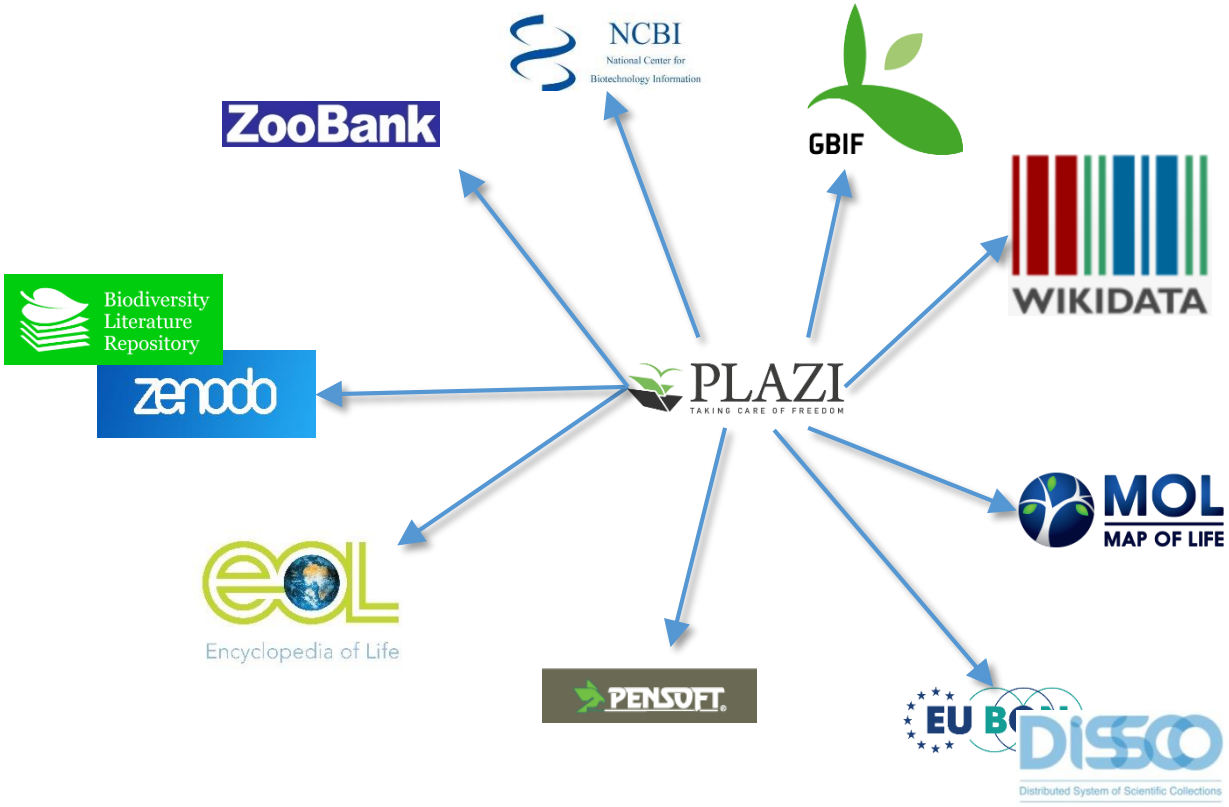
Plazi – on expedition to discover known biodiversity

Donat Agosti
Plazi
agosti@plazi.org; @myrmoteras
ORCID: 0000-0001-9286-1200

August 10, 2018, New York Botanical Garden, New York

DOI: 10.5281/zenodo.1343376

A partner in a global science network



Plazi promotes open science, is innovative, and complementary to existing activities.

