Filling Biodiversity Knowledge Gaps: GBIF supporting research for conservation management

Arturo H. Ariño, presenting

Co-authors: Andrea Pino del Carpio[1], Rafael Miranda, Ana Villarroya[2], Jordi Puig

University of Navarra
Department of Environmental Biology

[1] Currently Spanish Society for Icththyology; [2] Currently The Nature Conservancy

Research supported in part by the Friends of the University of Navarra Association

Universidad de Navarra

GBIF

Biodiversity Knowledge Gaps

- Biodiversity *exists.*
- Knowledge about biodiversity is *gained.*
  - Absolute gaps remain between *absolute* knowledge and *achievable* knowledge.
  - Relative gaps remain between *achievable* knowlegde and *desirable* knowledge.
  - Practical gaps remain between *desirable* knowledge and *gained* knowledge

Ariño, A.H., et al., in prep.: Biodiversity Knowledge Gaps.

The Biodiversity Knowledge Gap Conjectures

1. The more knowledge about biodiversity is gained, the less likely *practical* gaps will be reduced by a new source of data.
2. Sources of biodiversity data will increasingly overlap as knowledge is gained.

Ariño, A.H., et al., in prep.: Biodiversity Knowledge Gaps.

Universidad de Navarra

GBIF

GAP FILLING

		Access/process difficulty		
		Easy	Medium	Hard
Capture or existence form	Digital	- Databases - Indexes - Digital inventories - CS output	- Unstructured files - Maps, digital RS - Survey results	- Locked files - Unknown files
	Analogical	- Imaged Reports, tables - Imaged museum data labels, structured ledgers	- Unscanned papers - Old imagery - Unmarked literature - Field notes	- Locked, private collections - Forgotten or unknown collections
	Future	- Automated surveys and monitoring - New CS endeavours	- New field surveys - Planned RS	- Unsampled remote or inaccessible sites - Unknown organisms

Figure XX: Access to, and forms of, biodiversity data. Black: DAK. Blue: LK. Red: BK. Green: Mineable data. Grey: Future data.

DAK – Digitally Accessible Knowledge: Primary data that are both digital and accessible in standard formats

LK – Locked knowledge: Data that are known to exist, but cannot be accessed because of some barrier (e.g. paywall, obsolete digital systems, inability to digitize)

BK – Buried knowledge: Data that exist but whose existence is not known or cannot be ascertained by users.

Ariño, A.H., et al., in review: GBIF Best Practice Guide for 'Data Gap Analysis for Biodiversity Stakeholders'. GBIF, Copenhagen.

From Conjectures to Hypotheses: Testing Approaches

- Top-down
  - Set the maximum achievable extent
  - Measure relative and practical gaps found in separate sources
    - *E.g. Hjarding et al., 2014*
- Bottom-up
  - Compile separate sources independently
  - Measure the extent of overlap
    - *E.g. Pino del Carpio et al., 2011, 2013, 2014*

Hjarding A., Tolley K.A., Burguess N.D., 2014: Red List assessments of East African chameleons: a case study of why we need experts. *Oryx*, doi:10.1017/S0030605313001427.

TABLE 1 Museums from which locality records for East African chameleons were obtained for the expert and GBIF datasets.

Museum	Dataset	
	Expert	GBIF
Afrika Museum, The Netherlands	x	
American Museum of Natural History, USA	x	
Bishop Museum, USA		x
British Museum of Natural History, UK	x	
Brussels Museum of Natural Sciences, Belgium	x	
California Academy of Sciences, USA		x
Ditsong Museum, South Africa	x	x
Los Angeles County Museum of Natural History, USA		x
Museum für Naturkunde, Germany	x	
Museum of Comparative Zoology (Harvard University), USA		x
Naturhistorisches Museum Wien, Austria	x	
Smithsonian Institution, USA		x
South African Museum, South Africa	x	
Trento Museum of Natural Sciences, Italy	x	
University of Dar es Salaam, Tanzania	x	
Zoological Research Museum Alexander Koenig, Germany	x	

TABLE 2 Data quality and data cleaning requirements for the expert and GBIF databases, with numbers of records of raw data, cleaned data, requiring taxonomic update, for *Chamaeleo* species, with no geographical coordinates, and with no locality listed. Total cleaned data were the records used in the analysis.

	Total raw	Total cleaned	Requiring taxonomic update	*Chamaeleo** sp.	No geographical coordinates	No locality listed
Expert	263	254	0	0	7	0
GBIF	2304	172	2302	67	478	833

*Many species in this genus have been reassigned to *Kinyongia* and *Trioceros*, but these changes are not reflected in the GBIF dataset.

A Word on "GBIF Data"

- It's data *mobilized* by GBIF
- Data are supplied by data owners and producers
- General perception of data quality issues
- Can these data be used? Always? For any purpose?
- Can we afford ignoring existing data?

Ariño, A.H., Chavan, V., Faith, D.P., 2013. Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges. Biodiversity Informatics, 8: 59-93.

[You do] acces primary biodiversity data [through:]

List the ways in which you use Primary Biodiversity Data

	Your own field works/surveys	Hardcopy, literature survey (non-digital form)	Primary Publications (e.g. taxonomic monographs, maps of species observations)	Other web based data portals	The GBIF Data Portal (http://data.gbif.org)	Access to offline digital data sets (CDROM/DVD/Tapes etc.)	Free and open datasets within and outside of your institution	Reciprocal agreements with other groups/individuals	Institutional agreements	Payment basis	Others	FTP sites
Species diversity and populations	370	335	320	172	157	126	120	104	70	20	22	8
Taxonomy	369	335	328	172	158	122	120	100	65	18	23	8
Life histories and phenologies	366	318	310	176	140	146	128	106	70	20	26	8
Biogeographic studies	337	310	309	162	150	117	109	101	63	14	20	6
Endangered, migratory and invasive Species	270	250	243	136	124	107	108	84	60	19	21	9
Ecology, Evolution and Genetics	269	235	229	124	108	87	83	72	49	16	16	5
Conservation Planning	194	183	173	98	87	86	86	66	49	14	9	5
Natural Resources Management	171	154	149	77	68	84	68	60	50	14	9	5
Education and Public Outreach	159	149	151	91	73	73	68	55	41	13	16	4
Impact of Climate Change	160	150	150	75	77	73	65	65	45	11	7	3
Environmental Impact Management	139	128	131	62	51	62	61	52	35	9	8	5
Sustainable Use	98	98	86	51	52	51	44	33	33	11	7	4
Agriculture, Fisheries, Forestry and Mining	98	89	89	42	42	42	41	33	24	10	1	2
Environmental regionalisation	77	76	76	40	40	39	36	34	24	4	3	3
Ecotourism	71	64	58	26	23	38	29	24	22	7	5	2
Bioprospecting	45	41	42	18	23	21	19	15	12	4	3	2
Forensics	40	32	30	16	15	11	17	11	8	4	1	
Recreation	30	30	23	16	12	18	16	7	10	3	1	
Border Control and Wildlife Trade	22	22	24	12	9	14	12	8	10	3	1	1
Health and Public Safety	24	24	19	11	9	11	10	10	7	5	3	1
Society and Politics	18	18	18	12	13	19	11	10	9	1	1	1
Human Infrastructure Planning	19	17	18	13	8	14	13	10	9	3		
Industrial Use	7	11	10	7	8	6	7	2	2	1		
Nursery and Pet Industry	12	11	11	7	1	4	5	2	3	1	2	1
Others (please specify)	3	5	2	6	2	1	3	2	1			

respondents	55%	49%	44%	38%	33%	27%	22%	16%	11%	5%	none

Universidad de Navarra

GBIF

Ariño, A.H., Chavan, V., Faith, D.P., 2013. Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges. Biodiversity Informatics, 8: 59-93.

