



## DATA PAPER - UNE INCITATION À LA QUALIFICATION ET À LA RÉUTILISATION DES JEUX DE DONNÉES

Webinaire Groupe de travail inter-reseaux - Atelier données, 05 novembre 2020

### Exemple d'intégration du data paper à un workflow de publication de jeux données

*L'outil intégré de publication (IPT) du GBIF et l'ARPHA Writing Tool de Pensoft*

Sophie Pamerlon ([pamerlon@gbif.fr](mailto:pamerlon@gbif.fr) ou [gbif@gbif.fr](mailto:gbif@gbif.fr))



le Réseau Calcul



le Réseau Resinfo



La Direction de l'Information Scientifique et Technique du CNRS



le Réseau DEVlog



le Réseau SIST de l'INSU



le Réseau QeR



le Réseau Médici



le Réseau Renatis



le Réseau RBDD



Global Biodiversity Information Facility  
Système Mondial d'Information sur la Biodiversité

> Accès libre et gratuit aux données de la biodiversité



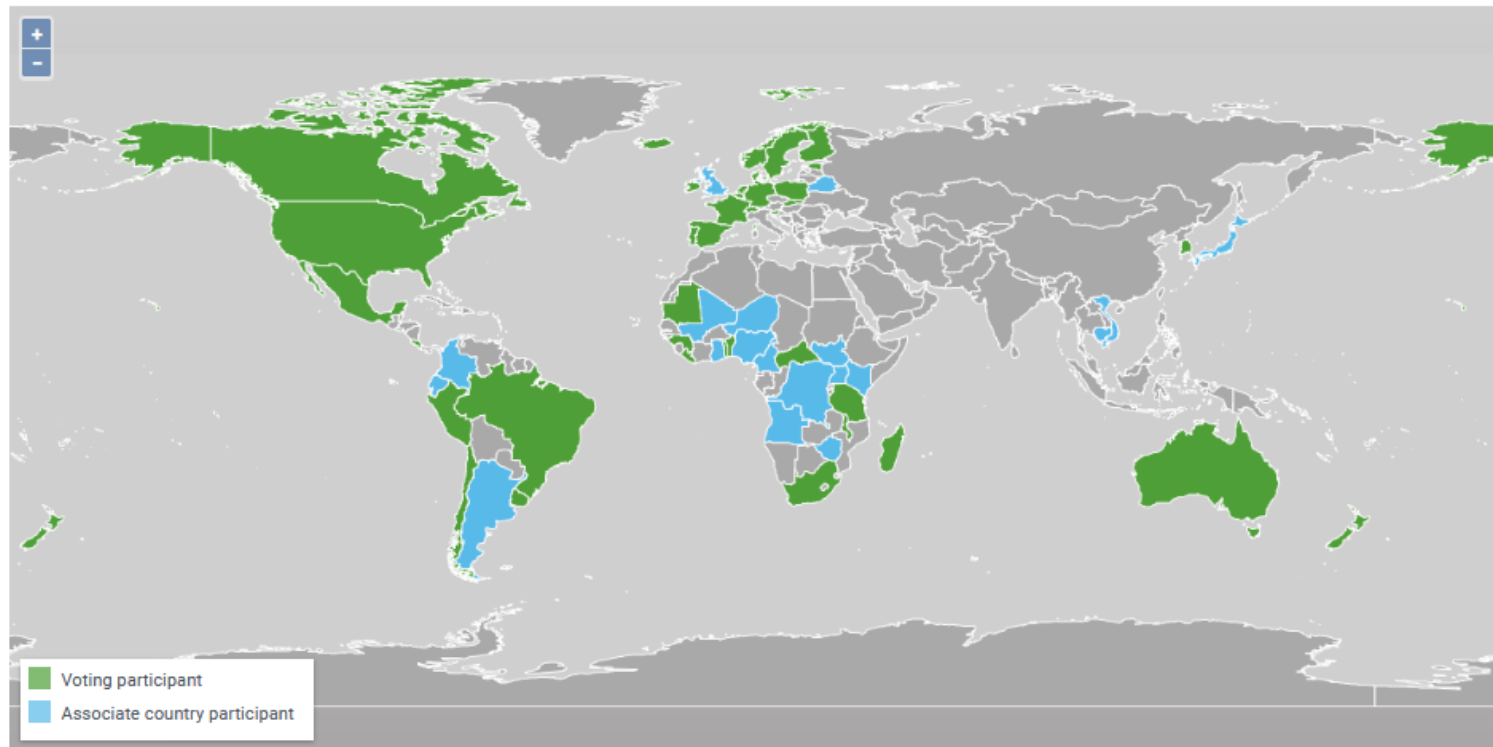
# Le réseau GBIF et ses outils

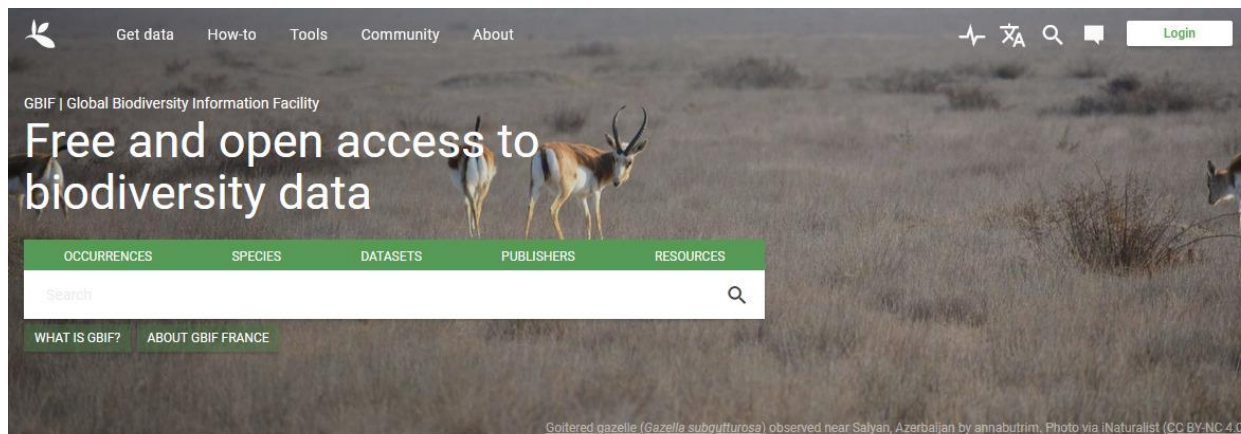


## Le réseau GBIF en 2020

- Programme intergouvernemental, créé en 2001, à l'initiative du comité scientifique de l'OCDE
- 60 pays et 38 organisations

