

# DATA PUBLISHING

Why and how?

**GBIF, Identifiers, publishing platforms**

*Hanna Koivula, Dmitry Schigel*

# GBIF IS...

- International open data research and policy infrastructure
- free and open access to biodiversity information online

# FOCUS ON

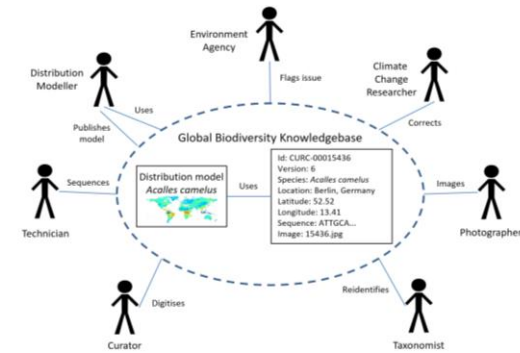
- access to global species occurrences through GBIF.org and web services
- common standards and free tools for biodiversity data management

# FOCUS ON

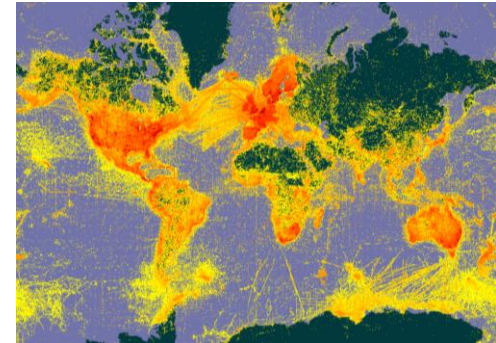
- mobilizing and sharing biodiversity data and information
- collaborative network at global and regional levels

# WHY WE DO IT

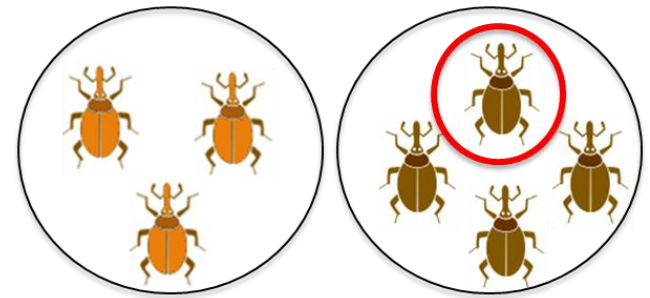
1. Collaboration in sharing and use of biodiversity data



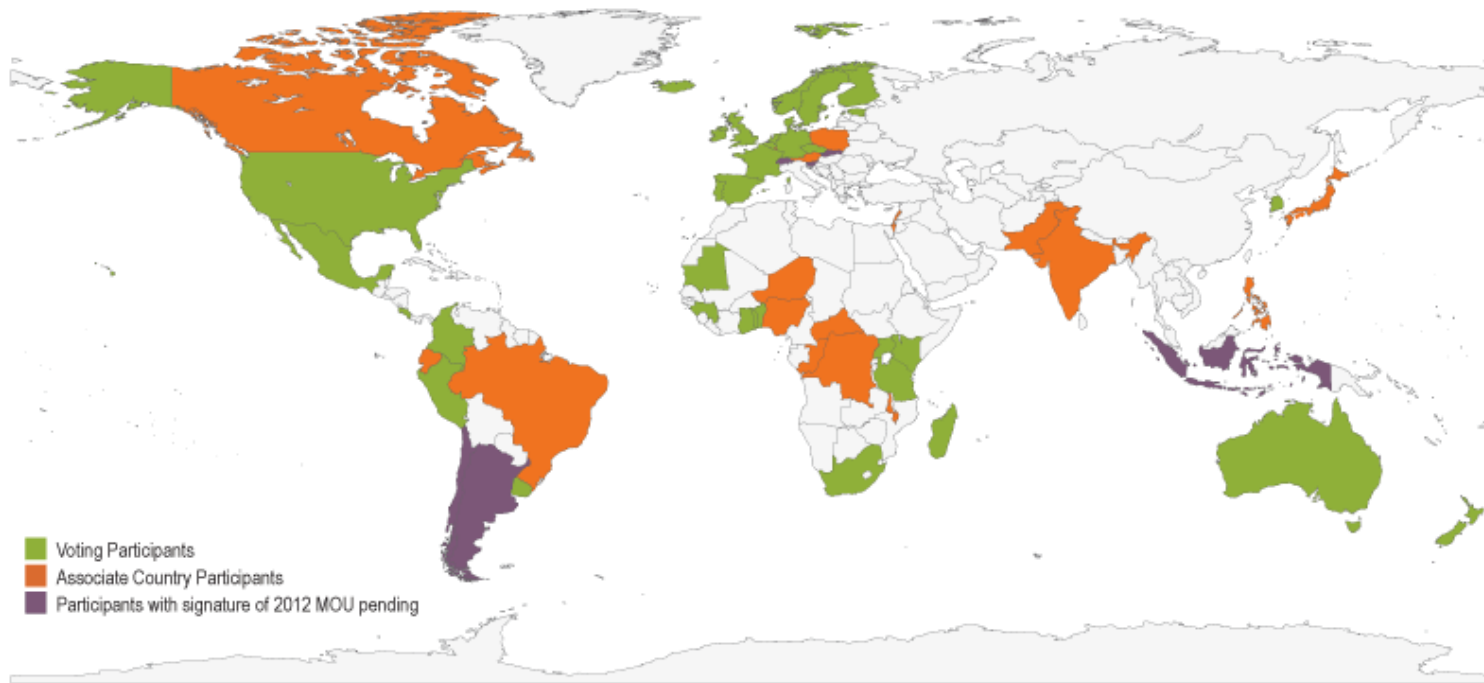
2. Organising evidence of occurrence of species in time and space



