



FÓRUM  
GESTÃO DE DADOS  
DE INVESTIGAÇÃO

8.º Fórum GDI

Curadoria de Dados em Repositórios:  
fluxos de trabalho e serviços

# Serviços temáticos: GBIF - the Global Biodiversity Information Facility

Rui Figueira, [ruifigueira@isa.ulisboa.pt](mailto:ruifigueira@isa.ulisboa.pt),  
GBIF Portugal, Instituto Superior de Agronomia



# GBIF

GBIF—the Global Biodiversity Information Facility — is an international network and data infrastructure funded by the world's governments and aimed at **providing anyone, anywhere, open access to data about all types of life on Earth.**

**[www.gbif.org](http://www.gbif.org)**



*Palinurus elephas* (Fabricius, 1787) Observed in Portugal  
by Ana Santos (licensed under <http://creativecommons.org/licenses/by-nc/4.0/>)

# GBIF BY THE NUMBERS | 18 NOVEMBER 2021

Species occurrence records

**1,901,456,957**



Country  
Participants

**61**

Organizational  
Participants

**40**

Avg records downloaded per month (2021)



**83.4 billion**

Datasets

**63,515**



Data-publishing  
institutions

**1,760**

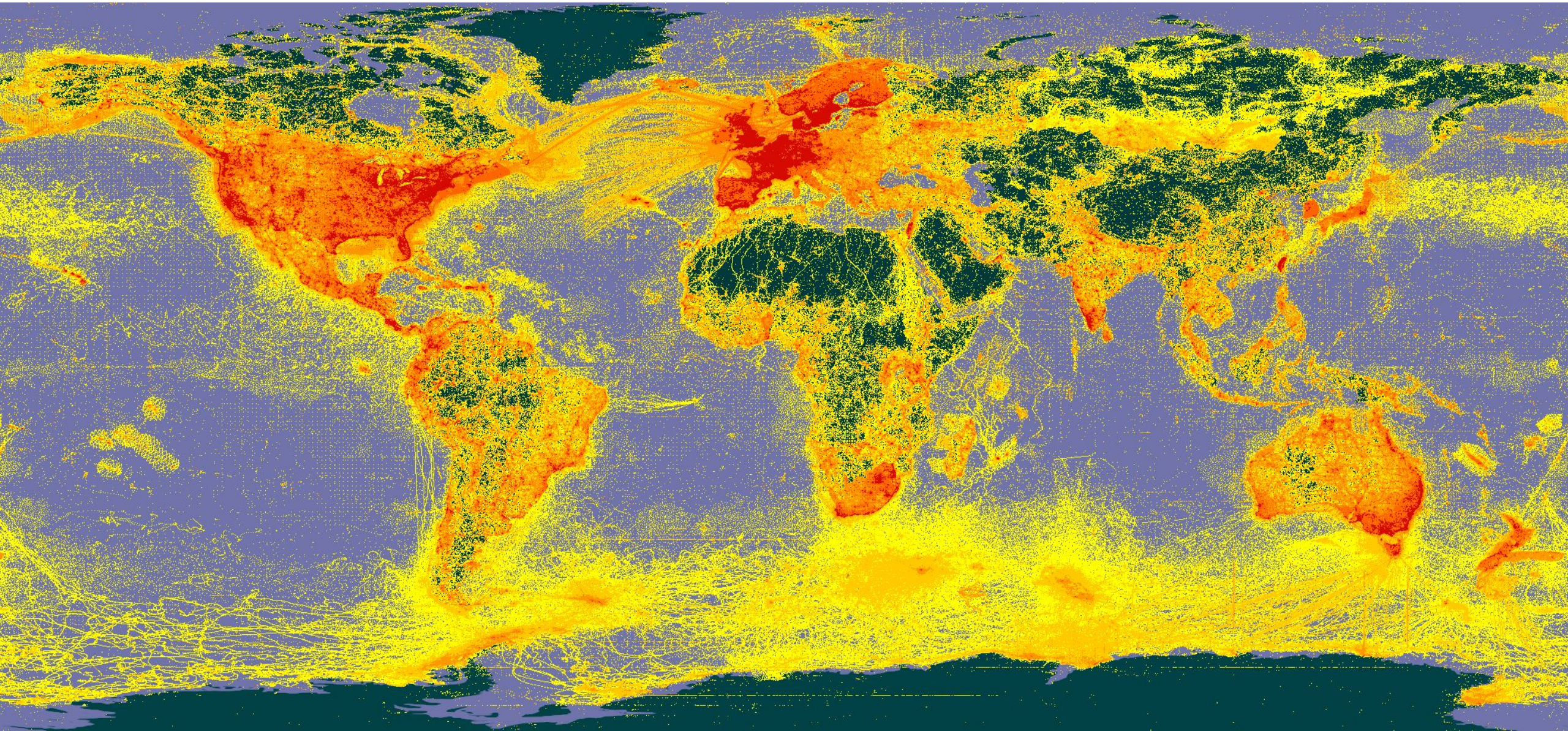
Peer-review papers  
using data

**6,429**



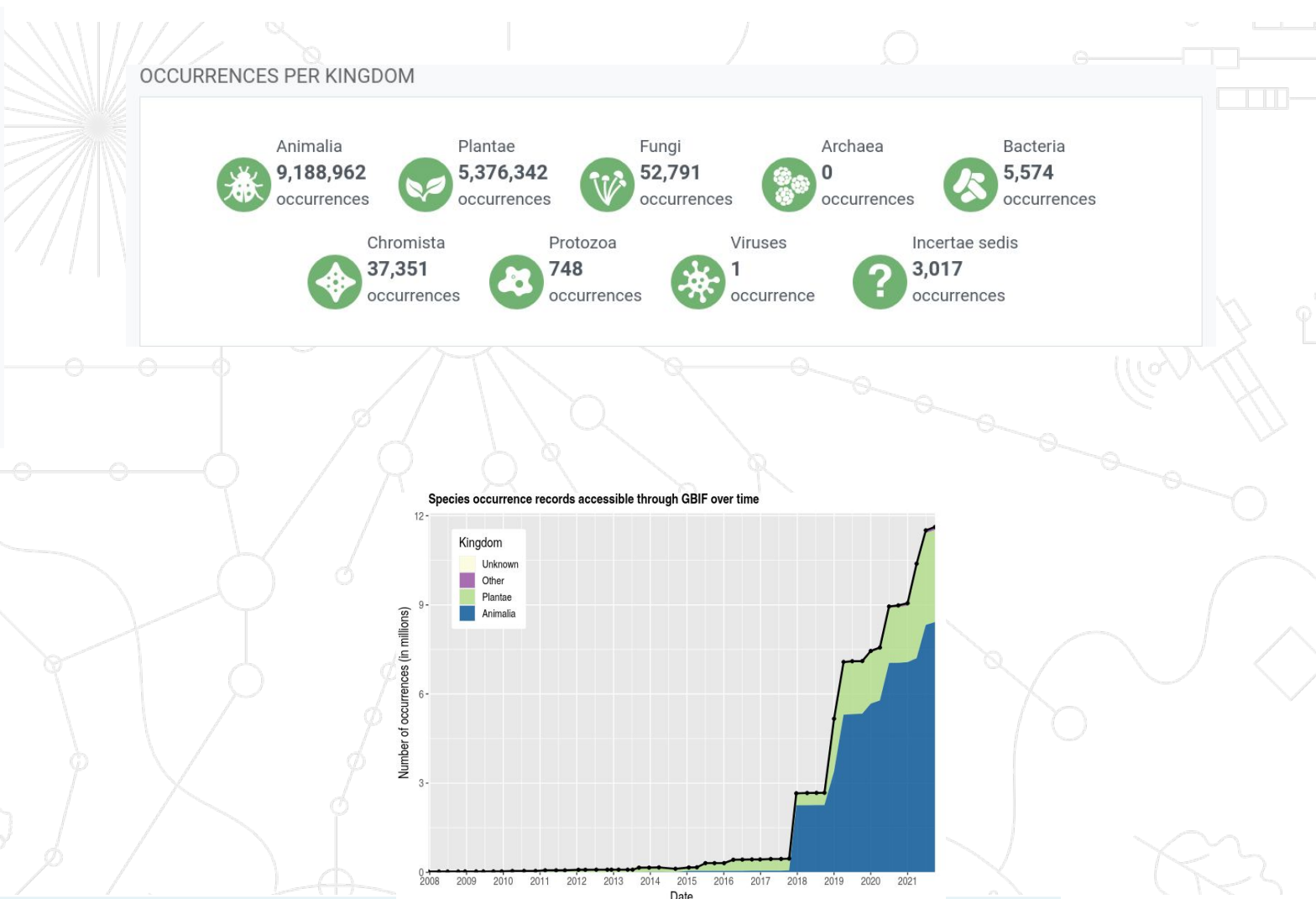
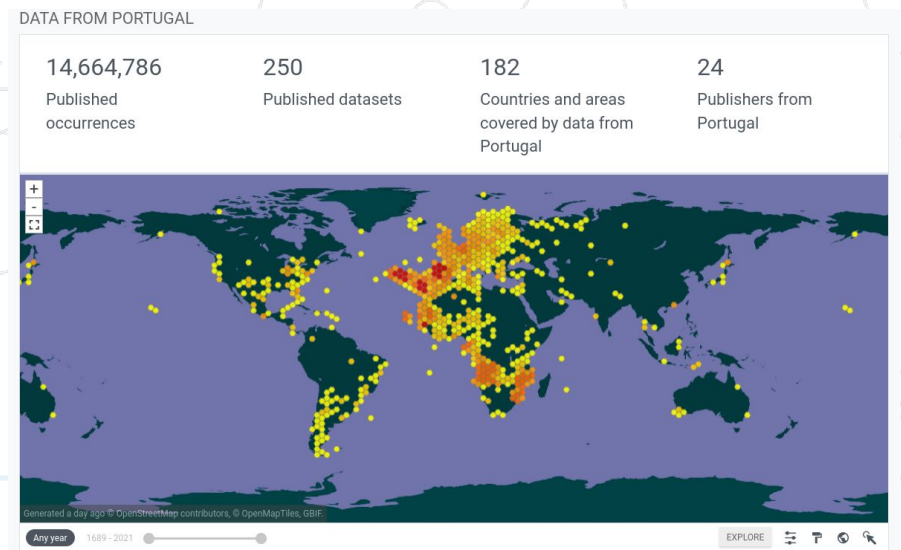


