



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

BIOGEOSDI

workshop

Using TDWG and OGC standards
together

Javier de la Torre
jatorre@imaste-ips.com

IMASTE^{IPS}

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



BIOGEOSDI
EXPERIMENT
Biodiversity
Information
Standards
T D W G

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-ips.com)



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

BioGeoSDI Code Fest in Campinas - April 2007

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



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

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

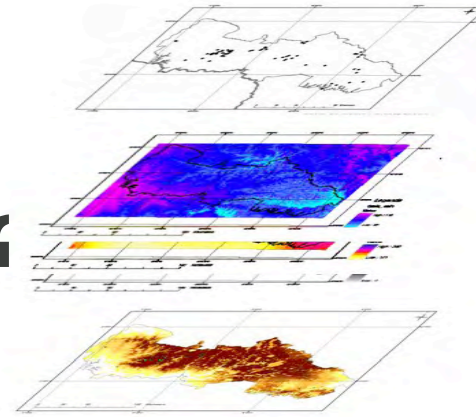
Biodiversity
Information
Standards
T D W G

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.



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 shelf you can find two beetles 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

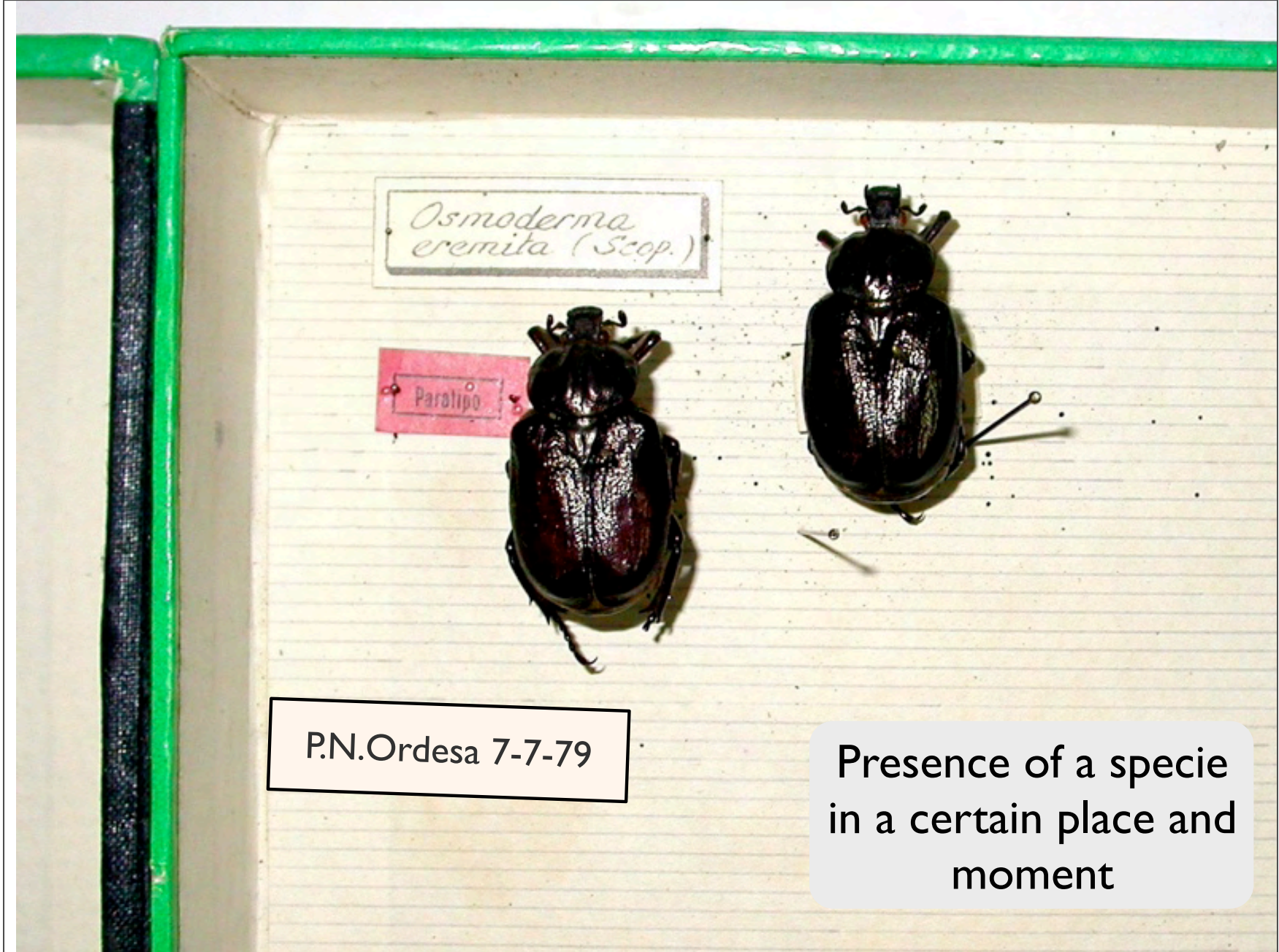


some biologist was
collecting ...



Osmoderma eremita

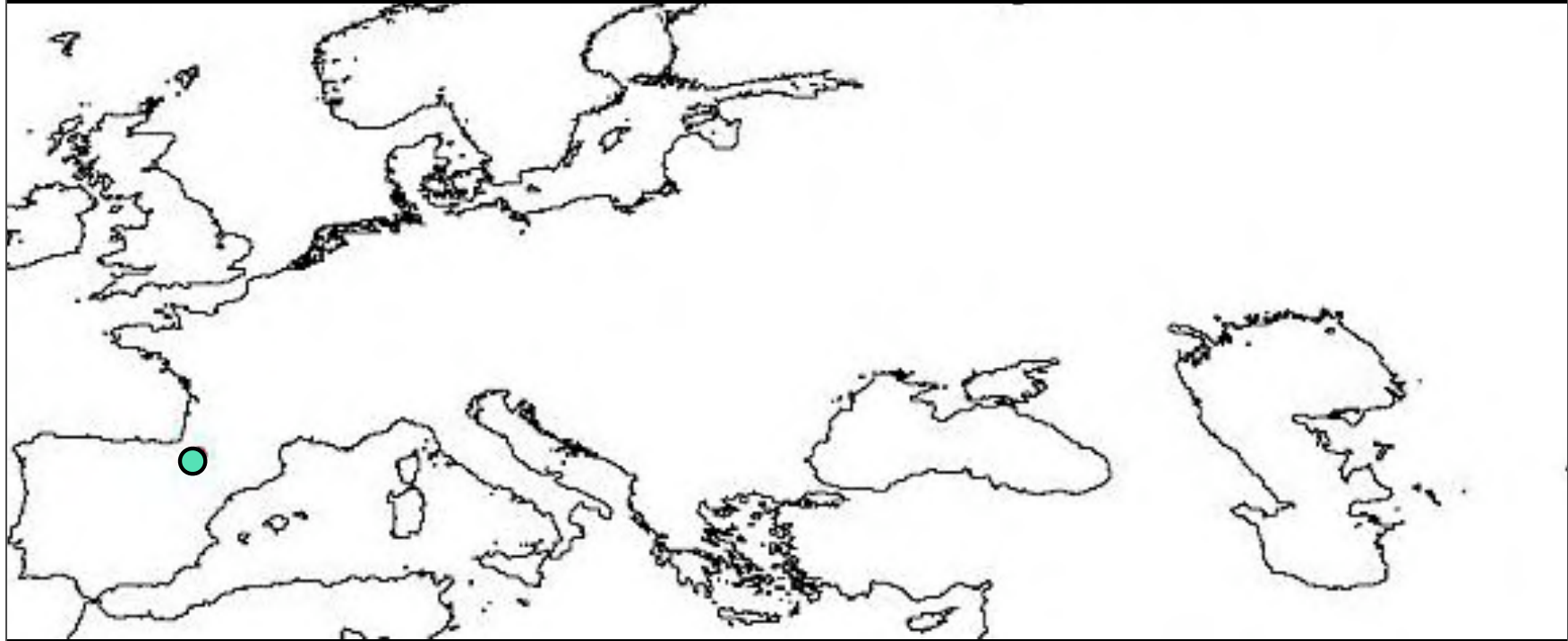
and found this beetle. Put it in a bottle and took it to the museum in madrid