3. Support global virtual natural history collection



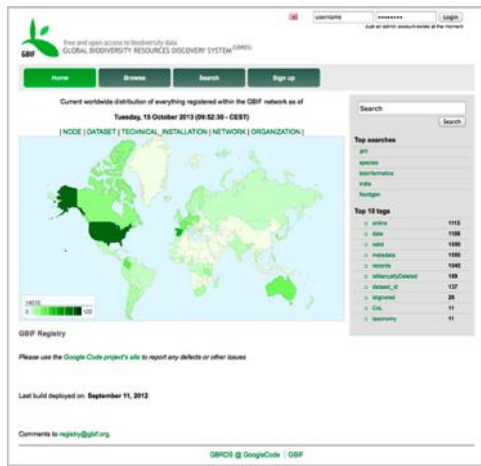
# 57 countries and 39 institutions participate in GBIF



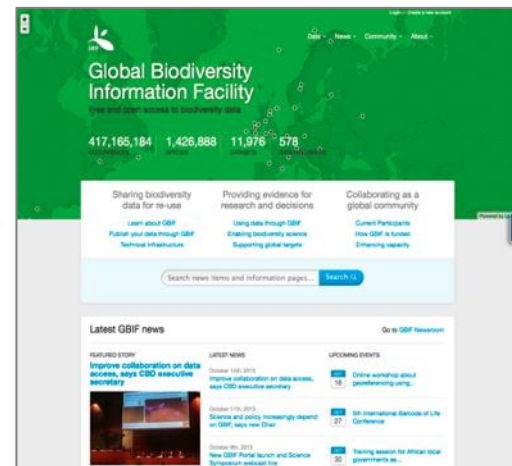
>800 institutions share data through GBIF network

# GBIF provides a data discovery system

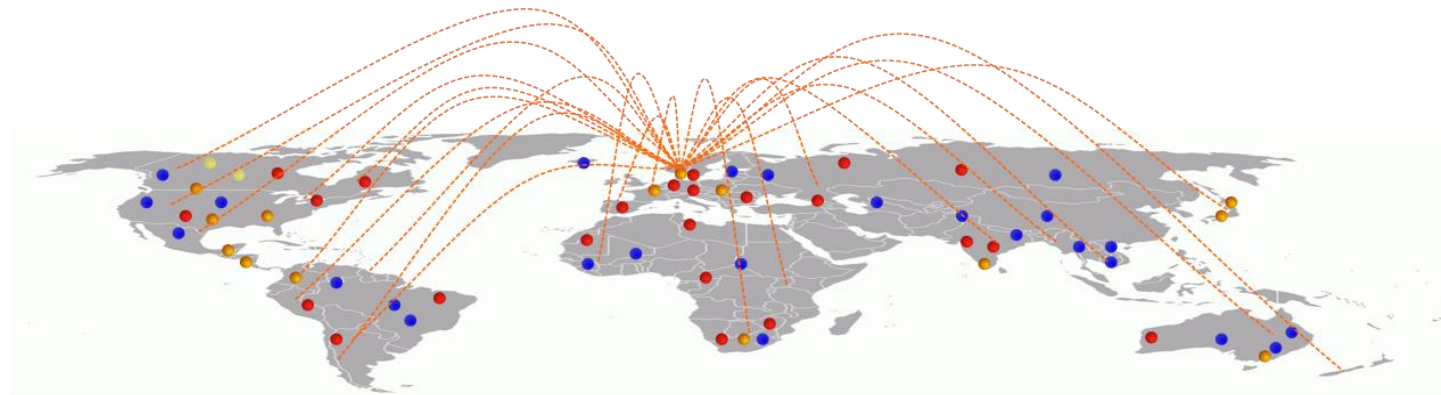
*that is dependent on resolvable stable identifiers for efficient functionality*



