



¿Cómo compartir datos biológicos?

Florencia Grattarola



Gestión y análisis de registros de fauna con
cámaras trampa utilizando software R

Montevideo, 12 de octubre 2018

Pasos para compartir mis datos

- 1. Elegir un repositorio de datos.**
- 2. Estandarizar la información.**
- 3. Elegir una licencia (derechos) para permitir su uso.**
- 4. Escribir metadatos descriptivos del conjunto de datos.**



1. Elegir un repositorio de datos

1. Elegir un repositorio de datos

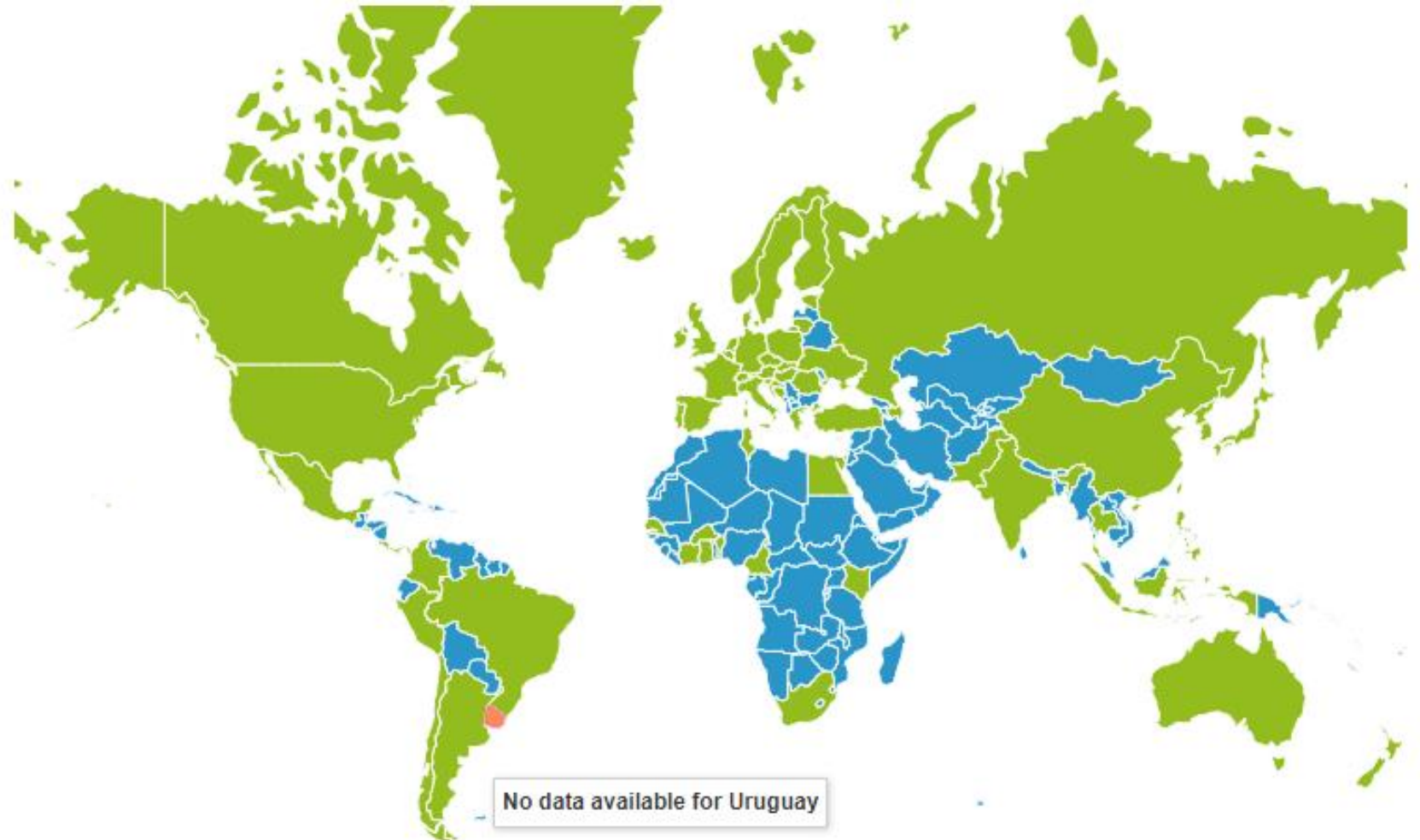
re3data.org
REGISTRY OF RESEARCH DATA REPOSITORIES