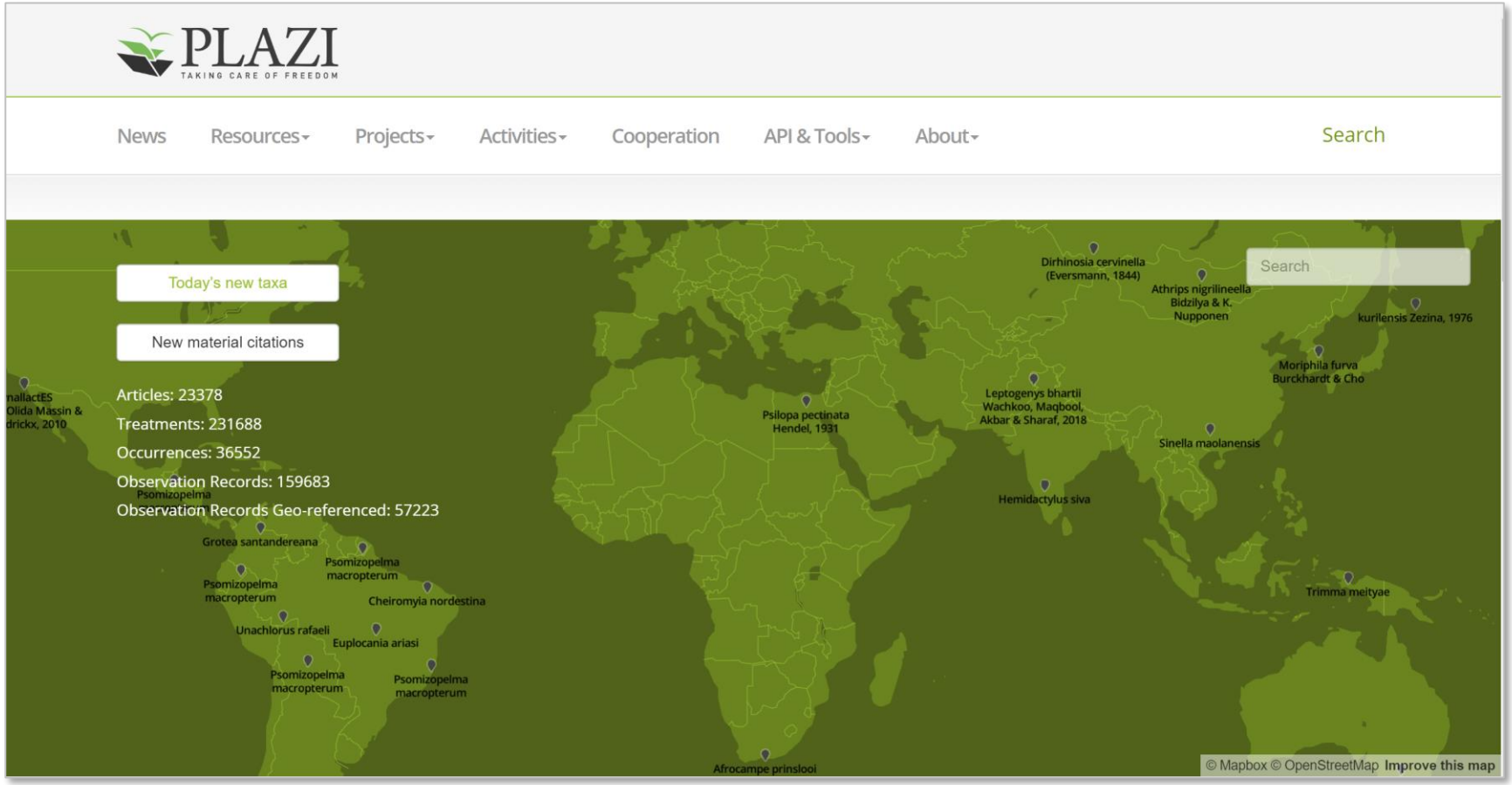
Plazi: Dissemination

Reuse of **subarticle** parts

Plazi text and data mining output 2017

Journals	132
Articles	8,545
Taxonomic treatments	69,539
New species	6,387
Observation records	45,381
Figures	56,692
Bibliographic references	267,728
Facts	>>10 M

Plazi: TreatmentBank



23,378 articles processed
231,688 taxonomic treatments
159,583 observation records
semi- to fully automated import treatments

Biolitrepo.org

All based on published results

DOI for all deposits

All deposits linked to source

- 34,238 articles
- 184,645 images

Collaboration with CERN and Pensoft

Source: BLR https://zenodo.org/communities/biosyslit/search?page=1&size=200&keywords=hymenoptera&q=aedeagus&file_type=png

JATS TaxPub

Journal **A**rticle **T**ag **S**uit, **Tax**onnomic **Pub**lication extension

Developed for and used to semantically enhance taxonomic literature

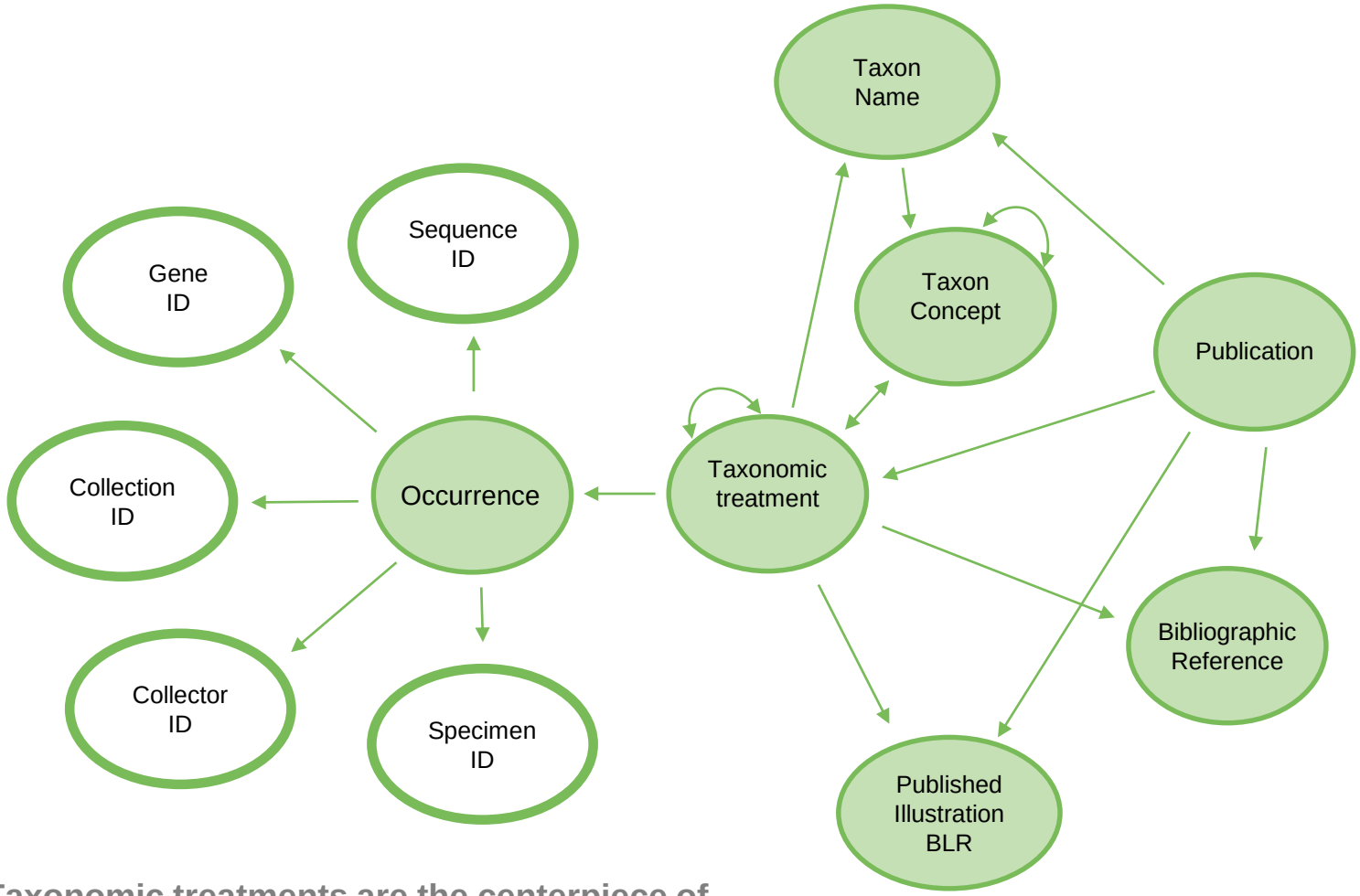
In collaboration with the National Library of Medicine
Semantic elements added for taxonomy
- if terms are in TWDG reused
- if not available, creation of complementing vocabulary