global registry



data portal



# GBIF BY THE NUMBERS

**624,470,102**  
species occurrence  
records

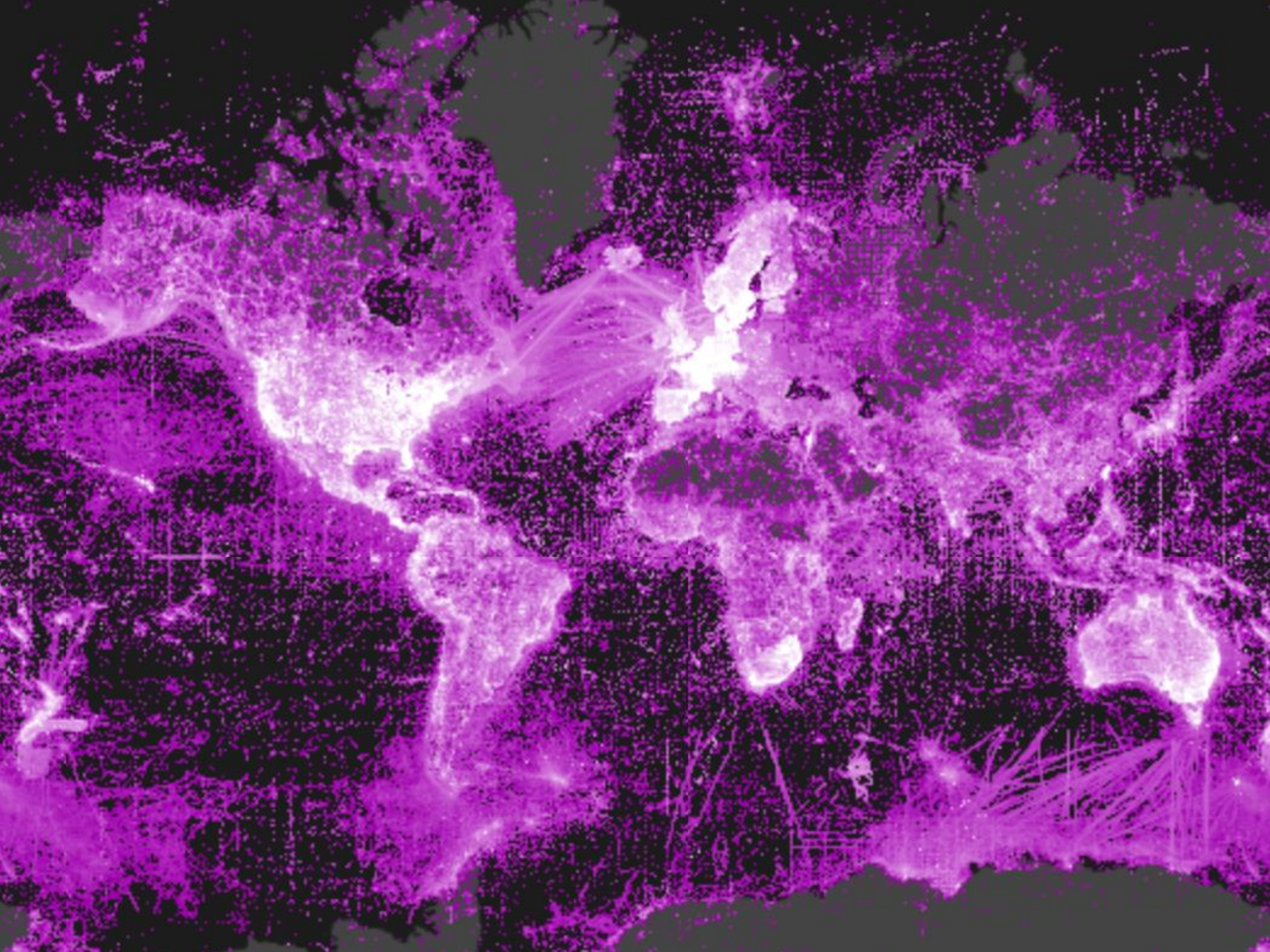
**1,643,948**  
species

**29,304**  
datasets

**821**  
data-publishing  
institutions













# Licensing for data published through GBIF

- GBIF Governing Board decision in 2014 - three CC licenses
- Status: set for all datasets published in GBIF since August 2016

	<b>CC0</b> possible	CC0 is recommended whenever	
	<b>CC BY</b>	with use guidelines specified by the owner	
	<b>CC BY-NC</b>	use is not for commercial purposes	




# **Mobilising and Publishing Data to GBIF portal**


# Data Publishing Options


Select a tool

GBIFS developed tools

Community tools

1  Integrated Publishing Toolkit

2  Spreadsheet Processor

3  Dwc-A Assistant

4  TAPIR

 DIGIR  BioCASE

Publish your data

Register with GBIF

Discovery through the Portal



# Using GBIF Data portal





# Global Biodiversity Information Facility

Free and Open Access to Biodiversity Data

Data ▾

News ▾

Community ▾

About ▾

[Explore occurrences](#)

[Explore species](#)

[Explore dataset](#)

[Explore by country](#)

[Explore data trends](#)

[Publishing data](#)

[Using Data](#)

[Infrastructure](#)

651,368,995

OCCURRENCES

1,634,951

SPECIES

22,051

DATASETS

805

DATA PUBLISHERS

## Sharing biodiversity data for re-use

[Learn about GBIF](#)

[Publish your data through GBIF](#)

[Technical infrastructure](#)

## Providing evidence for research and decisions

[Using data through GBIF](#)

[Enabling biodiversity science](#)

[Supporting global targets](#)

## Collaborating as a global community

[Current Participants](#)

[How GBIF is funded](#)

[Enhancing capacity](#)

## Explore **624,169,030** occurrences

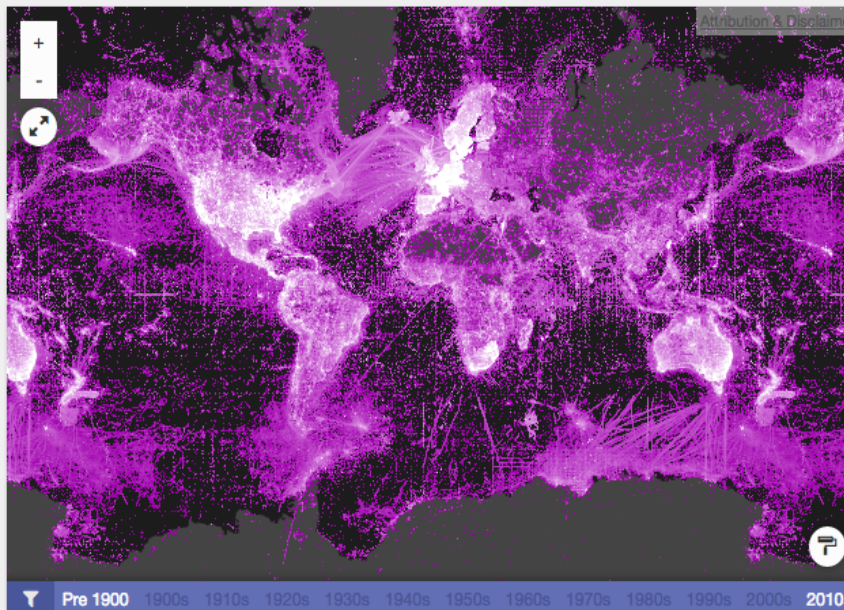
Occurrence records document evidence of a named organism in nature. Through this portal, you can [search](#), [view](#) and [download](#) records that are published through the GBIF network.

**624,169,030**

occurrences records

**546,208,754**

georeferenced records



### Georeferenced data

#### VIEW RECORDS

[All records](#) | [In viewable area](#)

#### ABOUT

This map shows the density of all 546,208,754 georeferenced occurrence records published through the GBIF network.

