

# ¿Crecen los datos en los árboles?

~~Dónde encontrarlos y cómo usarlos~~  
Como entenderlos y usarlos

Francisco Pando  
GBIF - España



CONABIO



# Sumario

- Como eran las cosas al principio
- Portales agregadores, GBIF
- Fuentes de datos por tipos
- Como leer esos datos
- El juego de los científicos
- Visibilidad, relevancia, sostenibilidad
- Dos reflexiones complementarias
- Conclusiones

# I. The Cactus Moth *Cactoblastis cactorum*. Pre GBIF

- Devours every single species of prickley-pear that has been tried.
- In the US and Mexico there are more than 90 species of *Platyopuntia*, many endangered, vital componente of arid ecosystems.
- In Mexico, *Opuntia* is the 10<sup>th</sup> product of agricultural importance

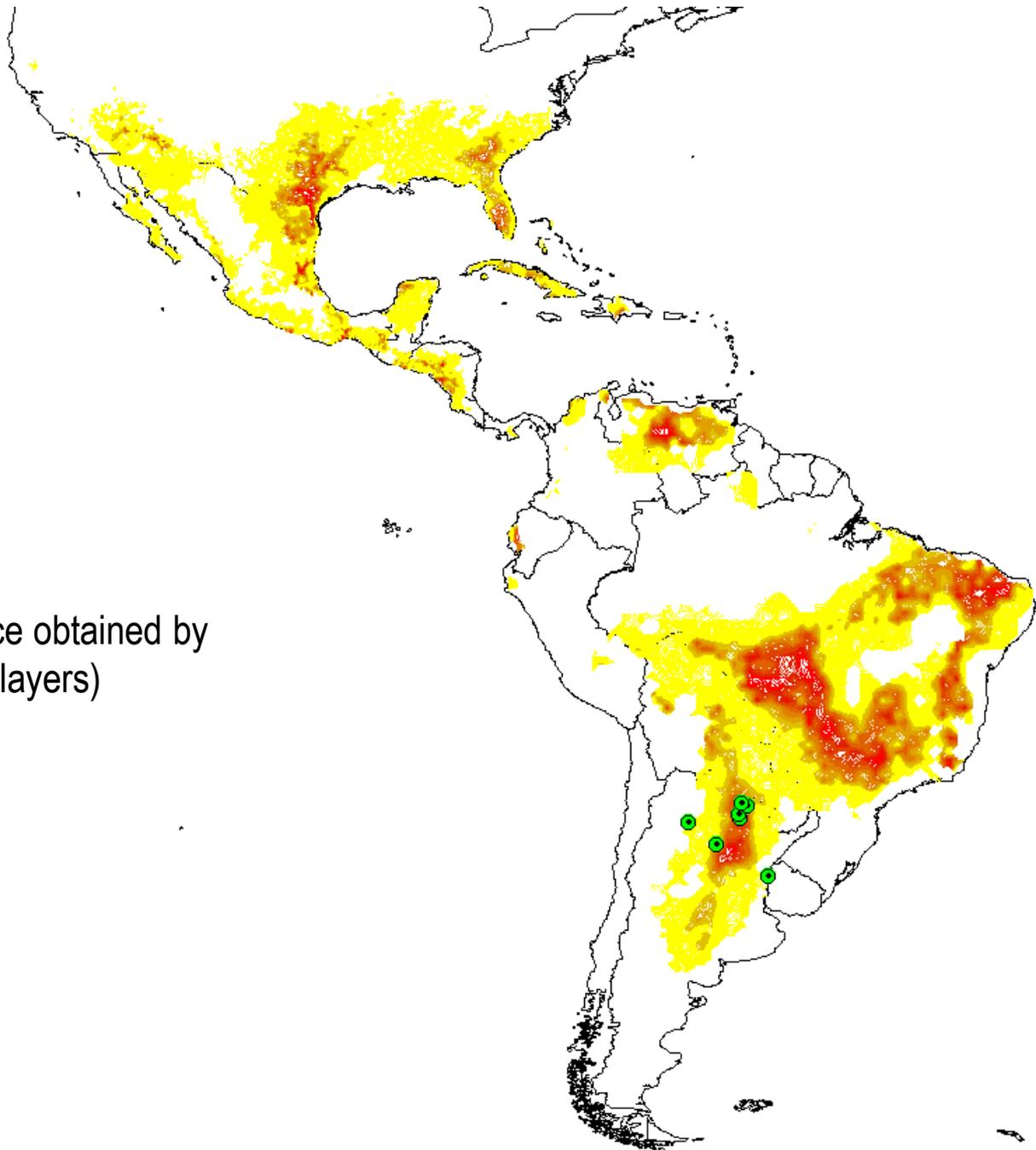
First problem: estimation of routes of invasibility



