



# **GBIF** for research stations

**Dmitry Schigel**GBIF Scientific Officer



Photo: skoppelo , Kilpisjarvi station, Finland <a href="https://www.flickr.com/photos/skoppelo/45281740/in/photolist-jjHSPg-515FE">https://www.flickr.com/photos/skoppelo/45281740/in/photolist-jjHSPg-515FE</a>

## The research data lifecycle

Generate / Access

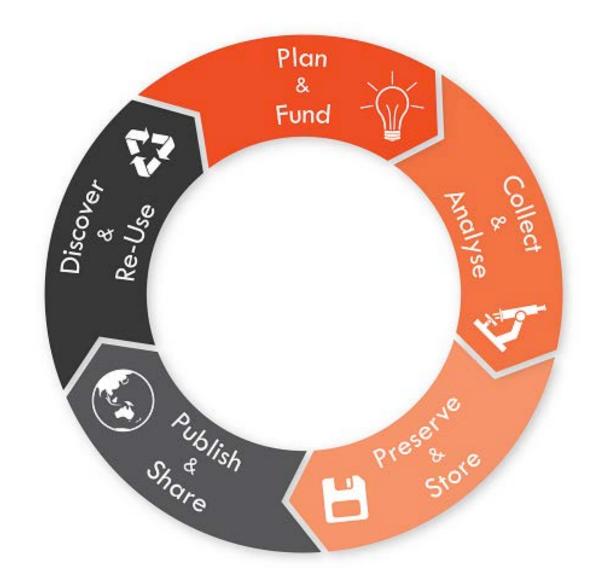
(re)Organize

Modify

**Analyze** 

**Archive** 

Cite





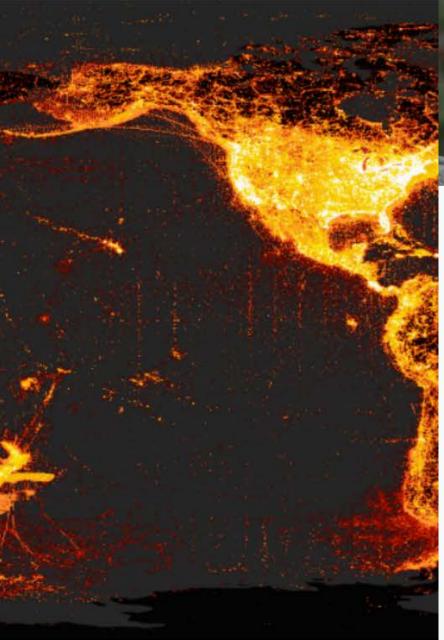
### All samples and specimens will be destroyed at some point

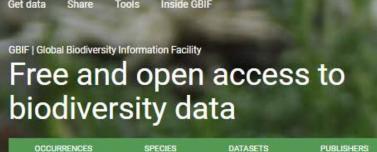
maybe data, too, but digitization increase the use and lifetime





# **GBIF.org**





Search

WHAT IS GBIF? ABOUT GBIF RUSSIAN FEDERATION

0ccurrence records 1,335,543,292



Belarus extends GBIF's European membership map eastward 16 July 2019

Datasets 45,635



Minimizing biodiversity loss in the Brazilian Cerrado 2 July 2019

Publishing institutions 1,439

RESOURCES



Science Review 2019
Stay up to date on the latest research investigations enhanced and supported by free and open access to biodiversity data.

Peer-reviewed papers using data 3,758



Programme seeks Biodiversity Open Data Ambassadors to expand best practices

10 July 2019



BIFA funds nine Asian data mobilization projects 26 June 2019



Atlantodesmus sierwaldae sp. nov.
Cladistic analysis and description of a new
species of the Brazilian genus Atlantodesmus
Hoffman, 2000 (Diplopoda: Polydesmida:
Chelodesmidae)



The GBIF network

Dozens of countries and organizations working together to make species data findable, accessible, interoperable and reusable.