GLOBAL AFRICA ASIA EUROPE AND CENTRAL ASIA LATIN AMERICA AND THE CARIBBEAN NORTH AMERICA OCEANIA PARTICIPANT ORGANISATIONS GBIF AFFILIATES





79,8 milliards de données téléchargées par mois (moyenne mensuelle 2018)

Occurrence records  
1 625 178 373

Datasets  
54 673

Publishing institutions  
1 665

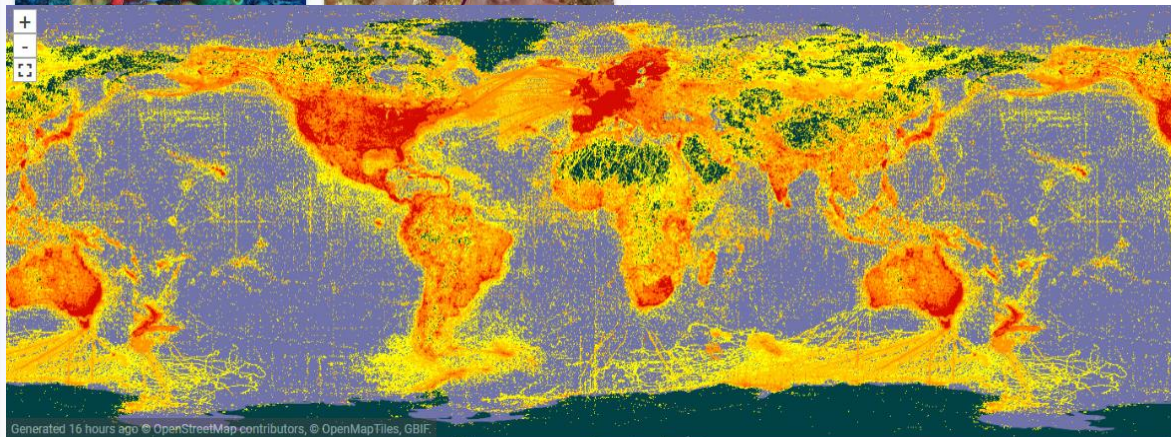
Peer-reviewed papers using data  
5 078



New task group to enhance GBIF-enabled research on species linked to human diseases



New video: an invitation to the private sector



ShinyBIOMOD wins 2020 GBIF Ebbe Nielsen Challenge

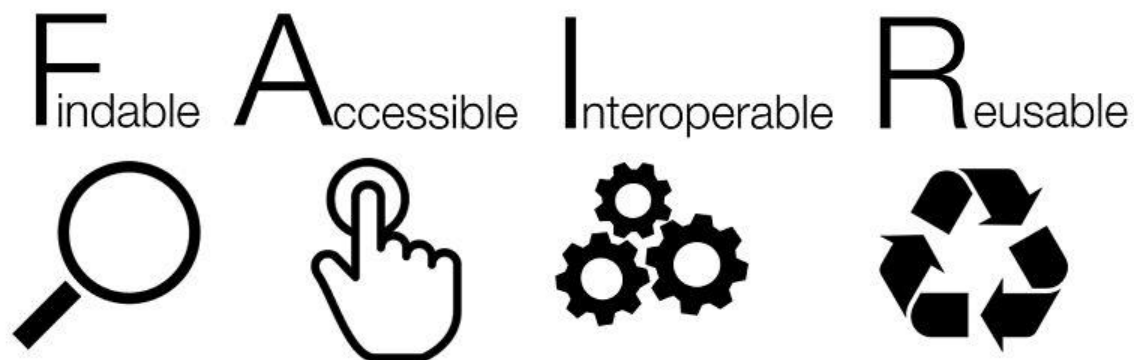


Newly described species: *Allenostreptus bicoloripes*

[www.gbif.org](http://www.gbif.org)



## Les principes FAIR







<https://www.go-fair.org/fair-principles>





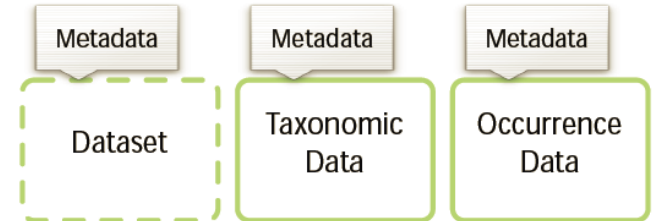
## Processus mis en place

<p><b>Attribution</b></p> 	<p>Digital Object Identifiers assignés pour chacun des :</p> <ul style="list-style-type: none"> <li>• Jeux de données publiés</li> <li>• Téléchargements par les utilisateurs</li> <li>• Data papers</li> </ul>
<p><b>Licences</b></p> 	<p>Les éditeurs choisissent une licence appropriée:</p> <ul style="list-style-type: none"> <li>• CC-0</li> <li>• CC-BY</li> <li>• CC-BY-NC</li> </ul>
<p><b>Standards</b></p> 	<p>Données: DarwinCore, ABCD, DublinCore</p> <p>Metadonnées: Ecological Metadata Language (EML)</p>
<p><b>Accessibilité</b></p> 	<p><a href="http://GBIF.org">GBIF.org</a> portail d'accès aux données</p> <ul style="list-style-type: none"> <li>• Requêtes interopérables</li> <li>• Cartes interactives</li> <li>• Téléchargements gratuits</li> <li>• web services</li> </ul>
<p><b>Publication</b></p>	<p>Integrated Publishing Toolkit (IPT), BioCASE, Spreadsheet templates</p>

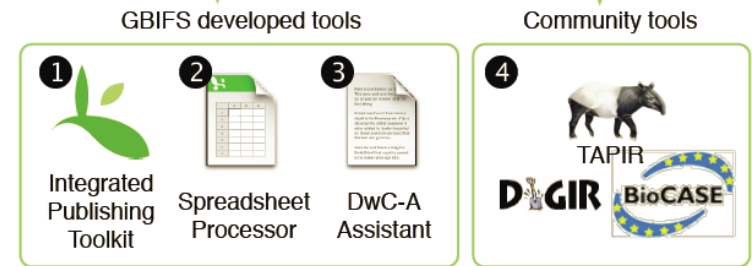


# GBIF.org : processus de publication des données et métadonnées associées

Which type of data?



