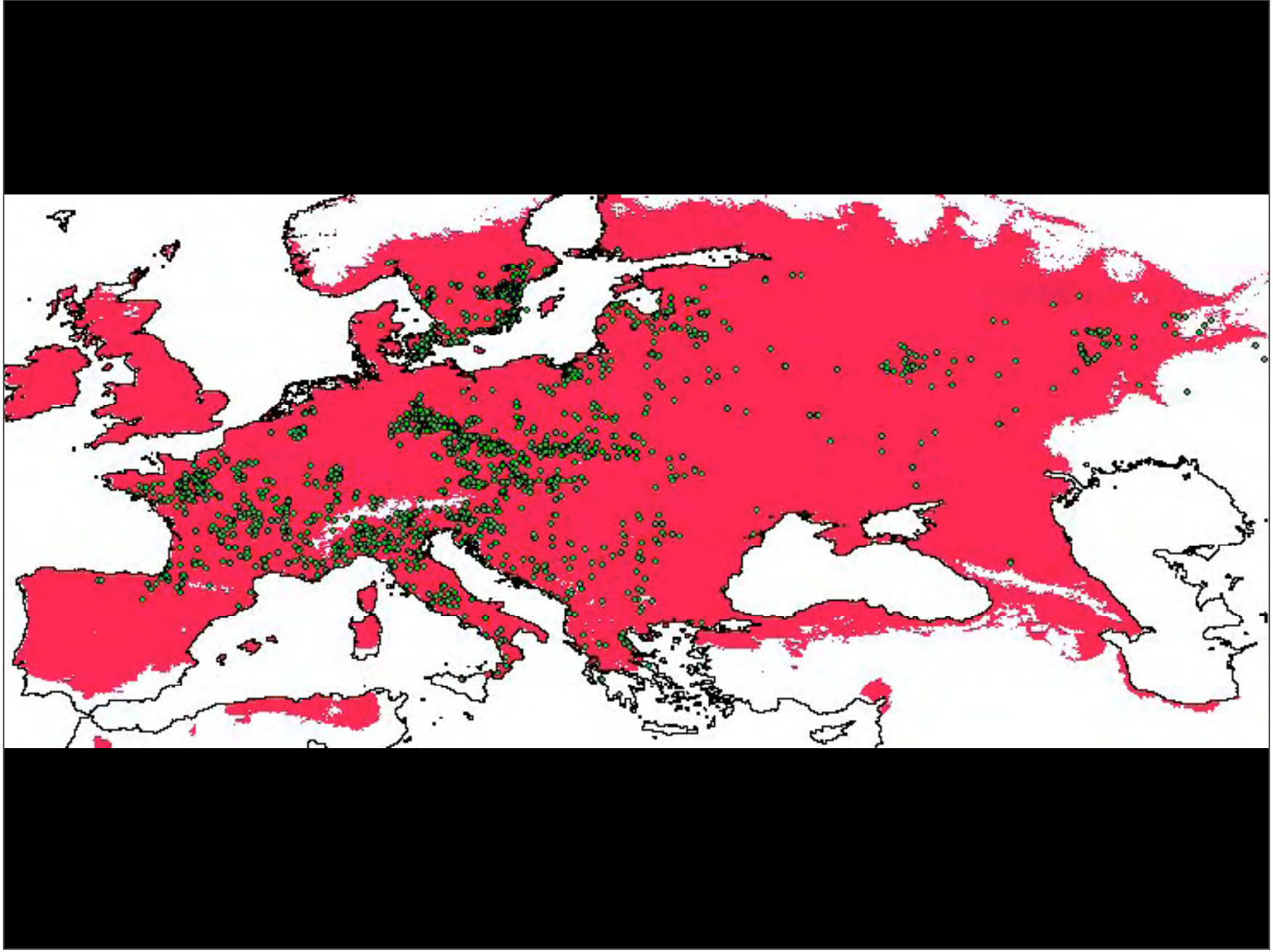


Average height

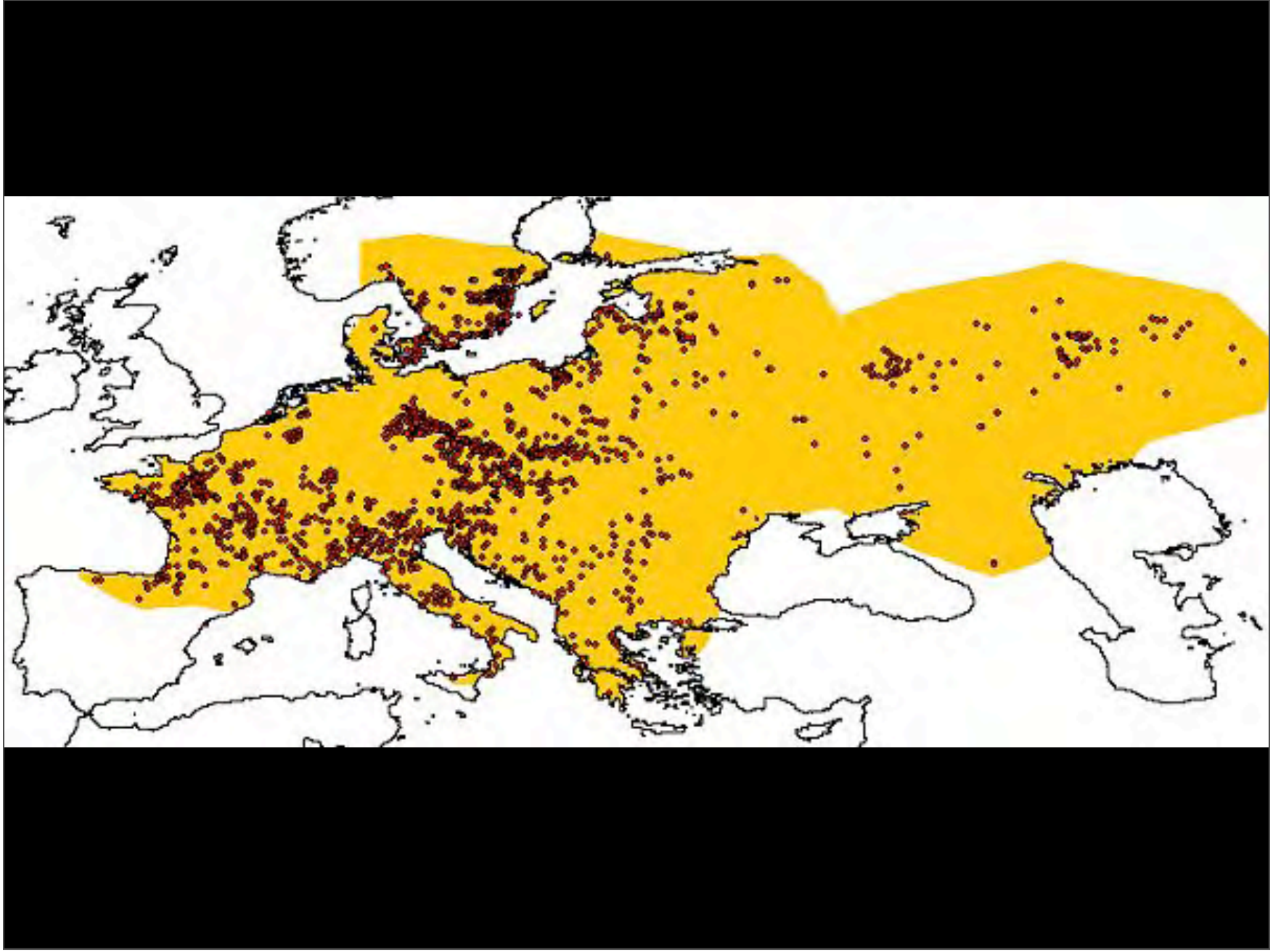
together with environmental data of the area...



