



Formation GBIF France dans le cadre d'Ecoscope - Valoriser ses données d'observation sur la biodiversité : qualité, standards et publication

Paris, 15-16 septembre 2015

Outils et ressources pour évaluer et améliorer l'aptitude des données à être utilisées

GBIF France (gbif@gbif.fr)

Basé sur la présentation de Nicolas Noé - niconoe@ulb.ac.be
pour GB18 training sessions - Buenos Aires, Argentine (sept 2011)

Aperçu

- Outils pour:
 - Métadonnées
 - Données spatiales
 - Données tabulaires
- Autres ressources
 - Données
 - Documents




Outils

Pour les métadonnées




Métadonnées et IPT

(Dwc-A)



free and open access to biodiversity data
GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

Logged in as [Account](#) [Logout](#) 
[gasc@gbif.fr](#)

[Home](#) [Manage Resources](#) [Administration](#) [About](#)


Basic Metadata: Collection entomologique Barthélémy


You must fill in at least these basic metadata before you can make this resource public.

For each contact you must supply at least a last name, a position, or an organisation. Title and Description are required.

Title

Description

Metadata Language 

Resource Language 

Subtype

Resource Contact

| | |
|---|---|
| First Name <input type="text" value="Marie-Françoise"/> | Last Name <input type="text" value="Faure"/> |
| Position <input type="text" value="Directrice-adjointe, Responsable des collections"/> | Organisation <input type="text" value="Muséum Henri-Lecoq"/> |
| Address <input type="text"/> | City <input type="text"/> |





free and open access to biodiversity data
GBIF INTEGRATED PUBLISHING TOOLKIT (IPT)

Logged in as [Account](#) [Logout](#)

Home

Manage Resources

Administration

About

Geographic Coverage: Collection entomologique Barthélémy

Drag the markers or fill in the fields to set the geographic bounding box of the area covered by the resource.



Set global coverage

☐

Minimum Longitude

-4.73

Maximum Longitude

8.59

Minimum Latitude

41.48

Maximum Latitude

50.98

Description

France métropolitaine

Section

[Basic Metadata](#)

[Geographic Coverage](#)

[Taxonomic Coverages](#)

[Temporal Coverages](#)

[Other Keywords](#)

[Associated Parties](#)

[Project Data](#)

[Sampling Methods](#)

[Citations](#)

[Collection Data](#)

[External links](#)

[Additional Metadata](#)



Couverture géographique

Darwin Core Archive Validator



Darwin Core Archive Validator

[home](#) [emi](#) [extensions](#) [api](#) [about](#)

Darwin Core Archive Validator

You can either copy paste a meta.xml descriptor into the form below, provide a url to an archive or upload a full darwin core archive including data files for validation.

Please note that we limit the size of uploaded files to 20MB, so reduce your data files if necessary. We will happily pull bigger archives from a url provided. If you need an archive for testing you can [download a test archive](#) first.

Copy paste meta.xml

```
<?xml version='1.0' encoding='utf-8'?>
<archive xmlns="http://rs.tdwg.org/dwc/text/"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="http://rs.tdwg.org/dwc/text/ http://rs.tdwg.org/dwc/text/tdwg_dwc_text.xsd">

  <core encoding="UTF-8" fieldsTerminatedBy="\t" linesTerminatedBy="\n" fieldsEnclosedBy="'" ignoreHeaderLines="0"
rowType="http://rs.tdwg.org/dwc/terms/Taxon">
    <files>
      <location>taxa.txt</location>
    </files>
    <id index="0" />
    <field index="2" term="http://rs.tdwg.org/dwc/terms/scientificName"/>
    <field index="3" term="http://rs.tdwg.org/dwc/terms/taxonomicStatus"/>
    <field index="4" term="http://rs.tdwg.org/dwc/terms/acceptedTaxonID"/>
    <field index="5" term="http://rs.tdwg.org/dwc/terms/acceptedTaxon"/>
```

Validate full online archive

Upload local archive or meta.xml

No file chosen

<http://tools.gbif.org/dwca-validator/>





DwC Archive Validation Result

Archive Source: http://www.gbif.fr:8080/ipt/archive.do?r=baillon_collection

Date validated: Oct 10, 2012 4:46:39 PM

This report has been written to <http://tools.gbif.org/dwca-reports/284-7808017097696686224.html> which will be deleted after one month. Until then you can revisit the report at your convenience.

Descriptor meta.xml

Validating against the dwc text guidelines [xml schema](#)

Validation successful

Archive read successfully

Metadata

An archive should (not required) have a metadata file bundled that informs about the whole dataset. GBIF recommends a [subset of EML](#), but simple [Dublin Core](#) is also permitted.

