



BIOGEOSDI
EXPERIMENT
Biodiversity
Information
Standards
TDWG

BioGeoSDI

workshop

Using TDWG and OGC standards
together

Javier de la Torre
jatorre@imaste-ips.com

IMASTE_{IPS}



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

The Campinas workshop



- 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-ipr.com)



<http://wiki.tdwg.org/twiki/bin/view/Geospatial/GeoApainter>

BioGeoSDI Code Fest in Campinas - April 2007

2

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



Biodiversity
Information
Standards
T D W G

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>

BioGeoSDI Code Fest in Campinas - April 2007

3

The objective 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.



BIOGEOSDI
EXPERIMENT

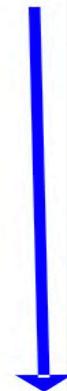
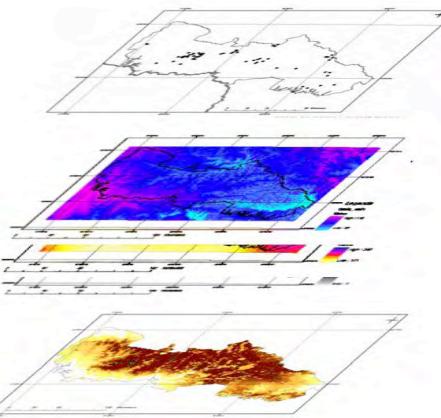
Biodiversity
Information
Standards
TDWG

A simple use case for a week

“

Save Biodiversity using TDWG and OGC standards

“



“

Online niche modelling experiments using web services

“

BioGeoSDI Code Fest in Campinas - April 2007

4

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.



5

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



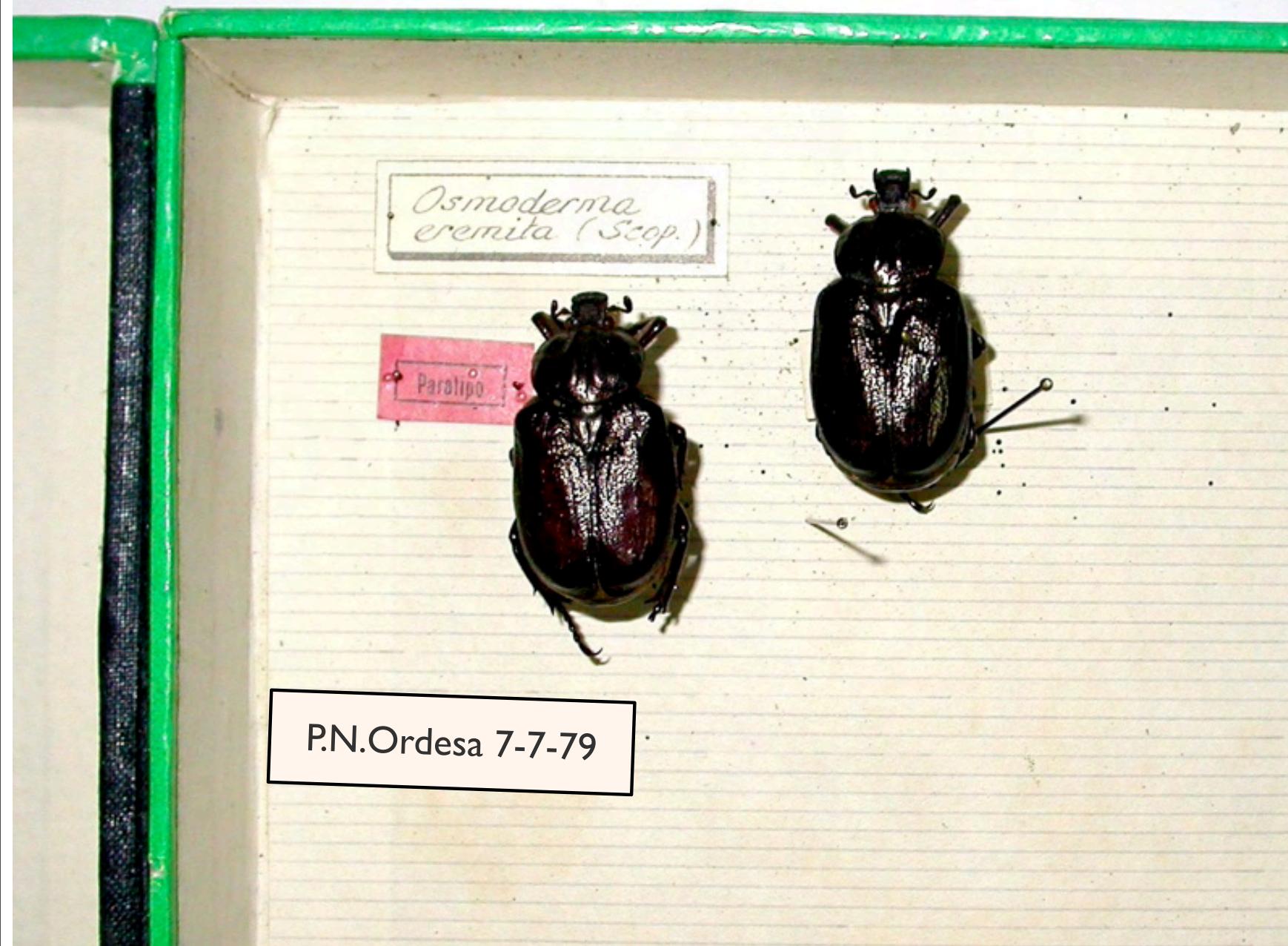
6

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....



7

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.



8

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.



9

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



7 de Julio de 1979

10

In the
Pyrenees



11

In a national
park



12

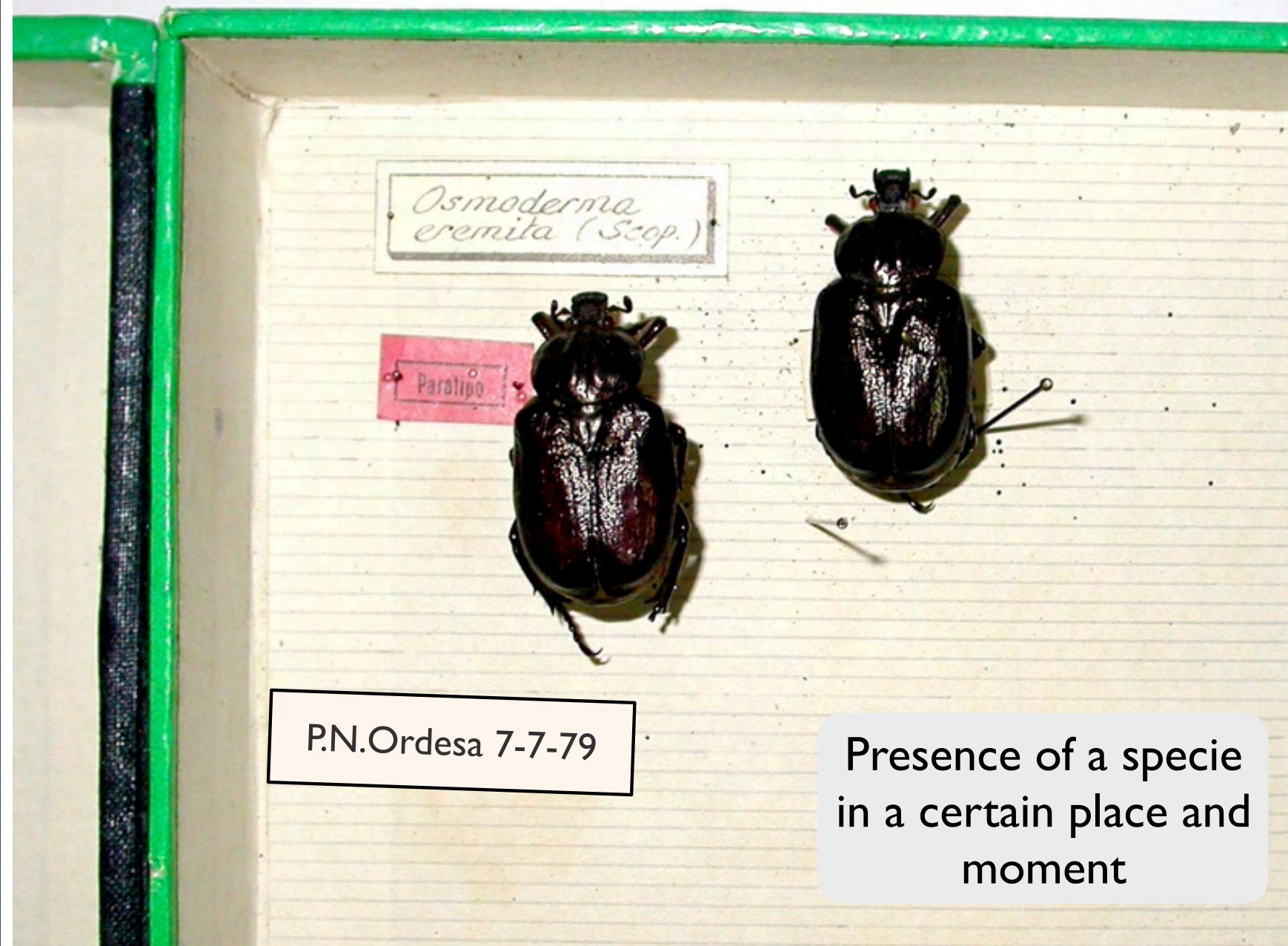
some biologist was
collecting ...