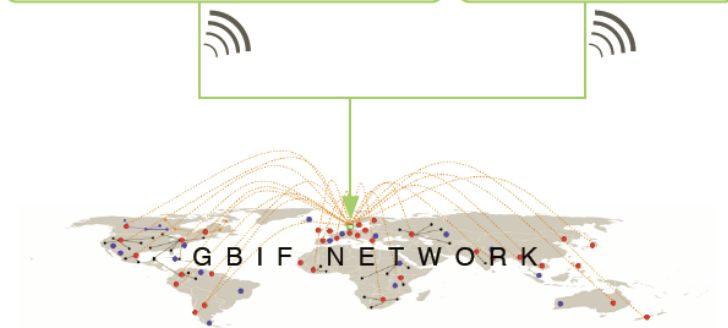
Select a tool



Publish your data

Register with GBIF

Discovery through the Portal





## Standards d'échange



Métadonnées



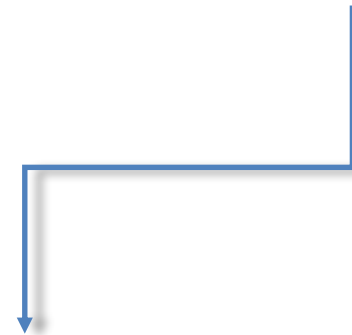
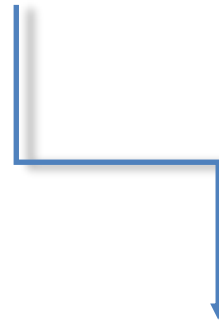
Données  
d'occurrences



Données  
taxonomiques



Données  
d'échantillonnage



### Ecological Metadata Language (EML)

- Description riche des jeux de données
- Profil de métadonnées du GBIF

### Darwin Core (TDWG)

- 180 concepts
- Ratifié en 2009
- Fichiers texte
- Extensible

### Ou ABCD (TDWG)

- >1200 concepts
- Utilisé par Biocase et Tapir





Global Biodiversity Information Facility  
Système Mondial d'Information sur la Biodiversité

> Accès libre et gratuit aux données de la biodiversité



# Les Data Papers



## Data Paper : qu'est-ce que c'est?

Une **publication scientifique\*** dont le but principal est de **décrire un ensemble de données** ou un groupe d'ensembles de données, plutôt que de rendre compte d'analyses et de résultats de recherche.

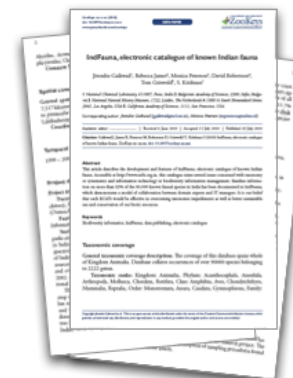
Son but est de **mettre en valeur les jeux de données**, ainsi que le temps passé par les gestionnaires de données à les structurer, nettoyer et mettre en ligne,

\* Publiée dans une revue scientifique à comité de lecture




The image shows a screenshot of the GBIF metadata form. It includes fields for 'Basic Metadata', 'Resource Contact', and 'Resource Description'. The form is titled 'GBIF METADATA PUBLISHING TOOLKIT' and contains various input fields and dropdown menus for entering data.

Document de métadonnées



Data Paper



## Une proposition concrète pour mettre en valeur les données et leurs gestionnaires



Depuis le développement des bases de données et des projets d'informatisation, la mauvaise valorisation du temps passé à informatiser des données et des connaissances est dénoncée par les scientifiques .

=> Une réflexion conjointe du **GBIF** et de l'éditeur de revues **PENSOFT** a abouti à une proposition en 2011 :

### Le Data Paper





## Data Paper : avantages



DOI : indexation et citation  
- indexés par Web of Knowledge (ISI), PubMedCentral, Scopus, Zoological Record, Google Scholar, CAB Abstracts, DOAJ, EBSCO.



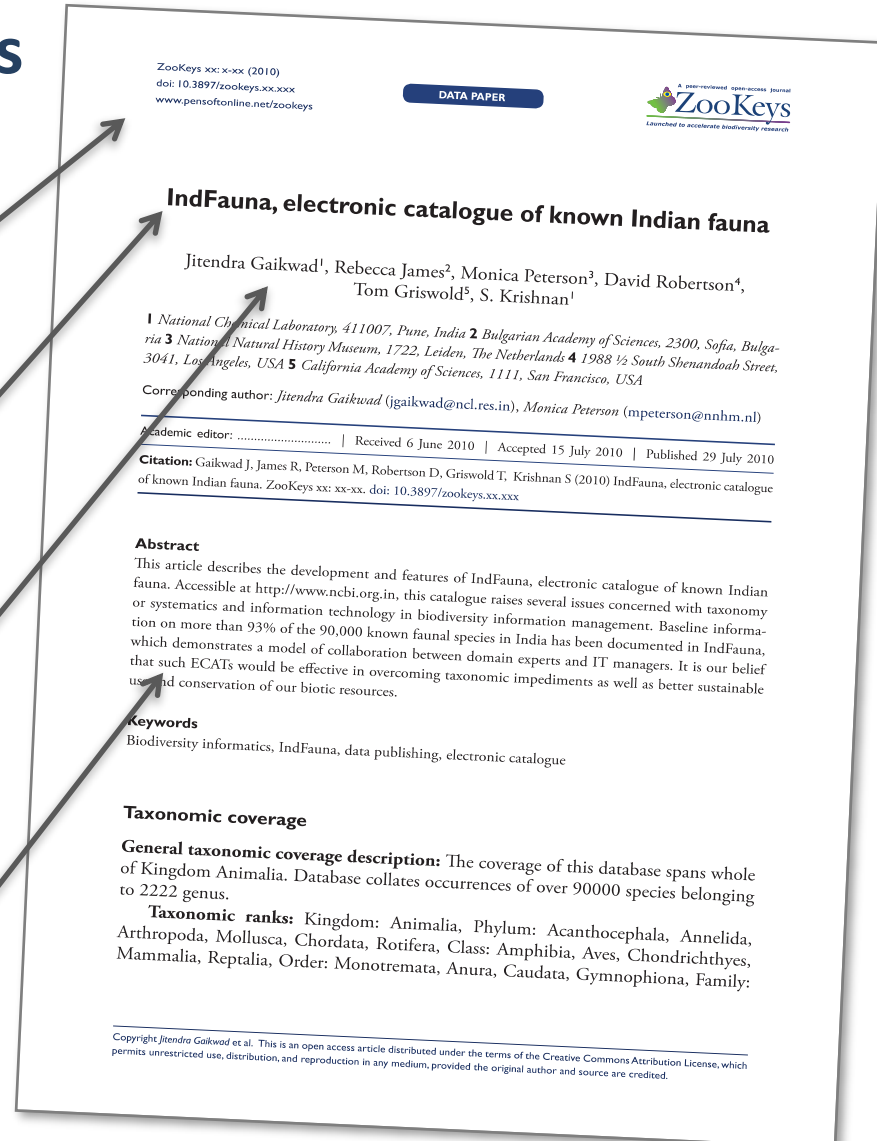
Promouvoir et faire connaître les données