Establishing a national biodiversity information facility in Chile

Experiences of setting up and running a node in Latin America



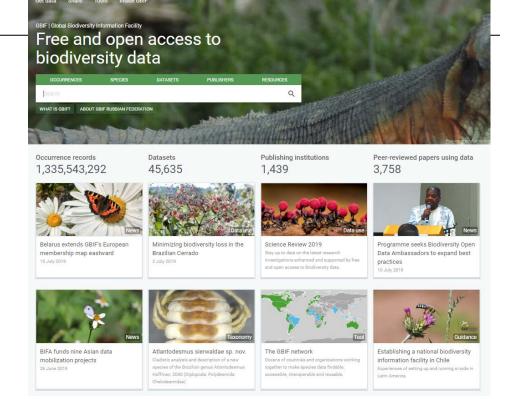
Intergovernmental open data infrastructure

Funded by the governments of the participant countries

Network for free and open access to biodiversity data

96 participants:

58 countries and 38 organisations

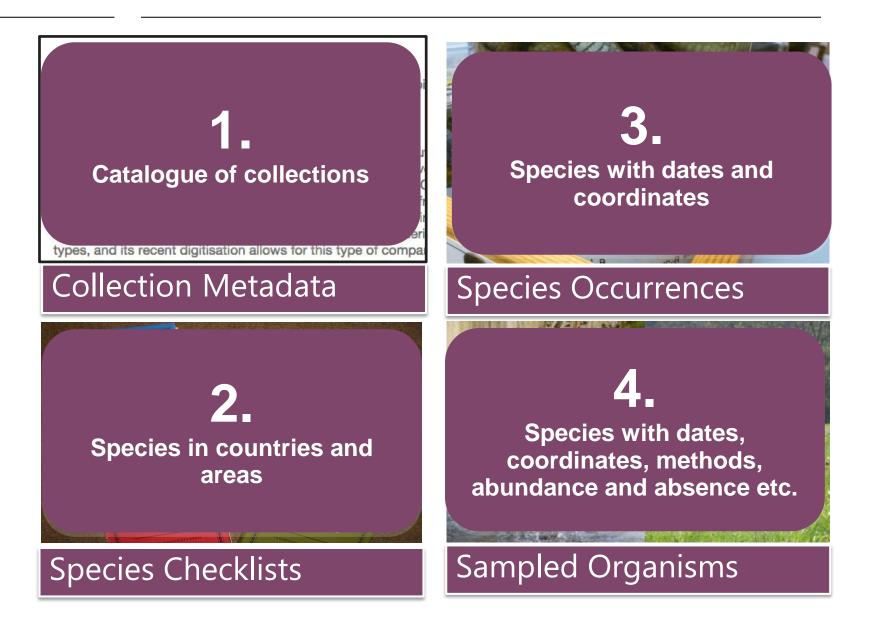


What species?
Where was it found?
When was it found?
What is the evidence?

Specimen data
Sampling event information
Sequences, images, etc.



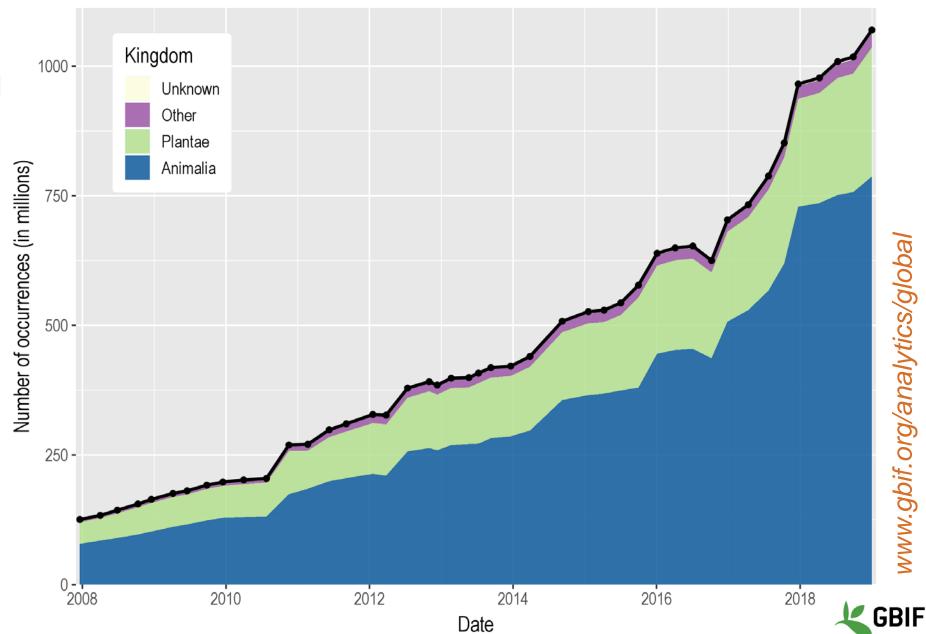
# Data richness levels supported by GBIF





