



Promote biodiversity data publishing and usage: the role of data papers

Rui Figueira
Nó Português do GBIF
rui.figueira@iict.pt



Apoio

FCT Fundação para a Ciência e a Tecnologia
MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA



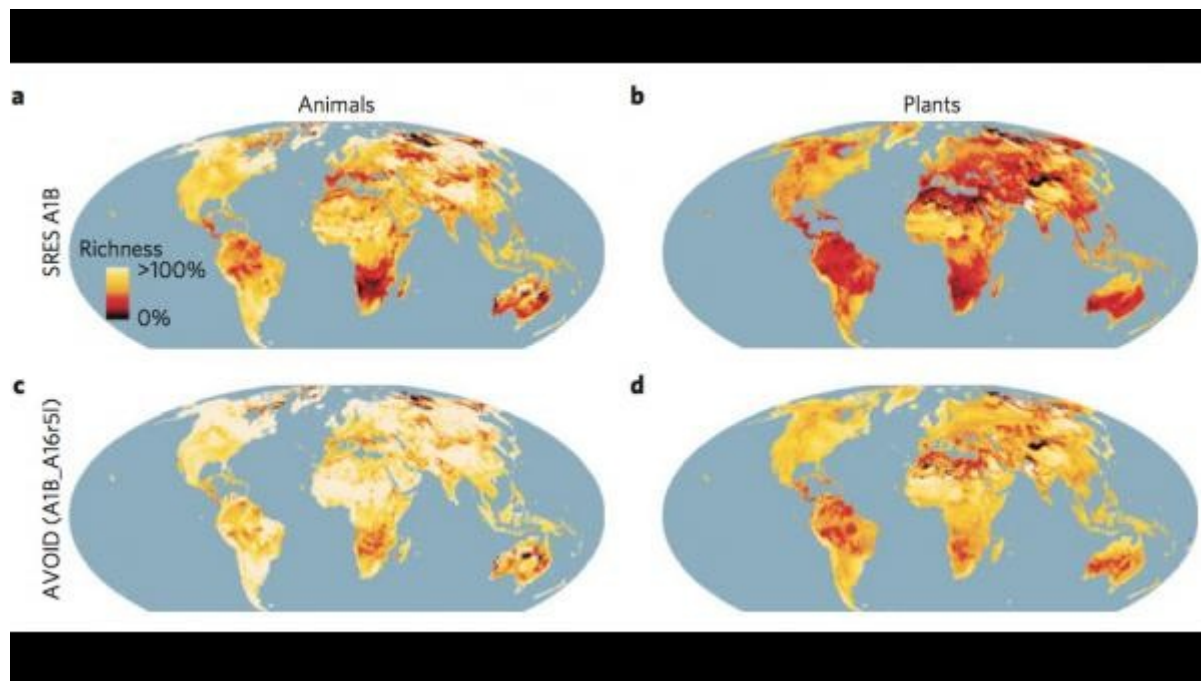
Outline

- Introduction
- Why publish data? - main motivations
- Data papers – a vehicle for data publication?
- How to publish biodiversity data? – options, tools, procedures





GBIF enables global study of climate impact on species



50 000 species
170 million records

More than half of the plants and over a third of animal species could lose more than half of their climatic range by 2080.



Relevance of shared data for global initiatives



Aichi Target 19 - By 2020, knowledge, the science base and technologies relating to biodiversity, its values, functioning, status and trends, and the consequences of its loss, are improved, widely shared and transferred, and applied.

The mobilization and access to primary biodiversity data, using common standards and protocols is recognised as a key action as recommended by GBO4.



The role of GBIF

The Global Biodiversity Information Facility - GBIF is an **intergovernmental organisation** funded in 2001 to facilitate sharing and free and open access to biodiversity data

The mission is to ensure that anyone, anywhere will have access to data about any kind of life on Earth, shared through the Internet.

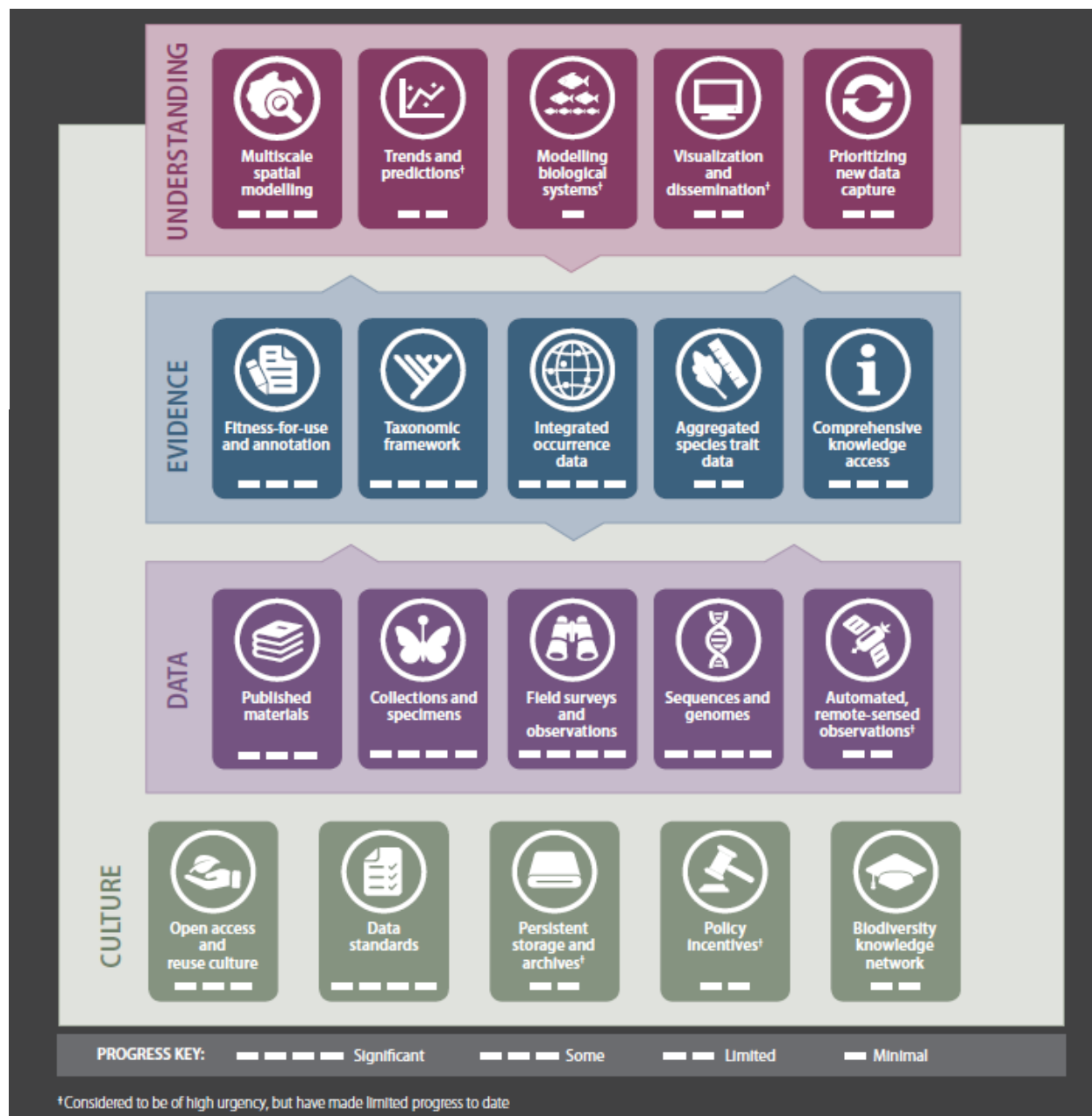
GBIF is, *de facto*, the global facility for the mobilisation and access to biodiversity data



