



Geo VII, Beijing, China

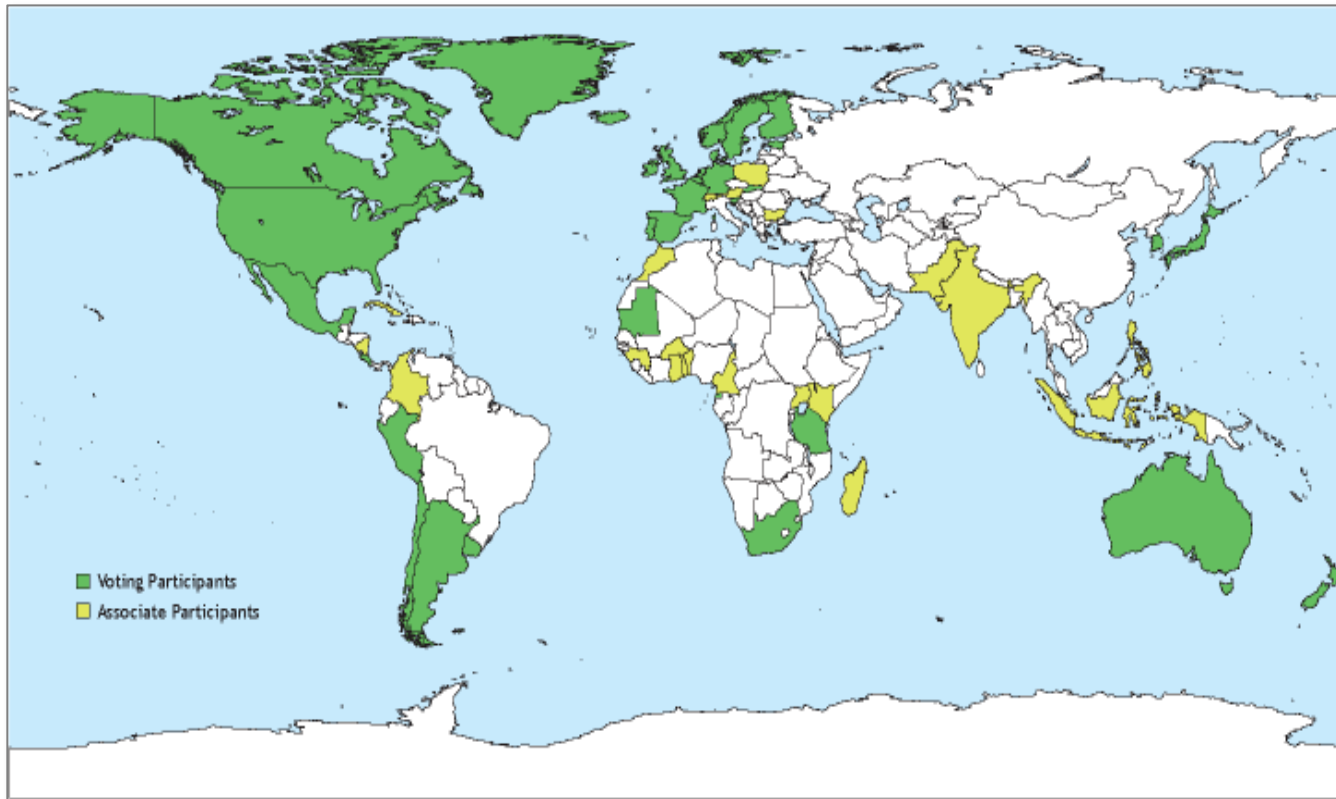
Benefits of Data sharing: the role of GBIF

Dr Nicholas King
Executive Secretary
Global Biodiversity Information Facility (GBIF)

November 2010



Current GBIF Participants



55 Countries

46 International Organisations

2007-11 Strategic Plan, to:

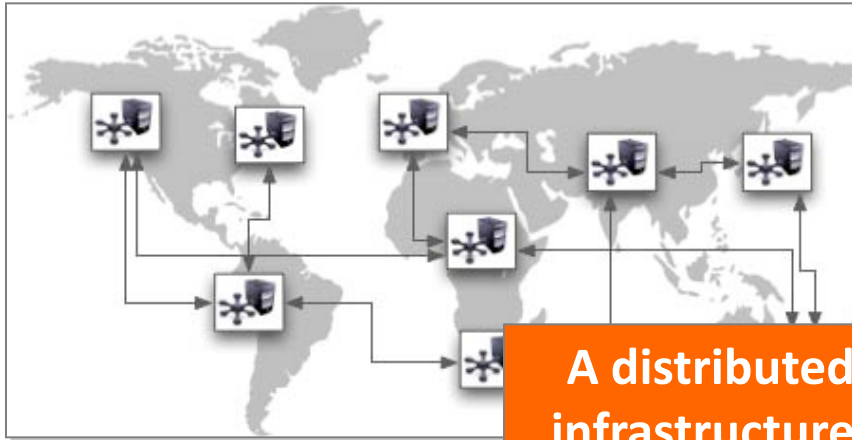
- 1. Make a whole world of biodiversity data that are currently exceedingly difficult to access freely and universally accessible via the Internet;*
- 2. Enable scientific research that has never before been possible; and*
- 3. Facilitate the use of scientific data in biodiversity policy- and decision-making.*

And new draft Strategic Plan 2012-2016 makes very explicit the need to build robust links from genomics to ecological data.....

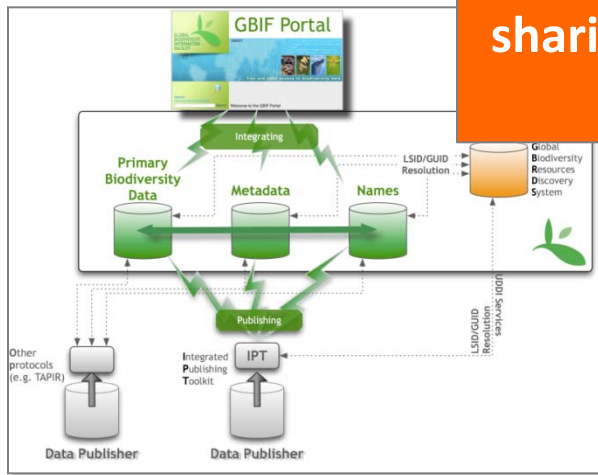
Key Data sharing Challenges

- **Common data management principles;**
- **Data sharing protocols, IPR, 'ownership'...;**
- **Data standards = compatibility, integration;**
- **Data types – what is needed?**
- **Data quality – fitness-for-use?**
- **Data volumes – how much is enough?**
- **Metadata, vocabularies, ontologies;**
- **Attribution, trace authorship, recognition, data owners vs data aggregators;**
- **Benefits/Incentives for data sharing?**
- **Data Security - Open Access vs Sensitive Data;**
- **Data download and use tracking;**
- **Analytical tools: web services – what good is access without the means to interrogate?**
- **Who pays?**

A global infrastructure for data exchange/sharing via a single portal



A distributed global infrastructure for the sharing of biodiversity data




GLOBAL BIODIVERSITY INFORMATION FACILITY

SPECIES COUNTRIES DATASETS OCCURRENCES SETTINGS ABOUT

... free and open access to biodiversity data

Welcome to the GBIF Data Portal

Access millions of data records shared via the GBIF network. To learn how to use this site, please see [About](#). To tune this site for smaller displays, see [Settings](#). Version 1.2.5 - [click here](#) to see what is new!