#### Species occurrence records accessible through GBIF over time

# Data published through GBIF.org





# BY THE NUMBERS 29 Aug 2019

Species occurrence records

1,338,052,443

Datasets

45,921

Country
Participants

Organizational Participants

**58** 

38

**Publishers** 

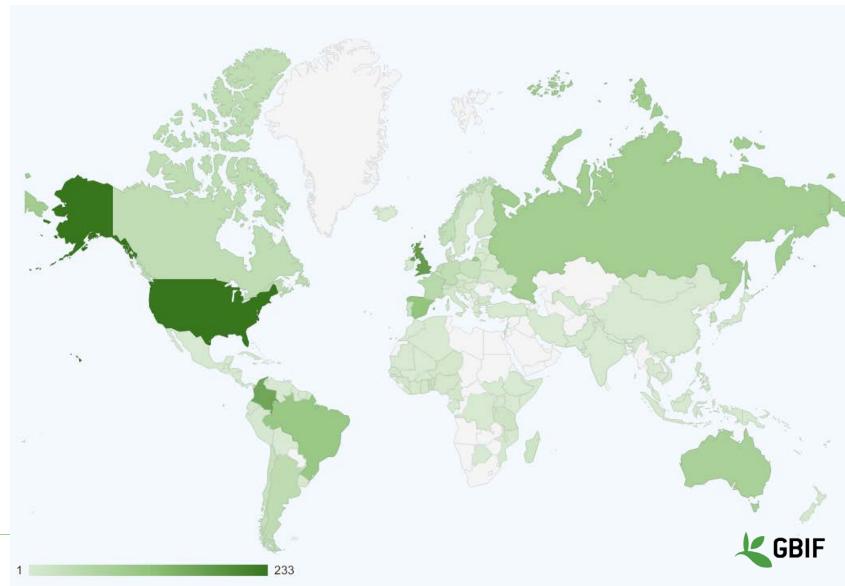
1,445

GBIF data publishers: organizations

*30 Aug 2019* 

129 countries and territories with GBIF publishers

115 countries and territories publish data

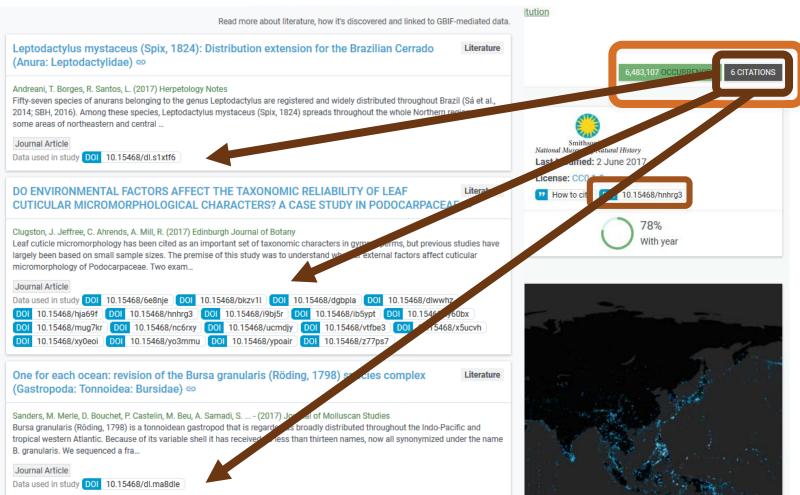


# Data citation: tracking and display

Get data Share Tools Inside GBIF

OCCURRENCE DATASET | REGISTERED 13 OCTOBER 2016

#### NMNH Extant Specimen Records





# GBIF enabled science today: peer-reviewed publications

Agriculture

Aliens and invasives

Biogeography

Citizen science

Climate change

Conservation

Data management

Ecology

Ecosystem services

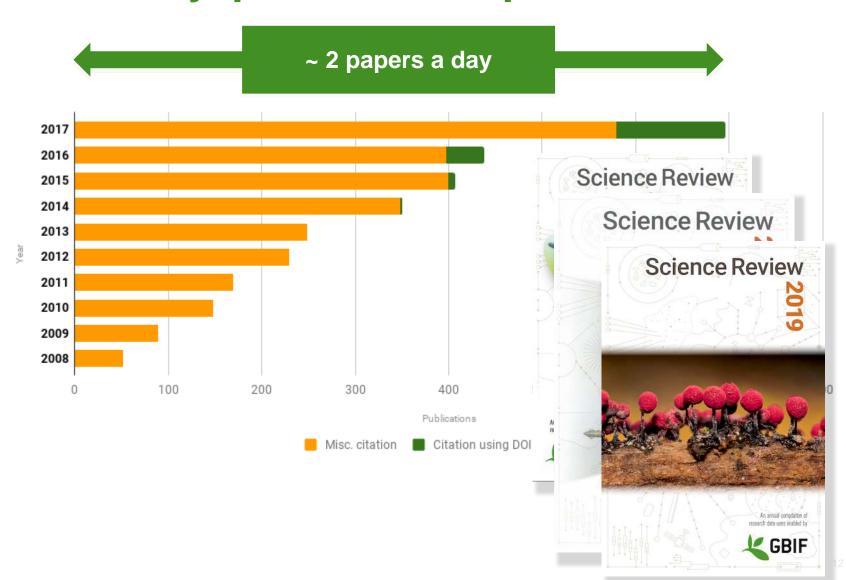
**Evolution** 

Human health

Species distributions

Phylogenetics

**Taxonomy** 



https://www.gbif.org/science-review

## **Benefits of openess**

OPEN & ACCESS Freely available online



#### Sharing Detailed Research Data Is Associated with Increased Citation Rate

Heather A. Piwowar\*, Roger S. Day, Douglas B. Fridsma

Department of Biomedical Informatics, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, United States of America

Reackgrowned. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or the data evaliable. Privilegal Fainingic. We examined the clatation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 46% of trials with publicly available data was significantly (p = 0.000) associated with microarray data received 85% of the aggregate citations. Publicly available data was significantly (p = 0.000) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. Significancer. This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Pierowar HA, Day RS, Fridsma DB (2007) Sharing Detailed Research Data is Associated with Increased Citation Rate, PLoS ONE 2(3): e308, doi:10.1371/journal.pone.0000308

#### INTRODUCTION

Sharing information facilitates science. Publicly sharing detailed research data-sample attributes, finited factors, patient outcomes, DNA regencies, raw mRNA microarray measurements-with other researchers allows these valuable resources to contribute far beyond their original analysis[1]. In addition to being used to confirm original results, are what can be used to explore related or new hypotheses, particularly when combined with other publicly available that sees. Red data is indispensable when investigating and developing study methods, analysis techniques, and software implementations. The largue scientific community also benefits tharing data encourages multiple perspectives, helps to identify and increases efficient use of fanding and patient population resources by avoiding duplicate data collection.