Global Biodiversity
Informatics Outlook

www.biodiversityinformatics.org

The GBIO framework





- Introduction
- Why publish data? - main motivations
- Data papers – a vehicle for data publication?
- How to publish biodiversity data? – options, tools, procedures



Ensure preservation and access to data

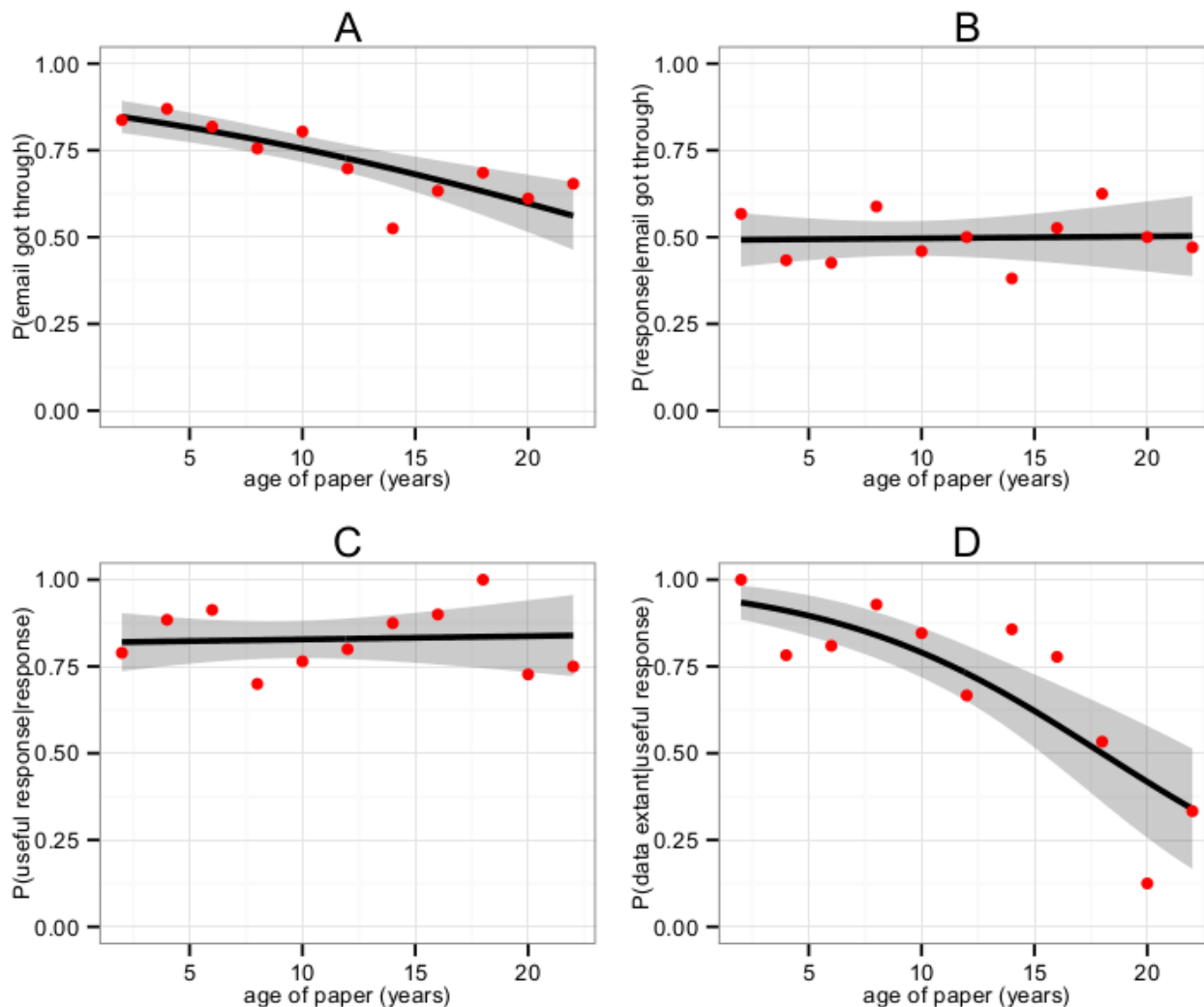
Current Biology

Volume 24, Issue 1, 6 January 2014, Pages 94–97

Report

The Availability of Research Data Declines Rapidly with Article Age

Timothy H. Vines^{1,2,*}, Arianne Y.K. Albert³, Rose L. Andrew¹, Florence Débarre^{1,4}, Dan G. Bock¹, Michelle T. Franklin^{1,5}, Kimberly J. Gilbert¹, Jean-Sébastien Moore^{1,6}, Sébastien Renaut¹, Diana J. Rennison¹



The availability of data
decreases rapidly with
time

Vines, Timothy H et al.. 2014. "The Availability of Research Data Declines Rapidly with Article Age." Current Biology 24 (1): 94–97. doi:10.1016/j.cub.2013.11.014.



Enforcing open access

21.7.2012

EN

Official Journal of the European Union

L 194/39

RECOMMENDATIONS

COMMISSION RECOMMENDATION

of 17 July 2012

on access to and preservation of scientific information

(2012/417/EU)

THE EUROPEAN COMMISSION,

scientific publishing and the preservation of research results, examining relevant organisational, legal, technical and financial issues.

FCT Fundação para a Ciência e a Tecnologia

MINISTÉRIO DA EDUCAÇÃO E CIÊNCIA

Política sobre Acesso Aberto a Publicações Científicas resultantes de Projetos de I&D Financiados pela FCT

(adotada em 5 de maio de 2014)

1. Introdução

A disponibilização aberta dos resultados da investigação realizada com recurso a financiamento público tem significativos benefícios sociais e económicos. Em linha com o que vem sendo crescentemente adotado por agências públicas de financiamento de I&D de outros países e com as recomendações da Comissão Europeia de 17 de Julho de 2012¹, a FCT – Fundação para a Ciência e a Tecnologia, IP – adota a presente política de obrigatoriedade de disponibilização em Acesso Aberto das publicações resultantes dos projetos de I&D que financia.



Increase impact

OPEN ACCESS Freely available online



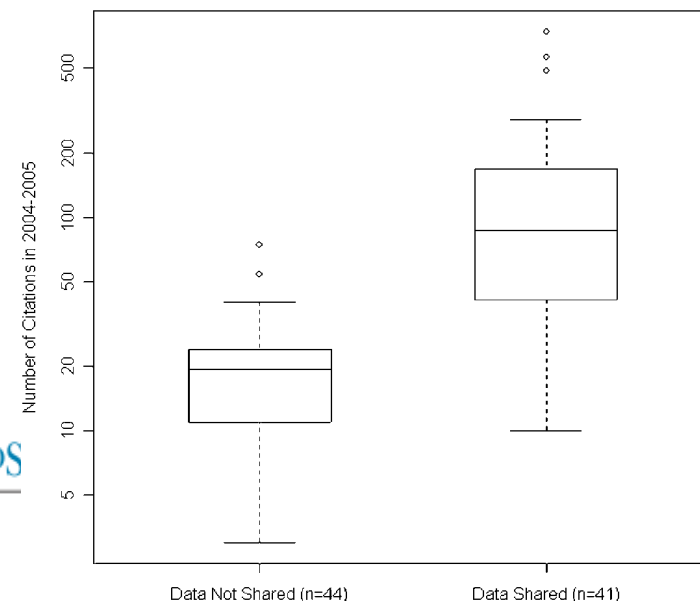
Sharing Detailed Research Data Is Associated with Increased Citation Rate

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

Background. Sharing research data provides benefit to the general scientific community, but the benefit is less obvious for the investigator who makes his or her data available. **Principal Findings.** We examined the citation history of 85 cancer microarray clinical trial publications with respect to the availability of their data. The 48% of trials with publicly available microarray data received 85% of the aggregate citations. Publicly available data was significantly ($p = 0.006$) associated with a 69% increase in citations, independently of journal impact factor, date of publication, and author country of origin using linear regression. **Significance.** This correlation between publicly available data and increased literature impact may further motivate investigators to share their detailed research data.