Osmoderma eremita

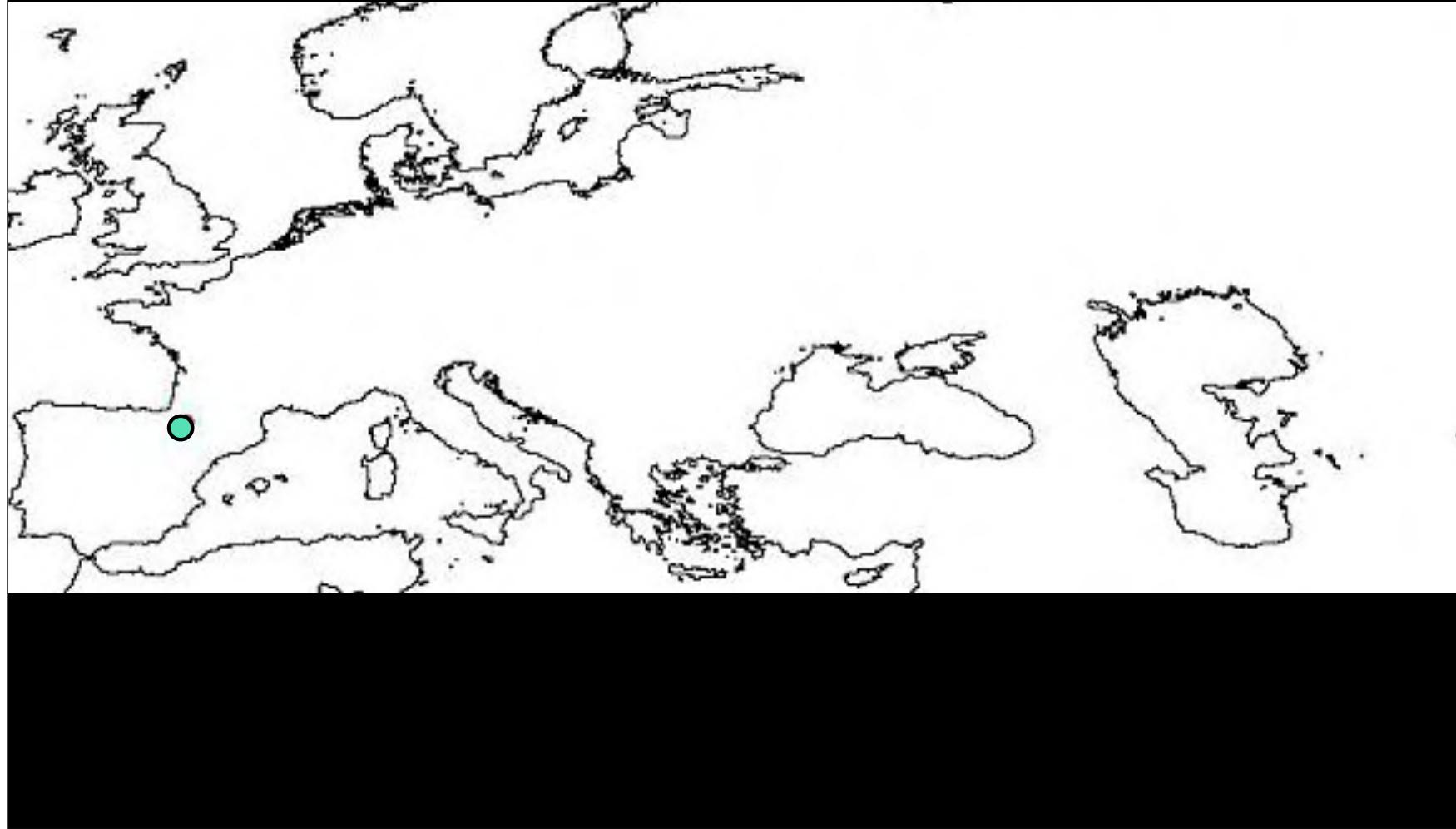
13

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



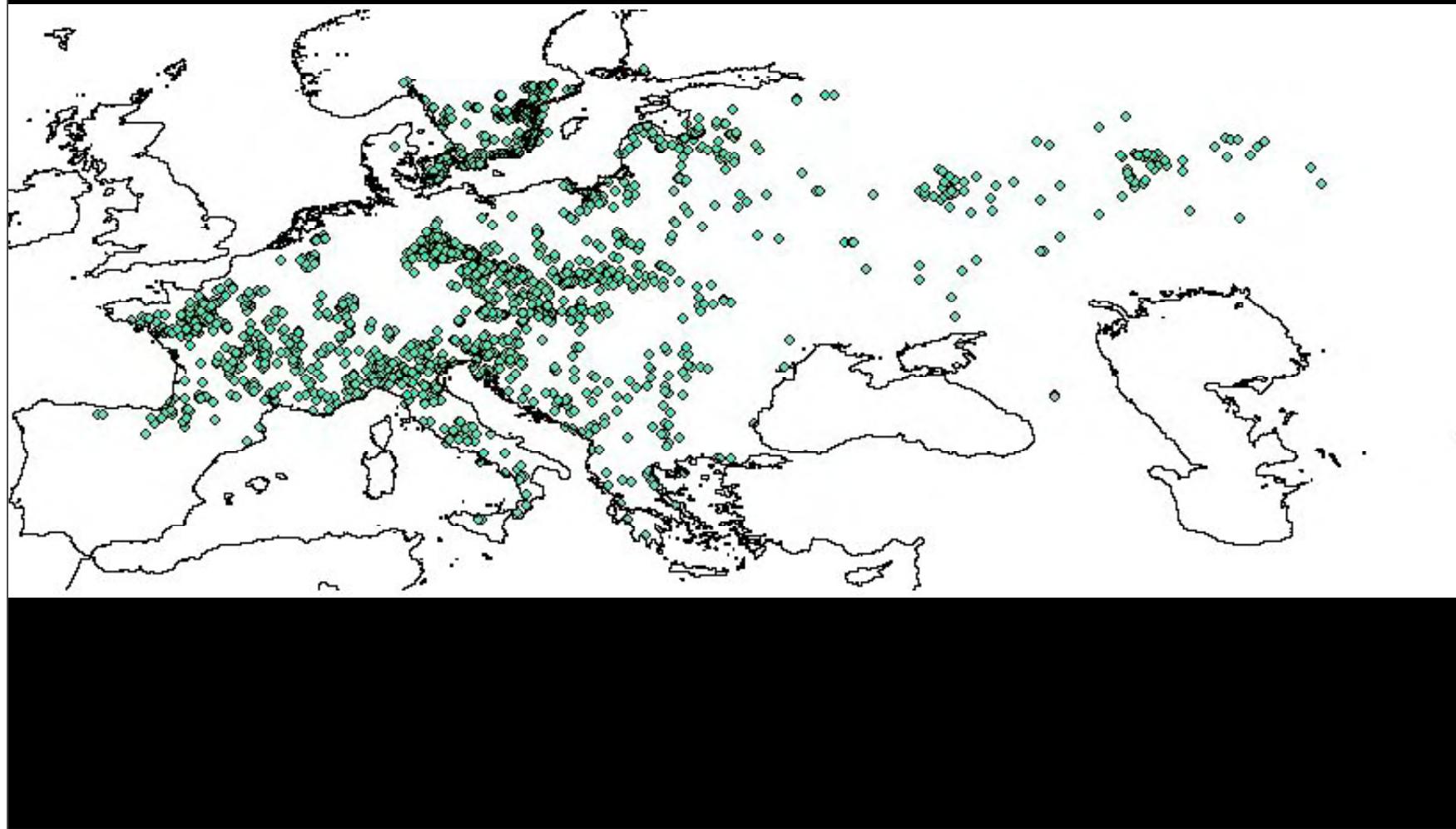
14

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



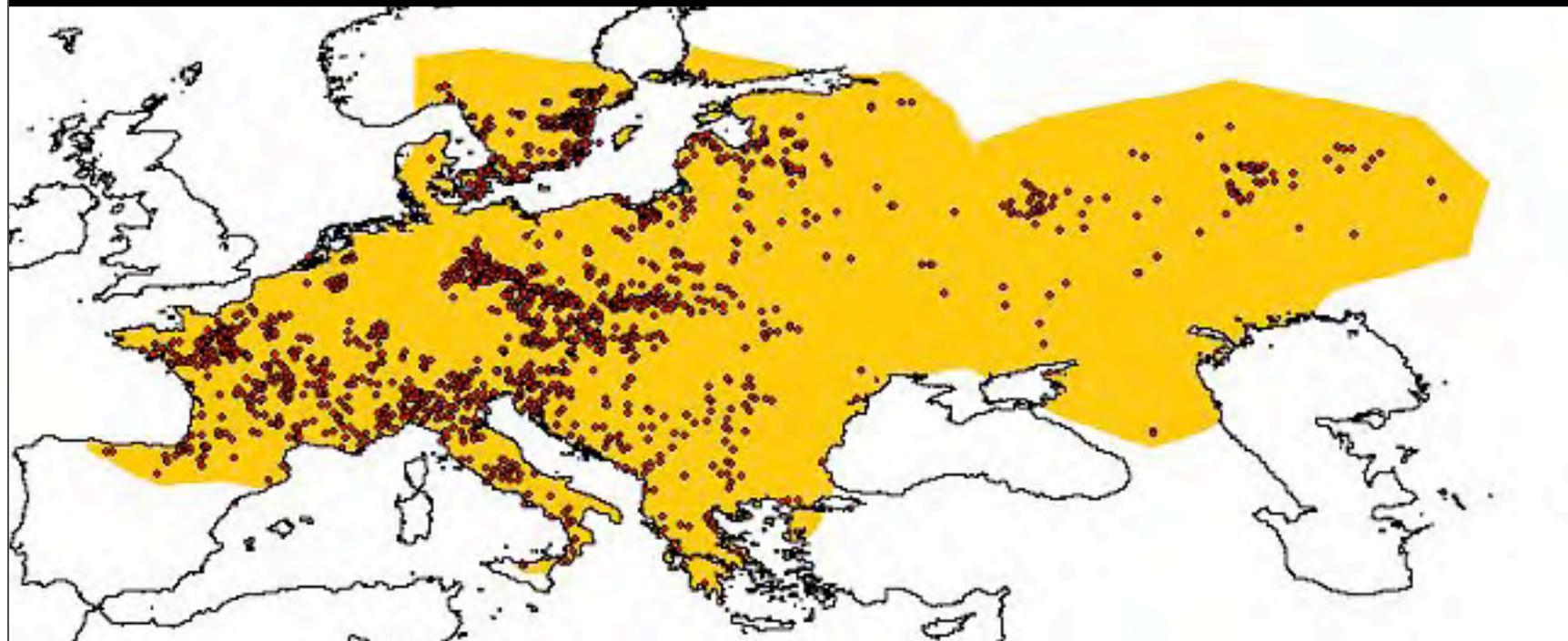
15

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



16

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



17

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



19

The most important initiative 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.