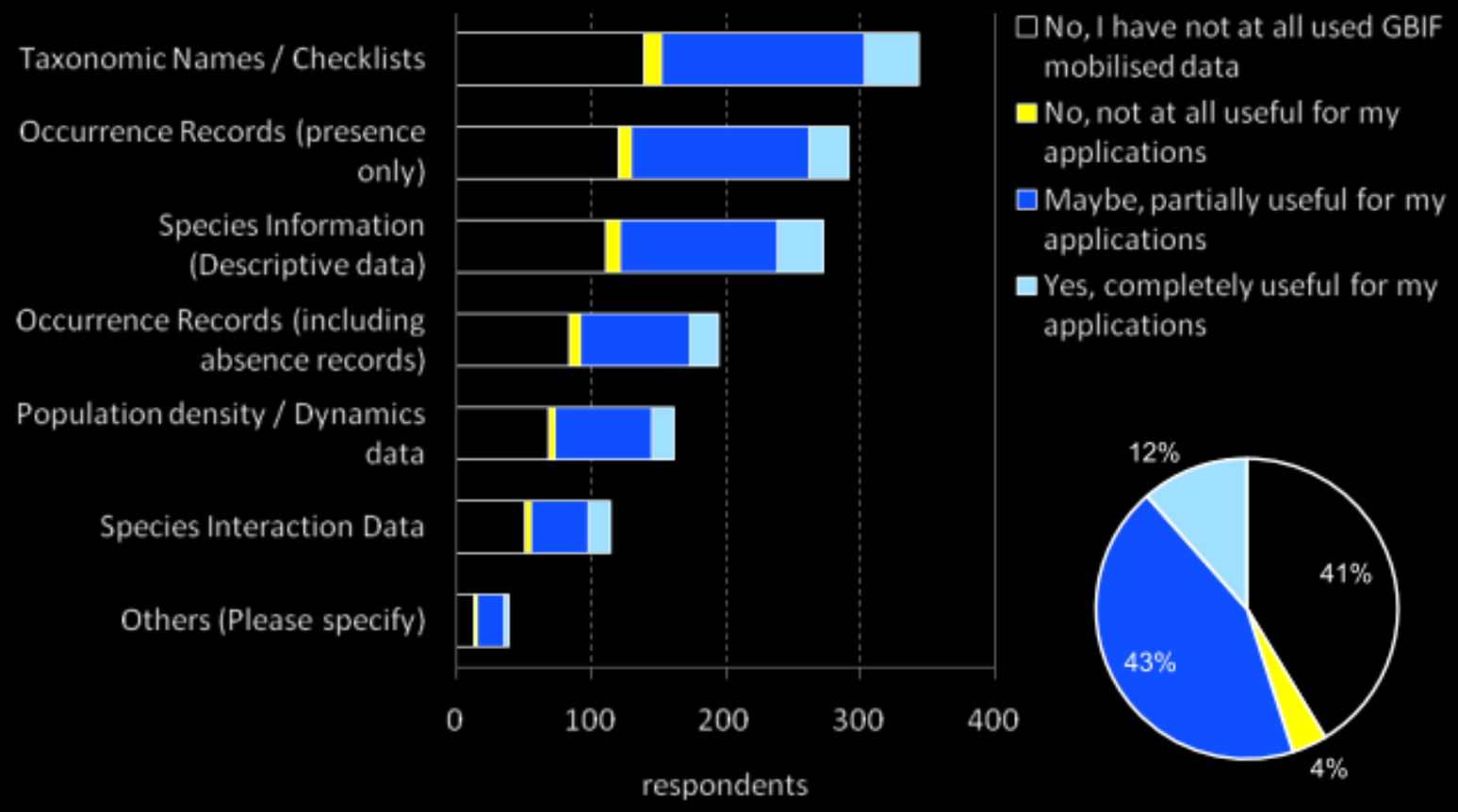
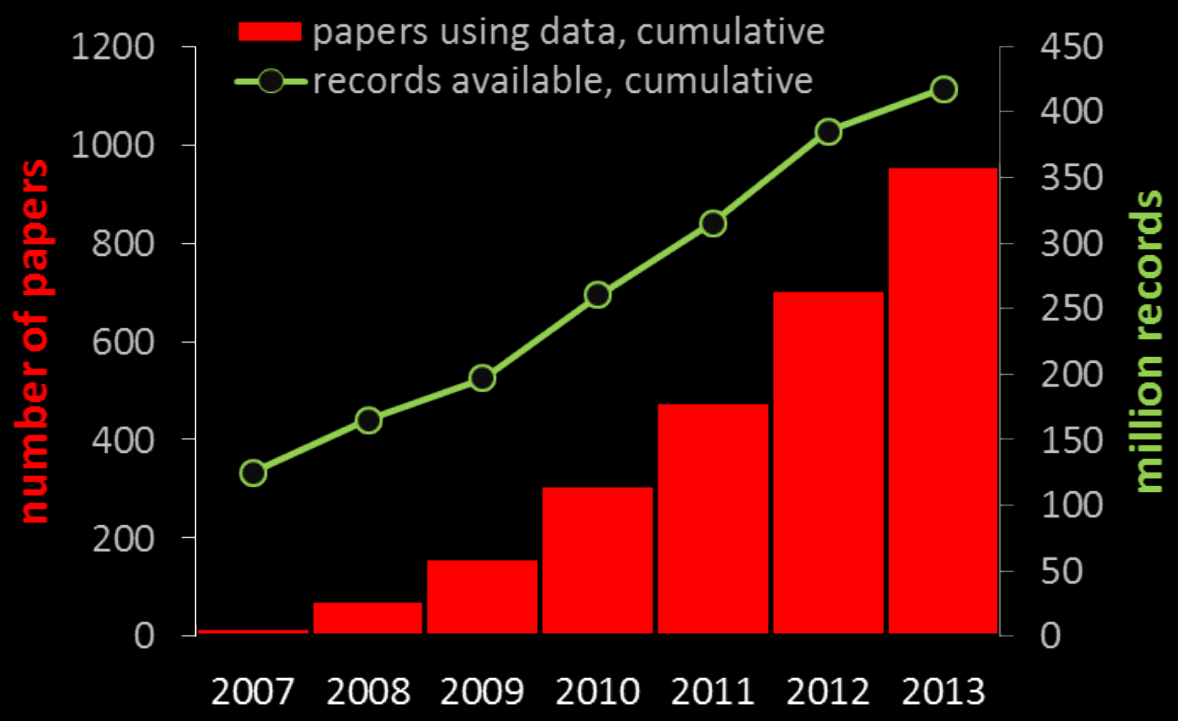
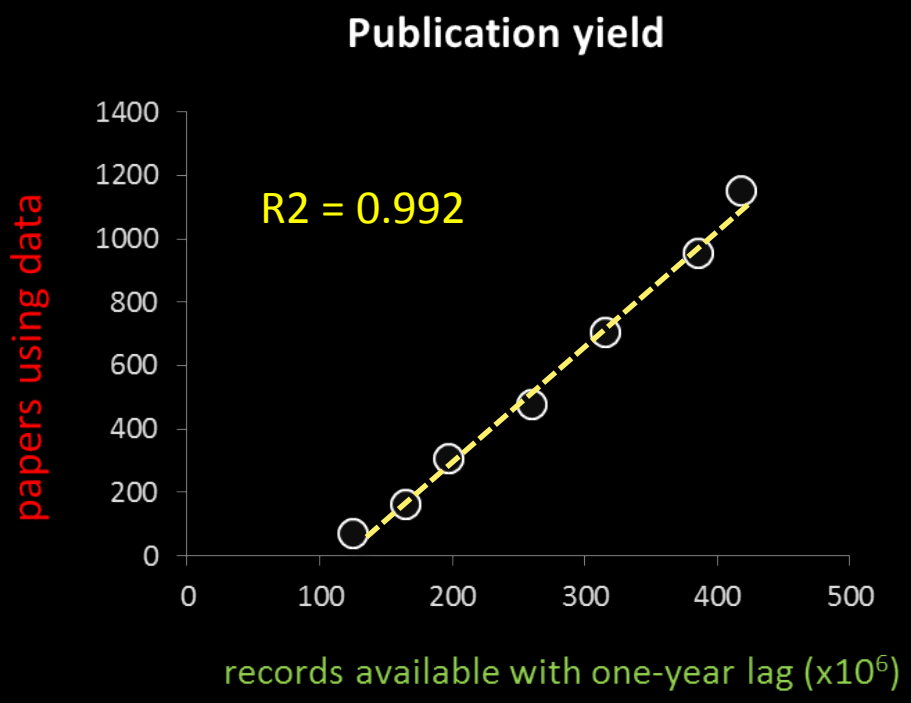