Dataset metadata description read from file [emi.xml](#).

| | |
|-------------|--|
| Title | Baillon Collection |
| Description | The Musée George Sand et de la Vallée Noire holds an important collection of mounted birds specimens started by Jean François Emmanuel Baillon during the 18th century and completed by his son Louis Antoine François Baillon until he died in 1855. Baillon father and son were 2 naturalists from northern France and have been in touch and exchanged specimens with some of the most famous naturalists and explorers of their time (e.g. Buffon, Cuvier, Temminck, Prince of Wied, Vieillot, Bonelli, Bruch, Leach, Ruppell, Levaillant, Verreaux, Leschenault...). The Baillon Collection houses 2480 specimens representing 1318 species of birds, collected in all continents including Antarctica, with more than 60 pre-1800 specimens as |
| Subject | |
| HomeUrl | |
| LogoUrl | |
| Published | Jun 22, 2012 |

Mappings

Inspecting the individual archive files and comparing the mapped concepts to the extensions registered with GBIF. An archive may have additional terms mapped than the ones declared by an extension. But those additions will not be understood widely so be careful!

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

The entity is mapped to source file *occurrence.txt*.

Core Record ID mapped to column 0
<http://rs.tdwg.org/dwc/terms/basisOfRecord> mapped to constant value "preservedspecimen"
<http://rs.tdwg.org/dwc/terms/eventRemarks> mapped to column 15
<http://rs.tdwg.org/dwc/terms/scientificName> mapped to column 2
<http://rs.tdwg.org/dwc/terms/eventDate> mapped to column 13
<http://rs.tdwg.org/dwc/terms/catalogNumber> mapped to column 3
<http://rs.tdwg.org/dwc/terms/class> mapped to column 7
<http://rs.tdwg.org/dwc/terms/order> mapped to column 12
<http://rs.tdwg.org/dwc/terms/country> mapped to column 16
<http://rs.tdwg.org/dwc/terms/genus> mapped to column 6
<http://rs.tdwg.org/dwc/terms/family> mapped to column 1
<http://rs.tdwg.org/dwc/terms/phylum> mapped to column 9
<http://rs.tdwg.org/dwc/terms/collectionCode> mapped to column 11
<http://rs.tdwg.org/dwc/terms/kingdom> mapped to column 14
<http://rs.tdwg.org/dwc/terms/recordedBy> mapped to column 8
<http://rs.tdwg.org/dwc/terms/institutionCode> mapped to column 5
<http://rs.tdwg.org/dwc/terms/locality> mapped to column 17
<http://rs.tdwg.org/dwc/terms/stateProvince> mapped to column 10
<http://rs.tdwg.org/dwc/terms/sex> mapped to column 4

Archive Data Files

Inspecting the archive using the dwca-reader library. The archive contains a core and 0 extension(s).

occurrence.txt

The data file contains 2,522 rows with 18 columns.

All rows in the data file have the same number of columns.

All core record ids are unique.



Outils

Pour les données spatiales



GeoLocate

<http://www.museum.tulane.edu/geolocate/>

- Géoreferencement 1 à la fois ou **par lot**
- **Géoref par nom de localité ou par coordonnées**
- Plusieurs **fonds de carte**
- **Correction** (déplacement du marqueur)
- Versions **en ligne, native** (téléchargeable) et **collaborative**

Home | Standalone App | Java Client | Web Application | Collaborative Georeferencing | Developer Resources | Workshops | Support and Contacts

GEOLocate Collaborative Georeferencing Web Client (BETA)

1 possible location(s) found.

baseLayer

- ☐ Google Hybrid
- ☐ Google Streets
- ☐ Google Satellite
- ☐ Virtual Earth Hybrid
- ☐ Virtual Earth Roads
- ☐ Virtual Earth Aerial
- ☐ OpenLayers WMS
- ☐ Mapnik
- ☐ Tiles@Home

overlays

- ☒ Polygon Layer
- ☒ Results
- ☒ Most Accurate Result

Lon: -92.947125
Lat: 35.17854

Google
Imagery ©2011 DigitalGlobe, GeoEye, State of Arkansas, USDA, Farm Service Agency - Terms of Use

Correct locality records

Community: Beta test community

- ☒ RDS 2489 Galla Creek 1 mi. NW Pottsville, dirt road.; USA; Arkansas; Pope; 35.178049; -92.947125
- ☐ Similar Records(0)
- ☐ Identical Records(0)

Next Record(s)
Correct
Skip Selected

place marker
draw polygon

logged in as: cogeguest
Calculated Coordinates
Lat: 35.178049

InfoXY

<http://splink.cria.org.br/infoxy?criaLANG=en>

specieslink

data & tools

infoXY

português

This tool was developed by CRIA with the aim of helping biological collections in validating geographic data. By entering data on geographic coordinates the tool returns information about the point, such as the name of the country, state or administrative region, and the name of the municipality or district. If the point is in the sea, the tool will calculate the distance to the closest coast, indicating the name of the country. The database used is GADM-Global Administrative Areas, a product of the BioGeoMancer project.
infoXY version 2.0

id , longitude , latitude (decimal degree)

exemploID, -41.5081, -20.7102

output:

HTML

☐ see map

Search

Results

| id | longitude | latitude | country | admin1 | typeadmin1 | admin2 | typeadmin2 | admin3 | typeadmin3 | admin4 | typeadmin4 |
|-----------|-----------|----------|---------|----------------|------------|--------|--------------|----------------|------------|--------|------------|
| exemploID | -41.5081 | -20.7102 | Brazil | Espírito Santo | State | Alegre | Municipality | Santa Angelica | District | | |

spOutlier

<http://splink.cria.org.br/outlier?&setlang=en>

specieslink

data & tools

spOutlier

Require input (id, longitude, latitude) - Optional (altitude)

import excel file (spreadsheet example): No file chosen

| | | |
|-----|---------|---------|
| 1, | -63.25, | -4.9167 |
| 2, | -64.15, | -4.9667 |
| 3, | -64.1, | -5 |
| 4, | -61.2, | -4.923 |
| 5, | -61.5, | -4.86 |
| 6, | -62.5, | -4.55 |
| 7, | -63.89, | -4.25 |
| 8, | -63.35, | -4.167 |
| 9, | -64.45, | -4.667 |
| 10, | -64.51, | -5 |

test

☒ see map

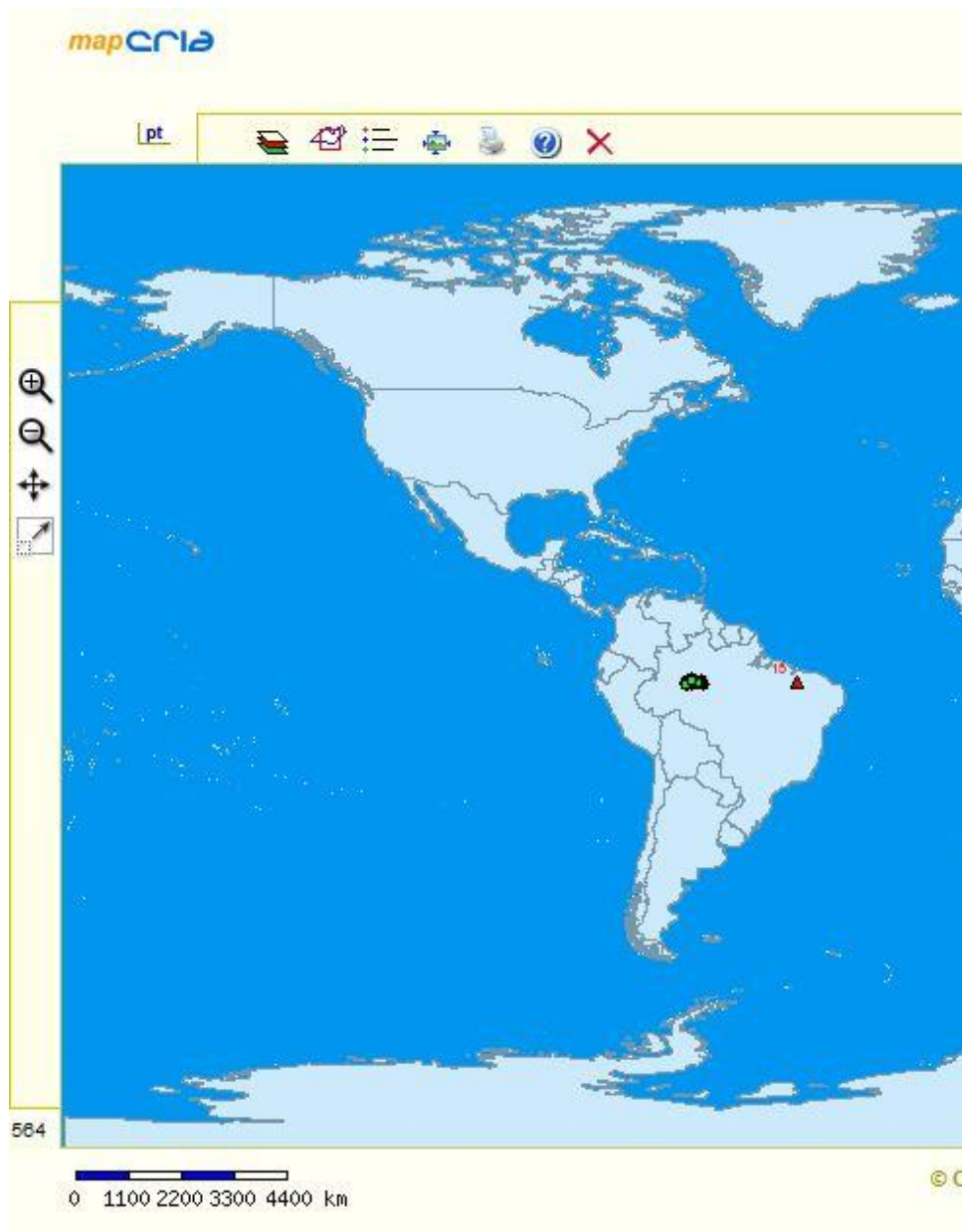
☒ show label for outliers on map

Result

15 -43.889 -4.586 Brazil

- Saisie lat / lon / altitude
- Données terrestres ou marines
- Carte en sortie
- Analyses statistiques





Georeferencing Calculator

Entrée:

Coordonnées,
offset, sources
d'erreurs

Sortie:

Coordonnées finales
Estimation de
l'erreur

English (local)

Georeferencing Calculator

Calculation Type

Coordinates and error - enter the Lat/Long for the named place or starting point