**Registro global de
repositorios de datos de
investigación**

Browse by subject

Browse by content type

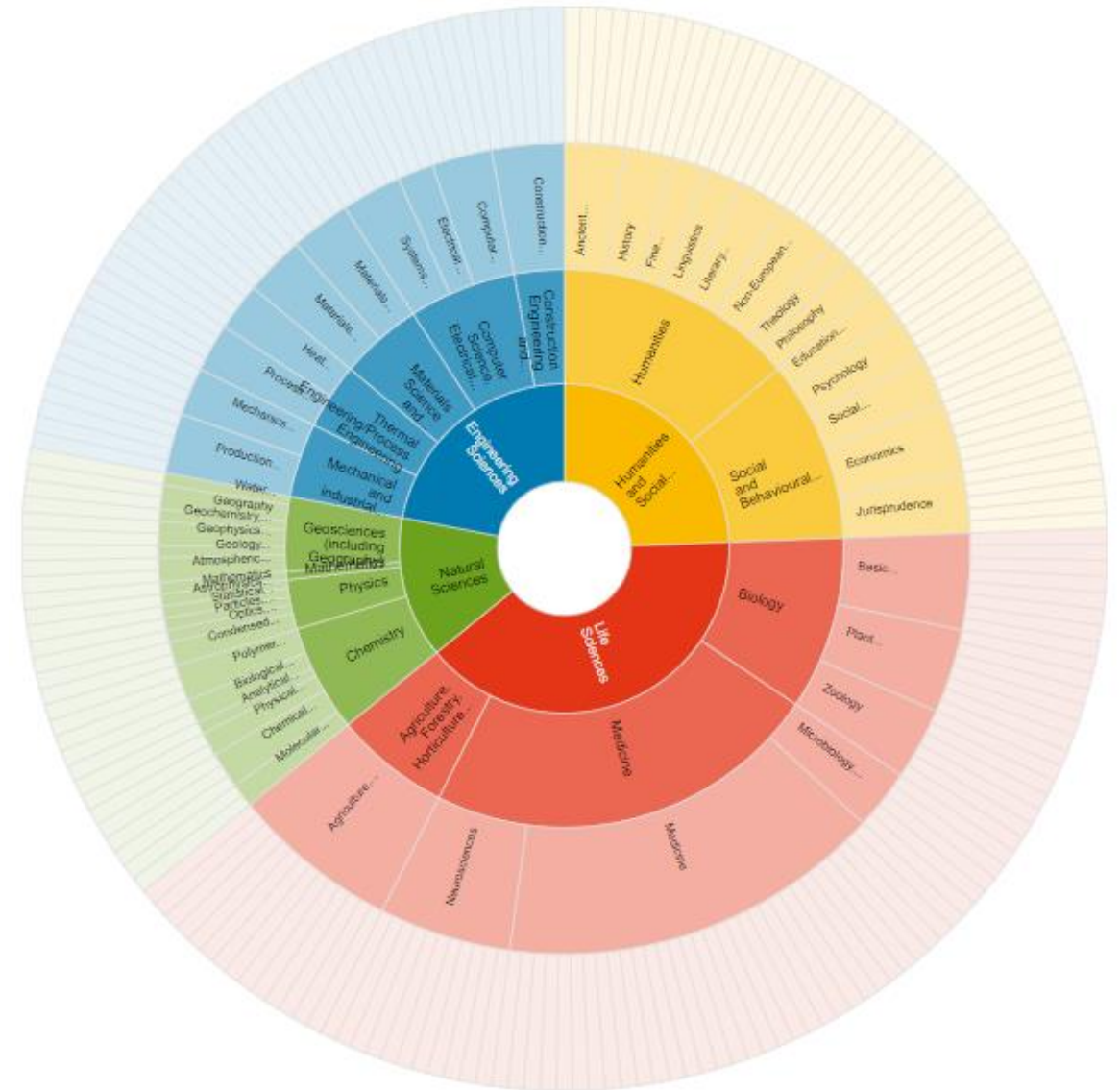
Browse by country



1. Elegir un repositorio de datos



- Natural sciences
- Engineering sciences
- Humanity & Social sciences
- Life sciences