Citation: Piwowar 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





Increase visibility and discoverability of the collection or database

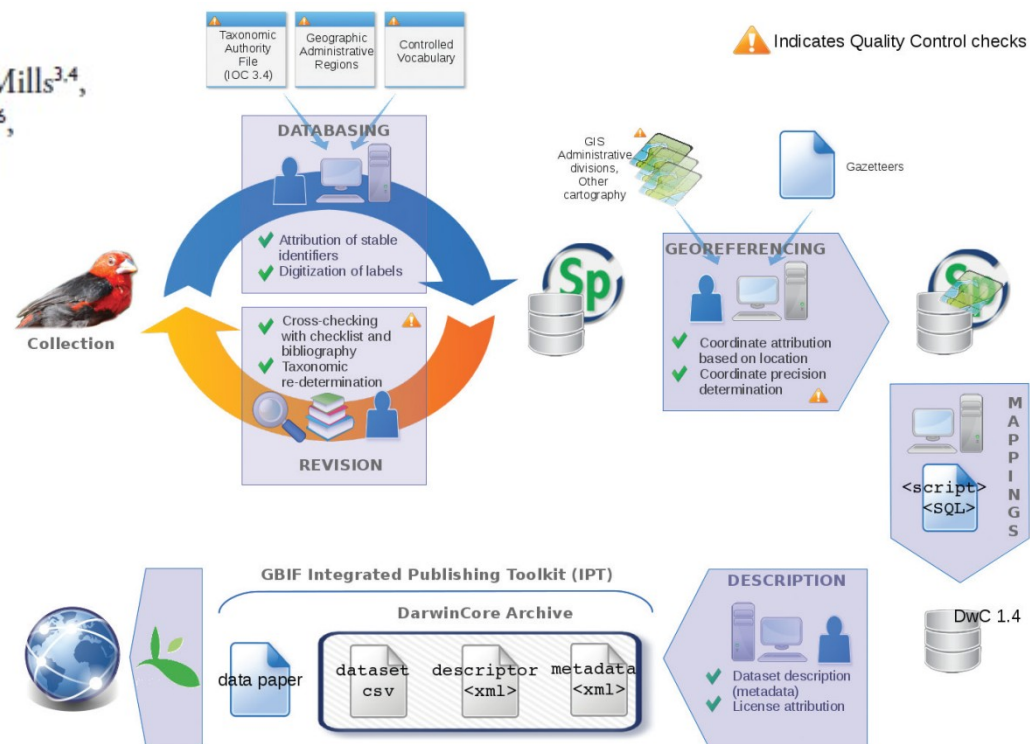
ZooKeys 387: 89–99 (2014)
doi: 10.3897/zookeys.387.6412
www.zookeys.org

DATA PAPER



The collection and database of Birds of Angola hosted at IICT (Instituto de Investigação Científica Tropical), Lisboa, Portugal

Miguel Monteiro^{1,2}, Luís Reino², Pedro Beja², Michael Stuart Lyne Mills^{3,4},
Cristiane Bastos-Silveira^{5,6}, Manuela Ramos¹, Diana Rodrigues⁶,
Isabel Queirós Neves^{5,6}, Susana Consciência¹, Rui Figueira^{1,2}

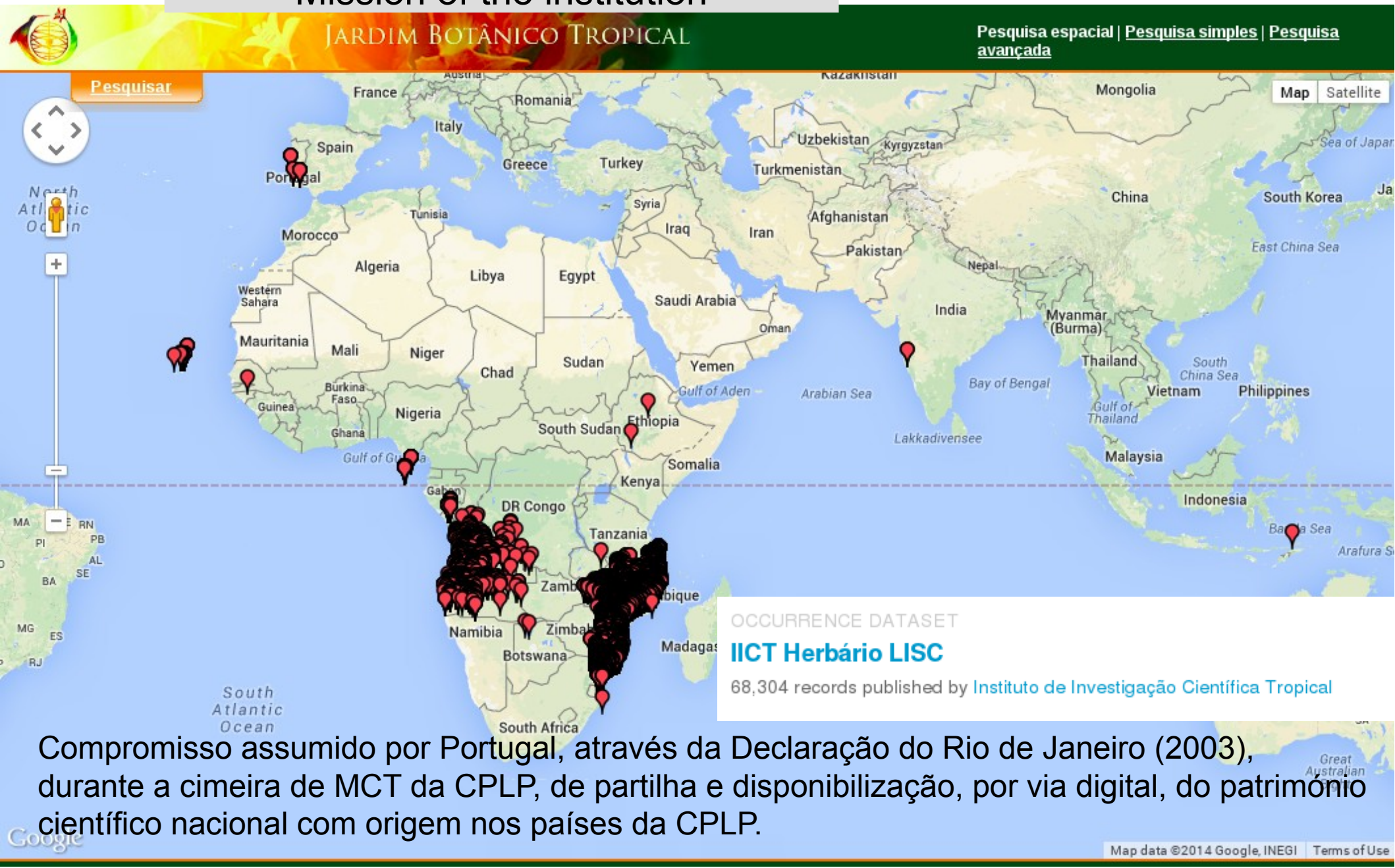




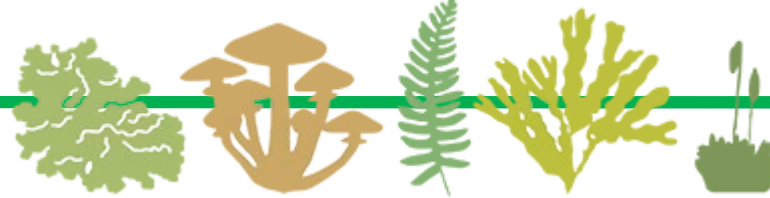
Mission of the institution

JARDIM BOTÂNICO TROPICAL

Pesquisa espacial | [Pesquisa simples](#) | [Pesquisa avançada](#)



Compromisso assumido por Portugal, através da Declaração do Rio de Janeiro (2003), durante a cimeira de MCT da CPLP, de partilha e disponibilização, por via digital, do património científico nacional com origem nos países da CPLP.



