Doi: 10.5281/zenodo.7081339

# The significance of taxonomic publications to understand biodiversity in the digital world

September 15, 2022

**Bauhin 2022 Conference**

**Basel, Switzerland**
**Donat Agosti**

ORCID: 0000-0001-9286-1200; @myrmoteras

**Plazi, Switzerland**

SIBILS
SWISS INSTITUTE OF BIOINFORMATICS
LITERATURE SERVICES

swissuniversities

101007492

ARCADIA

BiCIKL

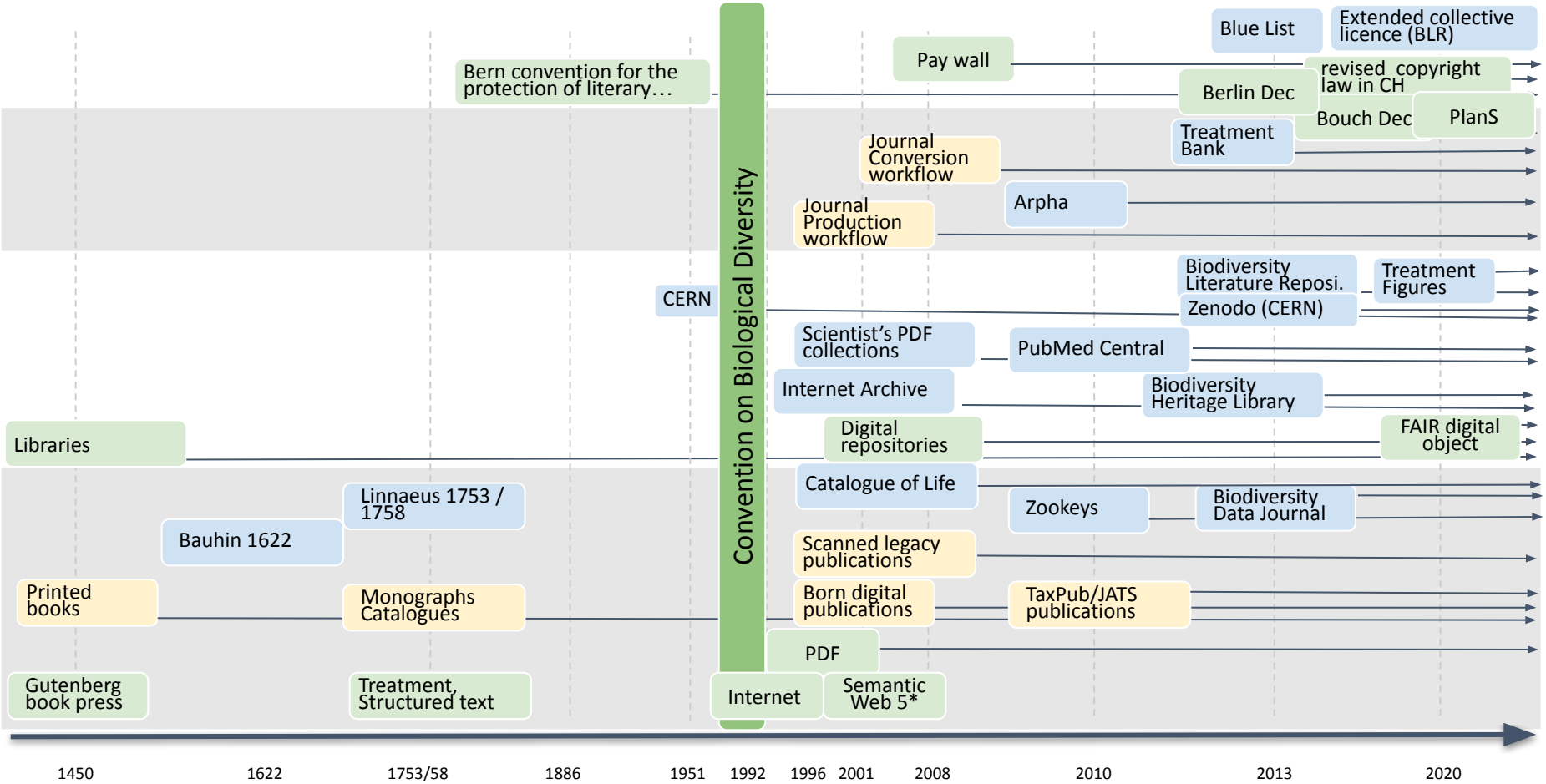
eBioDiv

Biodiversity Literature Repository

mobilise Mobilise Action
Mobilising Data, Policies and Experts in Scientific Collections

TreatmentBank

# Brief history of publishing in biodiversity / taxonomy



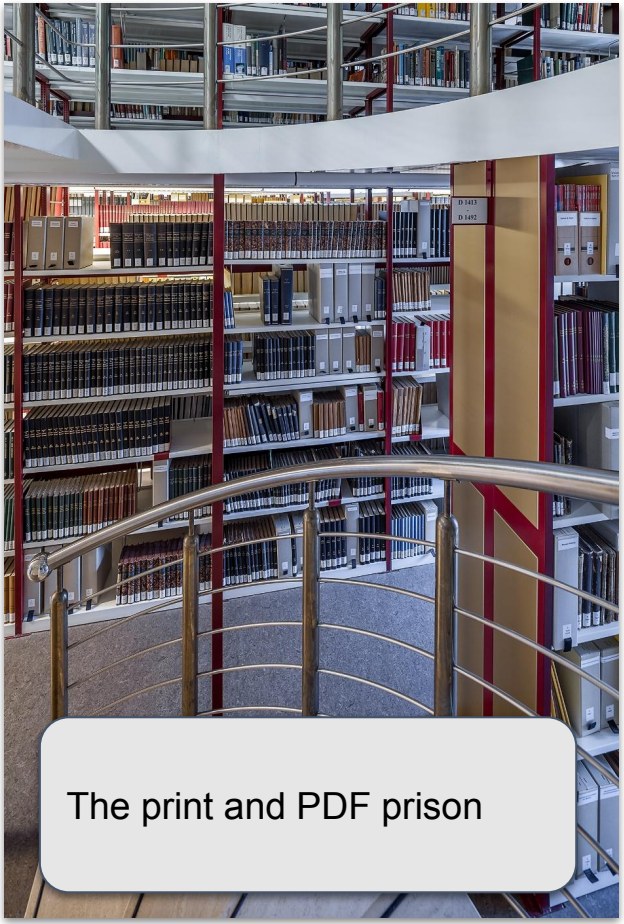
# Convention on Biological Diversity

# The challenge


United Nations

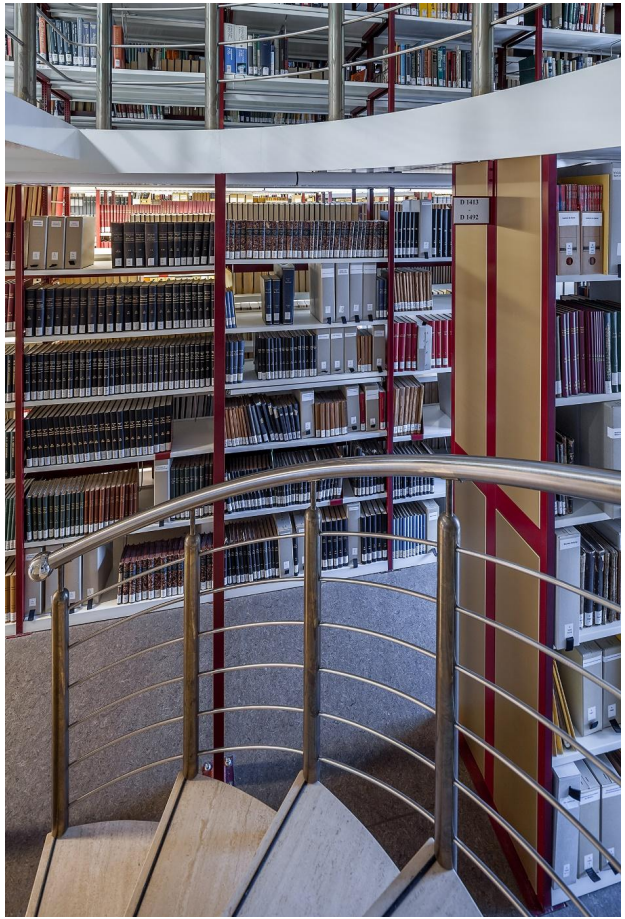
**Article 7. Identification and Monitoring**
How many species do we lose?
How many species do we know?
How many species are on Earth?