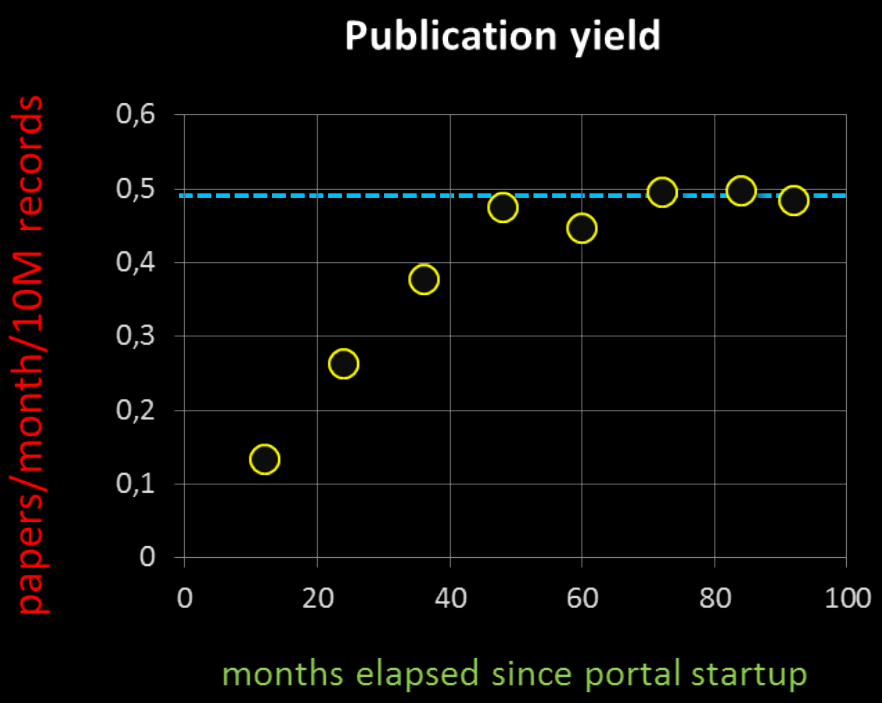
Figure 35: Percentages of people finding GBIF-mobilised data useful. Frequencies of responses to Q18: "Does GBIF mobilised data satisfy your needs?" and breakdown of frequencies according to types or nature of data required (Q10).

Data availability -> data use



Data use *rate* – stabilized?





Test Run: GBIF filling biological diversity knowledge gap in Biosphere Reserves

- A bottom-up approach to measure the price of ignorance
- Can conservation be informed by alternate knowledge sources?
- Do data alone suffice?

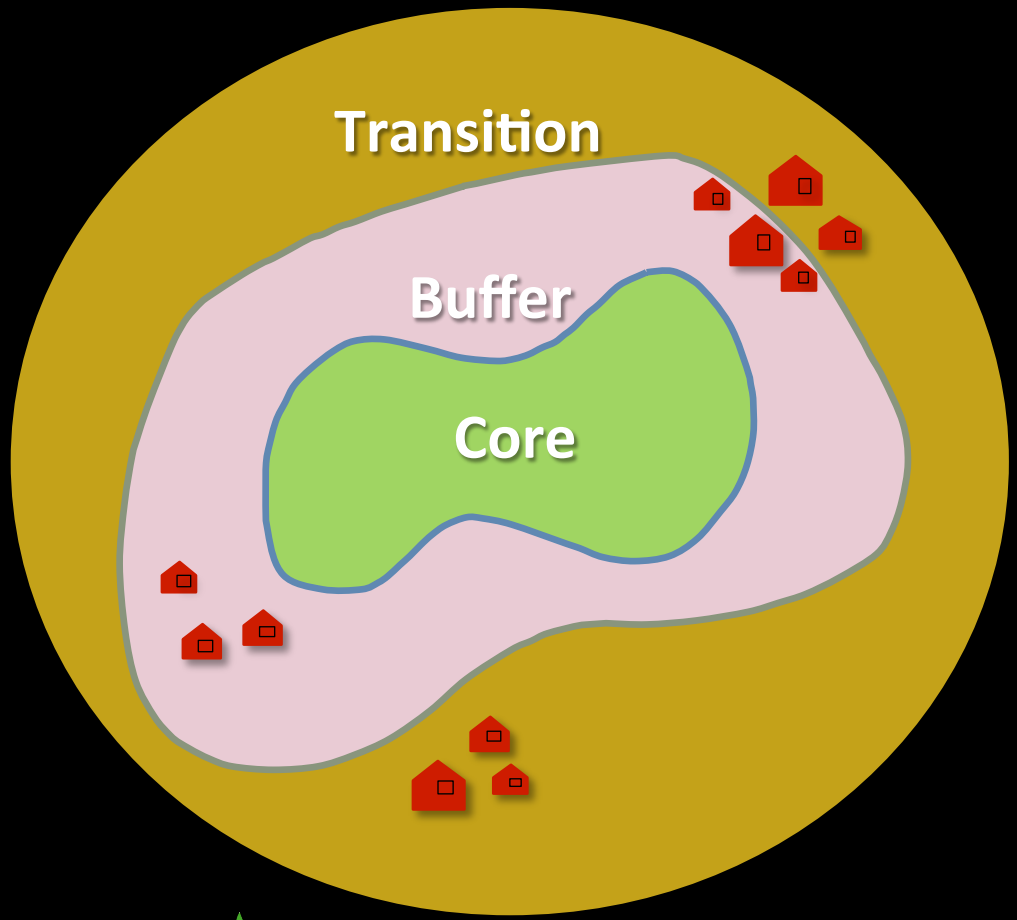
The Biosphere Reserves

- *"Sites established by countries and recognized under UNESCO's Man and the Biosphere (MAB) Programme to promote sustainable development based on local community efforts and sound science."* -- Man and Biosphere Programme
- **631 biosphere reserves in 119 countries all over the world.**
- tools to help countries implement the results of the World Summit on Sustainable Development and, in particular, the Convention on Biological Diversity and its Ecosystem Approach.

The Biosphere Reserves

Pino-del-Carpio A., 2013. The use of knowledge about biodiversity on the management of Biosphere Reserves. PhD Thesis by Andrea Pino del Carpio. Doctoral co-advisors: Rafael Miranda, Arturo H. Ariño

Area zones



TYPICAL ACTIVITIES

Core:

- Biodiversity conservation
- Monitoring
- Non-destructive research

Buffer:

- Cooperating activities
- Ecotourism
- Research

Transition:

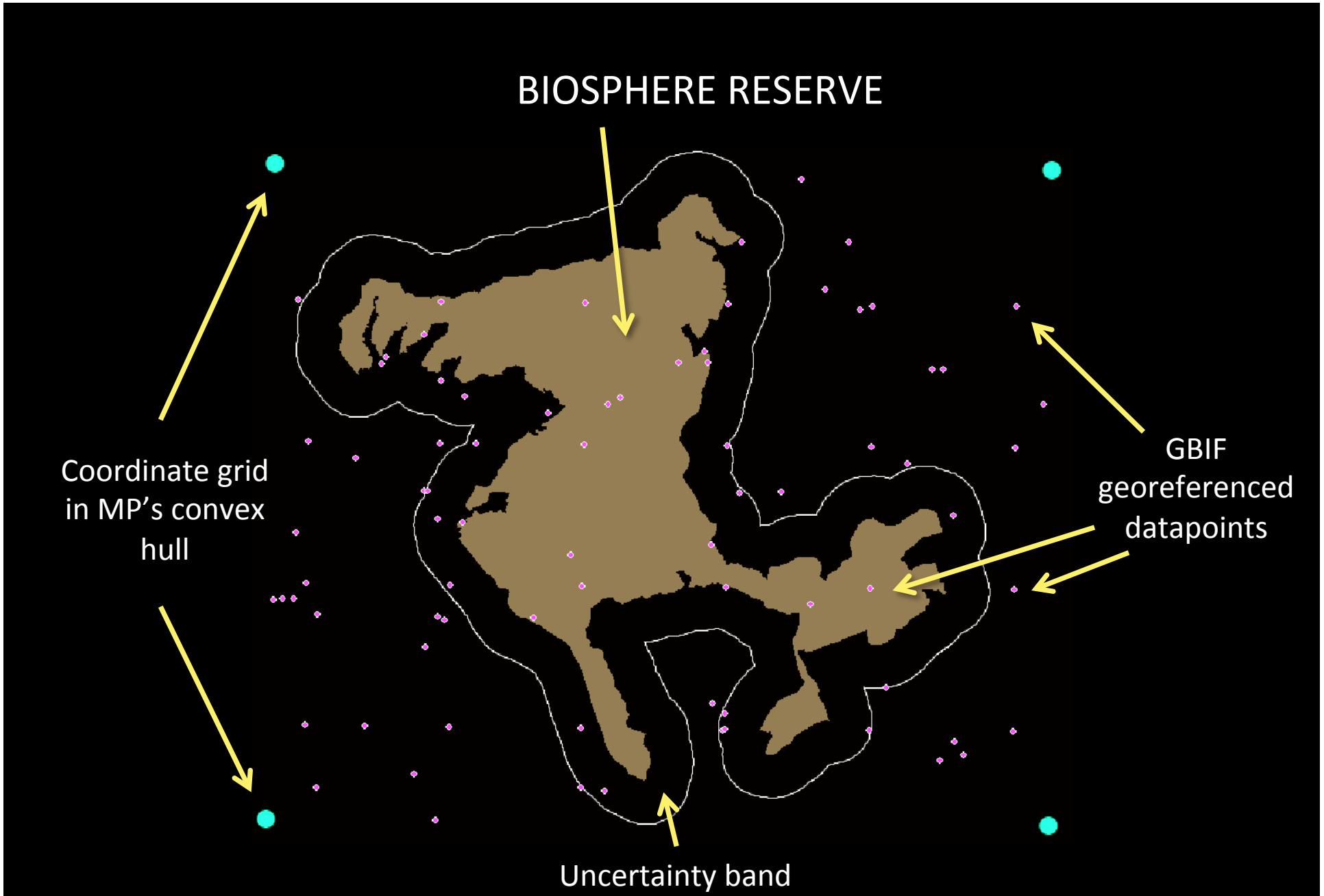
- Sustainable development
- Sustainable agriculture, settlements
- Community services

Universidad de Navarra

GBIF

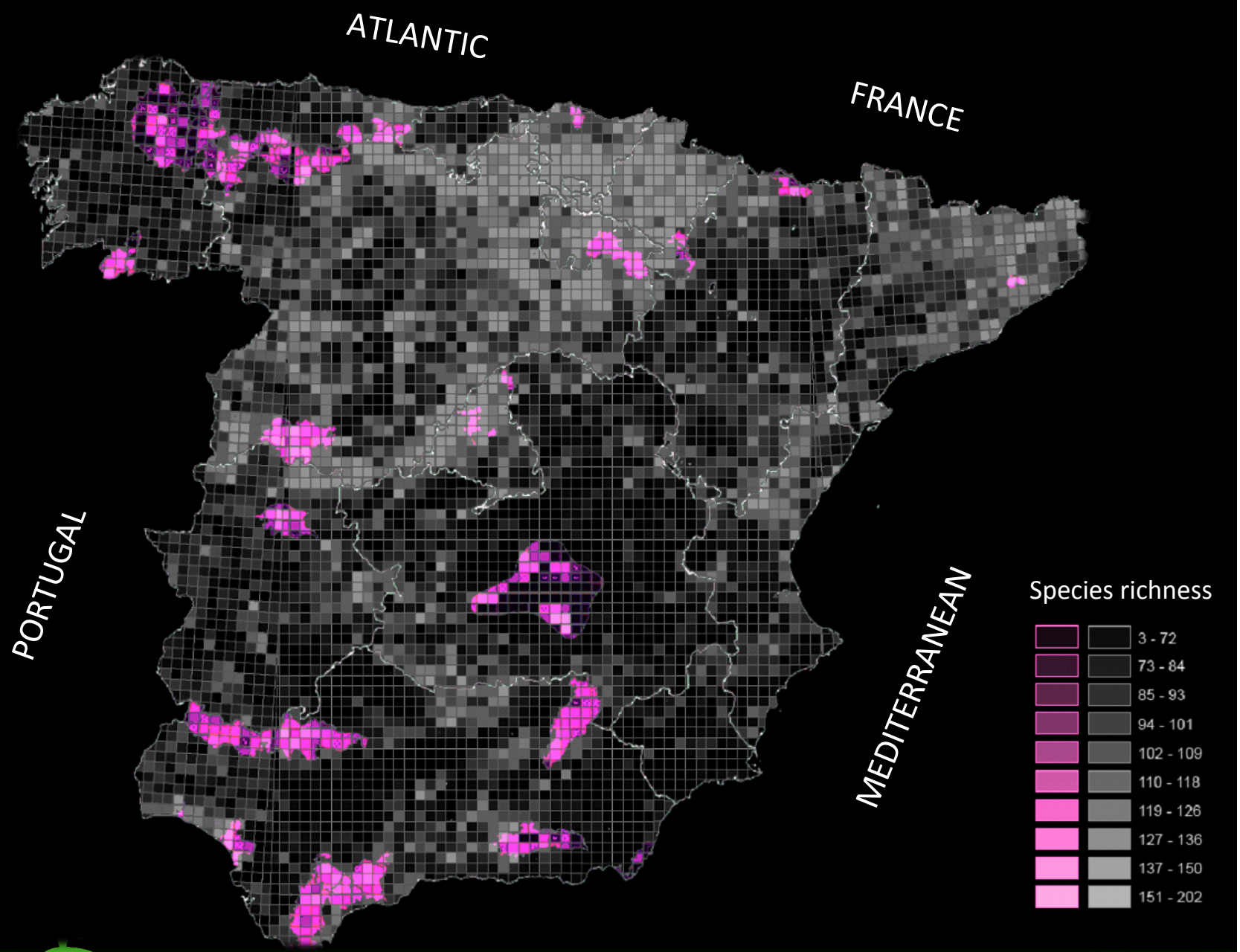
General Methods

1. Select Biosphere Reserves Networks
2. Extract information from sources:
   1. Management Plans and nomination proposals (MP)
   2. Scientific literature including books, redlists (LIT)
   3. Mutiple datasets made available through GBIF
3. Compile lists
4. Review taxonomy
5. Harmonize taxonomies using expert input
6. Tabulate lists
7. Compute intersections between lists
8. Calculate gaps


BIOSPHERE RESERVE
Coordinate grid in MP's convex hull
GBIF georeferenced datapoints
Uncertainty band