# DATA FROM THE GBIF NETWORK 30 Sept 2021 (publishing data since 2003)





# DATA TO/FROM PORTUGAL 18 Nov 2021



# GBIF MEDIATED DATA AND USES

Projects / EIA / NHM /  
Env DNA / Citizens



Photo by Braño on Unsplash

All environments



All ecologies



Many uses



# GBIF NETWORK OPERATIONAL UNITS

## SECRETARIAT

- Global coordination
- Infrastructure (global platform, tools)
- Capacity enhancement (training, documentation)
- Engagement

## PARTICIPANT NODES

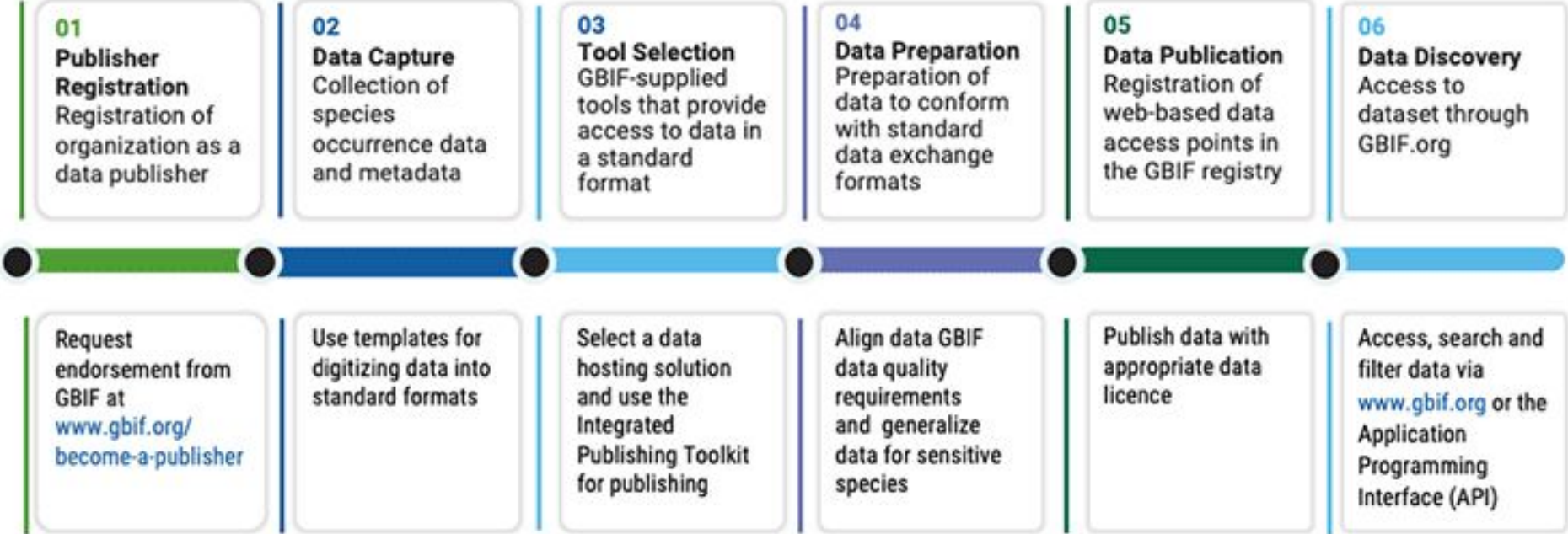
- National coordination
- Infrastructure (IPT, data portals)
- Training, documentation
- Institution engagement and endorsement

## INSTITUTIONS

- Publisher registration
- Data mobilization (digitization, data quality)
- Data standardization and meta-documentation
- Data publication



# DATA PUBLICATION WORKFLOW



ROLES	PUBLISHER	register	digitize	IPT, other tool	standardize	publish	
	NODE	endorse		IPT infrastructure	training	IPT infrastructure	
	SECRETARIAT	register			documentation	Infrastructure, register, index	Infrastructure, share