Locality Type

Distance along orthogonal directions (e.g., 2 mi E and 3 mi N of Bakersfield)

Step 3) Enter all of the parameters for the locality.

Coordinate Source

locality description

North or South Offset Distance

1

N

Coordinate System

degrees minutes seconds

East or West Offset Distance

3

W

Latitude

23

0

21

'

0

"

S

Extent of Named Place

0.5

Longitude

43

0

40

'

0

"

E

Measurement Error

0

Datum

datum not recorded

Distance Units

km

Coordinate Precision

nearest minute

Distance Precision

1 km

Decimal Latitude

-23.3409706

Decimal Longitude

43.6373286

Maximum Error Distance

6.048

km

Calculate

Promote

-23.3409706

43.6373286

6048

datum not recorded

degrees minutes seconds

Distance Converter:

0

km

=

0

km

Scale Converter:

0

mm

1:24000

=

0

km

Version 20110430en

copyright (c) 2001-2011 Regents of the University of California

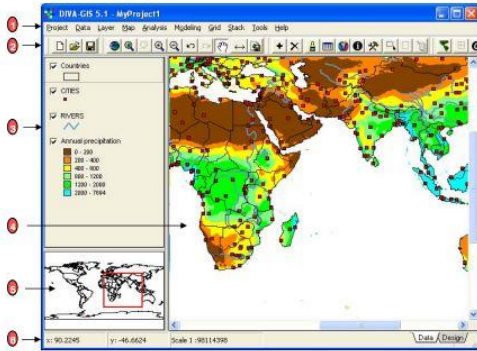
[Calculator Manual \(English\)](#)

[Georeferencing Guidelines \(English\)](#)

[Manual para el Uso de la Calculadora \(Español\)](#)



Logiciels SIG



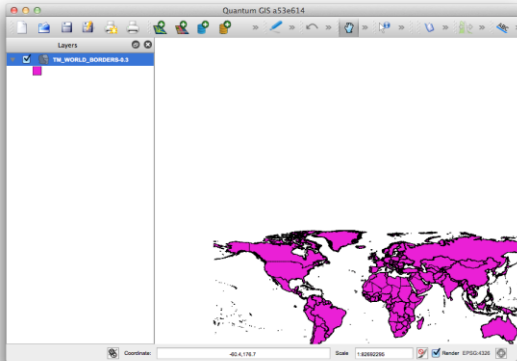
DIVA-GIS

Gratuit

Orienté biologie

Vectorielle/raster

Compatible avec de nombreux
format



Quantum GIS

SIG généraliste

Gratuit et Open-Source

Compatible avec de nombreux
formats / dbs / services

Mac/Linux/Windows

Vector/Raster

Extensible (plugins)



gvSIG

Open-Source, multiplateforme

Vector/Raster

Extensible (plugins)