Reconnaissance des éditeurs de données via une publication scientifique



Décrit les données sous forme structurée et lisible par un humain



ZooKeys xxx-xxx (2010)  
doi: 10.3897/zookeys.xxx.xxx  
www.pensoftonline.net/zookeys

**DATA PAPER**

**ZooKeys**  
An international open-access journal  
Launched to accelerate biodiversity research

### IndFauna, electronic catalogue of known Indian fauna

Jitendra Gaikwad<sup>1</sup>, Rebecca James<sup>2</sup>, Monica Peterson<sup>3</sup>, David Robertson<sup>4</sup>, Tom Griswold<sup>5</sup>, S. Krishnan<sup>1</sup>

<sup>1</sup> National Chemical Laboratory, 411007, Pune, India <sup>2</sup> Bulgarian Academy of Sciences, 2300, Sofia, Bulgaria <sup>3</sup> National Natural History Museum, 1722, Leiden, The Netherlands <sup>4</sup> 1988 ½ South Shenandoah Street, 3041, Los Angeles, USA <sup>5</sup> California Academy of Sciences, 1111, San Francisco, USA

Corresponding author: Jitendra Gaikwad (jgaikwad@ncl.res.in), Monica Peterson (mpeterson@ncl.res.in)

Academic editor: ..... | Received 6 June 2010 | Accepted 15 July 2010 | Published 29 July 2010

**Citation:** Gaikwad J, James R, Peterson M, Robertson D, Griswold T, Krishnan S (2010) IndFauna, electronic catalogue of known Indian fauna. ZooKeys xx: xx-xx. doi: 10.3897/zookeys.xx.xxx

**Abstract**  
This article describes the development and features of IndFauna, electronic catalogue of known Indian fauna. Accessible at <http://www.ncbi.org.in>, this catalogue raises several issues concerned with taxonomy or systematics and information technology in biodiversity information management. Baseline information on more than 93% of the 90,000 known faunal species in India has been documented in IndFauna, which demonstrates a model of collaboration between domain experts and IT managers. It is our belief that such ECATs would be effective in overcoming taxonomic impediments as well as better sustainable use and conservation of our biotic resources.

**Keywords**  
Biodiversity informatics, IndFauna, data publishing, electronic catalogue

**Taxonomic coverage**  
**General taxonomic coverage description:** The coverage of this database spans whole of Kingdom Animalia. Database collates occurrences of over 90000 species belonging to 2222 genus.  
**Taxonomic ranks:** Kingdom: Animalia, Phylum: Acanthocephala, Annelida, Arthropoda, Mollusca, Chordata, Rotifera, Class: Amphibia, Aves, Chondrichthyes, Mammalia, Reptalia, Order: Monotremata, Anura, Caudata, Gymnophiona, Family:

Copyright Jitendra Gaikwad et al. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.



# L'Integrated Publishing Toolkit (IPT) facilite le remplissage des métadonnées et la production automatisée d'un manuscrit de Data Paper



Logged in as pamerlon@gbif.fr ACCOUNT LOGOUT ENGLISH

Home

Manage Resources

Administration

About

Resource Title [Abundances and biological traits of the juveniles salmon sampled in the survey of Salmon abundance Indices in the Scorff river \(France\)](#)

## Basic Metadata

Please enter all the mandatory properties on the Basic Metadata page, and then continue entering metadata in the other pages that are applicable to your resource. The more metadata you provide, the greater the chance that your resource will be found, reused by other researchers, and cited.

Title\*

Abundances and biological traits of the juveniles salmon sampled in the survey of Salmon abundance Indices in

Publishing Organisation\*

Institut national de recherc

Type\*

Occurrence

Metadata Language\*

English

Update Frequency\*

Unknown

Subtype

Select a subtype

Data Language\*

English

Data Licence\*

Creative Commons Attribution (CC-BY) 4.0

## Section

### [Basic Metadata](#)

[Geographic Coverage](#)

[Taxonomic Coverage](#)

[Temporal Coverage](#)

[Keywords](#)

[Associated Parties](#)

[Project Data](#)

[Sampling Methods](#)

[Citations](#)

[Collection Data](#)

[External links](#)

[Additional Metadata](#)



This work is licensed under a [Creative Commons Attribution \(CC-BY\)](#)



# L'Integrated Publishing Toolkit (IPT) facilite le remplissage des métadonnées et la production automatisée d'un manuscrit de Data Paper



- Summary
- Data Records
- Downloads
- Versions
- How to cite
- Rights
- GBIF Registration
- Keywords
- Contacts
- Geographic Coverage
- Taxonomic Coverage
- Temporal Coverage
- Project Data
- Sampling Methods