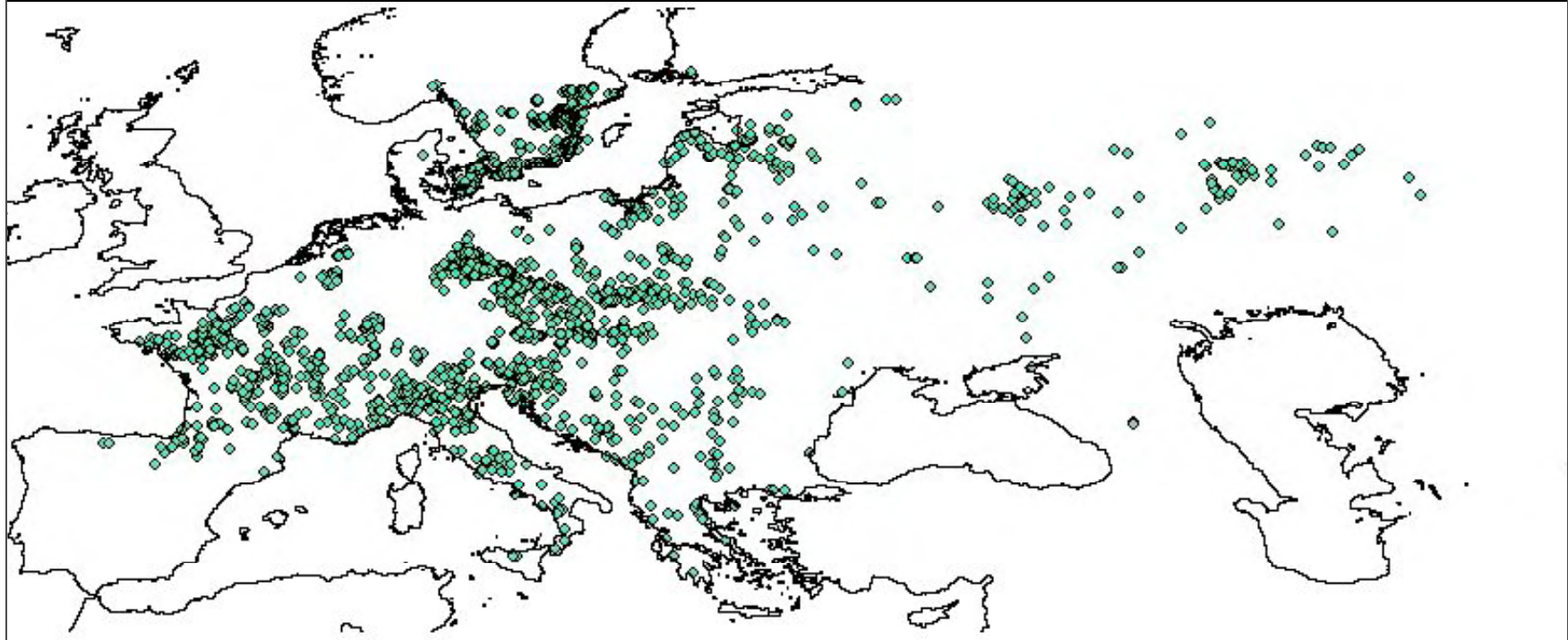
P.N.Ordesa 7-7-79

Presence of a specie
in a certain place and
moment

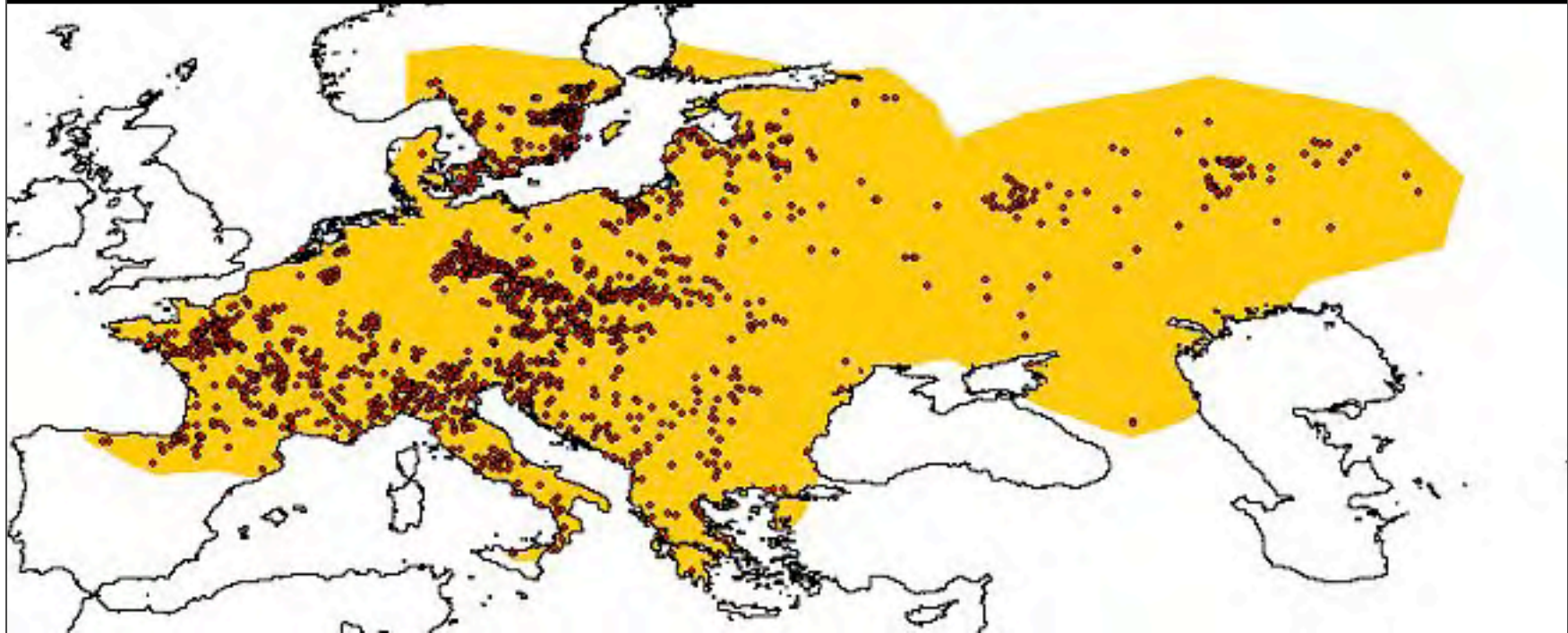
So this specimen at the museum is indicating that this specific specie was found in a certain place and moment.



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

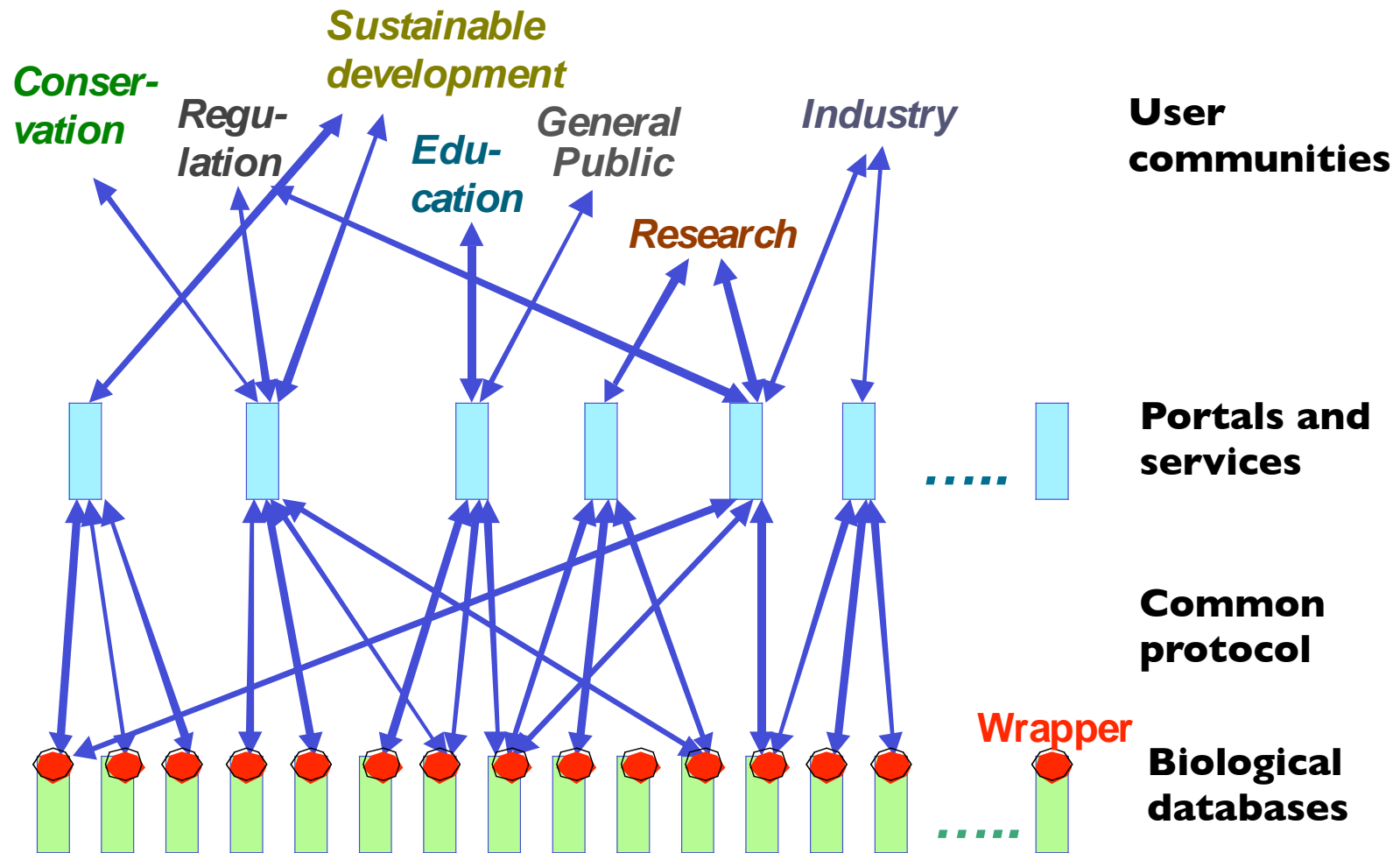
But to so we need lots of data



The most important 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.