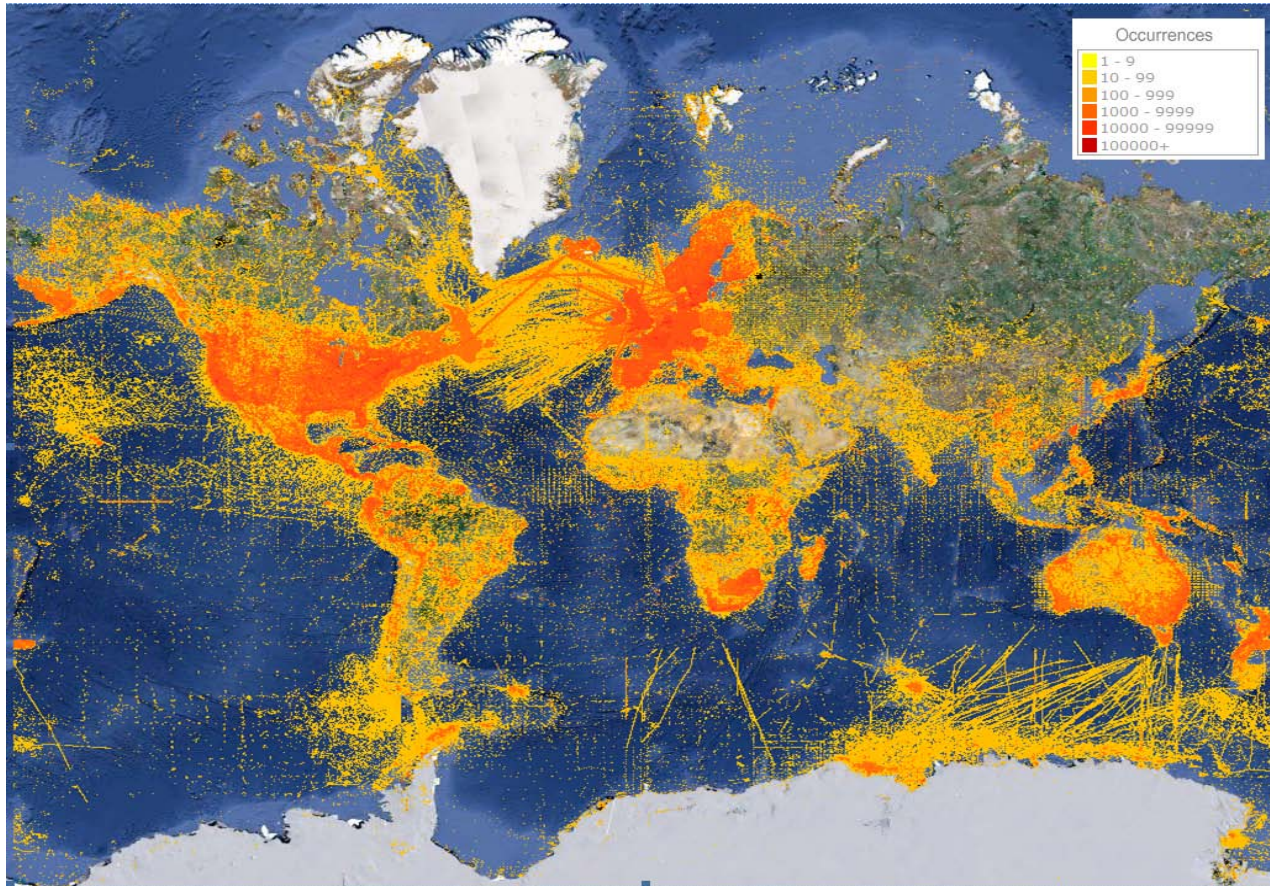
Explore Species: Find data for a species or other group of organisms. **Species** Information on species and other groups of plants, animals, fungi and micro-organisms, including species occurrence records, as well as classifications and scientific and common names. **Example species:** *Puma concolor* (Linnaeus, 1771)

Explore Countries: Find data on the species recorded in a particular country. **Countries** Information on the species recorded in each country, including records shared by providers from throughout the GBIF network. **See data for:** [Denmark](#)

Explore Datasets: Find data from a data provider, dataset or data network. **Datasets** Information on the data providers, datasets and data networks that share data through GBIF, including summary information on 8634 datasets from 299 data providers. **Latest dataset added:** (Table 1) Distribution of planktonic foraminifera from samples spanning the Cretaceous-Tertiary boundary at ODP Site 207-1259

Layout & design © GBIF. Data providers retain all rights to data. Contact us

Data shared via GBIF network



October 2010: >220 million occurrence records from >10,000 datasets from >300 publishers and spanning a wide range of geospatial, temporal and taxonomic coverages being shared

Data shared via GBIF network (2)

The greatest concentrations of biodiversity are found in developing countries but most of the data and information about it are located in developed countries.



Data sharing with countries of origin

An overview of data records available in the GBIF network

The table shows numbers of records, for countries (and in some cases regions within countries), shared by other countries and international networks. How to do searches:

- ▶ Using the drop-down menu select the country(ies) of interest and country(ies) that host data for them
- ▶ **Countries that host data:** Select a country at the top of the table (that is, by column) to get an overview of the data they share.
- ▶ **Countries to which the data are relevant:** Select a country on the left (that is, by row) for an overview of the numbers of records relevant to that country that are shared by other countries or international networks.
- ▶ **Intersecting** cells indicate the number of records hosted and shared by country X/international networks that are relevant to country Y.

Change view: Host: Country:

Download as tab file

HOST	AR	AT	AU	BE	CA	CH	CN	CO	CR	DE	DK	EE	ES	FI	FR	GB	HK	IE	IL	IN	IS	JP	KR	LU	MG	MX	NI	NL	NO	NZ	PE	PK	PL	PT	SE	SI	SK	TZ	UK	US	ZA	Int. networks					
COUNTRY	AR	AT	AU	BE	CA	CH	CN	CO	CR	DE	DK	EE	ES	FI	FR	GB	HK	IE	IL	IN	IS	JP	KR	LU	MG	MX	NI	NL	NO	NZ	PE	PK	PL	PT	SE	SI	SK	TZ	UK	US	ZA	Int. networks					
AD	X	X								X	X		X		X							X																				X		X			
AE		X	X	X		X				X			X	X	X					X		X						X					X		X	X					X	X		X	AE		
AF		X	X		X	X				X	X		X	X	X							X						X	X	X		X		X							X		X		X	AF	
AG		X	X		X					X					X	X												X					X								X		X		X	AG	
AI		X				X				X					X												X									X						X		X		X	AI
AL		X		X	X	X				X	X		X		X							X						X	X				X		X	X					X	X		X	AL		
AM		X	X	X	X	X				X	X		X	X	X							X					X	X	X				X		X						X		X		X	AM	
AN		X	X	X	X	X				X	X		X	X	X							X					X	X	X				X		X						X		X		X	AN	
AO		X	X	X	X	X				X	X		X	X	X	X						X					X	X	X				X		X		X	X				X	X		X	AO	
AQ	X	X	X	X	X	X				X	X		X	X	X	X				X		X				X		X	X	X		X		X							X	X		X	AQ		

As a result of the GBIF data sharing mechanism, countries now have quick and easy access to data about their own biodiversity served by institutions around the world.