## Abundances and biological traits of the juveniles salmon sampled in the survey of Salmon abundance Indices in the Scorff river (France)

Latest version published by Institut national de recherche pour l'agriculture, l'alimentation et l'environnement (INRAE) on Feb 28, 2020

A survey started in 1993 is conducted every year in early autumn (late September to early October) to quantify the abundance of juvenile Atlantic salmon in the Scorff In Brittany. The electric fishing protocol of Prévost and Baglinière (1995) is used. It targets Atlantic salmon young of the year (0+ parr), but older fish (juvenile salmon  $\geq 1+$ ) are also caught and included in this dataset. Sampling is restricted to areas with shallow running water flowing on coarse bottom substrate, i.e. the preferred habitat of young of the year salmon. The data consist of abundance indices and biological traits measured on the fish sampled: sex, maturity status, length, weight and age. The latter is ascertained from scale samples taken from the fish which size does not allow to determine their age unambiguously. The survey is carried out under the Research Observatory on Diadromes Fishes in Coastal Streamss (ERO DiaPFC) project. The data are stored in the database of the ERO. They are used to develop predictive models and tools for providing scientific advice to improve the management of this heritage species.



- [Home](#)
- [GBIF](#)
- [DwC-A](#)
- [EML](#)
- [RTF](#)
- [Versions](#)
- [Rights](#)
- [Cite this](#)

```

<?xml:namespace prefix="xsi" schemaLocation="http://ecoinformatics.org/eml-2.1.1 http://rs.gbif.org/schema/eml-gbif-profile/1.1/eml-2.1.1" />
<eml:eml xsi:schemaLocation="http://ecoinformatics.org/eml-2.1.1 http://rs.gbif.org/schema/eml-gbif-profile/1.1/eml-2.1.1" system="http://gbif.org" scope="system" xml:lang="eng">
  <dataset>
    <alternateIdentifier>89064e3a-aa3c-495d-b236-092e1dae7042</alternateIdentifier>
    <alternateIdentifier>http://ipt.gbif.fr/resource?r=inra_rennes_scorff</alternateIdentifier>
    <title xml:lang="eng">
      Abundances and biological traits of the juveniles salmon sampled in the survey of Salmon abundance Indices in the Scorff river (France)
    </title>
    <creator>
      <individualName>
        <givenName>Nicolas</givenName>
        <surName>Jeannot</surName>
      </individualName>
      <organizationName>
        INRAE U3E, Pôle OFB-INRAE-Agrocampus Ouest-UPPA pour la gestion des migrateurs amphihalins dans le Scorff
      </organizationName>
      <positionName>Researcher</positionName>
      <address>
        <deliveryPoint>Moulin des Princes</deliveryPoint>
        <city>Pont-Scorff</city>
        <postalCode>56620</postalCode>
        <country>FR</country>
      </address>
      <phone>+33297325429</phone>
      <electronicMailAddress>Nicolas.jeannot@inrae.fr</electronicMailAddress>
    </creator>
    <creator>
      <individualName>
        <givenName>Didier</givenName>
        <surName>Azam</surName>
      </individualName>
      <organizationName>
        INRAE U3E, Pôle OFB-INRAE-Agrocampus Ouest-UPPA pour la gestion des migrateurs amphihalins dans le Scorff
      </organizationName>
    </creator>
  </dataset>
  </eml:eml>
  
```



## www.gbif.org/data-papers

GBIF has been working with partners in academic publishing to promote the data paper as a means of bringing credit and recognition to all those involved in data publication; to alert the scientific community to the existence of biodiversity datasets and the value they can bring to particular research projects; and as a mechanism for quality assessment and control of data accessible through GBIF and other networks.

### Journals that publish biodiversity-related data papers

Please notify us at [communication@gbif.org](mailto:communication@gbif.org) with corrections or updates.

Journal	Publisher	Open Access (license)	APC estimate	Journal Impact Factor (2019)	Scopus CiteScore (2019)
<a href="#">Anxius de Miscel·lània Zoològica</a>	Nat Hist Museum of Barcelona	CC BY	0	-	0.7
<a href="#">Big Earth Data</a>	Taylor & Francis Online	CC BY	EUR 530	-	-
<a href="#">Biodiversity Data Journal</a>	Pensoft	CC BY	EUR 475	1.331	2.1
<a href="#">Biogeographia</a>	eScholarship	CC BY	0	-	1.9
<a href="#">BioInvasions Records</a>	REABIC Journals	CC BY	EUR 420 (EUR 70 pr. page)	1.504	1.9
<a href="#">BioRisk</a>	Pensoft	Yes	EUR 300	-	0.6



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

SCIENTIFIC DATA

Home | About | For Authors | Advisory and Editorial Board | Open Access | Contacts | FAQ

Helping you publish, discover,  
and reuse research data

<p><b>Credit</b> Credit, through a citable publication, for depositing &amp; sharing your data</p>	<p><b>Reuse</b> Complete, curated &amp; standardized descriptions enable the reuse of your data</p>	<p><b>Quality</b> Rigorous community based peer review</p>
<p><b>Discovery</b> Find datasets relevant to your research</p>	<p><b>Open</b> Promotes &amp; endorses open science principles &amp; available to all through a Creative Commons license</p>	<p><b>Service</b> In-house curation, rapid peer review &amp; publication of your data descriptions</p>







# L'outil Arpha Writing Tool (Pensoft)

*Un outil - parmi d'autres - d'aide à la rédaction de Data papers*



**Pensoft** : maison d'édition de littérature scientifique, fondée en 1992 en Bulgarie

A développé les Data Papers sur les données de biodiversité en partenariat avec le GBIF (Global Biodiversity Information Facility) en 2010.

