Formation GBIF sur la qualité, la publication et l'utilisation des données sur la biodiversité – Antananarivo, 04 – 05 avril 2016

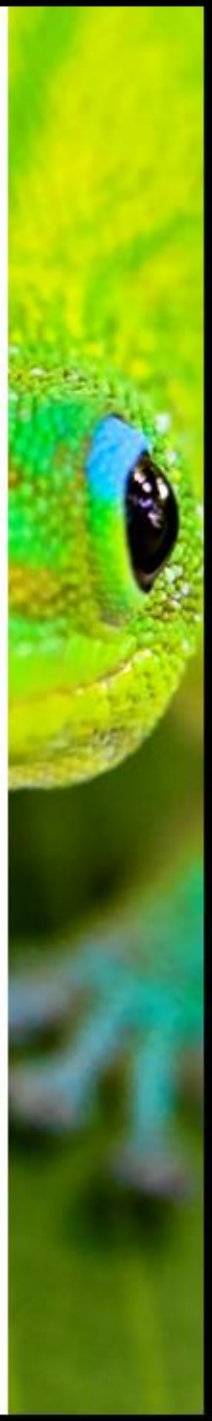
Outils et ressources pour évaluer et améliorer l'aptitude des données à être utilisées

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

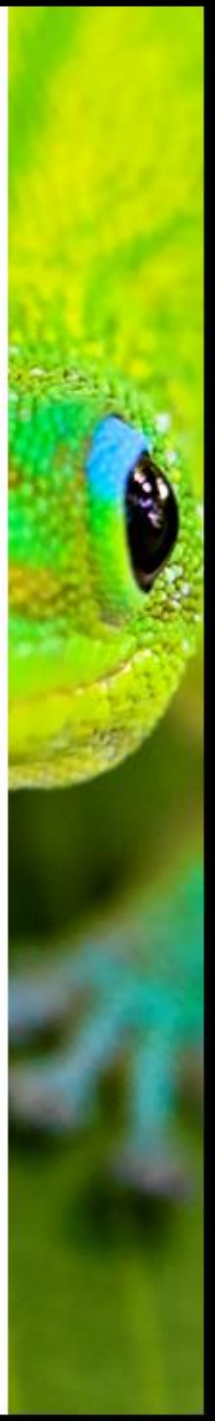
Aperçu

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



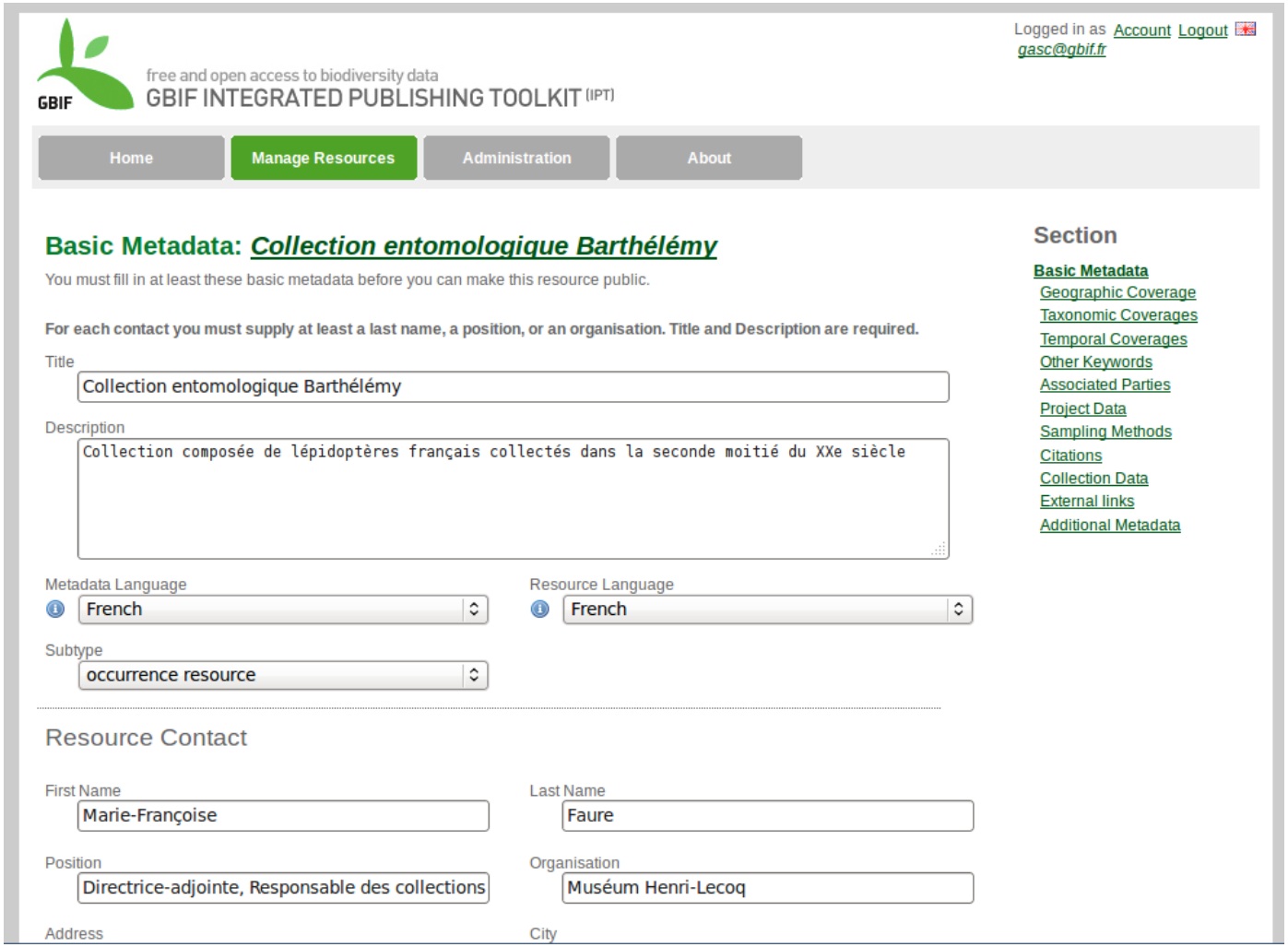
Outils

Pour les métadonnées

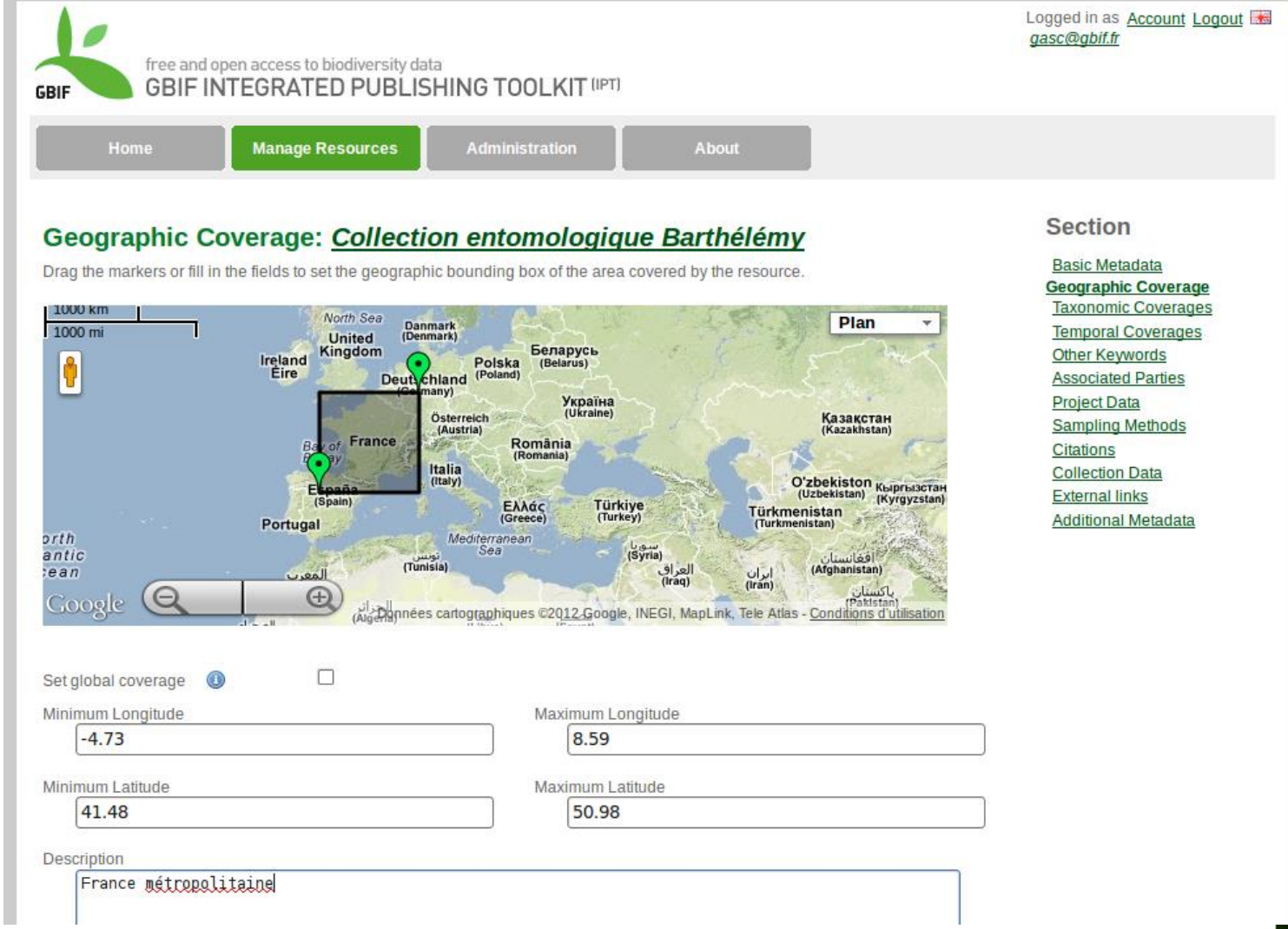


Métadonnées et IPT

(Dwc-A)