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

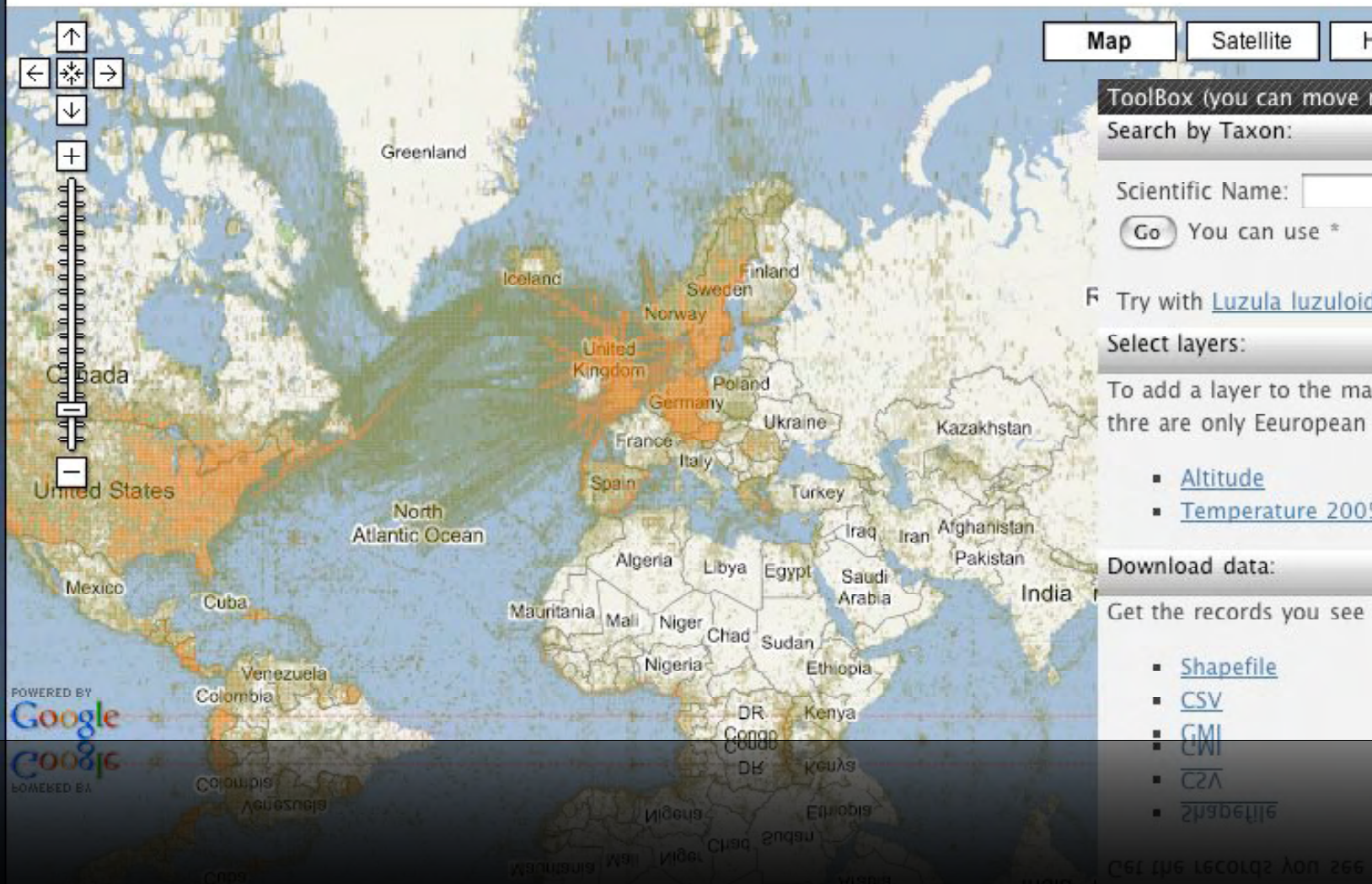
Synthesys NA-D 3.6 demonstration prototype

About

Google Map

MapBuilder

Mapping service DPS

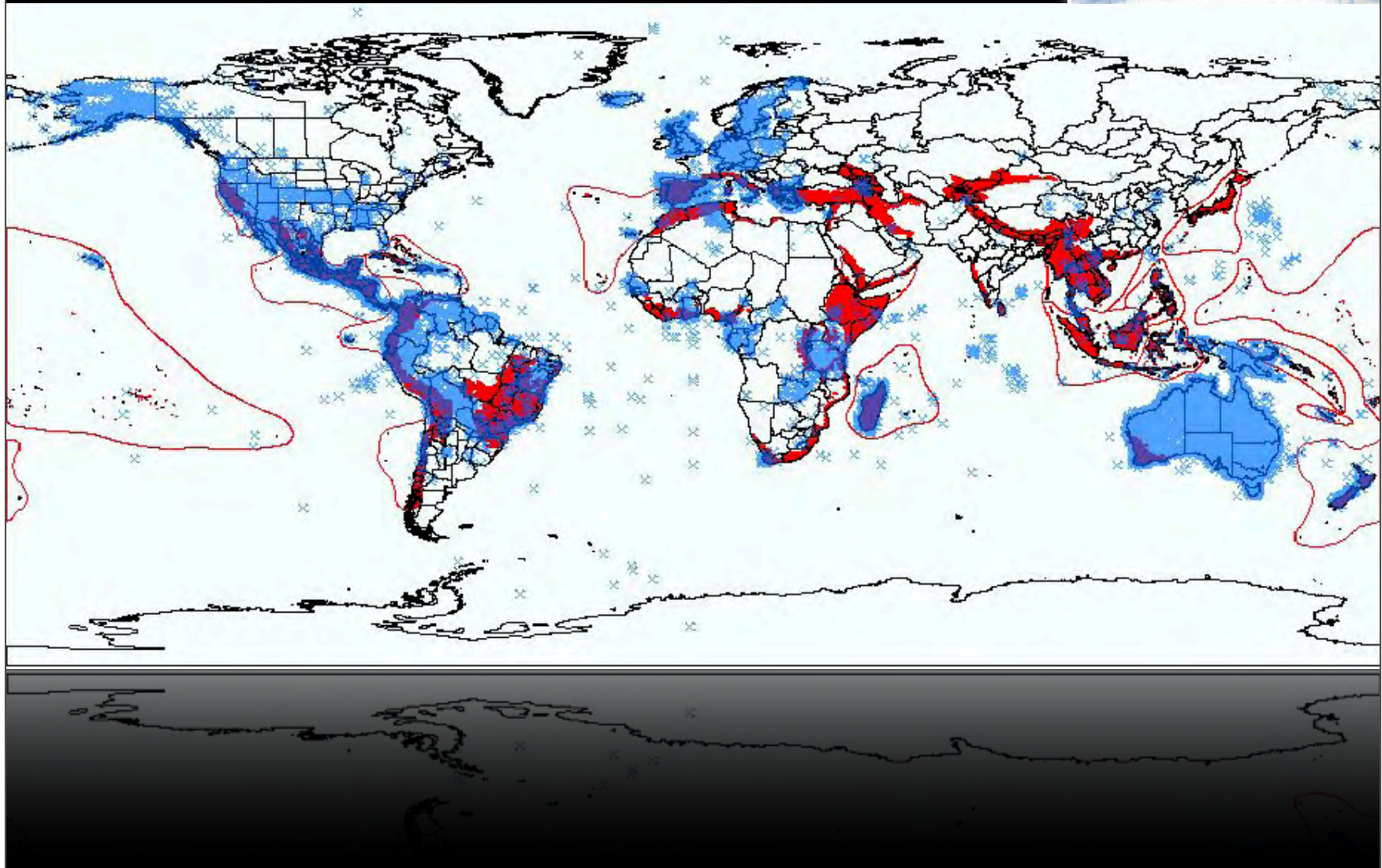


22

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.



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



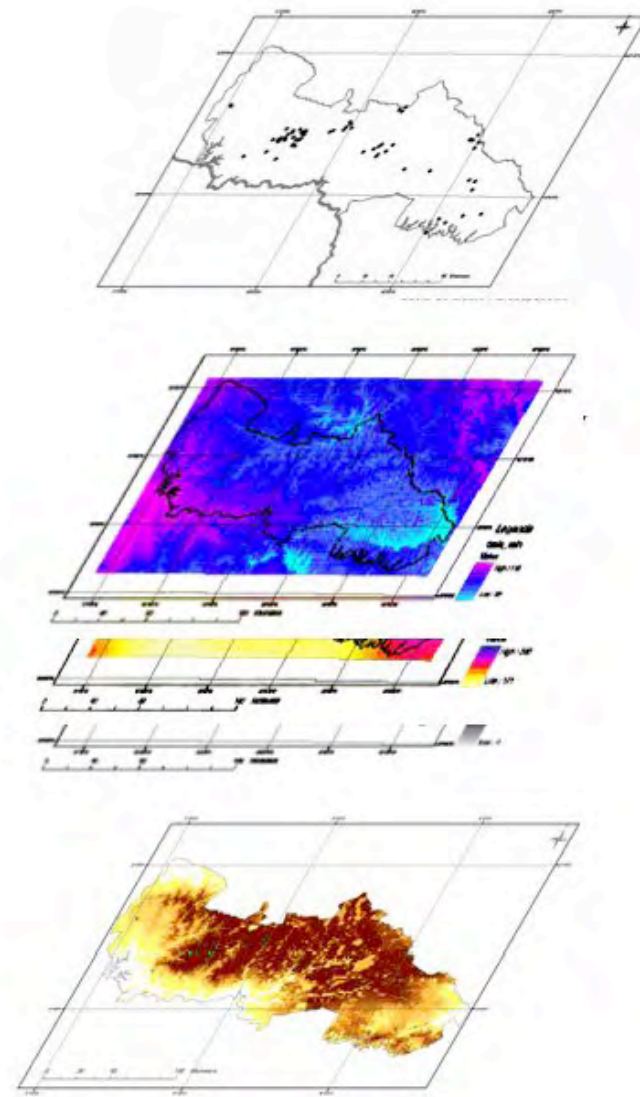
We can see that most data we have does not fall within them. Actually less than 10% of the data we have falls within a hotspot, and we know that at least 60% of world species occur there... so it is clear we don't have enough data actually to study.