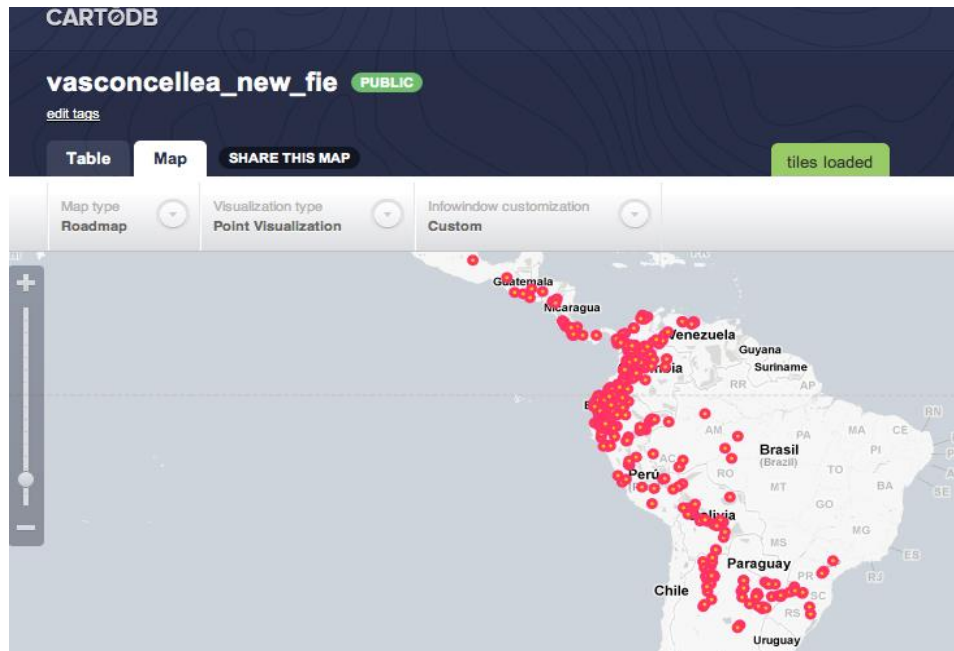
SIG on-line (II)

CartoDB

Importation efficace d'archives csv, shp, csv, xsl

Couches (et projections cartographiques) basées sur Google Maps

Edition des points (effacer ou déplacer) avec la possibilité d'exporter avec les modifications effectuées



SIG on-line (III)

Geonames

Enorme base de données (plus de 8 millions de noms géographiques) concernant la géolocalisation d'entités naturelles, culturelles, politiques, codes postaux...

Téléchargeable ou accessible via des services web

[GeoNames Home](#) | [Postal Codes](#) | [Download / Webservice](#) | [About](#)

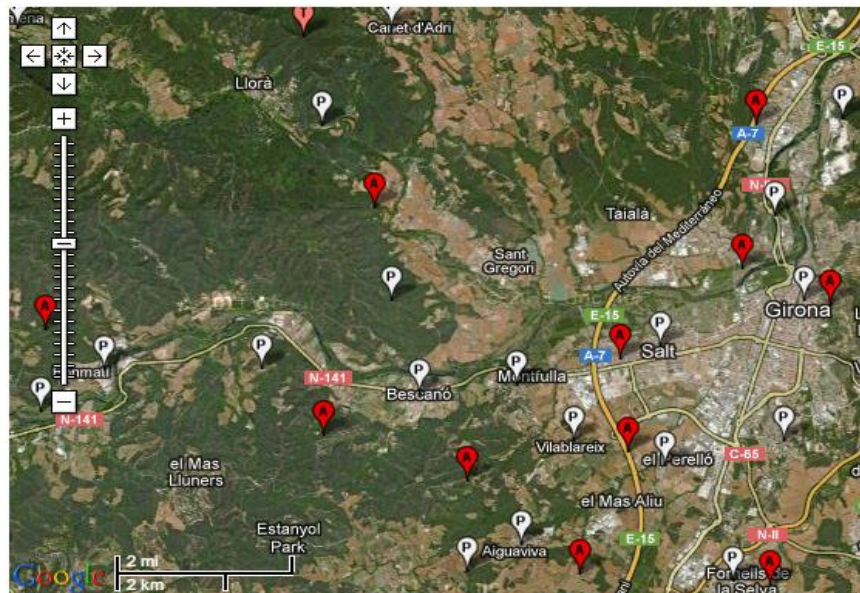
Map center : N 41° 58' 58" E 2° 49' 30"



[GeoNames Wikipedia](#)

features

- ☒ city, village,...
- ☒ mountain, hill, rock,...
- ☒ stream, lake, ...
- ☒ country, state, region,...
- ☒ parks, area, ...
- ☒ road, railroad
- ☒ spot, building, farm
- ☒ forest, heath, ...
- ☒ undersea



SIG on-line (IV)

Google Fusion Tables

- Cartes faciles à obtenir
- Format csv, txt, kml, Google spreadsheets
- Hébergement de données en ligne
- Geocode
- Fusion avec d'autres données
- Possibilité de faciliter la collaboration (édition avec multiples utilisateurs)