Main motivations for data sharing

- Increase scientific impact
- Increase visibility at the global level
- Facilitate collections' revision by specialists
- Ensure data preservation
- Comply with open data policies enforced by funding agencies
- Maximize recognition on data use
- Enable new research



Research Policy 43 (2014) 1621–1633



Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol



Open access to data: An ideal professed but not practised



Patrick Andreoli-Versbach^{a,b,*}, Frank Mueller-Langer^{a,c}

^a Max Planck Institute for Innovation and Competition

^b University of Munich, Department of Economics

^c International Max Planck Research School for Competition and Innovation

Andreoli-Versbach, Patrick, and Frank Mueller-Langer. "Open Access to Data: An Ideal Professed but Not Practised." Research Policy 43, no. 9 (November 2014): 1621–33. doi:10.1016/j.respol.2014.04.008.



Factor with negative effect

- Perception of higher competition disadvantage
- Researchers with lower rates of citation are less willing to share data
- The volume of data matters: larger datasets require more management efforts for data and metadata preparation
- Personal attitudes towards open science



- Introduction
- Why publish data? - main motivations
- Data papers – a vehicle for data publication?
- How to publish biodiversity data? – options, tools, procedures



First things first!

Any scientific paper should be also a data paper!

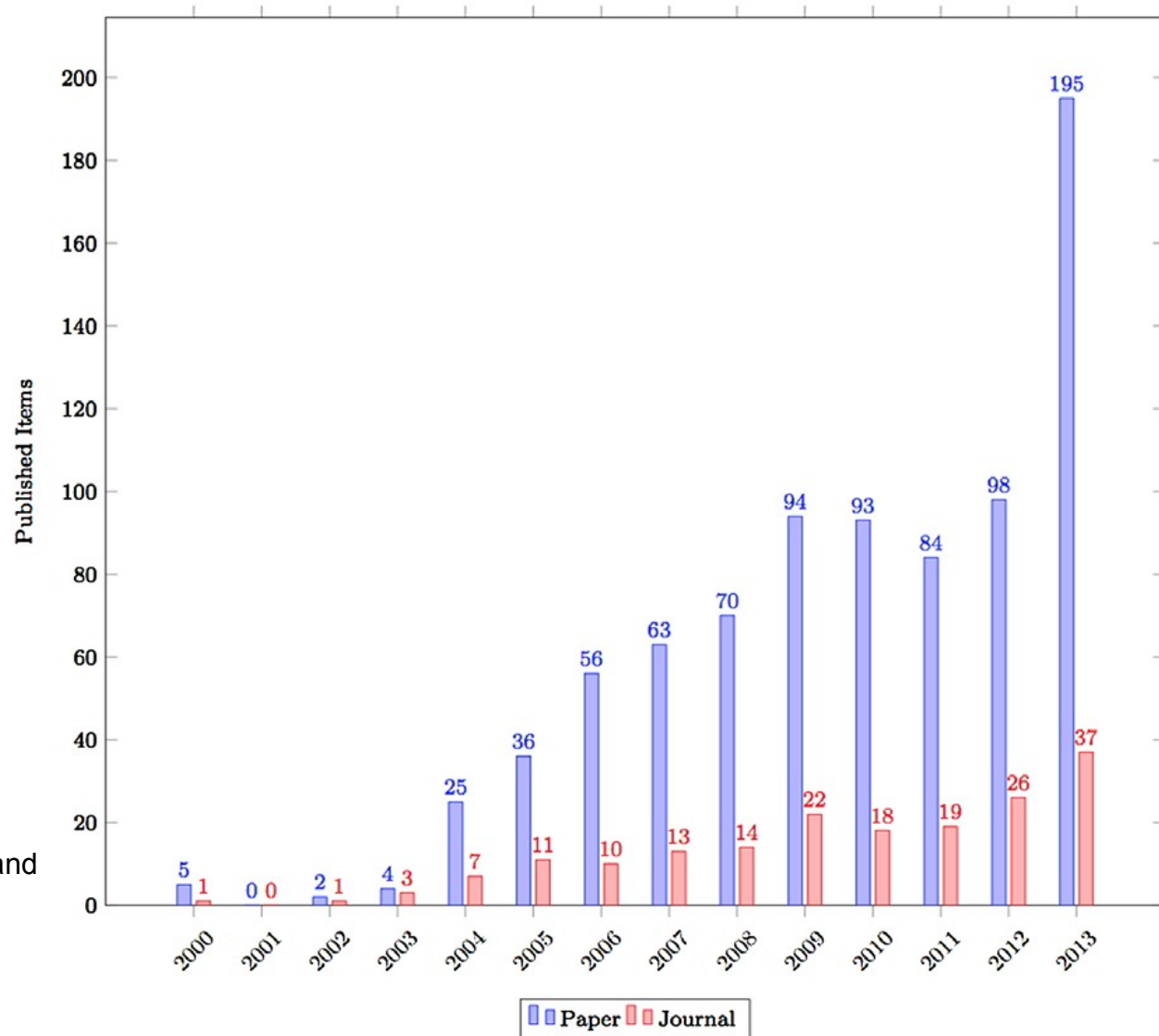
Meaning, it should be accompanied by the publication
of the raw data.



“Attraction” for data papers

- A recent review shows that exists more than 100 data journals
- And the number of data papers is quickly growing

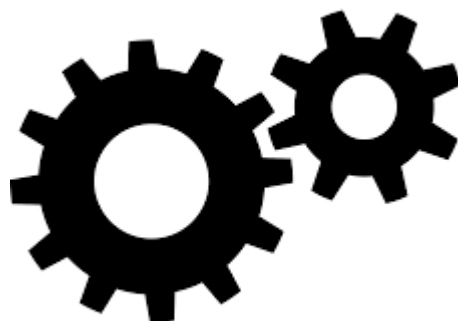
Candela, Leonardo, Donatella Castelli, Paolo Manghi, and Alice Tani. “Data Journals: A Survey.” Journal of the Association for Information Science and Technology, February 1, 2015, n/a – n/a. doi:10.1002/asi.23358.





Main benefits of publishing a data paper:

1. Increase of discoverability and accessibility to data
2. Promotes full documentation of data by metadata making easier to verify the fitness-for-use of data
3. Reviewing process may increase data quality
4. Facilitate proper attribution and citation
5. Promotes reuse





SCIENTIFIC DATA

[Home](#) | [About](#) | [For Authors](#) | [For Referees](#) | [Advisory and Editorial Board](#) | [Open Access](#) | [FAQ](#)

[Sign up for Scientific Data e-alert](#) [Facebook](#) [Twitter](#)

Submit to *Scientific Data* in three simple steps

DESCRIBE

Write a detailed description of your dataset. We have templates to help you and a detailed guide to authors.