We need solutions to lack of data

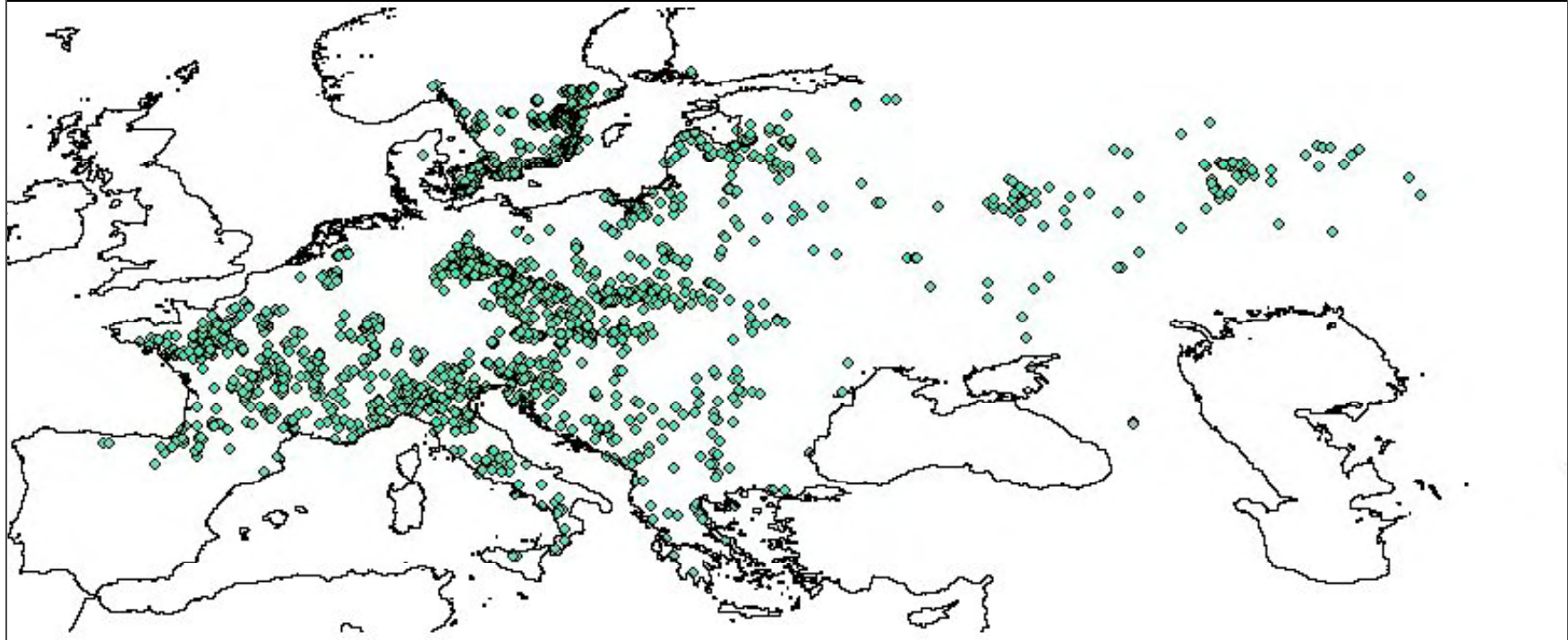
25

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.

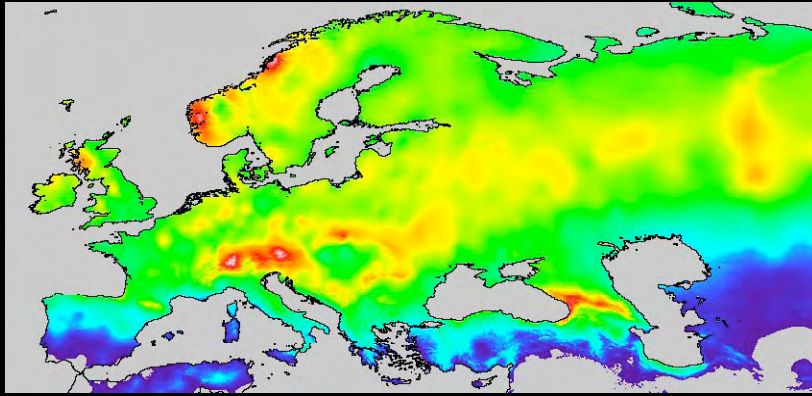
Niche Modelling



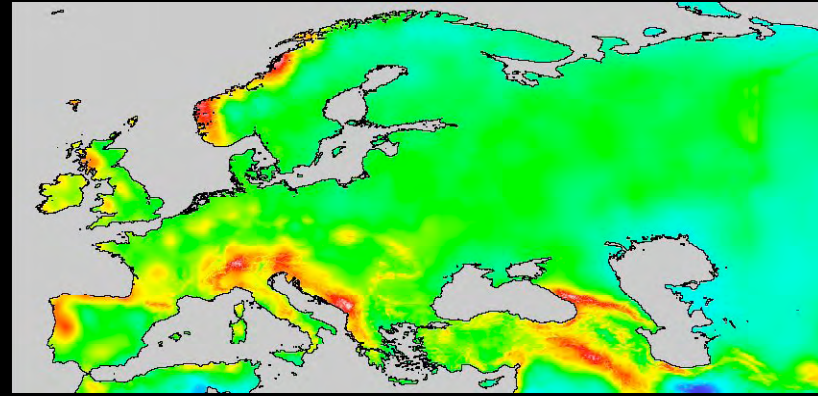
Niche modelling is one possible short term solution to the lack of data... lets take a look to it.



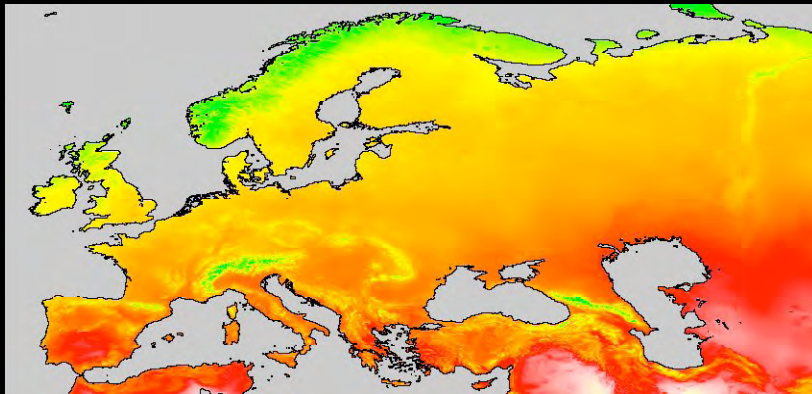
Using the incomplete data we already have...