**Article 6. General Measures for Conservation and Sustainable Use**
What do we know about the species?




BGBM.org

# The challenge



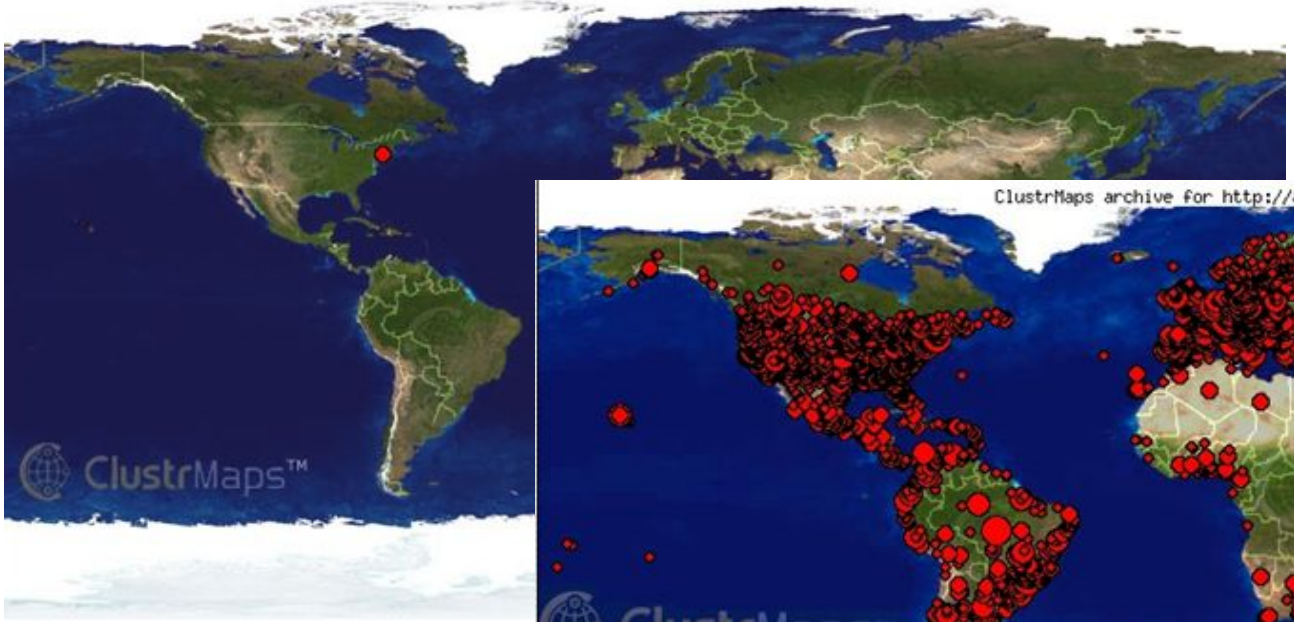
## Known biodiversity knowledge

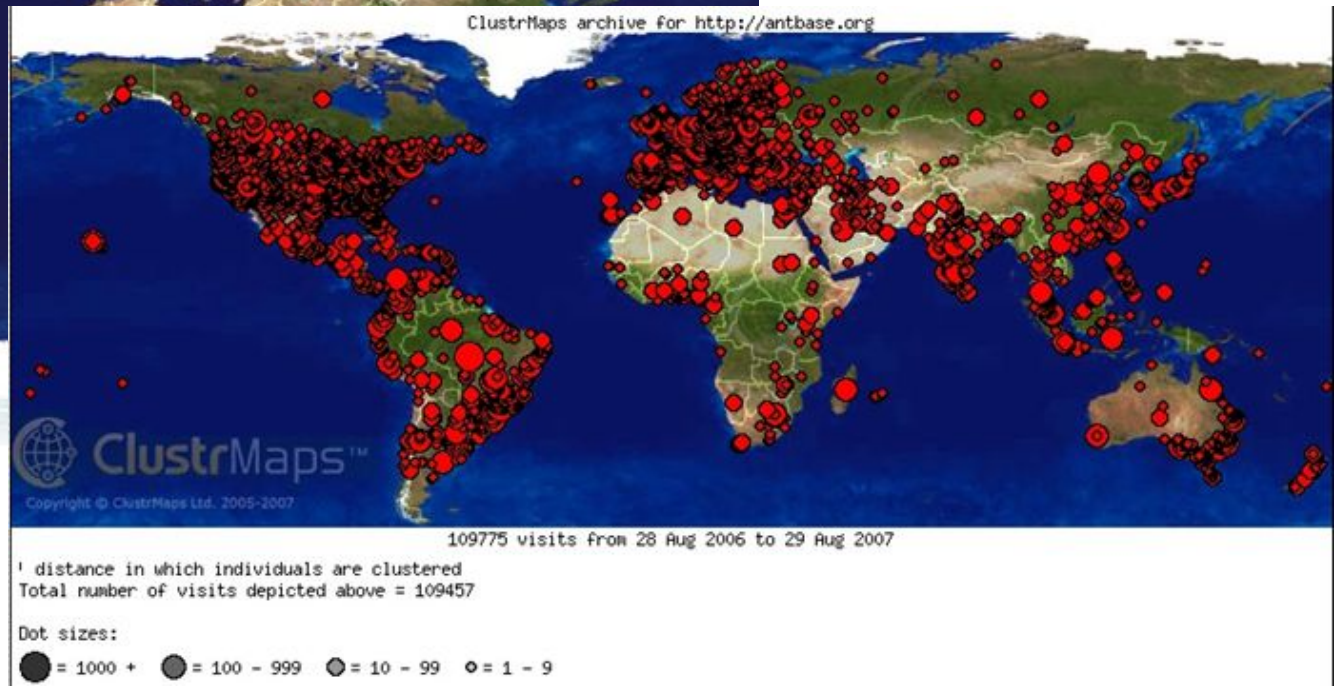
- Empiric science: all results published
- 500,000,000+ printed pages
- >> 1,000 journals publishing taxonomic content
- > 1,900,000 species described
- > 20,000,000+ taxonomic treatments
- approx. 17,000 new species discovered / year
- >> Millions of specimens identified by specialists (material citations)
- Billions of facts

BUT: only ca. 10-30% of the knowledge is digital. Most is “unknown known knowledge”, not Digital Accessible Knowledge (DAK)

# The impact of WWW in the 90ties

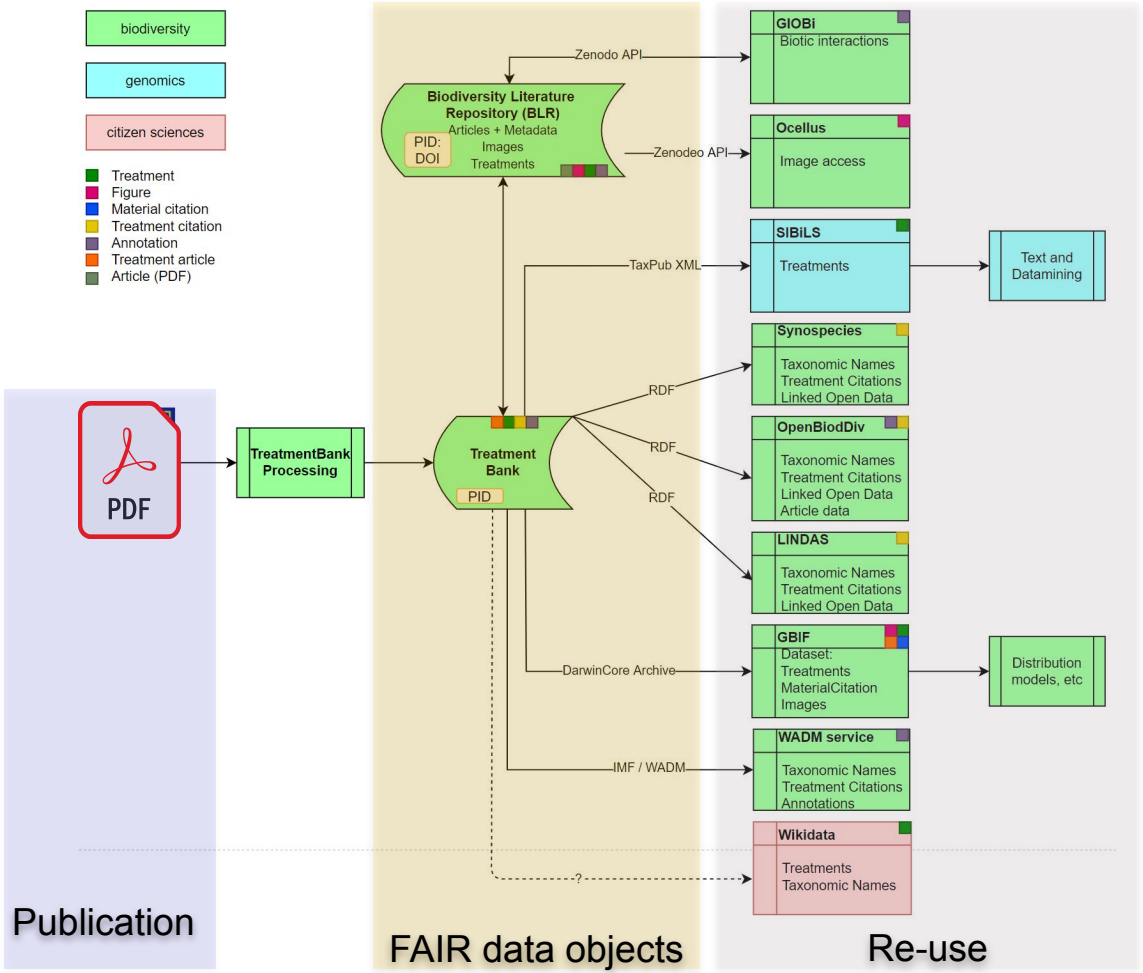
Before antbase.org, Harvard‘s Museum of Comparative Zoology could claim to be the only location with a complete set of ant systematics publications from 1758 - present.




Through antbase.org‘s digital library, access to this body of literature is worldwide, and it is actively used (>10,000 visits in one month only).

# Re-use of scholarly publication's FAIR data



# Plazi: Goal

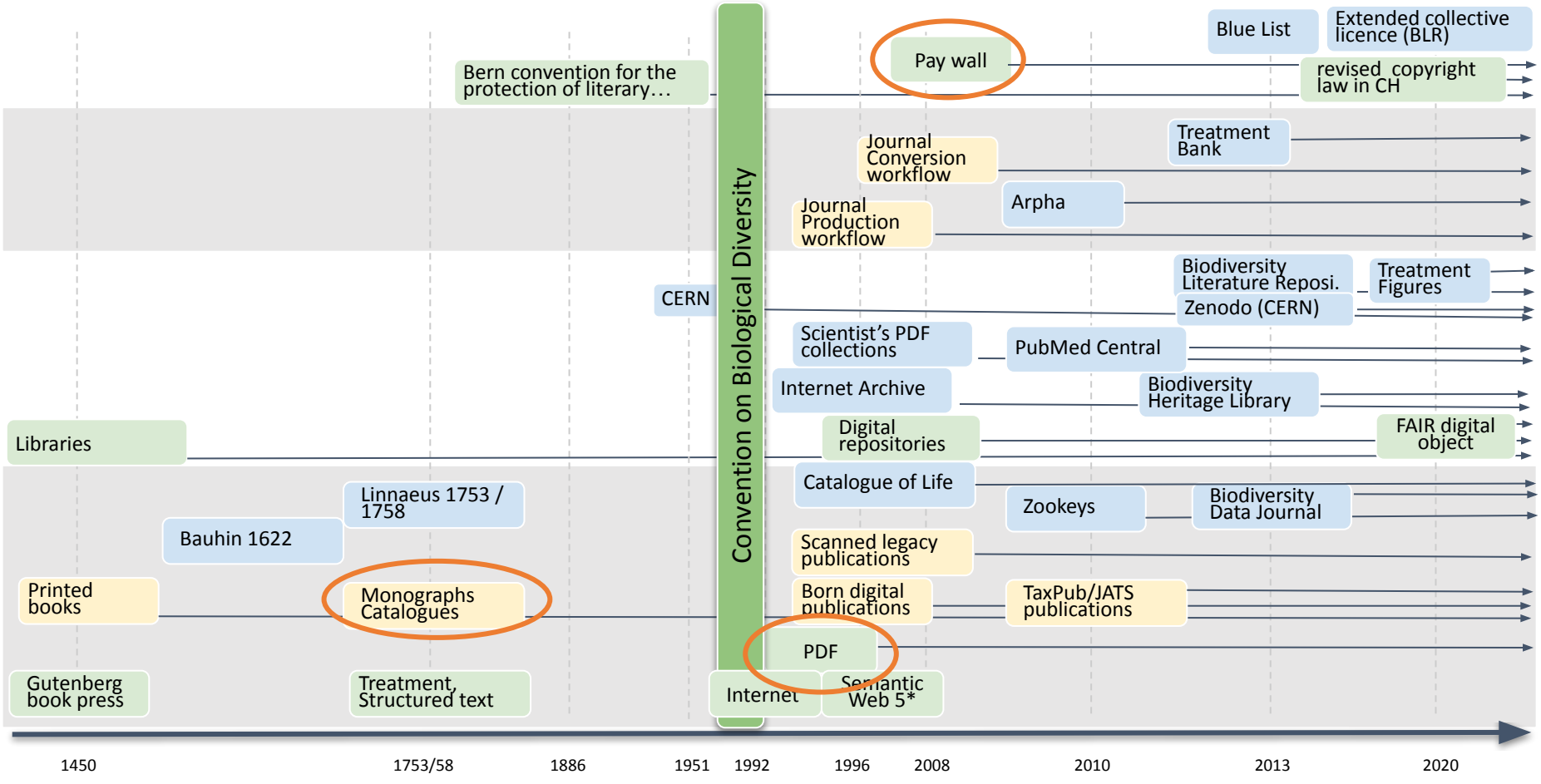
**Discovering known biodiversity**

Create a list of the Earth' known taxa, and related digital accessible knowledge embedded in scholarly publications as open findable, accessible, interoperable and reusable data about the Earth's species (FAIR digital objects), as input to the biodiversity knowledge graph, liberated from scholarly publications.

