



# CIÊNCIA ABERTA E DADOS ABERTOS: do compartilhamento ao reuso de dados científicos

**Renata G. Curty**

Professora Adjunta  
Departamento de Ciência da Informação  
Universidade Estadual de Londrina

**I Seminário Gestão de Dados de Pesquisa**  
**Realização: UNIFESP/UFABC**

São Paulo, SP

27 de agosto de 2018

À medida que a ciência continua a evoluir, vários indicadores sugerem uma **mudança da ênfase e da principal unidade de comunicação científica** para uma abordagem mais ampla, **voltada para conjunto de dados.**



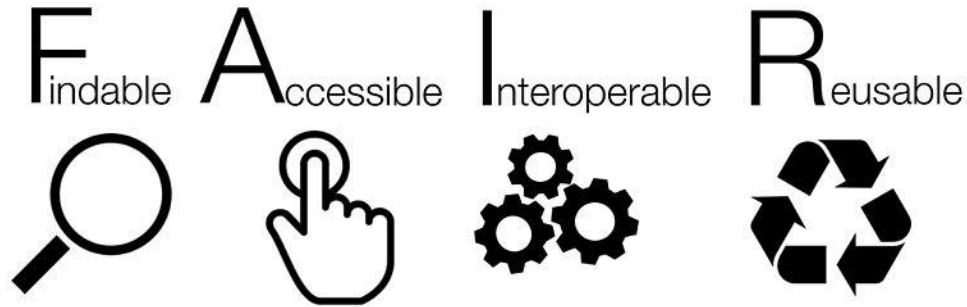


Dados científicos são **autônomos e de alto valor intrínseco**. Podem ser aplicados a diferentes contextos de modo a responder questões além das propostas e antecipadas pelos investigadores que os coletaram/produziram. Uma **riqueza de relações que podem ser extraídas de uma mesma coleção de dados**.



# Contextualizando

## Princípios do FAIR Force11 (2016)



**F1** Metadados são atribuídos a um identificador globalmente exclusivo e eternamente persistente.

**F2** Metadados ricos.

**F3** Metadados são registrados ou indexados em um recurso pesquisável.

**F4** Metadados especificam o identificador de dados.

**A1** Metadados recuperáveis por IDs com protocolo de comunicação padronizado.

**A1.1** O protocolo é aberto, gratuito e universal. **A1.2** Permite autenticação e autorização.

**A2** Metadados estão acessíveis, mesmo quando os dados não estão mais disponíveis.

**I1** Metadados usam uma linguagem formal, acessível, compartilhada e amplamente aplicável para a representação do conhecimento.

**I2** Metadados usam vocabulários que seguem os princípios do FAIR.

**I3** Os metadados incluem referências qualificadas a outros metadados.

**R1** Metadados têm uma pluralidade de atributos precisos e relevantes.

**R1.1** São liberados com licença de uso de dados clara e acessível. **R1.2** Estão associados à sua proveniência.

**R1.3** Atendem aos padrões da comunidade/domínio.

# Contextualizando

- Busca por maior **transparência**, **otimização de recursos injetados em C&T**, e **reprodutibilidade** em ciência.
- Crescimento vertiginoso da **disponibilidade de dados científicos** seja por **mandados** ou por **altruísmo** dos pesquisadores.



# Contextualizando



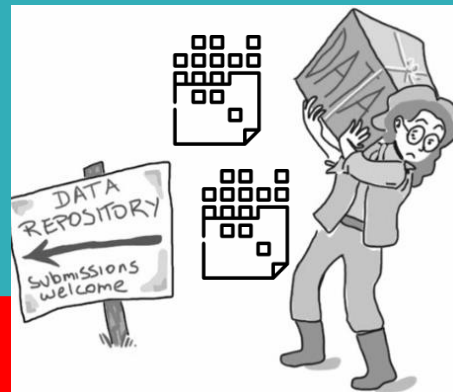
## Conscientização

Discussões sobre o ethos científico, propriedade e direitos aos dados



## Políticas e Mandatos

Diretivas para compartilhamento (Agências de Fomento, Editoras e Periódicos Científicos)