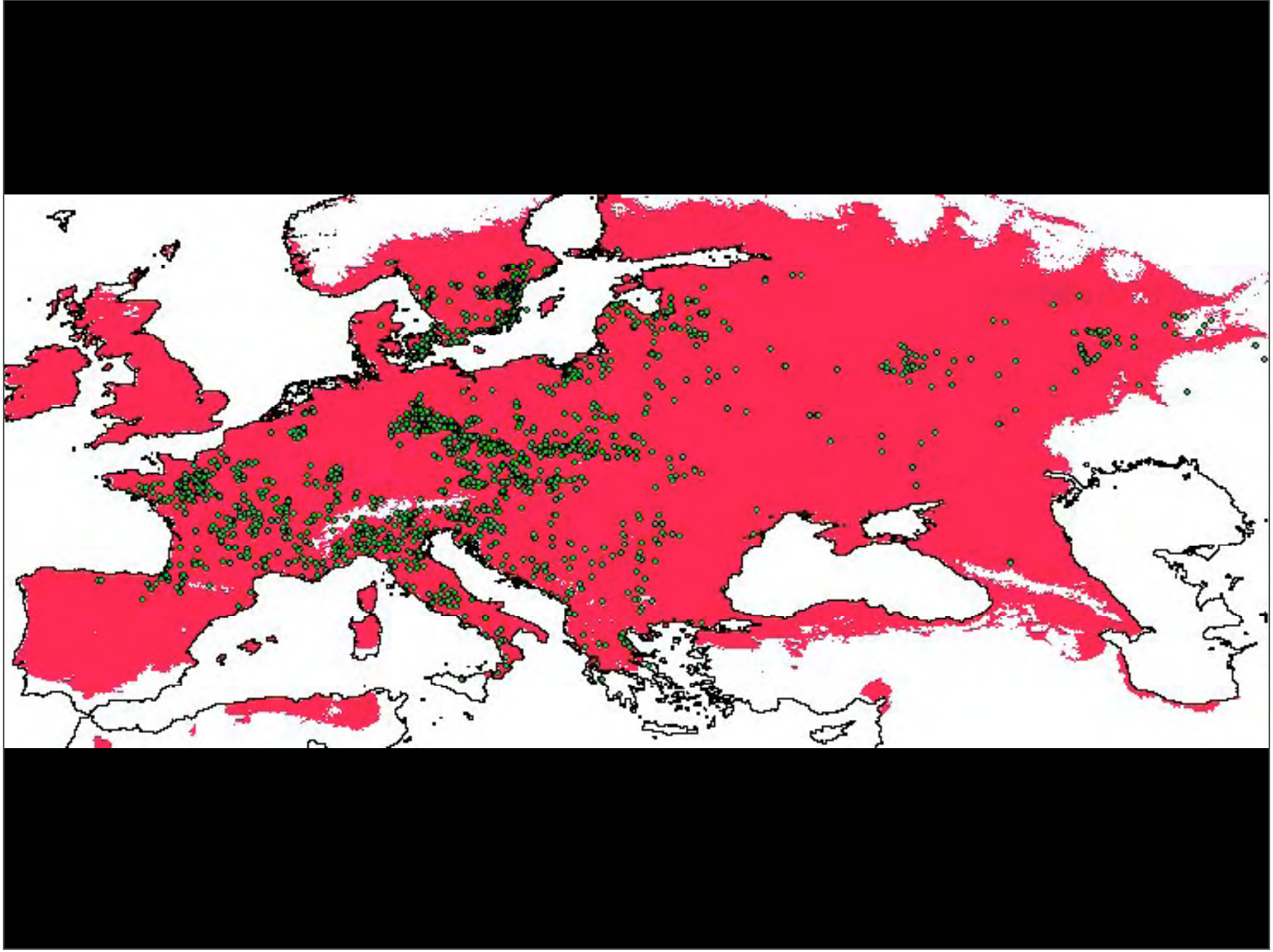
We can apply and algorithm to obtain a map of possible distribution of the specie based on the environmental where it lives.



Defining a threshold we can obtain a map of the potential distribution of this specie.

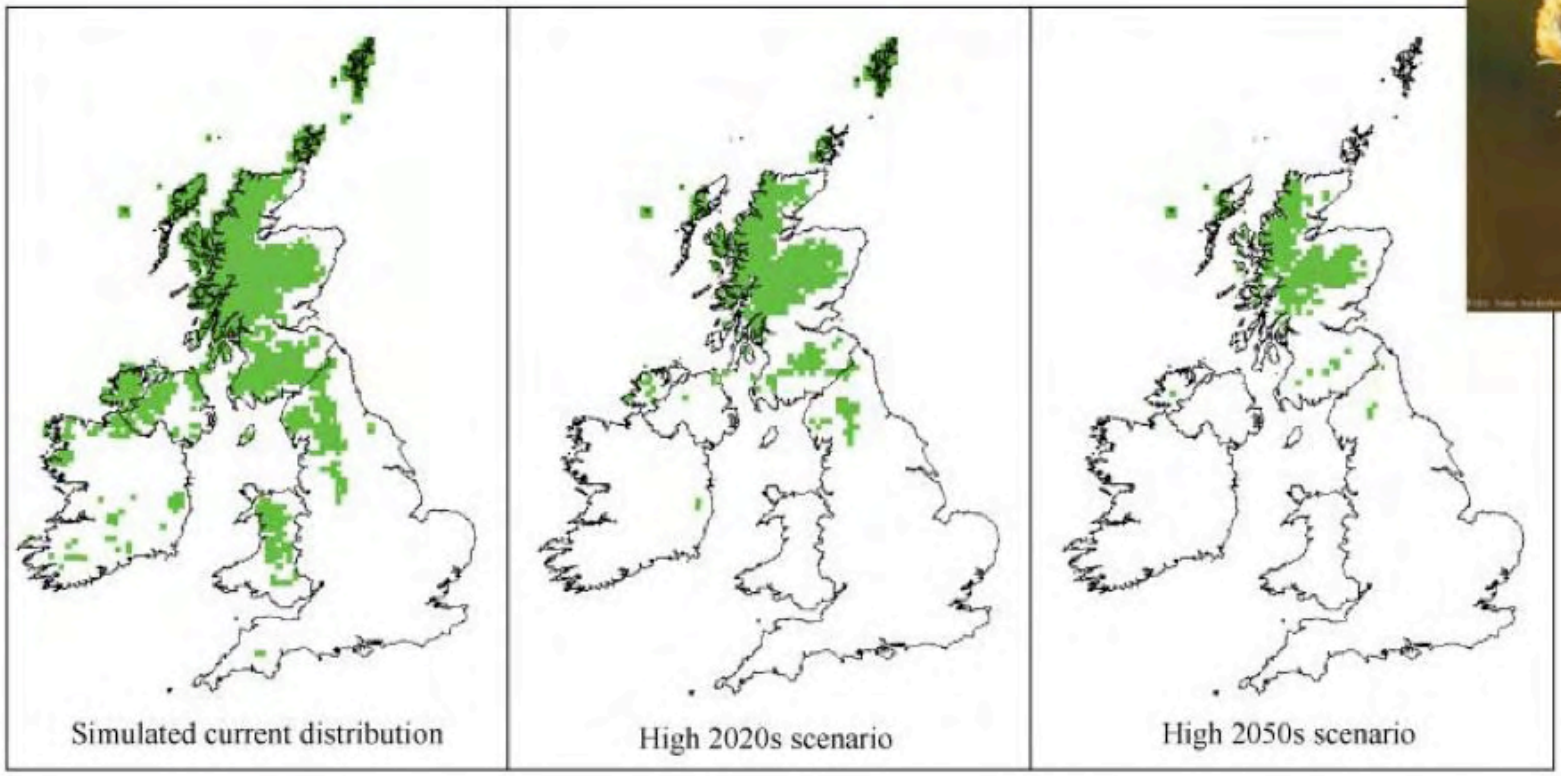


compared to the original
one



we can see that the distribution of the specie is potentially bigger. Probably if we go and collect on those places we will find this beatle. Because there is no enough information to get the real distribution of a specie using niche modelling we can infer it.