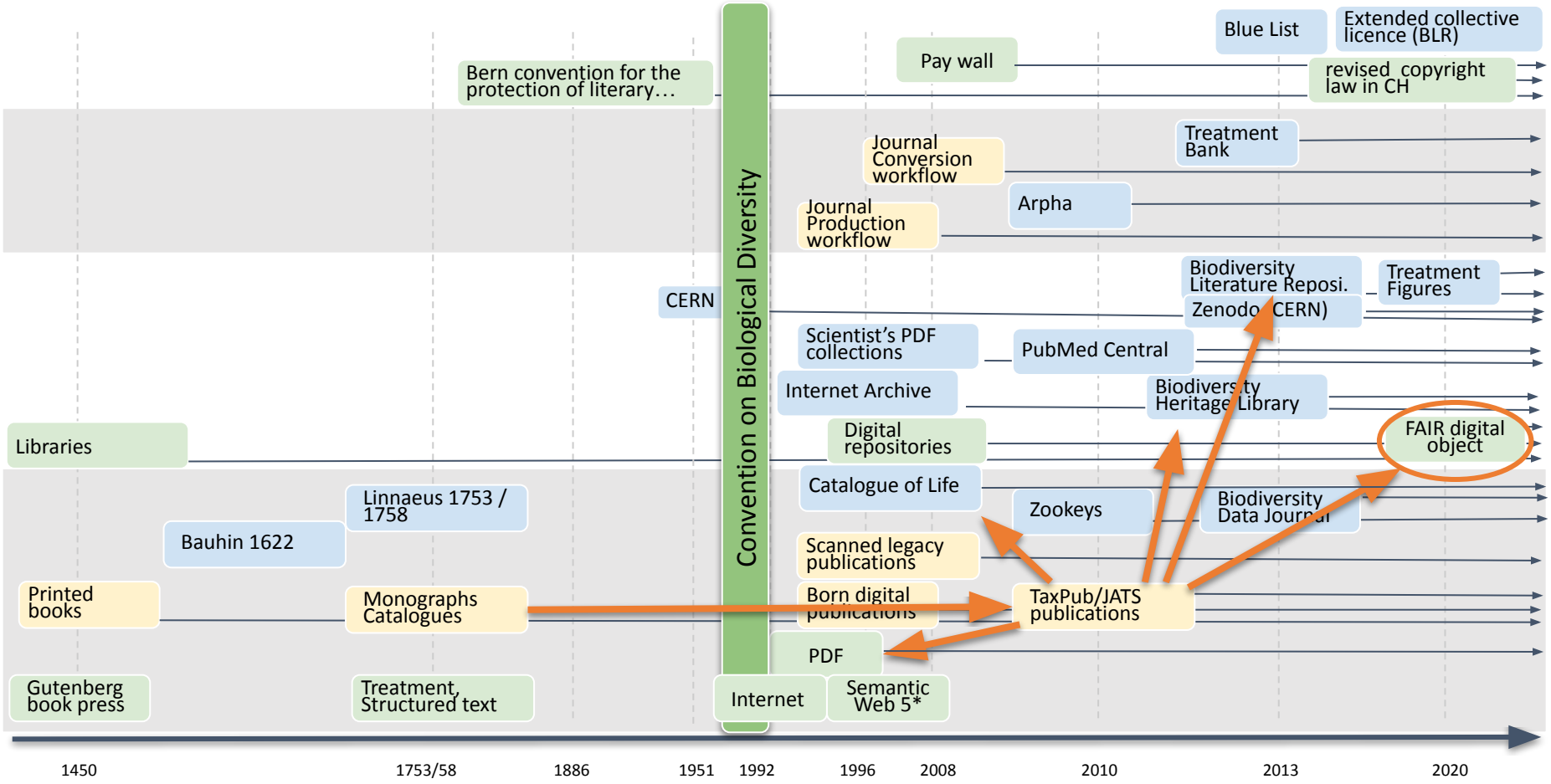
**Digital accessible knowledge in biodiversity (DAK)**

- Data understandable by human and actionable by machine reflecting the growth of our knowledge and interdependence of biodiversity.
- Editorial structure including textflow, paragraphs, sections such as title, authors and affiliations, materials and methods, etc., tables, figures, bibliographic references and their citations in the text.
- Semantic annotated (e.g. using TDWG standards) data at the base of the biodiversity knowledge graph:
    - data about a taxon (taxonomic treatments with their nomenclature section)
    - cited previous treatments (treatment citations)
    - cited specimens (material citations)
    - named entities (persons, taxonomic names, accession -, collection -, institution - or specimen codes
    - attributes including their persistent identifiers

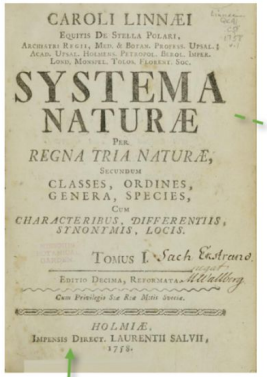
# Status of publishing in 2022



# Goal and future for publishing: also a reality



# Digital Accessible Knowledge: Taxonomic Treatment


CAROLI LINNÆI
EQUITIS DE STELLA POLARI,
ARCHIATRI REGII, MED. & BOTAN. PROFESS. UPSAL. &
ACAD. UPSAL. HOLMENS. PETROPOL. BEROL. IMPER.
LOND. MONSPEL. TOLOS. FLORENT. SOC.

SYSTEMA NATURÆ

PER
REGNA TRIA NATURÆ,
SECUNDUM
CLASSES, ORDINES,
GENERA, SPECIES,
CUM
CHARACTERIBUS, DIFFERENTIIS,
SYNONYMIS, LOCIS.

TOMUS I. Sach. Ecstrand

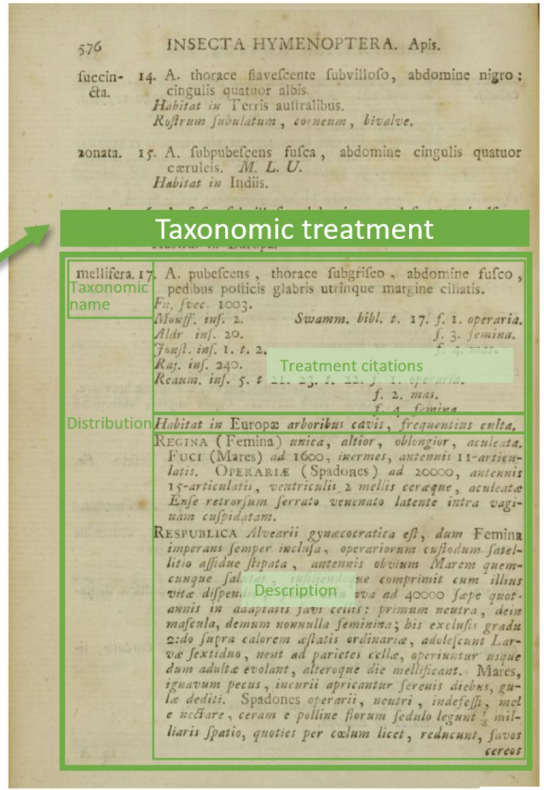
EDITIO DECIMA, REFORMATA.

Cum Privilegio Sac. Reg. Majtt. Suecia.

HOLMIÆ,
IMPENSIS DIRECT. LAURENTII SALVII,
1758.

Linnaeus, 1758
4,819 taxa

*Apis mellifera L.*



576 INSECTA HYMENOPTERA. Apis.

fuccinæta. 14. A. thorace flavescente subvilloso, abdomine nigro: cingulis quatuor albis.
*Habitat in* Terris australibus.
*Rostrum subulatum, corneum, bivalve.*

zonata. 15. A. subpubescens fusca, abdomine cingulis quatuor cæruleis. *M. L. U.*
*Habitat in* Indiis.

Taxonomic treatment

Taxonomic name

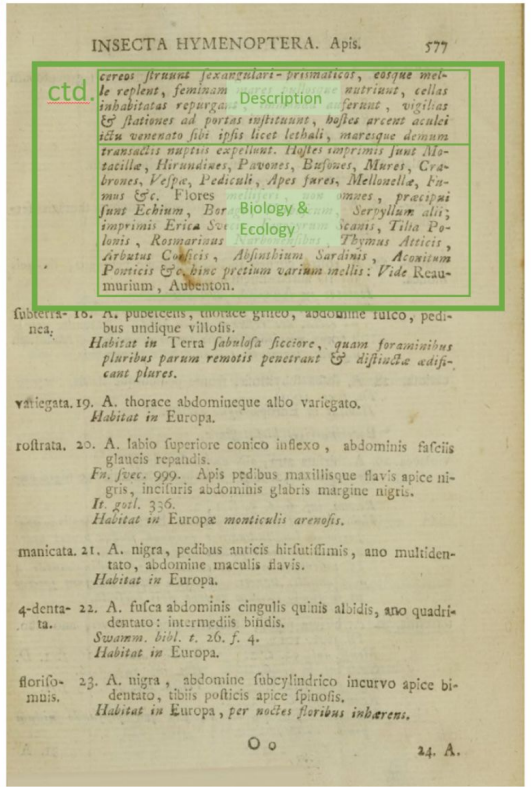
mellifera. 17. A. pubescens, thorace subgriseo, abdomine fusco, pedibus posticis glabris utrinque margine ciliatis.
*Fn. suec.* 1003.
*Mouff. inf.* 2. *Swamm. bibl. t.* 17. *f.* 1. *operaria.*
*Aldr. inf.* 20. *f.* 3. *femina.*
*Jonst. inf.* 1. *t.* 2.
*Raj. inf.* 240.
*Reaum. inf.* 5. *t.* *f.* 2. *mas.*
*f.* 4. *femina.*

Treatment citations

Distribution *Habitat in* Europæ *arboribus cavis, frequentius culta.*

Description

REGINA (Femina) *unica, altior, oblongior, aculeata.*
FUCI (Mares) *ad* 1600, *inermes, antennis* 11-*articulatis.*
OPERARIÆ (Spadones) *ad* 20000, *antennis* 13-*articulatis, ventriculis* 2 *mellis ceræque, aculeato Ense retrorsum serrato venenato latente intra vaginam cuspidatam.*
RESPUBLICA *Alvearii gynæcocratica est, dum Femina imperans semper inclusa, operariorum custodum satellitio assidue stipata, antennis obvium Marem quemcunque sal* *ne comprimit cum illius vitæ dispen* *ova ad* 40000 *sæpe quotannis in duabus anni vicibus: primum neutra, dein mascula, demum nonnulla feminina; his exclusis gradu 2do supra calorem æstatis ordinariæ, adolescunt Larvæ sextidno, nexit ad parietes cellæ, operiuntur usque dum adultæ evolant, alteroque die mellificant. Mares, ignavum pecus, incurii apricantur serenis diebus, gulæ dediti. Spadones operarii, neutri, indefessi, mel e nectare, ceram e polline florum sedulo legunt; milliaris spatio, quoties per cœlum licet, redeunt, favos cereos*


INSECTA HYMENOPTERA. Apis. 577

ctd. *cereos struunt sexangulari-prismaticos, eosque mellе replent, feminam* *nutriunt, cellas inhabitatas repurgant* *ferunt, vigilias & stationes ad portas instituunt, hostes arcent aculei ictu venenato sibi ipsis licet lethali, marasque demum transactis nuptiis expellunt. Hostes imprimis sunt Motacillæ, Hirundines, Pavones, Buſones, Mures, Crabrones, Vespæ, Pediculi, Apes fures, Mellonellæ, Fumus &c. Flores* *omnes, præsipui sunt Echium, Bor* *Serpyllum alii; imprimis Erica Sco* *canis, Tilia Polonis, Rosmarinus* *Thymus Atticis, Arbutus Corsicis, Absinthium Sardinis, Aconitum Ponticis &c. hinc pretium varium mellis: Vide* Reaumurium, Aubenton.