## Compartilhamento/ Publicação de Dados

Altruísmo e/ou coerção



## Acesso e Reuso

Maior quantidade de dados disponíveis para potencial reuso

# Publicação de Dados

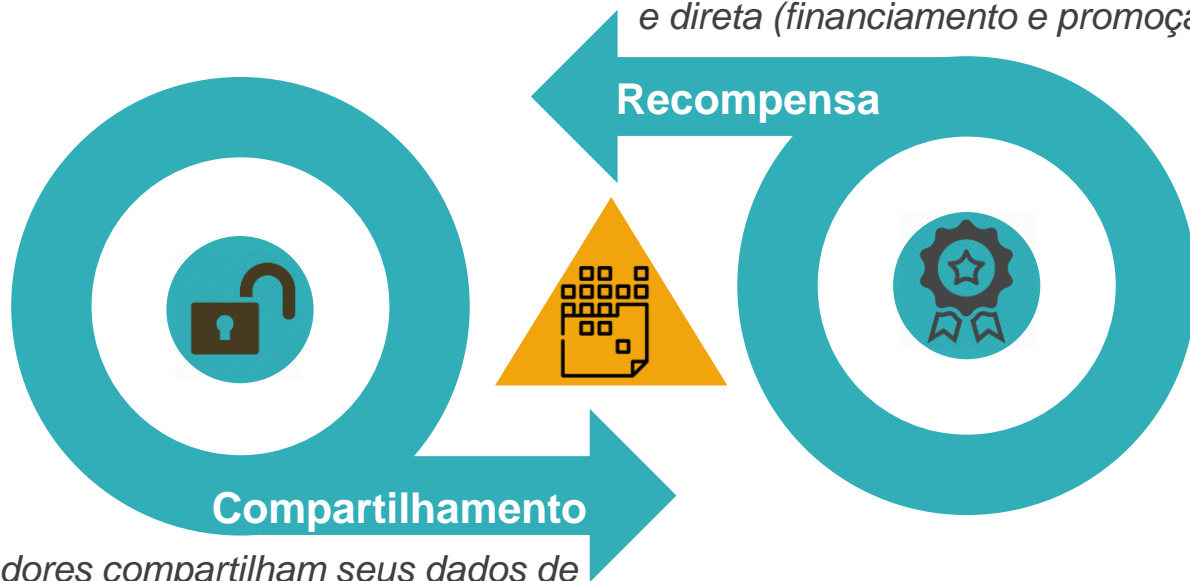
Divulgação ampla e pública dos dados de pesquisa de modo a torná-los **rastreáveis, acessíveis e passíveis de reuso** a outros pesquisadores para fins de verificação, e replicação, assim como para torná-los reutilizáveis além do propósito original para o qual foram coletados.

Devem estar acompanhados por **metadados associados, documentação e código de software**.  
(em casos de dados brutos processados ou manipulados)



# Publicação de Dados

*O reuso fica condicionado à atribuição de produtores/autores e reflete na lógica de recompensa simbólica (prestígio e reconhecimento) e direta (financiamento e promoção)*



*Pesquisadores compartilham seus dados de forma sistemática e organizada, seguindo padrões capazes de potencializar o reuso.*



# Publicação de Dados

## Principais Vantagens

01

### Maior Longevidade

Preservação e acesso à longo prazo

02

### Uso de Padrões

Requer a observância de padrões (mesmo que mínimos)

03

### Proveniência, Atribuição e Licenças

Garante a demarcação e atribuição de autoria/responsabilidade, bem como de embargo e licenças uso/reuso

04

### Maior Transparência

Possibilidade de verificação dos dados

05

### Redução de Esforços e de Custos

Evita duplicações em pesquisa

06

### Visibilidade e Citabilidade

Amplia a visibilidade e potencial de citação e impacto de pesquisas

# Vantagens para Citação

PeerJ

✓ PEER-REVIEWED [Bioinformatics and Genomics section](#)

## Data reuse and the open data citation advantage

View 52 tweets

Research article [Bioinformatics](#) [Science Policy](#)

Related research

Heather A. Piwowar<sup>1,2</sup>, Todd J. Vison<sup>1,2,3</sup>

Published October 1, 2013

PeerJ Picks - 2014 Highlighted in [PeerJ Picks 2014 Collection](#)

Note that a [Preprint of this article](#) also exists, first published April 4, 2013.