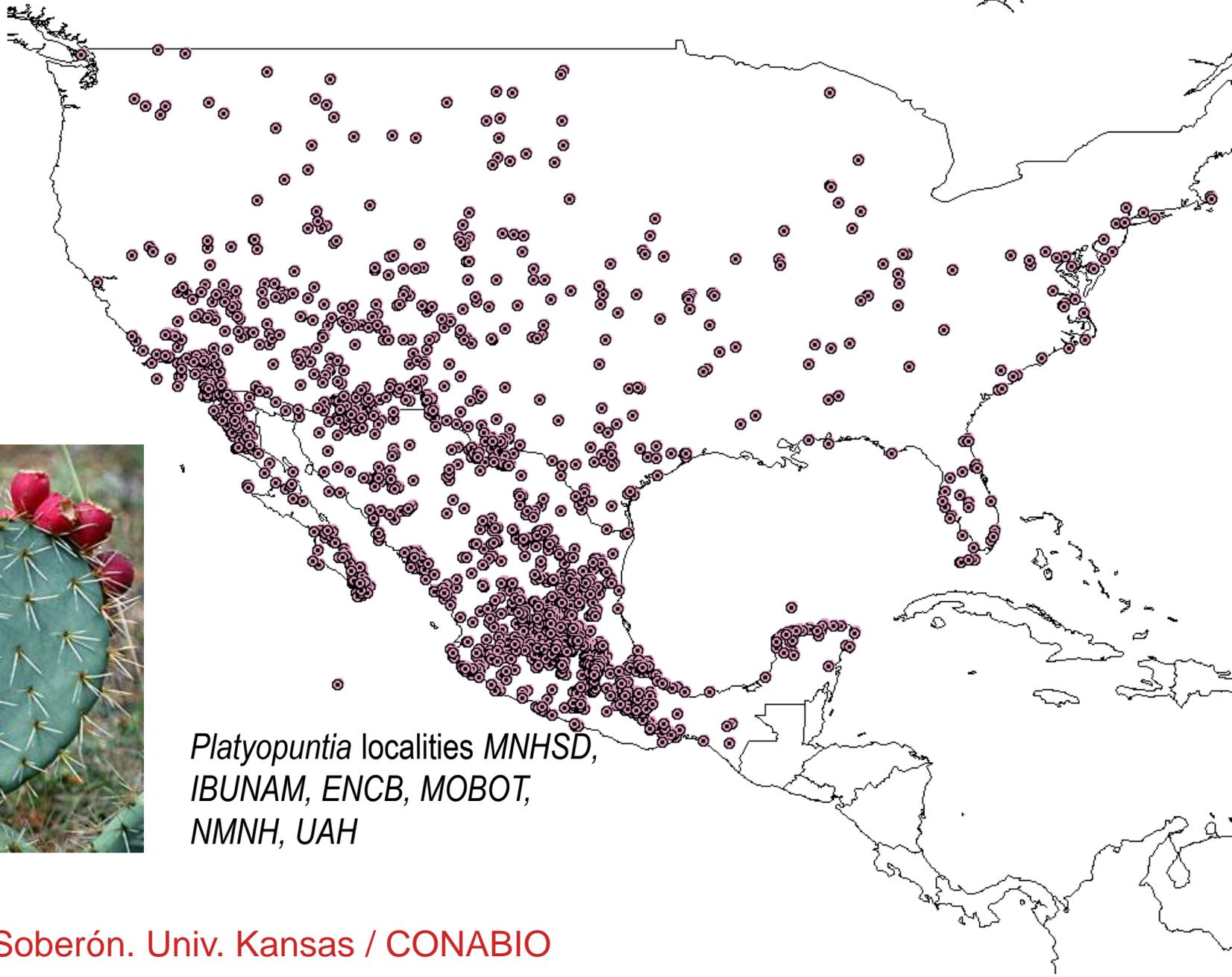
*Cactoblastis cactorum*

Data points obtained from  
the NMNH, USA

De Jorge Soberón. Univ. Kansas / CONABIO

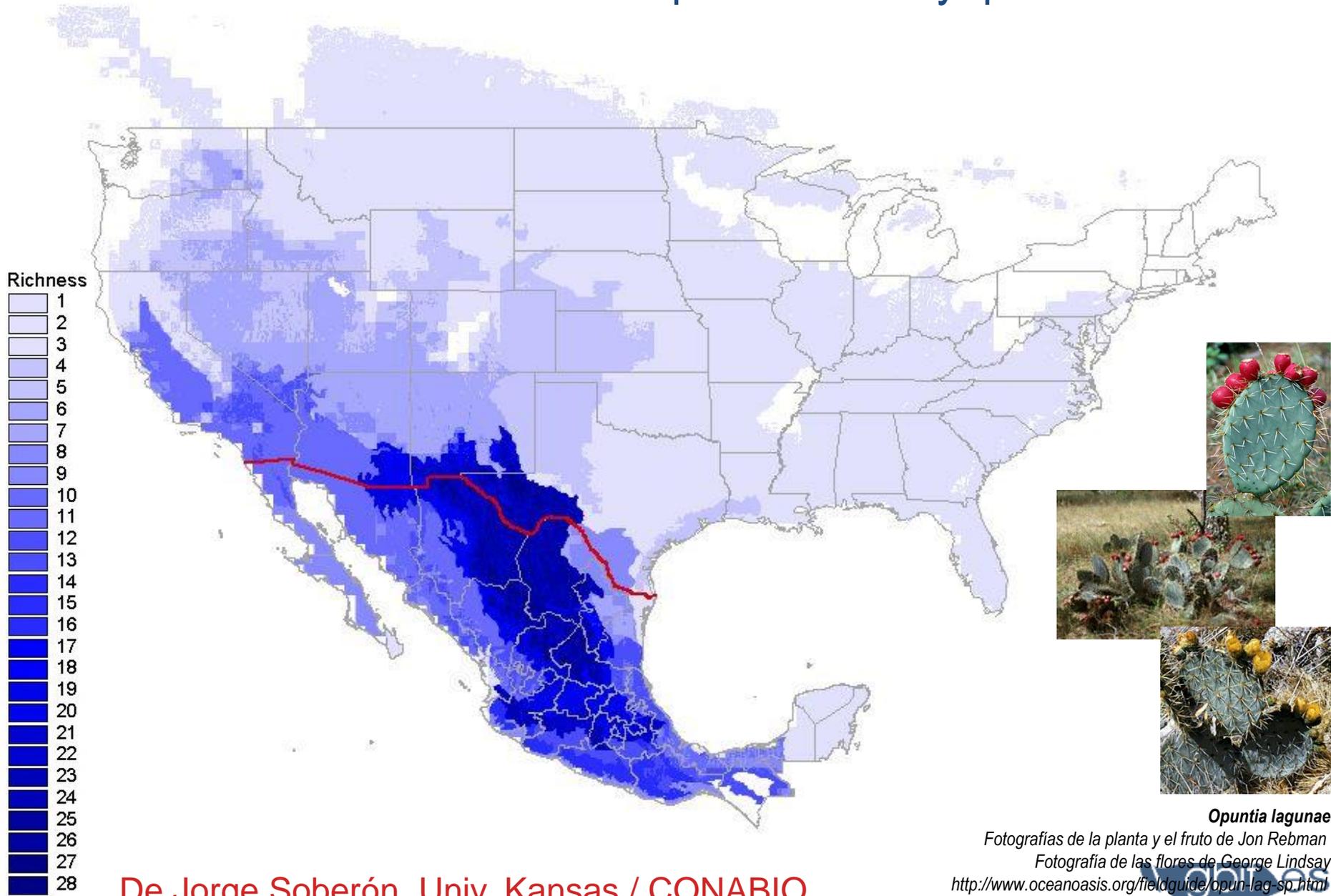


Climate surface obtained by Floramap (12 layers)



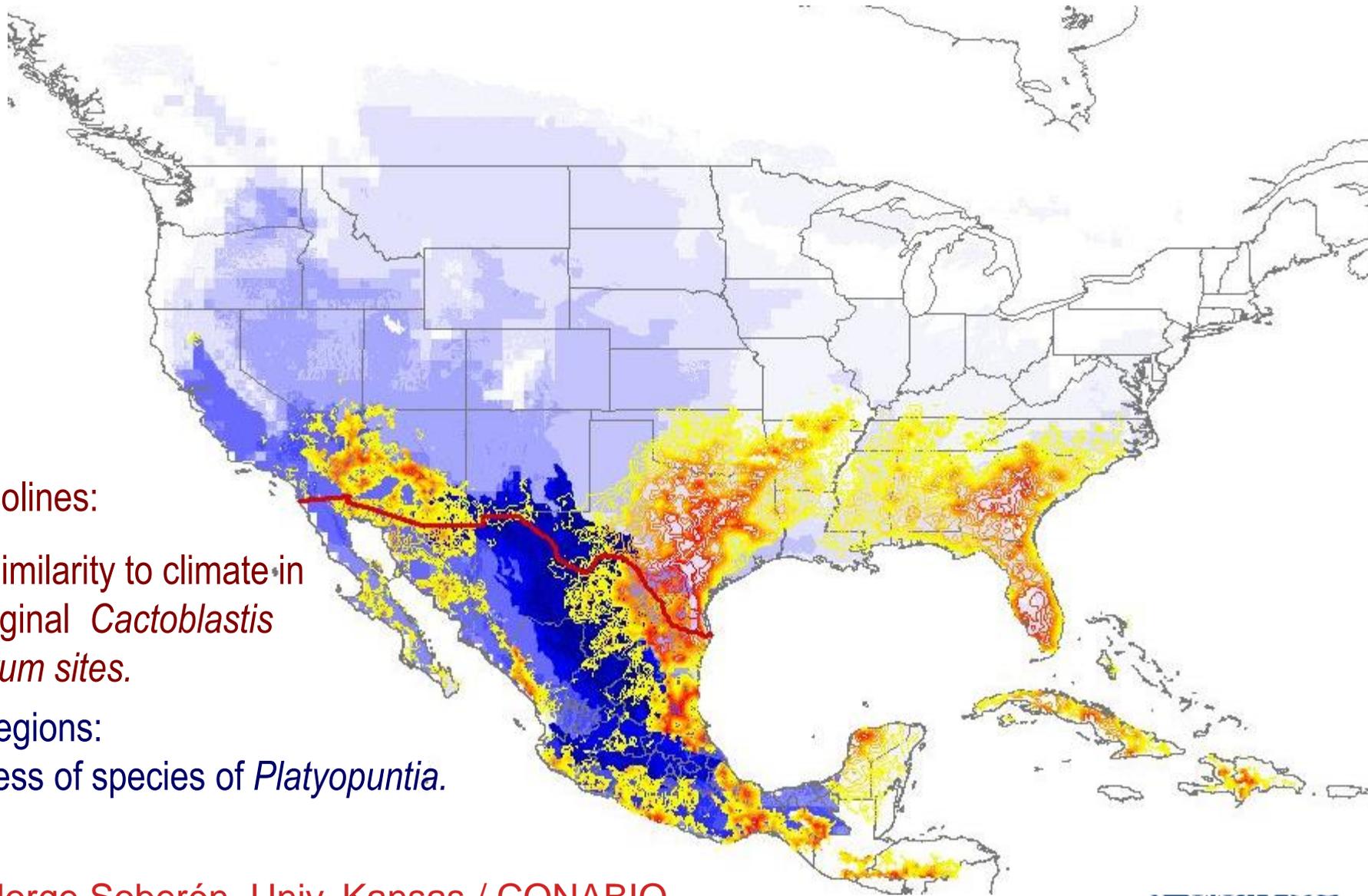
*Platyopuntia* localities MNHSD,  
IBUNAM, ENCB, MOBOT,  
NMNH, UAH

# Predicted number of species of *Platyopuntia*



De Jorge Soberón. Univ. Kansas / CONABIO

## Vulnerable areas to *Cactoblastis* (right climate and right food)



Red isolines:

High similarity to climate in  
the original *Cactoblastis*  
*cactorum* sites.