Description

Biology & Ecology

subterranea. 18. A. pubescens, thorace griseo, abdomine fusco, pedibus undique villosis.
*Habitat in* Terra *fabulosa ficciore, quam foraminibus pluribus parum remotis penetrant & distinctæ aedificant plures.*

variegata. 19. A. thorace abdomineque albo variegato.
*Habitat in* Europa.

rostrata. 20. A. labio superiore conico inflexo, abdominis fasciis glaucis repandis.
*Fn. suec.* 999. Apis pedibus maxillisque flavis apice nigris, incisuris abdominis glabris margine nigris.
*It. goth.* 336.
*Habitat in* Europæ *monticulis arenosis.*

manicata. 21. A. nigra, pedibus anticis hirsutissimis, ano multidentato, abdomine maculis flavis.
*Habitat in* Europa.

4-dentata. 22. A. fusca abdominis cingulis quinis albidis, ano quadridentato: intermediis binis.
*Swamm. bibl. t.* 26. *f.* 4.
*Habitat in* Europa.

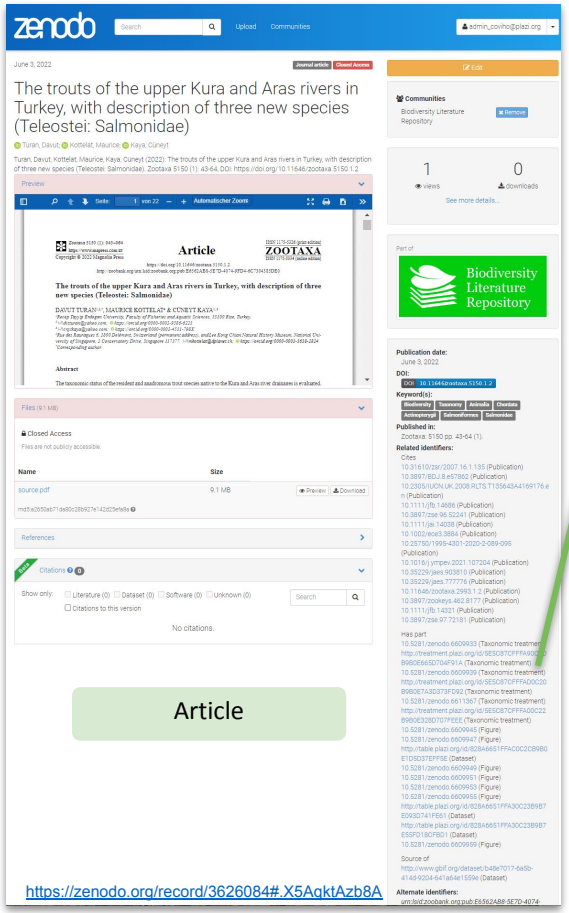
florisomnis. 23. A. nigra, abdomine subcylindrico incurvo apice bidentato, tibiis posticis apice spinosis.
*Habitat in* Europa, *per noctes floribus inhærens.*

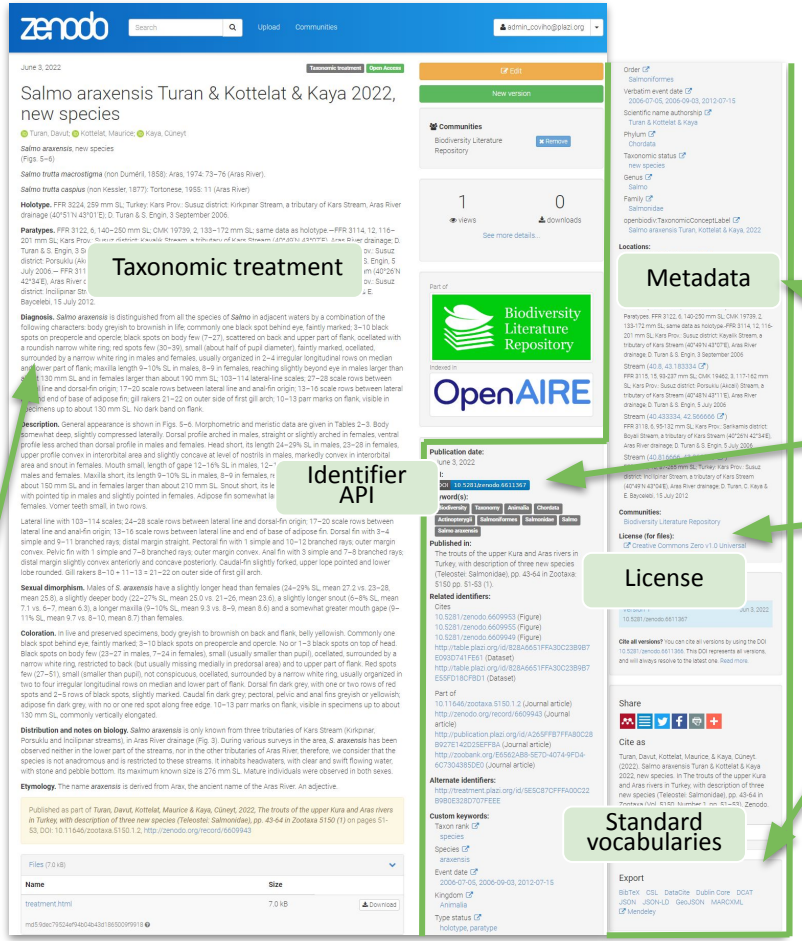
O o 24. A.

Each type has a published taxonomic treatment
Each taxonomic treatment is multiple times augmented

Tens of millions of treatments exist as part of ca 500 Million published pages of biodiversity literature
Each includes a numerous facts

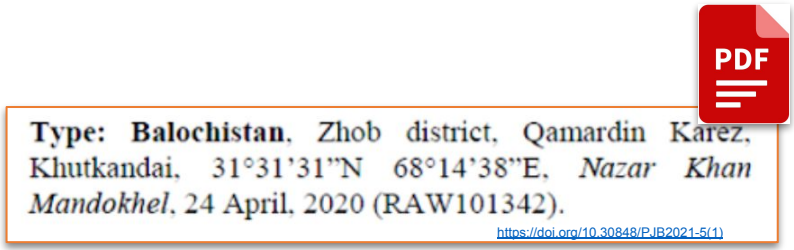
# Treatments as **FAIR digital objects** in the Biodiversity Literature Repository




Treatment DOI
Journal article BLR Record
Article DOI

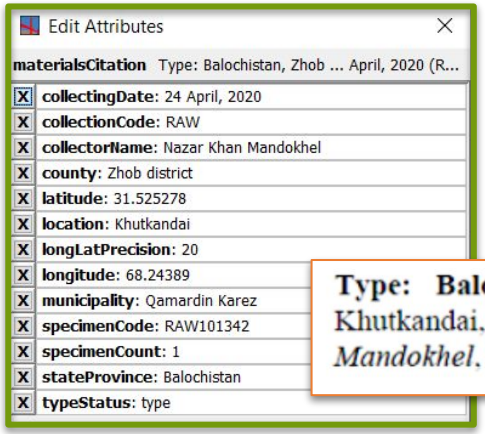
# Published materials citation links



**Type: Balochistan**, Zhob district, Qamardin Karez, Khutkandai, 31°31'31"N 68°14'38"E, *Nazar Khan Mandokhel*, 24 April, 2020 (RAW101342).

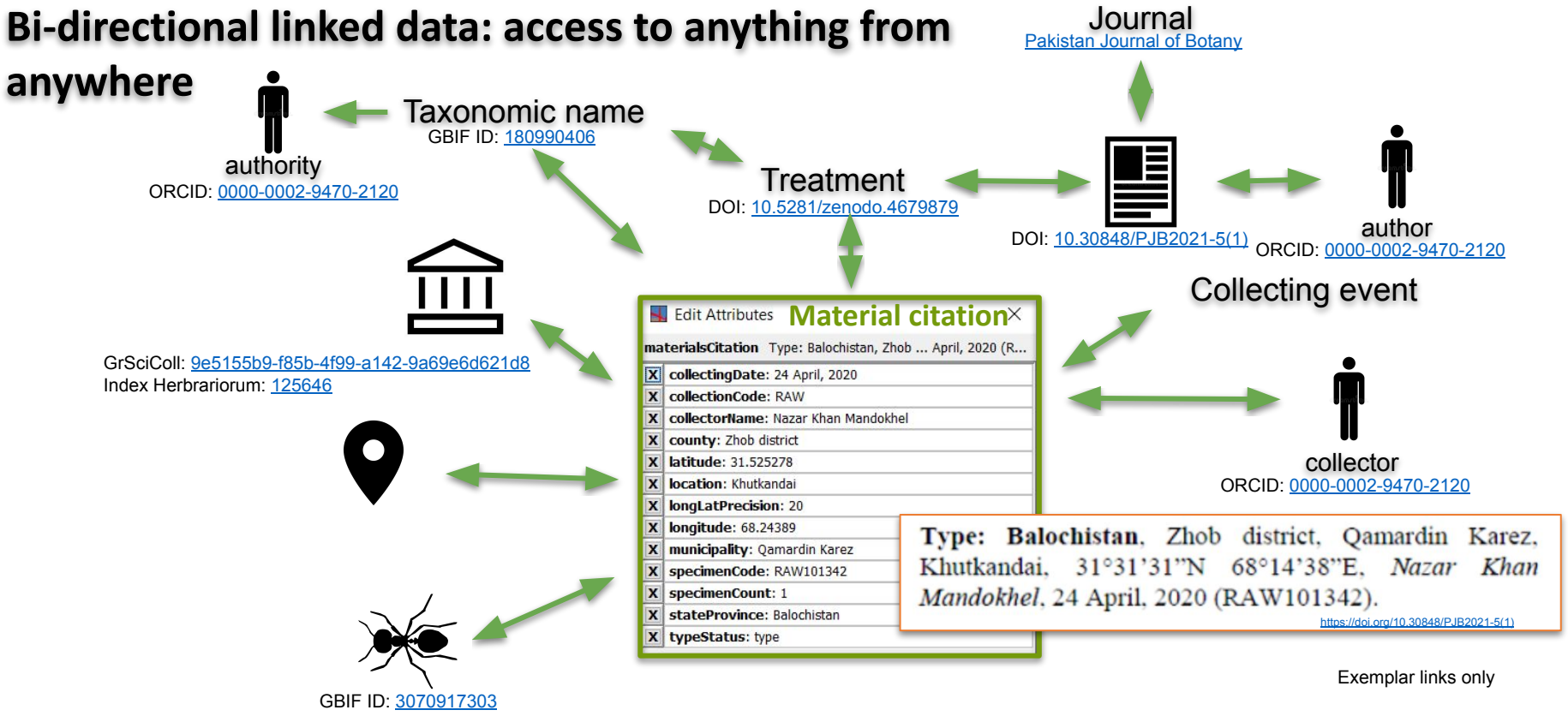
https://doi.org/10.30848/PJB2021-5(1)