JATS is used to import scholarly articles into Pub Med Central
-> make it part of the world largest corpus of biomedical literature

TaxPub is used by Pensoft publishers and European Journal of Taxonomy

(Supported by US NSF and German DFG binational digital library award, 2004-7)

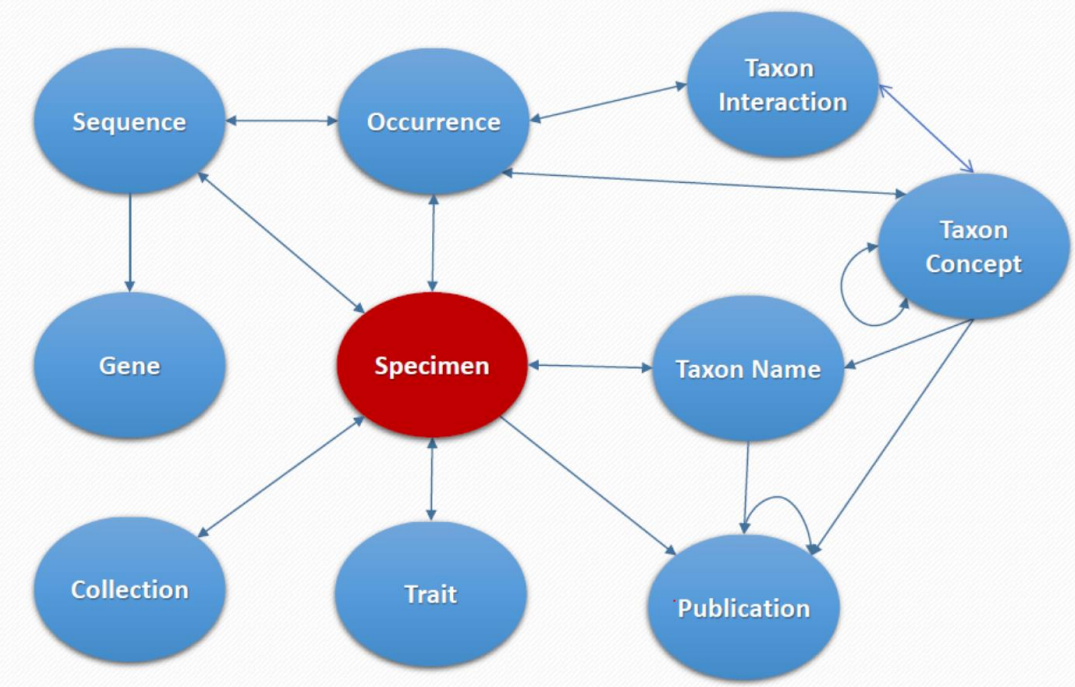
Ongoing digitization projects: Plazi structure



Taxonomic treatments are the centerpiece of the TreatmentBank (Plazi) and the Open Biodiversity Knowledge Management System (OBKMS)

