GBIF…Spain…data collection…data processing

Francisco Pando
GBIF - España

Building the European
Biodiversity
Observation Network

Joint Workshop Sierra Nevada LTER and EU BON
EU BON site data management workshop
Carmen de la Victoria, Granada, Spain. Jan . 2014

Summary

- Definitions & clarifications
- GBIF, the blue sky picture
- GBIF as a framework
- Gorillas in the room
- GBIF and LTERN
- How can we make it better (closing remarks)

Definitions & clarifications

- GBIF is not an european project
- GBIF is not even a project!
- What is data
  - (no answers, but stuff to get answers)
- What is biodiversity data
- What is accessible data (it is not available data)
- GBIF is not a data source (there is no such a thing as GBIF data)
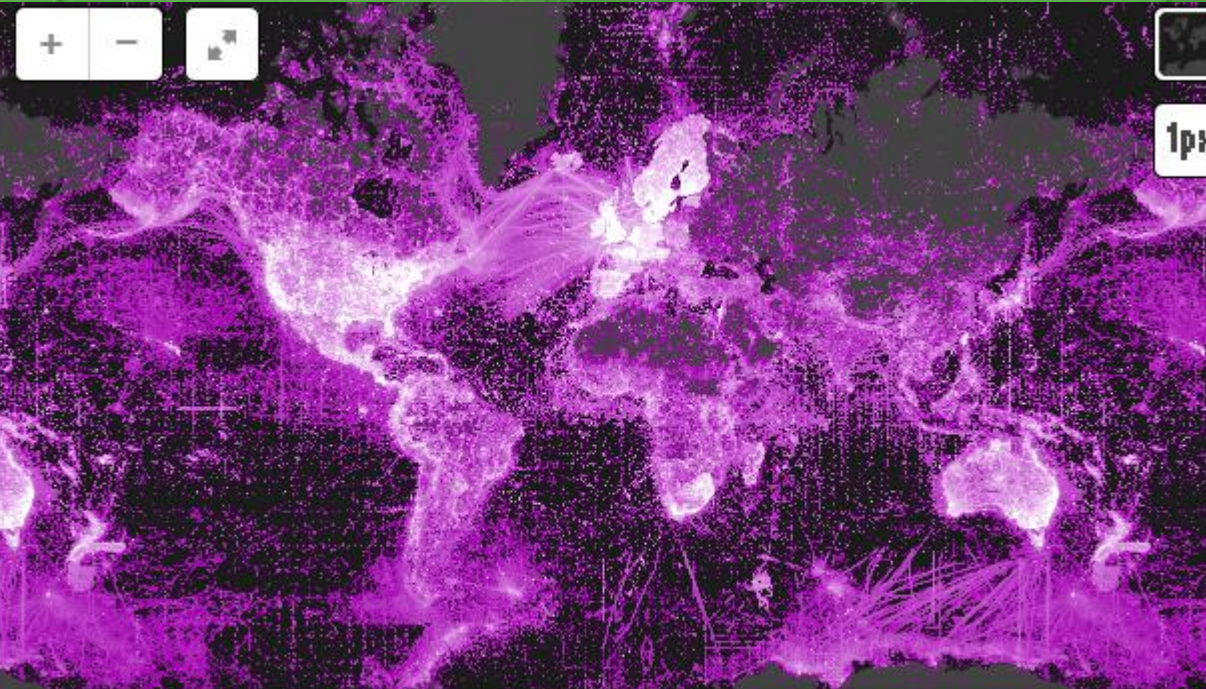
GBIF, the blue sky picture

GBIF

Global Biodiversity Information Facility
Free and open access to biodiversity data

428,364,089 OCCURRENCES | **1,454,695** SPECIES | **14,626** DATASETS

Taxonomic characteristics

The following provides a summary of number of records per kingdom. Further filters, such as a location or temporal fi... applied when **exploring the data**.

286.515.884 (66,8861%) Animalia records	121.523.887 (28,3693%) 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
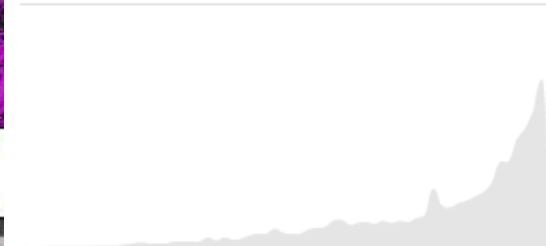
Records may originate from a variety of means, such as a scientist collecting a specimen or an individual recording th... an organism. This is classified by the **Darwin Core basisOfRecord** standard.

303.738.491 (70,907%) Observation records	89.465.185 (20,885%) Specimen records	31.381.460 (7,326%) Unknown evidence records	2.615.217 (0,611%) Fossil records
762.123 (0,178%) Living Specimen records	401.613 (0,094%) Literature Occurrence records		

Temporal characteristics

This visualization sh... growth in occurrence... **after 1950**. GBIF pro... to many older record... add date range filters... content for any period...