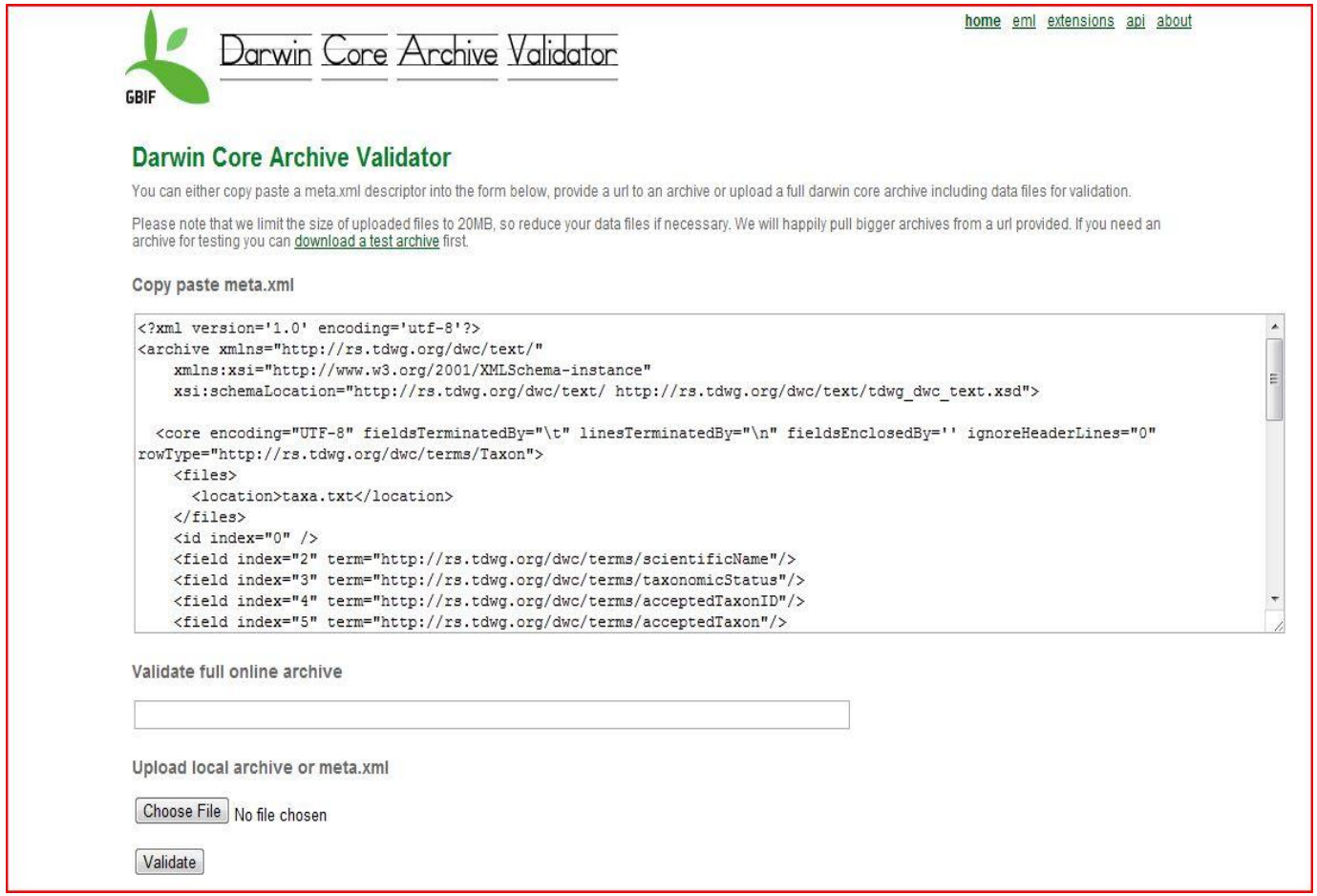


http://code.google.com/p/gbif-providertoolkit/



Couverture géographique

Darwin Core Archive Validator



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

Darwin Core Archive Validator

GBIF

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 *eml.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

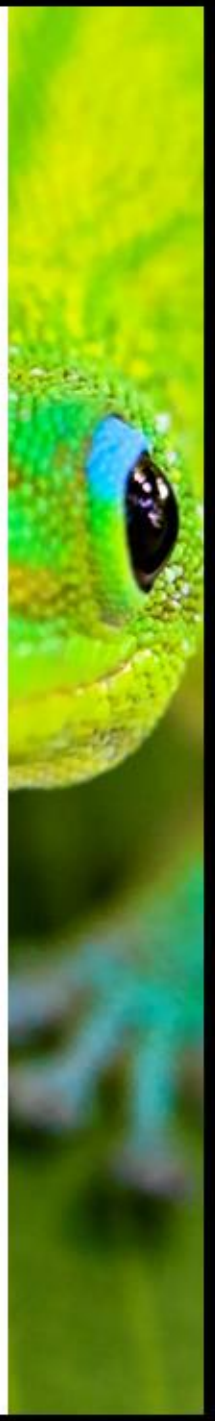
Inspecting the archive using the dwca-reader library. The archive contains a core and 0 extension(s).

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.