PubMed 24109559

Author and article information

Abstract

121 Total citations  
85 Recent citations  
22 Field Citation Ratio  
2.32 Relative Citation Ratio

Enter your institution  
To find colleagues at PeerJ

Enter to search

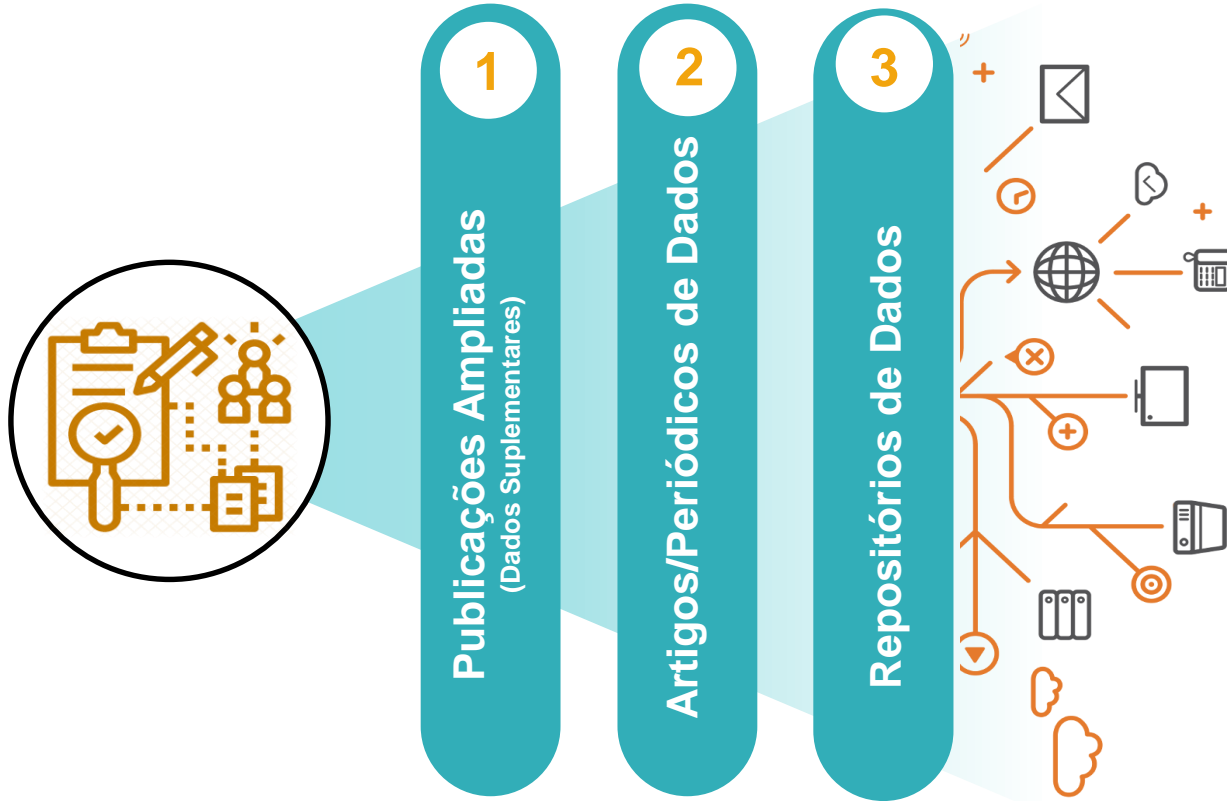
Download

Content Alert<sup>NEW</sup>  
Just enter your email

Tools & info  
Peer Review history  
See citing articles 120  
Ask questions 6

**Incremento de 9% em citações**

# Abordagens para a Publicação de Dados



# Publicações Tradicionais



# Publicações Ampliadas

- Complementação do artigo manuscrito por **módulos conectados de arquivos de dados executáveis e interligados**.
- Permitem processo de *peer review* de modo de modo **mais interativo**, e **potencializam o reuso de dados científicos**.
- Uma das primeiras tentativas para materializar a publicação de dados científicos **mais próxima do modelo de publicação científica tradicional**.



## Níveis de Sofisticação



1º Interligação de ativos externos e de estudos relacionados

2º Capacidades de interação por meio de recursos Web 2.0 pós-publicação (comentários, recomendação, ranking, avaliação)



3º Interligação de dados da própria pesquisa com objetivo de ilustrar e melhor evidenciar as informações



4º Possibilidade de reprodução com dados executáveis e recursos de interação para plotagem e visualização de dados

# Exemplo

The image shows a screenshot of a PeerJ article page. The article title is "Data reuse a advantage". A modal window titled "Articles citing this paper" is open, displaying a table of citation sources and a list of citing articles. The table shows 98 citations from CrossRef, 120 from Scopus, and a search option for Google Scholar. The list of citing articles includes:

- 2018 [Management, Archiving, and Sharing for Biologists and the Role of Research Institutions in the Technology-Oriented Age](#) *BioScience* Sébastien Renaut, Amber E Budden, Dominique Gravel, Timothée Poisot, Pedro Peres-Neto
- 2018 [Handbook of Knowledge Management for Sustainable Water Systems](#) Janet G. Hering, Lothar Nunnenmacher, Harald von Waldow
- 2018 [Practical Tips for Ethical Data Sharing](#) *Advances in Methods and Practices in Psychological Science* Michelle N. Meyer
- 2018 [Open Science Is Robust Science](#) *Industrial and Organizational Psychology* Samuel T McAfee, Joshua R Grubbs, Michael J Zickar

On the right side of the article page, there are statistics: 121 Total citations, 85 Recent citations, 22 Field Citation Ratio, and 2.32 Relative Citation Ratio. Below these are buttons for "Download" and "Content Alert". At the bottom right, there is a "Citing articles" button with a red badge showing "120". A blue circle with the number "01" and a yellow arrow points to this badge.

# Exemplo

investigating and comparing study methods, analysis techniques, and software implementations. The larger scientific community also benefits: sharing data from multiple perspectives, helps to identify errors, discourages redundancy, is useful for training new researchers, and increases efficiency of funding and patient population resources by avoiding duplicate data collection.

02

- 1 Making research data publicly available also has challenges and costs. Some costs are borne by society: For example, data archives must be created and maintained. Many costs, however, are borne by the data-collecting investigators: Data must be documented, formatted, and uploaded. Investigators may be afraid that other researchers will find errors in their results, or “scoop” additional analyses they have planned for the future.
- 1 Personal incentives are important to balance these personal costs. Scientists report that receiving additional citations is an important motivator for publicly archiving their data (Tenopir et al., 2011).
- There is evidence that studies that make their data available do indeed receive more citations than similar studies that do not (Gleditsch, Metelits & Strand, 2003; Piwowar, Day & Fridsma, 2007; Ioannidis et al., 2009; Pienta, Alter & Lyle, 2010; Henneken & Accomazzi, 2011; Sears, 2011; Dorch, 2012). These findings have been referenced by new policies that encourage and require data archiving (e.g., Rausher et al., 2010), demonstrating the appetite for evidence of personal benefit.

Questions / F1000 and other data publication models

Ask a question  
(Or learn more about O&A)

about Data reuse and the open data citation advantage

### F1000 and other data publication models

1 Do you see the F1000research data publications as a solution to the getting scooped issue you articulate here?

precedence data publications

Christopher Lortie · 1,702 · 5 years ago

Has there ever been an instance of scooping anyway? That was evident in the IR? - Christopher Lortie · 5 years ago

+ add comment note or answer instead

---

#### 4 Answers

Accepted answer

2 No, not really. The “scooped” issue is really just someone else examining and publishing on a research question before you do. This could happen after a F1000research data publication as well.