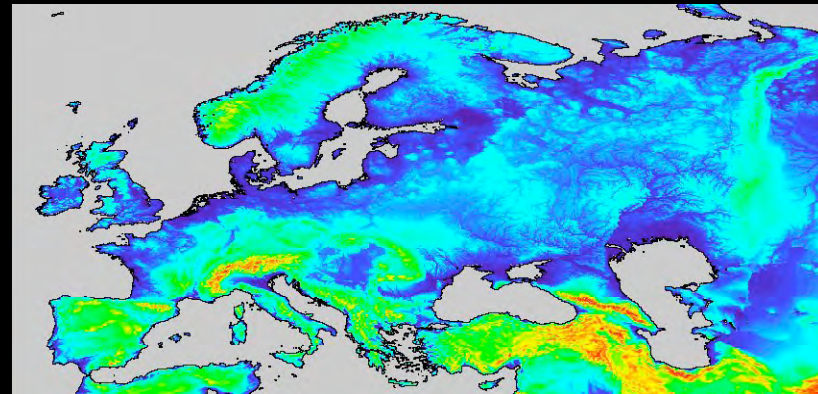
Summer precipitation



Winter precipitation

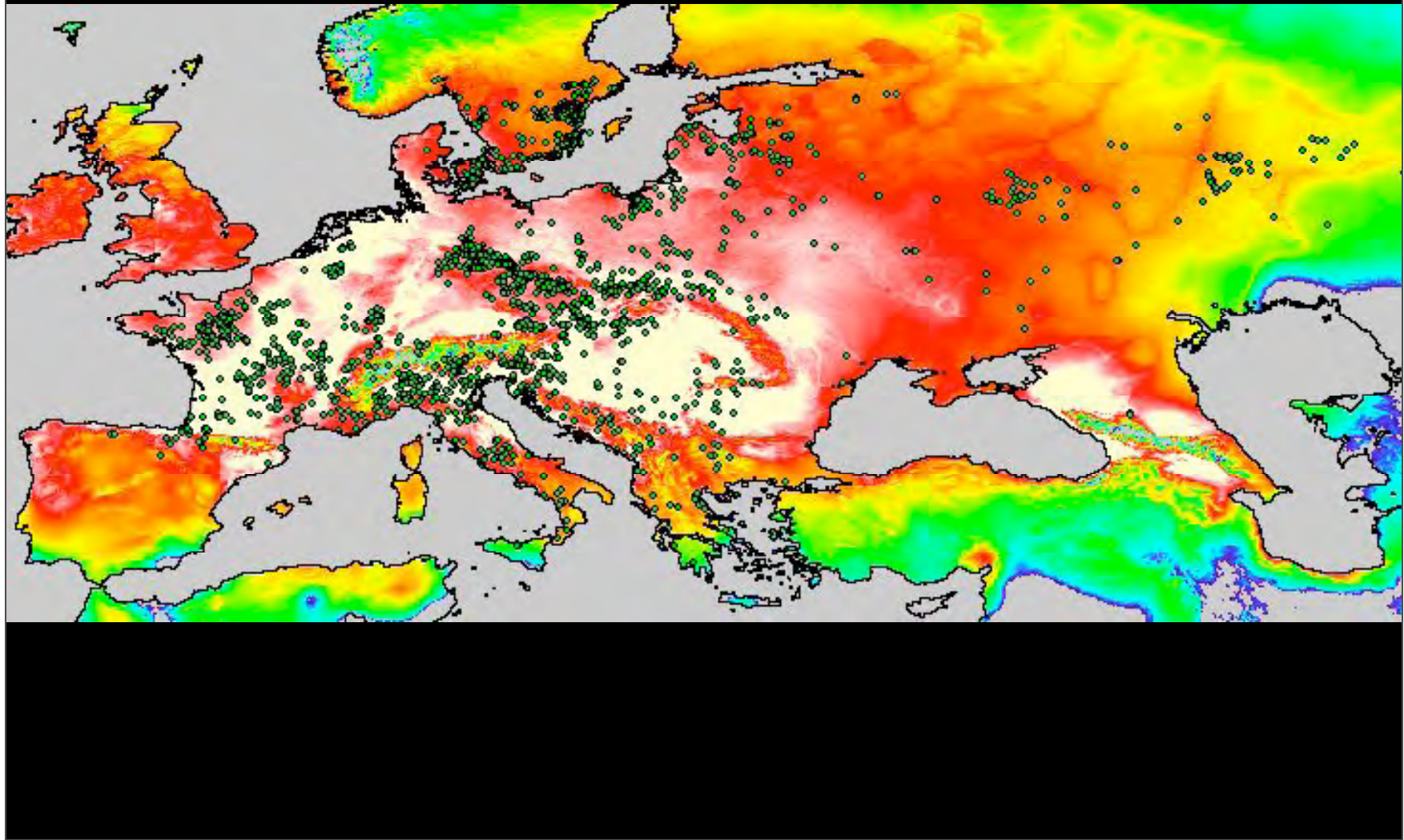


Maximum temp. summer

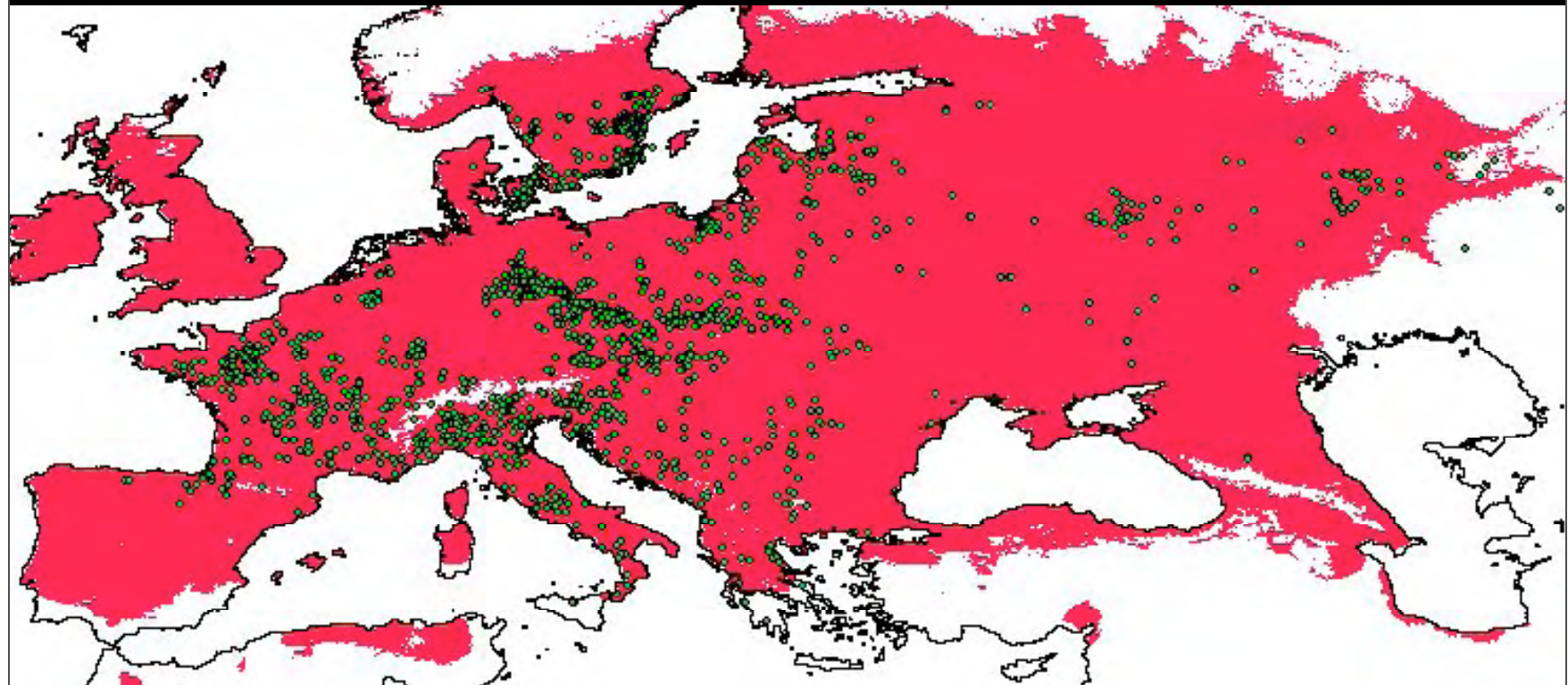


Average height

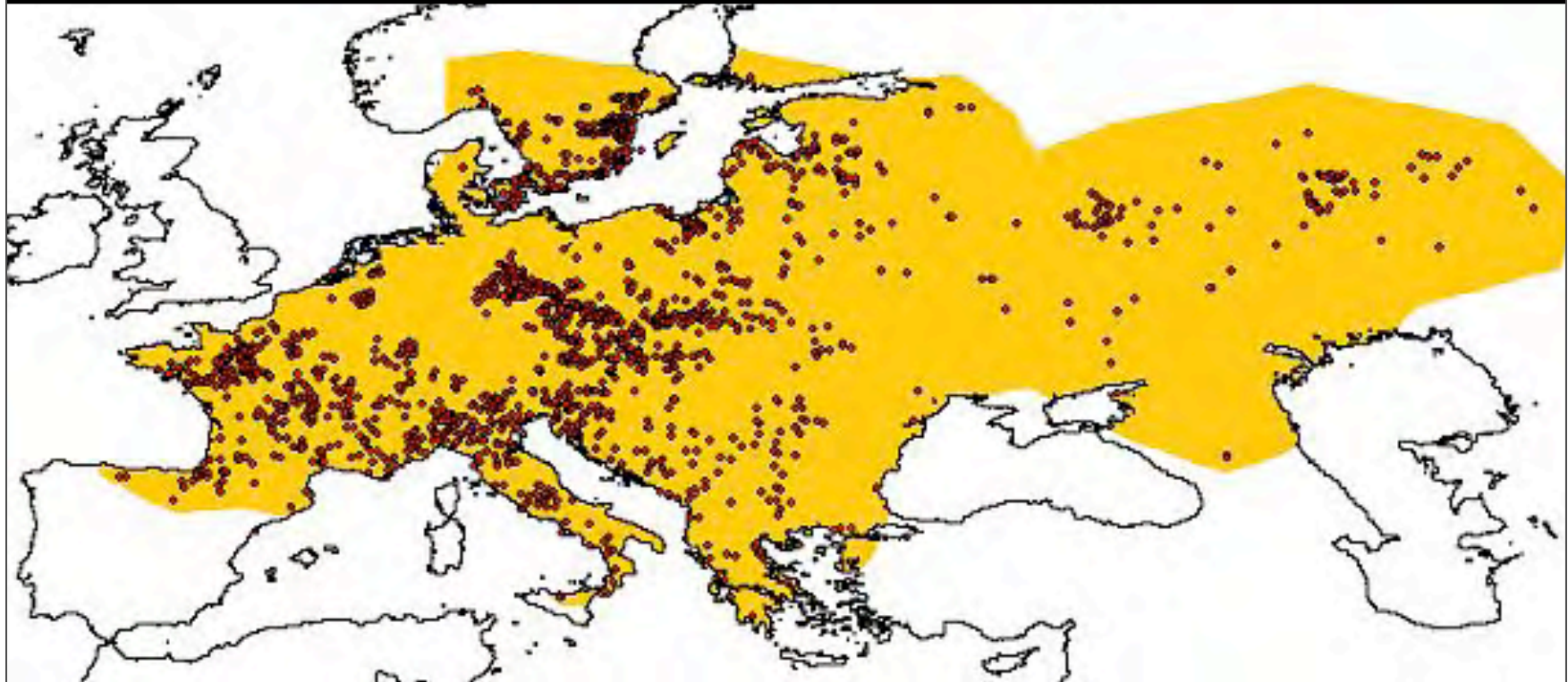
together with environmental data of the area...



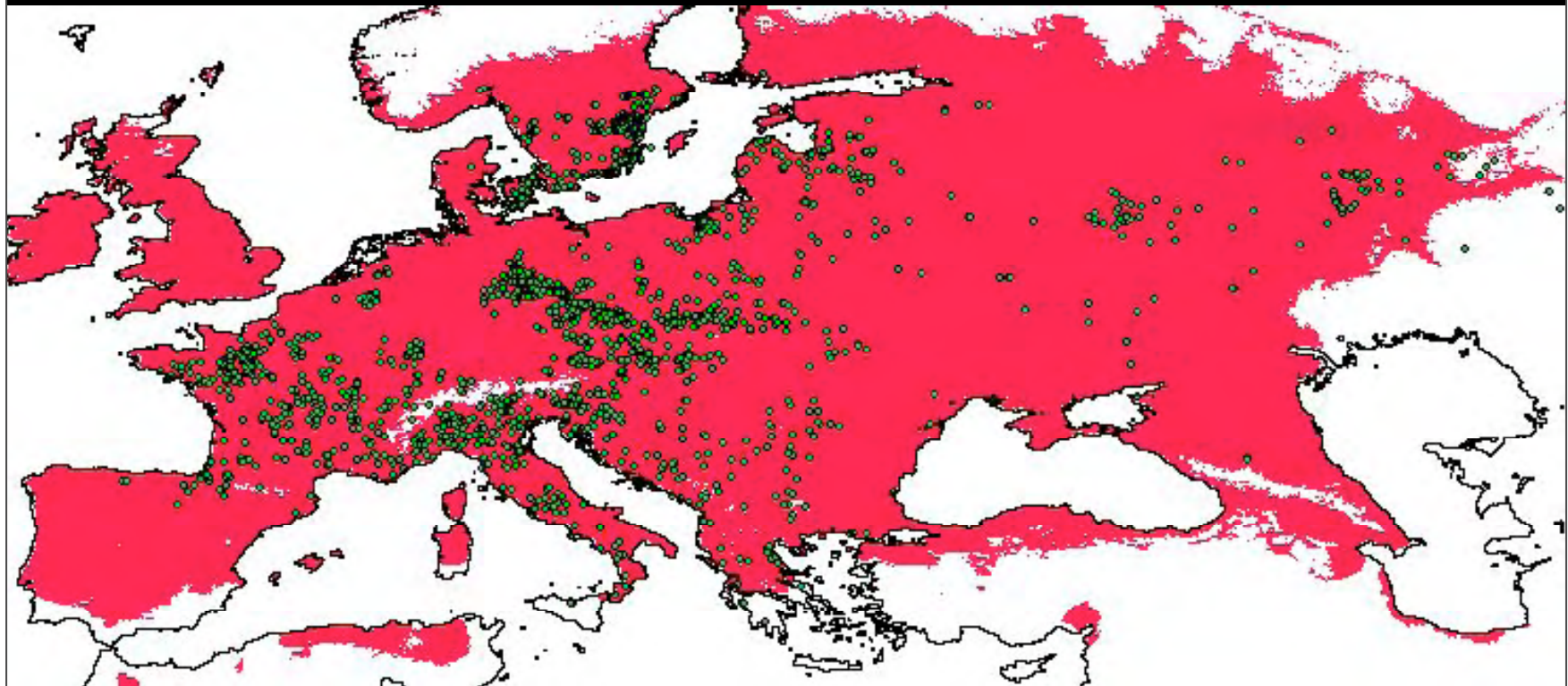
We can apply an algorithm to obtain a map of possible distribution of the species based on the environmental conditions where it lives.