GBIF promotes Data Sharing

... through:

- freely and openly distributable software that meets community standards;
- helpdesk services, “how-to” manuals, training workshops, etc;
- straight-forward data sharing and data use agreements that encourage responsible sharing and use;
- ensuring the data made available via the GBIF network belong to the data owners and remain fully under their control - Provision of a clear statement on **Intellectual Property Right** (IPR) issues in the **Memorandum of Understanding** that is signed between GBIF and each Participant;



GBIF IPR Principles

- **Access to data** - all users have equal access to data in databases shared in GBIF, although they are asked to respect conditions set by data providers;
- **Attribution** - GBIF works to ensure that the source of data is acknowledged by users and requires that attribution is maintained in subsequent uses of the data;
- **Rights** - GBIF does not assert any proprietary rights whatsoever to the data in the databases that are affiliated with GBIF;
- **Data collection** - GBIF asks for reasonable assurance from data providers that data collection was consistent with applicable laws, regulations and requirements for prior informed consent.

Interoperability is key


- *Open access is essential* to enable more effective decision making, and much effort has gone into overcoming technical, economic, socio-cultural and legal challenges;
- BUT: the value of individual datasets/information systems increasingly lies more in their *capacity to be cross-linked and used together with others*;
- Too little effort has gone into *building the 'cross-walks' between existing major information systems*;
- Only achievable thorough *partnerships* - GBIF has made *significant progress with primary biodiversity data and GEO BON provides the platform to extend to all biodiversity data types...*



Biodiversity metadata network established

Provides a standards-based mechanism for connecting **metadata catalogues** on the GBIF network

Provides a metadata catalogue for biodiversity resources for EuroGEOSS (and GEO BON)



GBIF
www.gbif.org
...free and open access to biodiversity data

Metadata - a Core Component of the GBIF Network

Key recommendations of the GBIF Metadata Implementation Framework Task Group (IMFTG) are guiding the implementation of the catalogue system. These include:

GBIF should adopt, or adapt, existing technology where possible and seek to collaborate on any new development in order to maximize impact of its development resources.

The metadata catalogue system should:

- support multiple metadata models natively;
- be able to return the original contributed metadata object;
- support replication and harvesting of metadata from providers;
- register with one or more host registries to advertise services available;
- expose one or more standard query Application Programming Interfaces (APIs) for programmatic access by client applications.

Metadata should be able to describe multiple types of primary biodiversity data.

Metadata should support data discovery, interpretation, and analytical reuse.

GBIF should accept metadata in multiple formats that are in common use.

GBIF should provide crosswalks to enable retrieval of metadata in multiple standards commonly in use in order to aid interoperability.

GBIF minimum requirements for metadata provision should be limited, but GBIF should accept very detailed metadata in any of the standard formats. The minimal acceptable metadata record might only include the Identifier, Title, Creator, Contact, Metadata Publisher, and Abstract.

The Global Biodiversity Information Facility (GBIF) is designing its Informatics Infrastructure as a scalable, distributed architecture that adheres to international standards for data exchange formats and protocols thereby enabling the maximum degree of **interoperability** across heterogeneous, distributed data holdings and applications. In large distributed networks featuring numerous providers and consumers of data, the key activities of inventory, discovery and access must be well coordinated through provision of registries and metadata catalogues, and through generation of specific indexes. Metadata is thus a central component in an expanded GBIF network that would offer many types of web services for delivering and consuming data.

The Initial Operating Capability (IOC) of the GBIF metadata Infrastructure (Fig. 1), planned for 2010, will feature a centralised, indexed cache of harvested metadata documents derived, through reciprocal sharing agreements, from 1) the GBIF data cache of primary occurrence records, 2) GBIF Integrated Publishing Toolkit Instances, 3) GBIF Participants' metadata catalogues, 4) other participating networks' catalogues. The GBIF catalogue, in turn, will link to clearing house portals such as the EuroGEOSS¹ GeoPortal which will demonstrate multidisciplinary interoperability across the domains of biodiversity, forestry and drought, and contribute to the establishment of the Group on Earth Observations Biodiversity Observation Network (GEO BON)². Beyond the IOC phase, there are plans, based on the recommendations of the GBIF Metadata Implementation Framework Task Group (IMFTG)³, to build a distributed system of regional metadata catalogues that support replication and harvesting.

About GBIF

GBIF makes digital biodiversity data openly and freely available on the Internet for everyone, and endorses both open source software and open data access.
www.gbif.org

GBIF provides scientific biodiversity data for decision-making, research endeavours and public use.
data.gbif.org

GBIF is a network of data publishers who retain ownership and control of the data they share. Linked datasets provide a more robust representation of biodiversity than any single dataset.

GBIF provides access to primary biodiversity data held in institutions in developed and developing countries. Data shared through GBIF are reprocessed data.

GBIF is a dynamic, growing partnership of countries, organisations, institutions and individuals working together to multibilise scientific biodiversity data.

GBIF invites you to download species occurrence data freely and openly from <http://data.gbif.org>

GBIF invites you to join the GBIF network and share your biodiversity data, as well as participate in developing new tools and services.

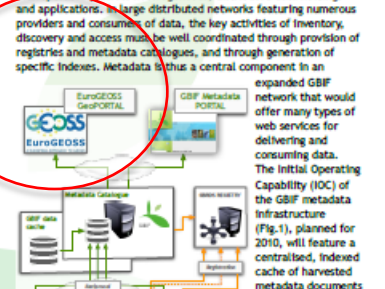

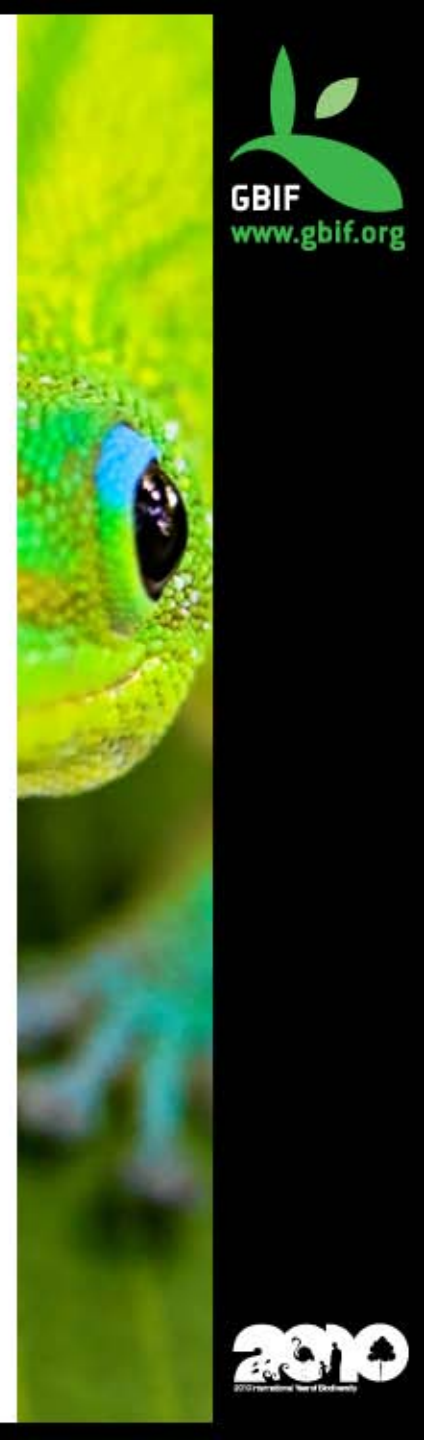


Figure 1: The main metadata related components envisaged for the Initial Operating Capability of the GBIF network.