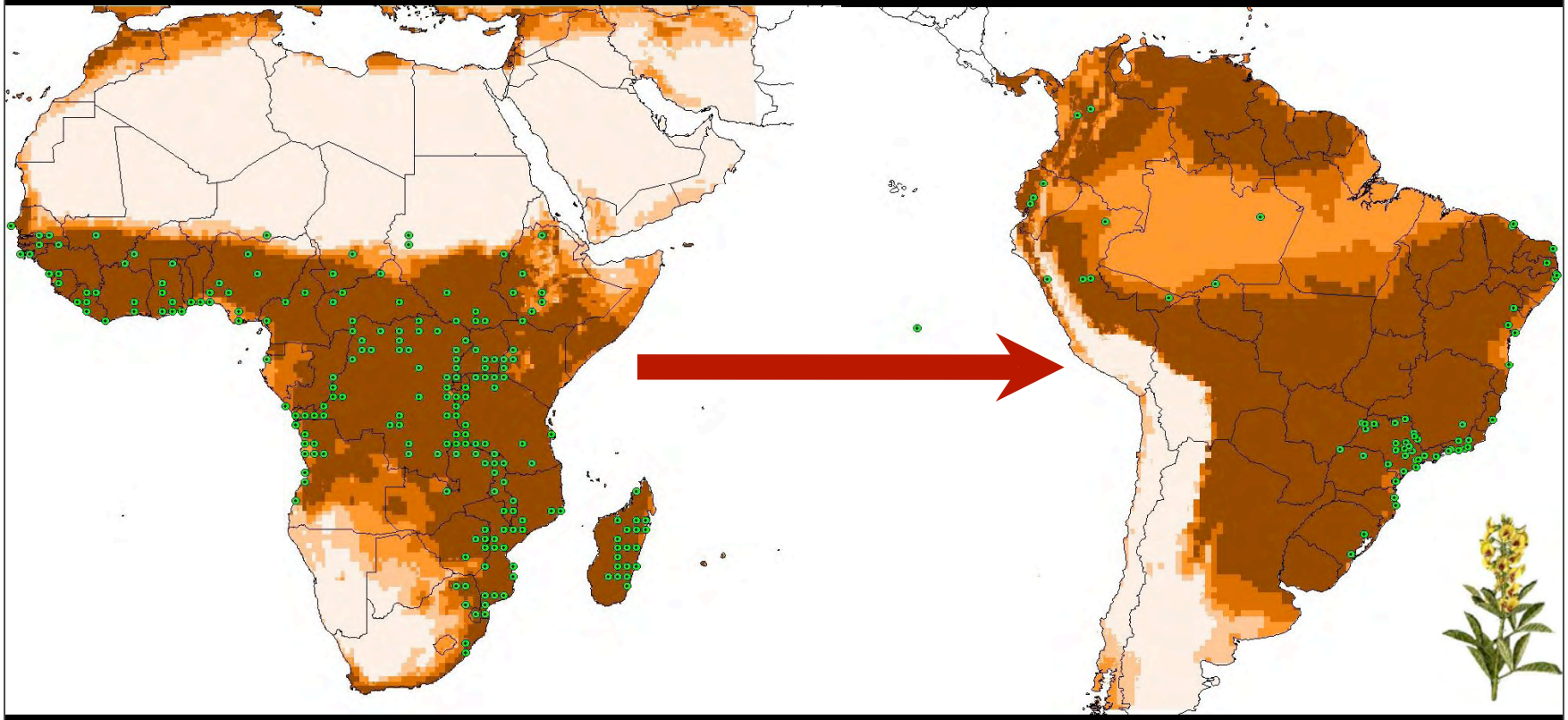
Distribution simulation of Carex bigelowii for different scenarios



Pearson et al., 2002

There is also a use for modelling for different scenarios, like future ones considering global change or warming. We can see how distributions are going to change.

Crotalaria pallida (FABACEAE)



Rafael Luís Fonseca

Or for example to analyze how and invasive specie is going to distribute in a new area...

Ocurrence data, specimen-observation data

+

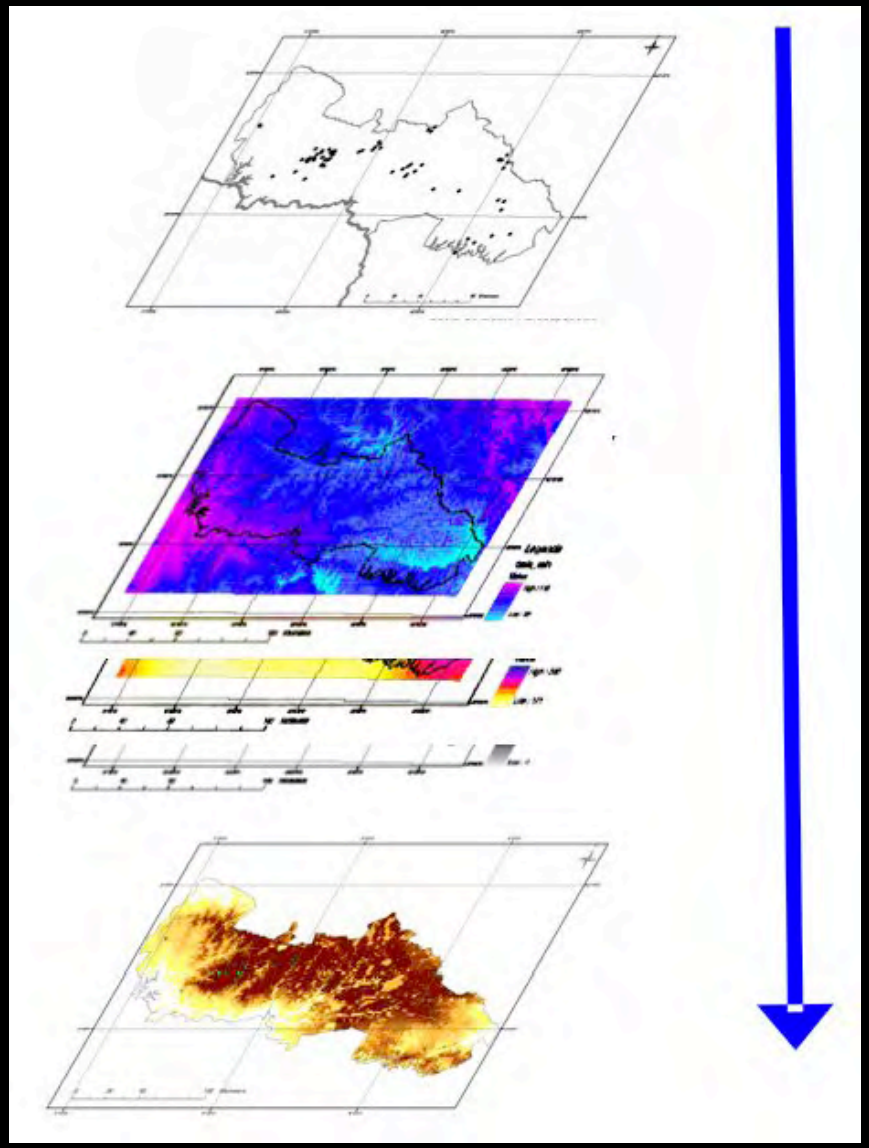
Environmental layers

+

Modelling algorithm

=

Distribution models



So at the end. Ocurrence data + environmental layers+modelling alorithms provide us distribution models.