DEPOSIT

See our list of recommended repositories. We will help you find the right place for your data.

SUBMIT

Submit online and get the credit you deserve for your data!

Get credit where credit's due and share your data

Scientific Data
is accepting
submissions

[Submit manuscript](#)



Scientific Data Updates

[Scientific Data conferences & events calendar](#)
April 3, 2014

[Endorsing the Joint Declaration of Data Citation Principles](#)
March 24, 2014

[Scientific Data's first publications](#)
March 24, 2014

[Scientific Data's metadata specification](#)
January 8, 2014

[New Advisory and Editorial Board names](#)

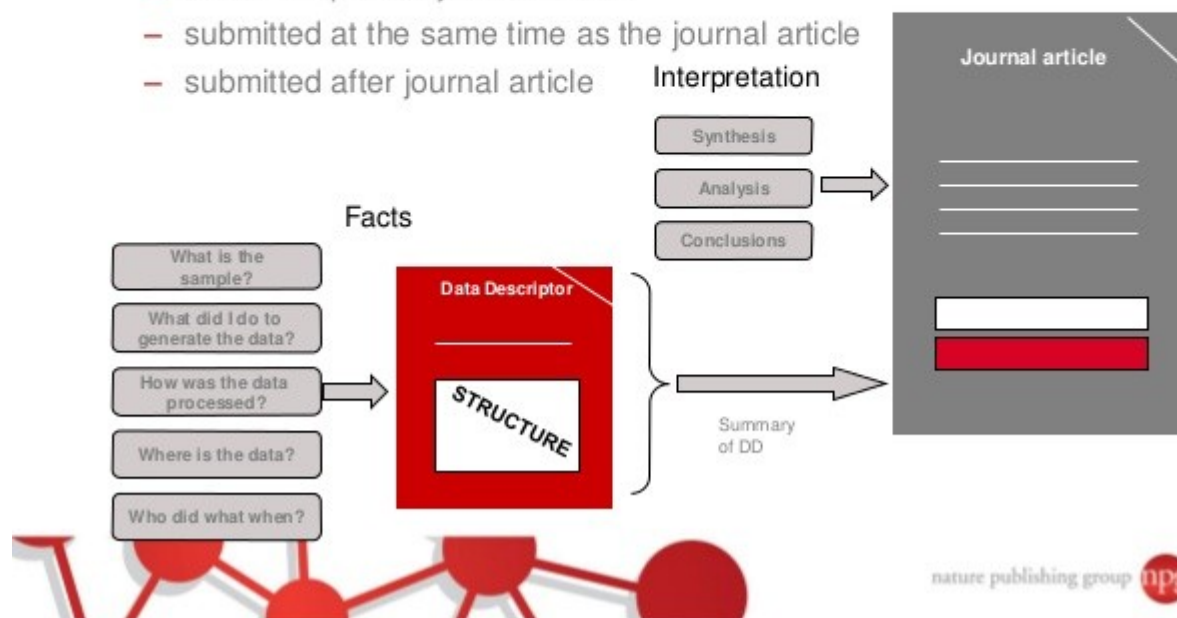
<http://www.nature.com/scientificdata/>



SCIENTIFIC DATA

Data Descriptor vs. Traditional Article

- The data descriptor is only concerned with the facts behind the methodology of data generation/collection and processing
- A data descriptor can be:
 - submitted prior to journal article
 - submitted at the same time as the journal article
 - submitted after journal article





- As data papers (in data journals or regular journals that accepts data papers

- Geoscience Data Journal
- Earth System Science Data
- Dataset Papers (Hindawi publishing)
- Journal of Chemical and Engineering Data
- GigaScience
- Journal of Physical and Chemical Research Data
- F1000 Research
- CODATA's Data Science Journal
- Journal of Open Archaeology Data
- Journal of Open Public Health Data
- Journal of Open Psychology Data
- Journal of Open Research Software
- Internet Archaeology



- As data papers (in data journals or regular journals that accepts data papers)

Biodiversity Data Journal

International Journal of Myriapodology

Journal of Hymenoptera Research

MycoKeys

Nature Conservation

NeoBiota

Nota Lepidopterologica

PhytoKeys

Subterranean Biology

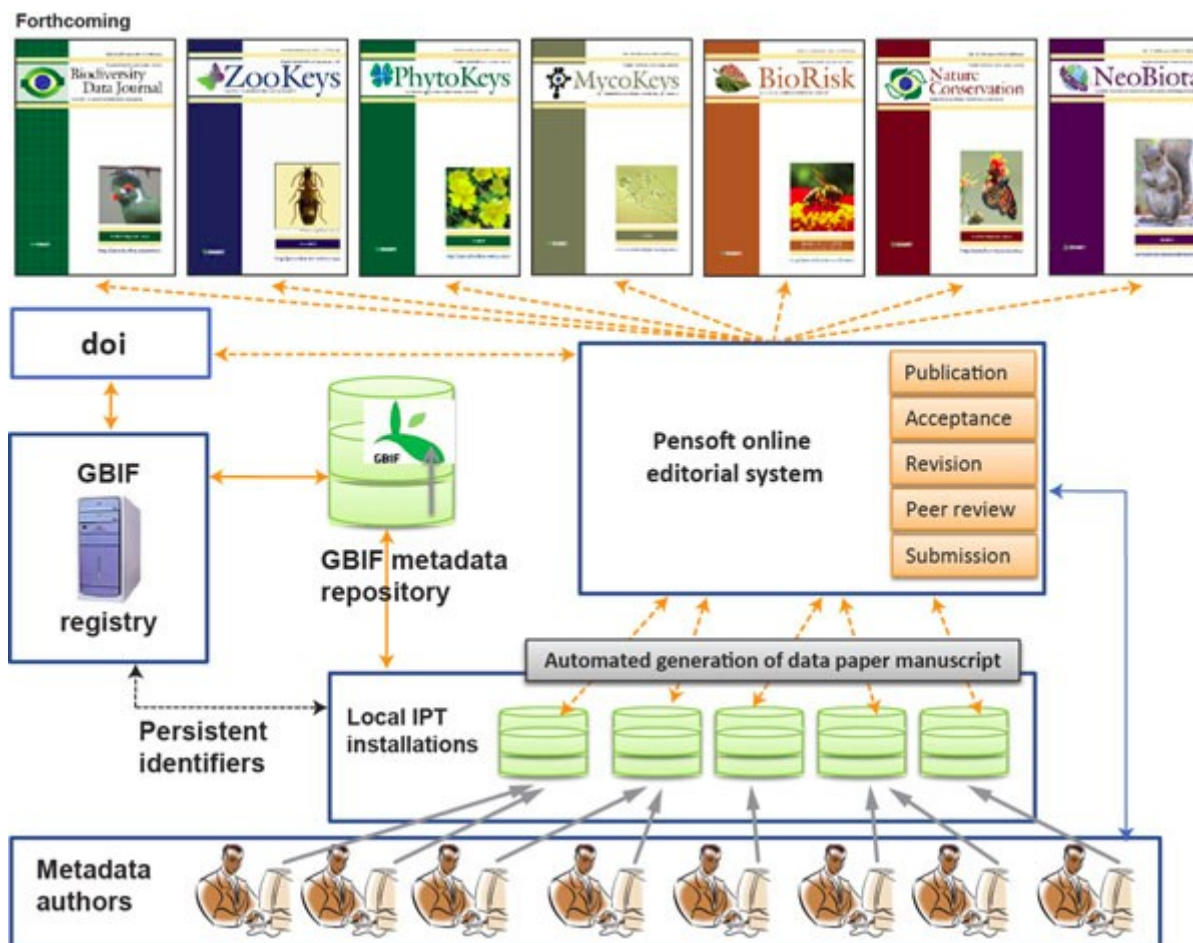
ZooKeys

Zoosystematics and Evolution



A peer-reviewed open-access journal
**Biodiversity
Data Journal**
Making your data count! ISSN 1314-2828 (online)