To explore the records, zoom into the map or click on the links above and add further filters to customize search results.

# Search 1,634,951 species


of the [GBIF Backbone Taxonomy](#)

 [Search Q](#)

[Mammals](#) · [Birds](#) · [Insects](#) · [Reptiles](#) · [Fishes](#) · [Butterflies](#) · [Fungi](#) · [Flowering Plants](#)



Confirmed species in the [Catalogue of Life](#)

Names under review 



# Hepatica nobilis Schreb.

Species in GBIF Backbone Taxonomy

Plantae · Tracheophyta · Magnoliopsida · Ranunculales · Ranunculaceae · Hepatica

73,870

Occurrences

6

Infraspecies

[View occurrences](#)

Information

## Overview

### FULL NAME

Hepatica nobilis Schreb.

### TAXONOMIC STATUS

accepted species

### COMMON NAMES

- hepatica eng
- Leberblümchen deu
- Leberblümchen

[more >](#)

### HABITAT

Not marine

### SYNONYMS

- [Anemone hepatica L.](#)
- [Anemone hepatica var. acuta Hitchc.](#)
- [Anemone hepatica var. acutiuscula Pritz.](#)
- [Anemone praecox Coste](#)
- [Anemone praecox Salisb.](#)

[more >](#)



GBIF ID [5371699](#)

### SEARCH LINKS

- [Encyclopedia of Life](#)
- [Catalogue of Life](#)
- [Biodiversity Heritage Library](#)



## Georeferenced data

### VIEW RECORDS

All 66,378 | In viewable area

### DISTRIBUTIONS

Text based [distributions](#) present in some sources.

[Add a filter](#)

Scientific name

Country

Location

Establishment Means

Basis of record

Type status

Multimedia types

Collector

Record number

OccurrenceID

OrganismID

Collection date

Year

Month

Last modified in GBIF

Catalogue number

Dataset

Institution code

Collection code

Publishing country

Elevation

Depth

Issues

Data repatriation

Locality

State/Province

Water body

Protocol

License

## Refine your search

[reset](#)

### BASIS OF RECORD

- [Human Observation](#) (56,830)
- [Specimen](#) (8,837)
- [Unknown evidence](#) (7,304)
- [Observation](#) (636)
- [Living Specimen](#) (232)
- [Literature Occurrence](#) (28)
- [Machine Observation](#) (2)
- [Fossil](#) (1)

### TYPE STATUS

- [type](#) (4)
- [holotype](#) (2)
- [syntype](#) (1)

### DATASET

- [Artdata](#) (27,984)
- [INPN - Données flore d...](#) (9,474)
- [Norwegian Species Obse...](#) (4,401)
- [Occurrence Data of Vas...](#) (4,315)
- [FloraCAT: Banco de dat...](#) (3,800)
- [Flora von Deutschland ...](#) (3,110)
- [Vascular Plants, Field...](#) (1,710)
- [Fundación Biodiversida...](#) (1,240)
- [VegetWeb - Repositori...](#) (1,152)
- [Vascular Plant Herbari...](#) (1,115)

# imr\_mareano\_beamtrawl

Occurrence dataset published by Institute of Marine Research

15,733

Occurrences

[View occurrences](#)

Information

Stats

Activity

All datasets in GBIF are assigned a dataset DOI

## Summary

### FULL TITLE

imr\_mareano\_beamtrawl

### DESCRIPTION

Data collected with a 2m beamtrawl towed about 200-300m. Dataset contains numbers and combined wet weight of specimen (down to 1 mg) within same species or higher taxonomic unit. Samples are preserved on formalin and then alcohol. After id the material is stored at the Bergen Museum, Bergen, [Norway](#).

### TEMPORAL COVERAGES

Date range: Jan 1, 2006 - Aug 5, 2013

### LANGUAGE OF METADATA

ENGLISH

### LANGUAGE OF DATA

ENGLISH

### ORIGINATOR

[Arne Hassel](#)

### METADATA AUTHOR

[Arne Hassel](#)

### ADMINISTRATIVE CONTACT

[Arne Hassel](#)

**DOI** [doi:10.15468/iomgkj](https://doi.org/10.15468/iomgkj)

### PUBLISHED BY

[Institute of Marine Research](#)

### PUBLICATION DATE

Jun 15, 2016

### REGISTRATION DATE

Apr 29, 2016

### SERVED BY

[IMR.NO-IPT](#)

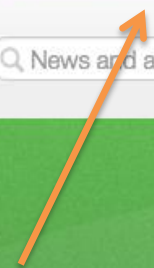
### ALTERNATIVE IDENTIFIERS

#### - UUID

99006141-bf02-4821-8e6e-5818a...

#### - [Source webpage](#)

<http://gbif.imr.no/ipt/resour...>



# Log in to GBIF

Login is required to download data

Account

**DISCLAIMER**  
By registering with GBIF you agree to abide by the terms of usage set out [here](#).



**Username \***

**First name \***  **Last name \***

**E-mail address \***

**Country \***

# Dag Endresen

User account and personal settings

Previous data downloads are listed from the user profile

Account

Downloads

## Your downloads

DOWNLOAD [doi:10.15468/dl.h6fdbn](https://doi.org/10.15468/dl.h6fdbn) 20th October 2016