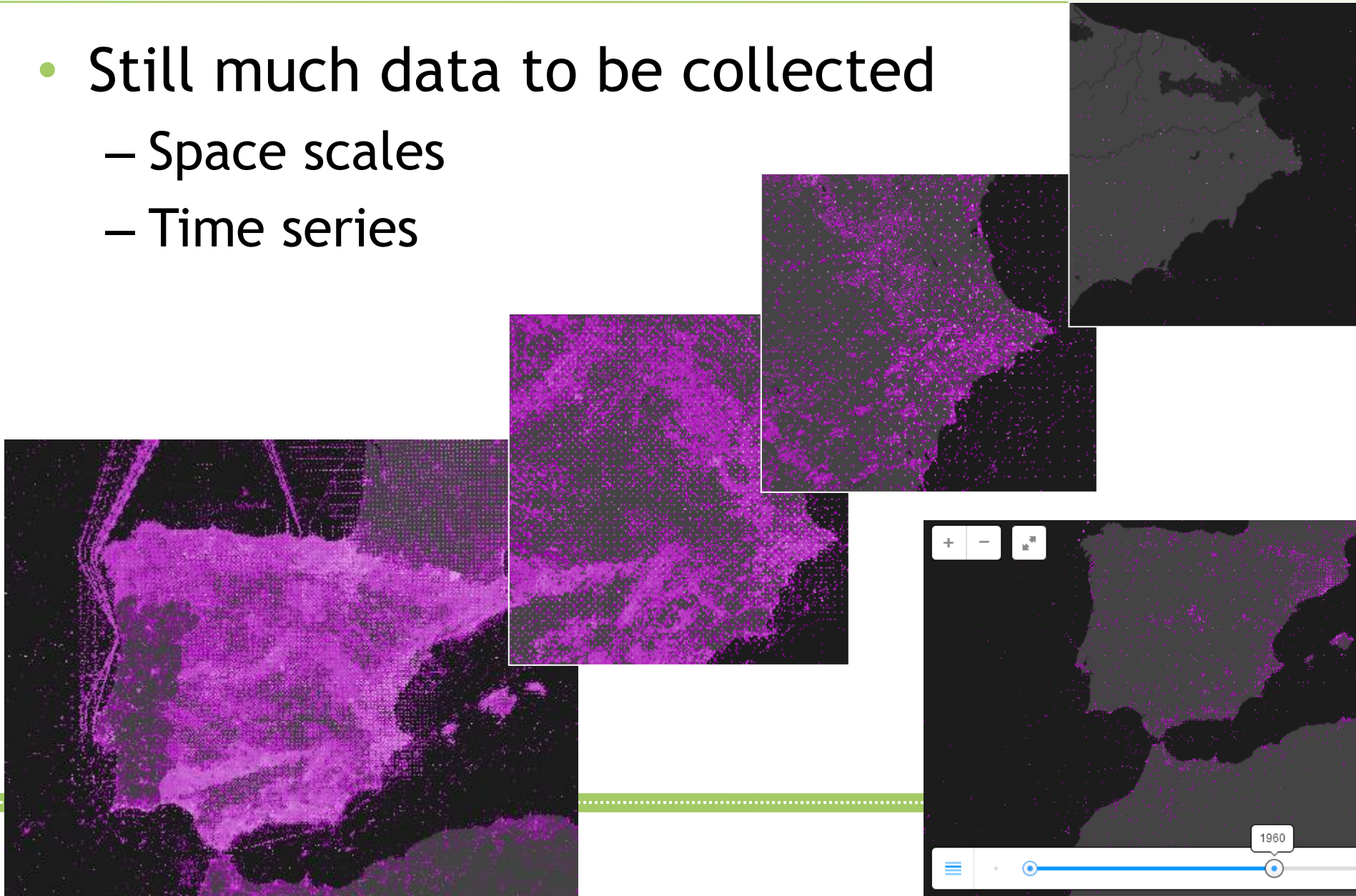For example, **here is** ... for records between ... 1950.