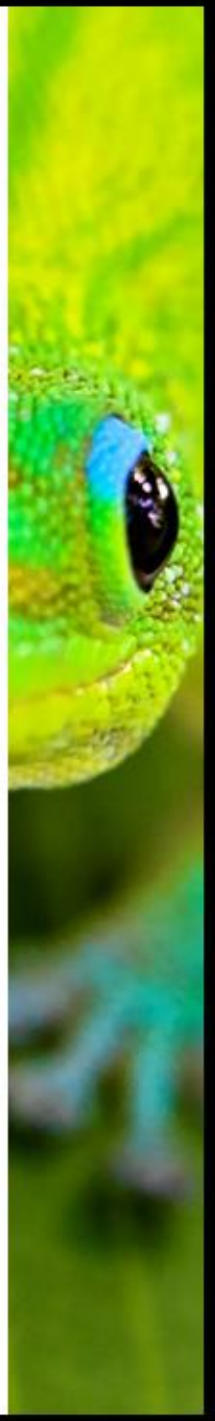


GBIF www.gbif.org

gbif France

Outils

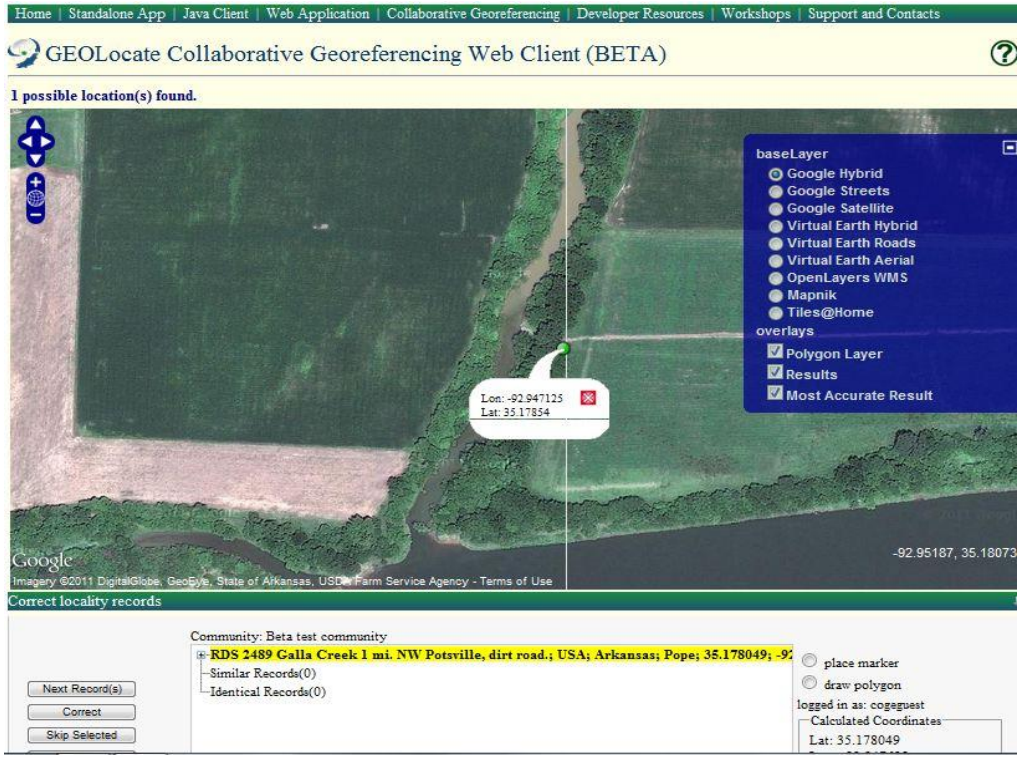
Pour les données spatiales



GeoLocate

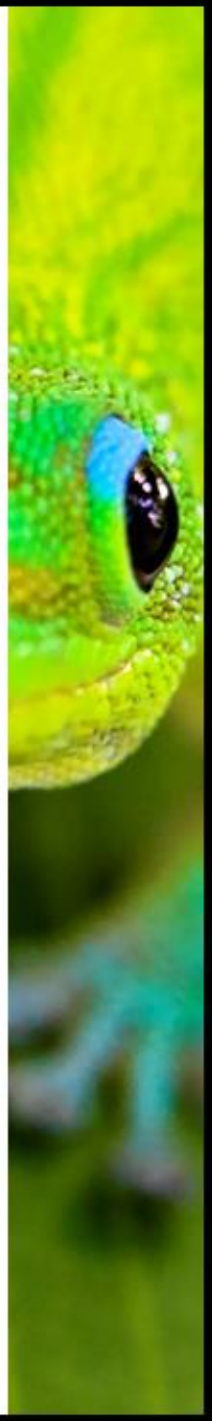
http://www.museum.tulane.edu/geolocate/

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



InfoXY

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



speciesLink

português

data & tools

infoXY

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

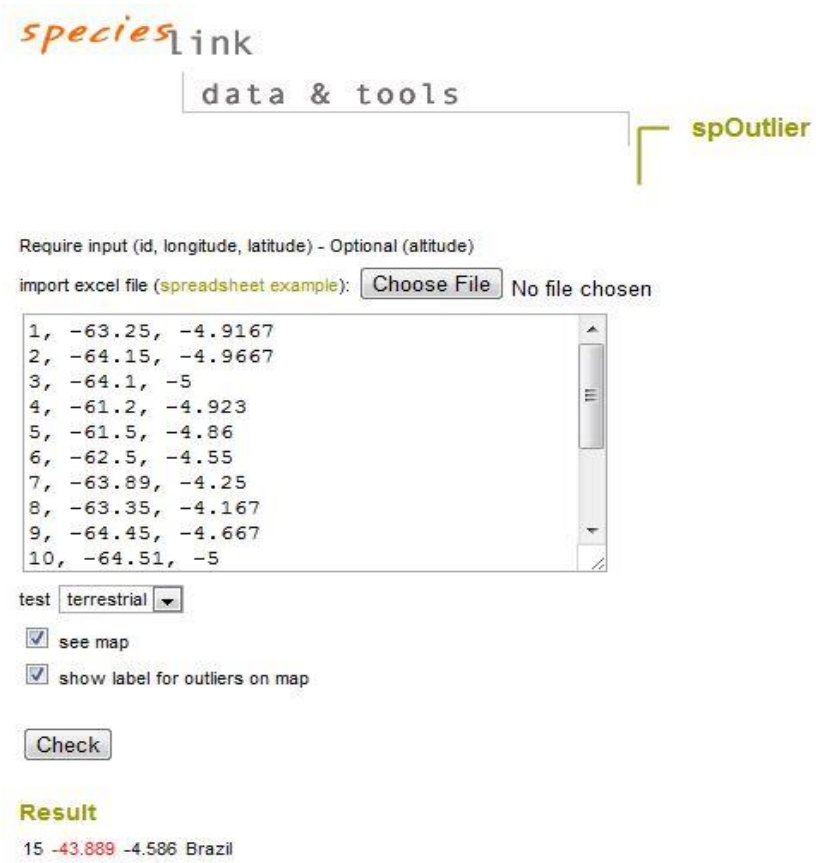
Results

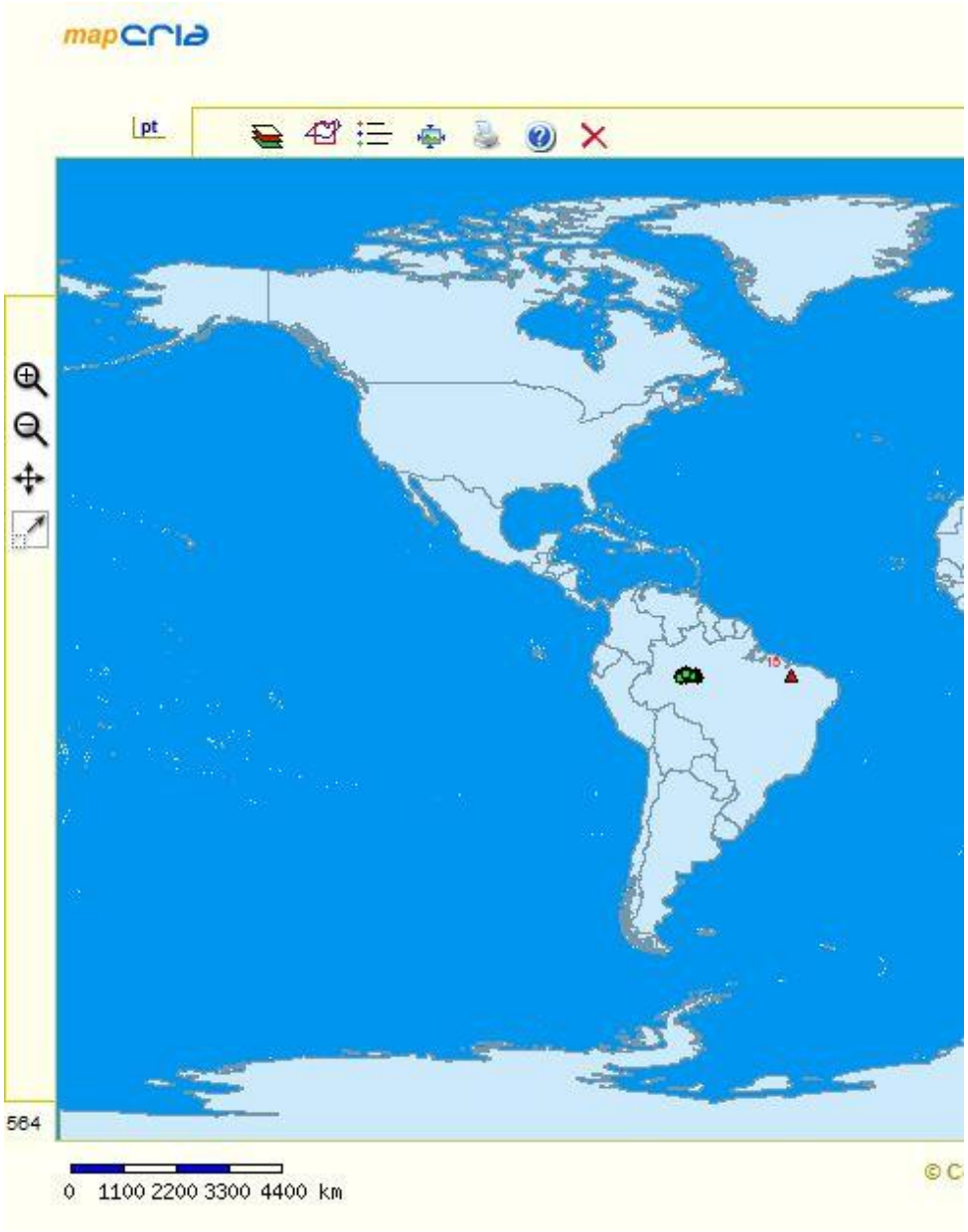
id	longitude	latitude	country	admin1	typeadmin1	admin2	typeadmin2	admin3	typeadmin3	admin4	typeadmin4