ROYAL TYRRELL
MUSEUM

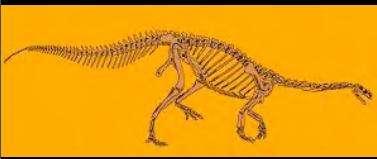
onature.ca
Canadian Museum of Nature



REAL JARDÍN BOTÁNICO
CONSEJO SUPERIOR DE INVESTIGACIONES CIENTÍFICAS



inatura
Erfahrung Naturschau Dornbirn



SAN DIEGO NATURAL HISTORY MUSEUM
Your Nature Connection

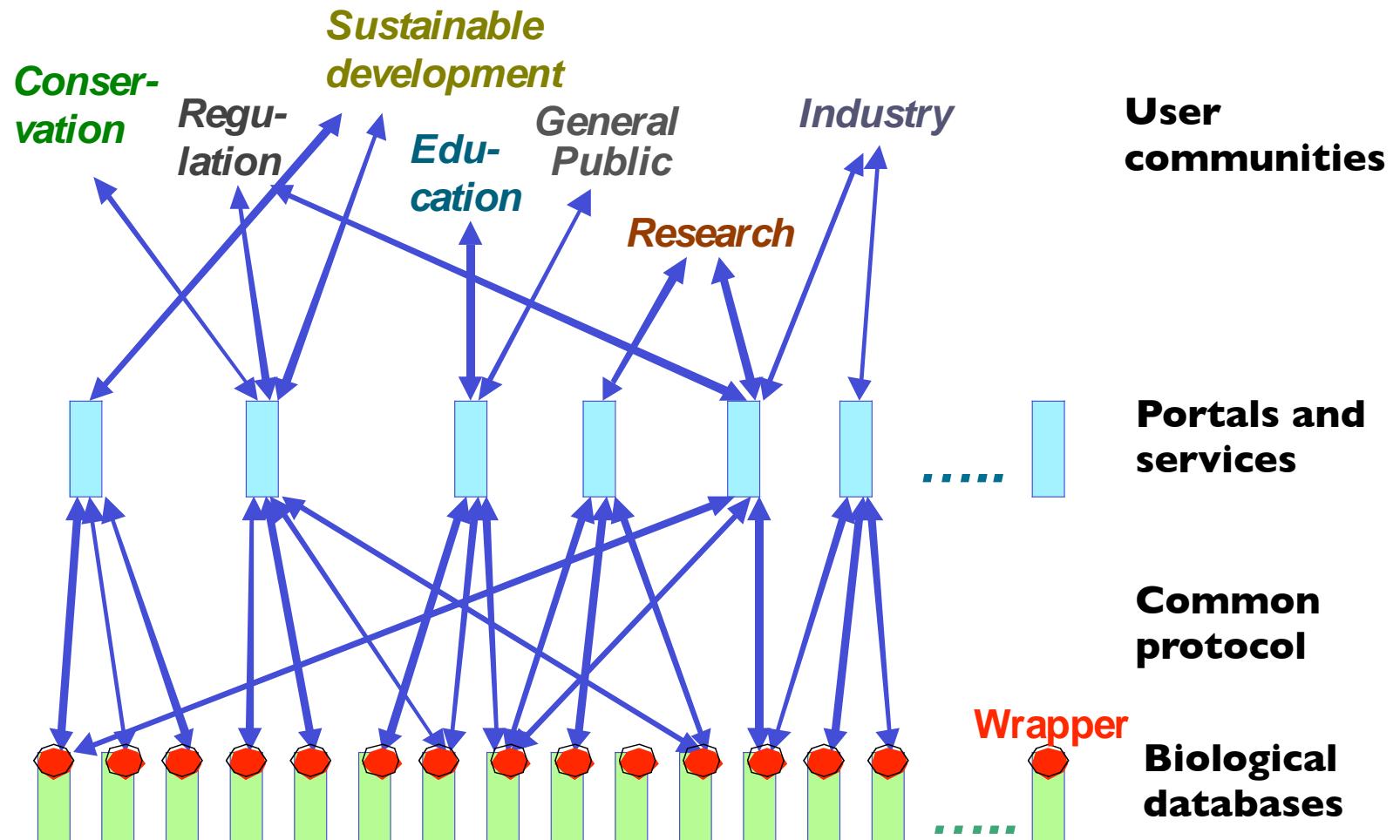


SIERRA
COLLEGE

The Raptor Center
College of Veterinary Medicine

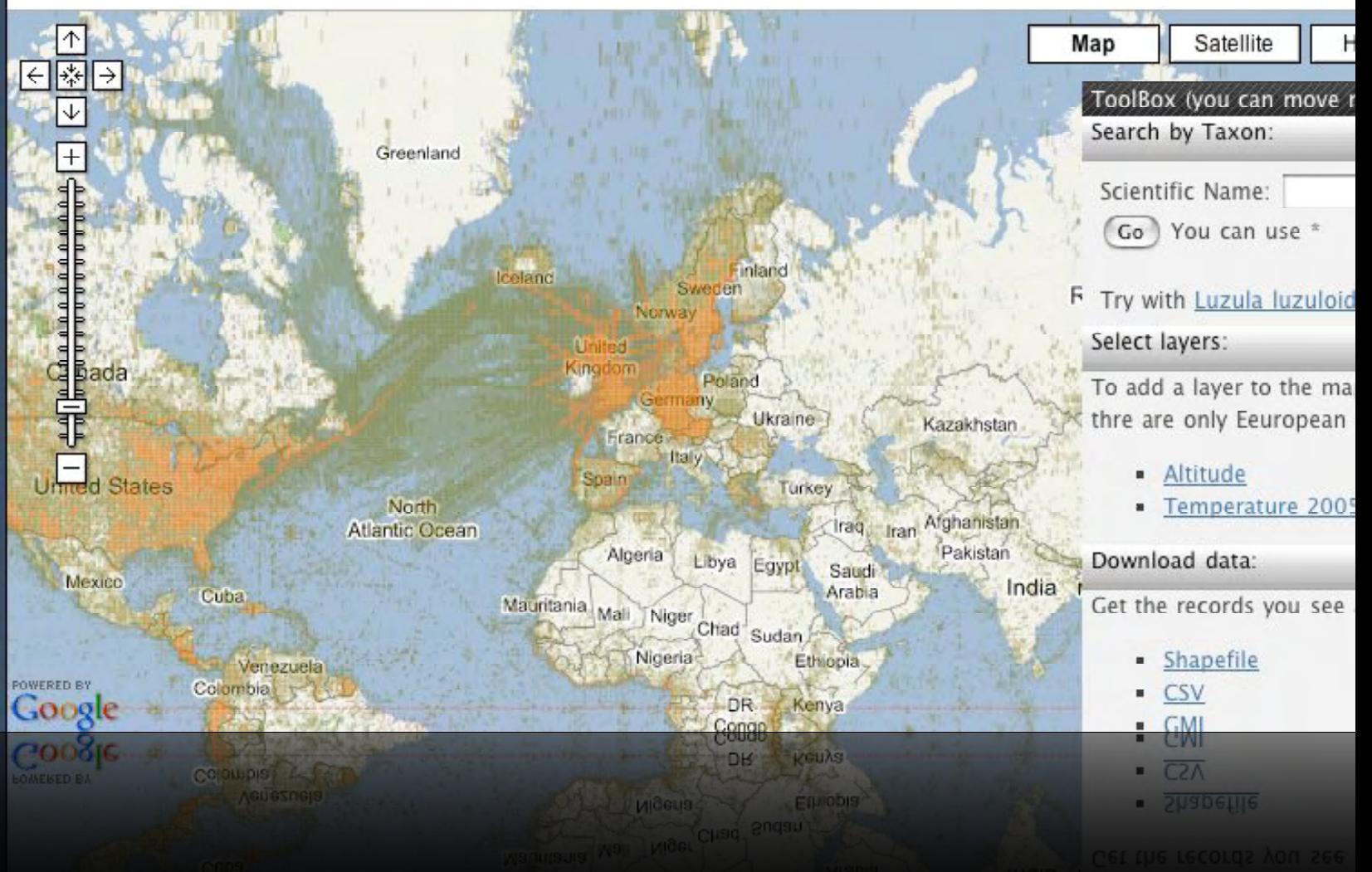
20

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 maintained by the source, the collections, and made available for portals like GBIF to query them... not getting much in detail here

Synthesys NA-D 3.6 demonstration prototype

[About](#)[Google Map](#)[MapBuilder](#)[Mapping service DPS](#)

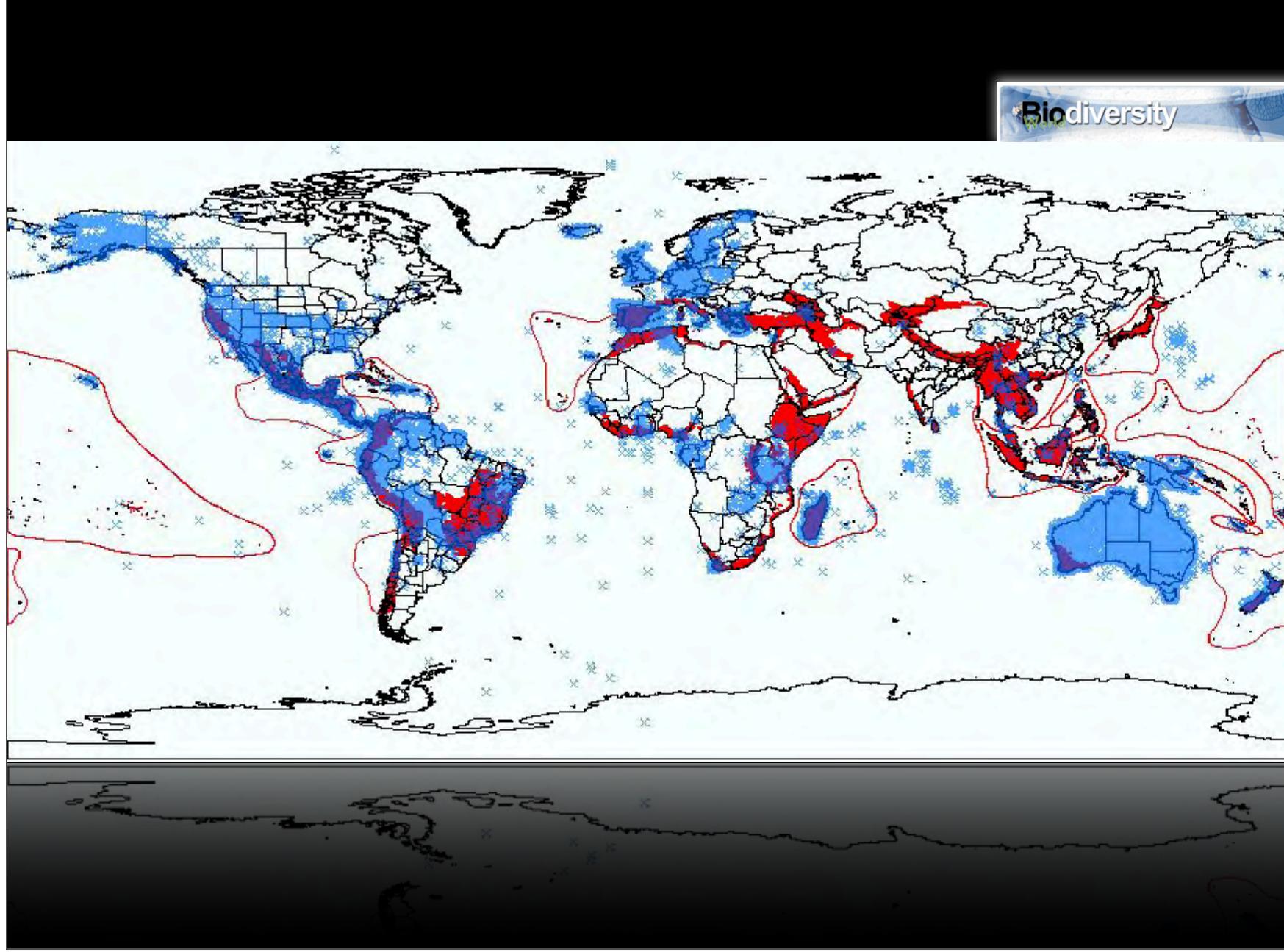
22

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.