# Published materials citation links




**Type: Balochistan**, Zhob district, Qamardin Karez, Khutkandai, 31°31'31"N 68°14'38"E, *Nazar Khan Mandokhel*, 24 April, 2020 (RAW101342).

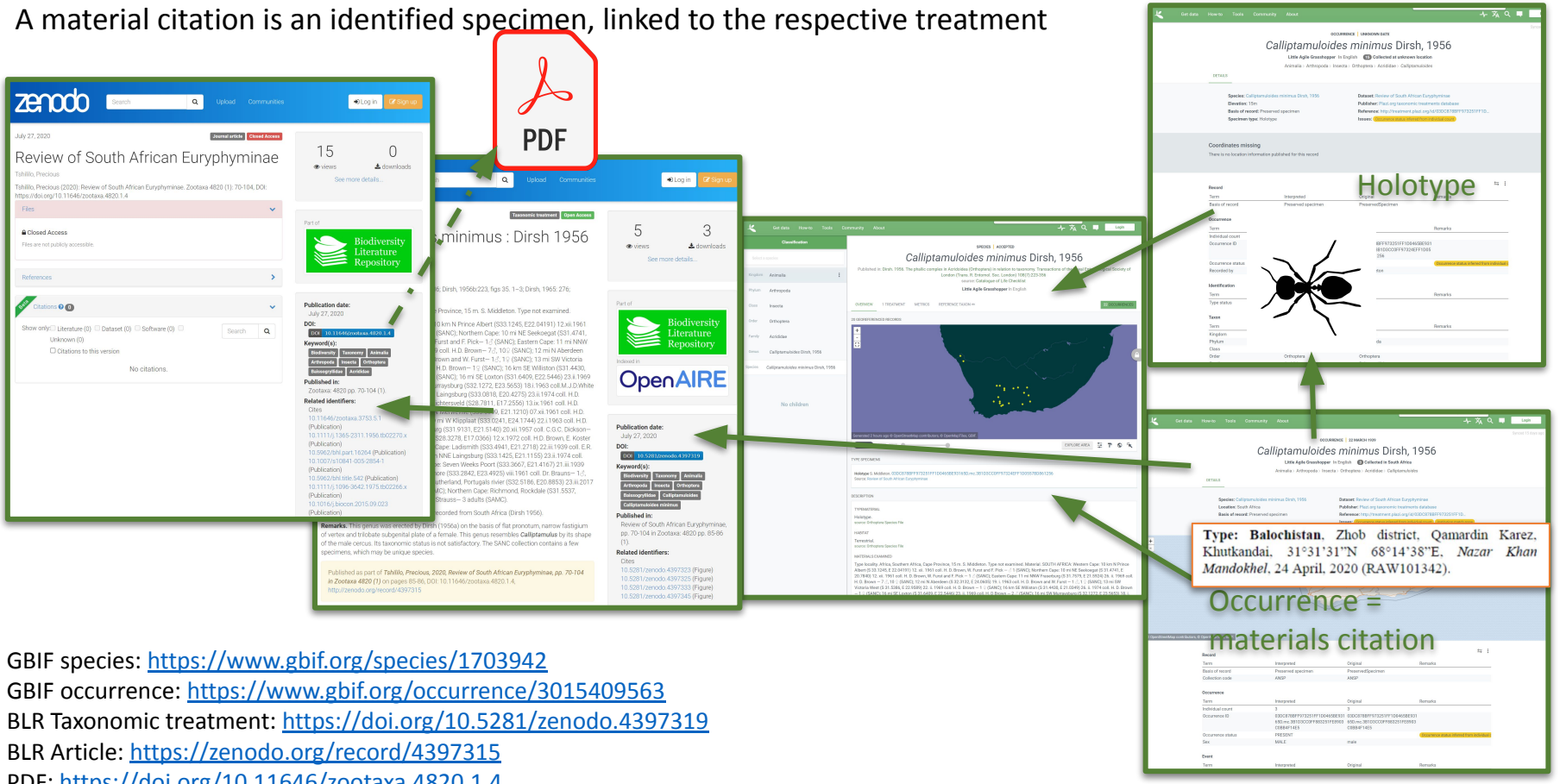
# Published materials citation links




Imagine the possible applications enabling making use of this big data?
Imagine the time saved if all these links are hyperlinks?

# Reuse of taxonomic data: GBIF

A material citation is an identified specimen, linked to the respective treatment




GBIF species: https://www.gbif.org/species/1703942
GBIF occurrence: https://www.gbif.org/occurrence/3015409563
BLR Taxonomic treatment: https://doi.org/10.5281/zenodo.4397319
BLR Article: https://zenodo.org/record/4397315
PDF: https://doi.org/10.11646/zootaxa.4820.1.4

# Details about a taxonomic name and link to justification

ChecklistBank

Eight new species and additional records of the Pselaphodes complex from Laos and Vietnam, with a key to

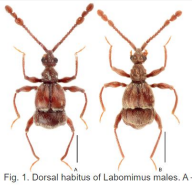
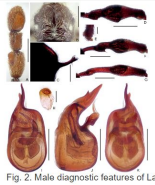
Login

Tools
COL: releases
Datasets
Dataset: 55892
About
Download
Browse
Search
Imports
Issues
References
Verbatim
Taxon: 03D13B6DE802513...

*Labomimus cavicornis* Yin & Li, 2021

Name details

About Verbatim

| | |
|---|---|
| Published in | Yin, Zi-Wei, Li, Ning (2021): Eight new species and additional records of the Pselaphodes complex from Laos and Vietnam, with a key to known species (Coleoptera: Staphylinidae: Pselaphinae). Acta Entomologica Musei Nationalis Pragae (Acta. Ent. Mus. Natl. Pragae) 61 (1): 35-53, DOI: 10.37520/aemnp.2021.002, URL: http://dx.doi.org/10.37520/aemnp.2021.002 |
| Type material | paratype PARATYPES: VIETNAM: Đ ÔNG NAI: 5 ♂♂, same label data as for holotype (MHNG, SNUC) Links: GBIF<br>holotype HOLOTYPE: VIETNAM: Đ ÔNG NAI: ♂, ' S-VIETNAM, Nam Cat Tien Nat. Park, 1. - 15.5.1994, Pacholatko & Dembicky' (MHNG) Links: GBIF |
| Status | accepted species |
| Classification | kingdom: Animalia > phylum: Arthropoda > class: Insecta > order: Coleoptera > family: Staphylinidae > genus: *Labomimus* > species: *Labomimus cavicornis* Yin & Li, 2021 |
| Media |  Fig. 1. Dorsal habitus of Labomimus males. A – L. cavicornis sp. nov.; B – L. cucphuong sp. nov. Scale bars: 1 mm.© Yin, Zi-Wei;Li, Ning https://doi.org/10.5281/zenodo.5037466<br>  Fig. 2. Male diagnostic features of Labomimus cavicornis sp. nov.A – left antennal club, in lateroventral view; B – pronotum; C – metaventral process, in lateral view; D – protrochanter and profemur; E – apex of protibia; F – mesotrochanter and mesofemur; G – metatrochanter and metafemur; H – sternite 7 (IX); I–K – aedeagus, in dorsal (I), lateral (J), and ventral (K) view. Scale bars: 0.3 mm in A, B, D, F, G; 0.2 mm in C, I, J, K; 0.1 mm in H; 0.05 mm in E.© Yin, Zi-Wei;Li, Ning https://doi.org/10.5281/zenodo.5037470 |
| Online resource | https://treatment.plazi.org/id/03D13B6DE802513E7F4BFA6A3BEDFB12 |
| Related names | |
| Origin | source |
| References | [1] Yin, Zi-Wei, Li, Ning (2021): Eight new species and additional records of the Pselaphodes complex from Laos and Vietnam, with a key to known species (Coleoptera: Staphylinidae: Pselaphinae). Acta Entomologica Musei Nationalis Pragae (Acta. Ent. Mus. Natl. Pragae) 61 (1): 35-53, DOI: 10.37520/aemnp.2021.002, URL: http://dx.doi.org/10.37520/aemnp.2021.002 |

https://www.checklistbank.org/dataset/55892/taxon/03D13B6DE802513E7F4BFA6A3BEDFB12.taxon

TreatmentBank

TREATMENTBANK BIODIVERSITY LITERATURE REPOSITORY SERVICES HOW TO PARTIC

## Labomimus cavicornis, Yin & Li, 2021

Yin, Zi-Wei & Li, Ning, 2021, Eight new species and additional records of the Pselaphodes complex from Laos and Vietnam, with a key to known species (Coleoptera: Staphylinidae: Pselaphinae), Acta Entomologica Musei Nationalis Pragae (Acta. Ent. Mus. Natl. Pragae) 61 (1), pp. 35-53 : 37

| | |
|---|---|
| publication ID | https://doi.org/ 10.37520/aemnp.2021.002 |
| publication LSID | lsid:zoobank.org:pub:0AEECD1F-1C71-4636-89A7-DA31D2DB73D6 |
| DOI | https://doi.org/10.5281/zenodo.5037454 |
| persistent identifier | https://treatment.plazi.org/id/03D13B6D-E802-513E-7F4B-FA6A3BEDFB12 |
| treatment provided by | Carolina (2021-06-28 18:05:09, last updated 2022-09-06 20:59:21) |
| scientific name | Labomimus cavicornis |
| status | sp. nov. |

Show all

### Treatment

*Labomimus cavicornis* sp. nov.