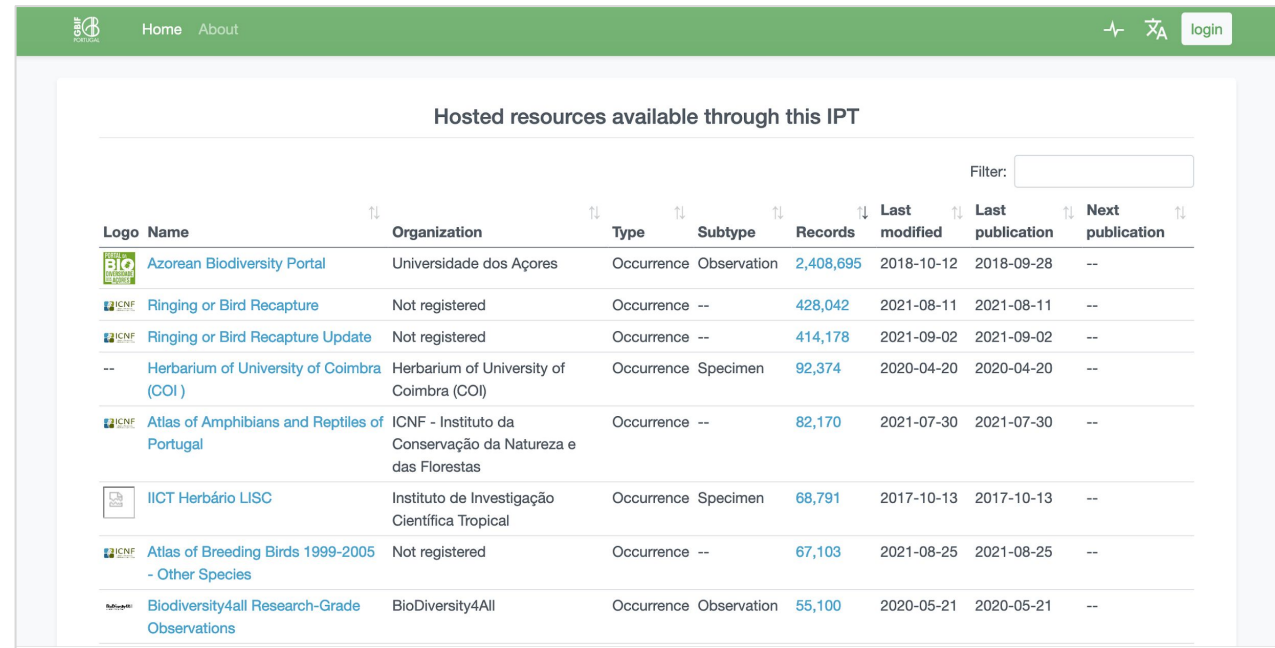
# DATA PUBLICATION WORKFLOW

03








## Tool Selection

### IPT – Integrated Publishing Toolkit

- Opensource, developed by GBIF Secretariat
- Data upload in standard format (Darwin Core)
- Data documentation (metadata) in Ecological Metadata Language (EML) – standard for metadata
- Standard License (CC0, CC-BY, CC-BY-NC)
- Wraps data and metadata in a fully documented Darwin Core Archive file (compressed)
- Supports versions for each dataset update
- Enables assigning DOIs to datasets
- 305 installations in 74 countries worldwide



The screenshot shows the GBIF IPT website interface. At the top is a green navigation bar with the GBIF logo, 'Home', 'About', and a 'login' button. Below the navigation bar is a section titled 'Hosted resources available through this IPT'. This section contains a table with columns for Logo, Name, Organization, Type, Subtype, Records, Last modified, Last publication, and Next publication. A 'Filter:' input box is located at the top right of the table. The table lists several datasets, including the Azorean Biodiversity Portal, Ringing or Bird Recapture, Herbarium of University of Coimbra (COI), Atlas of Amphibians and Reptiles of Portugal, IICT Herbário LISC, Atlas of Breeding Birds 1999-2005 - Other Species, and Biodiversity4all Research-Grade Observations.

Logo	Name	Organization	Type	Subtype	Records	Last modified	Last publication	Next publication
	<a href="#">Azorean Biodiversity Portal</a>	Universidade dos Açores	Occurrence	Observation	2,408,695	2018-10-12	2018-09-28	--
	<a href="#">Ringing or Bird Recapture</a>	Not registered	Occurrence	--	428,042	2021-08-11	2021-08-11	--
	<a href="#">Ringing or Bird Recapture Update</a>	Not registered	Occurrence	--	414,178	2021-09-02	2021-09-02	--
--	<a href="#">Herbarium of University of Coimbra (COI)</a>	Herbarium of University of Coimbra (COI)	Occurrence	Specimen	92,374	2020-04-20	2020-04-20	--
	<a href="#">Atlas of Amphibians and Reptiles of Portugal</a>	ICNF - Instituto da Conservação da Natureza e das Florestas	Occurrence	--	82,170	2021-07-30	2021-07-30	--
	<a href="#">IICT Herbário LISC</a>	Instituto de Investigação Científica Tropical	Occurrence	Specimen	68,791	2017-10-13	2017-10-13	--
	<a href="#">Atlas of Breeding Birds 1999-2005 - Other Species</a>	Not registered	Occurrence	--	67,103	2021-08-25	2021-08-25	--
	<a href="#">Biodiversity4all Research-Grade Observations</a>	BioDiversity4All	Occurrence	Observation	55,100	2020-05-21	2020-05-21	--

<https://github.com/gbif/ipt>

# DATA PUBLICATION WORKFLOW

04

## Data Preparation

### Data standard - Darwin Core

- Maintained by the Biodiversity Information Standards (TDWG) organization
- Contains in the core more than 150 terms arranged in nine classes to describe all aspects of biodiversity occurrence data
- Can be expanded to support other data classes with extensions created by the community and managed by GBIF
- GBIF use of the DwC requires unique identifiers for records
- GBIF (Secretariat and nodes) provide tools, manuals and training to help publishers