23

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



24

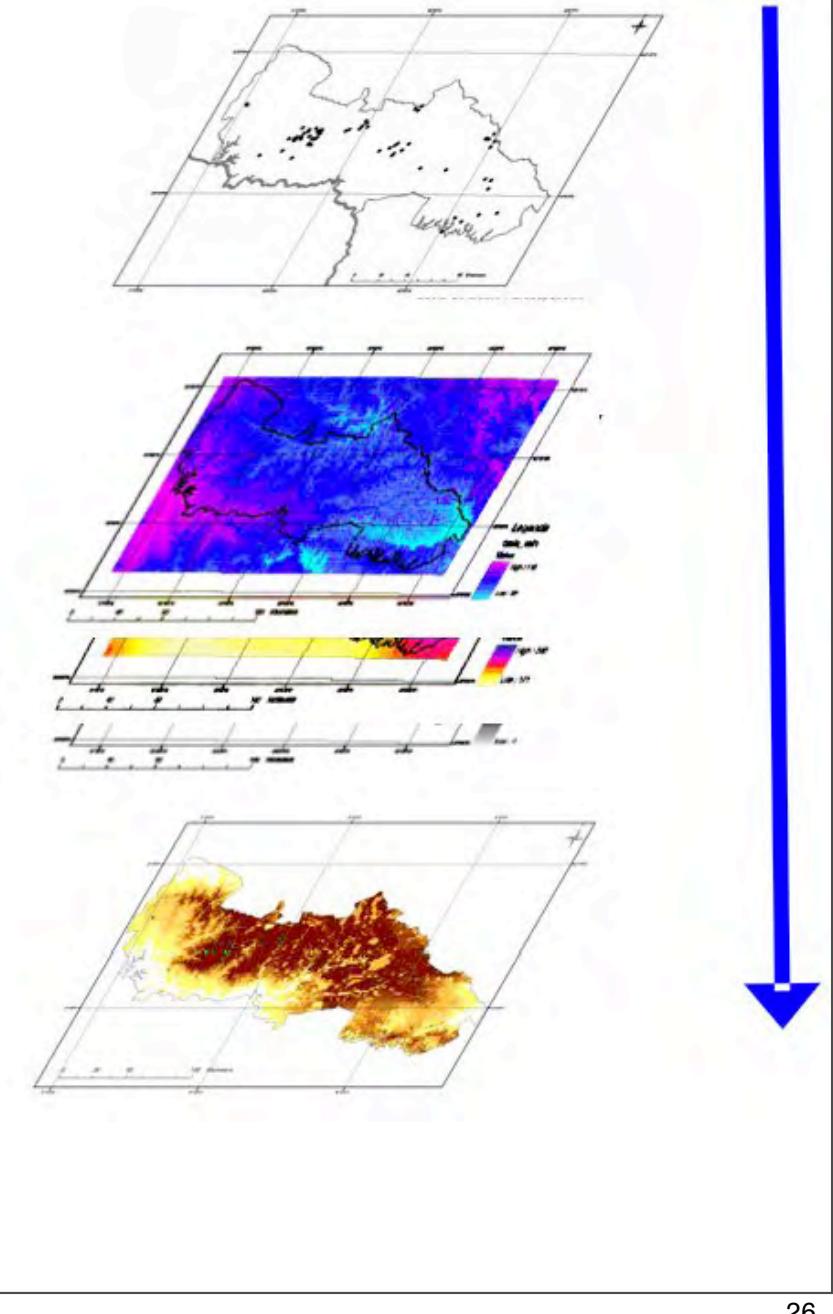
We can see that most data we have does not fall within them. Actually less than 10% of the data we have falls within a hotspot, and we know that at least 60% of world species occur there... so it is clear we don't have enough data actually to study.

We need solutions to lack of data

25

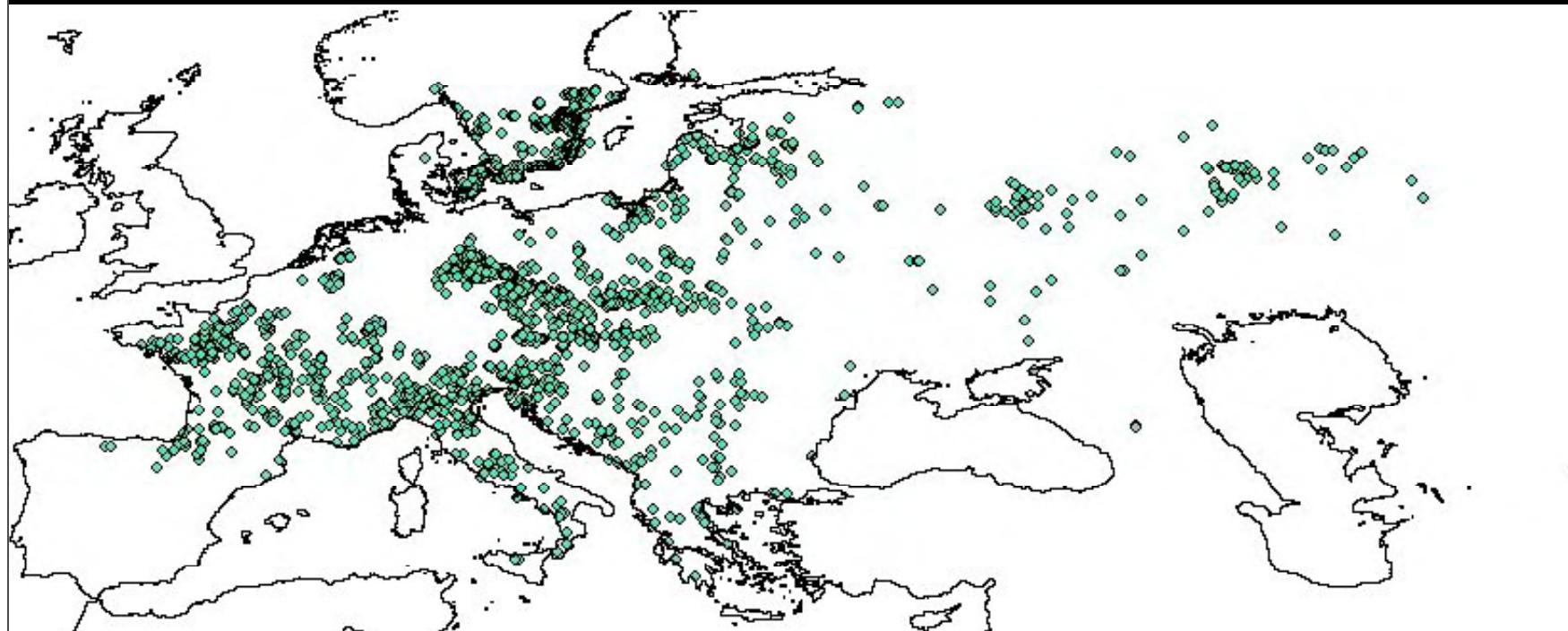
So we dont have much data about where species occur... how are we gonna protect biodiversity if we dont know where it happens? well, we have a problem and we need a solution.

Niche Modelling



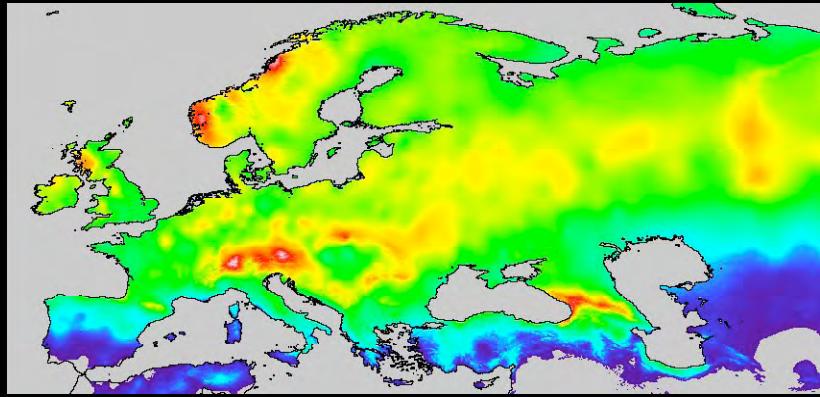
26

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

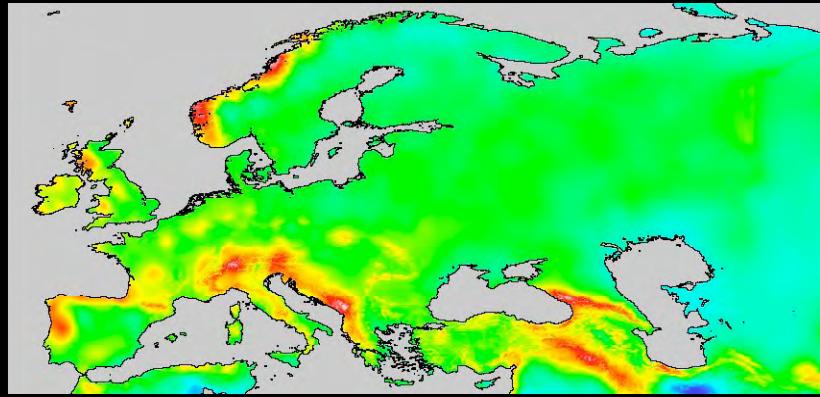


27

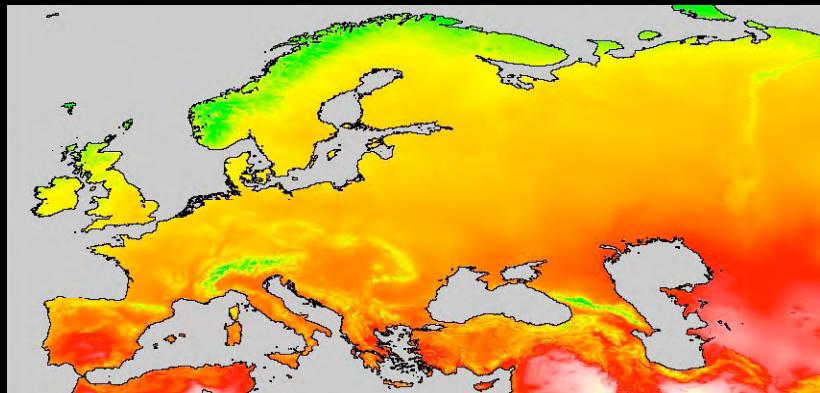
Using the incomplete data we already have...