Pensoft accepts IPT as a platform for the preparation of the manuscript
It as an agreement with GBIF



Pensoft Data Publishing Policies and Guidelines for
Biodiversity Data

<http://www.pensoft.net/page.php?P=23>



Option for publishing scientific data

Requirements of data repositories

- Accredited, internationally or institutionally
- guarantee long-term availability
- with unique object identifier (e.g. DOI)
- data sets available free of charge



GenBank

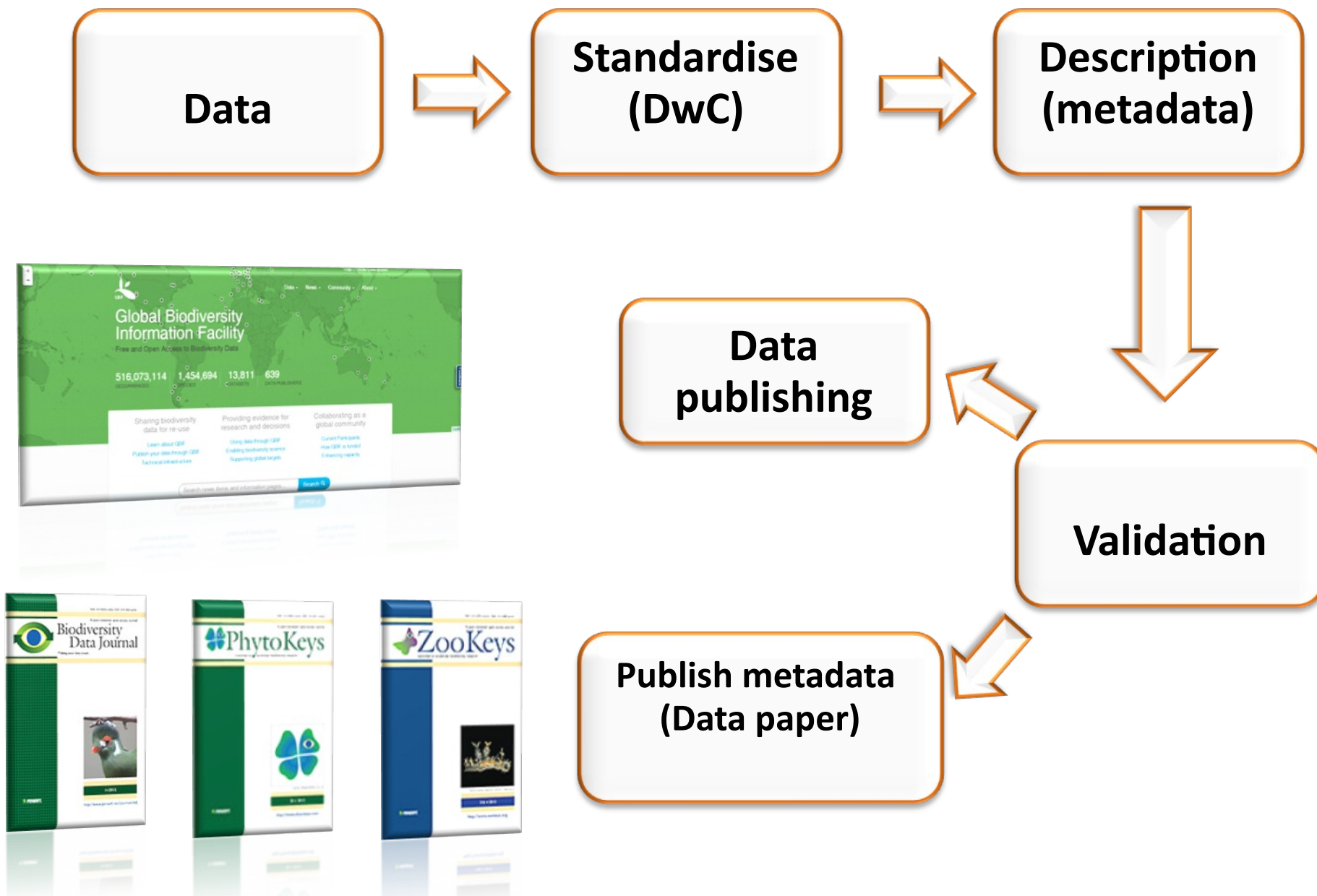




- Introduction
- Why publish data? - main motivations
- Data papers – a vehicle for data publication?
- How to publish biodiversity data? – options, tools, procedures



How to publish data through GBIF





Integrated Publishing Toolkit



GBIF PORTUGAL - INTEGRATED PUBLISHING TOOLKIT
acesso aberto e gratuito a dados de biodiversidade (IPT)



ENGLISH

Home

About

Hosted resources available through this IPT

Filter:

Logo	Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication
--	Bryophyte collection of Porto Herbarium (PO)	Museu de História Natural da Universidade do Porto	Occurrence	--	7,621	2014-07-25	2014-07-25	--
--	Checklist da Flora de Portugal (Continental, Açores e Madeira)	GBIF Portugal	Checklist	--	3,994	2014-11-18	2014-11-18	--
--	Moluscos Marinhos de Augusto Nobre	Museu de História Natural da Universidade do Porto	Occurrence	Specimen	881	2013-12-17	2013-12-17	--

Showing 1 to 3 of 3 resources

◀ previous next ▶

The most recently updated resources are also available as an [RSS feed](#).



Integrated Publishing Toolkit

Web application, open source software developed by GBIF

Allows registration of a dataset, for record and metadata indexing

One installation can host several datasets, managed by different curators, through the browser

Infrastructure - requirements

- Stable URL estável
- Java (and tomcat 7)
- Memory: 256 MB
- Storage: 20 MB + space for datasets



Data and metadata standards

Darwin Core

Introduction

References

Quick Reference Guide

Term Index

Record-level Terms

Occurrence

Event

Location

GeologicalContext

Identification

Taxon

ResourceRelationship

MeasurementOrFact

Term Definitions

Simple Darwin Core

Type Vocabulary

Namespace Policy

XML Guide

Text Guide

Complete History

Decision History

Mapping to ABCD

Mapping to Old Versions

Translations

Darwin Core Terms: A quick reference guide

Title: Darwin Core Terms: A quick reference guide

Date Issued: 2009-02-12

Date 2011-10-26

Modified:

Abstract: This document is a quick reference for all recommended Darwin Core terms. For complete historical term information, including version changes and pre-standard terms, see [\[HISTORY\]](#). For a comparative table of elements from pre-standard versions of Darwin Core to the current terms in standard, see [\[VERSIONS\]](#).

Contributors: John Wieczorek (MVZ), Markus Döring (GBIF), Renzo De Giovanni (CRIA), Tim Robertson (GBIF), Dave Viegas (KUNHM)

Legal: This document is governed by the standard legal copyright, licensing provisions and disclaimers is by the Taxonomic Databases Working Group.