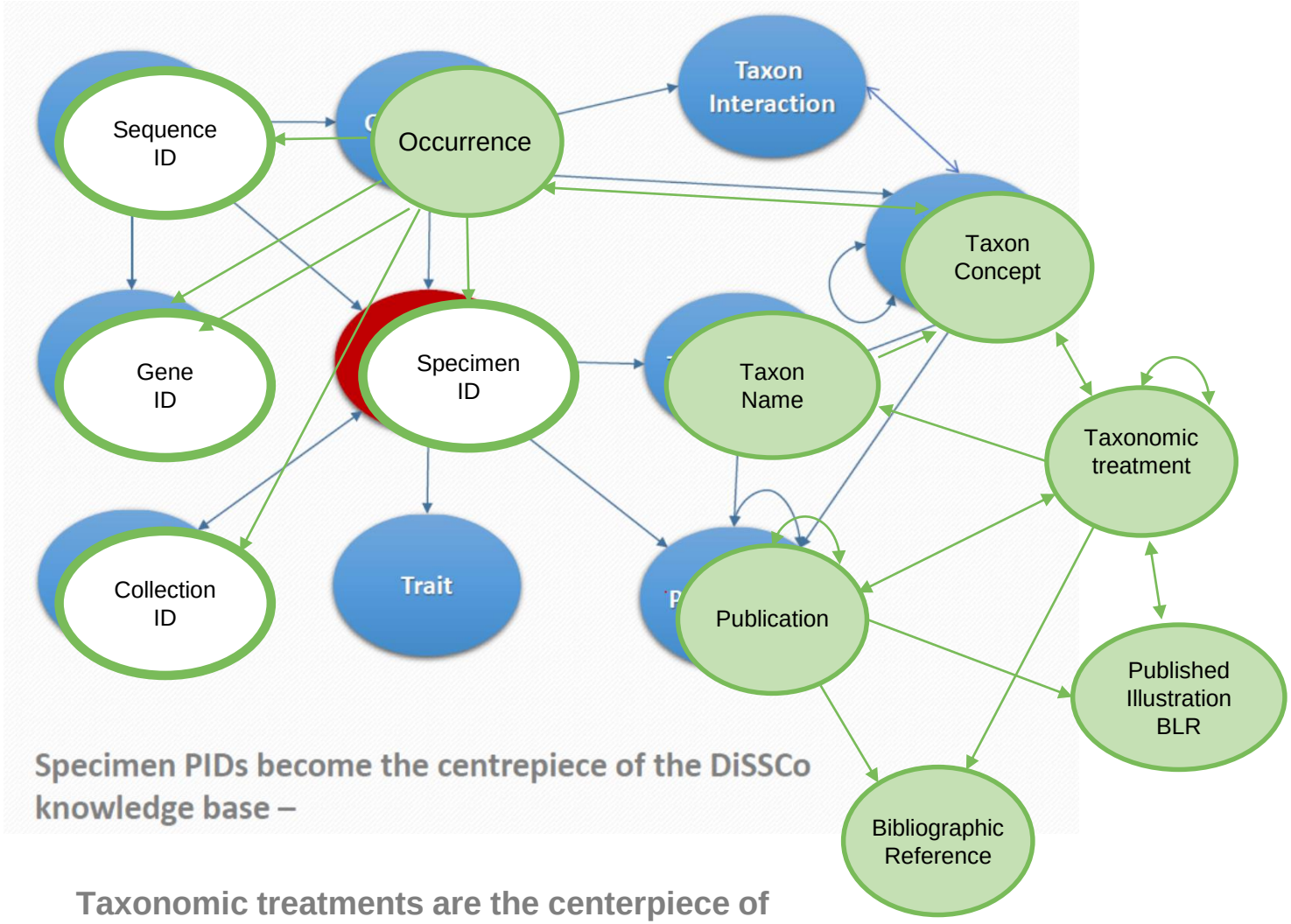
Ongoing digitization projects: DiSSCo overview

Distributed **S**ystem of **S**cientific **Co**llections
DiSSCo is a new pan-European Research Infrastructure initiative of 21 European countries with a vision to position European natural science collections at the centre of data-driven scientific excellence and innovation in environmental research, climate change, food security, one health and the bioeconomy.



Specimen PIDs become the centrepiece of the DiSSCo knowledge base –

Ongoing digitization projects: DiSSCo and Plazi overview



Taxonomic treatments are the centerpiece of Plazi's TreatmentBank and names of the Open Biodiversity Knowledge Management System (OBKMS)

Arcadia Fund award

Promote the *Biodiversity Literature Repository* as a way to liberate data from scholarly publications, especially b by building and index to publications through illustrations

Provide taxonomic treatments and related data for >50% of new described species annually

Make these data accessible through collaborations, promote TaxPub based publishing and Text and Data Mining of articles from highly relevant journals, including back issues

3 years (2018-21), 1.1M Euro

Plazi's uniqueness

Plazi is

- competent, activist, innovator, catalyst and leader in biodiversity informatics
- complementing ongoing efforts
- collaborator in the global biodiversity landscape
- building science infrastructure for science
- is independent, passionate and a "long" track record
- operating in Switzerland with copyright more favorable for text and data mining