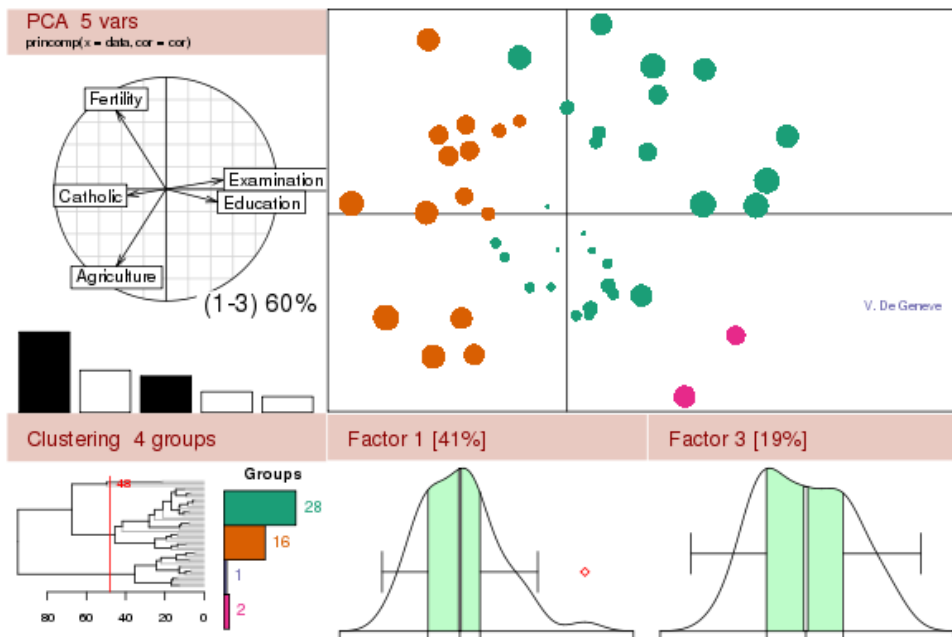


The R-project

<http://www.r-project.org>

Environnement et
langage de
programmation pour
l'analyse statistique

- Permet
l'**analyse** et la
visualisation.
- S'intègre avec
les SIG, les
langages de
programmation



Outils

Pour les données tabulaires



Name Parser



ECAT Name Parser

Name Parser

This is a simple html form to make use of the ECAT name parser. The parser takes a list of names and splits them into their components. It does only keep name parts required to reconstruct a full 3-part name. It ignores varieties and subspecies given for varieties. Please see our [API documentation](#) for details.

You can copy paste a list of names, one per row, or upload a text file with a name.

| | |
|--|------------------------------------|
| Names to parse: | Pseudocercospora |
| | Pseudocercospora Speg. |
| | Pseudocercospora Speg. 1910 |
| One per line or delimited by the pipe symbol " " | Pseudocercospora Spegazzini, 1910 |
| | Tridentella tangeroeae Bruce, 198? |
| | Ca Dyar 1914 |
| | Ea Distant 1911 |
| | Ge Nicéville 1895 |
| | Ia Thomas 1902 |
| | Io Lea 1831 |

Upload File: No file chosen

- Standardisation des champs
- “Découpage” des noms en 3 parties
- Ignore les variétés et autres subdivisions (en dessous de la sous-espèce)

Name Parser

Parsed Names

276 name parsed. 1 wellformed, 12 hybrid formulas and 14 doubtful names. See legend for [parsing types](#).

[Show](#) extended parsing

| Original | Genus | Infrageneric | Specific | Rank | Notho | InfraSpecific | Authorship | Year | (Authorship) | (Year) |
|---|-----------------|--------------|-------------|------|-------|---------------|-------------------------|------|--------------|--------|
| Asplenium Xinexpectatum (E.L. Braun 1940) Morton(1956) | Asplenium | | | | | | | | | |
| Aa Baker 1940 | Aa | | | | | | Baker | 1940 | | |
| Abacetulus laevicollis de Chaudoir, 1869 | Abacetulus | | laevicollis | | | | de Chaudoir | 1869 | | |
| Abelia 'Edward Goucher' | Abelia | | | cv. | | | | | | |
| Acanthobasidium delicatum (Wakef.) Oberw. ex Jülich 1979 | Acanthobasidium | | delicatum | | | | Oberw. ex Jülich | 1979 | Wakef. | |
| Acarospora cratericola 1929 | Acarospora | | cratericola | | | | | 1929 | | |
| Acripeza Guérin- Ménéville, 1838 | Acripeza | | | | | | Guérin- Ménéville | 1838 | | |
| Actinia stellula Hemprich and Ehrenberg 1834 | Actinia | | stellula | | | | Hemprich & Ehrenberg | 1834 | | |