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



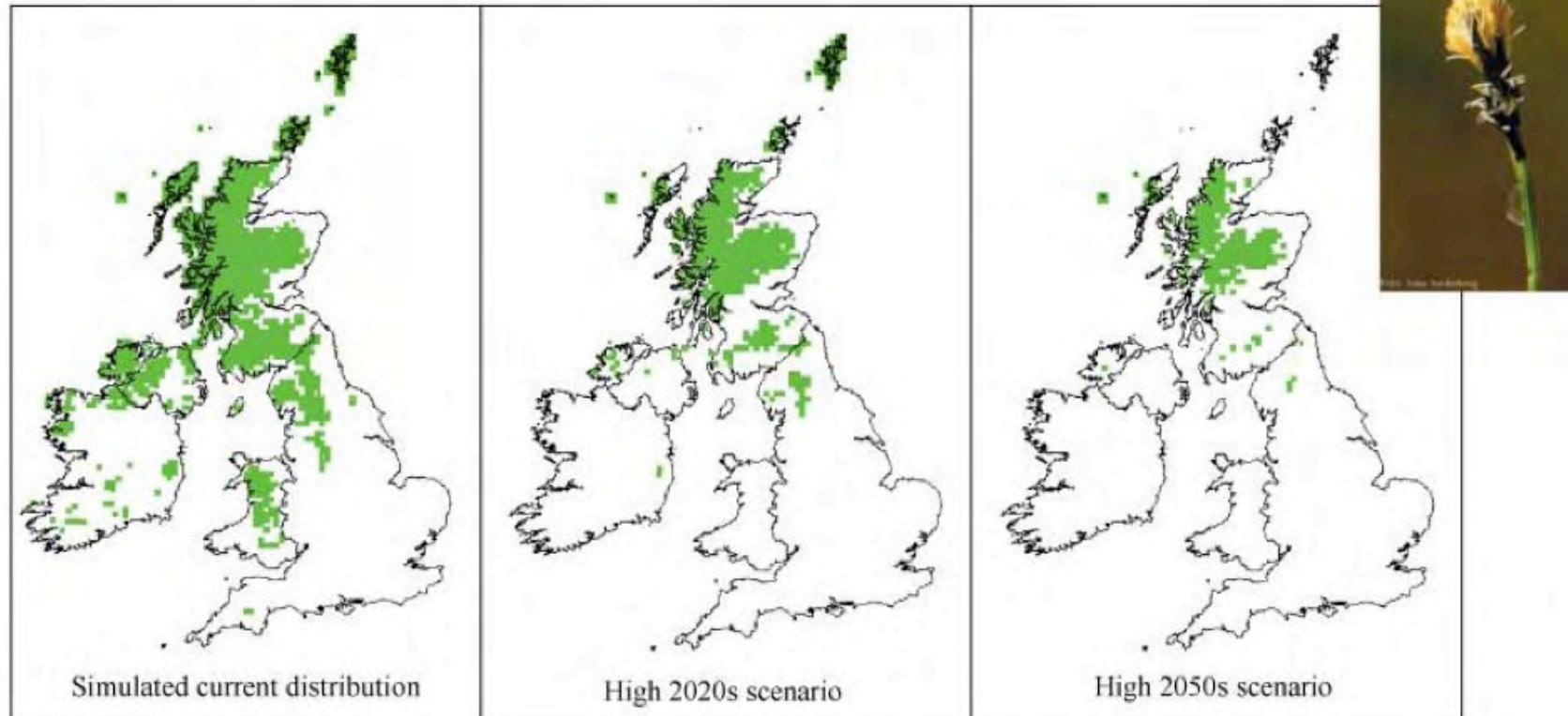
compared to the original
one



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

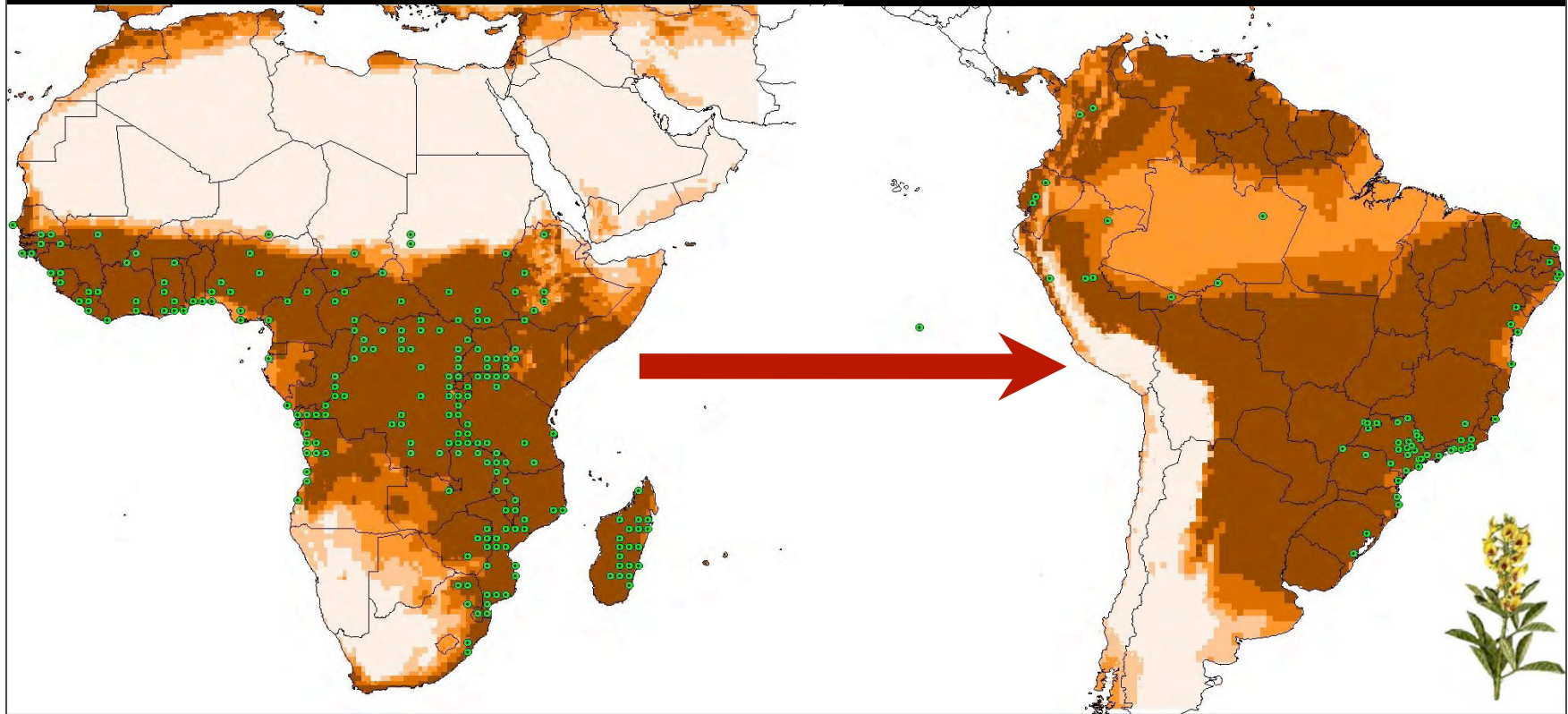


Pearson et al., 2002

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



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

Ocurrence data,
specimen-observation
data

+

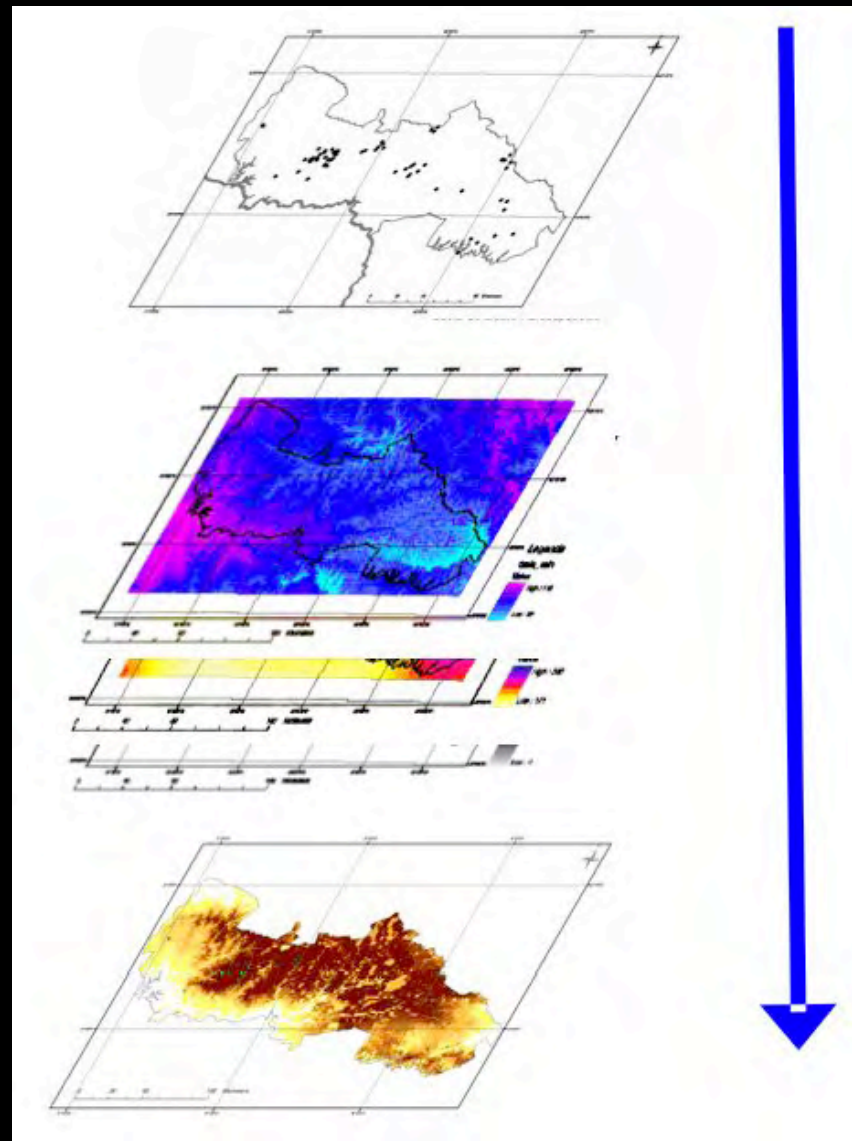
Environmental layers

+

Modelling algorithm

=

Distribution models



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



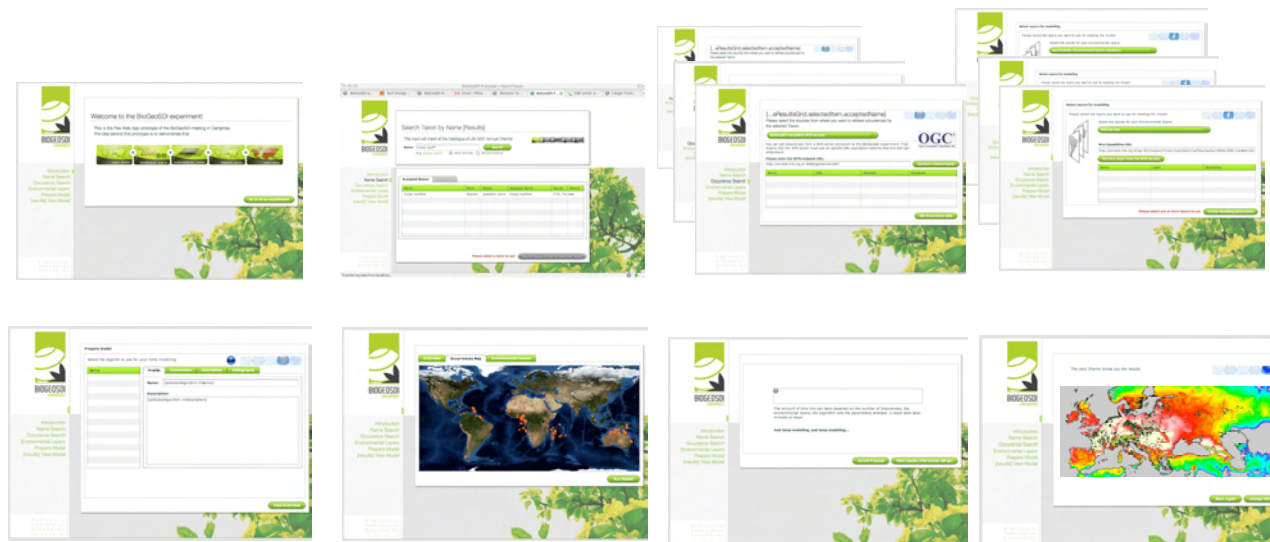
BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
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/>



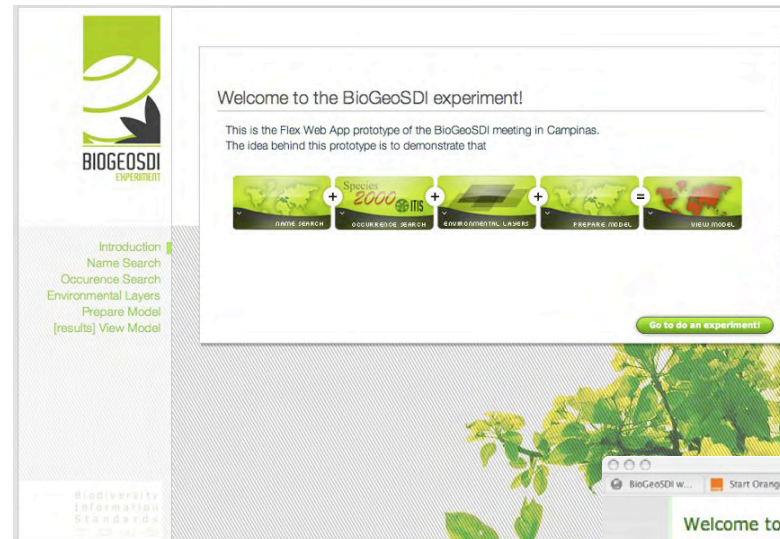
BioGeoSDI Code Fest in Campinas - April 2007

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.

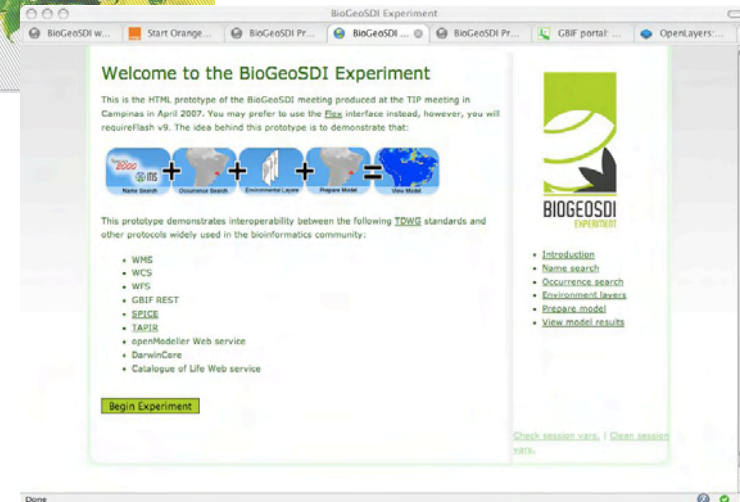


I) Start



Flex (flash) version

PHP-HTML version



BioGeoSDI Code Fest in Campinas - April 2007

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 S

2) Search for a scientific Name



SPICE protocol, COL REST service

Search Taxon by Name [Results]

This input will check at the Catalogue of Life 2007 Annual Checklist

Name:

e.g. [Cocos nucif*](#) Web Service SPICE Protocol

Name	Rank	Status	Accepted Name	Source	Details
Cocos nucifera	Species	accepted name	Cocos nucifera	ITIS: The www	

Please select a name to use

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
T D W S

3) Search for occurrences



WFS with our own GML app schema

The screenshot shows the BioGeoSDI web application interface. On the left is a navigation menu with the following items: Introduction, Name Search, Occurrence Search (highlighted), Environmental Layers, Prepare Model, and [results] View Model. The main content area contains a search form for WFS services. At the top of the form, it displays the selected taxon name: `{...eResultsGrid.selectedItem.acceptedName}`. Below this, it asks the user to select the sources for the selected taxon, with a dropdown menu currently set to "BioGeoSDI compliant WFS service". A note explains that the WFS server must use a specific GML application schema. The form includes a text input for the WFS endpoint URL, with the example `http://omtest.cria.org.br:8080/geoserver/wfs?`. A "Retrieve FeatureTypes" button is located to the right of the URL input. Below the input is a table with the following columns: Name, Title, Abstract, and Keywords. The table is currently empty. A "Get Occurrence data" button is located at the bottom right of the form. The OGC logo (Open Geospatial Consortium, Inc.) is visible on the right side of the form.`

BioGeoSDI Code Fest in Campinas - April 2007

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



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