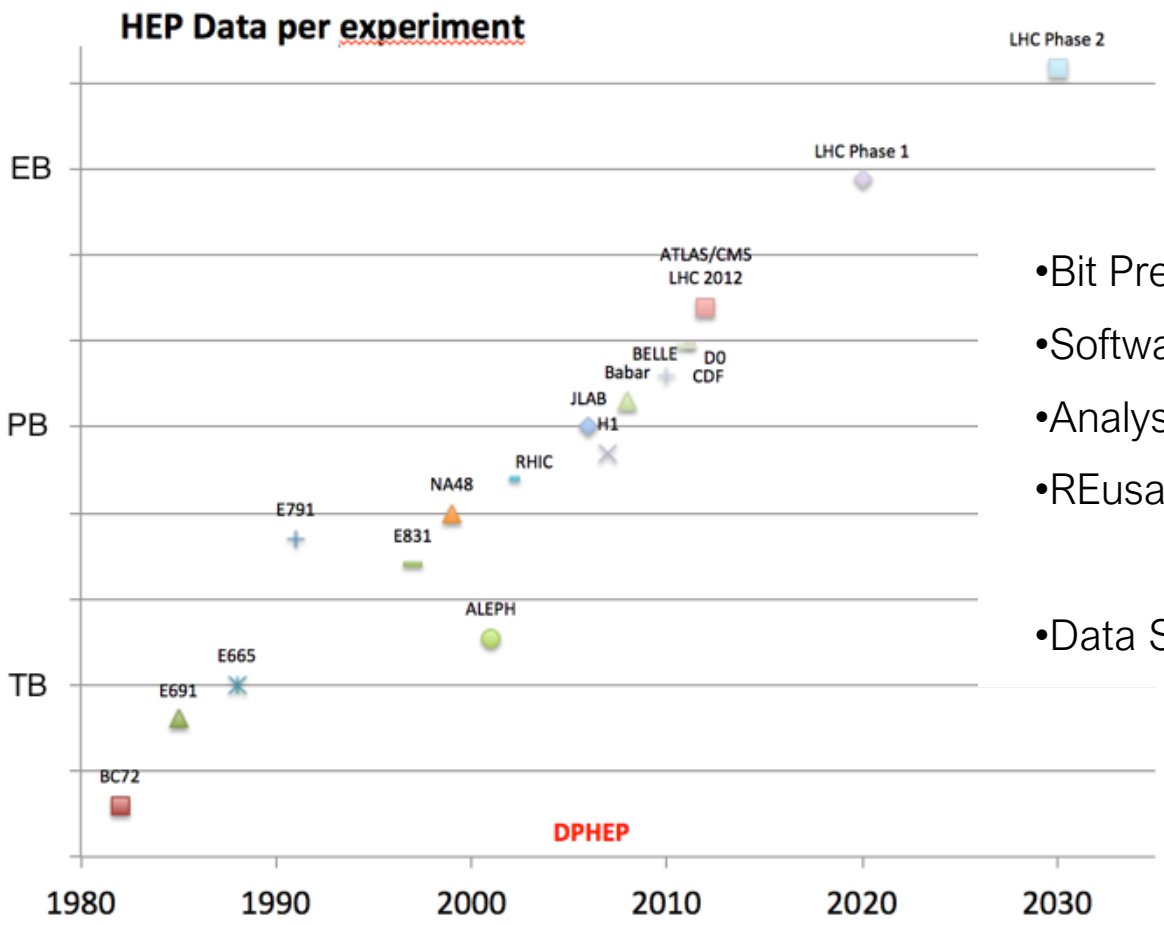
Plazi: Introduction

NEWS | 6 JULY 2018

Big data for biodiversity: GBIF.org surpasses 1 billion species occurrences

Milestone represents a collective effort to share evidence of our planet's biodiversity by more than 1,200 institutions in 123 countries—and more than 1 million researchers and citizen scientists worldwide

Preservation: Driven by Science



- Bit Preservation: ISO 16363 standard
- Software preservation
- Analysis Preservation
- REusable ANAlyses