QUERY

*Puccinellia fasciculata* (Torr.) E.P.Bicknell, *Puccinellia maritima* (Huds.) Parl., *Puccinellia distans* (Jacq.) Parl., *Sclerochloa dura* (L.) P.Beauv., *Stipa capensis* Thunb., *Pseudoroegneria spicata*, *Secale sylvestre* Host, *Stephanachne pappophorea* (Hack.) Keng, *Thinopyrum intermedium* (Host) Barkworth & D.R.Dewey, *Elymus tauri* (Boiss. & Balansa) Melderis, *Stipa pappiformis* Keng, *Psathyrostachys fragilis* (Boiss.) Nevski, *Sesleria caerulea* (L.) Ard., *Vahlodea atropurpurea* (Wahlenb.) Fr. ex Hartm., *Pleuropogon refractus* (A.Gray) Benth. ex Vasey, *Triniochloa stipoides* (Kunth) Hitchc., *Pleuropogon sabinei* R.Br., *Stipa juncea* L., *Vulpia fasciculata* (Forssk.) Samp., *Antinoria agrostidea* (DC.) Parl., *Psammochloa villosa* (Trin.) Bor, *Poa billardierei* Steud., *Ptilagrostis dichotoma* Keng ex Tzvelev, *Ptilagrostis mongholica* (Turcz. ex Trin.) Griseb., *Sesleria autumnalis* (Scop.) F.W.Schultz, *Schizachne purpurascens* (Torr.) Swallen, *Dryopoa dives* (F.Muell.) Vickery, *Stipa barbata* Desf., *Taeniatherum caput-medusae* (L.) Nevski, *Poa annua* L., *Trisetum cernuum* Trin., *Poa pratensis* L., *Stenostachys gracilis* (Hook.f.) Connor, *Puccinellia nuttalliana* (Schult.) Hitchc., *Elymus reflexiaristatus* subsp. *strigosus* (Schult.) Melderis, *Polypogon monspeliensis* (L.) Desf., *Elymus stipifolius* (Czern. ex Nevski) Melderis, *Polypogon viridis* (Gouan) Breistr., *Sorghum bicolor* (L.) Moench, *Scolochloa festucacea* (Willd.) Link, *Triticum aestivum* L., *Sphenopus divaricatus* (Gouan) Rchb., *Stipa lagascae* Roem. & Schult., *Rostraria azorica* S.Hend., *Trikeria hookeri* (Stapf) Bor, *Stipa parviflora* Desf., *Triplachne nitens* (Guss.) Link, *Poa alsodes* A.Gray, *Stephanachne nigrescens* Keng, *Triticum turgidum* L., *Vulpia myuros* (L.) C.C.Gmel., *Poa autumnalis* Muhl. ex Elliott, *Wangenheimia lima* (L.) Trin., *Trisetum flavescens* (L.) P.Beauv., *Thinopyrum bessarabicum* (Savul. & Rayss) Á.Löve, *Thinopyrum junceum* (L.) Á.Löve, *Torreyochloa pallida* (Torr.) Church, *Piptochaetium montevidense* (Spreng.) Parodi, *Poa palustris* L., *Ptilagrostis kingii* (Bol.) Barkworth, *Rostraria cristata* (L.) Tzvelev, *Periballia involucrata* (Cav.) Janka, *Koordersiochloa longiarista* (A.Rich.) Veldkamp, *Polypogon fugax* Nees ex Steud., *Trisetum bifidum* (Thunb.) Ohwi, *Polypogon maritimus* Willd., *Piptochaetium*

TAXON

# Occurrence download

doi:10.15468/dl.h6fdbn 20th October 2016

1,101,724  
Occurrences

Download

### PLEASE NOTE

The download result will be retained for as long as possible, but might be removed in the future.  
The [doi:10.15468/dl.h6fdbn](https://doi.org/10.15468/dl.h6fdbn) will always resolve to this page, even if the download is removed.



All downloads are assigned a DOI for citation



## Download details

IDENTIFIER	<b>DOI</b> <a href="https://doi.org/10.15468/dl.h6fdbn">doi:10.15468/dl.h6fdbn</a>
CITE AS	GBIF.org (20th October 2016) GBIF Occurrence Download <a href="http://doi.org/10.15468/dl.h6fdbn">http://doi.org/10.15468/dl.h6fdbn</a>
QUERY	<i>Puccinellia fasciculata</i> (Torr.) E.P.Bicknell, <i>Puccinellia maritima</i> (Huds.) Parl., <i>Puccinellia distans</i> (Jacq.) Parl., <i>Sclerochloa dura</i> (L.) P.Beauv., <i>Stipa capensis</i> Thunb., <i>Pseudoroegneria spicata</i> , <i>Secale sylvestre</i> Host, <i>Stephanachne pappophorea</i> (Hack.) Keng, <i>Thinopyrum intermedium</i> (Host) Barkworth & D.R.Dewey, <i>Elymus tauri</i> (Boiss. & Balansa) Melderis, <i>Stipa pappiformis</i> Keng, <i>Psathyrostachys fragilis</i> (Boiss.) (...)
GEOREFERENCED	<i>true</i>
SIZE	67.0 MB
FORMAT	CSV
STATUS	Succeeded <a href="#">query latest data</a>

# Occurrence download

1,101,724  
Occurrences

doi:10.15468/dl.h6fdbn 20th October 2016

## 820 datasets contributed data to this download

**DATASET** [Marie-Victorin Herbarium \(MT\) - Plantes vasculaires](#)  
**RECORDS** 333 records from this dataset included at time of download  
**IDENTIFIER** [doi:10.5886/rzav8bu2](#)  
**CITATION** Marie-Victorin Herbarium (MT) from University of Montreal Biodiversity Centre. <http://dx.doi.org/10.5886/rzav8bu2> (accessed on [date]).