( Figs 1 A View Fig , 2 View Fig )

Type material (6 specimens). HOLOTYPE: VIETNAM: ĐÔNG NAI: ♂, 'S-VIETNAM, Nam Cat Tien Nat. Park , 1.- 15.5.1994, Pacholatko & Dembicky' ( MHNG) . PARATYPES: VIETNAM: ĐÔNG NAI: 5 ♂♂, same label data as for holotype ( MHNG, SNUC) .

Diagnosis. Male body length 3.09–3.23 mm. Antennomere 9 with broad depression in apical half, with disc-like impression inside depression; antennomeres 10 broadly depressed and antennomere 11 deeply concave on ventral side. Horn-like metaventral processes in lateral view elongate and bifurcate at apices. Protibia with short, blunt preapical projection; mesotrochanter with acute ventral spine. Median lobe of aedeagus asymmetric, narrowing apically; parameres broad at bases and narrowed in apical halves; endophallus with spine-like structures and one elongate sclerite. Female unknown.

Description. Male ( Fig. 1 A View Fig ). Body length 3.09–3.23 mm. Head slightly broader than long, HL 0.56–0.61 mm, HW 0.59–0.63 mm; eyes prominent, each composed of about 41 facets. Antennomeres 2–8 of similar shape, each moniliform; antennomeres 9–11 ( Fig. 2 A View Fig ) forming distinct club; antennomere 9 with large depression in apical half, and one disc-like impression inside depression; antennomeres 10 broadly depressed and antennomeres 11 with

https://tb.plazi.org/GgServer/html/03D13B6DE802513E7F4BFA6A3BEDFB12

# Treatment citation: Catalogue of life


Currently accepted name

Synonymized taxonomic name

Original name or new combination

https://synospecies.plazi.org/#Kiotina+spatulata

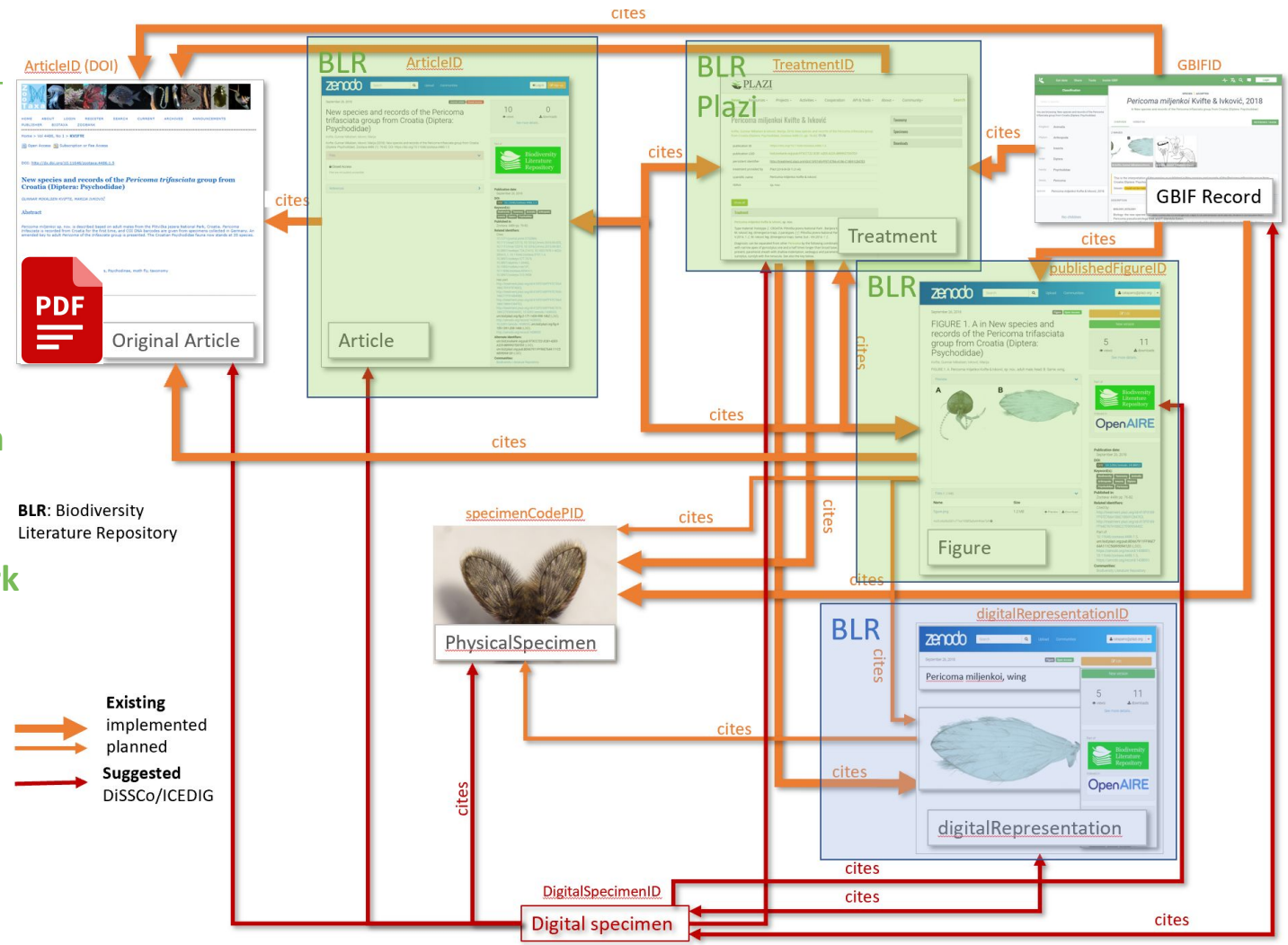
# **Open access to FAIR data**: A visual index to taxonomic publications

**Beyond PDF...**

https://ocellus.info/

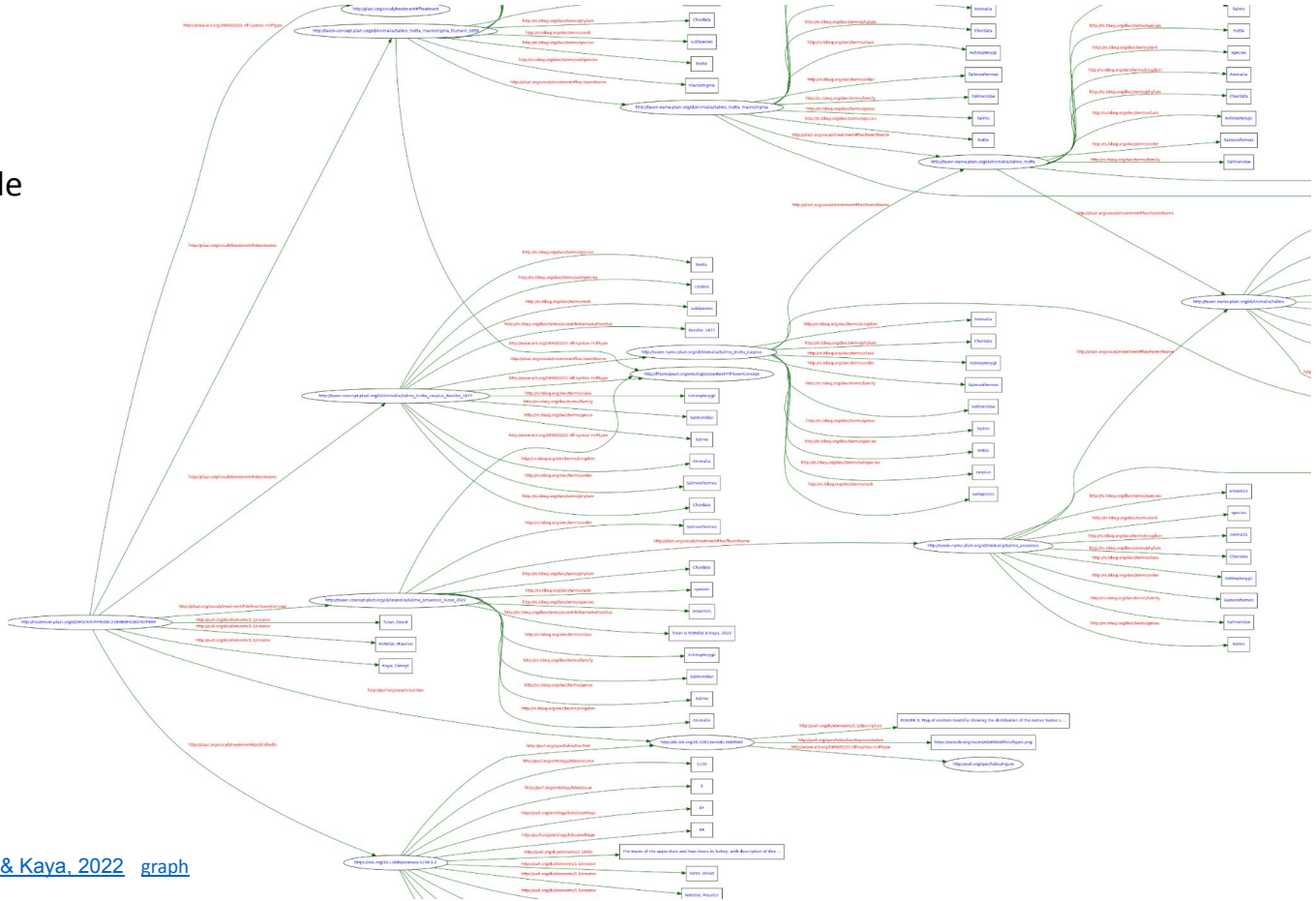
Focus on data in publications: What's inside?