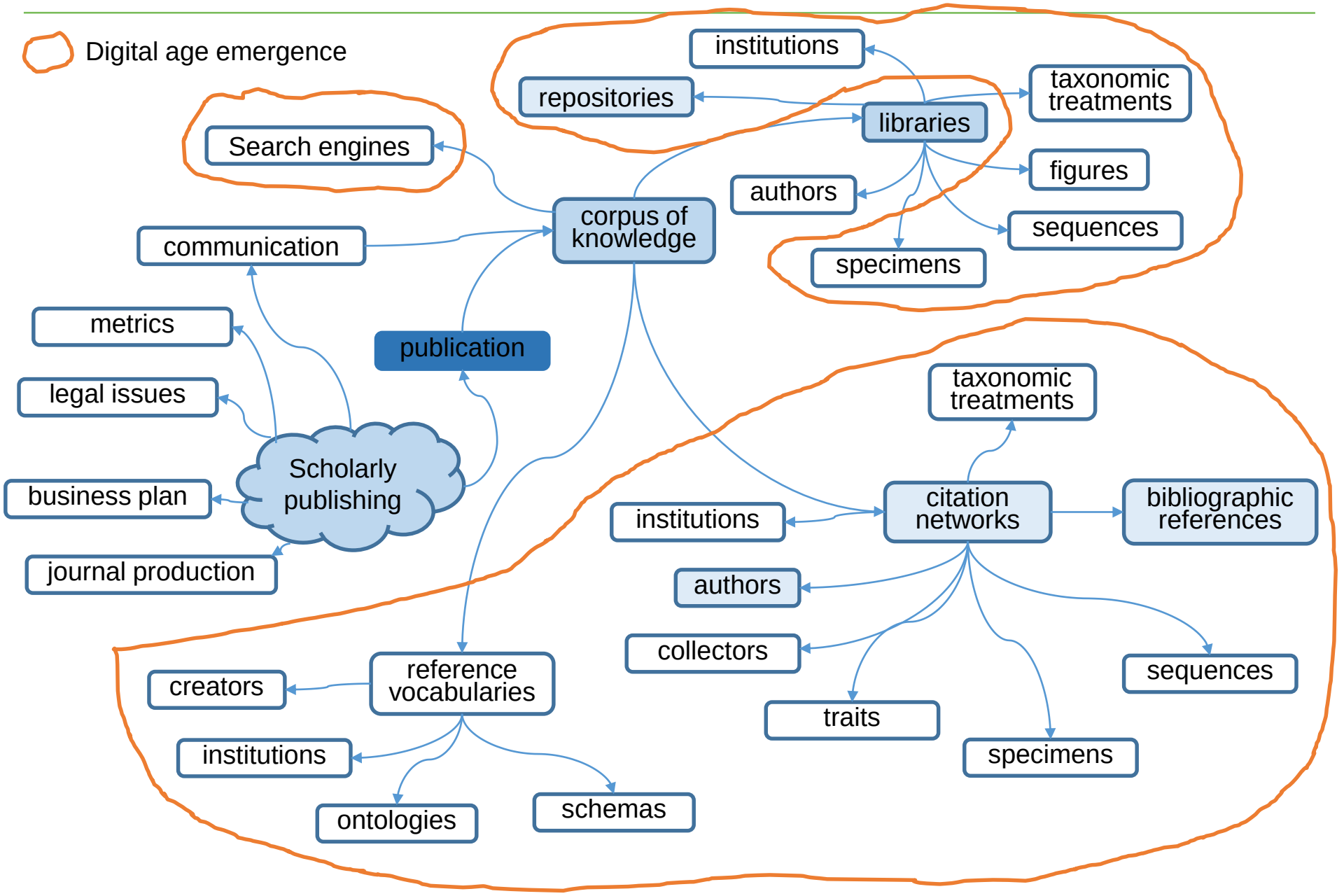
- Data Seal of Approval

HEP: Hight Energy Physics, DPHEP: Digital Preservation in Hight Energy Physics,

TB = Terabyte, PB = Petabyte, EB = Exabyte

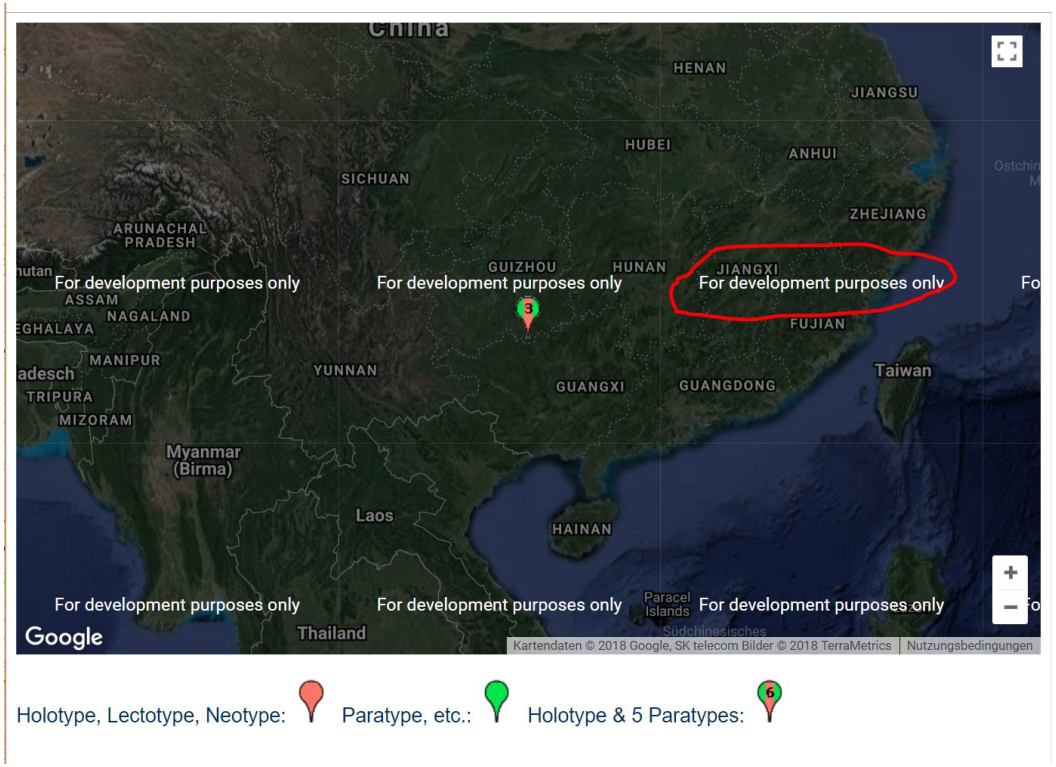
(Tim Smith)