Believing that the these benefits converging the costs of sharing

Refereing that these benefits outweight as costs of sharing. Refereing that these benefits outweight as costs of sharing to the control of the control of the control of the cost of sharing their data available. Some journal, including the PLoS family, require the submission of detailed bismerical data to publish available databases as a condition of publishment Size 2003, the NIII has required a data sharing plan for all large familing grants. The growing open-screen publishing movement will certain sixtense peer prosume to share data.

However, while the general research community benefits from shared data, much of the baselen for sharing the data falls to the study investigator. Are there benefits for the investigators themselves?

A currency of value to many investigators is the number of times their publications are cited. Although limited as a proxy for the scientific contribution of a paper[5], citation counts are often used in research funding and promotion decisions and have even been autigned a sulary-increase follar value [6]. Bootting citation rate is thus is a potentially supportunt motivator for publication authors.

In this study, we explored the relationship between the citation and whether it date was made publicly available. Using cancer microarray clinical trials, we addressed the following questions. Do trials which have their microarray discharacter microarray clinical trials, we addressed the following questions: Do trials which have their microarray clinical arcicles once (citation 20 trials true even within lower profile trials). What other data-daring variables are associated with an investigate causation, and White this mody in red able to investigate causation, these relationships. Clinical microarray data provides a united existence to first investigation despite being valuable for read extensively only to collect, in no yet universally absence.

#### RESULTS

We midied the citations of 85 cancer microarray clinical trials published between January 1969 and April 2003, a sidentified in a systematic review by Nirani and Isaamiolis[7] and Isaed in Supplementary Texa St. We found at 1 of the the clinical trials Supplementary Texa St. We found at 1 of the the clinical trials interest. Most data sets were located on lab verbries [280, with a few found on publisher websites [10], on within public distubates [6] in the Sandord Microarray Database [SMD][8], 6 in Gere Expression Dambous (EGD[9]), 2 in the NCI GeneExpression Dam Israel (GEDD[8]), 6 in Aprophysical (Sandassas Interest) and the Company of the C

authors were more many to many them cata.

The cobort of 18 trials was cited an aggregate of 6299 times in 2001-2005 by 3135 distinct articles (median of 1.0 cobort citation per article, range 1-23). The 44% of trials which sharred their data received a total of 5334 citations (85% of aggregate), distributed as shown in Figure 1.

Academic Editor: John loannidis, University of loanning School of Medicine, Greene

Received December 13, 2006; Accepted February 26, 2007; Published March 21, 2007

Copyright: © 2007 Piwower et al. This is an open-access article distributed under the terms of the Creative Commons Attribution Ucense, which permits unnesticated use, distribution, and expoduction in any medium, provided the original author and source are created.

Funding HAP was supported by MURT fairing Grant Number ST15-LM00195-19. The NBH had no role in shally design, data collection or analysis, writing the paper, or the decision to salarial it for publication. The publication centeria are solely the responsibility of the authors and do not necessarily represent the official views of the RBH.

Competing Interests: The authors have declared that no competing interest

\*To whom correspondence should be addressed. E-mail: hp/wcmarthcbm/.p

March 2007 | Issue 3 | e308

PLoS ONE | www.plosone.org

Increases the efficiency of research

Promotes scholarly rigor and quality of research

Enables tracking of data use and data citation through DOIs

Expands the spectrum of academic products through data papers

Enhances visibility and scope for engagement

Enables researchers to ask new research questions

Enhances collaboration and community-building

Increases the economic and social impact of research

International conventions and requirements from funding agencies



# Research stations in GBIF: increase visibility, collaboartion, impact through data

Only <50 stations represented in GBIF worldwide

Masses of high-quality and longterm data

Opportinity to get discovered through data searches

Measure research value also through data products and data citation

#### **Archbold Biological Station**

Joined 2 years ago

The mission of Archbold Biological Station is to build and share the scientific knowledge needed to protect the life, lands, and waters of the heart of Florida and beyond. Archbold is located in Venus...