**Citation network**



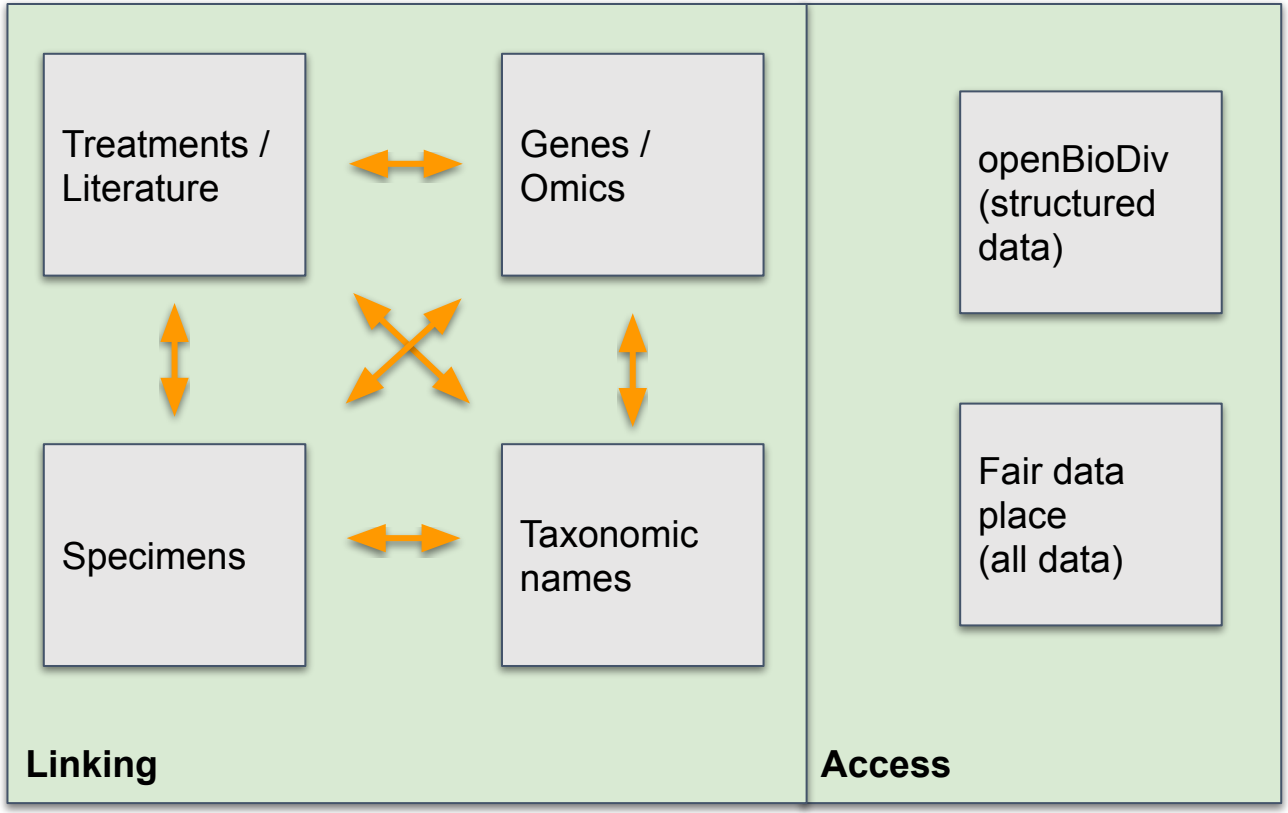
# Treatments as **FAIR digital objects** in the Biodiversity Literature Repository

Visualization of the knowledge graph of a single treatment (part)



Example: Salmo araxensis, Turan & Kottelat & Kaya, 2022 graph

# BiCIKL: Linking and access

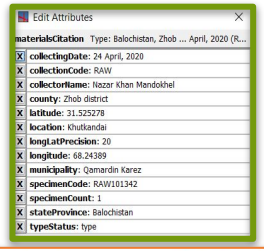


# Linking specimens and literature via material citations

**Bidirectional linking**

material citations

specimens / occurrences




**Type: Balochistan**, Zhob district, Qamardin Karez, Khutkandai, 31°31'31"N 68°14'38"E, *Nazar Khan Mandokhel*, 24 April, 2020 (RAW101342).

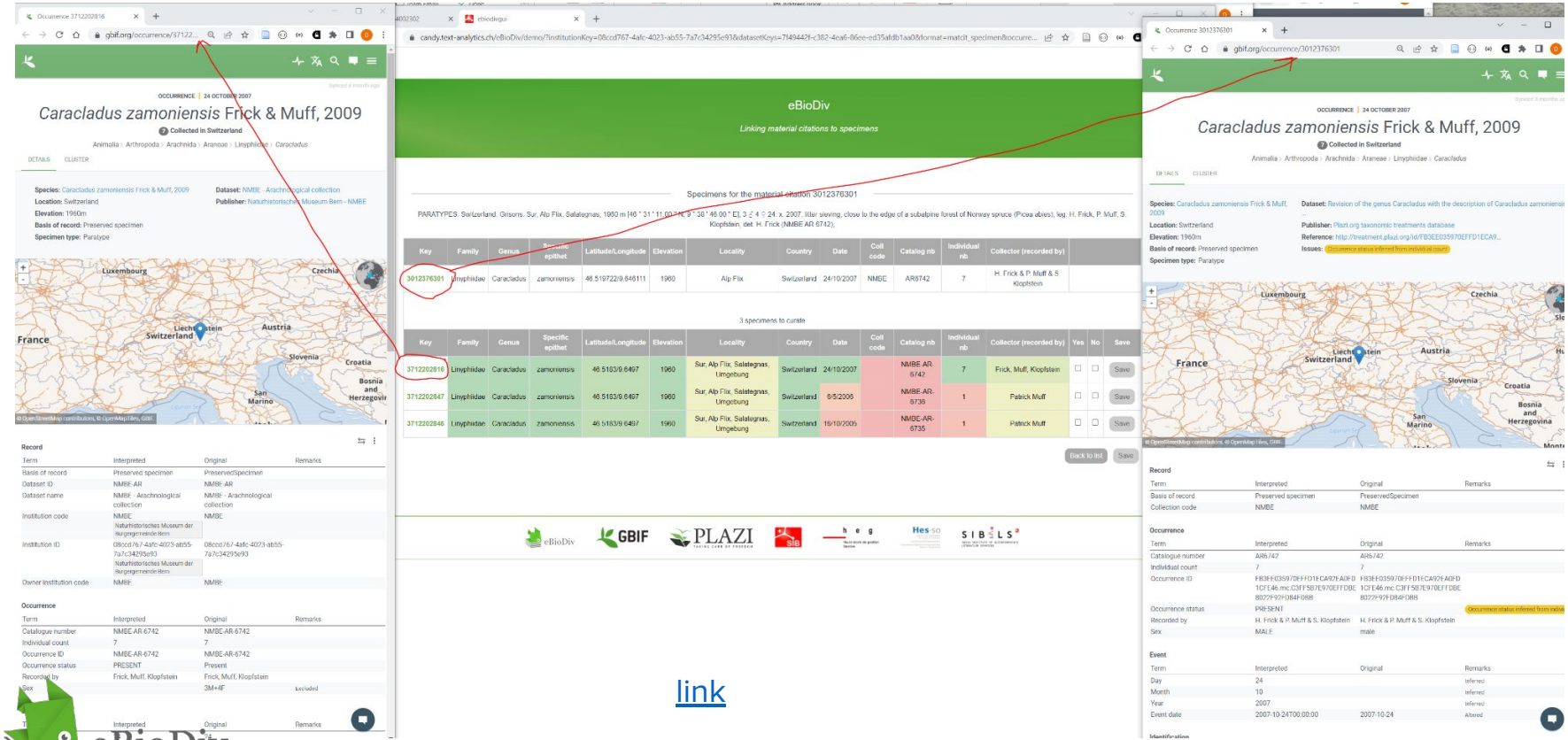
eBioDiv

Pilot matching service (link)