3) Search for occurrences

GBIF REST service

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

Biodiversity
Information
Standards
T D W S

BIOGEOSDI
EXPERIMENT

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

GBIF REST services

The GBIF REST services provides you access to the entire GBIF cache compiled from all its data providers.

GLOBAL
BIODIVERSITY
INFORMATION
FACILITY

Get Occurrence data

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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W G

3) Search for occurrences

TAPIR protocol providers

BIOGEOSDI
EXPERIMENT

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

`{...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

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 S

4) Get environmental layers



WCS

Select Layers for modelling

Please select the layers you want to use for creating the model

Select the source for your environmental layers:

WCS service

Wcs Capabilities URL:
`http://omtest.cria.org.br/cgi-bin/mapserv?map=/opt/data/mapfiles/hadley1990to1999.map&service=`

Retrieve layer from the WCS service

Name	Label	Description

Please select one or more layers to use **Define Modelling parameters**

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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

4) Get environmental layers

Searching on a catalog Service (not implemented)

BIOGEOSDI
EXPERIMENT

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

Select Layers for modelling

Please select the layers you want to use for creating the model!

Select the source for your environmental layers:

OGC Catalog Service

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.

Biodiversity
Information
Standards

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 :(



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

4) Get environmental layers

Using Modelling service local layers
(OMWS getAvailableLayers)

The screenshot shows the BIOGEOSDI EXPERIMENT web interface. On the left is a navigation menu with links: Introduction, Name Search, Occurrence Search, Environmental Layers, Prepare Model, [results] View Model. The main content area displays a dialog box titled 'Select Layers for modelling'. The dialog contains the following elements:

- A header: 'Select Layers for modelling'
- A prompt: 'Please select the layers you want to use for creating the model'
- A dropdown menu for 'Select the source for your environmental layers:' with the selected option 'openModeller Environmental layers repository'.
- A prompt: 'select one or several layers to use in your model (keep ctrl. press to select multiples)'
- A table with the following structure:

ID	Label	hasProjection

At the bottom of the dialog, there is a red error message: 'Please select one or more layers to use' and a green button labeled 'Define Modelling parameters'.