The screenshot displays the Darwin Core website interface. On the left, a sidebar lists various data classes: Record-level, Occurrence (highlighted), Organism, MaterialSample, Event, Location, GeologicalContext, Identification, Taxon, MeasurementOrFact, ResourceRelationship, UseWithIRI, LivingSpecimen, PreservedSpecimen, FossilSpecimen, HumanObservation, and MachineObservation. The main content area is divided into three sections: Record-level, Occurrence, and Event. Each section contains a list of terms in a grid format. The Record-level section includes terms like type, modified, language, license, rightsHolder, accessRights, bibliographicCitation, references, institutionID, collectionID, datasetID, institutionCode, collectionCode, datasetName, ownerInstitutionCode, basisOfRecord, informationWithheld, dataGeneralizations, and dynamicProperties. The Occurrence section includes occurrenceID, catalogNumber, recordNumber, recordedBy, individualCount, organismQuantity, organismQuantityType, sex, lifeStage, reproductiveCondition, behavior, establishmentMeans, occurrenceStatus, preparations, disposition, associatedMedia, associatedReferences, associatedSequences, associatedTaxa, otherCatalogNumbers, and occurrenceRemarks. The Event section includes eventID, parentEventID, fieldNumber, eventDate, eventTime, startDayOfYear, endDayOfYear, year, month, day, verbatimEventDate, habitat, samplingProtocol, sampleSizeValue, sampleSizeUnit, samplingEffort, fieldNotes, and eventRemarks. The Location section includes locationID, higherGeographyID, higherGeography, continent, waterBody, islandGroup, island, country, countryCode, stateProvince, county, municipality, locality, verbatimLocality, minimumElevationInMeters, maximumElevationInMeters, and verbatimElevation. A URL <https://dwc.tdwg.org/terms/> is displayed in an orange box at the top right of the main content area.



# DATA PUBLICATION WORKFLOW

06

Data Discovery

## Access by Users to the Global Infrastructure

### DATA VISUALIZATION

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

- access to all records
- no registration required
- raw and interpreted versions of records
- filters/facets for indexed fields, including quality issue tags
- logs for publishers (downloads, citations)

### DATA DOWNLOAD

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

- csv or DwC-A format
- login required
- asynchronous download
- no limit in volume
- DOI / recommended citation
- standard license

### DATA WEBSERVICE API

[api.gbif.org/v1/](http://api.gbif.org/v1/)

- registry, species, occurrences, maps, literature
- no registration required, except for download
- third-party tools for R or Python
- derived datasets downloads DOI

### OTHER DATA ACCESS

- **National Portals** with additional services (e.g. spatial analysis)
- snapshots available at Microsoft AI for Earth Program and AWS (of CC0 and CC-BY records only), for big data query support