Blue regions:

Richness of species of *Platyopuntia*.

# Costs and Benefits of the Analysis

- The *C. cactorum* analysis took more than one year. Visits were made to USA. Australia took almost four years to provide their data. San Diego, Missouri and Smithsonian queried their own databases quickly. Arizona was not computerized.
- The results were adopted by Mexico's Minister of Agriculture. Base for colaboration with the USA

# Portales agregadores de datos

- Acceso a los datos en un formato que permite su análisis
- Los pioneros
  - Species Analyst
  - REMIB (CONABIO)

**Red Mundial de Información en Biodiversidad REMIB**

[http://www.conabio.gob.mx/remib/doctos/remib\\_esp.html](http://www.conabio.gob.mx/remib/doctos/remib_esp.html)

- Los que siguieron: OBIS, GBIF, ...

La mayor red de datos de biodiversidad del mundo

## Global Biodiversity Information Facility

Free and open access to biodiversity data

441,819,230 OCCURRENCES	1,454,695 SPECIES	15,094 DATASETS
----------------------------	----------------------	--------------------

Map interface with zoom controls (+, -, Home) and a 1px scale indicator. A timeline at the bottom allows filtering data by date.

### Taxonomic characteristics

The following provides a summary of number of records per kingdom. Further filters, such as a location or temporal filter, applied will [exploring the data](#).

286,515,884 (66.8861%) Animalia records	121,523,887 (28.2893%) Plantae records	8,038,407 (1.8765%) Fungi records	4,332,760 (1.0115%) Protozoa records
2,107,154 (0.4919%) Chromista records	1,007,975 (0.2353%) Bacteria records	840,131 (0.1961%) Other records	14,228 (0.0033%) Archaea records
7,796 (0.0018%) Viruses records			

### Record type characteristics

Records may originate from a variety of means, such as a scientist collecting a specimen or an individual recording the occurrence of an organism. This is classified by the [Darwin Core basIORecord](#) standard.

303,738,491 (70.907%) Observation records	89,465,185 (20.885%) Specimen records	31,381,460 (7.305%) Unknown evidence records	2,615,217 (0.611%) Fossil records
762,123 (0.178%) Using Specimen records	401,613 (0.094%) Literal Occurrence records		

### Temporal characteristics



# Desde la ciencia

- muy formal
- muy fiable
- estandarizado
- muy integrable( $\pm$ )
- accesible,  $\pm$  disponible
- problemas de representatividad (se centra en lo nuevo)
- problemas de actualidad (lento)
- técnico (difícil de usar fuera del ámbito científico)



# Desde las administraciones

- representativo
- actual ( $\pm$ )
- centrado en un territorio
- centrado en un repertorio de especies limitado
- $\pm$  accesible,  $\pm$  disponible
- poco integrable



# Desde la sociedad civil

- actual
- en expansión
- reconocido
- fiabilidad discutida
- $\pm$  poco accesible,  $\pm$  poco disponible
- sesgos diversos



¿Donde encontramos estas características de los juegos de datos de GBIF?

# Swiss National Butterflies Databank

Occurrence dataset published by Centre Suisse de Cartographie de la Faune - CSCF

581.051

Occurrences

[View occurrences](#)

Information

Stats

Activity

## Summary

### FULL TITLE

Swiss National Butterflies Databank

### DESCRIPTION

Database with records of various origins, assembling literature data, digitized collection data and recent observational records provided to the CSCF. Proportionally, this latter resource represents the most important one and is the result of the network activities of voluntary collaborators, internally operated projects (Biodiversity Monitoring Z3-Z4, Red List Project) and impact studies, but also includes data from PhD and master theses. The data cover a period from 1835 to our days. Records have been reviewed according to a standardized validation procedure ([http://www.cscf.ch/page24164\\_en.html](http://www.cscf.ch/page24164_en.html)). Digitized collection data for specimens hosted in Swiss institutions have been excluded from this resource as already part of the corresponding institutional dataset. Nomenclature has been reviewed within the Swiss participation in the Pan-European Species Directory Infrastructure PESI (Fauna Europaea).

### PUBLISHED BY

[Centre Suisse de Cartographie de la Faune - CSCF](#)

### REGISTRATION DATE

29-may-2012

### SERVED BY

[BioCASE Installation Centre Suisse de Cartographie de la Faune - CSCF](#)

### LINKS

[Dataset homepage](#)

ALTERNATIVE IDEN [Dataset homepage](#)

# El estándar de datos de GBIF: Darwin Core

- Tiene una parte de metadatos, inspirada en el estándar Dublin Core
- Inicialmente orientado a datos de presencia (esta especie, en este sitio, en este momento).
- Ahora también le apunta a “check-lists” y atributos de especies
- Tiene una historia de versiones a veces confusa: 1.2 (AKA 2), 1.4, la de TDWG... tuvo “core” y “extensions”, ahora hay una herramienta IPT que tiene sus propias extensiones fuera de estándar....
- Busca la interoperabilidad con otros estándares del ámbito biológico.