<https://www.re3data.org/browse/by-subject/>

1. Elegir un repositorio de datos



Repository details



VertNet

General Institutions Terms Standards

Name of repository	VertNet
Repository URL	http://vertnet.org/
Subject(s)	Plant Sciences Zoology Biology Life Sciences
Description	<p>VertNet is a NSF-funded collaborative project that makes biodiversity data free and available on the web. VertNet is a tool designed to help people discover, capture, and publish biodiversity data. It is also the core of a collaboration between hundreds of biocollections that contribute biodiversity data and work together to improve it. VertNet is an engine for training current and future professionals to use and build upon best practices in data quality, curation, research, and data publishing. Yet, VertNet is still the aggregate of all of the information that it mobilizes. To us, VertNet is all of these things and more.</p>
Contact	http://vertnet.org/feedback/contact.html
Content type(s)	Standard office documents Databases Images Structured graphics Scientific and statistical data formats Plain text Raw data other
Keyword(s)	biodiversity vertebrates biology natural history invertebrates plants fungi paleo
Repository size	over 20,546,275 records from 282 data resources, containing 370 collections, shared by 112 publishers globally
Repository type(s)	disciplinary institutional
Mission statement for designated community	http://vertnet.org/about/about.html
Research data repository language(s)	eng
Data and/or service provider	dataProvider serviceProvider

1. Elegir un repositorio de datos



Repository details



Global Biodiversity Information Facility

General Institutions Terms Standards

Name of repository	Global Biodiversity Information Facility
Additional name(s)	Data index of the Global Biodiversity Information Facility GBIF.org
Repository URL	https://www.gbif.org/
Subject(s)	Biology Agriculture, Forestry, Horticulture and Veterinary Medicine Life Sciences
Description	GBIF is an international organisation that is working to make the world's biodiversity data accessible everywhere in the world. GBIF and its many partners work to mobilize the data, and to improve search mechanisms, data and metadata standards, web services, and the other components of an Internet-based information infrastructure for biodiversity. GBIF makes available data that are shared by hundreds of data publishers from around the world. These data are shared according to the GBIF Data Use Agreement, which includes the provision that users of any data accessed through or retrieved via the GBIF Portal will always give credit to the original data publishers.
Contact	https://www.gbif.org/contact-us
Content type(s)	Images Structured graphics Scientific and statistical data formats Structured text Software applications
Certificates and Standards	World Data System
Keyword(s)	biodiversity animal plant fungi protozoa chromista bacteria archaea virus
Persistent identifier(s) of the repository	RRID:SCR_005904 RRID:nlx_149475 FAIRsharing_doi:10.25504/FAIRsharing.zv11j3 OMICS_19240
Repository size	964.313.520 occurrence records; 37.614 datasets;
Repository type(s)	institutional disciplinary
Mission statement for designated community	https://www.gbif.org/what-is-gbif

1. Elegir un repositorio de datos

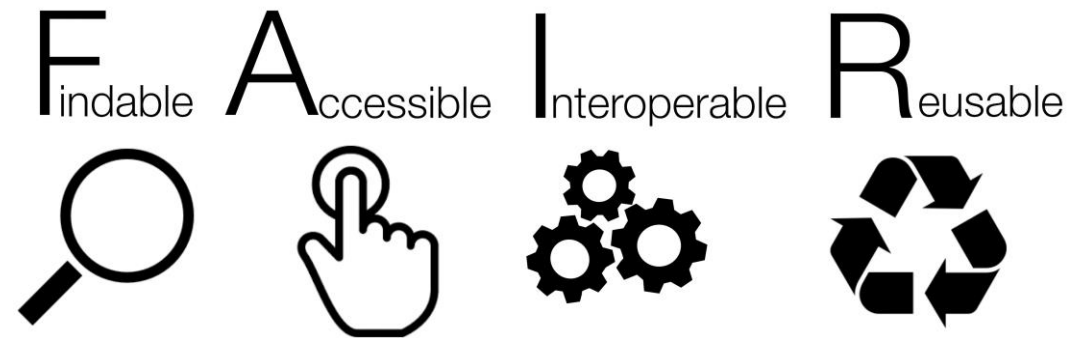


La red GBIF de países y organizaciones participantes, trabaja a través de nodos participantes, proporciona a las instituciones de almacenamiento de datos de todo el mundo estándares comunes y herramientas de código abierto que les permiten compartir información sobre dónde y cuándo se han encontrado las especies.