GBIF users agree with Data User Agreement

# DATA PUBLICATION: NATIONAL PORTAL

06

## Data Discovery

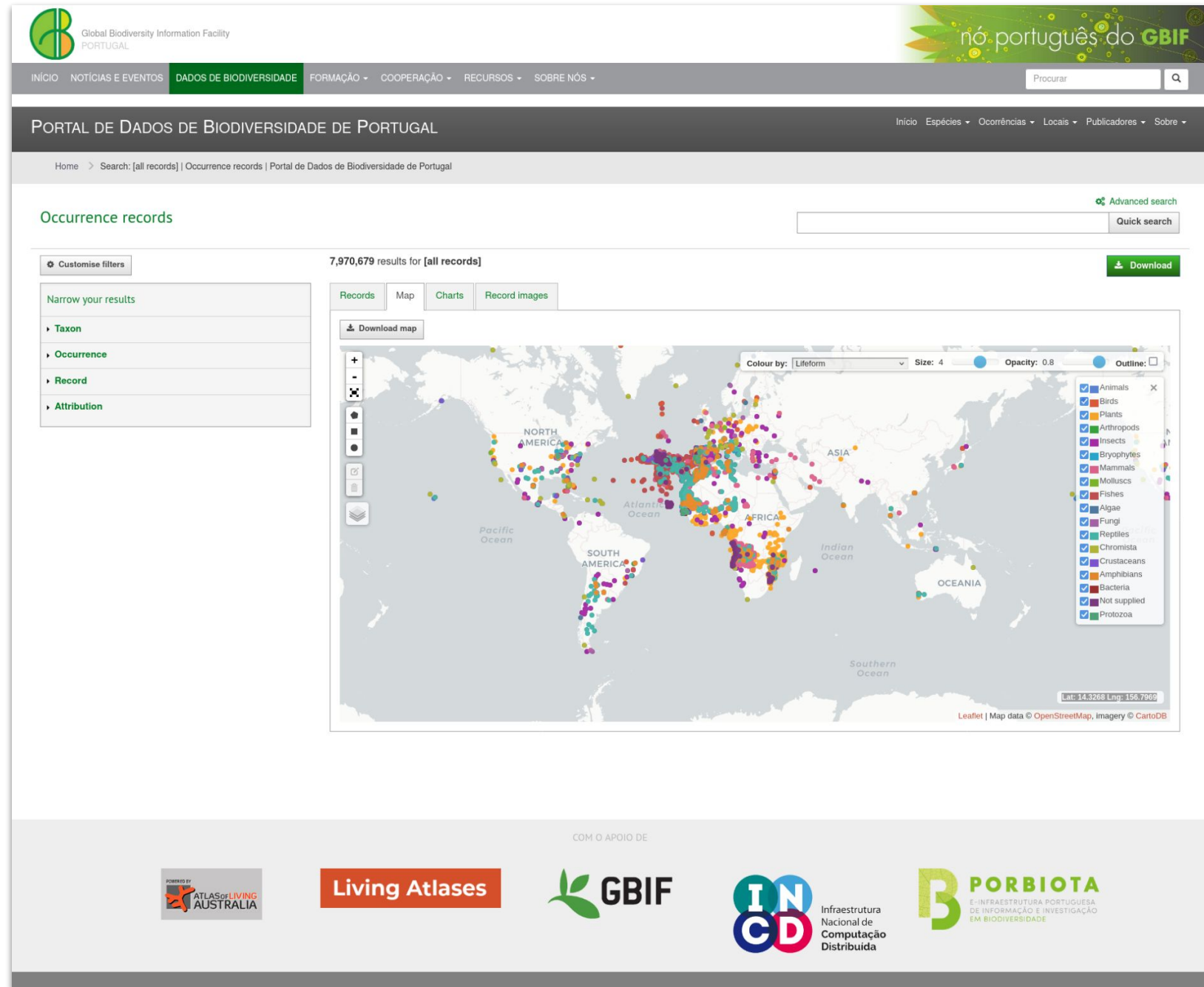
<https://dados.gbif.pt>

Result of a community effort by

- Living Atlases
- Atlas of Living Australia
- GBIF

Infrastructure support by

- PORBIOTA - Portuguese e-Infrastructure for Information and Research on Biodiversity
- INCĐ - National Distributed Computing Infrastructure