exemploID	-41.5081	-20.7102	Brazil	Espírito Santo	State	Alegre	Municipality	Santa Angelica	District		

spOutlier

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

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





GBIF
www.gbif.org

gbif
France

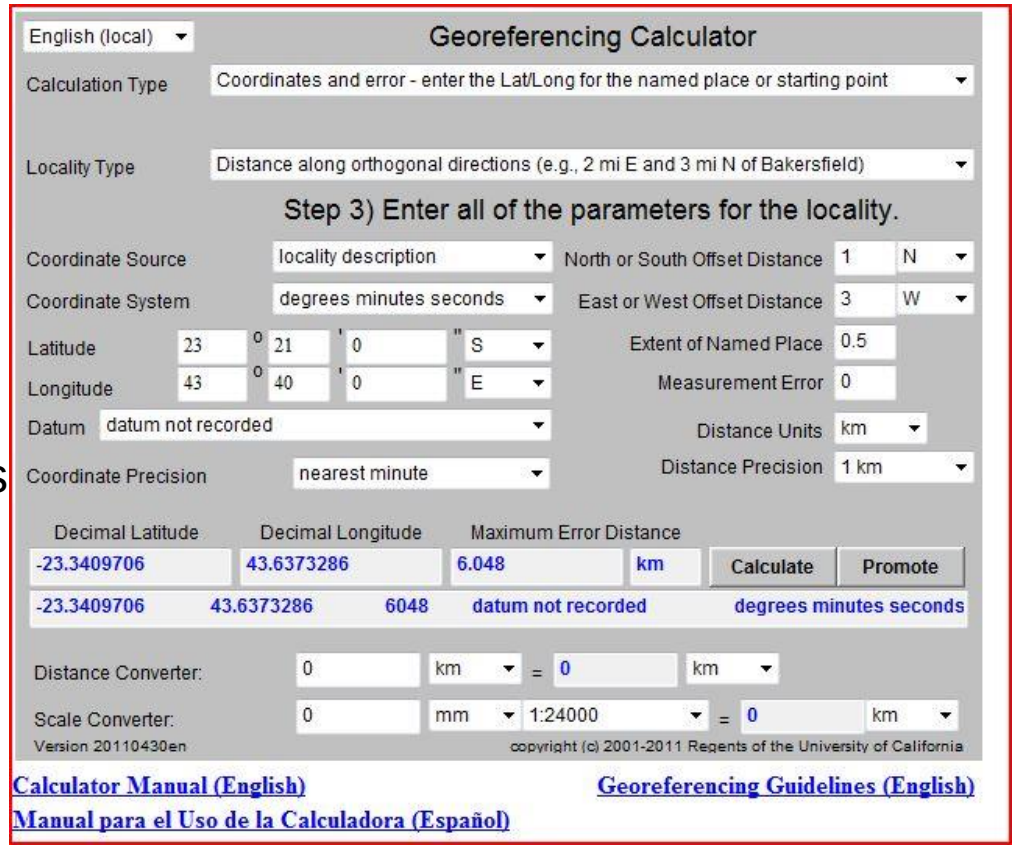
Georeferencing Calculator

Entrée:

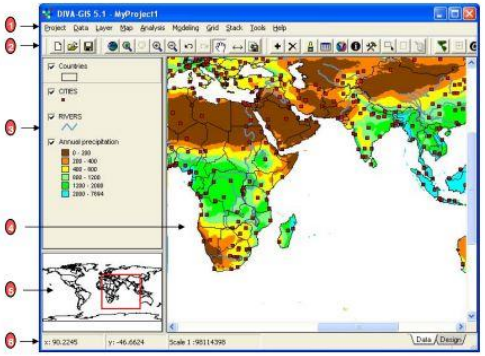
Coordonnées, offset, sources d'erreurs

Sortie:

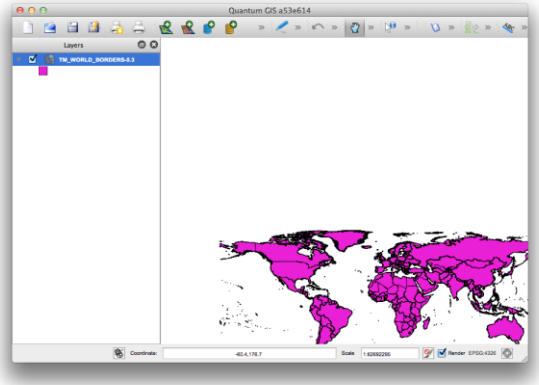
Coordonnées finales
Estimation de l'erreur



Logiciels SIG



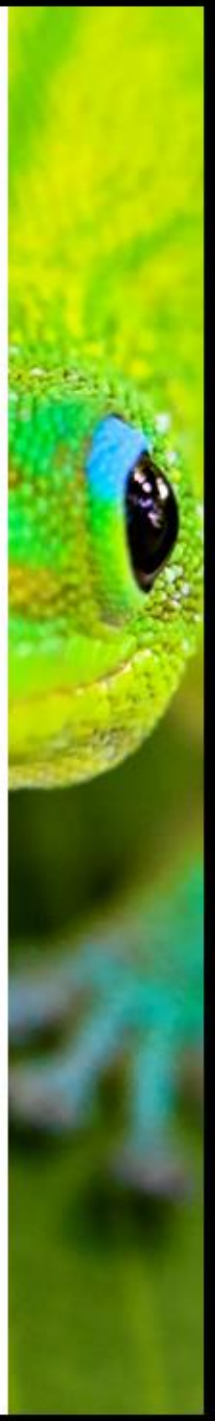
DIVA-GIS
Gratuit
Orienté biologie
Vectorielle/raster
Compatible avec de nombreux format