**DATASET** [National Trust - Wicken Fen nature reserve species data held by The National Trust](#)  
**RECORDS** 46 records from this dataset included at time of download  
**IDENTIFIER** [doi:10.15468/iqeemg](#)  
**CITATION** UK National Biodiversity Network: National Trust - Wicken Fen nature reserve species data held by The National Trust

**DATASET** [OEH Atlas of NSW Wildlife](#)  
**RECORDS** 4,137 records from this dataset included at time of download  
**IDENTIFIER** [doi:10.15468/14jd9g](#)  
**CITATION** Office of Environment and Heritage, Department of Premier and Cabinet representing the State of New South Wales: OEH Atlas of NSW Wildlife

**DATASET** [A global database for the distributions of crop wild relatives](#)  
**RECORDS** 28,096 records from this dataset included at time of download  
**IDENTIFIER** [doi:10.15468/jyrthk](#)  
**CITATION** Global Consortium of Crop Wild Relative Occurrence Data Providers. 2015. A global database for the distributions of crop wild relatives. Scientific data (in prep)

PLEASE NO  
 The downlo  
 The doi:10.1

### Downlo

IDENTIFI  
 CITE AS  
 QUERY

SIZE  
 FORMAT  
 STATUS

# **Data citation with DOI**



# Scholarly Data citation is good research practice

- **Data citation** integrated part of the scholarly system
- Supporting research **validation** and **reproducibility**
- **Enable reuse** of research output and results
- **Tracking impact** of published data
- Scholarly structure to **reward and credit people** involved in producing research data



# DOI – Digital Object Identifier

- Persistent resolvable identifiers
- Standard for scholarly journal papers
- Simplify citation of references
- Used for **measuring impact**



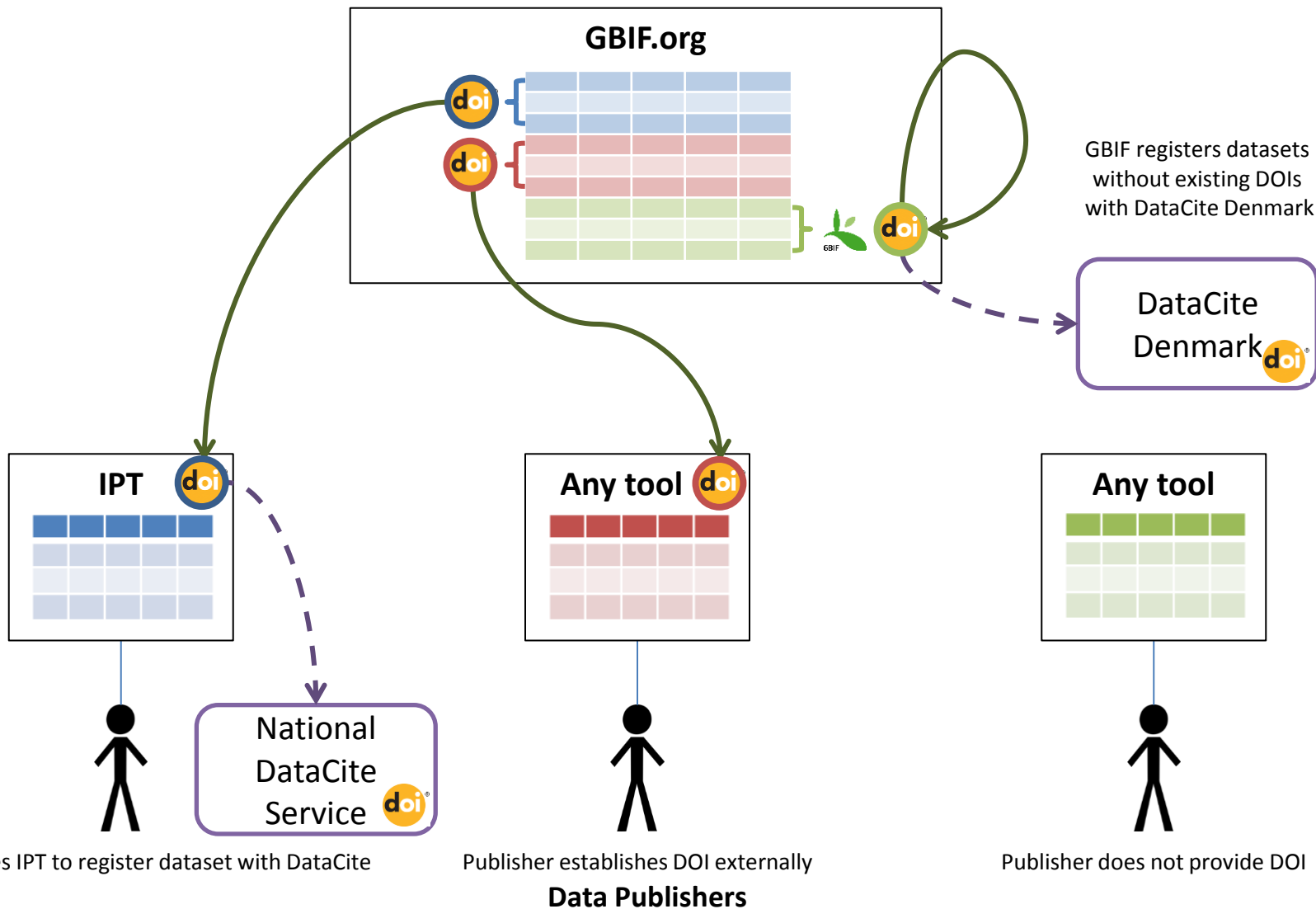
# Who has cited your data resource?