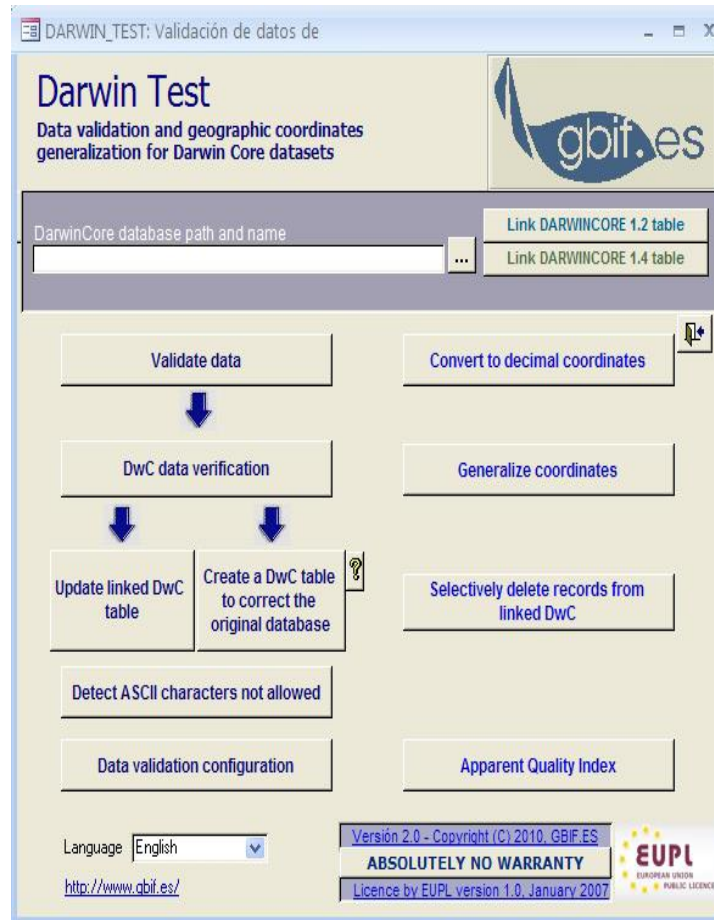


Darwin Test

DARWIN TEST est une application pour tester et valider les enregistrements de données au format Darwin Core 1.2 / 1.4 et Darwin Core Archive

- Chaque test peut être activé ou non
- Extensible (nouveaux tests)
- Conversions de coordonnées (UTM, decimal degrees, ...)
- Comparer les noms à des bases de données telles que Species2000
- Détection des erreurs d'encodage
- Généralisation des données géographiques (données sensibles).

Basé sur MS-Access
Open Source
Interface graphique



Open Refine

Démonstration



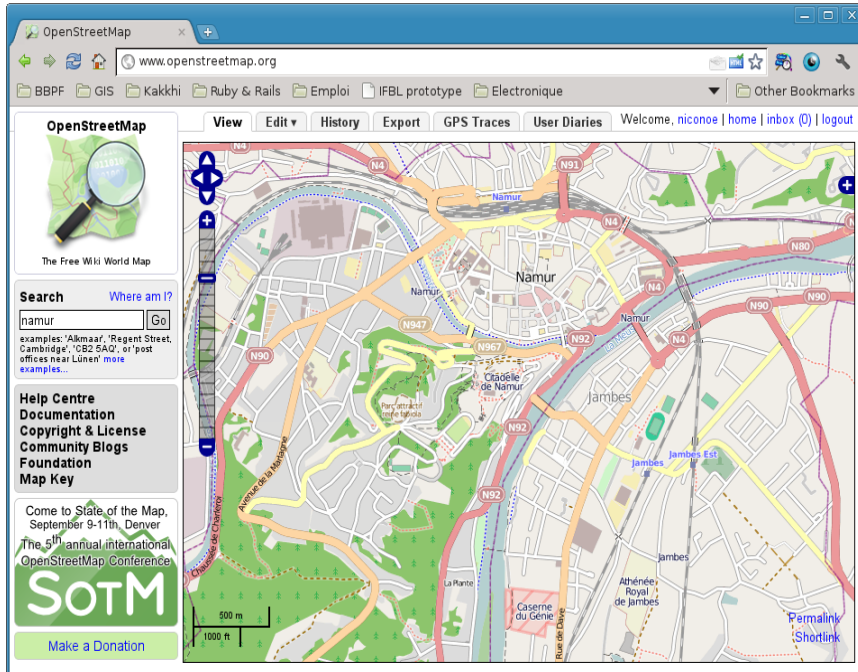
Autres ressources

Données



OpenStreetMap

ou "Wikipedia des cartes"



Source de données réutilisables:

- License libre
- Exportable dans différents formats

Négatif

- Qualité des données variable

<http://www.openstreetmap.org>

Thesauri

- Checklists thématiques:
 - Poissons : [Fishbase](#)
 - Animaux: [Index to Organism Name \(ION\)](#)
 - Mammifères: [Mammal Species of the World \(MSW\)](#)
 - Bactéries: [List of Bacteria with Standing in Nomenclature \(LBSN\)](#)
- Codes pays
 - ISO 3166-1 ou ISO3166-2, disponible par exemple au format Access
- ...