Scholarly publishing: a daunting task



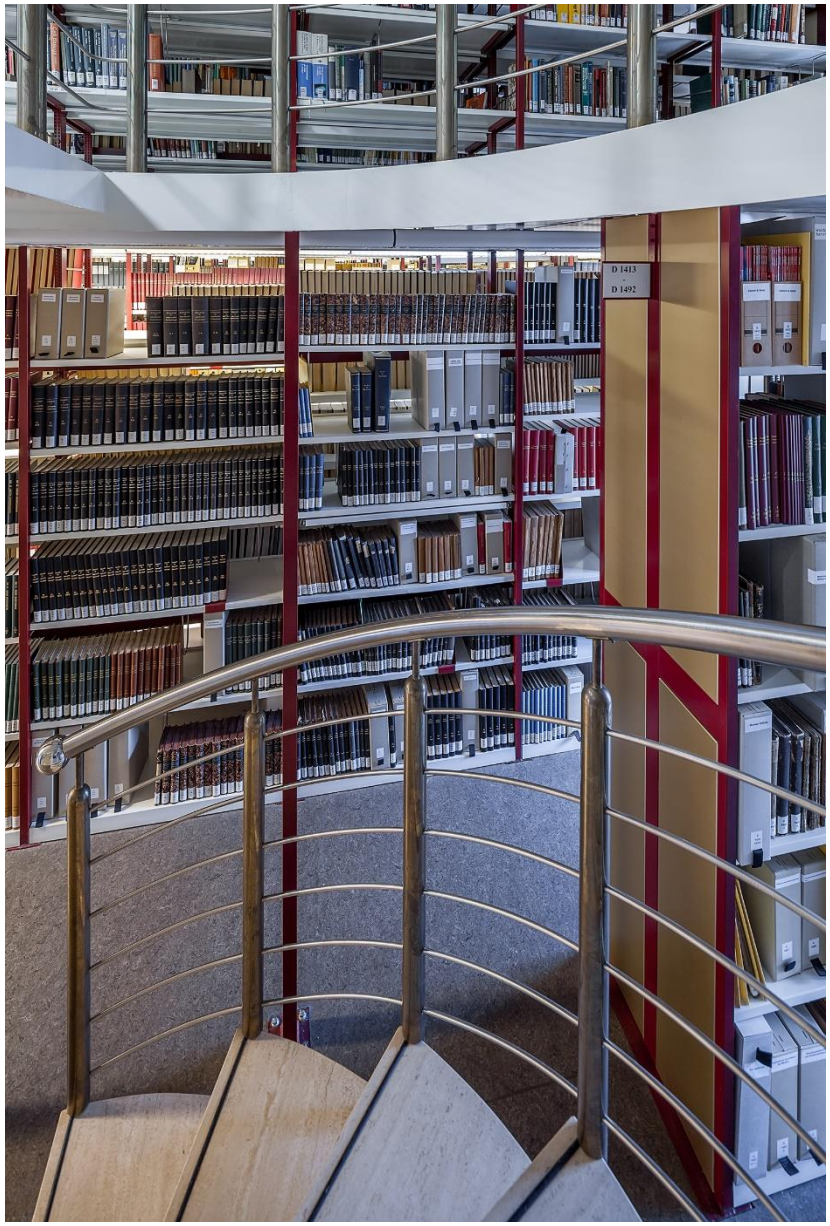
Knowledge management and infrastructure

We are in a global race to control and commercialize the management of global knowledge



The scientific community needs to make an effort to keep the baseline data of its own knowledge open and accessible for everybody

The challenge: Taxonomy as an empiric science



500,000,000+ printed pages
1,900,000 species described
20,000,000+ species treatments
18,000 new species discovered / year

BUT: Data are hidden

- Incomplete digitization
- Publications are not semantically enhanced
- Data are not linked
- Most data are not open

How can we query our accumulated biodiversity knowledge?