2. Estandarizar la información

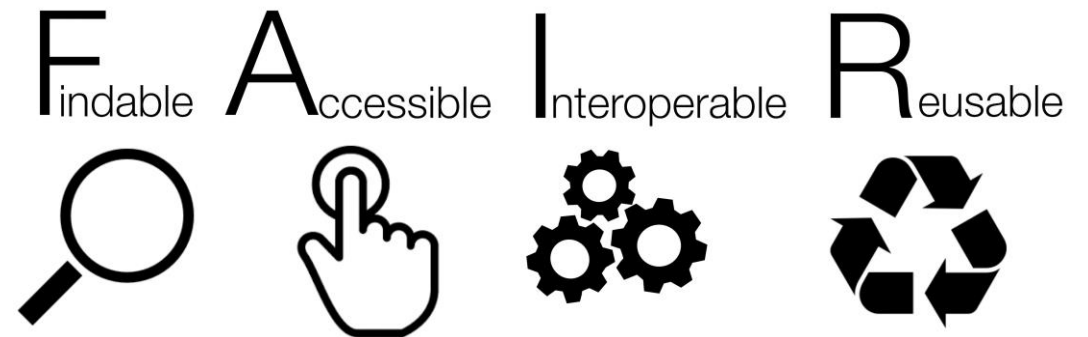
2. Estandarizar la información



FINDABLE – Encontrables

Esto incluye asignar un identificador persistente (como un DOI), tener metadatos enriquecidos para describir los datos y asegurarse de que se puedan encontrar a través de repositorios (locales e internacionales).

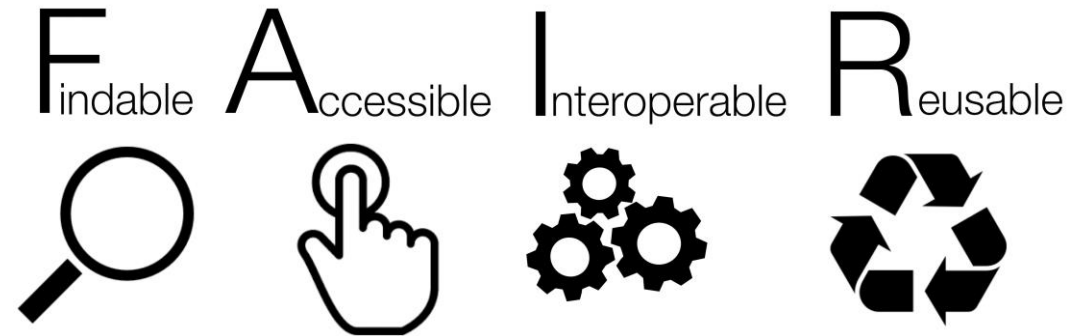
2. Estandarizar la información



ACCESSIBLE – Accesibles

Esto puede incluir hacer que los datos se abran usando un protocolo estandarizado. Sin embargo, los datos no necesariamente tienen que estar abiertos. A veces hay buenas razones por las cuales los datos no se pueden abrir, por ejemplo, preocupaciones de privacidad, seguridad nacional o intereses comerciales. Si no está abierto, debe haber claridad y transparencia en torno a las condiciones que rigen el acceso y la reutilización.

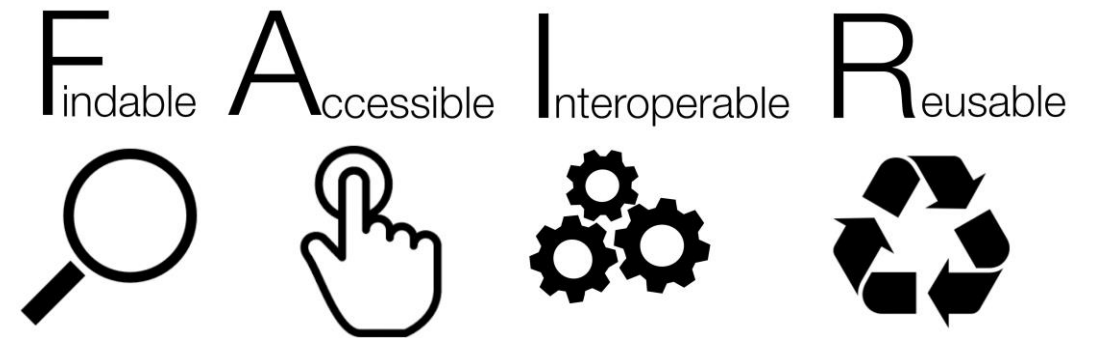
2. Estandarizar la información



INTEROPERABLE – Interoperables

Para ser interoperable, los datos deberán usar formatos, lenguaje y vocabularios acordados por la comunidad. Los metadatos también deberán usar estándares y vocabularios acordados por la comunidad, y contener enlaces a información relacionada utilizando identificadores.