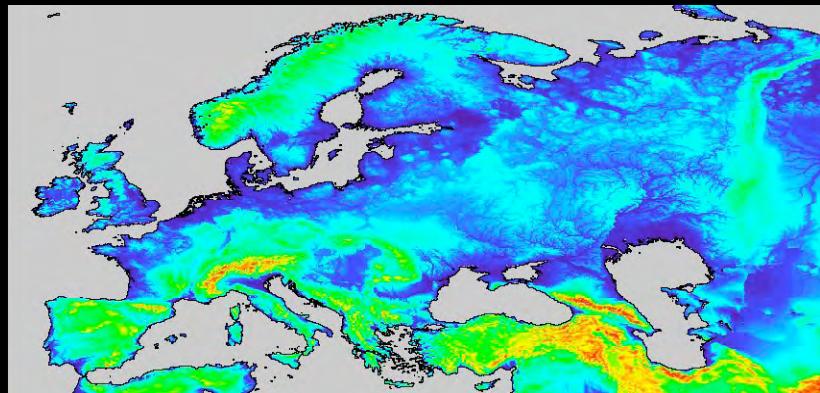
Summer precipitation



Winter precipitation

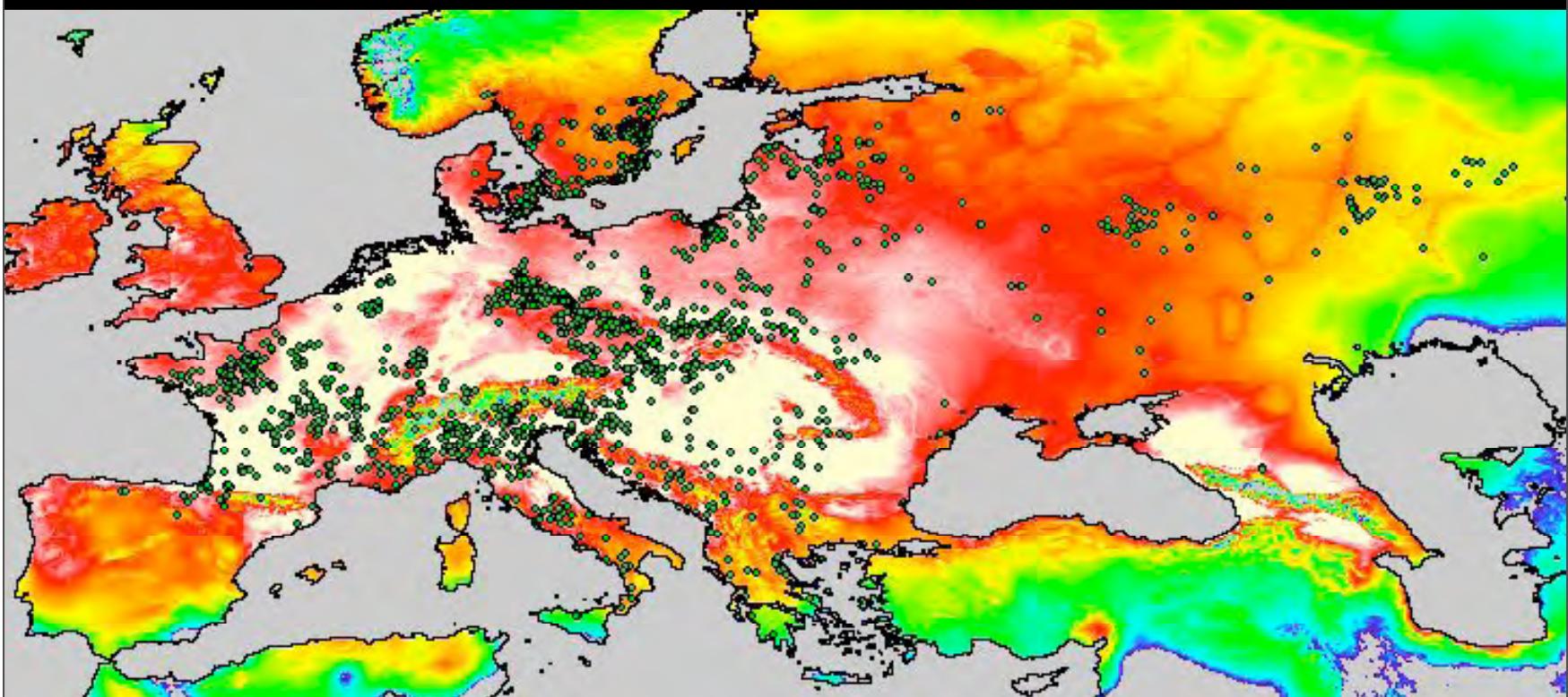


Maximum temp. summer



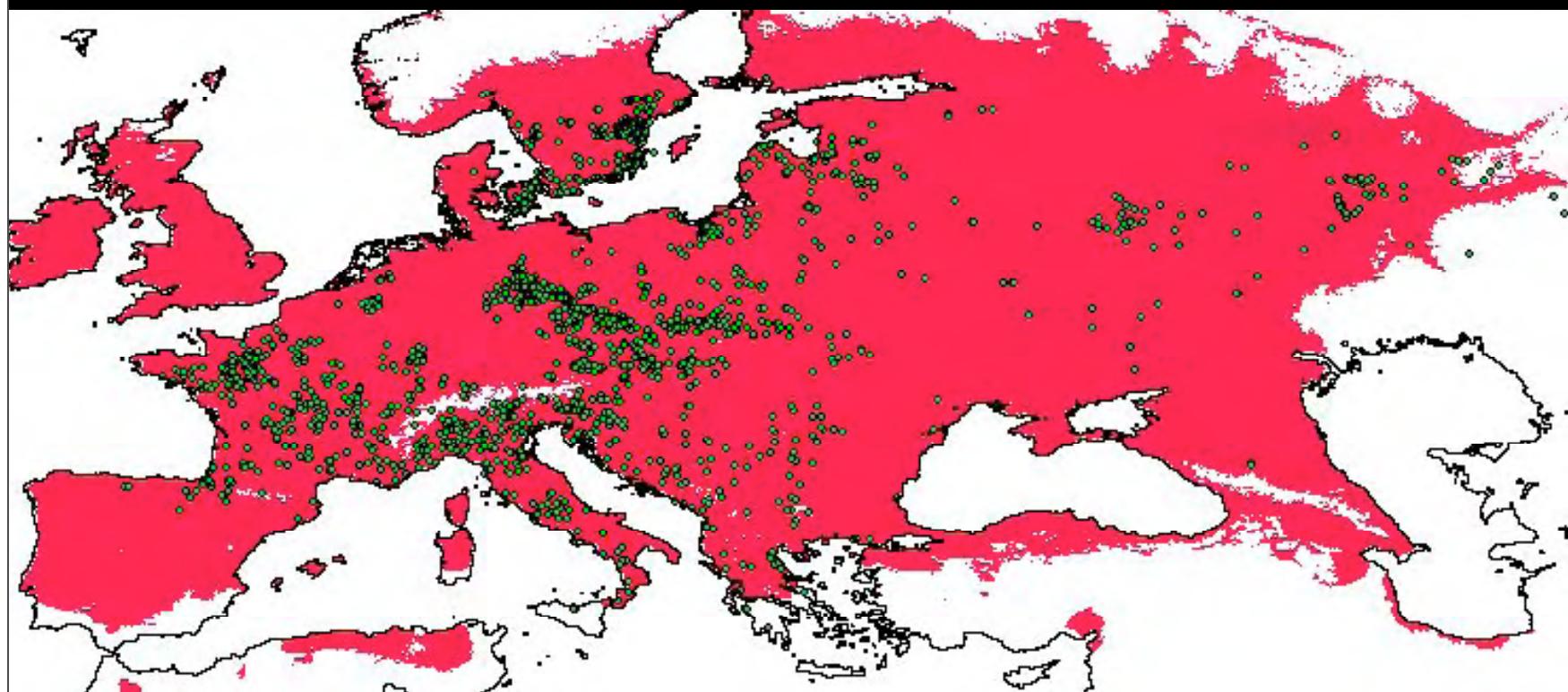
Average height

together with environmental data of the area...