MEXICO : BIOSPHERE RESERVES

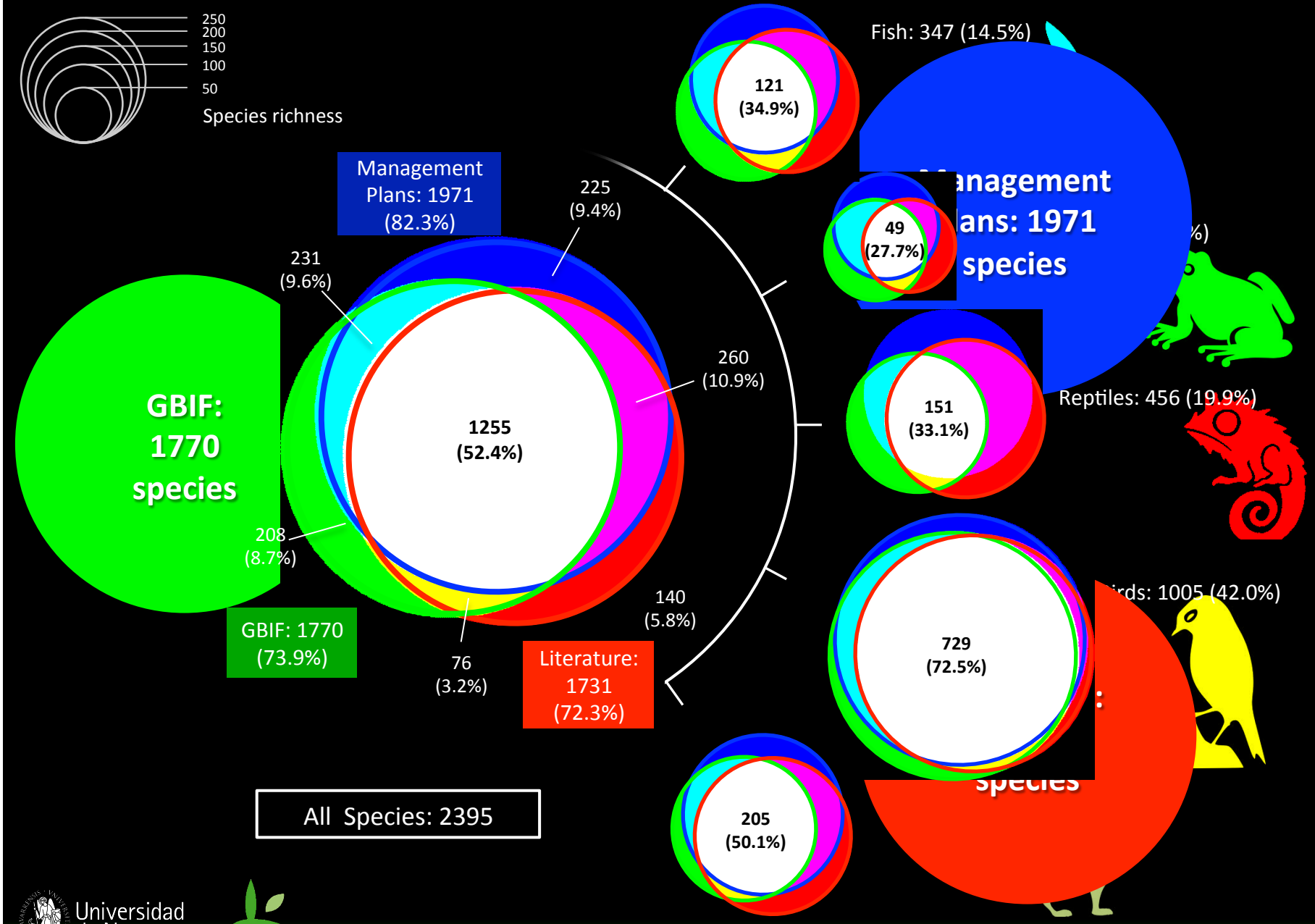
SPAIN : BIOSPHERE RESERVES



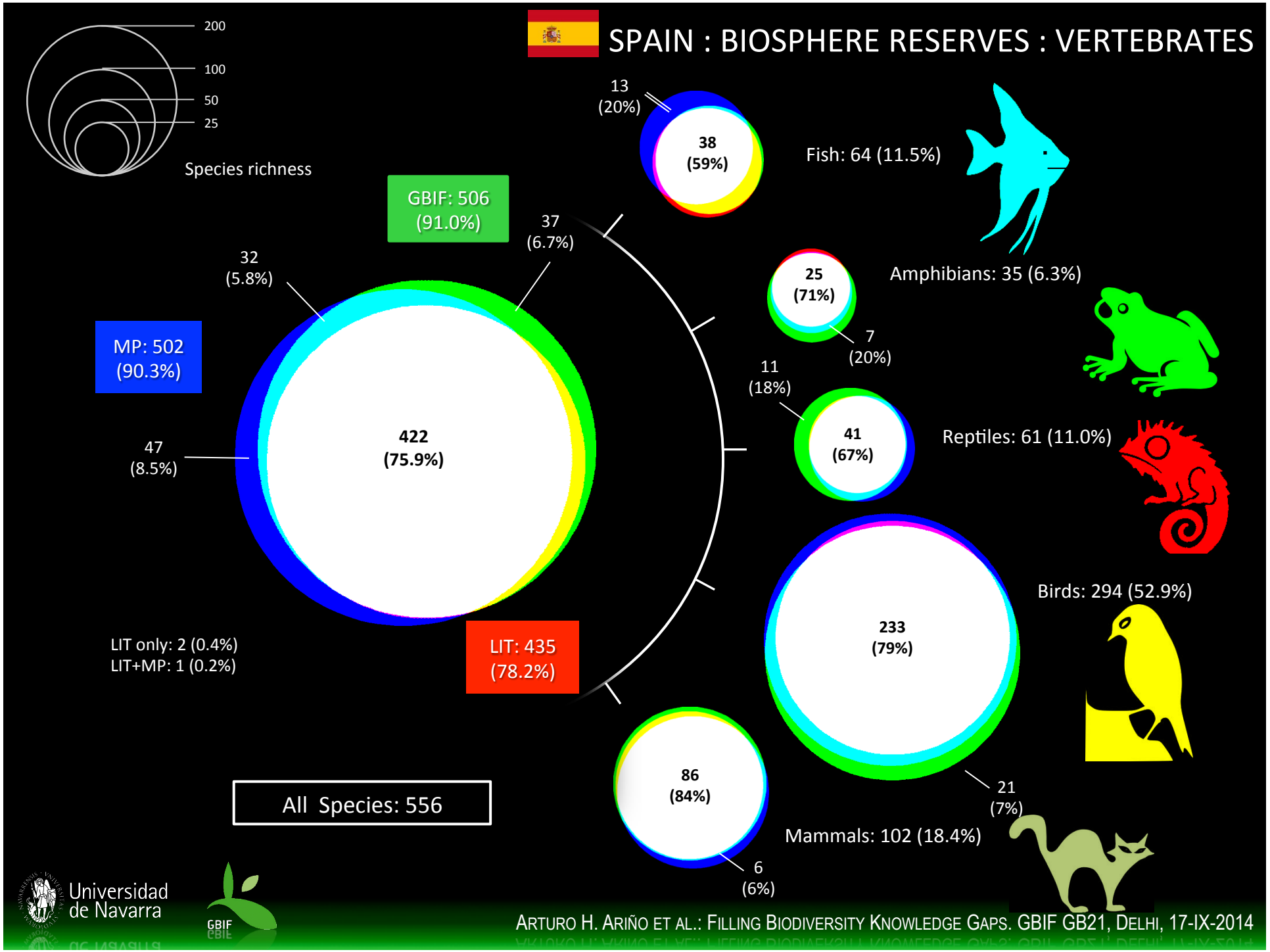
Universidad de Navarra

GBIF

MEXICO : BIOSPHERE RESERVES : VERTEBRATES



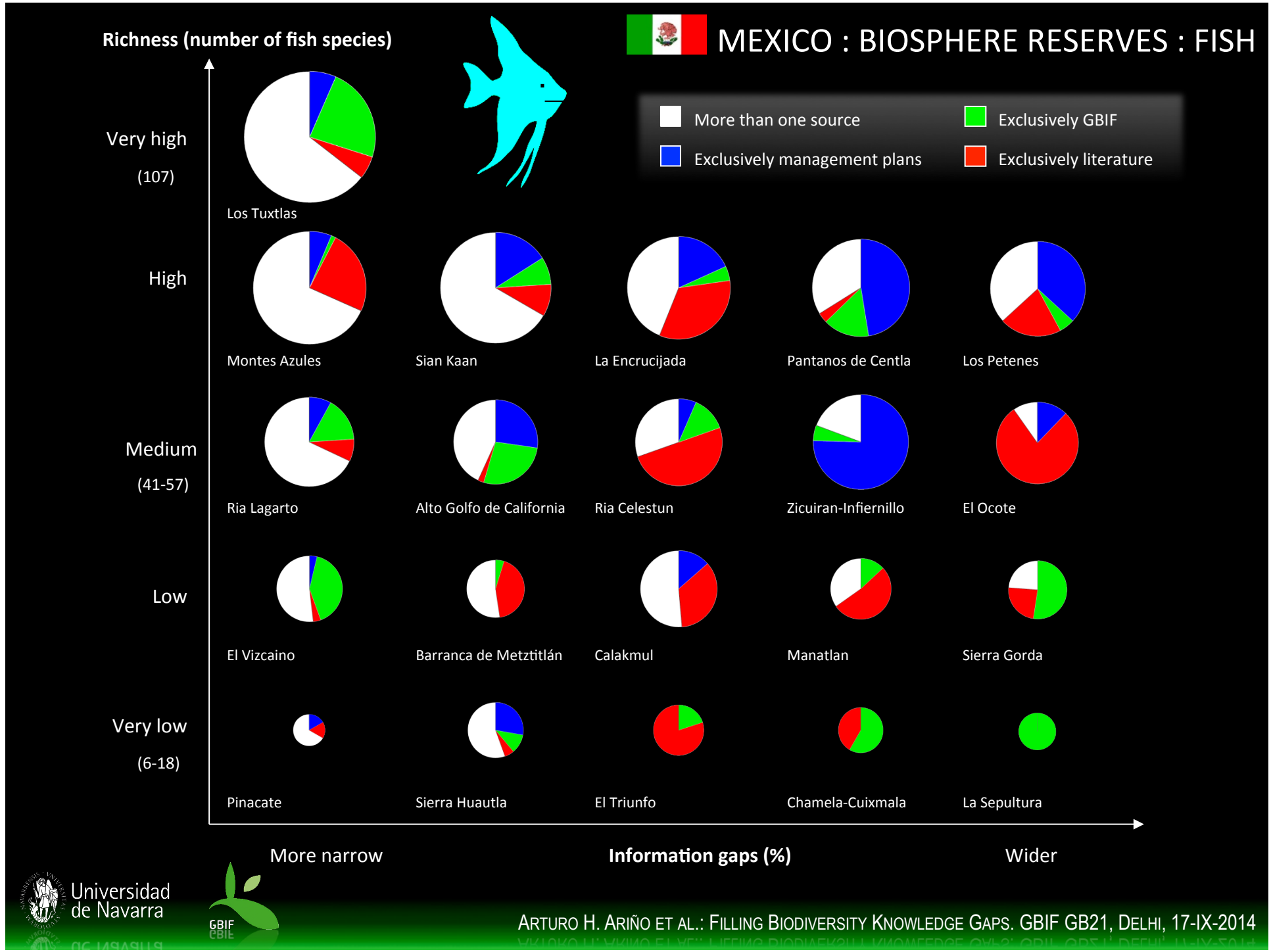
SPAIN : BIOSPHERE RESERVES : VERTEBRATES