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, multiplataforme
Vector/Raster
Extensible (plugins)



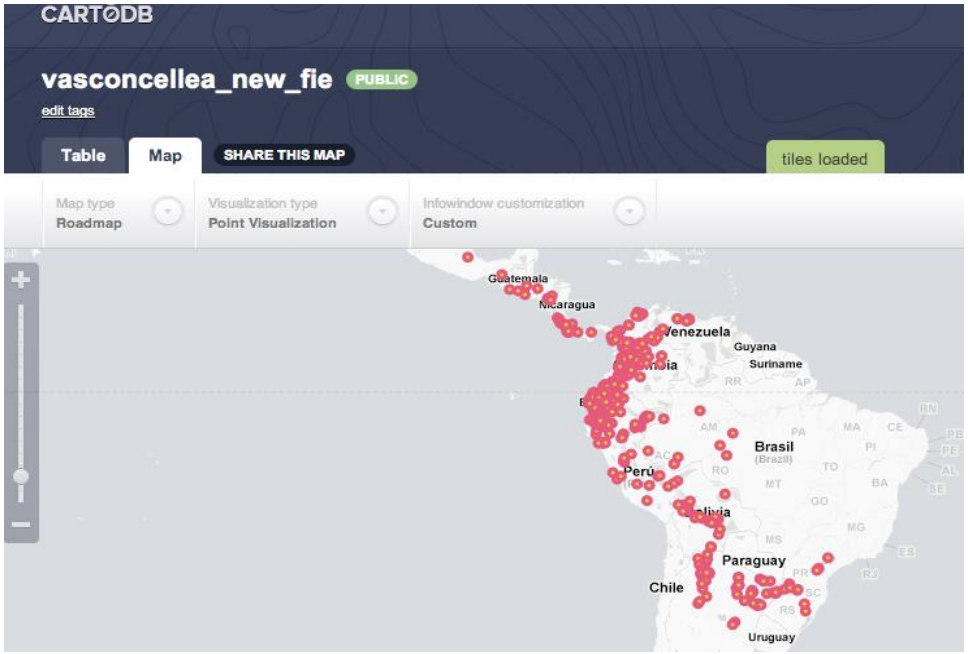
GBIF
www.gbif.org

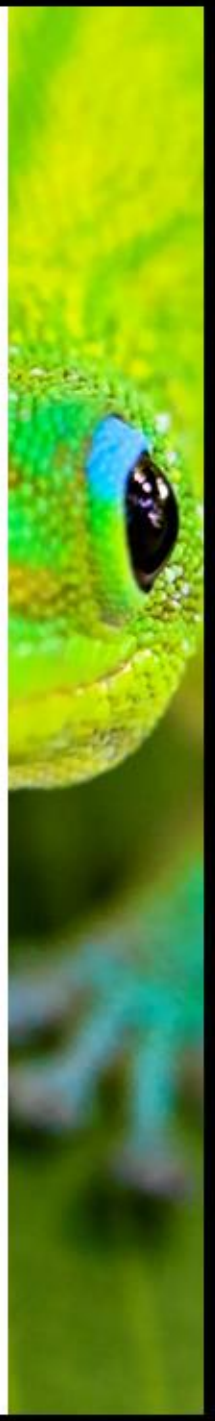
gbif France

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





GBIF
www.gbif.org

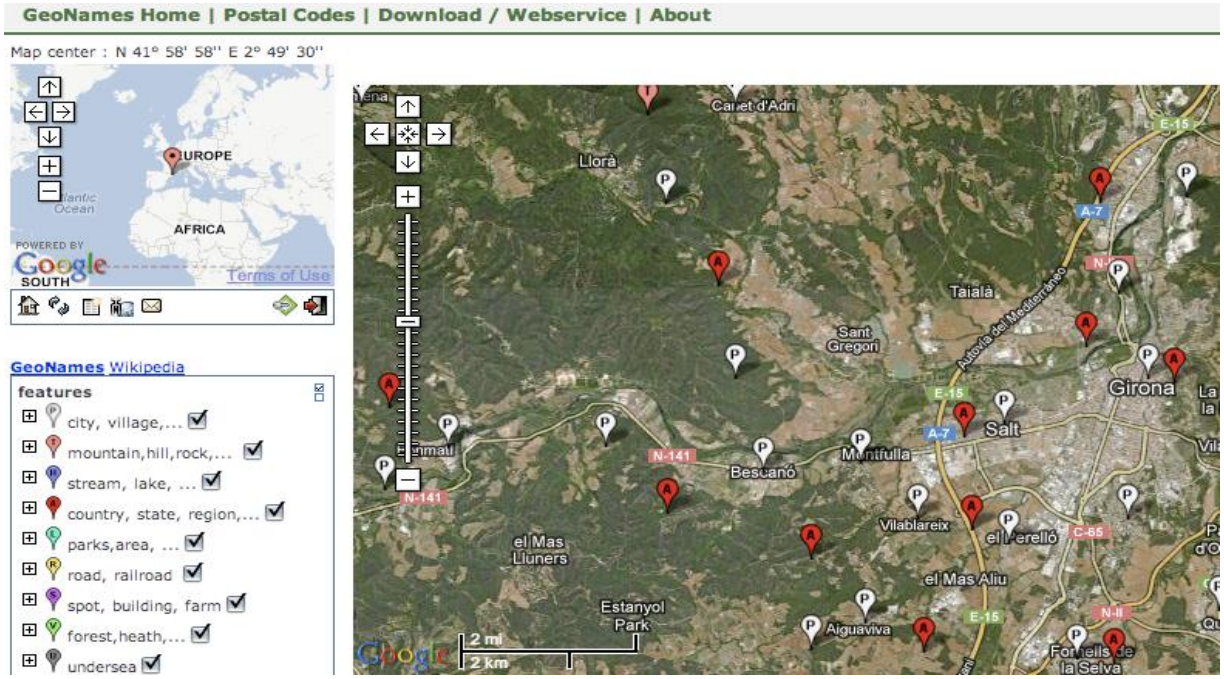
gbif France

SIG on-line (III)

Geonames

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

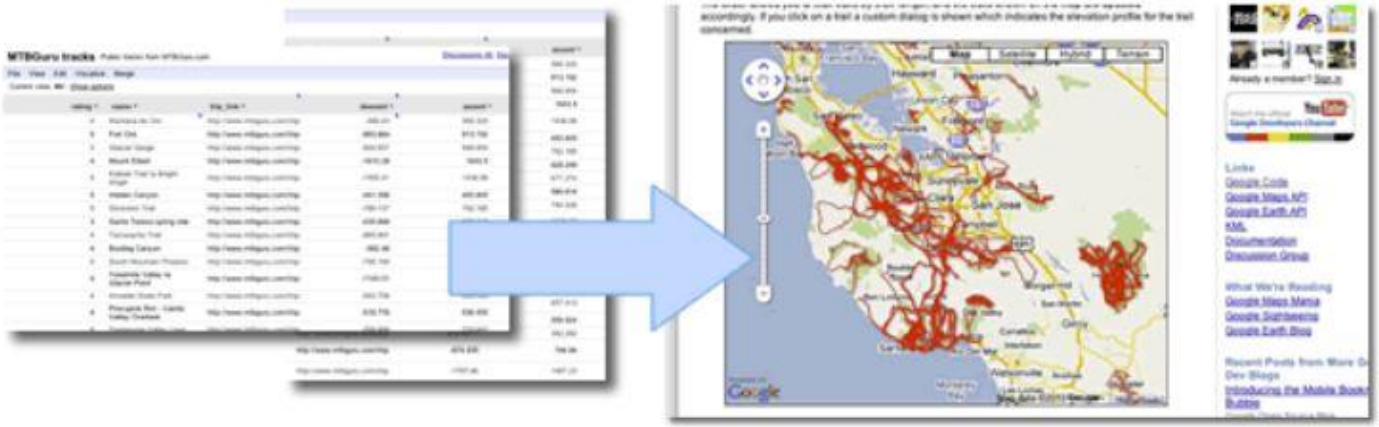
Téléchargeable ou accessible via des services web