¹ www.eurogeoss.eu
² www.earthobservations.org/geobon.shtml
³ Report of the GBIF Metadata Implementation Framework Task Group (IMFTG), Jones, M. et al., September 15, 2009 (draft available from ecotama@gbif.org)



2010
INTERNATIONAL YEAR OF BIODIVERSITY

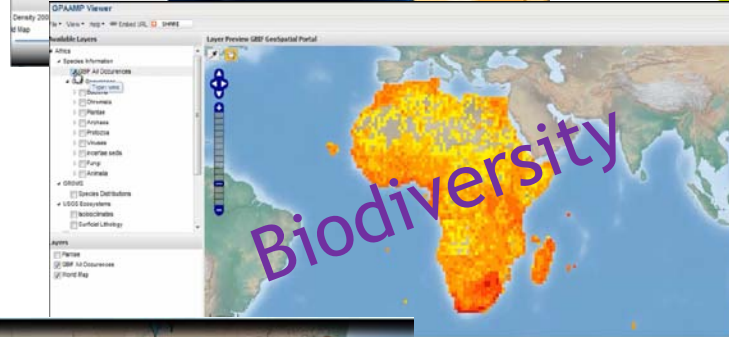
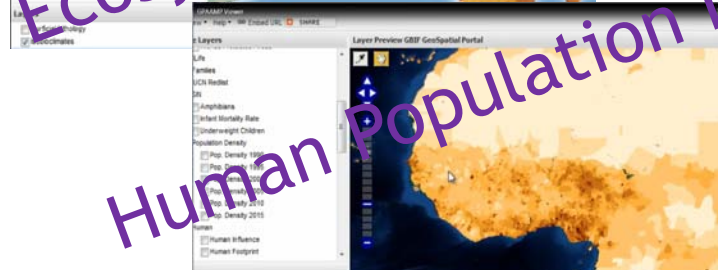


Delivering a GEO BON “Early Product”

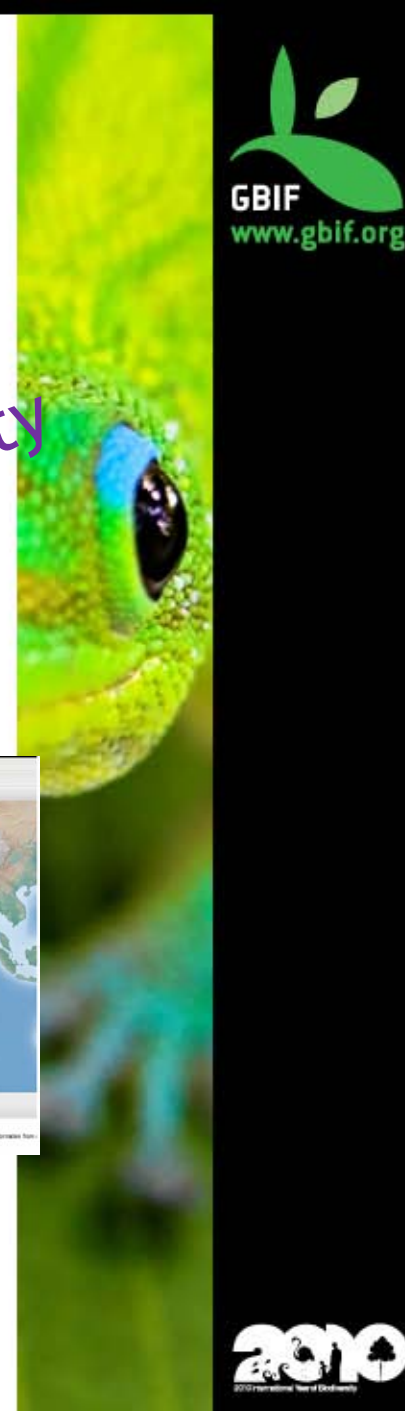
An Open Source, standards-based web GIS client.

A component of the EuroGEOSS DOPA.

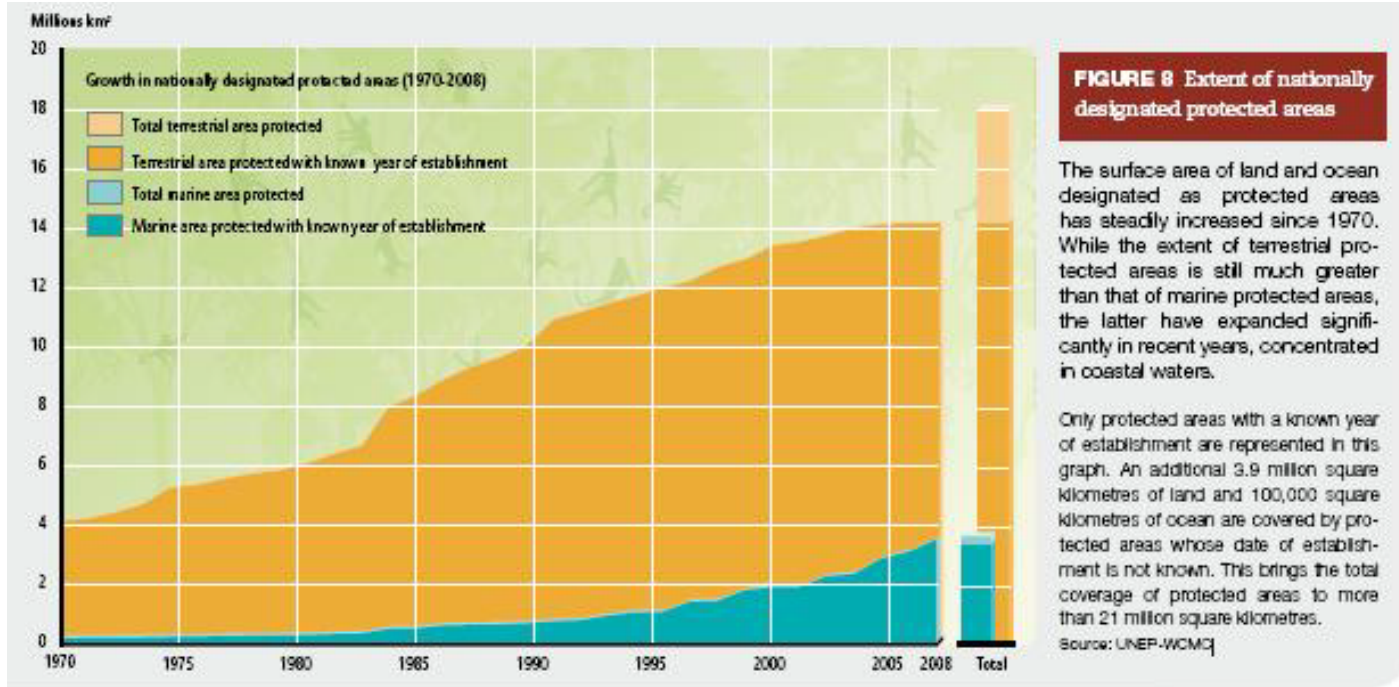
Provides viewing and download service for datasets from several sources including: GBIF, UNEP-WCMC, BirdLife, USGS, CIESIN, JRC.