BIOGEOSDI EXPERIMENT

Biodiversity Information Standards TDWG

Prototype

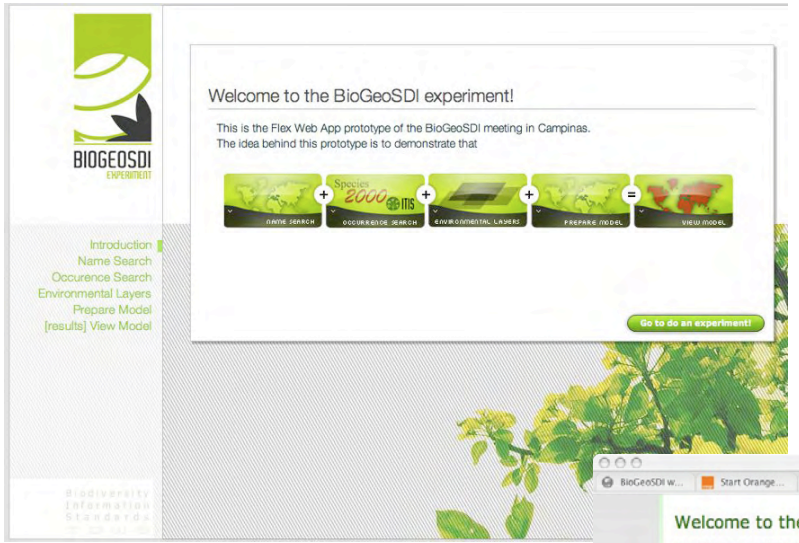
Prototype-proof of concept web application that uses several standards-technologies to do online niche modelling

http://omtest.cria.org.br/biogeosdi/

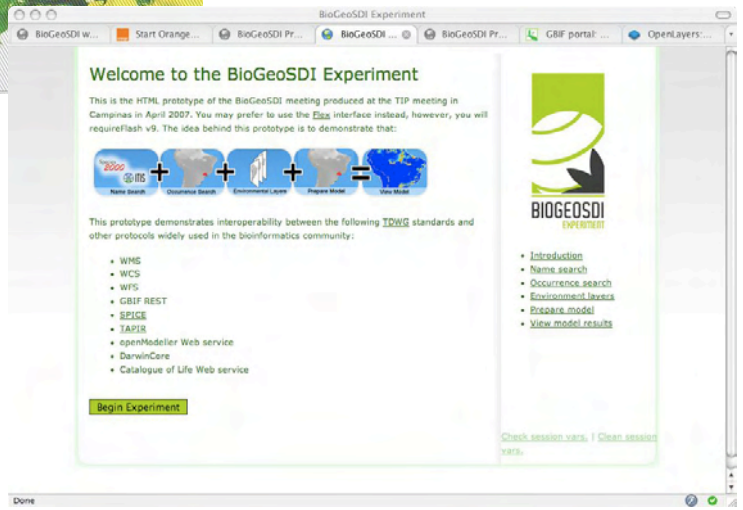
BioGeoSDI Code Fest in Campinas - April 2007

And this kind of experiments is the one we wanted to do in our prototype. In a web application that has a wizard style appearance we wanted to perform niche modelling experiments.

1) Start



Flex (flash) version

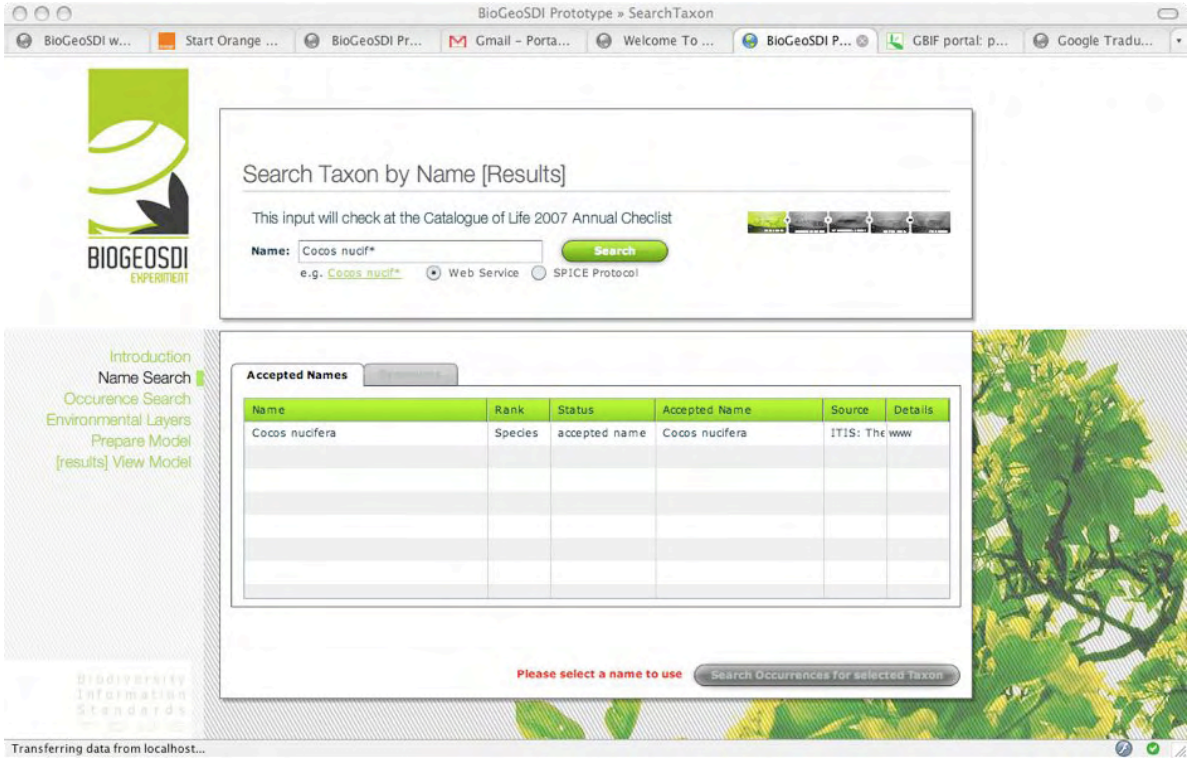


PHP-HTML version

Actually we had time to create two interfaces, one in Flex (flash) and another in PHP-HTML

2) Search for a scientific Name

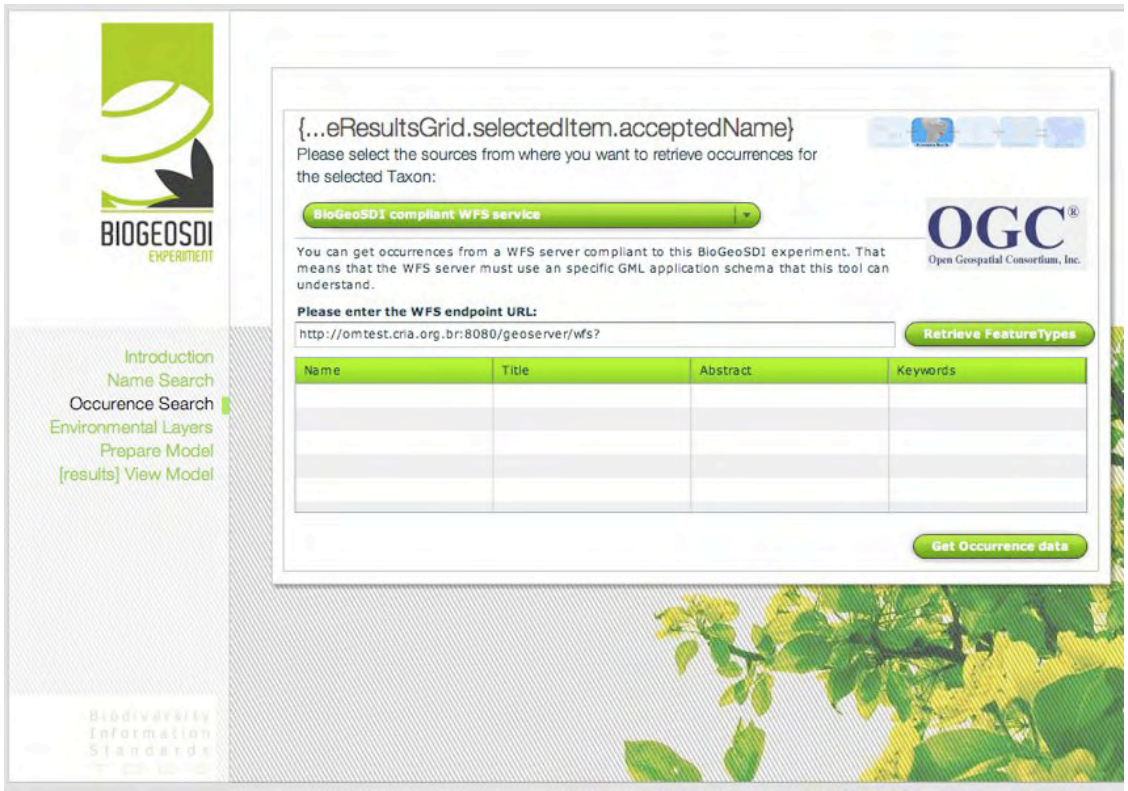
SPICE protocol, COL REST service



So the first you do when you enter the prototype is to search for a scientific name. Because you might be writing it wrong, or maybe it is not a valid name... we check for the name on a taxonomic service, actually two, in order to find the accepted name for the search that the uses do... this is the SPICE protocol and the Catalogue of Life REST services.