Embargoes (a fixed delay between the time data is deposited into a repository and the time it is made public) are an effective remedy for this. They should be used when the applicable research community decides that the disadvantages of scooping outweigh the advantages of the research field moving more quickly.

Heather Piwowar · 520 · 5 years ago

+ add comment note

---

2 I think the perception of getting scooped is so old school. Hopefully, the culture will change soon to give this notion up. I just read an article on networked discovery in Wired. Most major findings occur concurrently in many people once the correct conditions are in place.

Christopher Lortie · 1,702 · 5 years ago



# Exemplo

The screenshot shows the PLOS ONE website interface. At the top, there are navigation links for 'Publish', 'About', and 'Browse', along with a search bar. The article title is 'Analysis of Chaotic Resonance in Izhikevich Neuron Model' by Sou Nobukawa, Haruhiko Nishimura, Teruya Yamanishi, and Jian-Qin Liu. It is a research article published on September 30, 2015. The article has 7 saves, 8 citations, 2,098 views, and 1 share. A navigation menu includes 'Article', 'Authors', 'Metrics', 'Comments', and 'Related Content'. The 'Article' tab is selected, showing a table of contents with 'Abstract' highlighted. The abstract text is visible, starting with 'In stochastic resonance (SR), the presence of noise helps a nonlinear system amplify a weak (sub-threshold) signal.'

The Supporting Information panel displays a file tree for 'S1\_Data.zip'. The tree structure is as follows:

- S1\_Data.zip
  - S1
    - .DS\_Store
    - Fig1.xls
  - \_\_MACOSX
    - S1

03



# Exemplo

## CSD Entry: TOHTAN

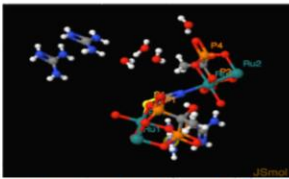
Your query was: DOI: doi:10.5517/cc12i5c6 and the search returned 1 record.

Database Identifier	Deposition Number
TOHTAN	1000567

Download -

TOHTAN : catena-[tetrakis(diaminomethaniminium) bis( $\mu$ -(1-hydroxyethane-1,1-diyl)bis(hydrogen (phosphonato)))-( $\mu$ -thiocyanato)-di-ruthenium tetrahydrate]  
Space Group:  $P \bar{1} (2)$ , Cell:  $a$  11.8622(13)Å  $b$  11.9892(13)Å  $c$  12.9119(14)Å,  $\alpha$  98.064(2) $^\circ$   $\beta$  110.1570(10) $^\circ$   $\gamma$  98.7280(10) $^\circ$

3D viewer

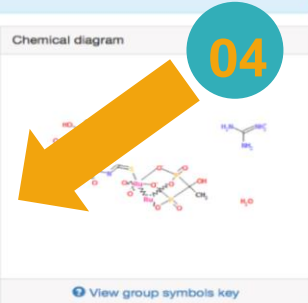


H Disorder  $\phi$  Menu Open -

Style: Ball and Stick Labels: All but C/H/N/O

Packing: None Measure: Distance

Chemical diagram



View group symbols key

Additional details

Deposition Number	1000567
Data Citation	Dan Wang, Huai-Qing Yang, Guang-Hao Wu, Xiao-Feng Hou, Jian-Hui Yang, Bin Liu CCDC 1000567: Experimental Crystal Structure Determination, 2014, DOI: 10.5517/cc12i5c6
Deposited on	01/05/2014

Associated publications

Dan Wang, Huai-Qing Yang, Guang-Hao Wu, Xiao-Feng Hou, Jian-Hui Yang, Bin Liu, *Inorganic Chemistry Communications*, 2014, 46, 241, DOI: 10.1016/j.inoche.2014.06.014

CCDC Home Deposit Structures Access Structures About This Service

## Recommended articles

Characterization of the Ribosome Biogenesi...  
Journal of Molecular Biology, Volume 425, Issu...  
Purchase PDF View details

Evolutionary morphology of the hemolymph ...  
Arthropod Structure & Development, Volume 41...  
Purchase PDF View details

Fluoroquinolone-metal complexes: A route t...  
Journal of Inorganic Biochemistry, Volume 138, ...  
Purchase PDF View details

1 2 Next >

## Citing articles (2)

Article Metrics

Captures

Exports-Saves: 1

Readers: 1

## Citations

Citation Indexes: Feedback

# Publicações Ampliadas | Limitações

- Demandam **sofisticada curadoria** de ativos científicos dispersos.
- Não permitem que os leitores encontrem e liguem dados **independentemente** do artigo científico.
- Muitas publicações ampliadas pertencem a **editores científicos comerciais**, o que torna questionável a garantia de acesso aberto aos dados e de potencial de reuso a longo prazo.



# Artigos de Dados

## Data Papers

Geralmente publicados em Periódicos dedicados a essa modalidade – Data Journals

- Artigos dedicados à descrição de conjunto(s) de dados (não analítico).
- Maior protagonismo destes ativos científicos, e na ampliação das condições de reuso.
- Preserva alguns preceitos essenciais do modelo de publicação de científica tradicional (condição de citação e atribuição de autoria aos criadores e geradores dos dados, e sistema de avaliação pelos pares).