## Schema Element Comparison Table

DwC 1.2	DwC 1.21	OBIS	DwC 1.4 (and extensions)	
Longitude	DecimalLongitude	Longitude	<a href="#">DecimalLongitude</a> (geospatial)	
	VerbatimLongitude		<a href="#">VerbatimLongitude</a> (geospatial)	
			StartLongitude	
			EndLongitude	
Latitude	DecimalLatitude	Latitude	<a href="#">DecimalLatitude</a> (geospatial)	
	VerbatimLatitude		<a href="#">VerbatimLatitude</a> (geospatial)	
			StartLatitude	
			EndLatitude	
			HorizontalDatum	<a href="#">GeodeticDatum</a> (geospatial)
			OriginalCoordinateSystem	<a href="#">VerbatimCoordinateSystem</a> (geospatial)
			GeorefMethod	<a href="#">GeoreferenceProtocol</a> (geospatial)
CoordinatePrecision	CoordinateUncertaintyInMeters	CoordinatePrecision	<a href="#">CoordinateUncertaintyInMeters</a> (geospatial)	
			Start_EndCoordinatePrecision	
	LatLongComments		<a href="#">GeoreferenceRemarks</a> (geospatial)	
BoundingBox	BoundingBox	BoundingBox	FootprintWKT ? (geospatial)	
MinimumElevation	MinimumElevationInMeters	MinimumElevation	<a href="#">MinimumElevationInMeters</a>	
MaximumElevation	MaximumElevationInMeters	MaximumElevation	<a href="#">MaximumElevationInMeters</a>	
	VerbatimElevation		<a href="#">VerbatimElevation</a> (curatorial)	
MinimumDepth	MinimumDepthInMeters	MinimumDepth	<a href="#">MinimumDepthInMeters</a>	
MaximumDepth	MaximumDepthInMeters	MaximumDepth	<a href="#">MaximumDepthInMeters</a>	
	VerbatimDepth		<a href="#">VerbatimDepth</a> (curatorial)	

# Darwin Core Georeference Terms (TDWG)

- decimalLatitude, decimalLongitude
- geodeticDatum
- coordinateUncertaintyInMeters
- coordinatePrecision
- pointRadiusSpatialFit *(Legal values: 0, greater than or equal to 1, or undefined)*
- footprintWKT, footprintSRS, footprintSpatialFit
- georeferencedBy, georeferenceProtocol
- georeferenceSources
- georeferenceVerificationStatus
- georeferenceRemarks

# El juego de los científicos

“Si paras de publicar te comen los buitres”

“Lo que hace falta es que las colecciones georeferencien más y publiquen esos datos”

“Si publico mis datos otros los van a usar”

# Data & Science

## Data-Intensive Scientific Discovery

- Data intensive. Basic science is data intensive in its own right, but data sources that support basic science are often insufficient to support applications. Localized impacts with global extent, such as intrusion of invasive species, are often difficult for centralized projects with small numbers of researchers to ascertain. **New applications-appropriate sources must be identified**, and new ways of observing (including the use of communities as data gatherers) must be developed.

Hey, T., Tansley, S., & Tolle, K. (Eds.) 2009. The fourth Paradigm. Data-Intensive Scientific Discovery. Microsoft Research. Redmon. E.E.U.U.

# La publicación de datos se está haciendo un requerimiento del proceso científico

<http://www.nsf.gov/nsb/publications/2011/nsb1124.pdf>



## **Digital Research Data Sharing and Management**

December 2011

Task Force on Data Policies  
Committee on Strategy and Budget  
National Science Board

This report recognizes the evolving role of data in science and society and strong and sustainable data sharing and management policies as a critical national need.



# Que se lee ahí

- Sharing can also be encouraged through the establishment of professional incentives such as promoting **the publication of data in a format that allows for citation and verification.**
- Data scientists and curators should be supported by funding agencies and by their home institutions by **providing pathways for advancement to tenure and other reward mechanisms.**
- ...researchers must be confident **when they share data that they will be properly attributed** and the provenance of the data is assured.
- new data **licensing mechanisms** can preserve intellectual property rights and provide researchers with incentives to make their data public.

# Que se está haciendo (1)

Publicar los datos de manera citable y verificable:  
“artículos de datos”, revistas de datos:



Council on  
Library and  
Information  
Resources

[My Profile](#) | [Communities](#) | [Directory](#) | [Re.Thinking Blog](#)

**Data Journals: Incentivizing Research Data Dissemination**

<http://connect.clir.org/blogs/katherine-akers/2013/12/12/data-journals-incentivizing-research-data-dissemination>

– Pensoft data papers:

<http://biodiversitydatajournal.com/>

– Nature Group: *Scientific Data*.

<http://www.nature.com/sdata/>

# Que se está haciendo (2)

- Citar los datos

- Mal

Con la finalidad de corroborar la información de su distribución, se consultaron las bases de datos de museos y colecciones disponibles a través de los portales electrónicos: Global Biodiversity Information Facility (GBIF; [www.gbif.net](http://www.gbif.net), julio del 2006), Ornis (<http://olla.berkeley.edu/ornisnet>, julio de 2006) y la Red Mundial de Información sobre Biodiversidad (REMIB; [http://www.conabio.gob.mx/remib/doctos/remib\\_esp.html](http://www.conabio.gob.mx/remib/doctos/remib_esp.html), julio de 2006).