# CITATION GUIDELINES

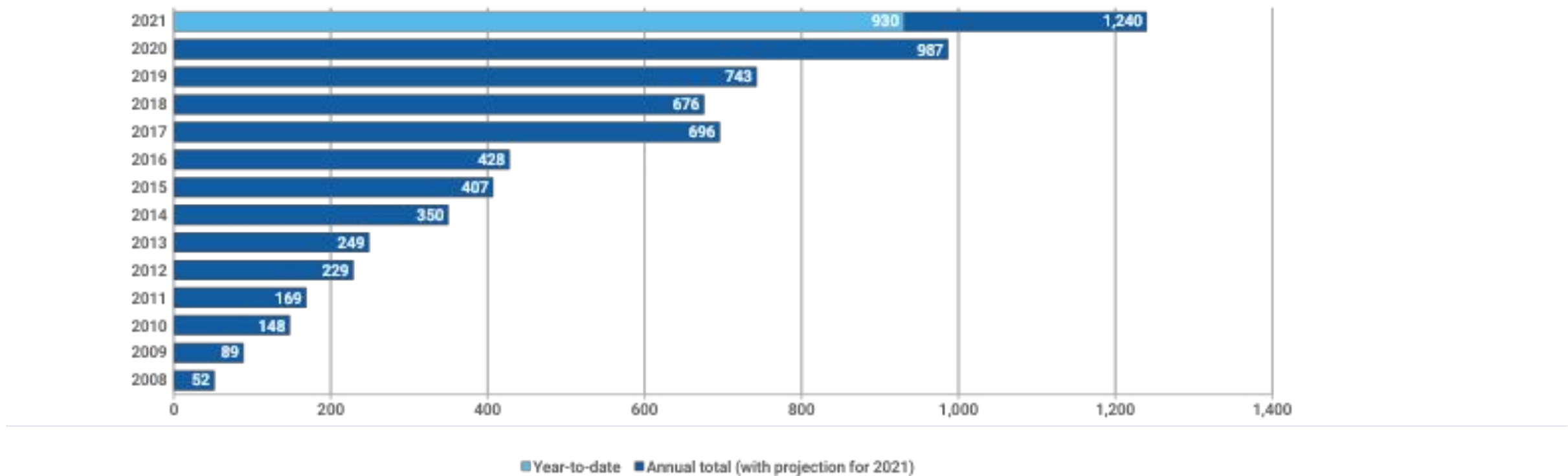
GBIF facilitates good citation practices by users, ensuring original publishers are credited

<b>Dataset DOI</b>	Data from single dataset	Single recommended citation which includes DOI given to dataset at the time of publication
<b>Download DOI</b>	Data from many datasets	Single recommended citation which includes DOI given to download. Attribution to original dataset by internal link
<b>Derived datasets DOI</b>	Access not via downloads Other cases	Users provide information to build a derived dataset DOI which reproduces the data used

Other services associated with citations:

- Citation tracking
- Report to publishers at dataset and publisher level
- Download dataset archiving – enable reproducibility

# PEER-REVIEWED PUBLICATIONS USING GBIF-MEDIATED DATA 30 Sept 2021





# COMPLIANCE OF GBIF WITH THE FAIR PRINCIPLES

GBIF compliance with FAIR principles		
F	Findable	<ul style="list-style-type: none"><li>Datasets are described with EML standard metadata</li><li>Each dataset receives a DOI and each record a unique ID</li><li>Each dataset version receives a version number</li></ul>
A	Accessible	<ul style="list-style-type: none"><li>Data access free to everyone on GBIF.org through a machine readable license</li><li>No restrictions on downloads of data (content or volume)</li><li>Access possible to through a documented webservice API (REST + JSON)</li><li>Open access licensed data available in third party cloud services</li></ul>
I	Interoperable	<ul style="list-style-type: none"><li>Data publishers use standards for data and metadata</li><li>Data available for download in Darwin Core standard format</li><li>Standard vocabularies used as defined in the Darwin Core standard</li></ul>
R	Reusable	<ul style="list-style-type: none"><li>Data available through one of the licenses: CC0, CC-BY or CC-BY-NC</li><li>No embargo after publication of data by institutions</li><li>Each download includes provenance and a recommended citation with DOI</li><li>Record level quality issues flags help to determine fitness for use</li></ul>

FAIR Data Maturity Model:  
specification and guidelines

DOI: 10.15497/RDA0050



FAIR Assessment in project



Summary Report	
Indicators for Findable	Compliant
Indicators for Accessible	Compliant
Indicators for Interoperable	Not Compliant
● Indicators Satisfied	9
● Indicators NOT Satisfied	3
● % Progress	75.00%
Indicators for Reusable	Compliant

# Thank you!

Rui Figueira, [ruifigueira@isa.ulisboa.pt](mailto:ruifigueira@isa.ulisboa.pt),  
GBIF Portugal, Instituto Superior de Agronomia

This presentation is based on:  
GBIF Secretariat (2021) GBIF overview slides, October 2021. Copenhagen: GBIF Secretariat. Available online at:  
<https://www.gbif.org/document/81771>.



PORTUGUESE NODE OF GBIF  
INSTITUTO SUPERIOR DE AGRONOMIA  
[node@gbif.pt](mailto:node@gbif.pt)



This presentation is available under license  
<https://creativecommons.org/licenses/by/4.0/>