1 dataset United States of America 2 citations

#### **Biological Field Station of Paimpont, University Rennes 1**

New publisher

Joined 2 months ago

Located in a rural, agricultural region in continental Brittany, North-West France, the Biological Field Station of Paimpont (Station Biologique de Paimpont, SBP) offers since 1967 an excellent outdoo...

1 dataset France

#### Station d'Ecologie de Lamto

Joined 2 years ago

The Ecology Station of Lamto is an ecological research station located in the center of Côte d'Ivoire. It is attached to Nangui Abrogoua University of Abidjan. Since its foundation, this station has b...

1 dataset Côte d'Ivoire

#### **Taiwan Forestry Bureau**

Joined 4 years ago

In order to manage the country's national forest and achieve the goal of management for multiple uses, the Bureau headquarters consists of five division and four offices. Responding to changes in the ...

4 datasets Taiwan 3 citations

#### Nikolai Pertsov White Sea Biological Station

Joined a year ago

WSBS MSU (Nikolai Pertsov White Sea Biological Station) is an educational and research centre, created for conducting marine scientific research and field student practices at the White Sea. WSBS is a...

Russian Federation

#### **Charles Darwin Research Station**



PUBLISHER | SINCE APRIL 26, 2018

## Research stations in GBIF Nikolai Pertsov White Sea Biological Station

⇔ HOME PAGE

PUBLISHER | SINCE APRIL 4, 2017

# **Archbold Biological Station**

ABOUT

METRICS

CO HOME PAGE

34,692 OCCURRENCES

1 DATASET

2 CITATIONS

Description: The mission of Archbold Biological Station is to build and share the scientific knowledge needed to protect the life, lands, and waters of the heart of Florida and beyond. Archbold is located in Venus, south-central Florida. Since inception in 1941, Archbold Biological Station has prioritized the development and curation of an on-site, multi-taxon, specimen-based, natural history collection. The collection of specimens is necessary for research at the Station and for outside investigators, emphasizing two essential activities the identification of species and documentation of biodiversity. After decades of steady growth the Archbold collection includes more than 250,000 well-preserved, and well-labeled specimens of plants, birds, fish, herptiles, mammals and arthropods. The Archbold collection is probably unrivalled in scope and size among biological field station collections in North America, and is likely one of the largest on-site collections encompassing the taxonomic diversity of a single (3,577-ha) site in the U.S.A. Our diverse natural history reference collection is a key component of the Station's infrastructure, serving a broad community of staff researchers, visiting investigators and students, and supplying collection material and information to outside investigators. Plant specimens have been used in studies of community ecology, such as the response of vegetation to fire. The vertebrate collection was designed for studies of variation, growth patterns, life histories, and population dynamics of local vertebrates. The arthropod collection contributes to numerous studies needing insect identification, as well as providing large numbers of specimens with ecological data.

Endorsed by: U.S. Geological Survey

Administrative contact: Stephanie Leon

Technical contact: Vivienne Sclater

Country or area: United States of America



# INTERACT: how to be visible in GBIF and publish data?

1. Work with CAFF and get your data published through CAFF: professional help

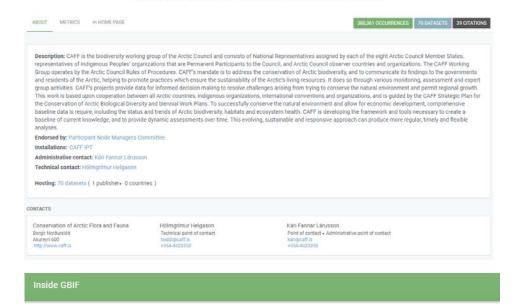
Kári Fannar Lárusson Point of contact Administrative point of contact kari@caff.is