2. Estandarizar la información



RE-USABLE - Re-utilizables

Los datos reutilizables deben mantener su riqueza inicial. Por ejemplo, no debe disminuirse a los efectos de explicar los hallazgos en una publicación en particular. Necesita una licencia legible por máquina clara e información de procedencia sobre cómo se formaron los datos.

2. Estandarizar la información

Biodiversity
Information
Standards
TDWG

Estándar diseñado con el propósito de crear un lenguaje común para publicar y documentar datos sobre registros biológicos (observaciones o ejemplares de colección), listas de especies y catálogos taxonómicos.

2. Estandarizar la información

Record table

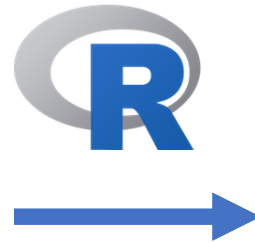


Tabla estandarizada

2. Estandarizar la información

Record table

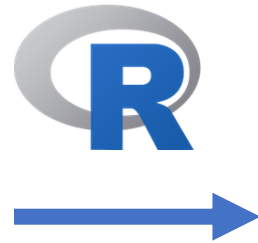


Tabla estandarizada



2. Estandarizar la información

Términos recomendados del estándar Darwin Core

Occurrence

occurrenceID	catalogNumber	recordNumber	recordedBy	individualCount	organismQuantity	organismQuantityType	sex
lifeStage	reproductiveCondition	behavior	establishmentMeans	occurrenceStatus	preparations	disposition	
associatedMedia	associatedReferences	associatedSequences	associatedTaxa	otherCatalogNumbers	occurrenceRemarks		

Taxon

taxonID	scientificNameID	acceptedNameUsageID	parentNameUsageID	originalNameUsageID	nameAccordingToID		
namePublishedInID	taxonConceptID	scientificName	acceptedNameUsage	parentNameUsage	originalNameUsage		
nameAccordingTo	namePublishedIn	namePublishedInYear	higherClassification	kingdom	phylum	class	order
genus	subgenus	specificEpithet	infraspecificEpithet	taxonRank	verbatimTaxonRank	scientificNameAuthorship	
vernacularName	nomenclaturalCode	taxonomicStatus	nomenclaturalStatus	taxonRemarks			

Más de 160
términos
(<http://rs.tdwg.org/dwc/terms/>)

2. Estandarizar la información

Términos recomendados del estándar Darwin Core

Occurrence

occurrenceID	catalogNumber	recordNumber	recordedBy	individualCount	organismQuantity	organismQuantityType	sex
lifeStage	reproductiveCondition	behavior	establishmentMeans	occurrenceStatus	preparations	disposition	
associatedMedia	associatedReferences	associatedSequences	associatedTaxa	otherCatalogNumbers	occurrenceRemarks		

Taxon

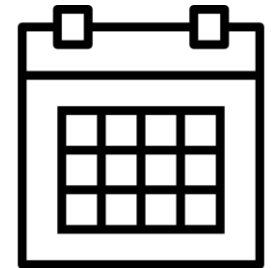
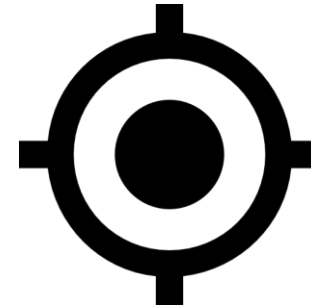
taxonID	scientificNameID	acceptedNameUsageID	parentNameUsageID	originalNameUsageID	nameAccordingToID			
namePublishedInID	taxonConceptID	scientificName	acceptedNameUsage	parentNameUsage	originalNameUsage			
nameAccordingTo	namePublishedIn	namePublishedInYear	higherClassification	kingdom	phylum	class	order	family
genus	subgenus	specificEpithet	infraspecificEpithet	taxonRank	verbatimTaxonRank		scientificNameAuthorship	
vernacularName	nomenclaturalCode	taxonomicStatus	nomenclaturalStatus	taxonRemarks				