Ecosystem Classification
Human Population Density
Biodiversity
Fire Drought Potentials



Increase in PA extent (GBO3)

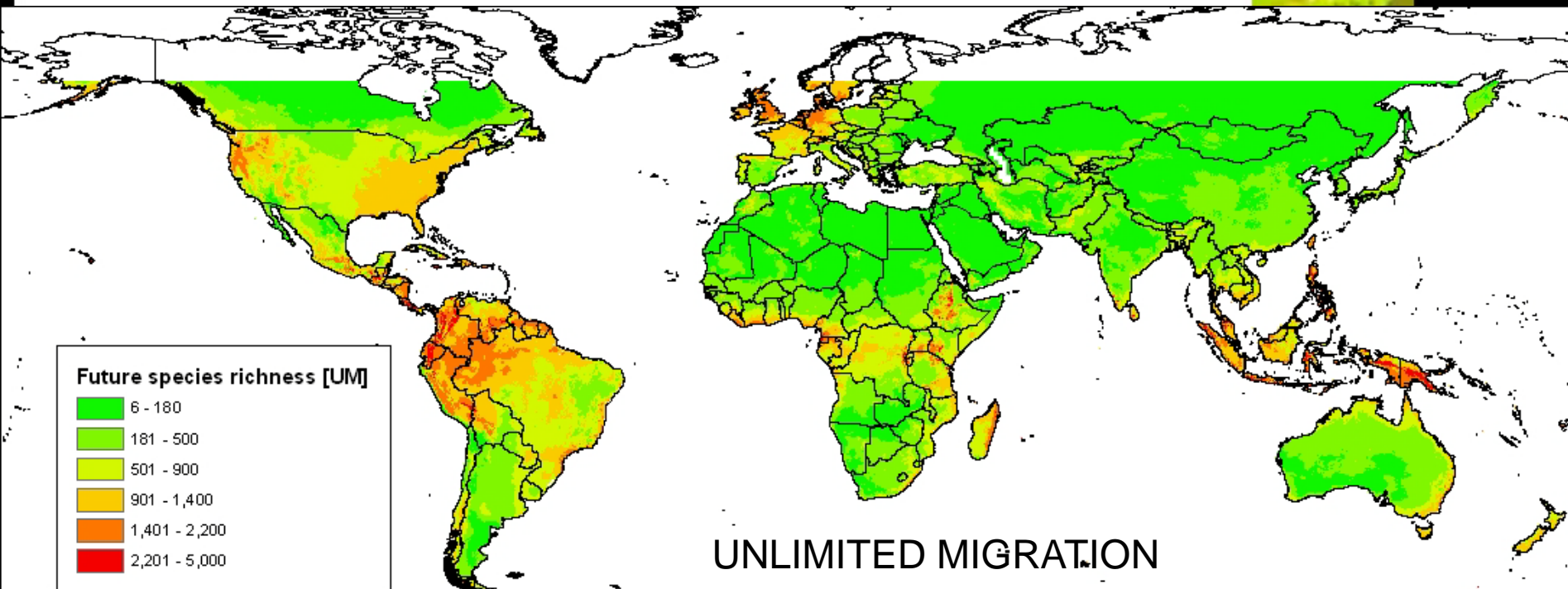


2020 PA targets?

How will we select additional areas? On current conservation value? On future value? And how do we connect the two?

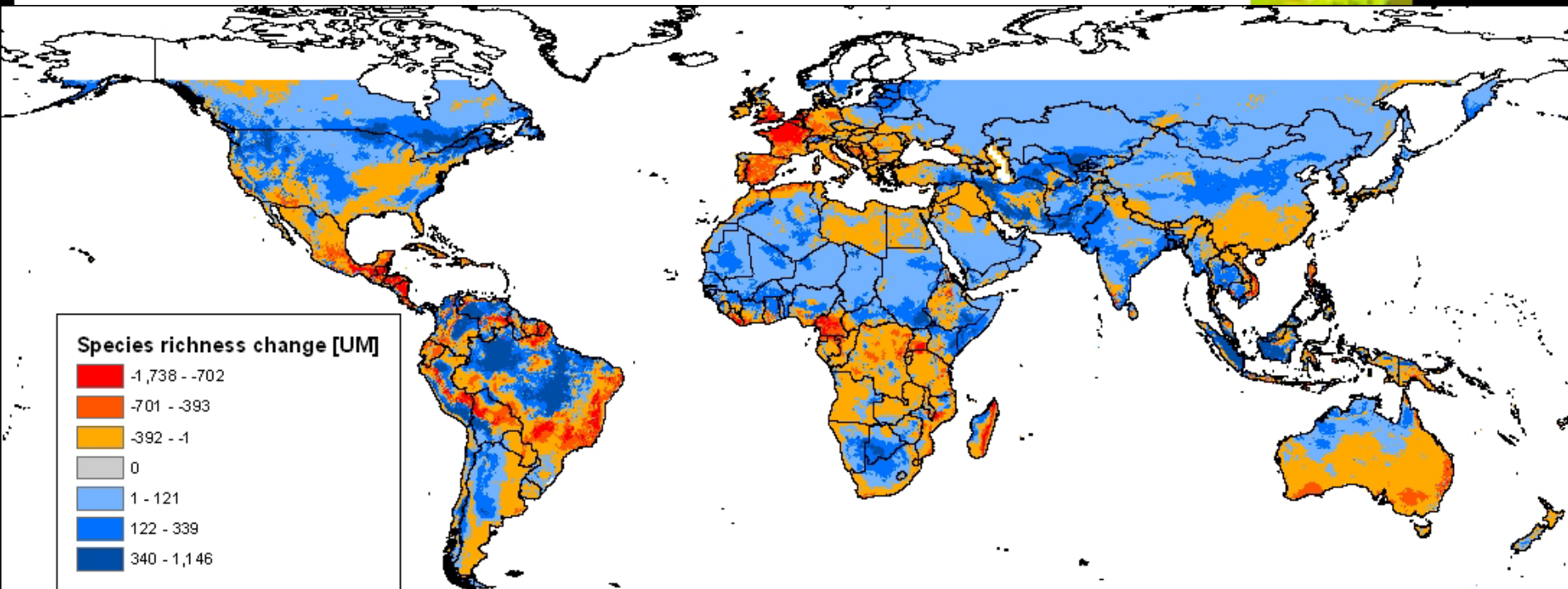
Current and future predicted plant species richness against 18 major climate change models