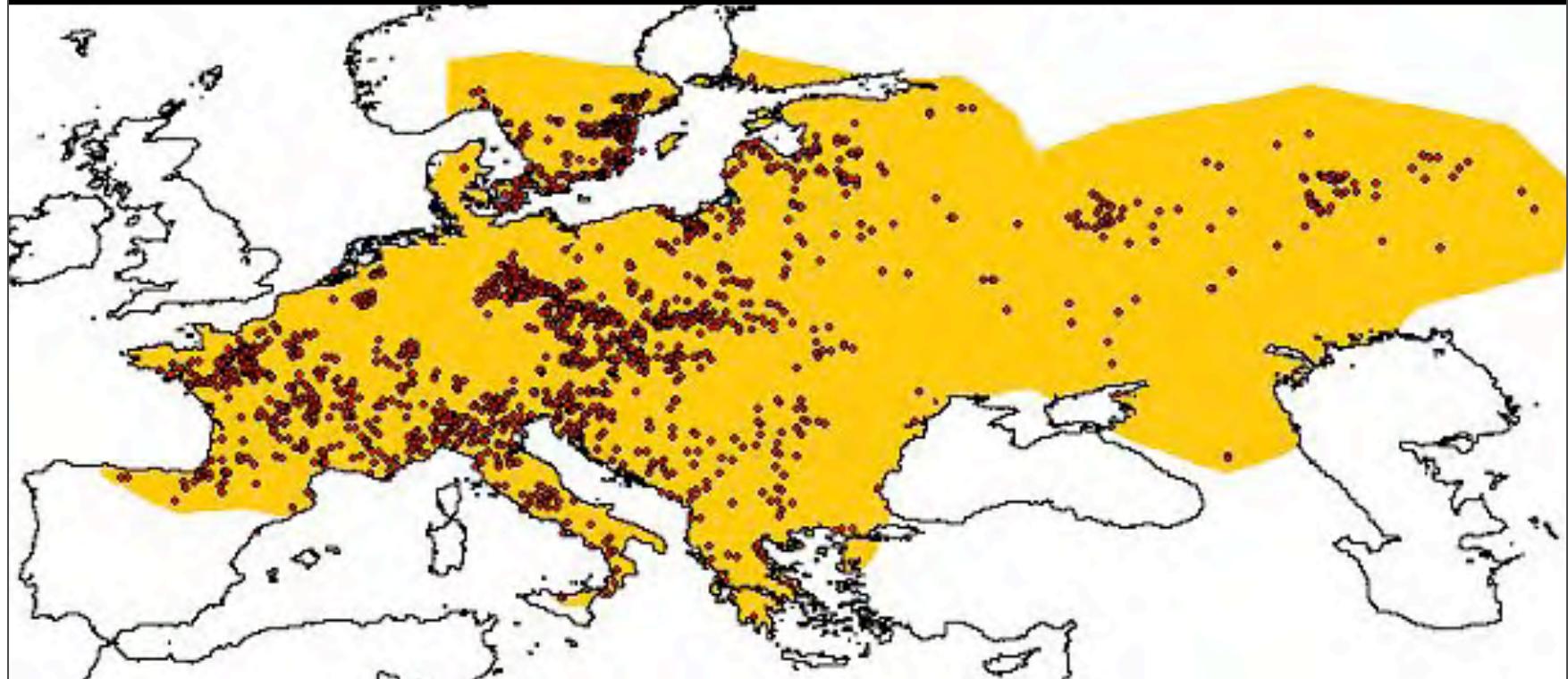
29

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



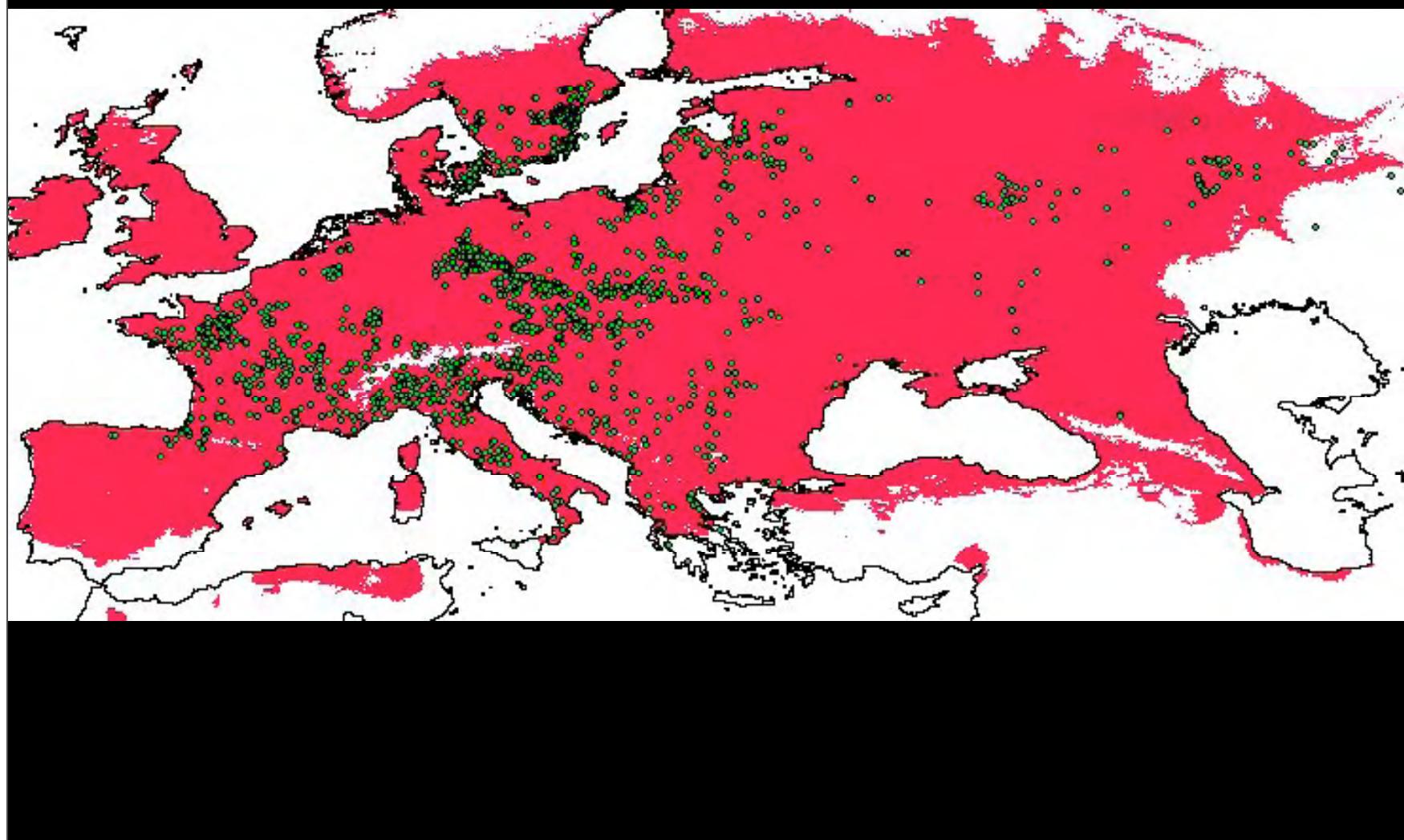
30

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



31

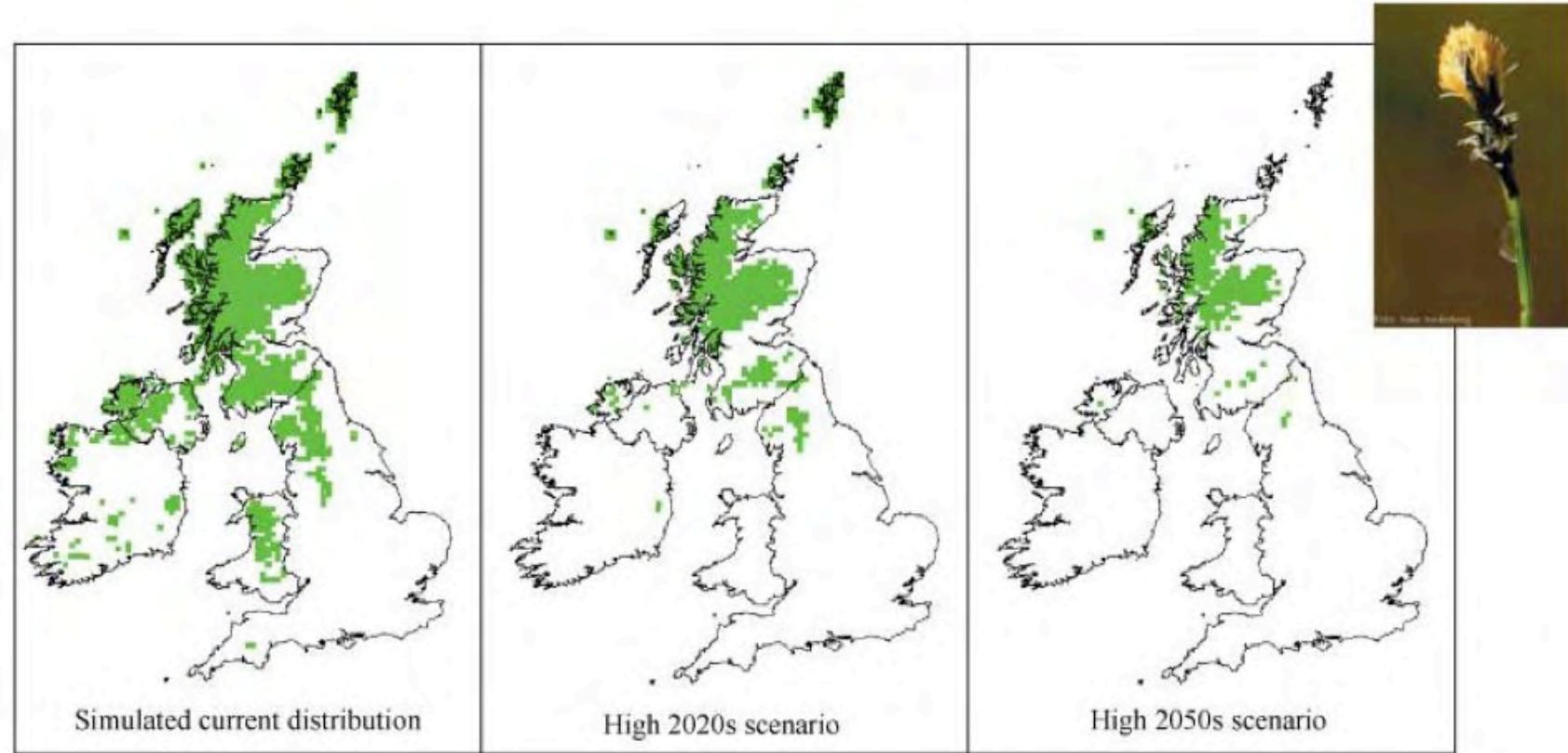
compared to the original
one



32

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 beetle. 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

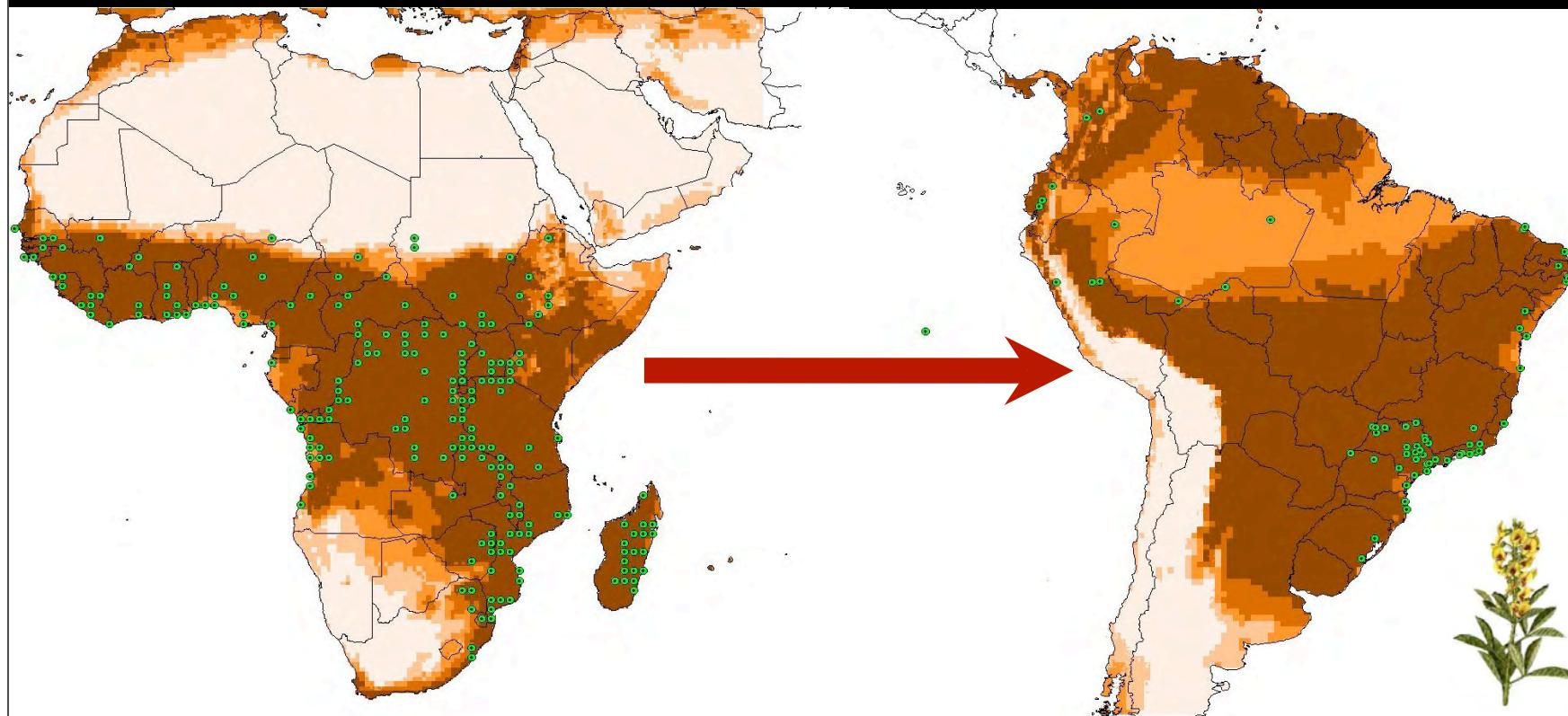


Pearson et al., 2002

33

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



34

Or for example to analyze how an invasive species is going to distribute in a new area...