# Que se está haciendo (2)

- Citar los datos

- Mal

Con la finalidad de corroborar la información de su distribución, se consultaron las bases de datos de museos y colecciones disponibles a través de los portales electrónicos: Global Biodiversity Information Facility (GBIF; [www.gbif.net](http://www.gbif.net), julio del 2006), Ornis (<http://olla.berkeley.edu/ornisnet>, julio de 2006) y la Red Mundial de Información sobre Biodiversidad (REMIB; [http://www.conabio.gob.mx/remib/doctos/remib\\_esp.html](http://www.conabio.gob.mx/remib/doctos/remib_esp.html), julio de 2006).

- Mejor

ANEXO I  
Lista de instituciones que contribuyen con registros de colecciones de murciélagos usados en este análisis

APPENDIX I  
List of institutions providing bats specimen data used in this analysis

GBIF

- Mammal specimens, <http://data.gbif.org/datasets/resource/181>
- Mammals, <http://data.gbif.org/datasets/resource/8439>
- Registros biológicos en Áreas protegidas obtenidos de documentos impresos, <http://data.gbif.org/datasets/resource/10869>
- TCWC Vertebrate Collections, <http://data.gbif.org/datasets/resource/8962>
- Tissues Specimens, <http://data.gbif.org/datasets/resource/9163>
- MCZ Mammalogy Collection, <http://data.gbif.org/datasets/resource/581>
- Collection Mammalia SMF, <http://data.gbif.org/datasets/resource/8313>
- Colección Nacional de Mastozoología - Museo Argentino de Ciencias Naturales 'Bernardino Rivadavia', <http://data.gbif.org/datasets/resource/9115>
- NMNH Vertebrate Zoology Mammals Collections, <http://data.gbif.org/datasets/resource/1837>
- Colección de Mastozoología, <http://data.gbif.org/datasets/resource/13547>
- Mammal specimens, <http://data.gbif.org/datasets/resource/657>
- Mammal specimens, <http://data.gbif.org/datasets/resource/582>

# Que se está haciendo (2)

- Citar los datos

- Mal

Con la finalidad de corroborar la información de su distribución, se consultaron las bases de datos de museos y colecciones disponibles a través de los portales electrónicos: Global Biodiversity Information Facility (GBIF; [www.gbif.net](http://www.gbif.net), julio del 2006), Ornis (<http://olla.berkeley.edu/ornisnet>, julio de 2006) y la Red Mundial de Información sobre Biodiversidad (REMIB; [http://www.conabio.gob.mx/remib/doctos/remib\\_esp.html](http://www.conabio.gob.mx/remib/doctos/remib_esp.html), julio de 2006).

- Mejor

ANEXO I  
Lista de instituciones que contribuyen con registros de colecciones de murciélagos usados en este análisis

APPENDIX I  
List of institutions providing bats specimen data used in this analysis

GBIF

- Mammal specimens, <http://data.gbif.org/datasets/resource/181>
- Mammals, <http://data.gbif.org/datasets/resource/8439>
- Registros biológicos en Áreas protegidas obtenidos de documentos impresos, <http://data.gbif.org/datasets/resource/10869>
- TCWC Vertebrate Collections, <http://data.gbif.org/datasets/resource/8962>
- Tissues Specimens, <http://data.gbif.org/datasets/resource/9163>

- Ideal

```
1 Please cite this data as follows, and pay attention to the rights documented in the rights.txt:
2 Please respect the rights declared for each dataset in the download:
3 The New York Botanical Garden: Herbarium of The New York Botanical Garden
4 WAHerb
5 National Biodiversity Data Centre: Irish Records from the Fungal Records Database of Britain and Ireland
6 Use of these data should be credited as follows:[compiler's name(s)]. Flora Mycologica Iberica fungus distribu
7 University of Arkansas: Planetary Biodiversity Inventory Eumycetozoon Databank
8 -
9 Estrada Torres, A. 2004. Hongos ectomicorrizógenos y myxomycetes asociados con Picea chihuahuana en la Sierra
10 Pando, F. et al. (2003). MA Cryptogamic collections online databases. http://www.rjb.csic.es/herbario/crypto/c
11 V. L. Komarov Botanical Institute, St. Petersburg, The Myxomycetes Collection at the V. L. Komarov Botanical I
12 -
13 BeBIF Provider: National Botanic Garden Belgium - Myxomycetes
14 National Museum of Nature and Science, Japan: Myxomycetes Collection of National Museum of Nature and Science
15 Centre d'estudis de la neu i de la muntanya d'Andorra (CENMA), Institut d'Estudis Andorrans: Fongs d'Andorra
```

# cambiar la cultura para hacer el sistema sostenible

- algo ya podemos hacer
  - Citar las fuentes adecuadamente
- más va a venir
  - DOIs para descargas (“generate and store a citation file for each data download, and issue a DOI for the citation file”. GBIF working doc)
- más hace falta
  - que sea una exigencia del sistema (A là Genbank)
  - que haya métricas y se usen
  - que sea un mérito científico

# Dos reflexiones

- Una que es un inciso y corolario:
  - aportar nuestros datos
- En busca de las ausencias
  - El caso del software Ornitho

# La plataforma *ornitho*

**ICO**

Gabriel Gargallo [Administrador]

Pàgina d'inici d'ornitho.cat

Entitats col·laboradores

- Observacions
- Observatori
  - El darrer 2 dies
  - El darrer 3 dies
  - El darrer 15 dies
  - Les darreres observacions de...
  - Totes les zones observades
  - Costa del aragonesa
  - Reunir del
- Dades i anàlisi
  - El darrer 24 hores
  - Totes les imatges i els sons
- Participa
  - Introduir observacions
  - Les dades i l'èxit
- Informació

Benvingut a [www.ornitho.cat](http://www.ornitho.cat)

ornitho.cat és el portal web dedicat a l'intercanvi d'informació sobre les observacions d'ocells de Catalunya. És una iniciativa de l'Institut Català d'Ornitologia vinculada amb el Banc de Dades de Biodiversitat de Catalunya i promoguda conjuntament amb el departament de Medi Ambient i Habitatge i diverses entitats ornitològiques i naturalistes de caire més local.