BioGeoSDI Code Fest in Campinas - April 2007

45

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.



BIOGEOSDI
EXPERIMENT

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 downloaded to the modelling server to have it ready when doing the experiment.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

5) Select Model algorithm & parameters

OMWS getAlgorithms



Prepare model

Select the algorithm to use for your niche modelling

Name	Profile	Parameters	Description	Bibliography

Name: {selectedAlgorithm.mName}

Description: {selectedAlgorithm.mDescription}

[View Overview](#)

BioGeoSDI Code Fest in Campinas - April 2007

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
T D W S

6) Overview before modelling

The image displays three overlapping screenshots of the BIOGEOSDI web application interface, specifically the 'Overview' tab. The interface includes a navigation menu on the left with options like 'Introduction', 'Name Search', 'Occurrence Search', 'Environmental Layers', 'Prepare Model', and 'View Model'. The main content area shows search results, including 'Species 2000' and 'Catalogue of Life: 2007 Annual Checklist'. A world map with orange dots is visible in the middle screenshot, and a color-coded environmental layer map is shown in the bottom screenshot. The BIOGEOSDI logo and 'EXPERIMENT' text are present in the top left of each screenshot.

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...



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

7) Run model



OMWS runModel, getLog (SOAP service)

The screenshot shows the BioGeoSDI web interface. On the left is a navigation menu with the following items: Introduction, Name Search, Occurrence Search, Environmental Layers, Prepare Model, [results] View Model. The main content area displays a loading screen with a progress indicator and the text: "The amount of time this can take depends on the number of Occurrences, the environmental layers, the algorithm and the parameters selected. It could take minutes or days!" Below this text is the phrase "Just keep modelling, just keep modelling...". At the bottom of the loading area are two buttons: "Cancel Process" and "View results (this button will go)".

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!



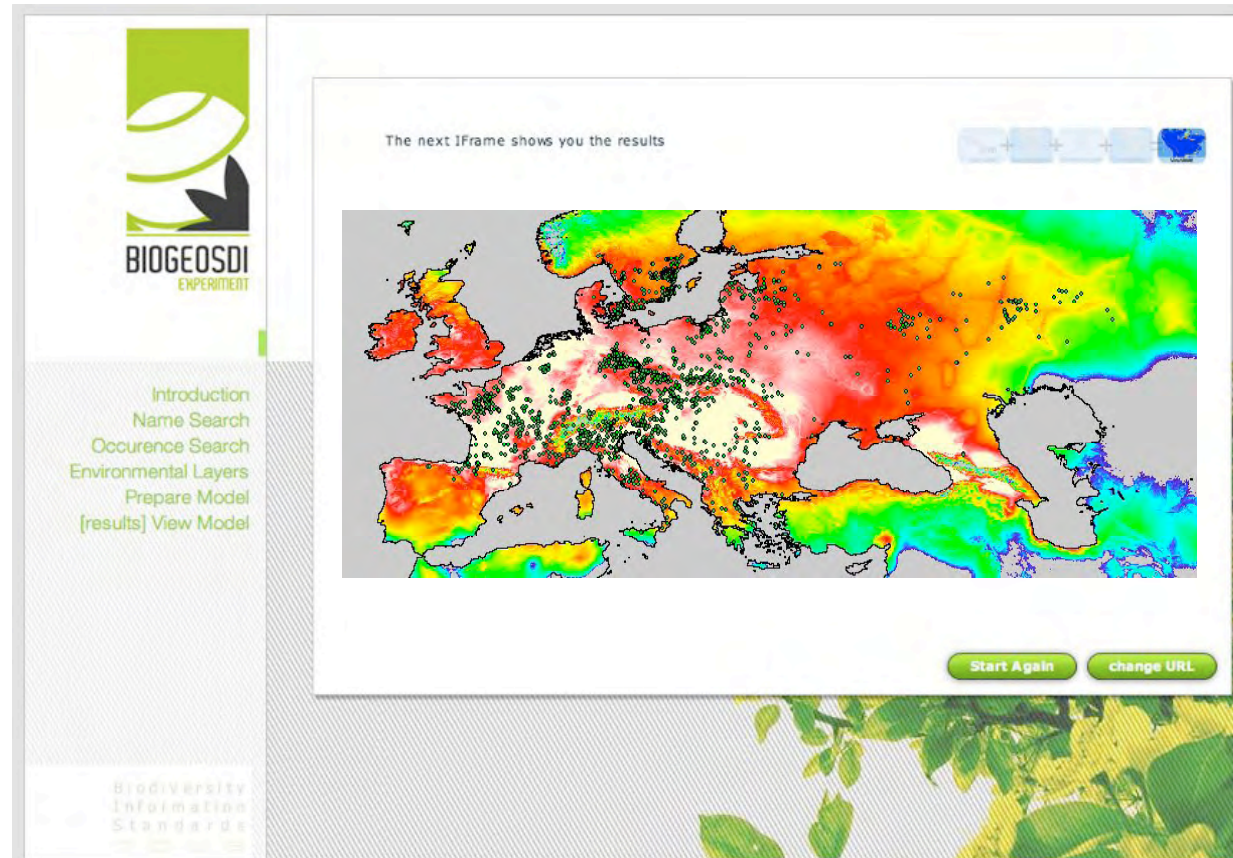
BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

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.



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W S

Technologies used

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

- PHP, Flex, HTML, Bash, Python

BioGeoSDI Code Fest in Campinas - April 2007

So this are the technologies we used...



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
TDWG

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

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



BIOGEOSDI
EXPERIMENT

Biodiversity
Information
Standards
T D W G



Thanks!

Workshop Wiki

<http://wiki.tdwg.org/twiki/bin/view/Geospatial/InteroperabilityWorkshop1>

Prototype

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

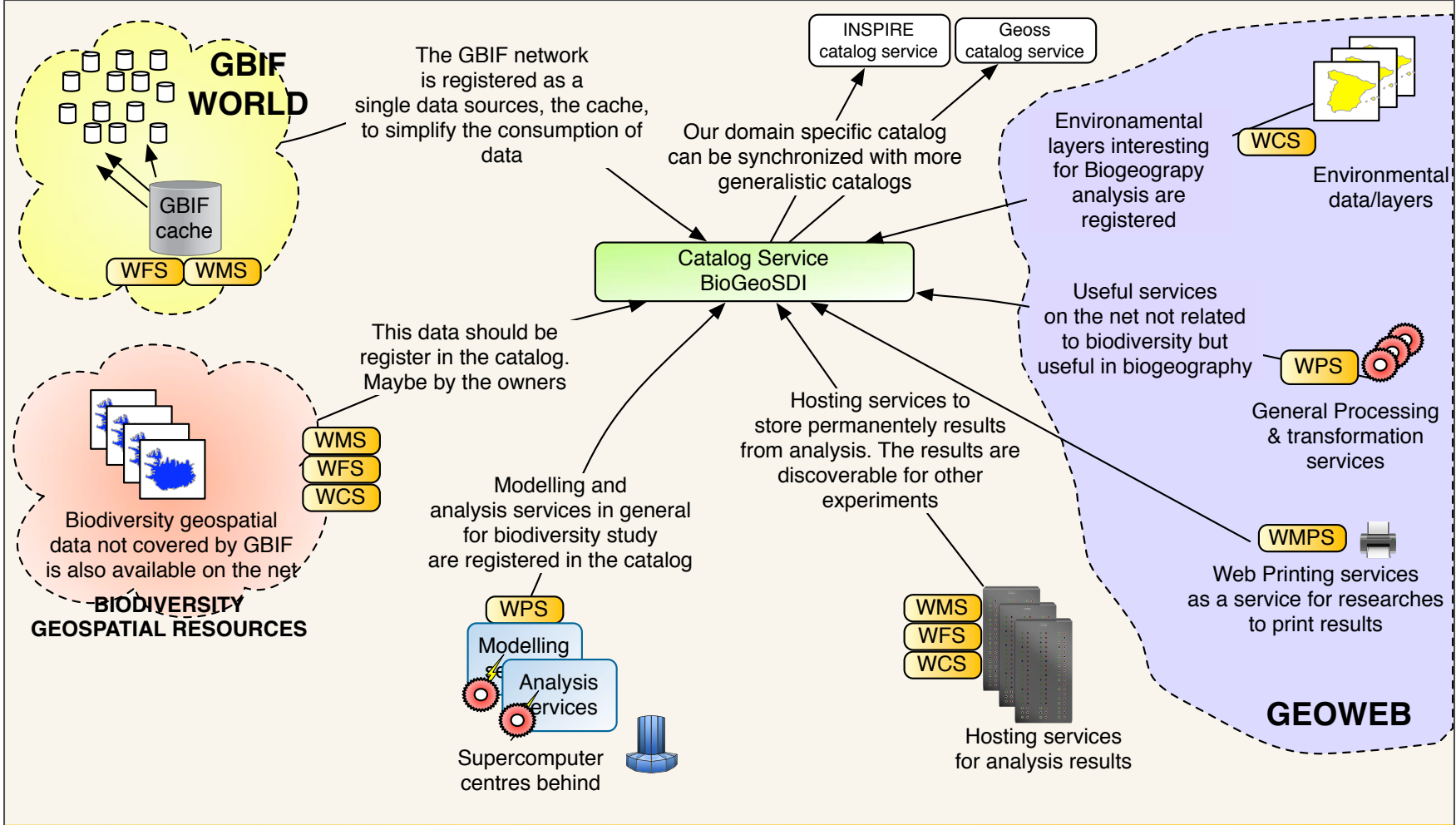
Source Code and documentation

<http://biogeosdi.org/>

Report

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

BioGeoSDI Code Fest in Campinas - April 2007



Biogeography Spatial Data Infrastructure

BioGeo Spatial Data Infrastructure

BioGeoSDI Code Fest in Campinas - April 2007

This not to be seen...