Ocurrence data,
specimen-observation
data



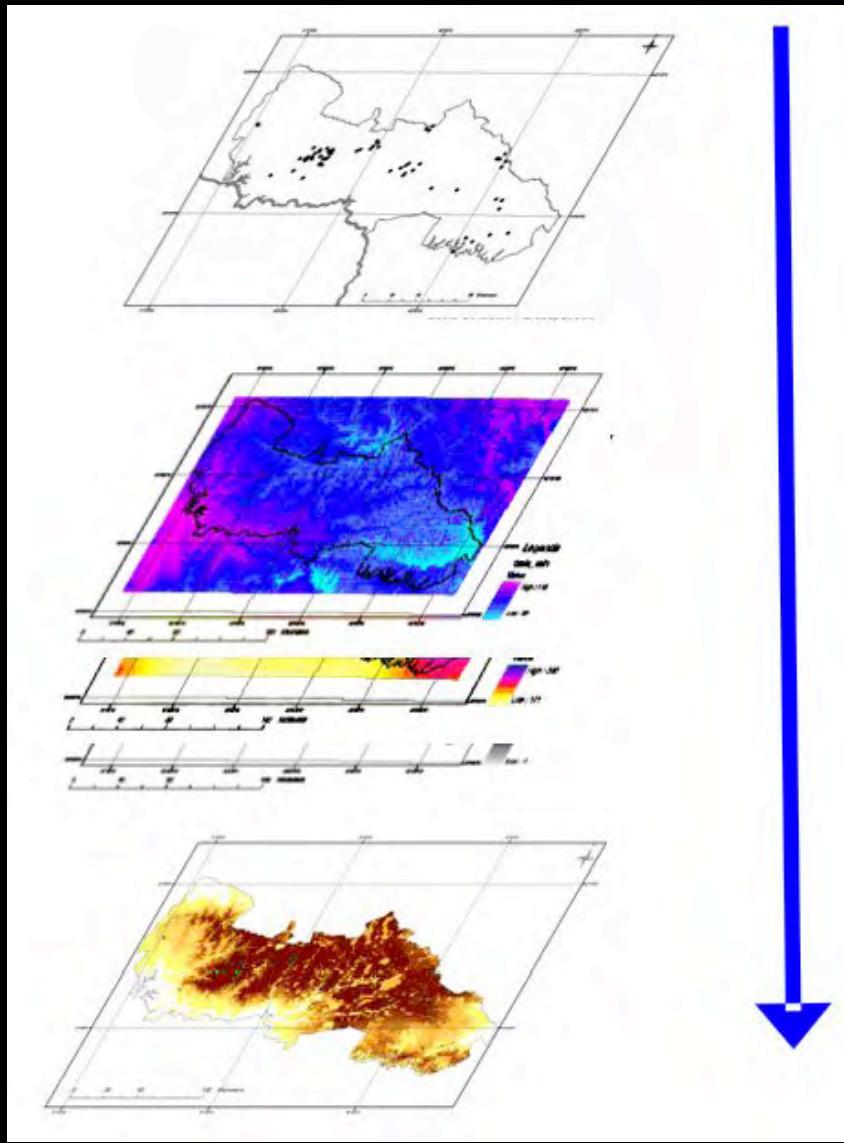
Environmental layers



Modelling algorithm



Distribution models



35

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

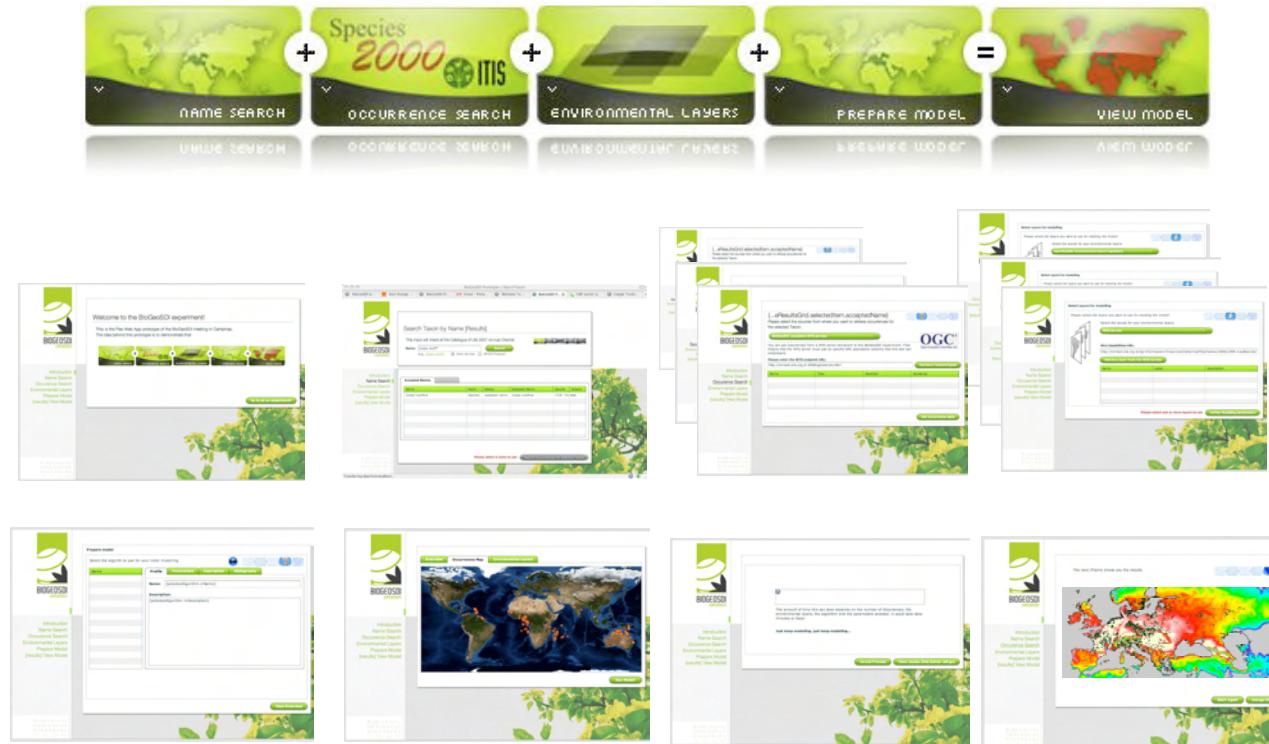


Biodiversity
Information
Standards
T D W S

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

36

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.



Biodiversity
Information
Standards
TDWG

I) Start

Welcome to the BioGeoSDI experiment!

This is the Flex Web App prototype of the BioGeoSDI meeting in Campinas. The idea behind this prototype is to demonstrate that:

Species 2000 + ITS + Occurrence search + Environmental layers + Prepare model = View model

Go to do an experiment!

Introduction
Name Search
Occurrence Search
Environmental Layers
Prepare Model
[results] View Model

PHP-HTML version

Flex (flash) version

Welcome to the BioGeoSDI Experiment

This is the HTML prototype of the BioGeoSDI meeting produced at the TIP meeting in Campinas in April 2007. You may prefer to use the [Flex](#) interface instead, however, you will requireFlash v9. The idea behind this prototype is to demonstrate that:

Species 2000 + ITS + Occurrence search + Environmental layers + Prepare model = View model

This prototype demonstrates interoperability between the following TDWG standards and other protocols widely used in the bioinformatics community:

- WMS
- WCS
- WFS
- GBIF REST
- SPICE
- TAPIB
- openModeller Web service
- DarwinCore
- Catalogue of Life Web service

Begin Experiment

Check session vars | Clean session vars

BioGeoSDI Code Fest in Campinas - April 2007

37

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



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W G

2) Search for a scientific Name



SPICE protocol, COL REST service

The screenshot displays the BioGeoSDI Prototype's SearchTaxon interface. At the top, there's a search bar with the placeholder "Search Taxon by Name [Results]" and a note about checking the Catalogue of Life 2007 Annual Checklist. Below the search bar, the "Name:" field contains "Cocos nucifera", and the "Search" button is visible. A radio button for "Web Service" is selected. To the right of the search bar, there are three small icons: a magnifying glass, a tree, and a globe. The main content area shows a table titled "Accepted Names". The table has columns for Name, Rank, Status, Accepted Name, Source, and Details. One row is shown: "Cocos nucifera", "Species", "accepted name", "Cocos nucifera", "ITIS: The www", and a link icon. At the bottom of the table, there's a message "Please select a name to use" and a "Search Occurrences for selected Taxon" button. On the left side of the interface, there's a sidebar with links: Introduction, Name Search, Occurrence Search, Environmental Layers, Prepare Model, and buttons for [results] and View Model. The footer of the interface says "Transferring data from localhost...". To the right of the interface, there's a decorative image of a tropical tree.

BioGeoSDI Code Fest in Campinas - April 2007

38

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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

3) Search for occurrences



WFS with our own GML app schema

BioGeoSDI Code Fest in Campinas - April 2007

39

After selecting the name what we want is to find occurrences of this species. Occurrences can be modeled 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 there is no official one now we created our own. Therefore the prototype can only talk with ourselves... not much of interoperability.



Biodiversity
Information
Standards
T D W G

3) Search for occurrences

GBIF REST service

The screenshot shows a web application interface. On the left, there's a sidebar with the BioGeoSDI logo at the top, followed by a list of links: Introduction, Name Search, Occurrence Search (which is highlighted in green), Environmental Layers, Prepare Model, [results], and View Model. Below this is another section with links: Biodiversity, Information, Standards, T D W G. The main content area has a header with placeholder text: "...eResultsGrid.selectedItem.acceptedName". It asks the user to select sources for retrieving occurrences for a selected taxon. A dropdown menu is open, showing "GBIF REST services" as the selected option. Below this, a message states: "The GBIF REST services provides you access to the entire GBIF cache compiled from all its data providers." To the right of this message is a small image of the Global Biodiversity Information Facility logo with the text "Get Occurrence data" below it. The background of the main content area features a photograph of green leaves.

BioGeoSDI Code Fest in Campinas - April 2007

40

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.



Biodiversity
Information
Standards
TDWG

3) Search for occurrences

TAPIR protocol providers

...eResultsGrid.selectedItem.acceptedName}
Please select the sources from where you want to retrieve occurrences for the selected Taxon:

TAPIR provider with Darwin Core 1.4 or ABCD 2.06 mapped

You can get occurrences from a TAPIR provider that has mapped his database to either Darwin Core 1.4 or ABCD 2.06. Please provide the TAPIR access point to the desire provider

Please enter the TAPIR accessPoint to the desire resource:

Get Occurrence data

BioGeoSDI Code Fest in Campinas - April 2007

41

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



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

3) Search for occurrences (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....

BioGeoSDI Code Fest in Campinas - April 2007

42

So, in reality behind the scene after retrieving the data from one of the services we were actually storing it temporarily 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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W G

4) Get environmental layers



WCS

BioGeoSDI Code Fest in Campinas - April 2007

43

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 experiment.