GBIF is not a data source, it is a framework

- Physical infrastructure
  - Physical infrastructure, cyberinfrastructure, human resources, and expertise, and program management and coordination.
- Information infrastructure
  - Data workflows
  - Datasets
  - Data openness
  - Data products
- Capability infrastructure
  - Knowledge base and management, training
  - Standard development
  - Data gateways development

Gorillas in the room (1)

- Still much data to be collected
  - Space scales
  - Time series

Gorillas in the room (2)

- Global, unique, persistent & resolvable identifiers for data units (records, datasets, etc.)
  - Without them, adding and connecting content is very inefficient
  - Tracking usage is also difficult

Gorillas in the room (3)

But it is all too easy to **take for granted** that these data will be stored, managed and kept freely available for researchers to query. ...and generated

EMBL
40 YEARS
European Molecular Biology Laboratory

- Data creation and publishing should be properly recognized
  - Without that, the knowledge system is not sustainable

Citing data use…

Nature (2013) doi:10.1038/nature12872

GBIF NEWS

How plants weather the cold

NATURE | LETTER

Three keys to the radiation of angiosperms into freezing environments

Amy E. Zanne, David C. Tank, William K. Cornwell, Jonathan M. Eastman, Stephen A. Smith, Richard G. FitzJohn, Daniel J. McGlinn, Brian C. O'Meara, Angela T. Moles, Peter B. Reich, Dana L. Royer, Douglas E. Soltis, Peter F. Stevens, Mark Westoby, Ian J. Wright, Lonnie Aarssen, Robert I. Bertin, Andre Calaminus, Rafaël Govaerts, Frank Hemmings, Michelle R. Leishman, Jacek Oleksyn, Pamela S. Soltis, Nathan G. Swenson, Laura Warman &

1038/nature12872

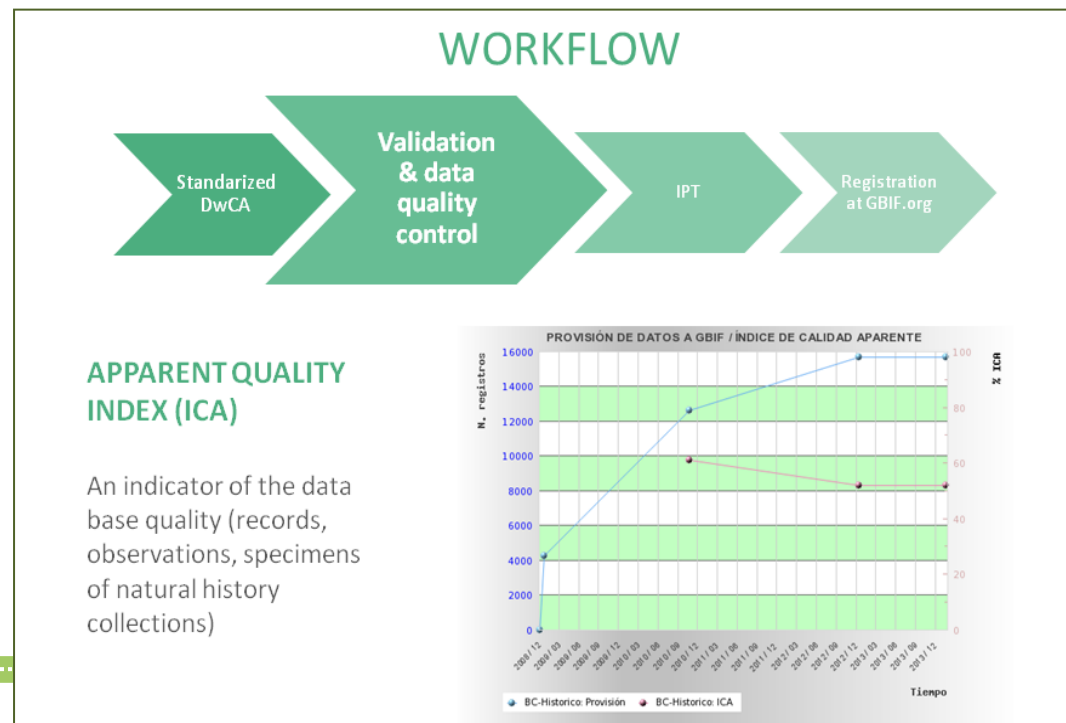RESEARCH | SUPPLEMENTARY INFORMATION

Supplementary Table 4 Global Biodiversity Information Facility (GBIF) data providers. A list of data providers to GBIF of plant georeference points used in this manuscript.

39 pages!