- CrossRef services link citations between scientific journal articles
- *With CrossRef you can discover who has cited your research paper!*
  
- Services to link dataset citations (DataCite DOI) with CrossRef systems
- *Towards services in CrossRef to discover who cited your research data ...*

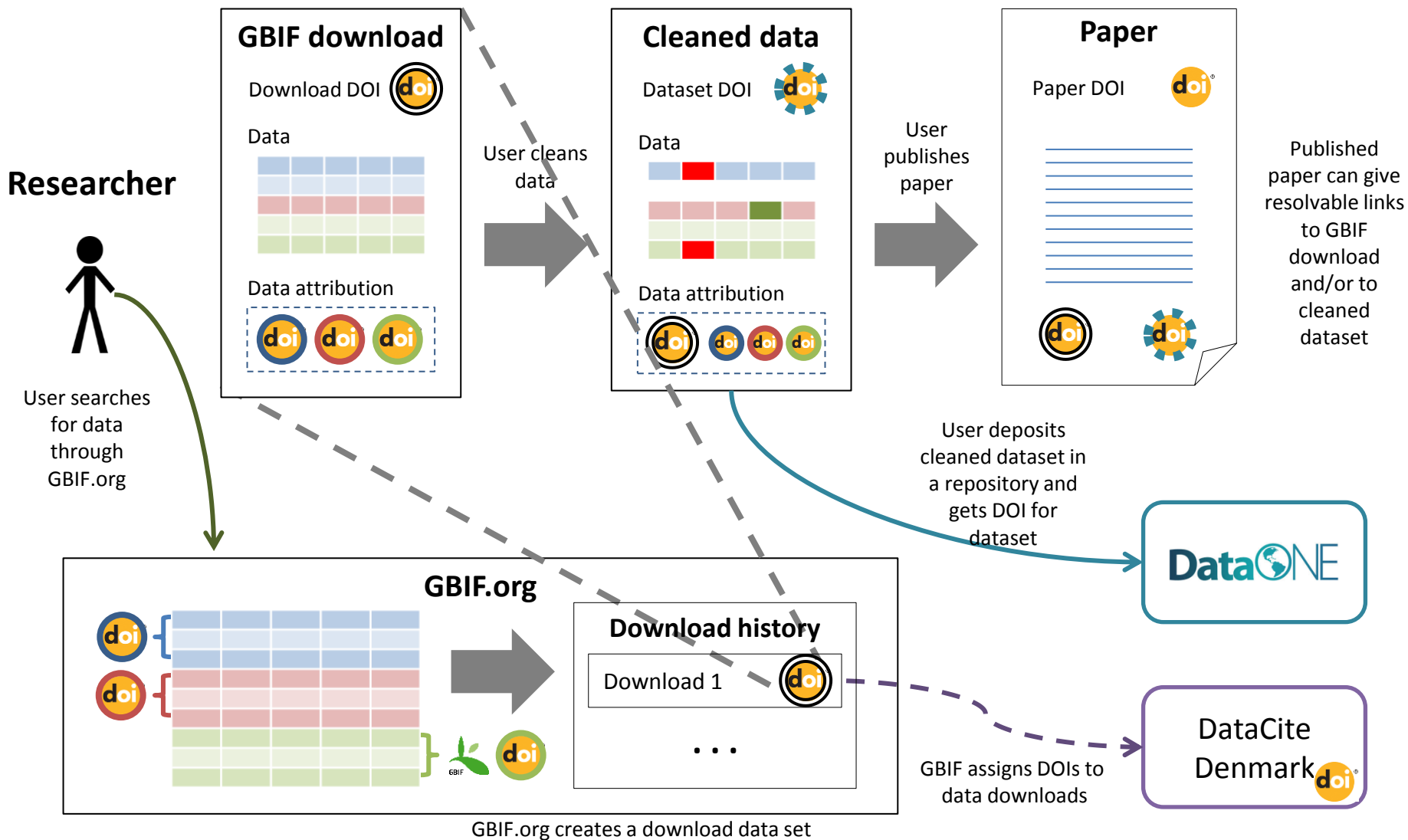


<http://blog.crossref.org/?s=DataCite>

# GBIF & DIGITAL OBJECT IDENTIFIERS (DOI)



# GBIF & DIGITAL OBJECT IDENTIFIERS (DOI)



# BIBSYS is the Norwegian member in DataCite



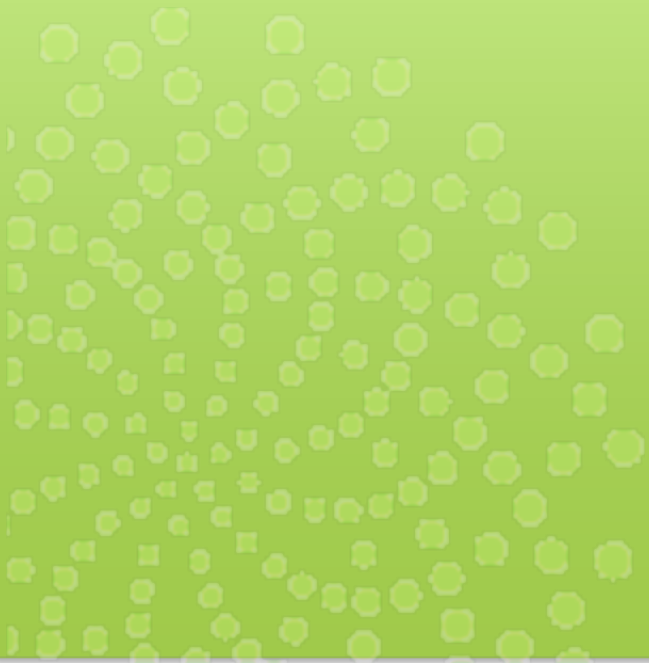
- "As a *DataCite* member **BIBSYS** act as a national DataCite representative in *Norway* and thereby allow all of Norway's higher education and research institutions to use *DOI on their research data*".

- Resource must be citation-worthy
- Resource must have good metadata
- Long-term archiving and accessibility for data

- <http://www.bibsys.no/rammeverk-for-doi/>



# **Publishing platforms And Data papers**



# Publishing platforms and archives

- **GBIF => Connected** (through IPT, DwC-A, BioCASE, TAPIR or DiGIR protocols)
  - Member institutions databases or DwC-Archives
  - Data networks using IPT
  - Biodiversity Data Journal – Pensoft “data papers” <http://bdj.pensoft.net/>
- **Not Yet Automatically Connected:** (You can place repository ID:s to DwC occurrence data)
  - BOLD <http://www.boldsystems.org/index.php/databases>
  - GenBank / INSD <http://www.insdc.org/>
  - DataOne <https://search.dataone.org/>
  - Dryad, Eudat, ...
- **Not Connected:**
  - Scientific Data – Nature <http://www.nature.com/sdata/>
  - JSTOR (GPI Virtual Herbarium) Not open access!