<https://pensoft.net>

# Growing family of Pensoft open access journals

The collage features the following journals and logos:

- Zoologia: An International Journal for Zoology
- Rethinking Ecology
- Italian Botanist
- Myriapodology
- Nature Conservation
- NeoBiota
- Subterranean Biology
- Biodiversity Data Journal
- PhytoKeys
- MycoKeys
- ZooKeys
- Check List: the journal of biodiversity data
- EVOLUTIONARY SYSTEMATICS
- Journal of Orthoptera Research
- Alpine Entomology
- COMPARATIVE Cytogenetics
- BioRisk
- Journal of Hymenoptera
- NOVA LEPIDOPTEROLOGICA
- RIO: Research Ideas and Outcomes
- MBMG: Metabarcoding & Metagenomics
- BioDISCOVERY
- One Ecosystem: Ecology and Sustainability Data Journal
- BISS: Biodiversity Information Science and Standards



**Outil de rédaction ARPHA** : facilite la mise en page, la soumission, le processus de relecture, la publication, l'hébergement et l'archivage d'articles scientifiques.

# ARPHA

an end-to-end journal publishing platform

Authoring, Reviewing  
Publishing, Hosting, Archiving



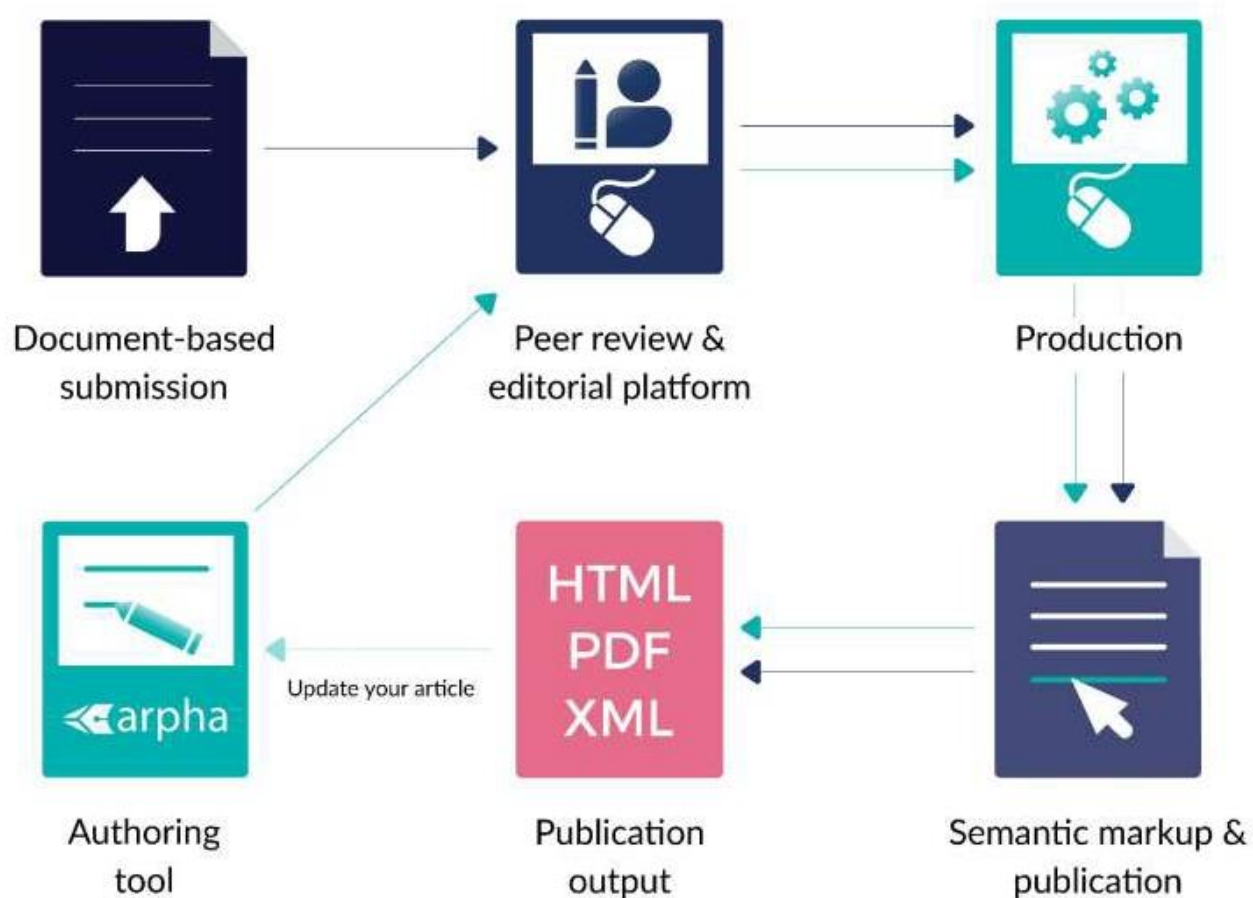
# Interface utilisateur

The screenshot displays the GBIF user interface for editing a 'Data Paper'. The interface is divided into several sections:

- Top Navigation:** Includes the 'arpha' logo, a 'writing tool' icon, and navigation links for 'Email co-authors and contributors', 'helpdesk', 'Tips and tricks', 'tutorial', 'Revision history', and the user profile 'Ms Sophie Pamerlon'.
- Left Sidebar:** A navigation menu with categories like 'Data Paper (Biosciences)', 'Authors', 'Contributors', 'Article metadata' (with sub-items: Title, Abstract & Keywords, Classifications, Funder), 'Introduction' (with sub-item: First section), 'General description', 'Project description', 'Sampling methods', 'Geographic coverage', 'Taxonomic coverage', 'Traits coverage', 'Temporal coverage', 'Collection data', 'Usage rights', 'Data resources', 'Additional information', and 'Acknowledgements'.
- Main Content Area:**
  - Header: 'Biodiversity Data Journal : Data Paper (Biosciences)' with a 'Print' icon.
  - Title: 'Data Paper' in a large, bold, blue box with an edit icon on the left and a comment icon on the right.
  - Author: 'Sophie Pamerlon' with a plus sign icon.
  - Affiliation: 'GBIF France, Paris, France'.
  - Corresponding author section: 'Corresponding author:' followed by a copyright notice '© 2018 Sophie Pamerlon' and a citation 'Citation: () . <https://doi.org/>'. An 'OPEN ACCESS' logo is also present.
  - Section Headers: 'Abstract', 'Background', 'New information', and 'Introduction' are listed. Under 'Background', there is a link 'Marchand 2017'. Under 'New information', the text 'test' is visible.
  - Section Headers: 'Introduction' and 'First section' are listed.
- Right Sidebar:** Contains a 'Resolved' status indicator and a 'Filter' button.