3) Search for ocurrences

WFS with our own GML app schema



After selecting the name what we want is to find ocurrences of this specie. Ocurrences can be model as features in a WFS service and therefore we gave the possibility for the user to provide a WFS endpoint. Make a getCapabilities, and search on a specific feature type. Of course the prototype only works with a specific GML app schema... because the is no official one now we created our own. Therefore the prototype can only talk with ourselves... not much of interoperability.

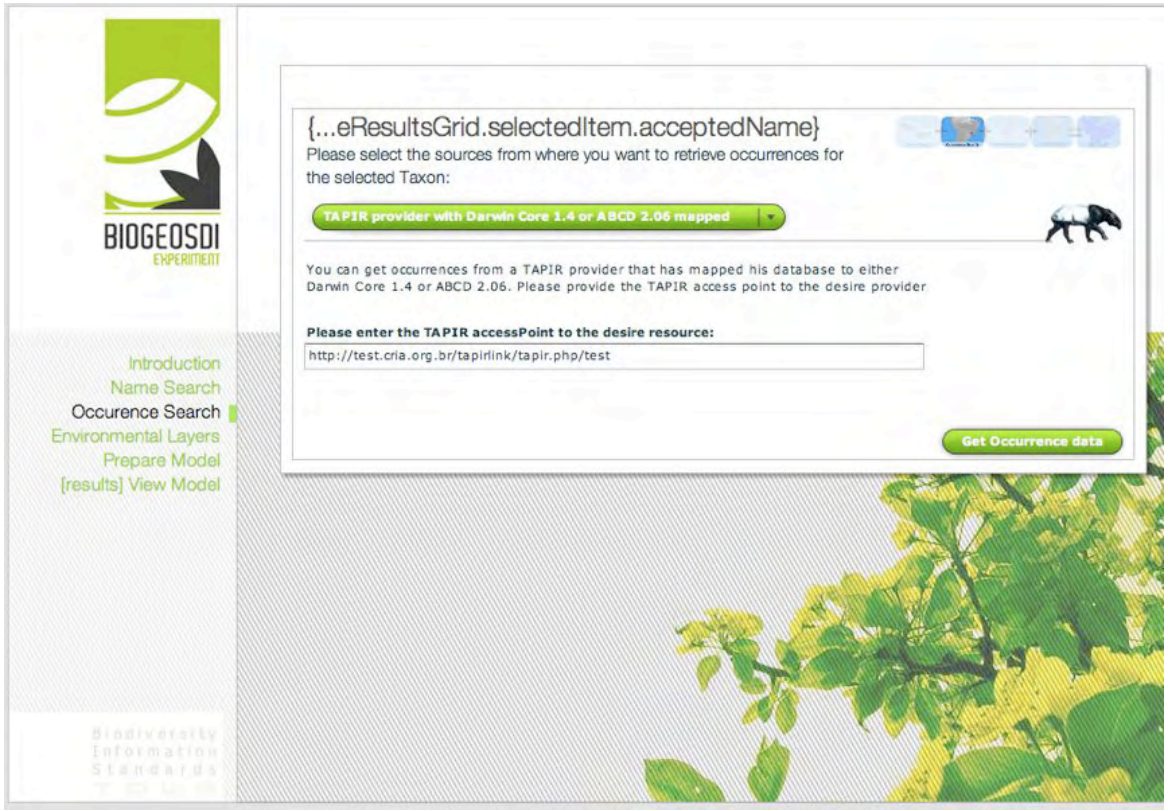
3) Search for ocurrences

GBIF REST service

We also gave the possibility to retrieve data from GBIF directly. They provide a nice and well documented REST service. The XML format behind the service its their invention. But definelty that was very easy to use.

3) Search for ocurrences

TAPIR protocol providers



Fianally the 3th posibility we provided was to use a specific TDWG transfer protocol TAPIR. Kind of a WFS but without the need to use GML...

3) Search for ocurrences (behind the scene)

TAPIR

GBIF

WFS

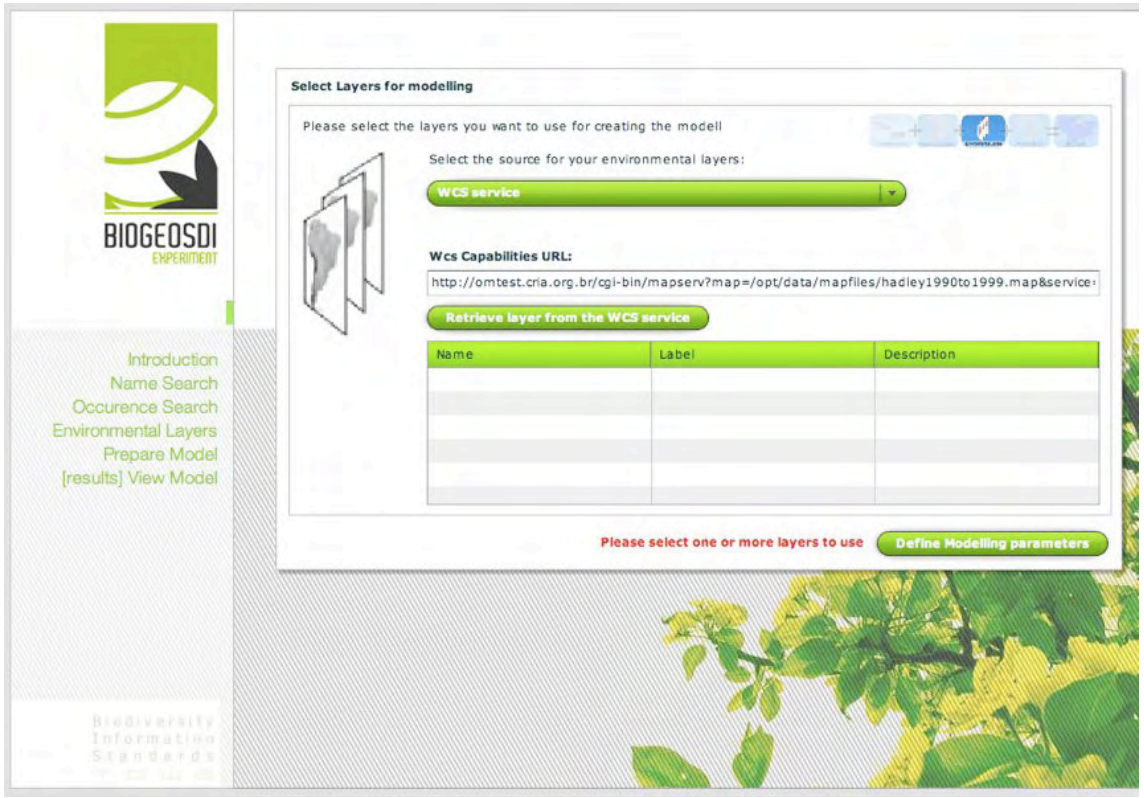
1. Retrieve from service
2. Store in a PostGIS table
3. Register in GeoServer as FeatureTypes (WMS,WFS) for later use

WMS,WFS, KML, PDF, PNG....

So, in reality behind the scene after retrieving the data from one of the sevrices we were actually storing it temporaily in a PostGIS table and register them in Geoserver as FeatureTypes. This way we could later make use of the data locally and generate KML, PDF, etc.