Universidad de Navarra

GBIF

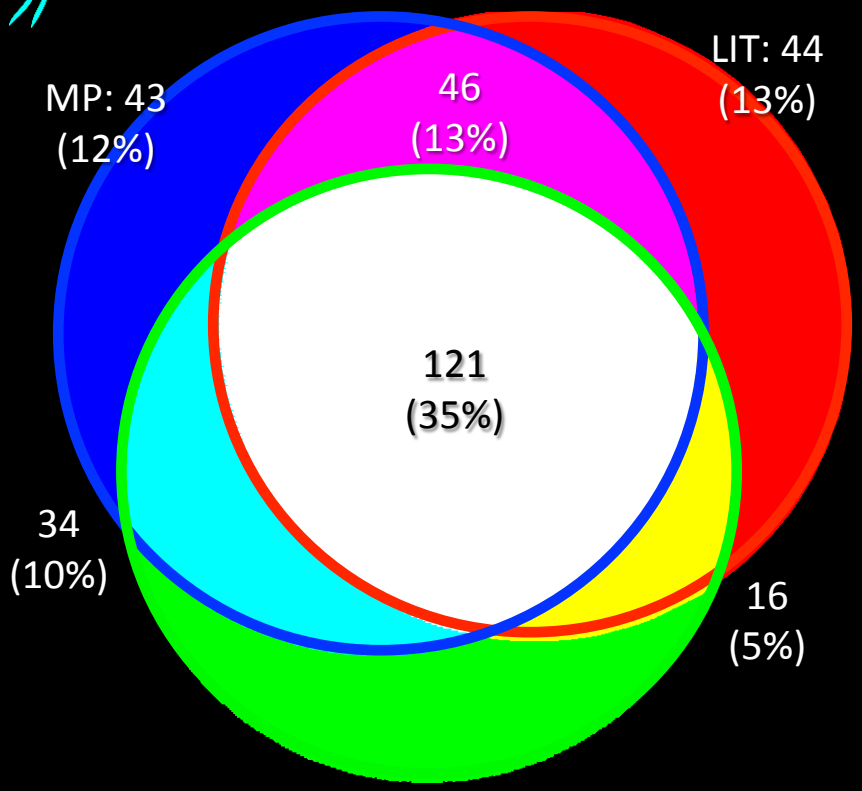
MEXICO : BIOSPHERE RESERVES : FISH



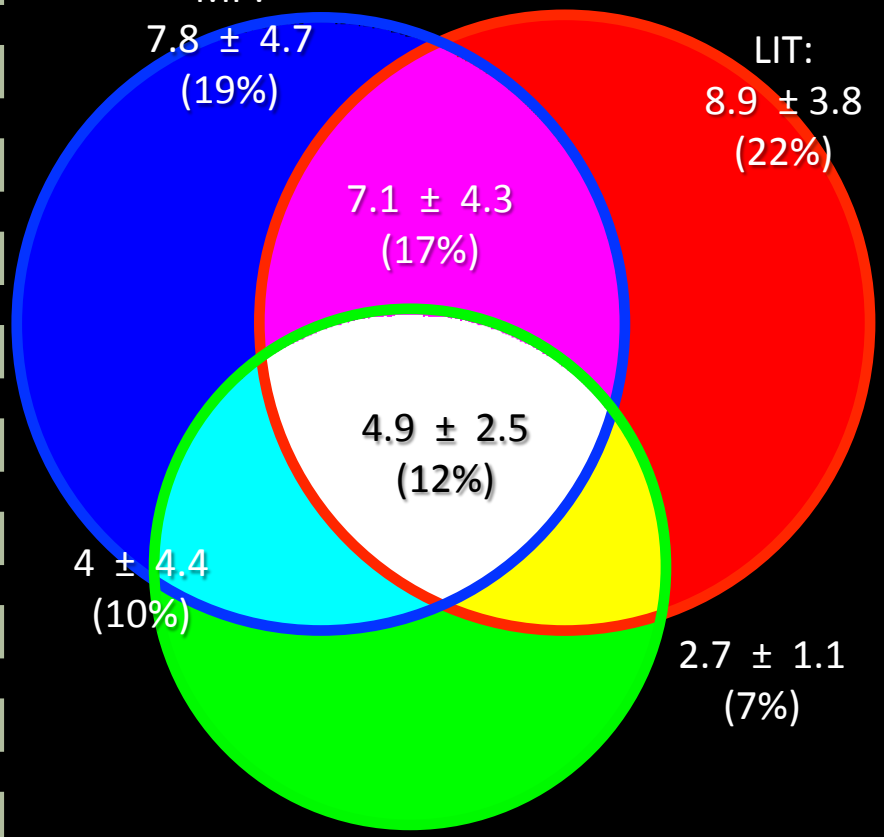
Universidad de Navarra

GBIF

MEXICO : BIOSPHERE RESERVES : FISH

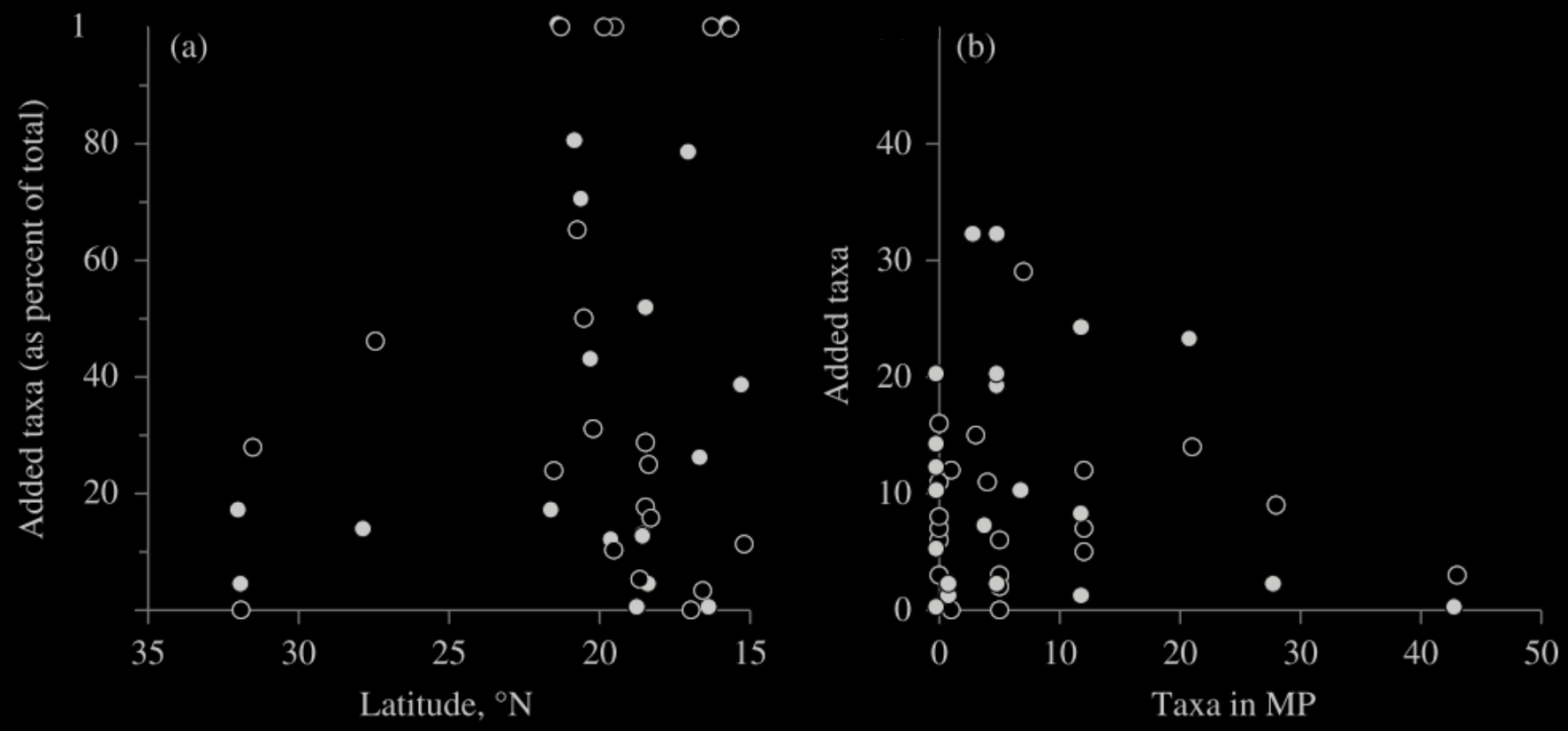


ALL RESERVES COMBINED



AVERAGE PER RESERVE

MEXICO : BIOSPHERE RESERVES : FISH