Como dados suplementares (apêndices)

Dados integrados à narrativa descritiva

Apontamento para repositórios de dados

# Data Journals

## Estudo Empírico com Periódicos/Principais Seções de Conteúdo

Geoscience Data Journal

Biodiversity Data Journal

Journal of Open Psychological Data

Phytokeys

Data in Brief

Journal of Open Health Data

(Giga)<sup>n</sup> Science

Scientific Data

- Indicação da composição, formato, localização e formas de acesso e manipulação dos dados, incluindo sistemas e software para processamento.
- Notas sobre o uso dos dados, caso os dados relatados tenham já sido utilizados em outras modalidades de publicação e que indiquem as condições de uso e para uso futuros dos dados documentados no *data paper*.
- Potencial de reuso dos dados, articulando como eles podem ser reutilizados dentro e fora do domínio em que foram gerados/coletados, incluindo exemplos de agregação, verificação e replicação e uso dos dados no contexto de ensino de pesquisa.

[doi.org/10.5281/zenodo.842213](https://doi.org/10.5281/zenodo.842213)

# Artigos de Dados

- Permitem maior valorização dos dados gerados em pesquisa, por elevarem seu *status* a uma **publicação científica legítima** e capaz de ser **indexada por bases de dados**.
- Possibilitam a **descrição minuciosa dos dados**, facilitando a **verificação, replicação e reprodutibilidade** em pesquisa
- Aumentam a **visibilidade** de materiais suplementares que muitas vezes ficam encobertos e são de difícil localização
- Aumentam o tráfego de acesso a diferentes produções associadas ao conjunto de dados, possibilitando **mais citações** e descortinando **mais possibilidades para colaboração** entre pesquisadores com interesses comuns.

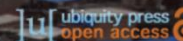




# Journal of open health data

Start Submission

Become a Reviewer



## Next Steps (formerly known as the Longitudinal Study of Young People in England)

Lisa Calderwood, Carole Sanchez

Follow on Twitter

Follow Via RSS

### About this Journal

Open Health Data features peer-reviewed data papers describing health datasets with high reuse potential. We are working with a number of specialist and institutional data repositories to ensure that the associated data are professionally archived, preserved, and openly available. Equally importantly, the data and the papers are citable, and reuse will be tracked.

#### LATEST ARTICLES

#### POPULAR ARTICLES



#### Seasonal Abundance of Fecal Indicators and Opportunistic Pathogens in Roof-Harvested Rainwater Tanks

Ahmed et al. — 03 Jul 2018

Share:



#### VBORNET gap analysis: Mosquito vector distribution models utilised to identify areas of potential species distribution in areas lacking records.

Schaffner et al. — 19 Dec 2016

Share:



#### VBORNET Gap Analysis: Sand Fly Vector Distribution Models Utilised to Identify Areas of Potential Species Distribution in Areas Lacking Records

Alten et al. — 10 Dec 2016

<https://openhealthdata.metajnl.com/articles/10.5334/ohd.28/>

# Exemplo

Data Paper

ZooKeys 489: 15-24 (23 Mar 2015)  
<https://doi.org/10.3897/zookeys.489.9292>

PDF XML Share article

## General taxonomic coverage description

All the recorded specimens in the dataset were identified to species. The identification of spider mites to species often requires the examination of male genitalia and specimens identified to genus were generally single females and have been discarded. Unidentified specimens have also been discarded. The dataset contains 175 species, i.e. 14 % of the species known in this family. Jean Gutierrez described 50 species (Table 1). Types of 49 are deposited in his collection.

Table 1.

[Download as CSV](#)

*Tetranychidae* species described by Jean Gutierrez with respective title and source of the publication and present nomenclature (Blommers and Gutierrez 1975; Gutierrez 1966; 1967a; b; 1968a; b; 1969; 1970; 1972; 1977; 1978; 1982; Gutierrez and Bolland 1973; 1986; Gutierrez and Etienne 1981; Gutierrez and Helle 1971). The presence of the type specimen and the number of types and paratypes in the dataset is also indicated.

Original genus	Species	Author	Present combination	Publication title	Publication source	Type
<i>Eonychus</i>	<i>grewiae</i>	Gutierrez,	<i>Eonychus grewiae</i>	Tetranychidae nouveaux	Acarologia, 11:	yes

Contents Article info Citation Metrics Comment Related

Figs Tabs Taxa Refs Cited

Article title

Abstract

Keywords

Data published through GBIF

Project details

Taxonomic coverage

— General taxonomic coverage description

Taxonomic ranks

— Spatial coverage

— Temporal coverage

— Natural collections description

Methods

— Method step description

— Uncertainty issues



# Exemplo

The screenshot shows the GigaScience website interface. At the top left is the GigaScience logo, and at the top right is the BGI logo (华大基因). A green navigation bar contains links for 'Articles', 'Submit', 'Alerts', and 'About', along with a search bar and a dropdown menu for 'All gigaScience'. The main article title is 'The draft genome sequence of a desert tree *Populus pruinosa*'. Below the title are the authors' names, the journal information (GigaScience, Volume 6, Issue 9, 1 September 2017), and the DOI link. A 'Published' date and an 'Article history' link are also present. A row of icons allows users to view the article in different formats: Views, PDF, Cite, Permissions, and Share. The abstract text is displayed in a light blue box. On the left side, there is a sidebar with 'Article Contents' including links to 'Abstract', 'Background', 'Data description', 'Availability of supporting data', 'Additional files', 'Abbreviations', 'Competing interests', 'Funding', 'References', and 'Supplementary data', which is highlighted with a red circle. On the right side, there are sections for 'View Metrics', 'Email alerts' (with options for new issue, in progress, advance article, and article activity alerts), and 'Related articles in' (with links to Google Scholar and CrossRef). At the bottom right, there are links for 'Latest', 'Most Read', and 'Most Cited'.