4) Get environmental layers

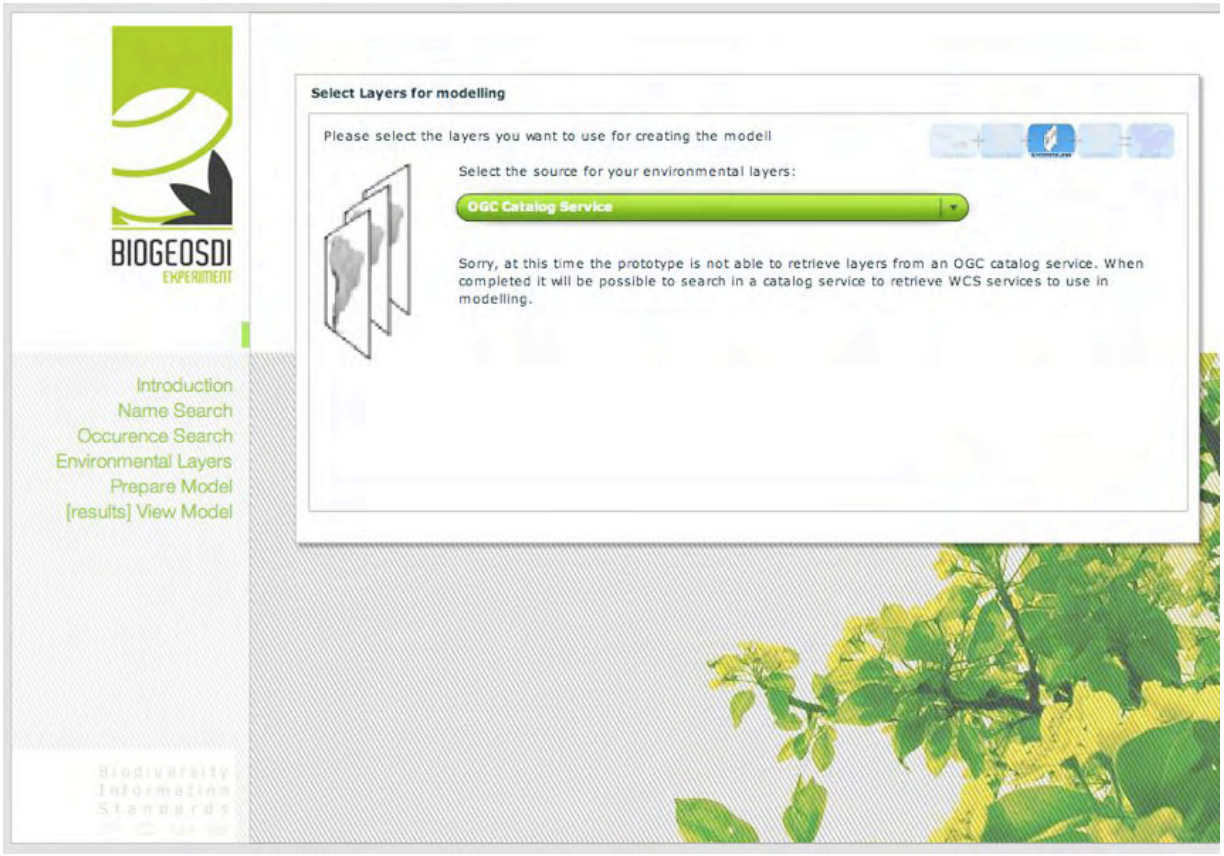
WCS



The next step is to get environmental layers. Again we provided 3 different ways to do so. First get them from a WCS server. Just introduce the endpoint, we do a getCapabilities and you can select which ones you wanna use in the expdriment.

4) Get environmental layers

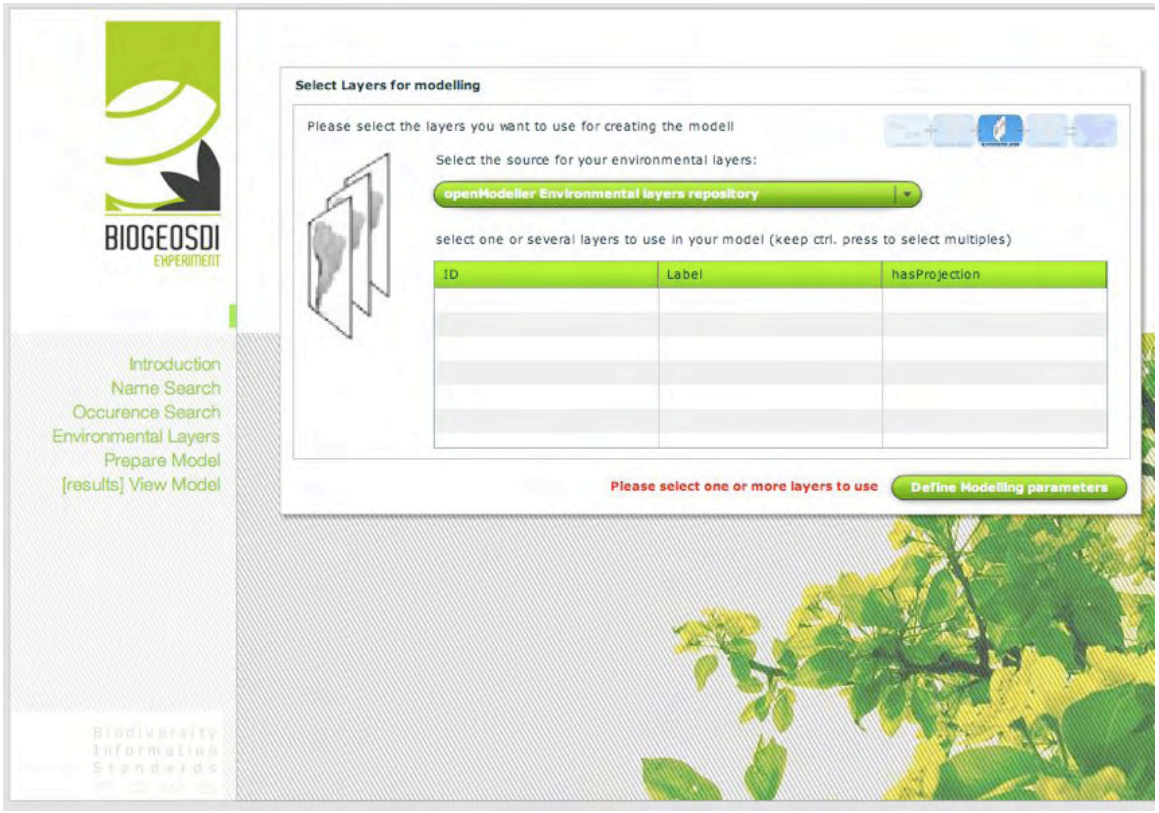
Searching on a catalog Service (not implemented)



We also envision the possibility that the user searches on a Catalog Service... biut we did not have time to implement anything like this :(

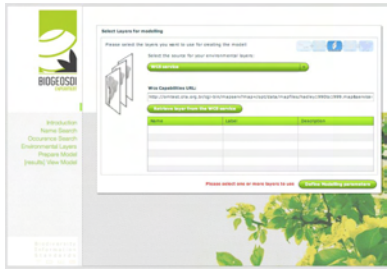
4) Get environmental layers

Using Modelling service local layers (OMWS getAvailableLayers)



The last way to retrieve layers is to connect to the Modelling web service and ask for a list of locally avaialble layers to make use of.