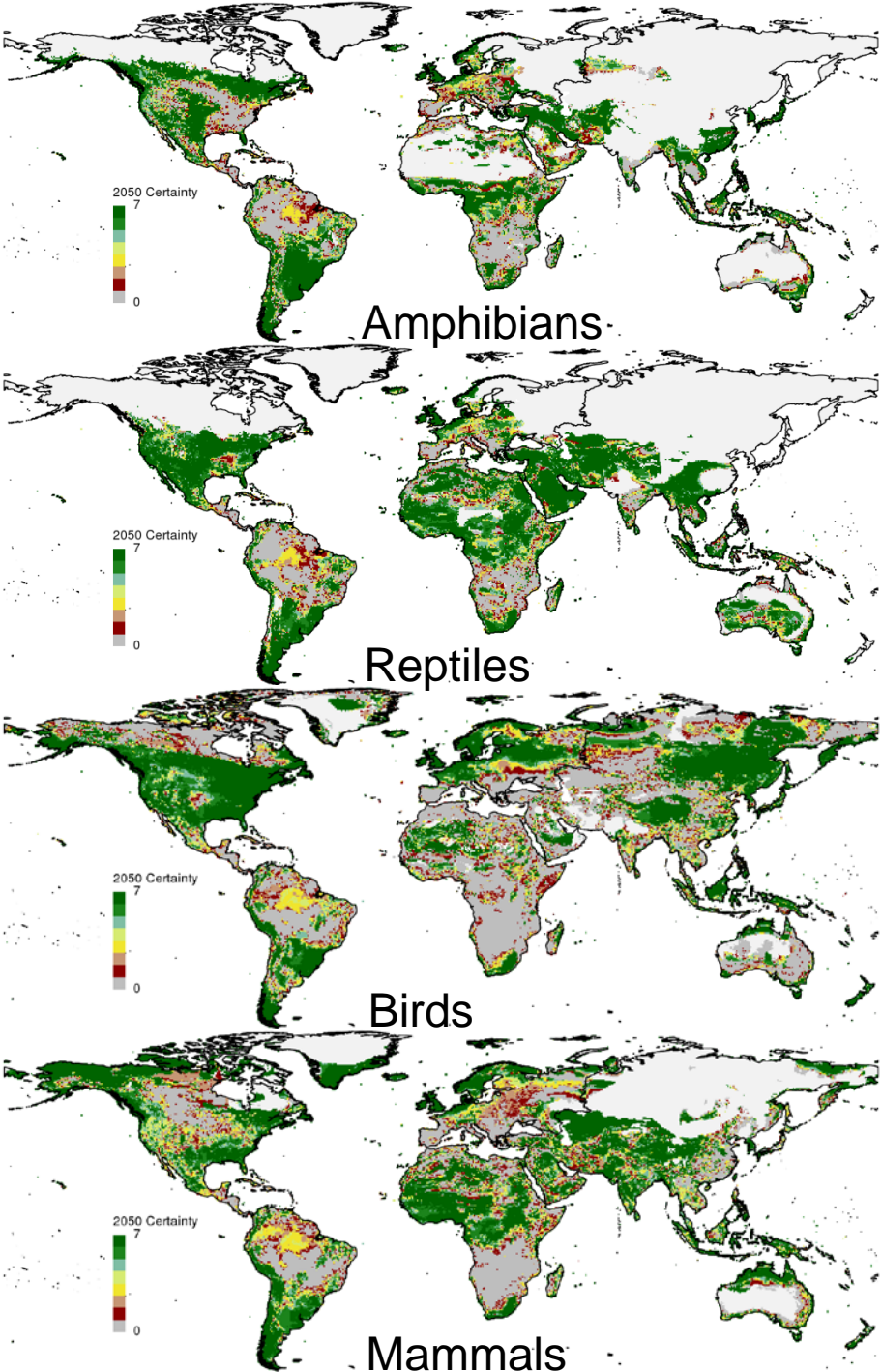
- Important hotspots in Latin America, Europe, Australasia and Central Africa
- Displacement and loss of niches



Results: changes in species richness

- Unlimited migration: mostly displacement
- Null migration: losses everywhere



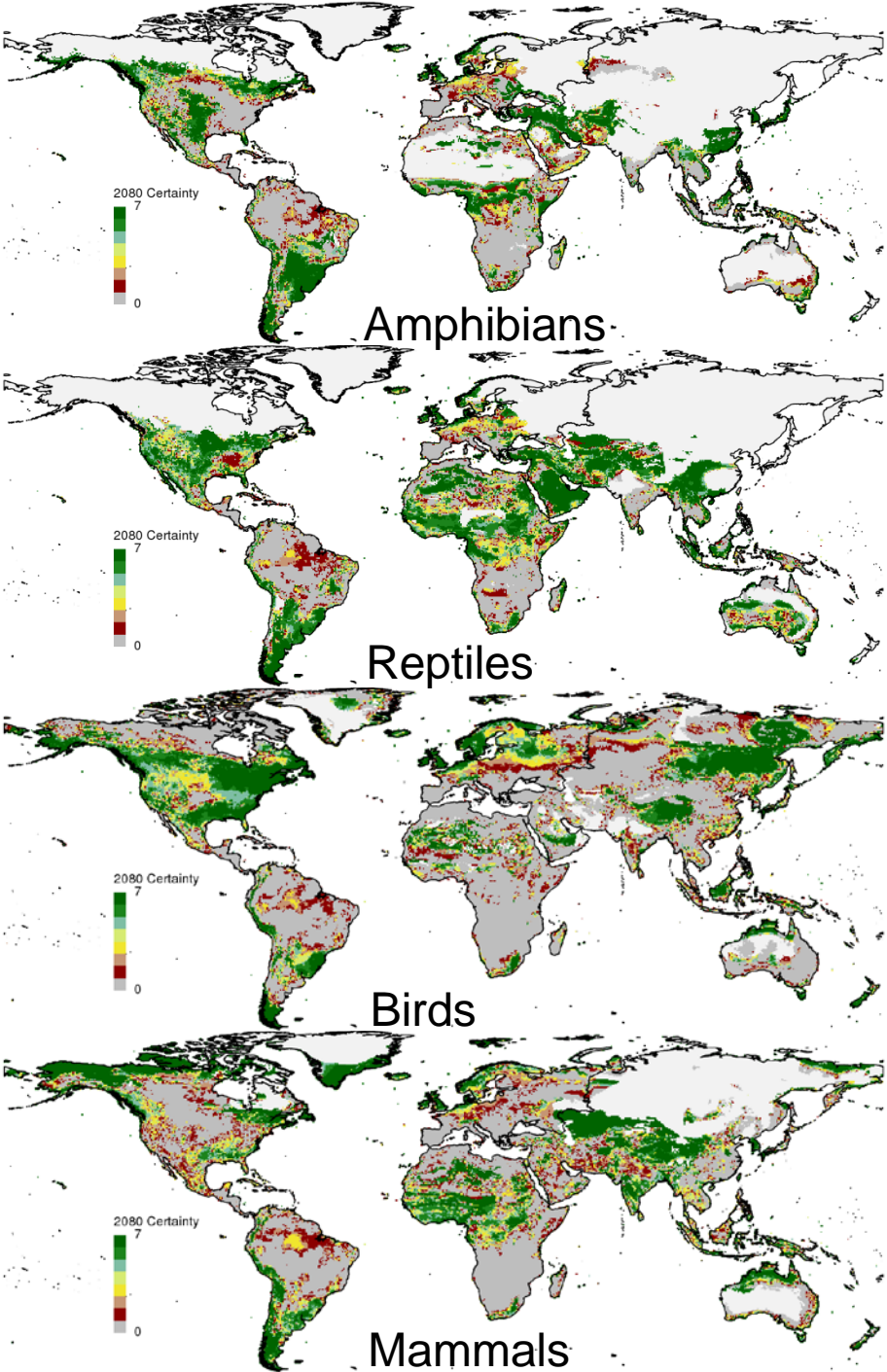


Potential Refugia at 2°C rise:

White - little or no data;
 Gray - no refugia;
 Other colors are areas projected to lose fewer than 25% of the species modeled.