Más de 160
términos

(<http://rs.tdwg.org/dwc/terms/>)

2. Estandarizar la información

Términos recomendados del estándar Darwin Core



especie

localización geográfica

fecha

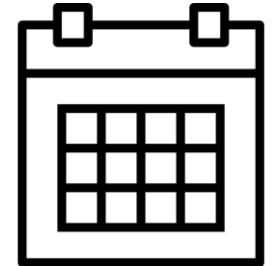
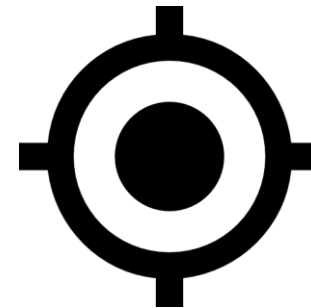
campos mínimos

2. Estandarizar la información

Términos recomendados del estándar Darwin Core

Tablas para descargar con los campos como columnas:

<https://github.com/tdwg/dwc/tree/master/dist>



especie

localización geográfica

fecha

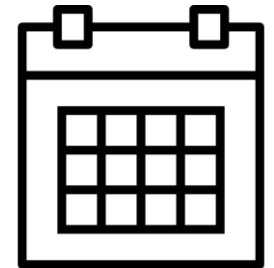
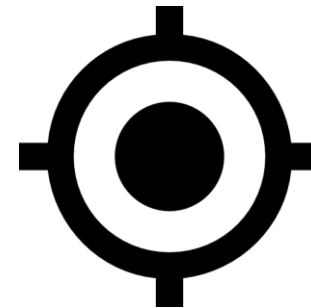
campos mínimos

2. Estandarizar la información

occurrenceID
basisOfRecord
institutionCode
collectionCode
catalogNumber
recordNumber
recordedBy
eventDate
habitat
stateProvince
county
minimumElevationInMeters
maximumElevationInMeters
verbatimLatitude
verbatimLongitude
decimalLatitude
decimalLongitude
scientificName
measurementType
measurementValue
measurementUnit



GBIF URUGUAY



especie

localización geográfica


fecha

campos mínimos


2. Estandarizar la información

1

Branch: master **dwc / dist /**

 peterdesmet Rebuild with lowerCamelCase labels for Dublin Core terms ...

..

 [simple_dwc_horizontal.csv](#) Rebuild with lowerCamelCase labels for Dublin Core terms


 [simple_dwc_vertical.csv](#) Rebuild with lowerCamelCase labels for Dublin Core terms

<https://github.com/tdwg/dwc/tree/master/dist>

2

Branch: master **dwc / dist / simple_dwc_horizontal.csv**

Find file Copy path

 peterdesmet Rebuild with lowerCamelCase labels for Dublin Core terms

8141947 on Oct 17, 2017

2 contributors  

2 lines (1 sloc) | 2.7 KB

 Search this file...

	1	type	modified	language	license	rightsHolder	accessRights	bibliographicCitation	references	institutionID	collec

Open link in new tab
Open link in new window
Open link in incognito window

Save link as...

Copy link address

Inspect

Ctrl+Shift+I

Souce	Function prefix
Encyclopedia of Life	eol
Taxonomic Name Resolution Service	tnrs
Integrated Taxonomic Information Service	itis
Global Names Resolver	gnr
Global Names Index	gni
IUCN Red List	iucn
Tropicos	tp
Theplantlist dot org	tpl
Catalogue of Life	col
National Center for Biotechnology Information	ncbi
CANADENSYS Vscan name search API	vascan
International Plant Names Index (IPNI)	ipni
Barcode of Life Data Systems (BOLD)	bold
National Biodiversity Network (UK)	nbn
Index Fungorum	fg
EU BON	eubon
Index of Names (ION)	ion
Open Tree of Life (TOL)	tol
World Register of Marine Species (WoRMS)	worms
NatureServe	natserv
Wikipedia	wiki





3. Elegir una licencia (derechos) para permitir su uso

3. Elegir una licencia (derechos) para permitir su uso

ABIERTO



1. Licencia abierta ←
2. Formato abierto legible por máquinas
3. Descargar todo el conjunto de datos
4. Actualizado (al día)
5. Disponible públicamente
6. Disponible gratis

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

No

Quiero compartir mis contenidos sin restricción

Si

Si, mientras se comparta bajo la misma licencia

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

No



Todos los derechos reservados

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Quiero compartir mis contenidos sin restricción



Dominio Público

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Si

Si, mientras se comparta bajo la misma licencia

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Si

Si, mientras se comparta bajo la misma licencia

¿Quieres permitir el uso de tus contenidos con fines comerciales?