Data providers
GEO-Tag der Artenvielfalt: Spandau HBO
Jagiellonian University, Institute of Zoology: Weevils of Wales and England
Levy Tacher, S. I. 1999. Contribución al conocimiento de la flora útil de la selva Lacandona. Conservation International México A.C. Bases de datos SNIB2010-CONABIO. Proyecto No. M002. México, D.F.
UK National Biodiversity Network: Glasgow Museums BRC - The Changing Flora of Glasgow: Orchid Dataset
GEO-Tag der Artenvielfalt: Rotes Steigle (Panzerübungplatz Böblingen)
Dickoré B. The Himalayan Uplands Plant database (HUP Version 1). Global Mountain Biodiversity Assessment GMBA
GEO-Tag der Artenvielfalt: Wälder bei Nordkirchen
Téllez Valdés, O. y J. Martínez. 2000. Base de datos de la flora de la Reserva de la Biosfera Chamela-Cuixmala, Jalisco, México. Universidad Nacional Autónoma de México. Instituto de Biología. Bases de datos SNIB2010-CONABIO proyecto No. L289. México, D.F.
GEO-Tag der Artenvielfalt: Feriendorf des Kreises Gedern (Ober-Seemen)
Centre d'estudis de la neu i de la muntanya d'Andorra (CENMA), Institut d'Estudis Andorrans: Fongs d'Andorra
Guardia, R. et al. (2007). Bases de dades de l'Herbari BCN http://www.ub.es/cedocbiv/bancdade.htm
Mwanga Mwanga I, Mergen P, Theeten F (2013) Herbarium Specimens of LW, CRSN, RMCA
GEO-Tag der Artenvielfalt: Dreilinden Gymnasium-Schulgelände
GEO-Tag der Artenvielfalt: Hochschule Zittau/Görlitz
University of British Columbia Herbarium (UBC). http://www.biodiversity.ubc.ca/museum/herbarium/database.html. (consulted on [date]), http://www.biodiversity.ubc.ca/museum/herbarium/database.html
National Museum of Nature and Science, Japan: Herbarium Specimens of Tokushima Prefectural Museum, Japan

Data quality and workflows

- – Data annotation, dataset flagging
  - Linked data
  - Semantic web
- – Data cleaning and DQ assurance practices

WORKFLOW

Standarized DwCA → Validation & data quality control → IPT → Registration at GBIF.org

APPARENT QUALITY INDEX (ICA)

An indicator of the data base quality (records, observations, specimens of natural history collections)

PROVISIÓN DE DATOS A GBIF / ÍNDICE DE CALIDAD APARENTE

More on data quality

✅ Metadata

- EML-GMP-IPT

Linking data quality and incentives

Data papers:

- Academic article
- Human readable
- Make data findable, assessable DOI-> URL
- Based in EML
- GBIF's IPT produces draft

Literature based species occurrence data of birds of northeast India

Sujit Narwade[1], Mohit Kalra[1], Rajkumar Jagdish[1], Divya Varier[1], Sagar Satpute[1], Noor Khan[1], Gautam Talukdar[2], V. B. Mathur[2], Karthikeyan Vasudevan[2], Dinesh Singh Pundir[2], Vishwas Chavan[3], Rajesh Sood[3]

1 *Bombay Natural History Society (BNHS), Shaheed Bhagatsingh Road, 400001, Mumbai, India* 2 *Wildlife Institute of India (WII), Post Box No 18, Chandrabani, 248001, Dehradun, India* 3 *Global Biodiversity Information Facility (GBIF), Universitetsparken 15, DK 2100, Copenhagen, Denmark*

Corresponding author: *Sujit Narwade* (bnhs@envis.nic.in)

GBIF & LTERN

Centre d'Estudis Avançats de Blanes – CSIC.

– http://www.gbif.org/dataset/81119a40-f762-11e1-a439-00145eb45e9a
Centre d'Estudis Avançats de Blanes. Macròfits Pirineu

– http://www.gbif.org/dataset/810f3e62-f762-11e1-a439-00145eb45e9a
Centre d'Estudis Avançats de Blanes. Macroinvertebrados Limno

– http://www.gbif.org/dataset/8119a2da-f762-11e1-a439-00145eb45e9a
Centre d'Estudis Avançats de Blanes. Limnological Observatory of the
Pyrenees, Diatomeas

– http://www.gbif.org/dataset/81165620-f762-11e1-a439-00145eb45e9a
Centre dEstudis Avancats de Blanes. Limnological Observatory of the
Pyrenees Zooplancton

– http://www.gbif.org/dataset/811076f6-f762-11e1-a439-00145eb45e9a
Centre dEstudis Avancats de Blanes. Limnological Observatory of the
Pyrenees Fitoplancton

Observatorio de cambio global Sierra Nevada

– http://www.gbif.org/dataset/db6cd9d7-7be5-4cd0-8b3c-fb6dd7446472
Sinfonevada: Floristic diversity in Forest of Sierra Nevada

– Phenology of flora of mediterranean high-mountains meadows (Sierra
Nevada) in IPT work in progress.

Closing remarks

- GBIF is not a competitor; is an essential component in our environmental knowledge infrastructure

- A component that:
  - Goes beyond data
  - Can be reused in many ways and contexts
  - You –as users of it, as EU-BON– can help to develop it, to make it more complete and usable

Francisco [Paco] Pando

Coordination Unit. GBIF-Spain
Real Jardín Botánico - CSIC
Claudio Moyano 1
28014 Madrid, Spain

pando@gbif.es
www.gbif.es
Phone: + 34 91 420 3017
Fax: + 34 91 420 0157