(Colors correspond to number of climate models agreeing that area might be a refugia)





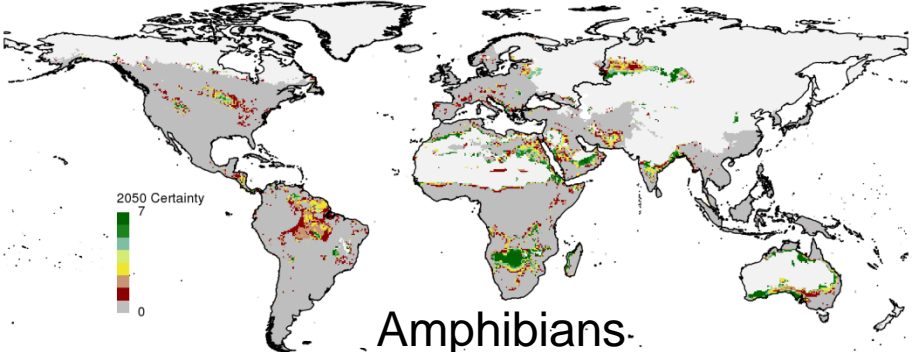
Potential Refugia at 3.5°C rise:

White – little or no data;
 Gray – no refugia;
 Other colors are areas projected to lose fewer than 25% of the species modeled.

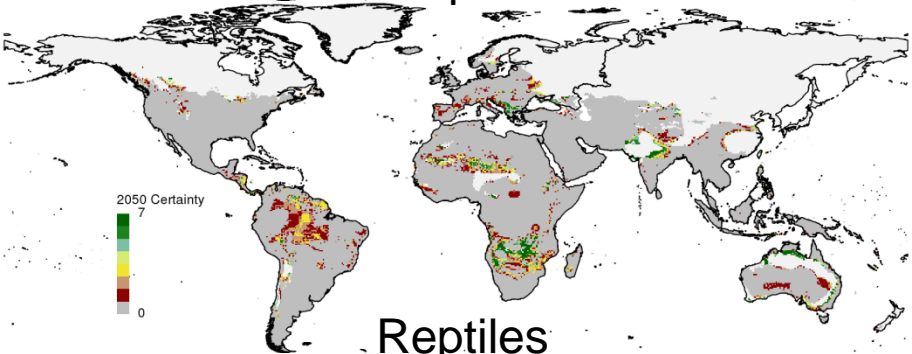
(Colors correspond to number of climate models agreeing that area might be a refugia)