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)

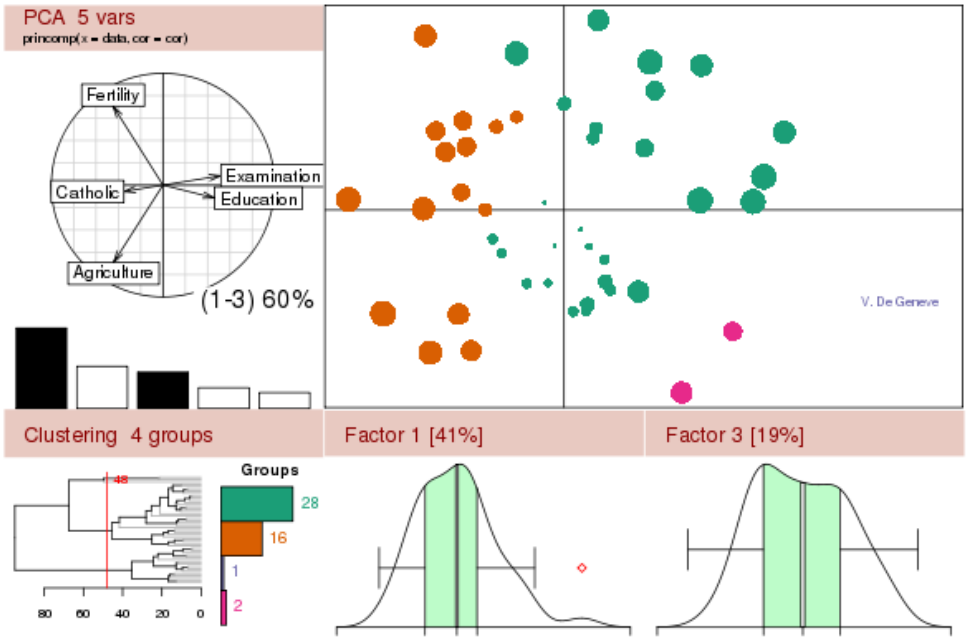


The R-project

http://www.r-project.org

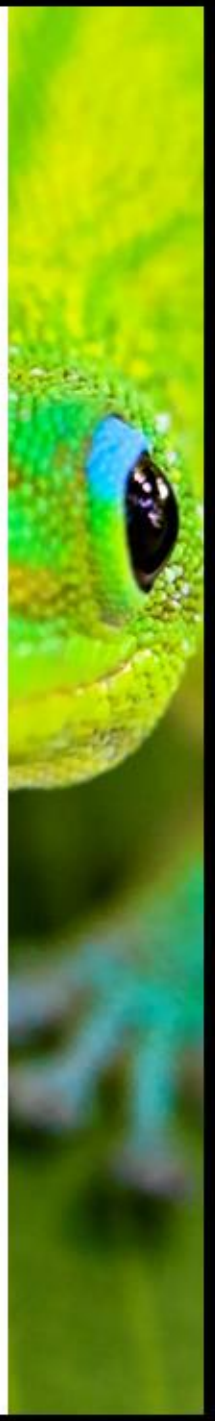
Environnement et language de programmation pour l'analyse statistique

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

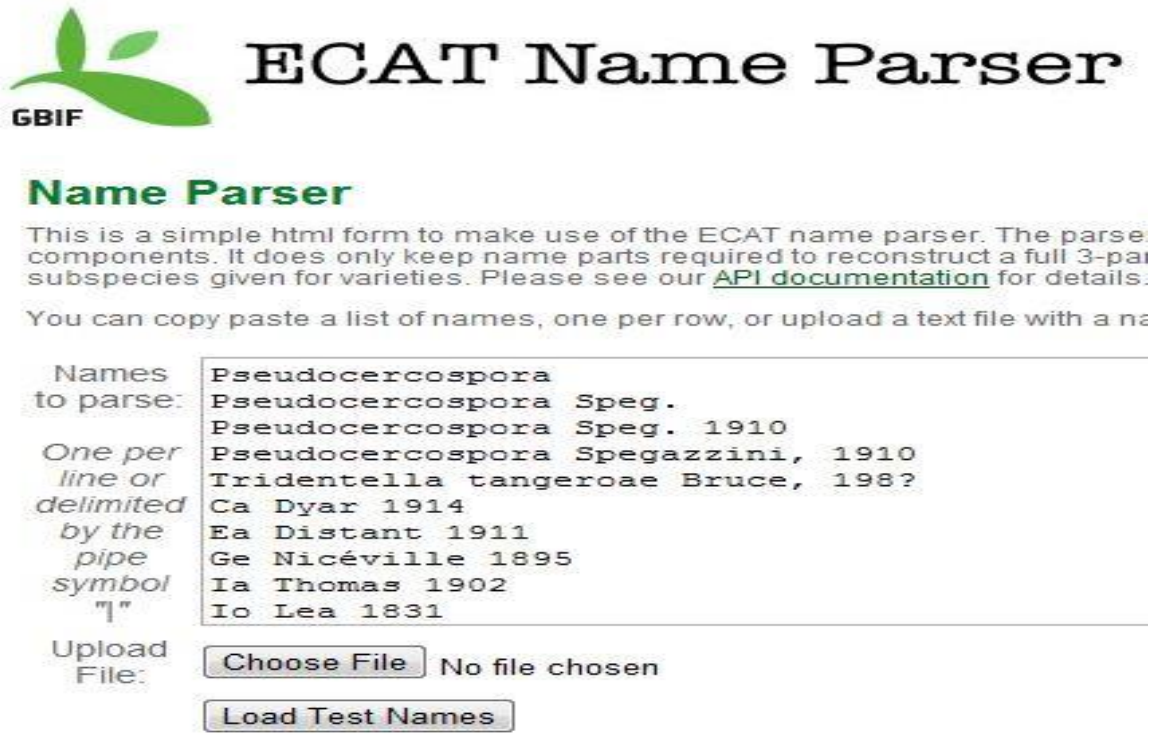


Outils

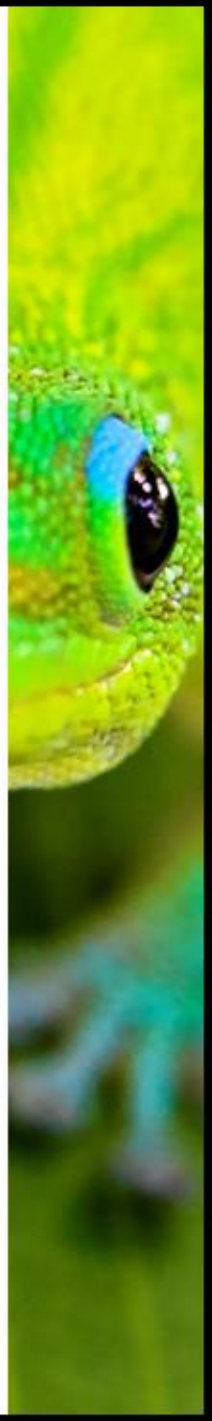
Pour les données tabulaires



Name Parser



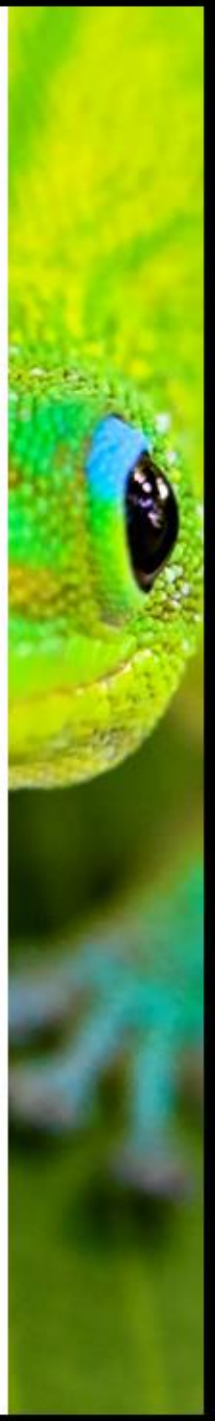
- Standardisation des champs
- “Découpage” des noms en 3 parties
- Ignore les variétés et autres subdivisions (en dessous de la sous-espèce)



GBIF
www.gbif.org

gbif France

Name Parser



Parsed Names

276 name parsed. 1 wellformed, 12 hybrid formulas and 14 doubtful names. See legend for parsing types.