**Darreres notícies**

**divendres 9 març 2012**

El Banc de Dades d'ORNITHO.CAT. En aquest portal podreu registrar totes les vostres observacions d'ocells i consultar les dades disponibles en temps real.

**divendres 17 febrer 2012**

Se celebra l'acte de presentació de l'edició catalana de l'Atlas de les Aves de Catalunya. La festa comença el dimarts 16 de febrer a les 19h i acaba el dimarts 17 de febrer a les 12h.

[altres notícies]

**Titula**  
divendres 17 març 2012  
Plena (24 dies)  
Set a les 13:40 i s'amaga a les 14:50

Set i sortida a les 04:20 i posta a les 16:43.  
Dia i sol a les 14:50 i crepuscle a les 16:43

Què veure avui: [Llista de punts d'observació]

Les darreres observacions [VCL]

[Mostra les observacions recents]

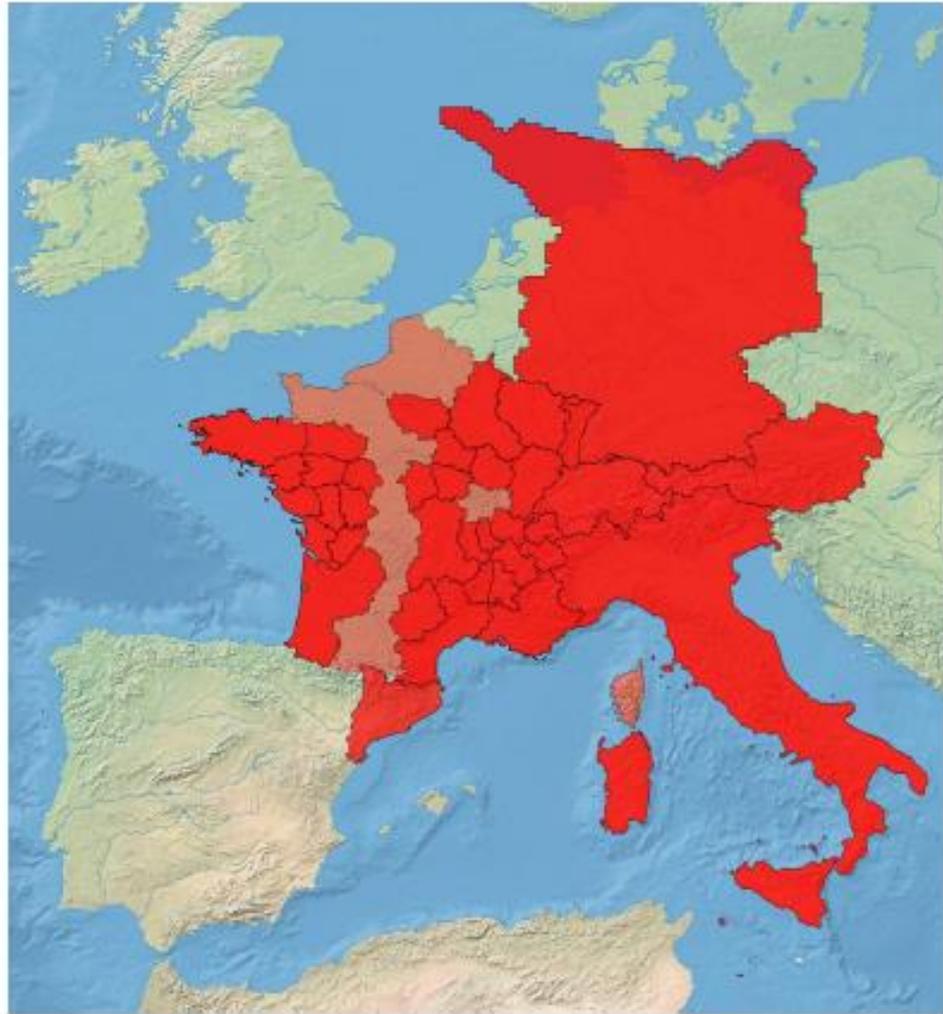
**divendres 16 març 2012**

**Avifauna de les Illes Balears (2011/2012)**

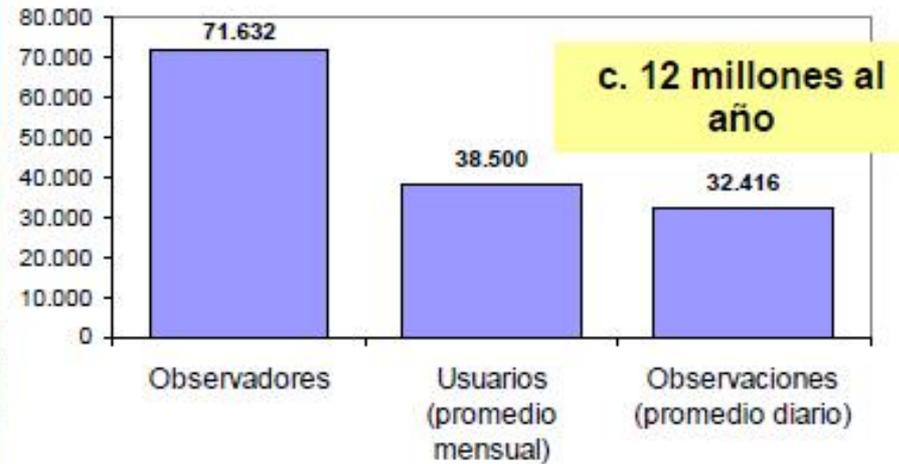
- Fardus roger
- Falco
- Euphonia
- Fregata comuna
- Gambusia mediterrània
- Colinus squamatus
- Turdus
- Turdus merula
- Columba palumbus
- Musca domestica
- Falco tinnunculus
- Turdus
- Cyanus cyaneus
- Falco
- Columba palumbus