Part of TDWG Standard: <http://www.tdwg.org/standards/450/>

Creator: Darwin Core Task Group

Identifier: <http://rs.tdwg.org/dwc/2011-10-26/terms/>

Latest Version: <http://rs.tdwg.org/dwc/terms/>

Replaces: <http://rs.tdwg.org/dwc/2009-12-07/terms/>

Document Status: Current Standard

locationID	higherGeographyID	higherGeography	continent	waterBody	islandGroup	island	country	countryCode	stateProvince	county	municipality	locality	verbatimLocality	minimumElevationInMeters	maximumElevationInMeters	verbatimElevation	minimumDepthInMeters	maximumDepthInMeters	verbatimDepth	minimumDistanceAboveSurfaceInMeters	maximumDistanceAboveSurfaceInMeters	locationAccordingTo	locationRemarks	decimalLatitude	decimalLongitude	geodeticDatum	coordinateUncertaintyInMeters	coordinatePrecision	pointRadiusSpatialFit
----------------------------	-----------------------------------	---------------------------------	---------------------------	---------------------------	-----------------------------	------------------------	-------------------------	-----------------------------	-------------------------------	------------------------	------------------------------	--------------------------	----------------------------------	--	--	-----------------------------------	--------------------------------------	--------------------------------------	-------------------------------	---	---	-------------------------------------	---------------------------------	---------------------------------	----------------------------------	-------------------------------	---	-------------------------------------	---------------------------------------



Darwin Core - <http://rs.tdwg.org/dwc/terms/>

Term Name: occurrenceID

Identifier: <http://rs.tdwg.org/dwc/terms/occurrenceID>

Class: <http://rs.tdwg.org/dwc/terms/Occurrence>

Definition: An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, a combination of identifiers in the record that will most closely make the occurrenceID globally unique.

Comment: For a specimen in the absence of a bona fide global unique identifier, for example, use the form: "urn:catalog:[institutionCode]:[collectionCode]:[specimenNumber]". For discussion see <http://terms.tdwg.org/wiki/dwc:occurrenceID>

Details: [occurrenceID](#)

Term Name: catalogNumber

Identifier: <http://rs.tdwg.org/dwc/terms/catalogNumber>

Class: <http://rs.tdwg.org/dwc/terms/Occurrence>

Definition: An identifier (preferably unique) for the record within the data set or collection.

Comment: Examples: "2008.1334", "145732a", "145732". For discussion see <http://terms.tdwg.org/wiki/dwc:catalogNumber>

Details: [catalogNumber](#)

Term Name: recordNumber

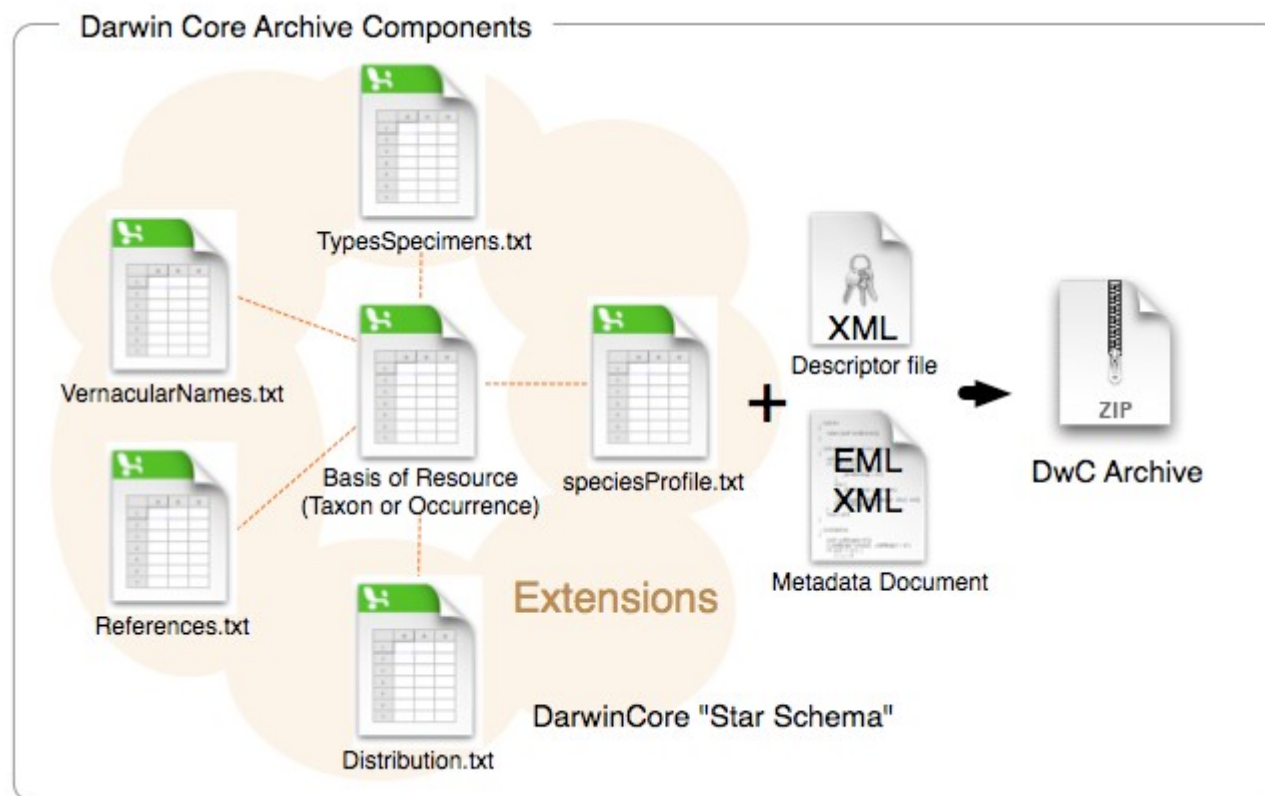
Identifier: <http://rs.tdwg.org/dwc/terms/recordNumber>

Class: <http://rs.tdwg.org/dwc/terms/Occurrence>

Definition: An identifier given to the Occurrence at the time it was recorded. Often serves as a link between field notes and an Occurrence record, such as a specimen number.

Darwin Core – Star schema with support for several information types:

Sample based data, germoplasm, multiple determinations, types and specimens, vernacular names, alternative identifiers, species profile, bibliographic references, taxon description, traits, multimedia, etc

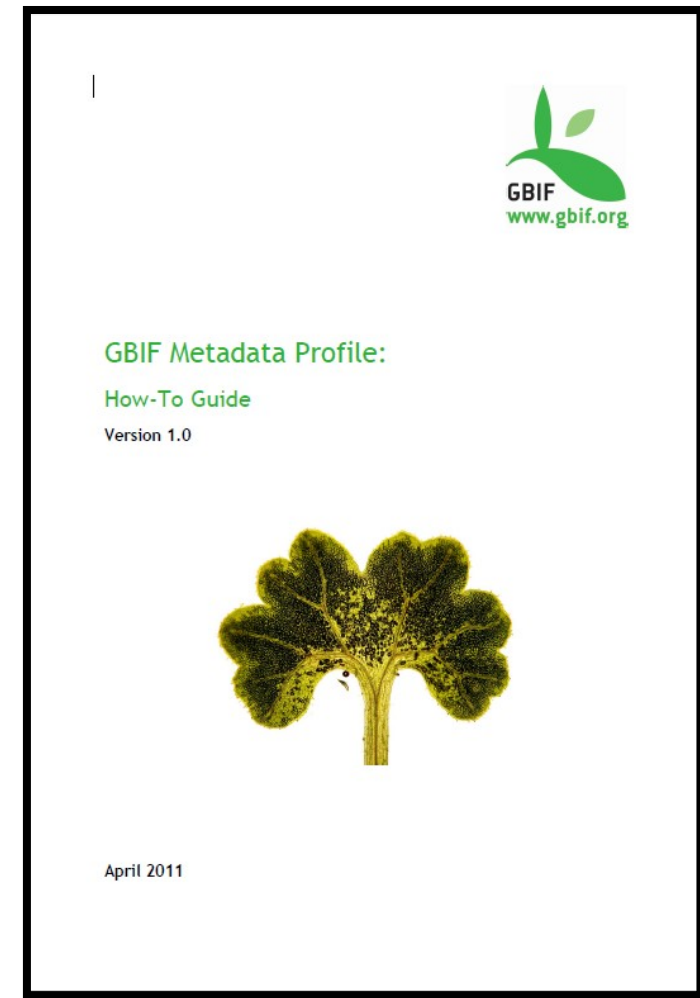




GMP (GBIF Metadata profile):