**(GIGA)<sup>n</sup> SCIENCE**

Articles Submit Alerts About

All gigaScience Search

**华大基因 BGI**

**The draft genome sequence of a desert tree *Populus pruinosa***

Wenlu Yang, Kun Wang, Jian Zhang, Jianchao Ma, Jianquan Liu, Tao Ma

GigaScience, Volume 6, Issue 9, 1 September 2017, Pages 1–7,  
<https://doi.org/10.1093/gigascience/gix075>

Published: 08 August 2017 Article history

Views PDF Cite Permissions Share

**Abstract**

*Populus pruinosa* is a large tree that grows in deserts and shows distinct differences in both morphology and adaptation compared to its sister species, *P. euphratica*. Here we present a draft genome sequence for *P. pruinosa* and examine genomic variations between the 2 species. A total of 60 Gb of clean reads from whole-genome sequencing of a *P. pruinosa* individual were generated using the Illumina HiSeq2000 platform. The assembled genome is 479.3 Mb in length, with an N50 contig size of 14.0 kb and a scaffold size of 698.5 kb; 45.47% of the genome is composed of repetitive elements. We predicted 35 131 protein-coding genes, of which 88.06% were functionally annotated. Gene family clustering revealed 224 unique and 640 expanded gene families in the *P. pruinosa* genome. Further evolutionary analysis identified numerous genes with elevated values for pairwise genetic differentiation between *P. pruinosa* and *P. euphratica*. We provide the genome sequence and gene annotation for *P. pruinosa*. A large number of genetic variations were recovered by comparison of the genomes between *P. pruinosa* and *P. euphratica*. These variations will provide a valuable resource for studying the genetic bases for the phenotypic and adaptive divergence of the 2

**Article Contents**

Abstract

Background

Data description

Availability of supporting data

Additional files

Abbreviations

Competing interests

Funding

References

**Supplementary data**

< Previous Next >

**View Metrics**

**Email alerts**

New issue alert

In progress issue alert

Advance article alerts

Article activity alert

Receive exclusive offers and updates from Oxford Academic

**Related articles in**

Google Scholar

**Citing articles via**

Google Scholar

CrossRef

**Latest | Most Read | Most Cited**

# Exemplo


**Geoscience Data Journal** [Open Access](#)  
[Explore this journal >](#)

[PDF](#)  
[Info](#)  
[References](#)  
[Figures](#)

[Open Access](#) [Creative Commons](#)

Data Paper

## Forest cover mask from historical topographic maps based on image processing

K. Ostafin , M. Iwanowski, J. Kozak, A. Cacko, U. Gimmi, D. Kaim, A. Psomas, C. Ginzler, K. Ostapowicz

First published: June 2017 [Full publication history](#)  
DOI: 10.1002/gdj.3.46 [View/save citation](#)

Cited by (CrossRef): 0 articles [Check for updates](#) [Citation tools](#) ▼

[Open Access](#) 3

Funding Information

Research carried out within the FORECOM project ( *Forest cover changes in mountainous regions - drivers, trajectories and implications* . PSRP-008/2010), supported by a grant from Switzerland through the Swiss contribution to the enlarged European Union.

### Abstract

This study aimed to obtain accurate binary forest masks which might be directly used in analysis of land cover changes over large areas. A sequence of image processing operations was conceived, parameterized and tested using various topographic maps from mountain areas in Poland and Switzerland. First, the input maps were filtered and binarized by thresholding in Hue-Saturation-Value colour space. The second step consisted of a set of morphological image analysis procedures leading to final forest masks. The forest masks were then assessed and compared to manual forest boundary vectorization. The Polish topographical map published in the 1930s showed low accuracy which could be attributed to methods of cartographic presentation used and degradation of original colour prints. For maps published in the 1970s, the automated forest extraction performed very well, with accuracy exceeding 97%, comparable to accuracies of manual vectorization of the same maps performed by nontrained operators. With this method, we obtained a forest cover mask for the entire area of the Polish Carpathians, easily readable in any Geographic Information System software.

### Dataset

Identifier: <https://doi.org/10.4171/journal-ar3740ad19e6947cd193ra-71fr04f3309>

[View issue TOC](#)  
Volume 4, Issue 1  
June 2017  
Pages 29-39

[Abstract](#)  
[Dataset](#)  
[Introduction](#)  
[1 Materials](#)  
[2 Data Production Method](#)  
[3 Results](#)  
[4 Discussion](#)  
[5 Conclusions](#)  
[6 Data Location and Form](#)  
[Acknowledgements](#)  
[References](#)  
[Related Content](#)

# Artigos de Dados | Restrições

- Não são adequados para disseminar todos os tipos de dados. São adequados apenas para descrever dados relativamente **estáveis e em menor escala**
- Demandam **tempo e custo** dos pesquisadores que tendem a dar prioridade à publicação



# Repositórios de Dados

Essenciais para o ecossistema da publicação de dados, e constituem por si só como uma abordagem de *data publishing*, uma vez que tornam públicas coleções de dados acompanhadas por recursos que otimizem seu potencial de reuso.