Biodiversity
Information
Standards
TDWG

4) Get environmental layers

Searching on a catalog Service (not implemented)

The screenshot shows a software interface for 'BIOGEOSDI EXPERIMENT'. On the left, there's a sidebar with links: Introduction, Name Search, Occurrence Search, Environmental Layers, Prepare Model [results], and View Model. The main area has a title 'Select Layers for modelling' with the instruction 'Please select the layers you want to use for creating the model'. It shows a dropdown menu set to 'OGC Catalog Service'. A note below says: 'Sorry, at this time the prototype is not able to retrieve layers from an OGC catalog service. When completed it will be possible to search in a catalog service to retrieve WCS services to use in modelling.' In the bottom right corner of the main window, there's a small image of a plant.

BioGeoSDI Code Fest in Campinas - April 2007

44

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



4) Get environmental layers

Using Modelling service local layers (OMWS getAvailableLayers)

BioGeoSDI Code Fest in Campinas - April 2007

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



Biodiversity
Information
Standards
T D W S

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

BioGeoSDI Code Fest in Campinas - April 2007

46

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



Biodiversity
Information
Standards
T D W G

5) Select Model algorithm & parameters



OMWS getAlgorithms

The screenshot shows the 'Prepare model' interface of the OMWS system. On the left, there's a sidebar with links: Introduction, Name Search, Occurrence Search, Environmental Layers, Prepare Model [results], and View Model. The main area is titled 'Prepare model' and says 'Select the algoritm to use for your niche modelling'. It has tabs for Profile, Parameters, Description, and Bibliography. The Parameters tab is selected. There are input fields for 'Name' and 'Description', both containing placeholder text. At the bottom right of the main area is a 'View Overview' button.

BioGeoSDI Code Fest in Campinas - April 2007

47

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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

6) Overview before modelling

The figure displays three screenshots of the BioGeoSDI Experiment software interface. The top-left screenshot shows the 'Overview' section with a search bar for 'Species 2000' and a link to the 'Catalogue of Life: 2007 Annual Checklist'. The bottom-left screenshot shows the 'Environmental Layers' section with a 'LayerId' input field and a preview image of a world map with environmental data. The right side of the figure shows a large 'Occurrences Map' displaying geographical distribution points across the world.

BioGeoSDI Code Fest in Campinas - April 2007

48

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



Biodiversity
Information
Standards
T D W G

7) Run model



OMWS runModel, getLog (SOAP service)

BioGeoSDI Code Fest in Campinas - April 2007

49

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 asynchronous mechanism... modelling can take days to be performed!

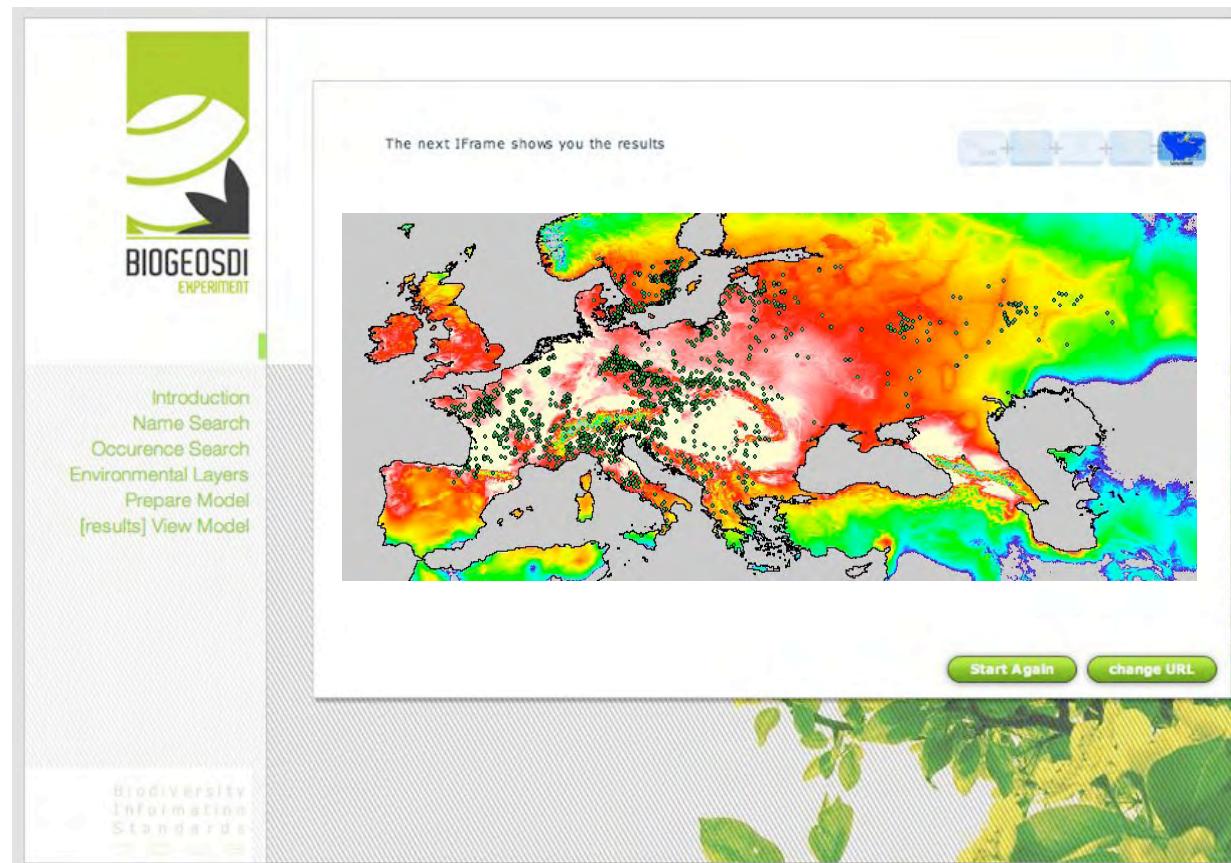


Biodiversity
Information
Standards
T D W G

8) View Result

WMS,WFS

OpenLayers



BioGeoSDI Code Fest in Campinas - April 2007

50

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 beginning.



Technologies used

- Geoserver (WMS,WFS, KML)
- PostGIS
- Mapserver (WCS)
- OpenModeller (OMWS)
- PHP, Flex, HTML, Bash, Python

BioGeoSDI Code Fest in Campinas - April 2007

51

So this are the technologies we used...



Biodiversity
Information
Standards
T D W G

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

BioGeoSDI Code Fest in Campinas - April 2007

52

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



Biodiversity
Information
Standards
T D W G



Thanks!

Workshop Wiki

[http://wiki.tdwg.org/twiki/bin/view/Geospatial/
InteroperabilityWorkshop1](http://wiki.tdwg.org/twiki/bin/view/Geospatial/InteroperabilityWorkshop1)

Prototype

<http://omtest.cria.org.br/biogeosdi/frontend/>

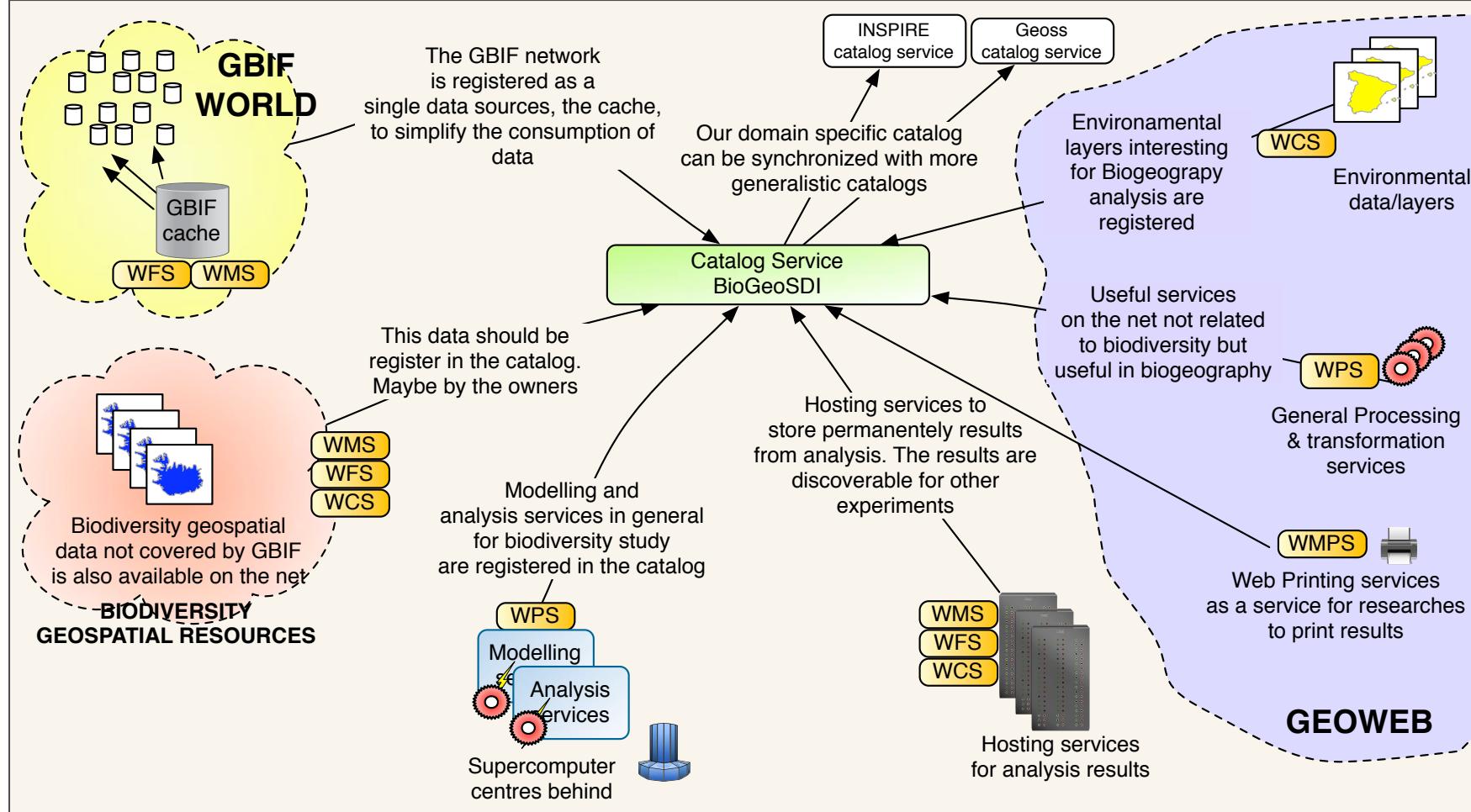
Source Code and documentation

<http://biogeosdi.org/>

Report

[http://omtest.cria.org.br/biogeosdi/doc/report/
BioGeoSDIreport.html](http://omtest.cria.org.br/biogeosdi/doc/report/BioGeoSDIreport.html)

BioGeoSDI Code Fest in Campinas - April 2007



Biogeography Spatial Data Infrastructure BioGeo Spatial Data Infrastructure

BioGeoSDI Code Fest in Campinas - April 2007

54

This not to be seen...