4) Get environmental layers (behind the scene)

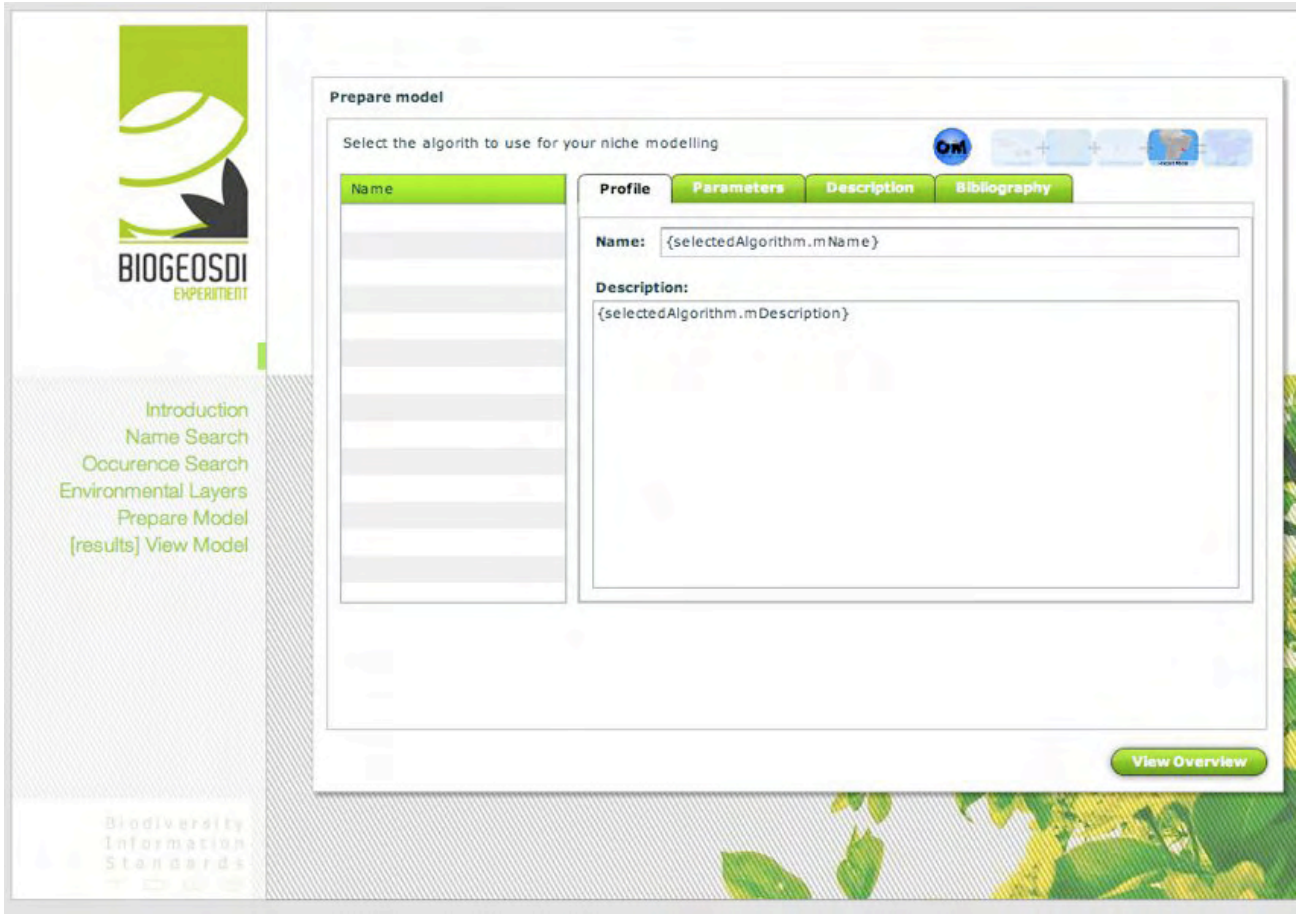


1. If it is a WCS layer retrieve it
2. Save it in Server
3. Register it in OpenModeller service to make it available for modelling

Again what we were doing behind the scenes... if the layer is avaialble in a remote WCS we doanlowaded to the modelling server to have it ready when doing the experiment.

5) Select Model algorithm & parameters

OMWS getAlgorithms



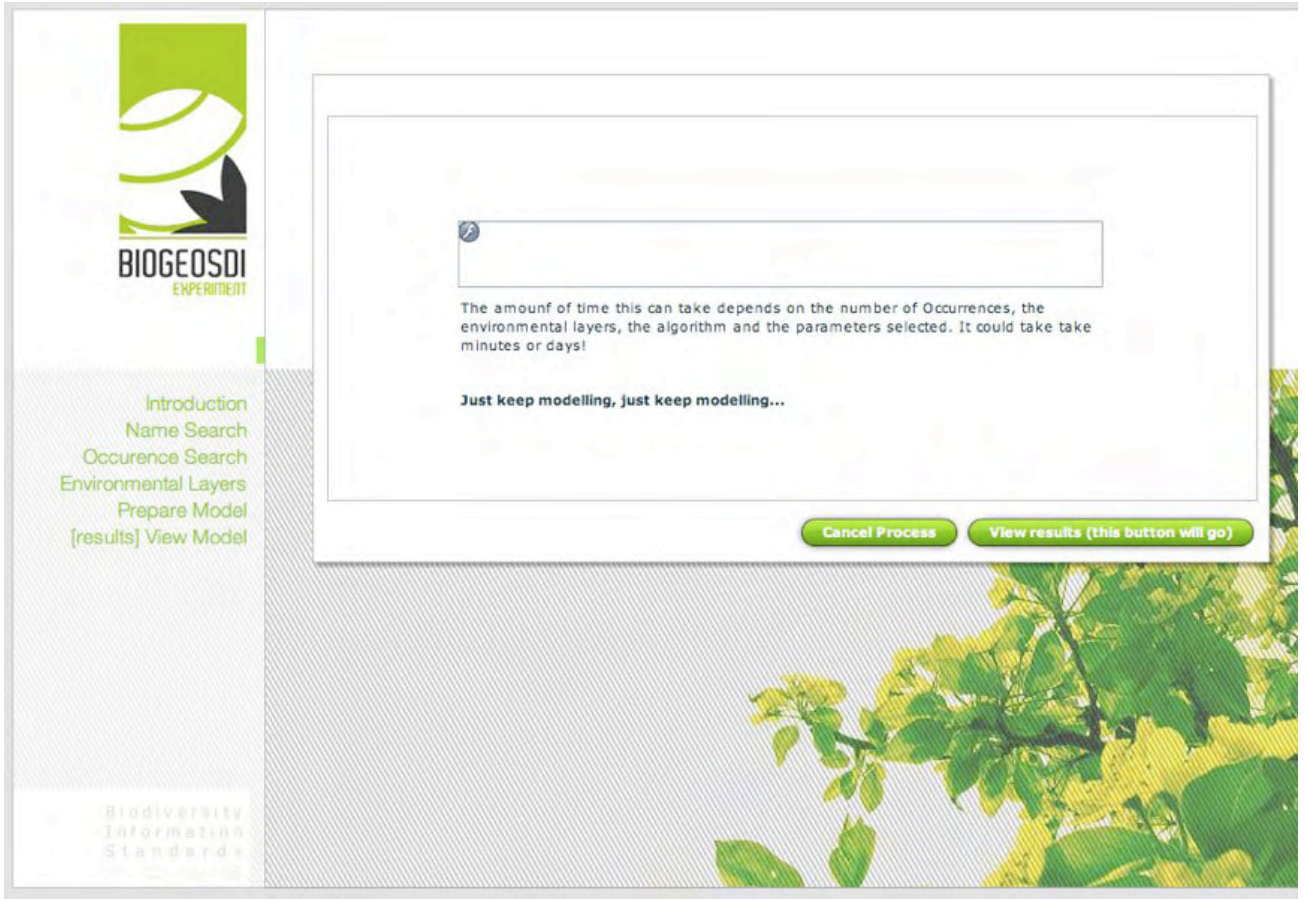
Next step is to select the Algorithm you want to use for modelling, there quite some developed in the last years. You should be able to select some parameters but for this prototype we always take the default ones... just a simplification.

6) Overview before modelling

Before going to modelling we wanted to give an overview to the user of the data he has collected for the experiment...