# PANORAMA DOS REPOSITÓRIOS DE DADOS

## Repositórios

1º Estados Unidos

981

2º Alemanha

324

3º Grã-Bretanha

290

4º Canadá

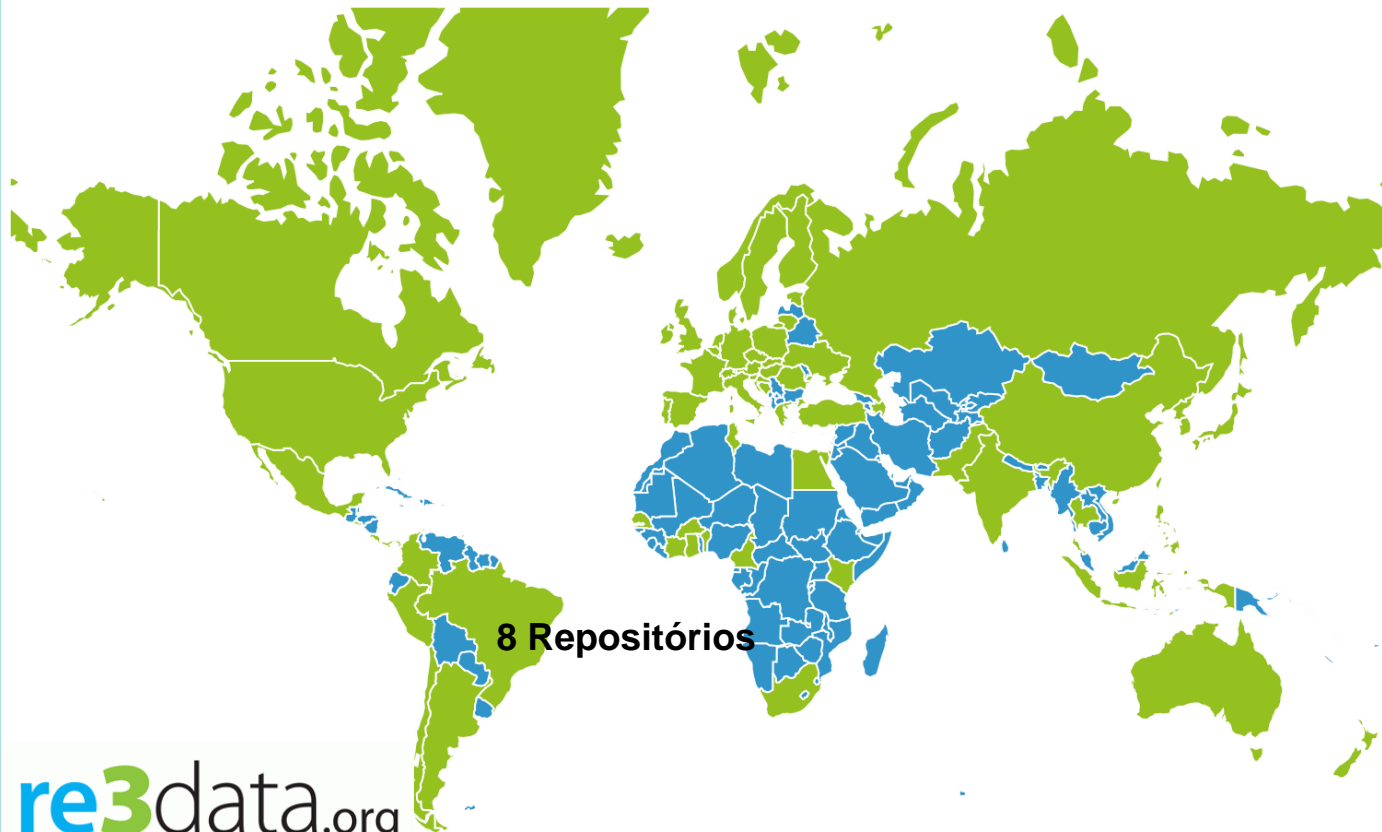
145

5º França

97

6º Austrália

84



# Repositórios de Dados | Exemplo - Figshare

This website uses cookies to help you have a better on-line experience. By using this website, you are agreeing to the use of cookies as explained in our [privacy policy](#).

Accept cookies



search on figshare



Browse

Upload

Sign up

Log in

store, share, discover **research**

get more citations for all of the outputs of your academic research  
over 5000 citations of figshare content to date

ALSO FOR INSTITUTIONS & PUBLISHERS

"figshare wants to open up scientific data to the world" **WIRED**

- Genérico (todas as áreas)
- Não exclusivo para dados e Datasets
- Baixa validação e exigência de metadados



## Repositórios de Dados | Limitações

**Abordagem de publicação de dados** muitas vezes considerada **menos atrativa** do ponto de vista daqueles que compartilham os dados, em termos de **recompensa** e **crédito científico**.



- Fornece ferramentas de busca, de visualização e de exportação dos dados.
- Segue padrões rigorosos de *workflow* interno para validação dos dados, e atribuição de licença de reuso.
- Opera sob princípios de endossamento buscando a qualidade dos dados, que eles sejam relevantes ao escopo e objetivos para a comunidade.
- Custódia e curadoria dos dados estável e persistente.
- *Feedback* de avaliação.