Gabriel Gargallo  
Institut Català d'Ornitologia

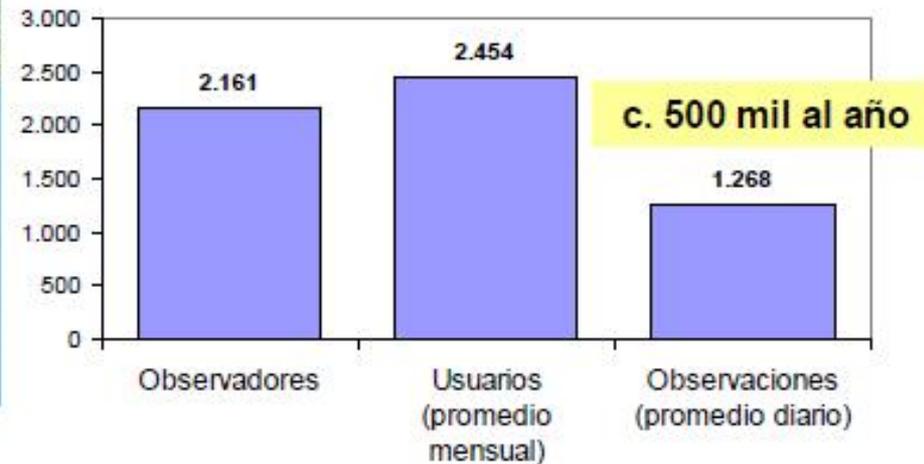
# Implantación y uso



ornitho (global)



ornitho.cat



# Funcionamiento: introducción de datos

Introducir observaciones - Aves

\* Los campos marcados con un asterisco son obligatorios.

\* Fecha: 21.05.2014 ✓Fecha y hora

\* Localidad: Àger [310/656] altitud 1056 m (preds tanto como puedas en áreas de montaña)

Tu última observación: Àger [310/656] domingo 27 abril 2014

\* Especie: Sylvia atricapilla ✓Especie y número

\* Número total de individuos: Conteo exacto 1

Los campos de abajo no son obligatorios

Comentario: Comentario oculto ✓Comentarios y otros datos complementarios (e.g. sexo, código atlas...)

Otros datos/información:  Dato oculto ✓Ocultación del dato

Código Atlas: Proporcional a la tasa de la observación

Puedes añadir una imagen JPEG (máx. 430 píxeles en el lado pequeño) o un sonido MP3 (máx. 1 Mb) relativo a esta observación

Subir imágenes y sonidos ✓Subir imágenes y sonidos

✓Opción 1: entrada de datos casuales

# Funcionamiento: introducción de datos

Introducir observaciones

Localidad: Áger [310/636] / Áger (NOA)  
Fecha: miércoles 21 mayo 2014 / 10h00-12h00

[cambiar la localidad]

Por favor, diga sobre todas las especies observadas (cuando una hilera se vuelve amarilla indica que se ha elegido la especie). Cuando utilizas las listas completas tus datos son más valiosos pero aun puedes incrementar más su valor si das información del número de aves observadas. Para añadir otros detalles de la observación diga sobre +.

+ Anax platyrhynchos	▼	Código Atlas: ningún ▼
+ Lagopus muta	▼	Código Atlas: ningún ▼
+ Tetrao urogallus	▼	Código Atlas: ningún ▼
+ Alcedo rufa	▼	Código Atlas: ningún ▼
+ Perdix perdix	▼	Código Atlas: ningún ▼
+ Coturnix coturnix	▼	Código Atlas: ningún ▼
+ Phasianus colchicus	▼	Código Atlas: ningún ▼
+ Bubulcus ibis	▼	Código Atlas: ningún ▼
+ Egretta garzetta	▼	Código Atlas: ningún ▼
+ Ardea cinerea	▼	Código Atlas: ningún ▼
+ Ardea purpurea	▼	Código Atlas: ningún ▼
+ Ciconia nigra	▼	Código Atlas: ningún ▼
+ Ciconia ciconia	▼	Código Atlas: ningún ▼
+ Pernis ptilorhynchus	▼	Código Atlas: ningún ▼
+ Milvus migrans	▼	Código Atlas: ningún ▼

✓ Opción 2: entrada de listas completas

# Para concluir y debatir

- Menos "que bien lo hace GBIF" y más apoyo real a los procesos de publicación de datos: desde los financiadores, desde las políticas, desde la cultura de los científicos, desde la educación
- Ser parte de la solución: publicando los propios datos; citar las fuentes apropiadamente

Francisco Pando

Unidad de coordinación, GBIF España

Real Jardín Botánico - CSIC

Claudio Moyano 1, 28014 Madrid

[pando@gbif.es](mailto:pando@gbif.es)

[www.gbif.es](http://www.gbif.es)



<http://creativecommons.org/licenses/by-sa/3.0/es/>