## Deux voies de publication possibles : document ou XML





## Outils et fonctionnalités interoperables

Integrated with other systems via  
**web services & data exchange**

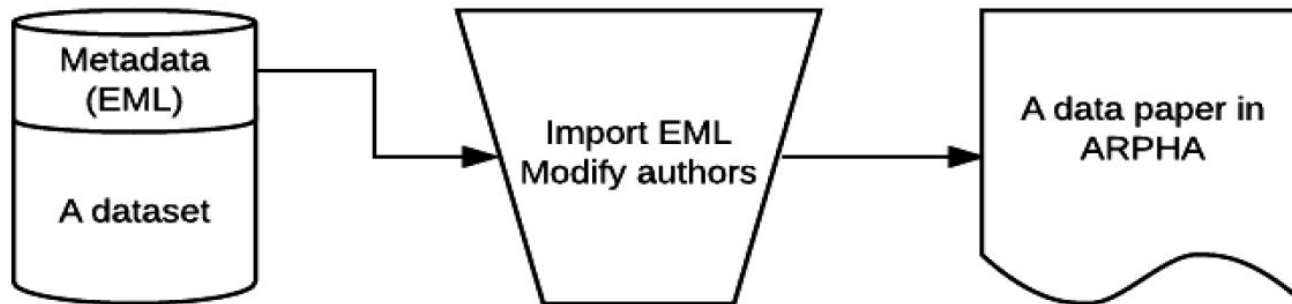




## Outils et fonctionnalités interopérables

Exemple : métadonnées GBIF (IPT) → Data paper (ARPHA)

Generate and import an **entire manuscript**



DataONE



GBIF

LTER

Name	Email	Include	Is Submitting
Viktor Senderov	datascience@pensoft.net	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
NoName PleaseChange	NoName@PleaseChange.This	<input type="checkbox"/>	<input type="checkbox"/>
Oscar Schofield	oscar@marine.rutgers.edu	<input type="checkbox"/>	<input type="checkbox"/>

Submit

Manuscript title page: DataONE (DataONE)

Photosynthetic pigments of water column samples analyzed using High Performance Liquid Chromatography (HPLC), sampled during Palmer LTER field season at Palmer Station Antarctica, 1991 - 2009.

Oscar Schofield<sup>1</sup>

<sup>1</sup>Rutgers University, ...

Corresponding author: oscar.schofield@marine.rutgers.edu

© 2010 Oscar Schofield

October 2, 2010

ARPHA



## Exemple de Data Paper publié via Arpha Writing Tool :



Articles About

About Pensoft

Books

E-Books

Blog

Journals



Data Paper

Biodiversity Data Journal 5: e15125 (09 Nov 2017)  
<https://doi.org/10.3897/BDJ.5.e15125>



### Abundance indices and biological traits of juvenile salmon (*Salmo salar*) sampled in three rivers on the Atlantic and Channel coasts (France)

▼ Frédéric Marchand, Laurent Beaulaton, Etienne Prévost, Richard Delanoë, Jean-Pierre Destouches, François Gueraud, Yoann Guilloux, Nicolas Jeannot, Emmanuel Huchet, Frédéric Lange, Jacques Rives, Julien Tremblay, Nadine Herrard, Didier Azam

#### Abstract ▲

#### Background

Atlantic Salmon (*Salmo salar*) is an anadromous migratory species adapted to cool temperatures. It is protected by the Bern convention and by the European Habitats Directive. It has been listed as vulnerable by the French IUCN Red List. Salmon decline is the result of combined and cumulated, mainly anthropic, causes: climate change, increasingly high number of impoundments, degradation of water quality and habitat and over-exploitation by fisheries. Monitoring of this species has been carried out on three rivers in France (Southern part of the distribution area) to produce data and knowledge (growth, precocious maturity, survival) for stock management.

For 24 years, a specific and standardised electric fishing protocol has been used to target young-of-the-year (0+ parr) Atlantic salmon. Sampling was restricted to areas with shallow running water that flows over a coarse bottom substrate, i.e. the preferred habitat of young salmon. This monitoring and inventory of growing areas thus allows assessment of juvenile recruitment and provides baseline data required to calculate total allowable catches (TACs).

Contents	Article info	Citation	Metrics	Comment	Related
Figs	Taxa	Data	Refs	Cited	

- Article metadata
- Introduction
- Project description
- Sampling methods
- Geographic coverage
- Taxonomic coverage
- Temporal coverage
- Collection data
- Usage rights
- Data resources
- Additional information
- Acknowledgements
- References
- Supplementary files





## Un des jeux de données correspondants sur [www.gbif.org](http://www.gbif.org) :



Get data Share Tools Inside GBIF

OCCURRENCE DATASET | REGISTERED 6 MARCH 2017

### Abundances indices and biological traits of juvenile salmon sampled in the Scorff river (France)

Published by [Institut National de la Recherche Agronomique \(INRA\)](#)

Nicolas Jeannot • Didier Azam • Yoann Guilloux • Etienne Prévost • Frédéric Marchand

DATASET PROJECT STATS ACTIVITY [DOWNLOAD](#) [DATASET HOMEPAGE](#)

34 940 OCCURRENCES 1 CITATION

A survey started in 1993 is conducted every year in early autumn (late September to early October) to quantify the abundance of juvenile Atlantic salmon in the Scorff in Brittany. The electric fishing protocol of Prévost and Baglinière (1995) is used. It targets Atlantic salmon young of the year (0+ parr), but older fish (juvenile salmon  $\geq 1+$ ) are also caught and included in this dataset. Sampling is restricted to areas with shallow running water flowing on coarse bottom substrate, i.e. the prefe... [more](#)