# IPT GBIF

Get data Share Tools Inside GBIF

## The Integrated Publishing Toolkit (IPT)

*A free open source software tool used to publish and share biodiversity datasets through the GBIF network.*



© OpenStreetMap contributors, © OpenMapTiles

229 INSTALLATIONS

YouTube BR

### INTEGRATED PUBLISHING TOOLKIT (IPT)

https://ipt.gbif.org/manage/

Apps The Kurator Nelson... GBIF Nodes 50 Countries Qu... Collage Maker | Be... GBIF Engage - For...

GBIF INTEGRATED PUBLISHING TOOLKIT TEST MODE

Logged in as kowless@gbif.org Assist Logout

Home Manage Resources Administration About

#### Resources you have rights to manage

Name	Organisation	Type	Subtype	Records	Last modified	Last publication	Next publication	Visibility	Author
Dromas, Emerald Dromas, Emerald	Not registered	Occurrence	Specimen	1	2017-06-13	2017-06-13	2017-06-20 12:37:36	Public	Kyle Brink
FPS, Great Plains Collection	Not registered	Occurrence	Specimen	0	2017-06-13	Not published	—	Private	Laura Russell
Birds at Darwin Lighthouse, Darwin	Not registered	Occurrence	Specimen	1,212	2017-06-12	2017-06-12	—	Public	Kurtis Mathews
Collection of Occurrence Events of Canada, CANADA	Not registered	Checklist	—	27,880	2017-06-12	2017-06-12	—	Public	Kyle Brink
Emerald Dromas, Emerald Dromas	Not registered	Occurrence	Specimen	9	2017-06-12	2017-06-12	—	Public	Kyle Brink
Natural History Collections of Romania	Not registered	Occurrence	—	1,212	2017-06-12	2017-06-12	—	Public	Kurtis Mathews
Dromas, Red Dromas	Not registered	Occurrence	—	9	2017-06-12	2017-06-12	—	Public	Daniel Noveckoff
Lighthouses, Birds	Not registered	Occurrence	—	1,212	2017-06-12	2017-06-12	—	Public	Kurtis Mathews
Great Lakes, Lake Erie	Not registered	Occurrence	—	249	2017-06-11	2017-06-11	—	Public	Simone A. ...
...	...	...	...	...	...	...	...	Public	...

0:14 / 24:16

### Introduction to publishing using the GBIF Integrated Publishing Toolkit (IPT)

1,311 views

1 0 SHARE

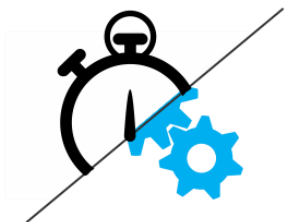
Gbif Secretariat  
Published on Jul 10, 2017

SUBSCRIBE 79

This demo shows how to publish a species occurrence dataset and register it with GBIF.org using the Integrated Publishing Toolkit (IPT) with special emphasis on how to properly map fields to Dwc terms while complying with vocabularies and making sure to satisfy the IPT's

# Artigos de dados automáticos

- Oferecem uma alternativa de incentivo aos pesquisadores, com relação à **obtenção de créditos, capaz de envolvê-los e mobilizá-los** mais efetivamente a compartilharem e publicarem seus dados.
- Além de poupar o tempo e esforço dos autores, **uniformiza e garante melhor documentação e preservação dos dados**, ampliando seu potencial de reuso.
- Mecanismo para publicação **mais ágil**, que confere créditos endossados pelo ecossistema científico, e também garante a qualidade dos dados e metadados publicados por meio do processo de revisão pelos pares.



# Sintetizando as abordagens

- **Publicações ampliadas:** possibilitam a interligação de módulos de dados fragmentados, acessórios e dependentes do artigo científico e muitas vezes de editores científicos comerciais.
- **Artigos de dados:** elevam os dados científicos à condição de protagonistas, pois se dedicam a descrever exaustivamente a coleção de dados, e quando publicados em periódicos de dados conservam características de ampla aquiescência entre os membros da comunidade científica, como *peer-review*, periodicidade e corpo editorial, ampliando seu potencial de receptividade no meio científico.
- **Repositórios de dados:** têm investido em ferramentas para citação e encontrabilidade dos dados, e algumas iniciativas buscam minimizar a falta de visibilidade e legitimidade da publicação por meio de artigos de dados automáticos.

# Desafios & Oportunidades

Contexto Brasileiro

Políticas de  
Compartilhamento  
integrado ao processo de  
geração de dados  
(DMPs/Comitês de  
ética/Plataforma Brasil)



Integração com Lattes/vínculo com  
demais produções

Participação Ativa dos  
Comitês/Colégios na  
Avaliação de Rep. de Dados



## Publicação de Dados



Política de Incentivo e  
Recompensas – Agências de  
Fomento e Instituições de  
Ensino e Pesquisa

Consolidação de  
Repositórios de Dados  
(disciplinares) –  
arquivamento automático e  
buscas federadas



Sofisticação de nossos  
periódicos científicos  
(plataformas e políticas)



Ampla adoção de  
Identificadores Persistentes  
(PIDs)



# Obrigada!

recurty@gmail.com