# Linking specimens with material citations via GBIF

Preserved specimen

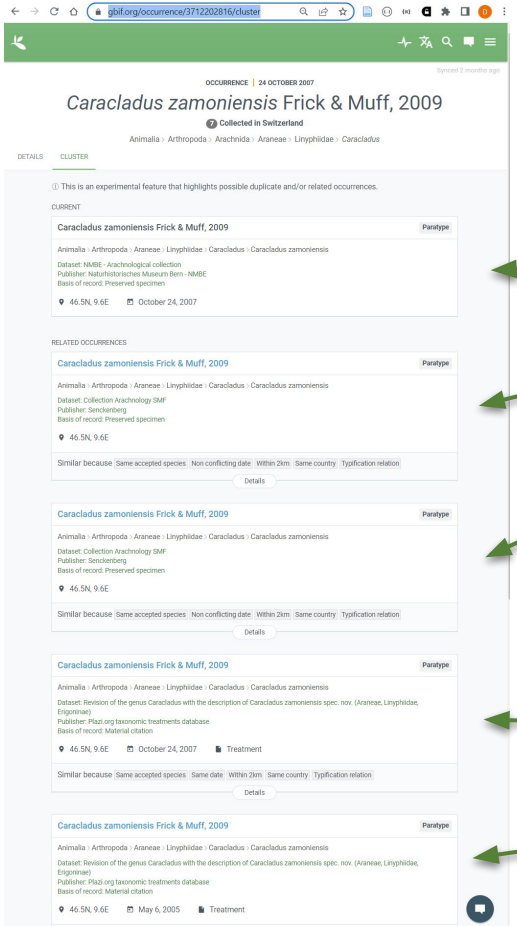
Material citation




link

eBioDiv

# Linking specimens with material citations via GBIF; Clusters




Preserved specimen

Material citation

# Plazi workflow


**Get publications:**
- PDF (library access)
- Zenodo via microservice

**Receive publications:**
- PDF
  - Born digital
  - Scanned
- XML
  - JATS
  - TaxPub/JATS
- HTML / XHTML

PDF

PENSOFT

**Decode**
- manual
- automated

**Enhance**
- semantics
- link

**Data Quality Control**
- Criteria
- Corrections

**Create**
- Open FAIR data

FAIR Digital Objects

TreatmentBank

**Access / Metrics**
- User Interface
- API
- Dashboards

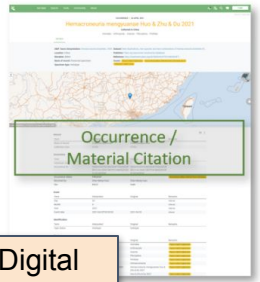
SynoSpecies
Taxonomic names
Treatments

**Disseminate**
- GBIF
- NCBI

Reuse

GBIF

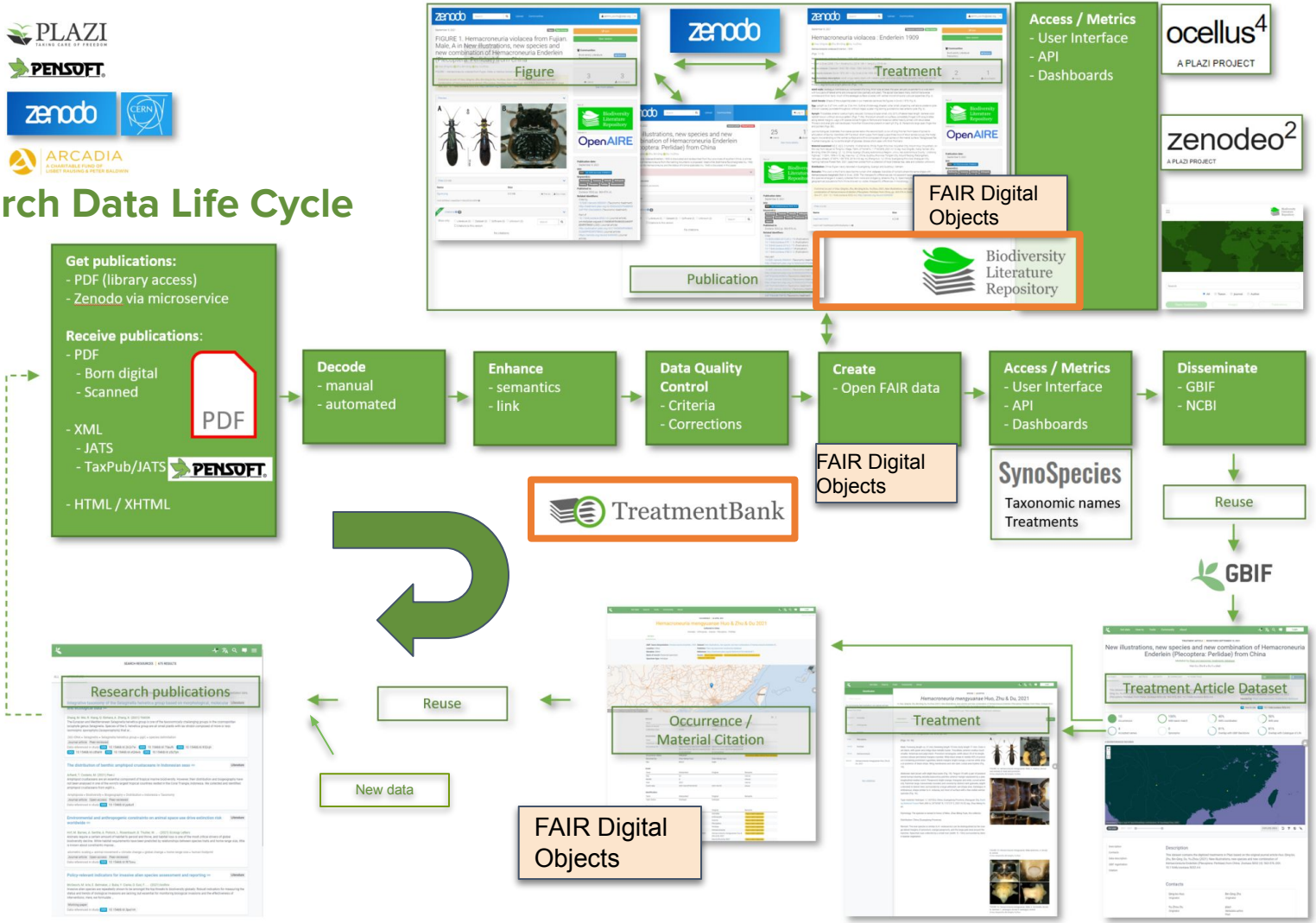




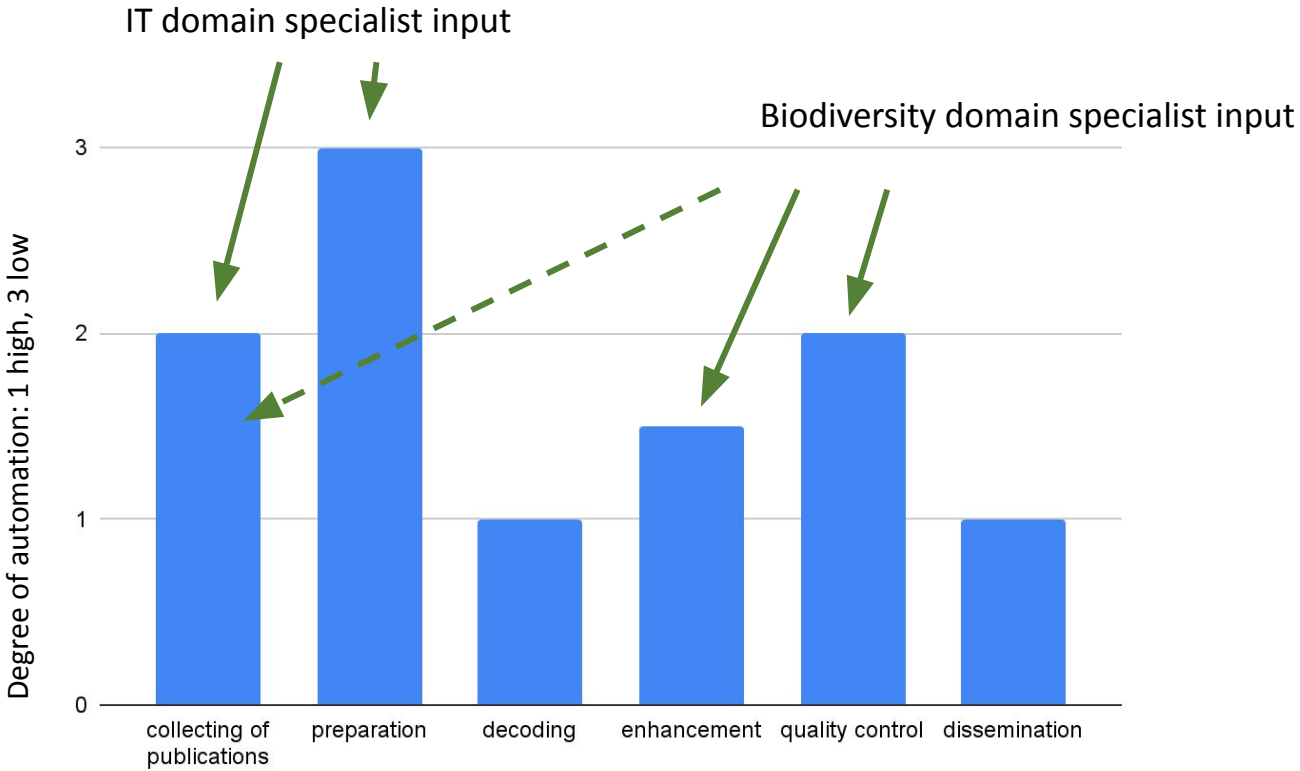

FAIR Digital Objects

Reuse

# Research Data Life Cycle



# Degree of automation / human input




This is a very expensive effort, has to be shared by partners avoiding duplications as much as possible, and needs be avoided by changing the way we publish.

# Infrastructures maintained by Plazi

**TreatmentBank**

Data conversion and access service
73,000 articles
762,000 taxonomic treatments
1,118,000 materials citations
>50% of annually described new species
production in 2021: 25,000 articles, 224,000 treatments, 180,000 images

**Biodiversity Literature Repository**

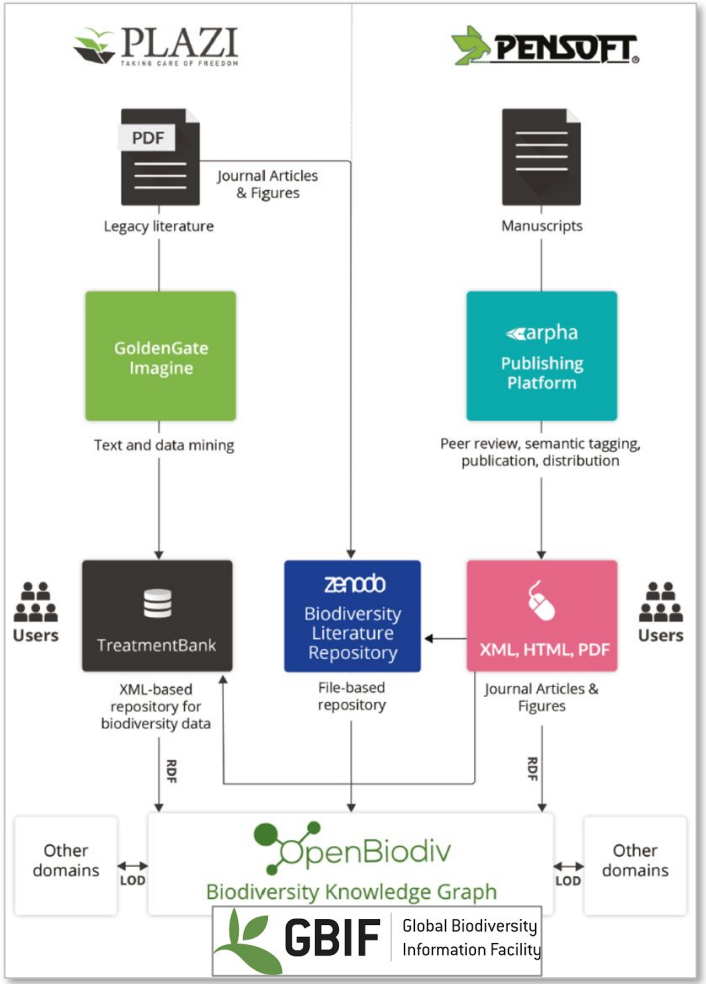
Repository for data liberated from publications
453,000 images
72,000 articles
400,000 taxonomic treatments
Collaboration with Zenodo /CERN
Recognized as EU research infrastructure
Mints DOI for treatments and figures

**Global Biodiversity Information Facility**

Reuse of treatment articles mediated by Plazi
41,000 treatment article data sets (54% of total data sets in GBIF)
377,000 taxonomic treatments (90,000 unique species)
224,000 figures
632,000 materials citations (occurrences)
The figures are lower in GBIF because of Quality Control measures
800 publications site data from TreatmentBank mediated literature

# The future will be open, semantically enhanced publishing

Legacy publications




Prospective publishing

# Thank you!

Questions, answers, participation https://github.com/plazi/community

Introduction to digitizing taxonomic literature with Plazi DOI

Donat Agosti

agosti@plazi.org

# Linking specimens and literature via material citations

**Further reading:**

- Plazi: further reading
- Pensoft: doi: 10.3897/zookeys.50.543 (e.g. Zookeys, BDJ)
- CETAF: doi: 10.5252/adansonia2018v40a1 (e.g. European Journal of Taxonomy)

**Data usage:**

- Rivera-Quiroz et al. 2020, doi: 10.1038/s41598-020-72549-8
- Dikow & Agosti, 2015, doi: 10.3897/BDJ.3.e5707

**Data access:**

- Brief introduction into Treatmentbank stats: PDF
- Treatment statistics: https://tb.plazi.org/GgServer/srsStats
- Article statistics: https://tb.plazi.org/GgServer/dioStats
- Biodiversity Literature Repository API introduction: https://developers.zenodo.org/
- Biodiversity Literature Repository: https://zenodo.org/communities/biosyslit/search?q=

**Applications based on and reuse of TreatmentBank and BLR data:**

- Images via Ocellus: https://ocellus.info/
- Treatment citations via Synospecies: https://synospecies.plazi.org/
- TreatmentBank data in GBIF: https://www.gbif.org/publisher/7ce8aef0-9e92-11dc-8738-b8a03c50a862

# Plazi

Plazi is a Swiss based international association supporting and promoting the development of persistent and openly accessible scholarly digital taxonomic publications

NGO, SME owned by the NGO; Founded in 2008 as spin-off from a former US/DFG binational digital library award (2003-06); Supported by service contracts, EU-research funding, philanthropic funds, voluntary contributions. Plazi GmbH SME as service provider.

13 persons working for Plazi in Brazil, France, Germany, Spain, Switzerland, USA

Collaborations with Global Biodiversity Information Facility (GBIF), Zenodo at CERN, Pensoft Publishers Ltd, Consortium of European Taxonomic Facilities (CETAF), Swiss Institute of Bioinformatics (SIB), National Center for Biotechnology Informatics (NCBI), Muséum nationale d'Histoire Naturelle, Paris, Data Futures.

Support from Arcadia Fund, EU projects, service contracts, and voluntary contributions.

A mission of Plazi is to **discover, make accessible, and disseminate known biodiversity data,** not publications *per se* and to promote semantic enhanced publishing (TaxPub/JATS).