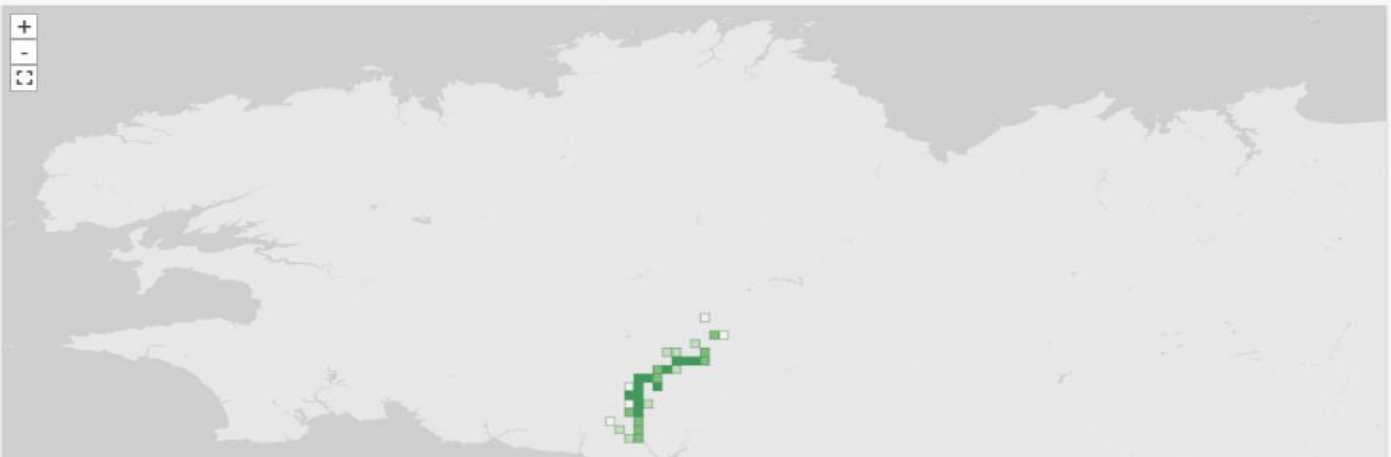
Last Modified: 28 September 2017

License: CC BY 4.0

[How to cite](#) [DOI](#) [10.15468/mz4lyw](https://doi.org/10.15468/mz4lyw)

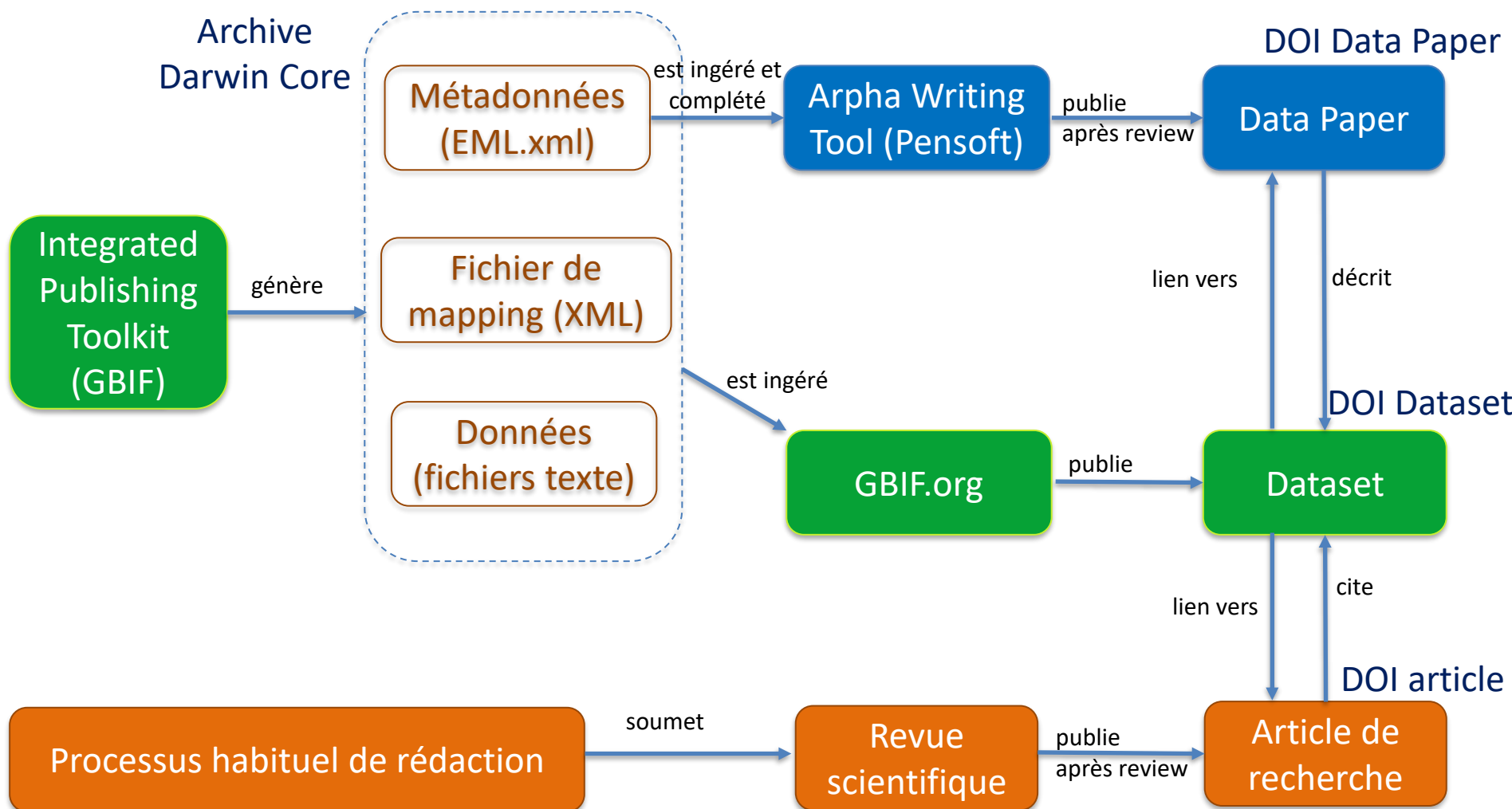


34 940 GEOREFERENCED RECORDS





## En résumé





## Contacts

Equipe GBIF France  
[gbif@gbif.fr](mailto:gbif@gbif.fr)

Site web

<http://www.gbif.fr>

Twitter

@gbiffrance