Show extended parsing

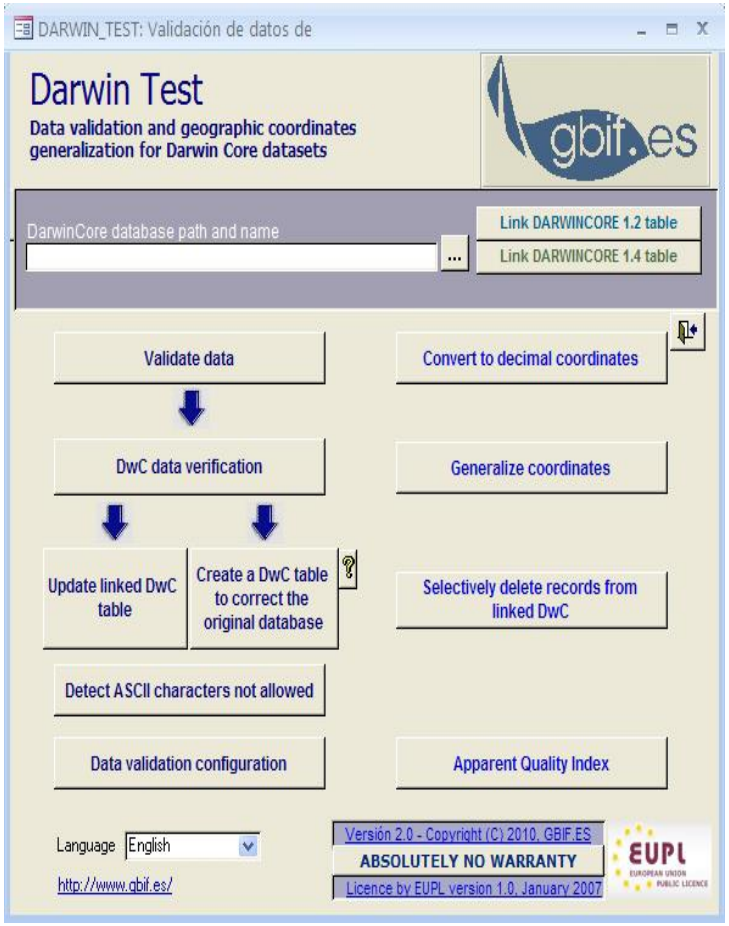
Original	Genus	Infrageneric	Specific	Rank	Notho	InfraSpecific	Authorship	Year	(Authorship)	(Year)
Asplenium Xinexpectatum (E.L. Braun 1940) Morton(1956)	Asplenium									
Aa Baker 1940	Aa						Baker	1940		
Abacetus laevicollis de Chaudoir, 1869	Abacetus		laevicollis				de Chaudoir	1869		
Abelia 'Edward 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-Méneville, 1838	Acripeza						Guérin-Méneville	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

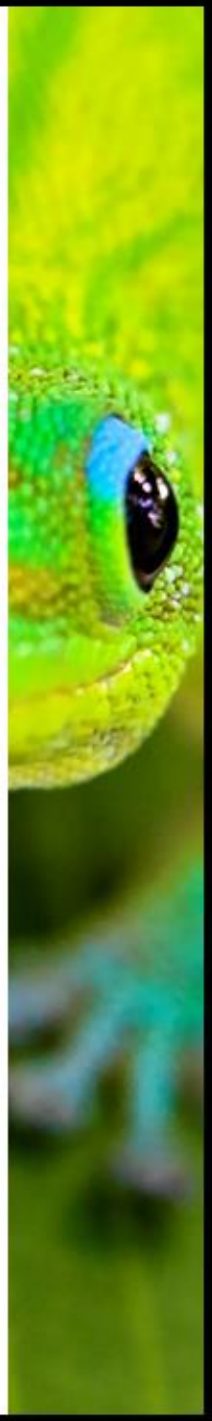
- Chaque test peut être activé ou non
- Extensible (nouveaux tests)
- Conversions de coordonnées (UTM, decimal degrees, ...)
- Comparer les noms à des bases de données telles que Species2000
- Détection des erreurs d'encodage
- Généralisation des données géographiques (données sensibles).

Basé sur MS-Access
Open Source
Interface graphique



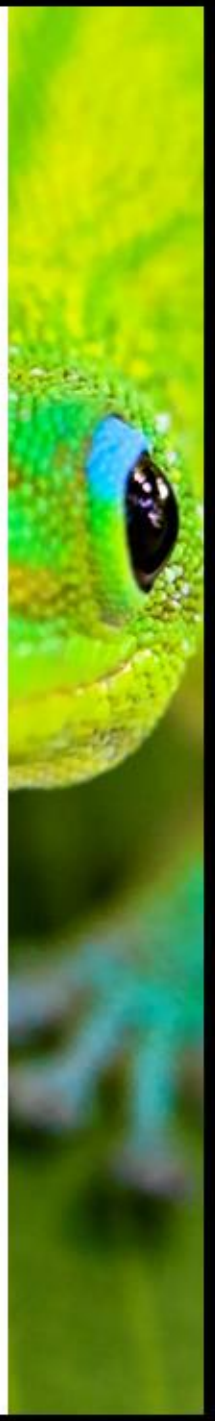
Open Refine

Démonstration



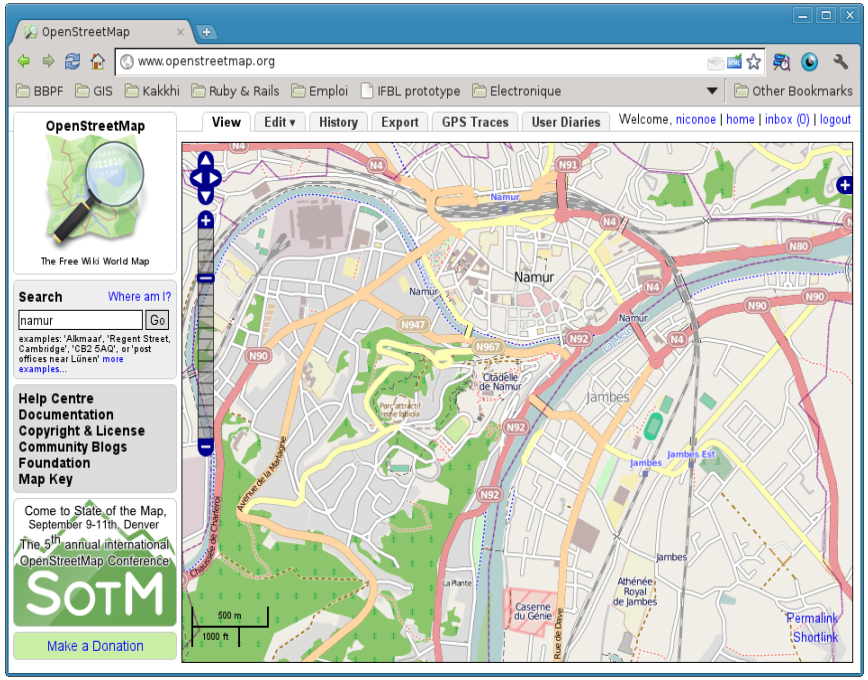
Autres ressources

Données



OpenStreetMap

ou "Wikipedia des cartes"



Source de données réutilisables:

- License libre
- Exportable dans différents formats

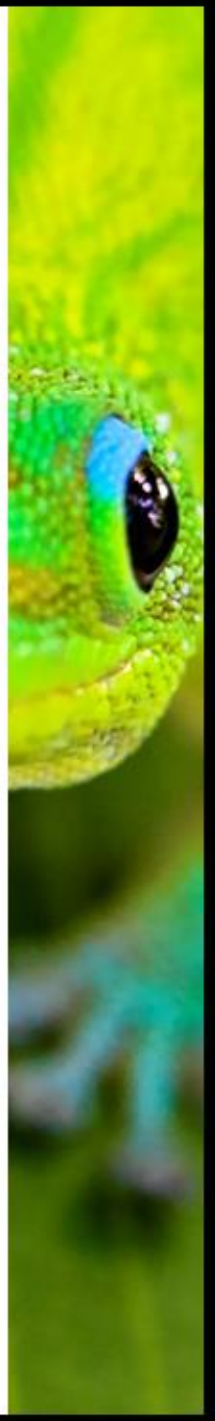
Négatif

- Qualité des données variable

http://www.openstreetmap.org

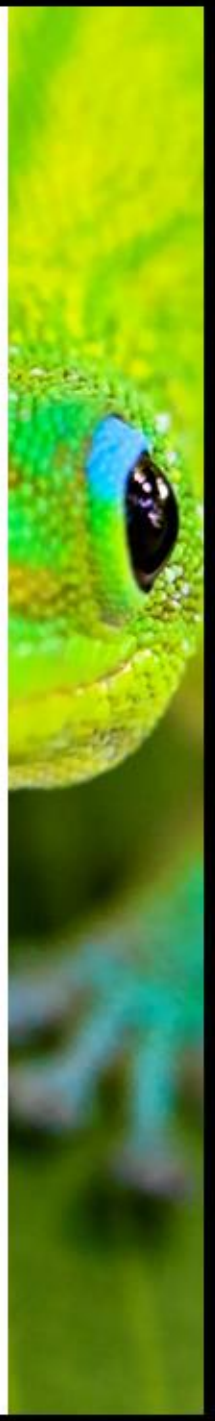
Thesauri

- 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
- ...

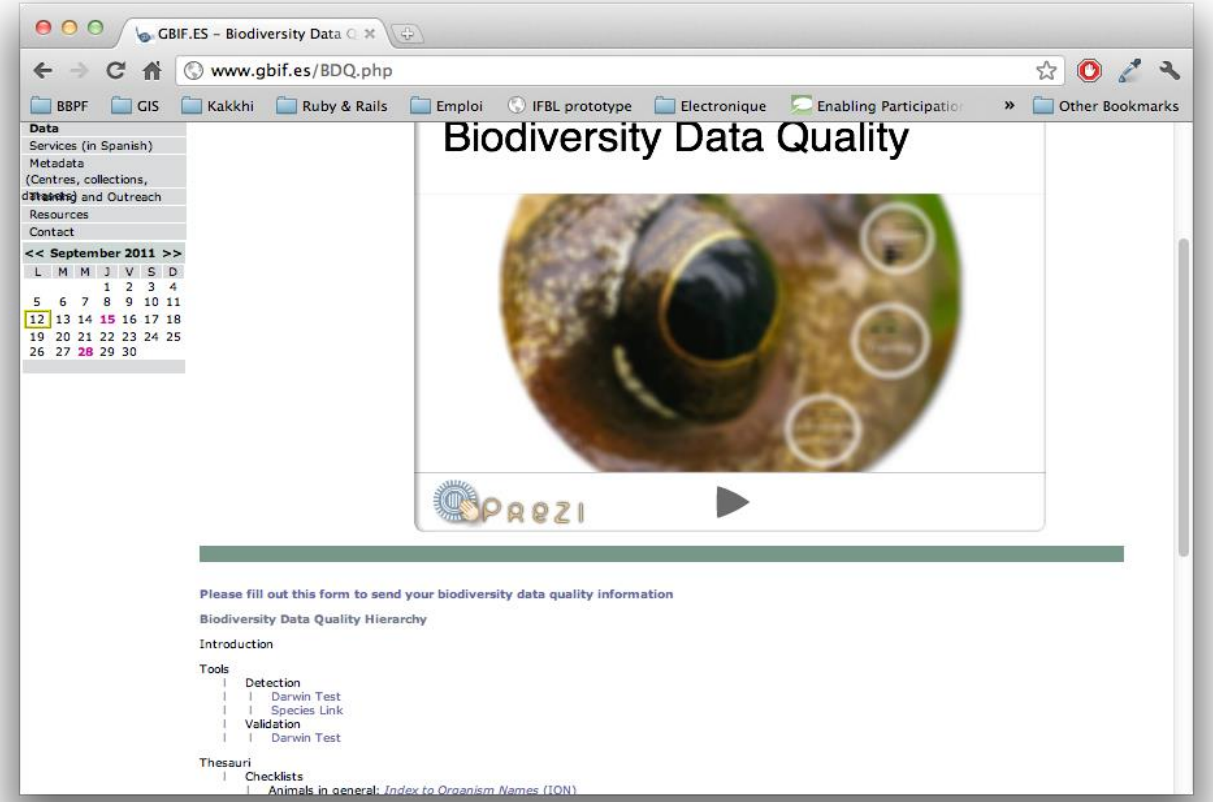


Autres ressources

Documents



Inventaire BDQ (GBIF Espagne)



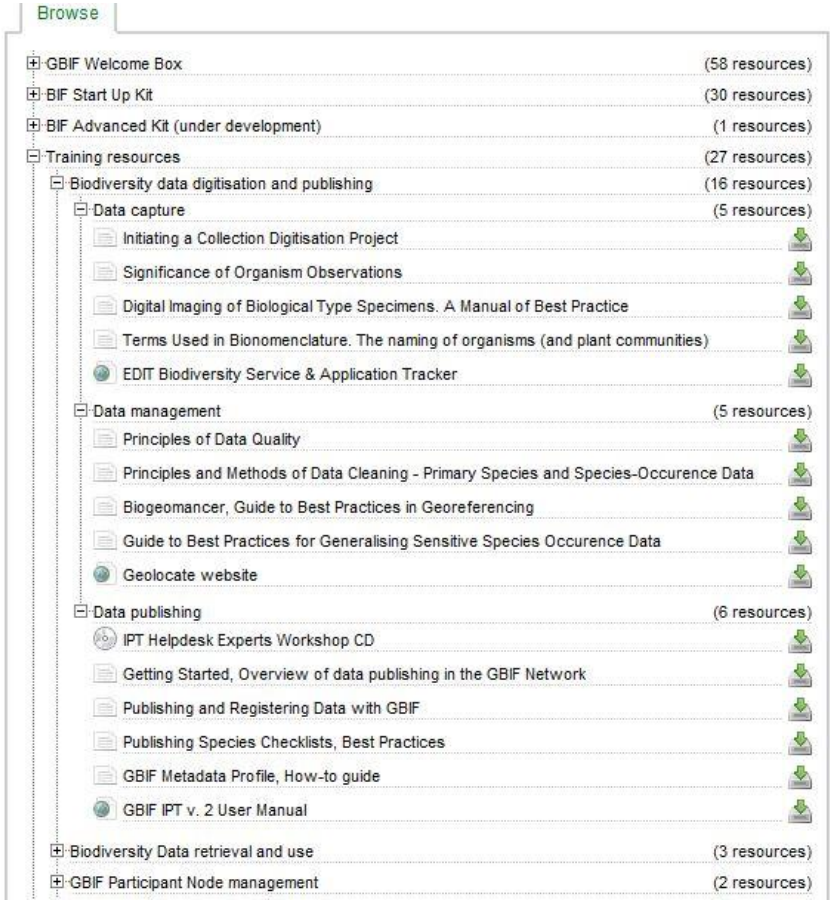
http://www.gbif.es/BDQ.php

Centre de resources GBIF

http://www.gbif.org/resources

Recherche et téléchargement

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



Merci de votre attention