#### OR

1. Register own GBIF account and go ahead: independent statistics and citations

https://www.gbif.org/publishing-data

#### PUBLISHER | SINCE SEPTEMBER 28, 2015 Conservation of Arctic Flora and Fauna



SHAB

#### Become a publisher

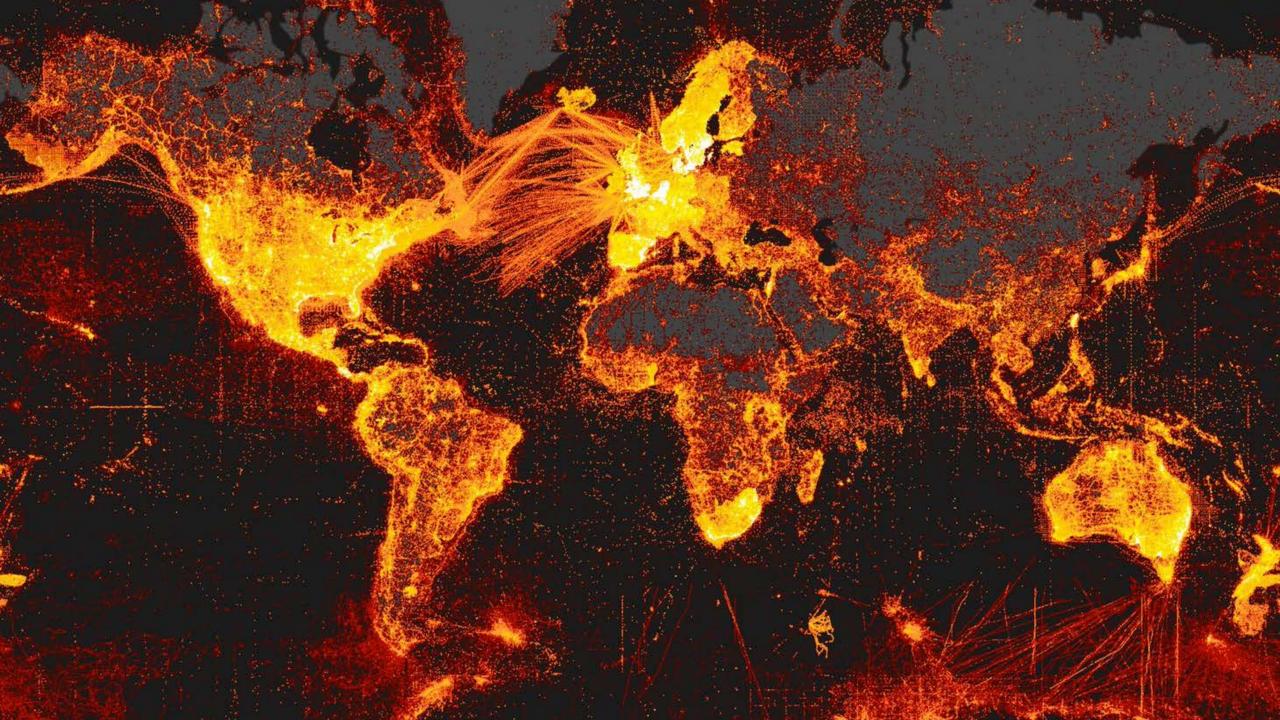
Organizations wishing to share data through GBIF can register here to request endorsement as a data publisher

To prepare for publishing data through GBIF, we ask prospective new publishers to complete the online form below.

Your answers will help us to give proper credit and attribution for the datasets you share. They will also help users to understand more about the provenance of data shared through the GBIF network.

Before GBIF indexes your datasets, your institution must receive endorsement as a data publisher from one of the Participant nodes that coordinate activities of the national and organizational Participants in the GBIF network. If your country is not yet participating in GBIF, we will seek endorsement on your behalf from the GBIF community.







# 風の谷のナケシナ

# Thank you

**Dmitry Schigel**Scientific Officer

**GBIF** 

Global Biodiversity Information Facility (GBIF)

Secretariat
Universitetsparken 15
DK-2100 Copenhagen Ø
DENMARK

dschigel@gbif.org