Taxa added to MP by GBIF

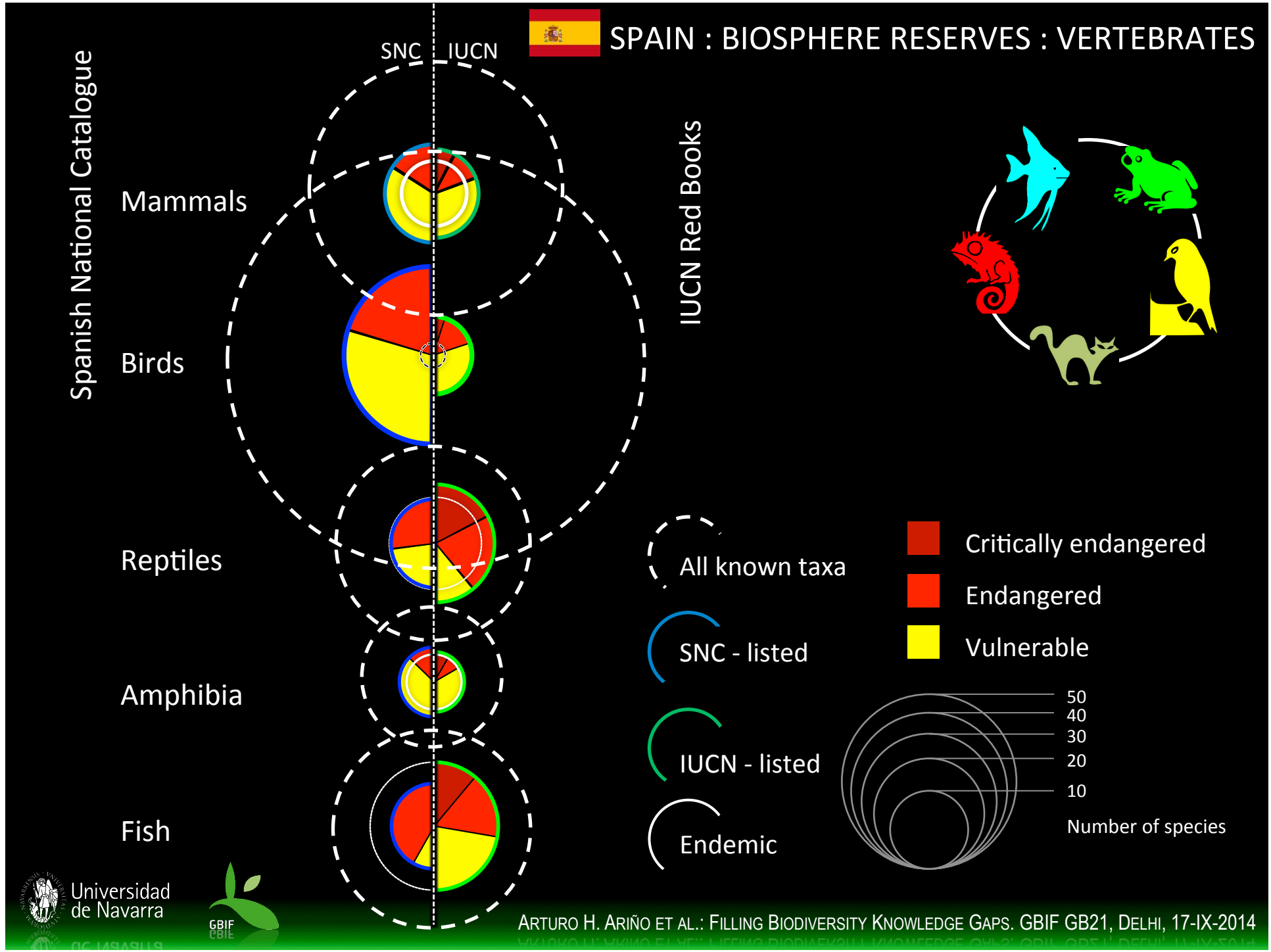
Taxa added to MP by literature

Pino-del-Carpio A., Villarroya A., Ariño A.H., Puig J., Miranda R., 2011. Communication gaps of knowledge of freshwater fish biodiversity: implications for the management and conservation of Mexican Biosphere Reserves. Journal of Fish Biology, 79(6): 1563-1591.