# Analyses platforms





# Analysis platforms read GBIF DATA PORTAL API

An interface to access  
data published through  
the GBIF network using  
**web services.**

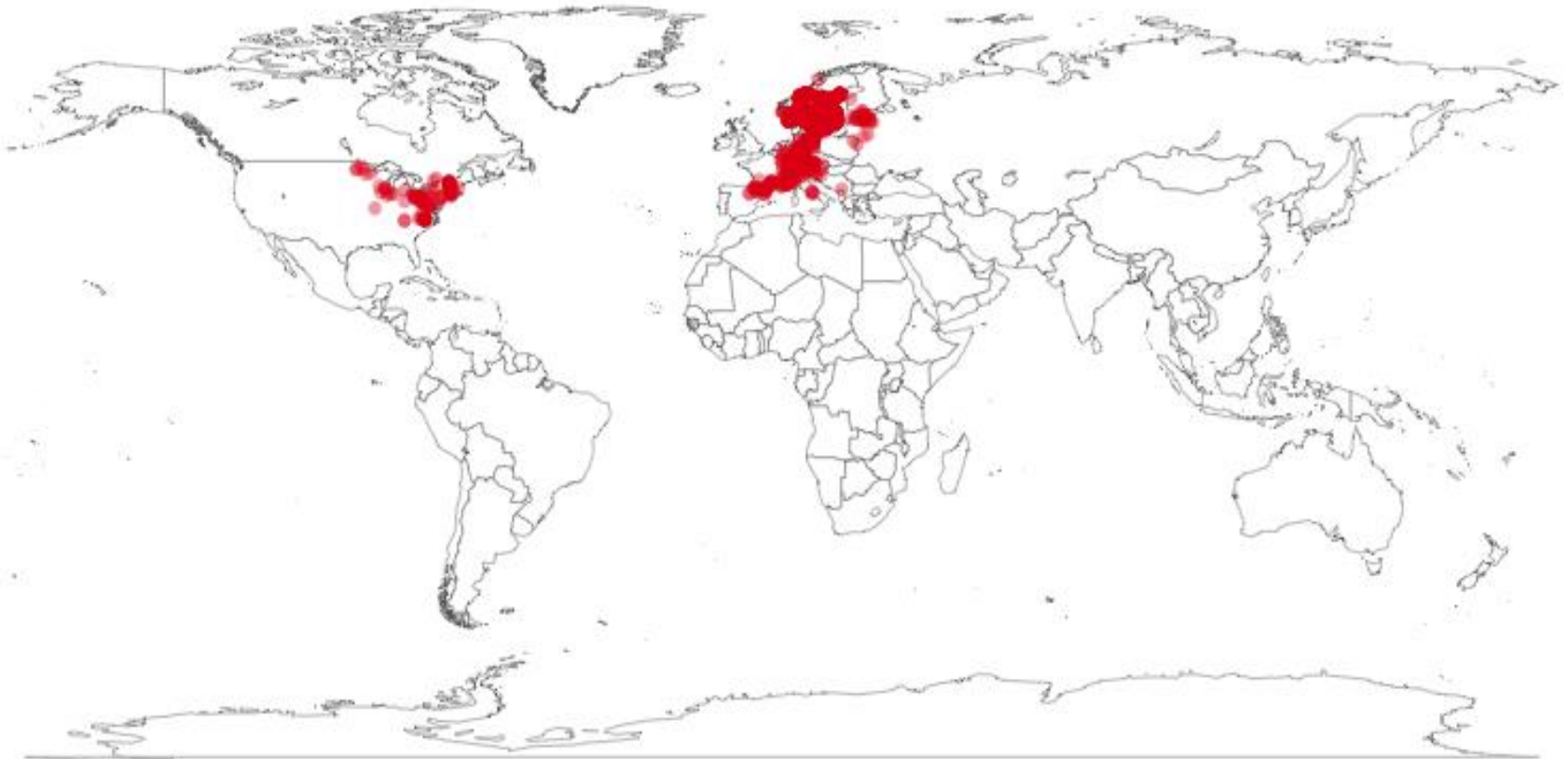
<http://api.gbif.org/v1/>



# rOpenSci : rgbif



- `library(rgbif)`
- `key <- name_backbone(name='Hepatica nobilis', kingdom='Plantae')$speciesKey`
- `sp <- occ_search(taxonKey=key, return='data', hasCoordinate=TRUE, limit=1000)`
- `gbifmap(sp)`



# Analyses tools and platforms

- LifeWatch

<https://www.analysisportal.se/>

- Ignorance maps

<http://alejandroruete.github.io/IgnoranceMaps/>

- Zonation

<http://cbig.it.helsinki.fi/software/zonation/>

# Identifiers?

- DOIs for datasets and downloads
  - Citation
  - Cross ref
- Linked Open Data
  - Identifiers for all **data items** and **vocabularies**
  - URI-format <http://tun.fi/MX.37890>

# References:

- GBIF Secretariat (2015). Getting Started: **An overview of data publishing in the GBIF network**, version 1.1. Copenhagen: Global Biodiversity Information Facility, 17 pp. ISBN: 87-92020-28-3 (for version 1.0). Available online at: <http://www.gbif.org/resource/80635>
- <http://biodiversity-informatics-training.org/>
- <http://gbif.ru/> GBIF Russia website
- GBIF Documents in Spanish: [http://www.gbif.org/resources?f\[0\]=gr\\_language%3A1074&searched=1](http://www.gbif.org/resources?f[0]=gr_language%3A1074&searched=1)