Wallace Initiative

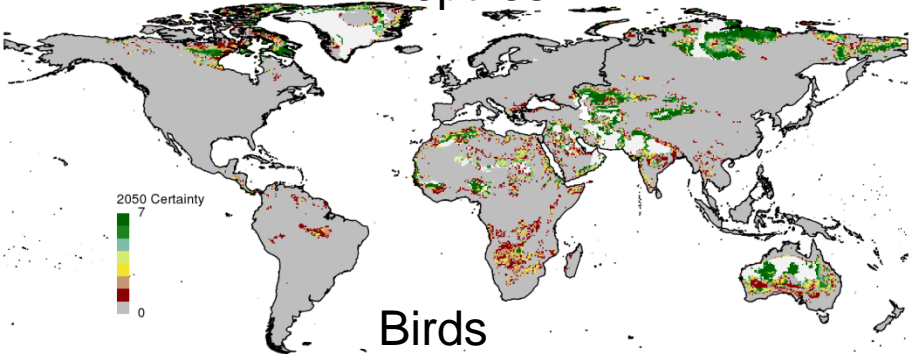




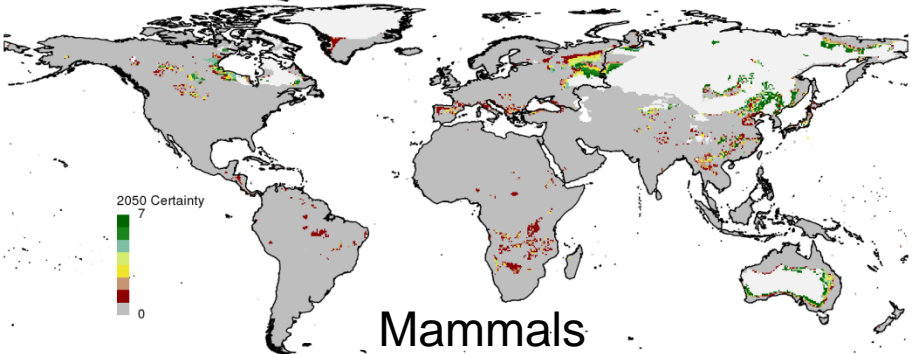
Amphibians



Reptiles



Birds



Mammals

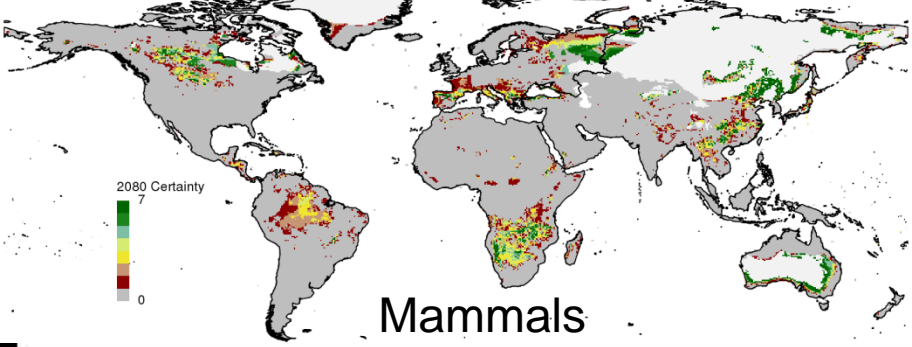
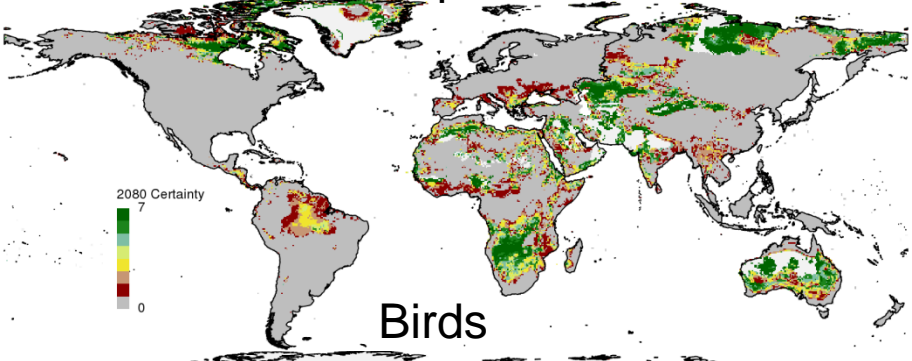
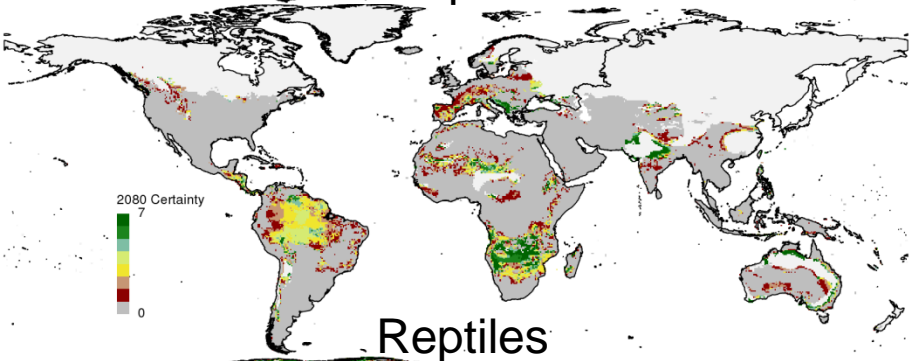
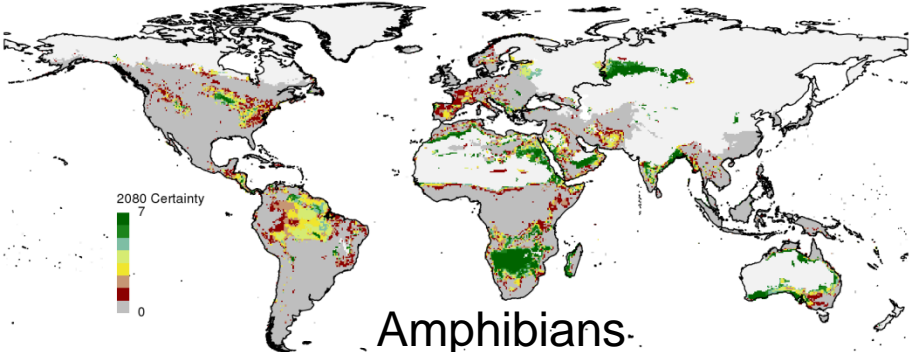
Areas of Highest Concern at 2°C rise – mass extinctions:

White – little or no data;
 Gray – species change >25% and <75%,
 Other colors are areas projected to lose more than 75% of the species modeled.

(Colors correspond to number of climate models agreeing that area might be climate sensitive and require significant adaptation action).

Wallace Initiative





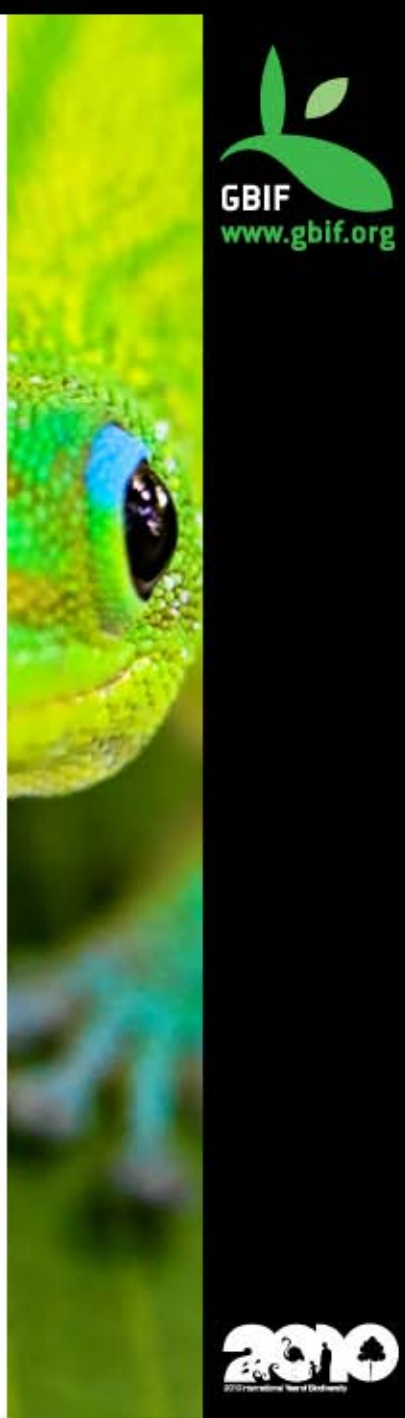
Areas of highest Concern at 3.5°C rise – mass extinctions:

White – little or no data:

Gray – species change >25% and <75%;

Other colors are areas projected to lose more than 75% of the species modeled.

(Colors correspond to number of climate models agreeing that area might be climate sensitive and require significant adaptation action).



GBIF - contributing 10 years of biodiversity informatics 'construction' expertise to GEO BON (and GEOSS)



 GROUP ON
EARTH OBSERVATIONS | GEO BON

PRESS RELEASE

**GEO Biodiversity Observation Network
to monitor and study the global biosphere**

Implementation Plan launched for International Day for Biodiversity

Geneva, 21 May 2010 – The Group on Earth Observations Biodiversity Observation Network (GEO BON) is releasing today its detailed, 175-page Implementation Plan for a coordinated global campaign to gather and share information on biodiversity, provide tools for data integration and

WG8 on **Data Integration & Interoperability**

Key outputs:

1. Detailed Implementation Plan
2. Principles of the GEO BON Information Architecture

GEO BON Detailed Implementation Plan

Approach - to avoid duplication and ensure compatibility with GEOSS:

1. Build on existing systems;
2. Coordinate, standardise and manage data held by a variety of disparate institutions and individuals for many different purposes;
3. Build permanent IT structures and linkages that will support the delivery of products from thematic groups;
4. Use guidance from GEOSS Architecture and Data Committee (ADC) and the concepts of the GEOSS Common Infrastructure (GCI) into design of GEO BON;
5. Develop separate companion document "*Principles of the GEO BON Information Architecture*" for introducing the GCI.

Principles of the GEO BON Information Architecture (1)

(companion document to the Detailed Implementation Plan)

- Documents the "*diversity of biodiversity networks*" and their chief characteristics;
- Highlights how GEO BON can *leverage* the work of *existing networks and initiatives*;
- Proposes approach to *informatics design* based on a Service Oriented Architecture as described in the *GEOSS Common Infrastructure*.



Principles of the GEO BON Information Architecture (2)

Covers:

- 41 existing global, regional and national networks, discovery services and registries (GBIF, ILTER, KNB, NASA GCMD, NBII);
- 20 standards for metadata, data exchange and transfer protocols;
- Ontologies, vocabularies, dictionaries for semantic mediation;
- Biological names and habitat classifications;
- Workflow of services and integration of applications with an example of a climate change scenario using GBIF data;
- Portals, search engines, querying and harvesting including GBIF Data Portal, LTER/ILTER, NBII, KNB and NASA GCMD facilities;
- Open Access Issues and GEOSS Data Sharing Principles.



Key messages

- Highly complex, non-trivial challenges;
- 10 years experience to date, and much already accomplished, allowing GEO BON to 'jump start' on GBIF;
- Based on partnerships, common purpose and 'public good' philosophy;
- Early product developed quickly based on GBIF and partners expertise to date....
- Watch this space!!!



Thank you!