7) Run model

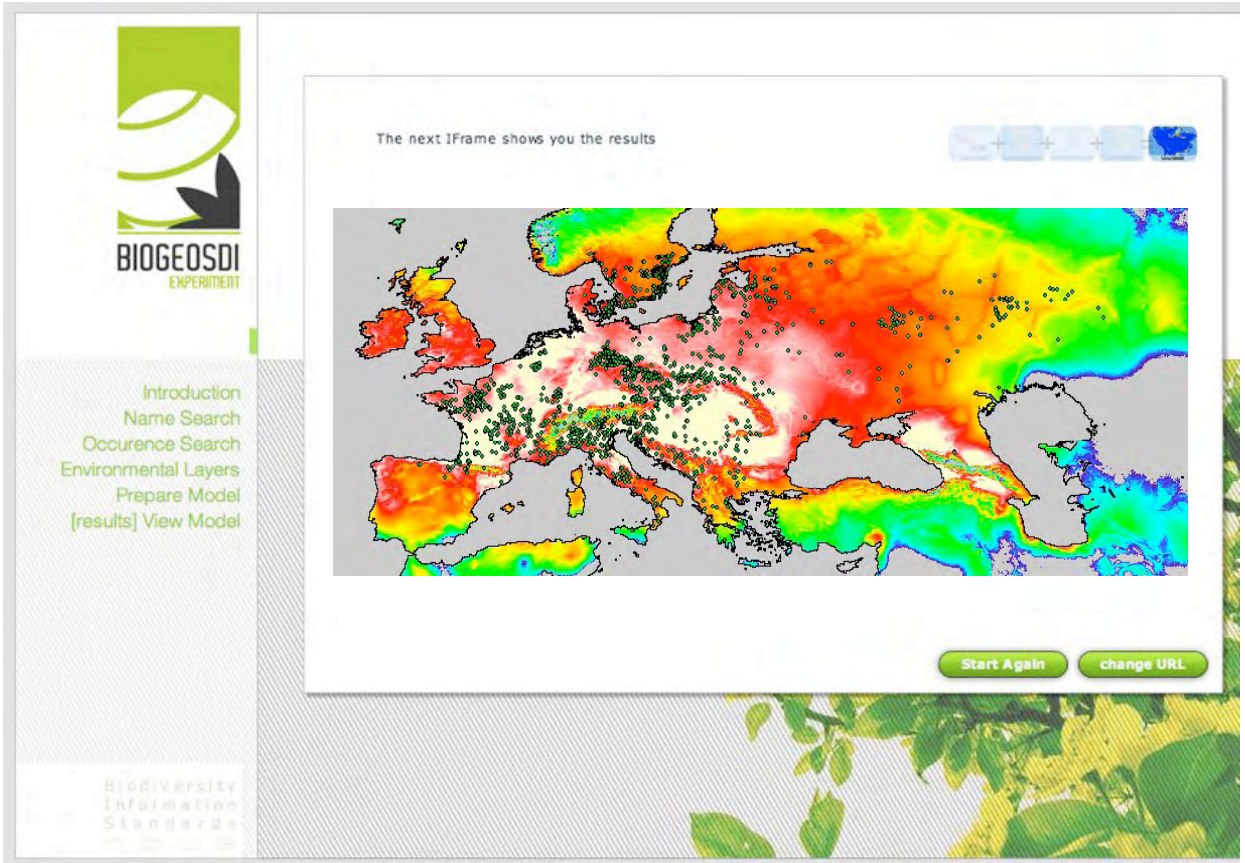
OMWS runModel, getLog (SOAP service)



And... finally run the model... For this we are using the OpenModeller web service... this can take a lot of time... the service is implemented in a SOAP service, but one cool thing here would be to make it available as a WPS service... The SOAP service has already an asychronous mecahnism... modelling can take days to be performed!

8) View Result

WMS, WFS



So finally... this is a fake screen I have to recognize... the results of the model would appear in a map... again in behind the scenes what we are doing is registering the modelling result as a WMS service and overlay it in an OGC client with the points we registered at the beggining.

Technologies used

- Geoserver (WMS, WFS, KML)
- PostGIS
- Mapserver (WCS)
- OpenModeller (OMWS)

- PHP, Flex, HTML, Bash, Python

So this are the technologies we used...

Report

- OGC standards documentation complicate for consumers
- WFS issues with lot of data
  - No paging
  - Huge messages
  - Capabilities do not show GML app schema behind a FeatureType
- GML app schemas needed because we cant use our own from TDWG.
- Filters in WFS GET request are difficult.
- WMS slow when using together with Google Maps, need to cache tiles.
- SPICE protocol issues
- OMWS, use of SOAP Document/literal complicate!
- Cool REST services from GBIF
- Lack of TAPIR registries
- Quality of data

And here are some hints on things we found during developing this thing...

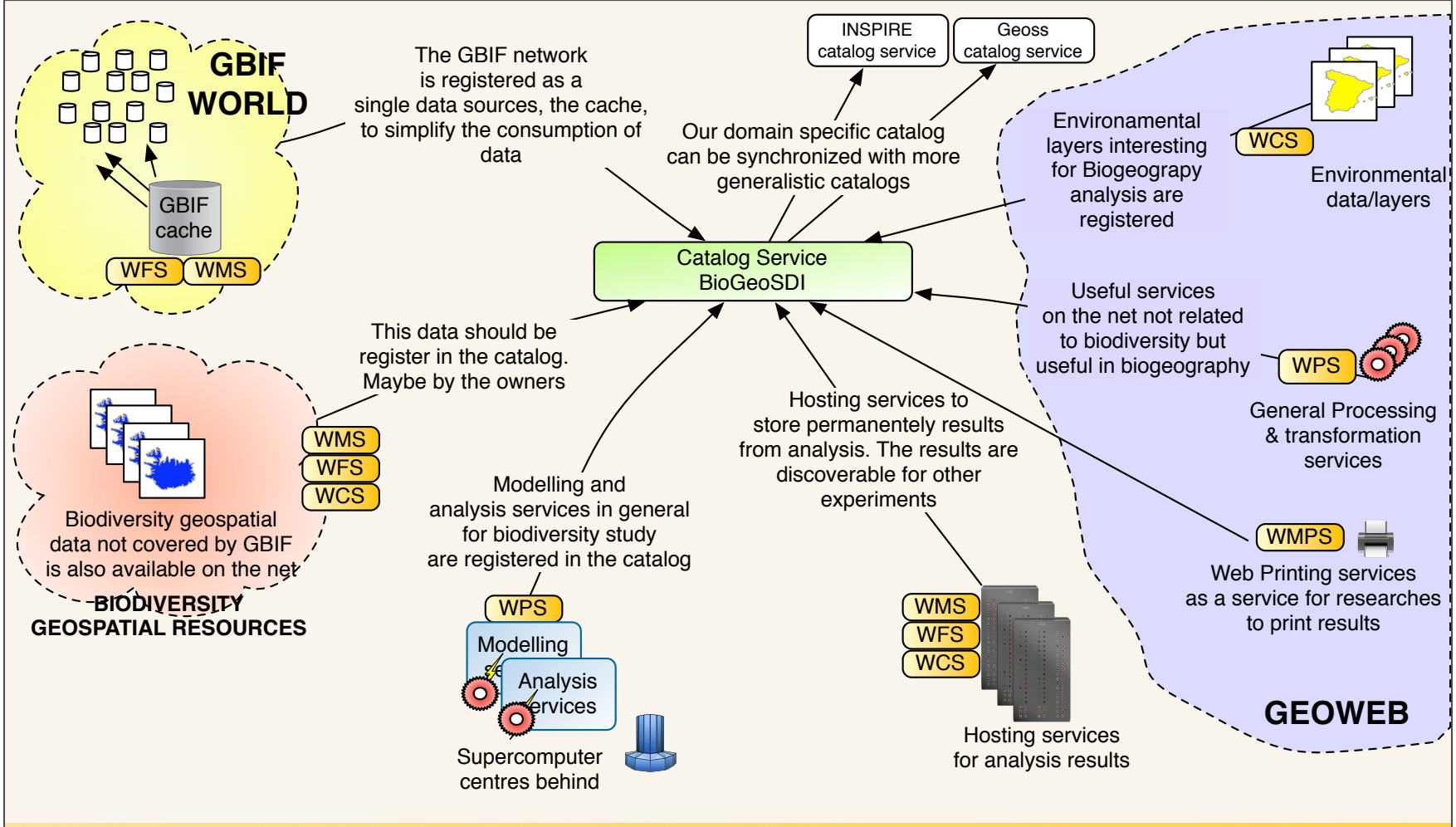
Thanks!

Workshop Wiki
http://wiki.tdwg.org/twiki/bin/view/Geospatial/InteroperabilityWorkshop1

Prototype
http://omtest.cria.org.br/biogeosdi/frontend/

Source Code and documentation
http://biogeosdi.org/

Report
http://omtest.cria.org.br/biogeosdi/doc/report/BioGeoSDIreport.html



Biogeography Spatial Data Infrastructure
BioGeo Spatial Data Infrastructure

This not to be seen...