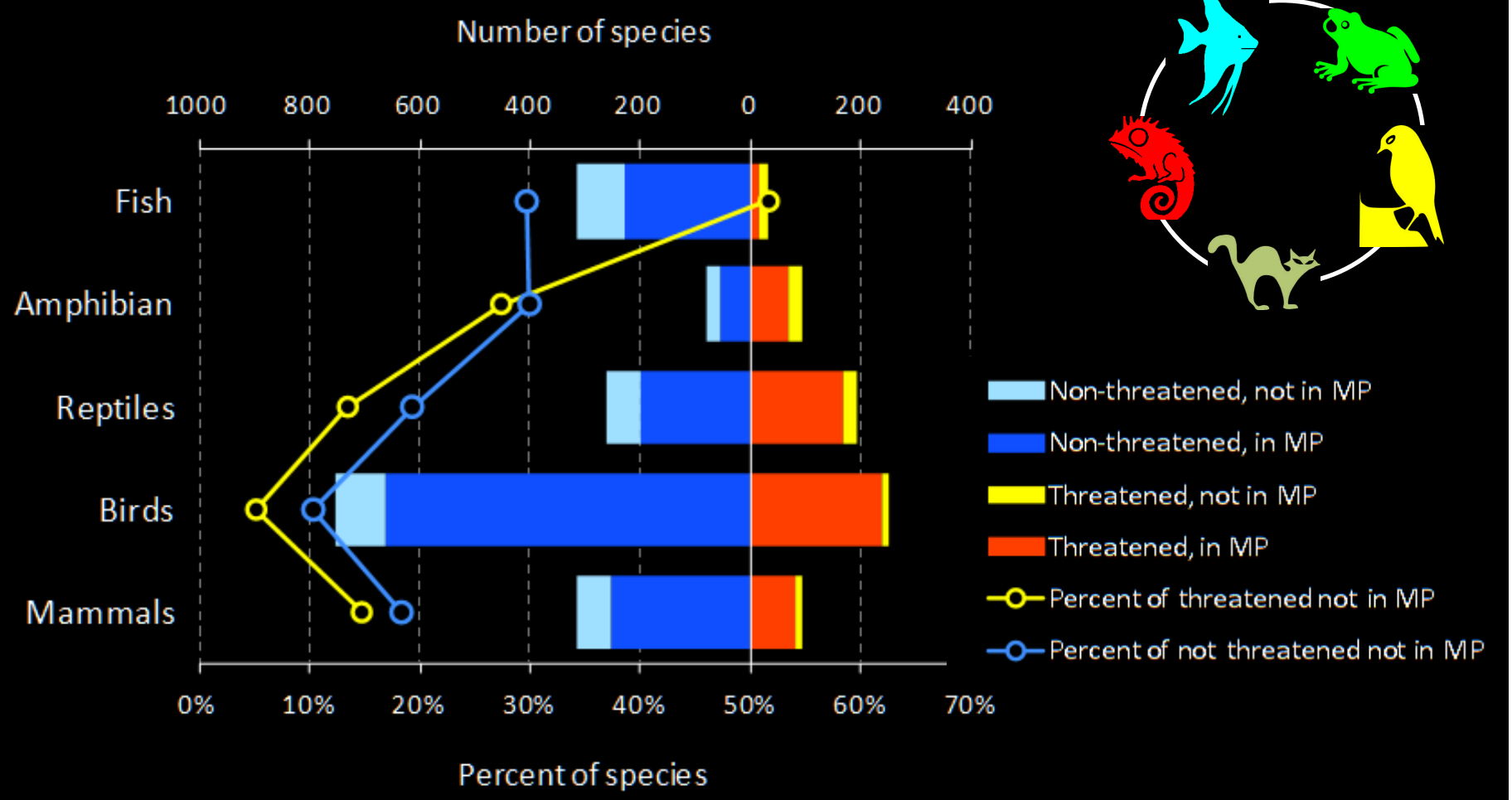
Universidad de Navarra

GBIF

SPAIN : BIOSPHERE RESERVES : VERTEBRATES



MEXICO : BIOSPHERE RESERVES : VERTEBRATES



Ariño, A.H., Chavan, V., Faith, D.P., 2013. Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges. Biodiversity Informatics, 8: 59-93.

IS GBIF USEFUL FOR CONSERVATION?

- Yes it is. It provides a practical source of information and can help filling gaps.
- GBIF conveys data from data publishers. Information not shared does not inform anything.
- Knowledge depends on information existing and exposed, whatever the means. GBIF is one practical means to expose existing information.
- Conservation depends on knowledge. Less knowledge -> success by accident. More knowledge -> failure by accident.
- Reliability and fitness-for-use of GBIF-enabled data remains the main perceived issue for conservation purposes. But lack of knowledge can only be alleviated by sharing data.
- Cross-checking sources enables correction and curation. Reliance on single sources is dangerous.
- The only "thing" that can substitute an expert is another expert. But experts have to rely on information sources too.

Works covered or cited

- PINO-DEL-CARPIO A., 2013. The use of knowledge about biodiversity on the management of Biosphere Reserves. PhD Thesis by Andrea Pino del Carpio. Doctoral co-advisors: Rafael Miranda, Arturo H. Ariño
- PINO-DEL-CARPIO A., **ARIÑO A.H**., MIRANDA R., 2014. Data exchange gaps in knowledge of biodiversity: implications for the management of Biosphere Reserves. *Biodiversity & Conservation*, DOI: 10.1007/s10531-014-0718-2.
- PINO-DEL-CARPIO A., VILLARROYA A., **ARIÑO A.H**., PUIG J., MIRANDA R., 2014. The biodiversity data knowledge gap: Assessing information loss in the management of Biosphere Reserves. *Biological Conservation,* 173: 74-79. doi: 10.1016/j.biocon. 2013.11.020
- FAITH, D.P., COLLEN, B., **ARIÑO, A.H.,** KOLEFF, P., GUINOTTE, J., KERR, J., CHAVAN, V., 2013. Bridging the biodiversity data gaps: Recommendations to meet users' data needs. *Biodiversity Informatics*, 8: 41-58.
- **ARIÑO, A.H**., CHAVAN, V., FAITH, D.P., 2013. Assessment of user needs of primary biodiversity data: Analysis, Concerns, and Challenges. *Biodiversity Informatics*, 8: 59-93.
- GAIJI, S., CHAVAN, V., **ARIÑO, A.H.**, OTEGUI, J., HOBERN D., SOOD R., ROBLES, E., 2013. Content assessment of the primary biodiversity data published through GBIF network: Status, Challenges and Potentials. *Biodiversity Informatics*, 8: 94-172
- **ARIÑO, A.H.,** OTEGUI, J., VILLARROYA, A., PÉREZ DE ZABALZA, A.I., 2012. Primary biodiversity data records in the Pyrenees. *Environmental Engineering and Management Journal*, 11(6): 1059-1075.
- OTEGUI, J., VILLARROYA, A., **ARIÑO, A.H**., 2012. Protected areas in the Spanish Pyrenees: A meaningful way to preserve biodiversity? *Environmental Engineering and Management Journal*, 11(6): 1133-1140.
- PINO-DEL-CARPIO A., VILLARROYA A., **ARIÑO A.H.,** PUIG J., MIRANDA R., 2011. Communication gaps of knowledge of freshwater fish biodiversity: implications for the management and conservation of Mexican Biosphere Reserves. *Journal of Fish Biology,* 79(6): 1563-1591.
- **ARIÑO, A.H**., CHI-JEN KO, B., 2014. Free-Text Answers to the CNA-TG Survey. Annex to ARIÑO ET AL., 2013, *Biodiversity Informatics*, 8(2): 59-93, dataset available: https://www.researchgate.net/publication/264346556_Free-text_answers_to_GBIFS_CNA-TG_Survey?ev=prf_pub
- Hjarding A., Tolley K.A., Burguess N.D., 2014: Red List assessments of East African chameleons: a case study of why we need experts. Oryx, doi:10.1017/S0030605313001427.


LA BIODIVERSIDAD DE MÉXICO
DESARROLLO SUSTENTABLE
PROFEPA

Thank You