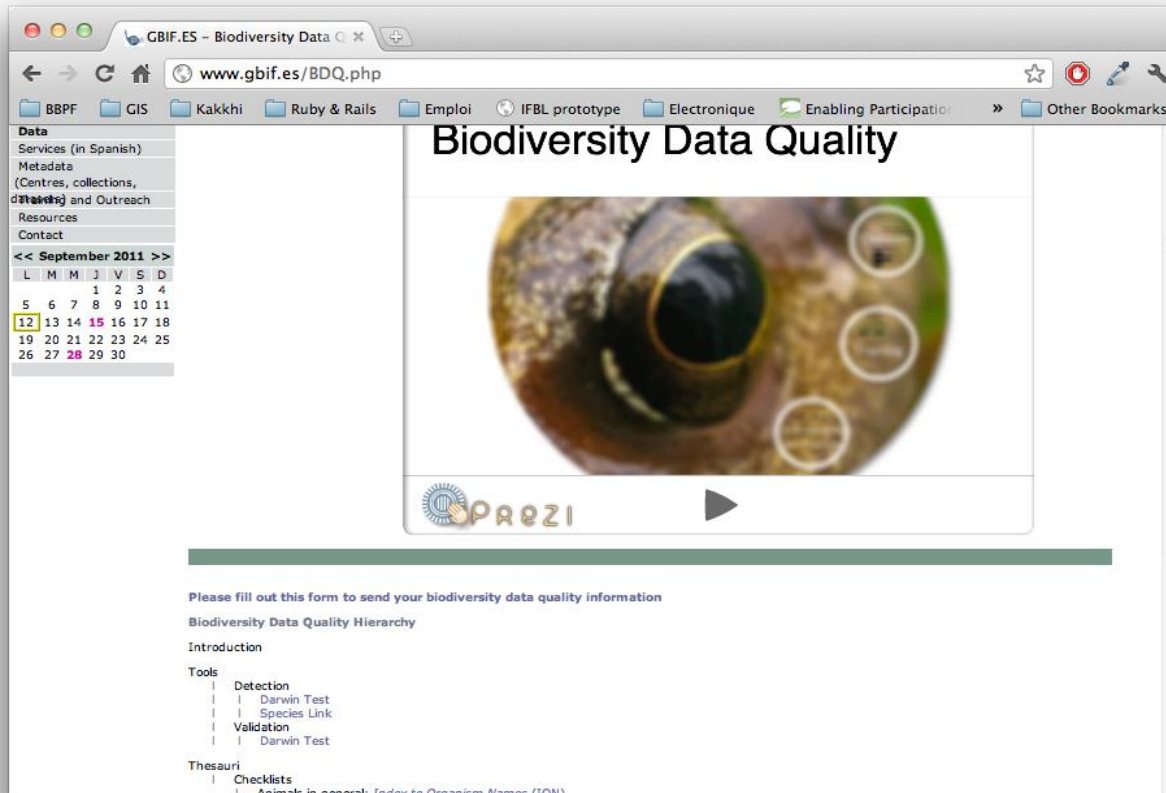


Autres ressources

Documents



Inventaire BDQ (GBIF Espagne)



<http://www.gbif.es/BDQ.php>

Centre de ressources GBIF

<http://www.gbif.org/resources>

Recherche et téléchargement

- Adéquation à l'usage
- Meilleures pratiques
- Manuels de formation

| | |
|---|----------------|
| Browse | |
| GBIF Welcome Box | (58 resources) |
| GBIF Start Up Kit | (30 resources) |
| GBIF Advanced Kit (under development) | (1 resources) |
| Training resources | (27 resources) |
| Biodiversity data digitisation and publishing | (16 resources) |
| Data capture | (5 resources) |
| Initiating a Collection Digitisation Project | |
| Significance of Organism Observations | |
| Digital Imaging of Biological Type Specimens. A Manual of Best Practice | |
| Terms Used in Bionomenclature. The naming of organisms (and plant communities) | |
| EDIT Biodiversity Service & Application Tracker | |
| Data management | (5 resources) |
| Principles of Data Quality | |
| Principles and Methods of Data Cleaning - Primary Species and Species-Occurrence Data | |
| Biogeomancer, Guide to Best Practices in Georeferencing | |
| Guide to Best Practices for Generalising Sensitive Species Occurrence Data | |
| Geolocate website | |
| Data publishing | (6 resources) |
| IPT Helpdesk Experts Workshop CD | |
| Getting Started, Overview of data publishing in the GBIF Network | |
| Publishing and Registering Data with GBIF | |
| Publishing Species Checklists, Best Practices | |
| GBIF Metadata Profile, How-to guide | |
| GBIF IPT v. 2 User Manual | |
| Biodiversity Data retrieval and use | (3 resources) |
| GBIF Participant Node management | (2 resources) |



**Merci de votre
attention**