Si

No



Licencia Atribución

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Si

Si, mientras se comparta bajo la misma licencia

¿Quieres permitir el uso de tus contenidos con fines comerciales?

Si

No



Licencia Atribución, No comercial

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Si

Si, mientras se comparta bajo la misma licencia

¿Quieres permitir el uso de tus contenidos con fines comerciales?

Si

No



Licencia Atribución, Compartir Igual

3. Elegir una licencia (derechos) para permitir su uso

¿Quieres que otras personas puedan usar y compartir tus contenidos?

Si

Si, mientras se comparta bajo la misma licencia

¿Quieres permitir el uso de tus contenidos con fines comerciales?

Si

No

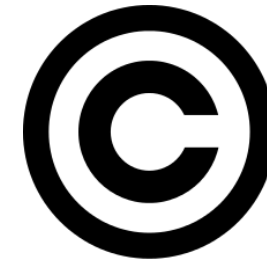


Licencia Atribución, No comercial, Compartir Igual

3. Elegir una licencia (derechos) para permitir su uso



Copyleft



Copyright

3. Elegir una licencia (derechos) para permitir su uso



Compartir sin restricciones



Compartir, con atribución



Compartir, con atribución y sin fines comerciales



Compartir, con atribución y obras derivadas de igual licencia



Compartir, con atribución, obras derivadas de igual licencia y sin fines comerciales





4. Escribir metadatos descriptivos del conjunto de datos

4. Escribir metadatos descriptivos del conjunto de datos

Parámetro	Descripción
title	Descripción del conjunto de datos.
summary	Párrafo que describe el conjunto de datos.
project	Nombre del proyecto que generó los datos.
fund	Detalle de la financiación.
investigators	Investigadores involucrados en la generación del conjunto de datos.
contact	Contacto para más información.
dateCreated	Fecha en la que fue creado el conjunto de datos.
keyWords	Lista de palabras claves para asociar el archivo.
qualityLevel	Descripción textual del nivel de control de calidad de los datos, incluyendo problemas conocidos que pueden limitar el uso de datos y/o razones de valores faltantes (NAs).
definitions	Etiquetas y descripciones de nuevas variables, definiciones de acrónimos o terminología especializada.

¿Cómo puede ayudarme R a liberar mis datos?

1. Elegir un repositorio de datos.
2. Estandarizar la información. 
3. Elegir una licencia (derechos) para permitir su uso.
4. Escribir metadatos descriptivos del conjunto de datos. 

Estándares de datos y metadatos

occurrenceID
basisOfRecord
institutionCode
collectionCode
catalogNumber
recordNumber
recordedBy
eventDate
habitat
stateProvince
county
minimumElevationInMeters
maximumElevationInMeters
verbatimLatitude
verbatimLongitude
decimalLatitude
decimalLongitude
scientificName
measurementType
measurementValue
measurementUnit

Parámetro	Descripción
title	Descripción del conjunto de datos.
summary	Párrafo que describe el conjunto de datos.
project	Nombre del proyecto que generó los datos.
fund	Detalle de la financiación.
investigators	Investigadores involucrados en la generación del conjunto de datos.
contact	Contacto para más información.
dateCreated	Fecha en la que fue creado el conjunto de datos.
keyWords	Lista de palabras claves para asociar el archivo.
qualityLevel	Descripción textual del nivel de control de calidad de los datos, incluyendo problemas conocidos que pueden limitar el uso de datos y/o razones de valores faltantes (NAs).
definitions	Etiquetas y descripciones de nuevas variables, definiciones de acrónimos o terminología especializada.

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



rgbif

https://ropensci.org/tutorials/rgbif_tutorial/

paquete R de rOpenSci que permite buscar y recuperar datos de GBIF.

envuelve código R alrededor de la API de GBIF para permitirle hablar con GBIF desde R y acceder a metadatos, nombres de especies y ocurrencias.



fGrattarola@lincoln.ac.uk



flograttarola.netlify.com



Fotos: Flickr JulanaJulana <https://www.flickr.com/photos/julana>
Íconos hechos por: Smashicons, Freepik y srip en
www.flaticon.com

Esta presentación se puede descarga