EML (Ecological Metadata Language)

- Dataset (Resource)
 - Project
 - People and Organisations
 - Keyword Set (General Keywords)
 - Coverage
 - o Taxonomic Coverage
 - o Geographic Coverage
 - o Temporal Coverage
 - Methods
 - Intellectual Property Rights
 - Additional Metadata + NCD (Natural Collections Descriptions Data)
- Related





Data paper

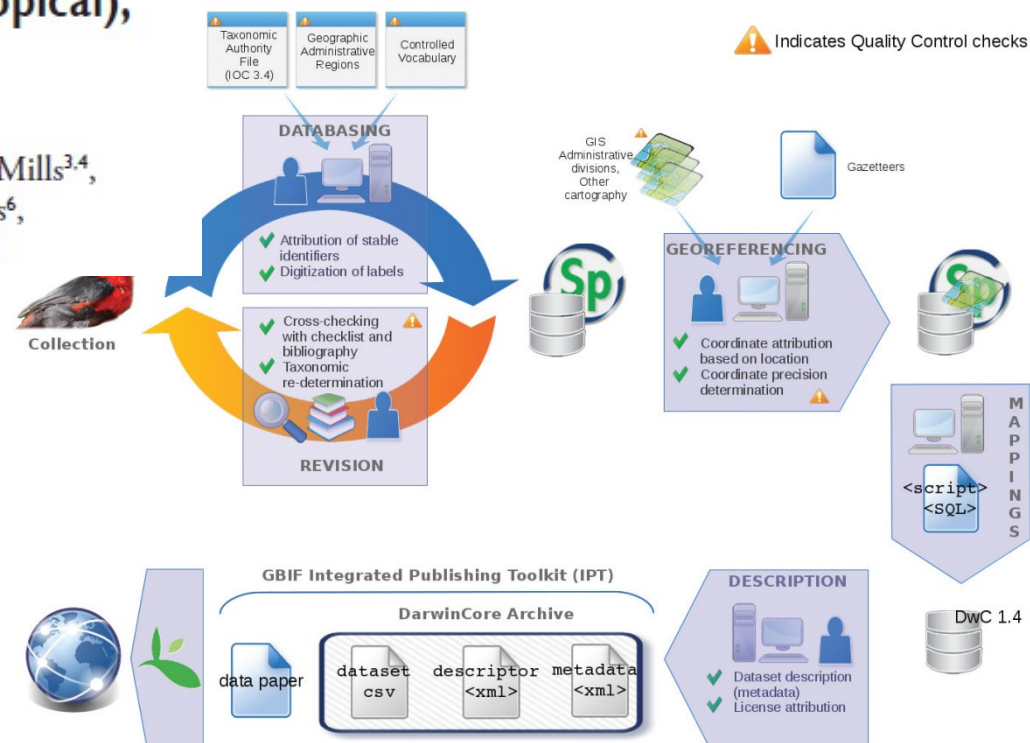
ZooKeys 387: 89–99 (2014)
doi: 10.3897/zookeys.387.6412
www.zookeys.org

DATA PAPER



The collection and database of Birds of Angola hosted at IICT (Instituto de Investigação Científica Tropical), Lisboa, Portugal

Miguel Monteiro^{1,2}, Luís Reino², Pedro Beja², Michael Stuart Lyne Mills^{3,4},
Cristiane Bastos-Silveira^{5,6}, Manuela Ramos¹, Diana Rodrigues⁶,
Isabel Queirós Neves^{5,6}, Susana Consciência¹, Rui Figueira^{1,2}





1st Data poster!

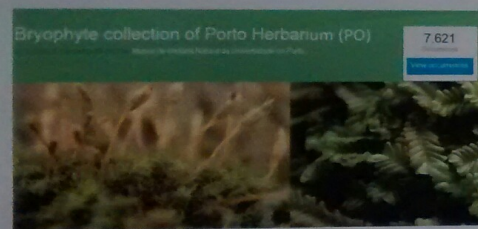
PUTTING THE BRYOPHYTES OUT THERE:

publishing the bryophyte collection data of Porto Herbarium (PO) with GBIF

Helena Hespanhol¹, Cristiana Vieira¹

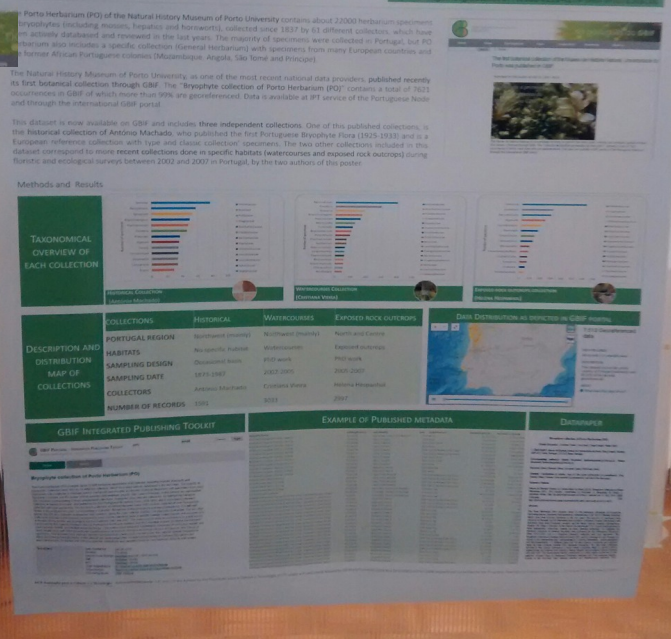
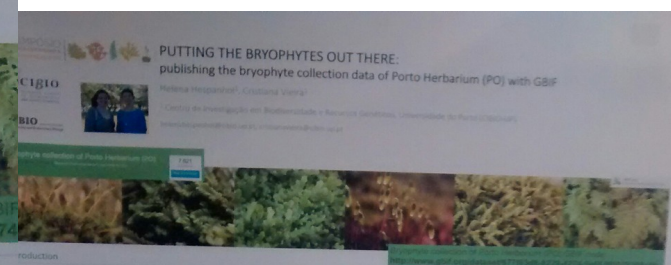
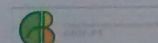
¹ Centro de Investigação em Biodiversidade e Recursos Genéticos, Universidade do Porto (CIBIO-UP).

helenahespanhol@cibio.up.pt, cristianavieira@cibio.up.pt



Introduction

The Porto Herbarium (PO) of the Natural History Museum of Porto University contains about 22000 herbarium specimens of bryophytes (including mosses, hepatics and hornworts), collected since 1837 by 61 different collectors, which have been published in the GBIF dataset.





Conclusions

- Publishing data is a good to attract attention to your work, while promoting reuse of data and new science
- Data papers are not to discuss science, only to describe data
- Data papers facilitates data indexing, discoverability, and reuse of data
- Data paper allows proper attribution and facilitates citation
- The process of creating a data paper can also help to improve the quality of your data, and of the description of the data



Obrigado!

Rui Figueira
Instituto de Investigação Científica Tropical
Nó Portugêses do GBIF
Rua da Junqueira, 86-1º
1300-344 Lisboa, Portugal
rui.figueira@iict.pt
www.gbif.pt