The PDF prison

Highly **structured** research data

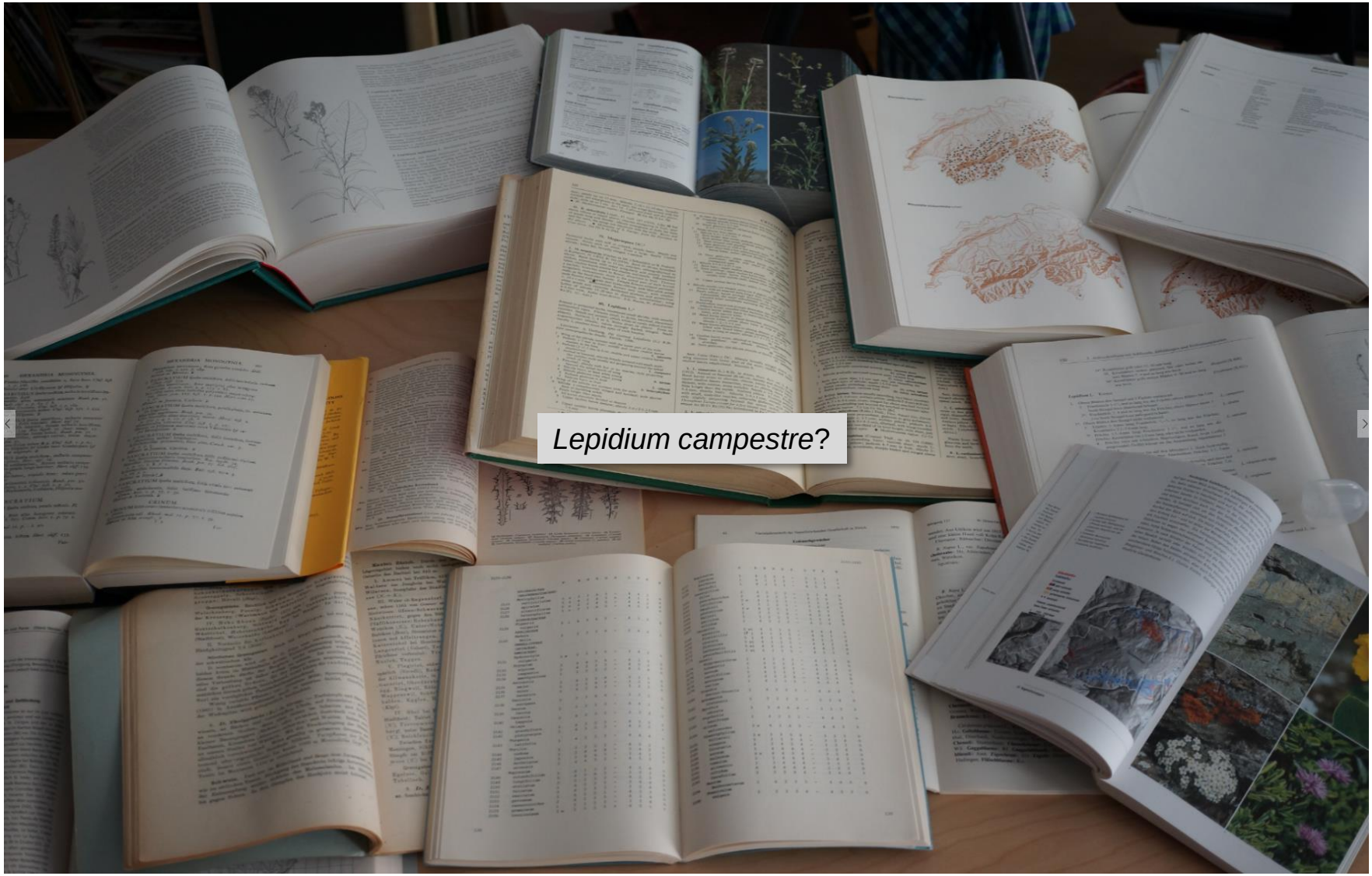
publishing

Semantically **unstructured** text

data mining

Semantically enhanced, **structured** data

Biodiversity Knowledge: 20 years ago, and still today



Social and behavioral constraints

TreatmentBank: Visualization of data from one article

PLAZI
TAKING CARE OF FREEDOM

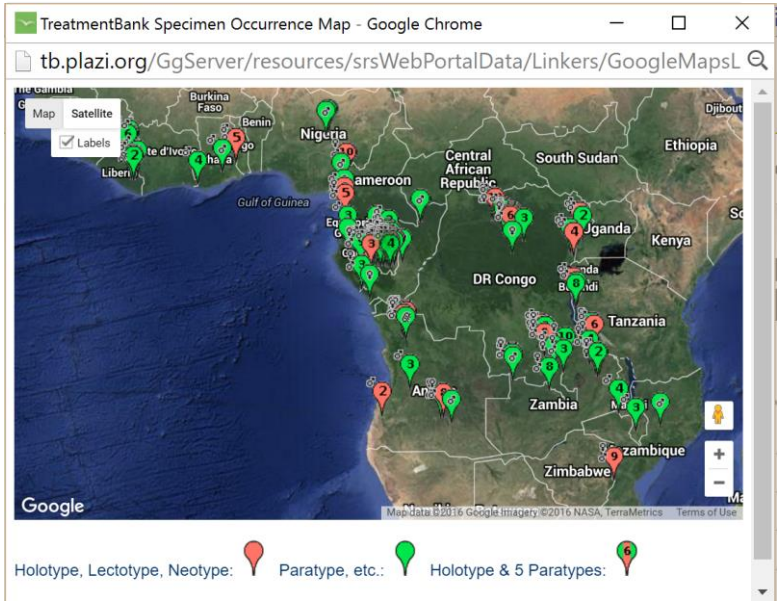
News | Resources | Projects | Activities | Cooperation | Developers | About | Search

Dijkstra, Klaas-Douwe B., Kipping, Jens & Mézière, Nicolas, 2015, Sixty new dragonfly and damselfly species from Africa (Odonata)

Dijkstra, Klaas-Douwe B., Kipping, Jens & Mézière, Nicolas, 2015, Sixty new dragonfly and damselfly species from Africa (Odonata), Odonatologica 44 (4), pp. 447-678

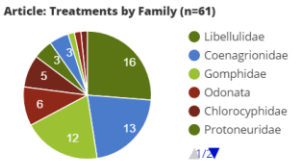
publication ID	http://dx.doi.org/10.5281/zenodo.35388
link to original document	http://dx.doi.org/10.5281/zenodo.35388
document provided by	Donat (2015-12-11 15:22:31, last updated by Admin 2015-12-18 12:58:08)

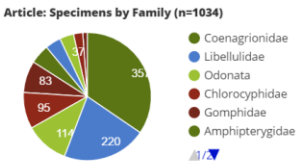
Treatments (61)

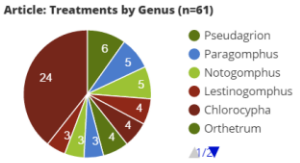


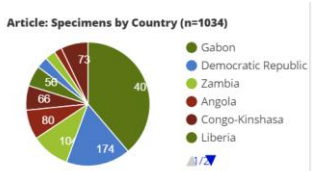
Distribution Map

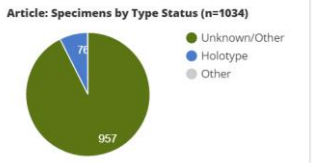
Specimens

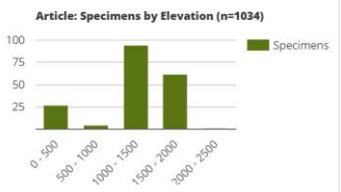






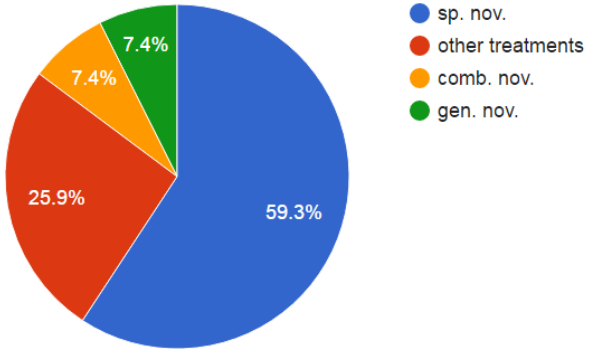




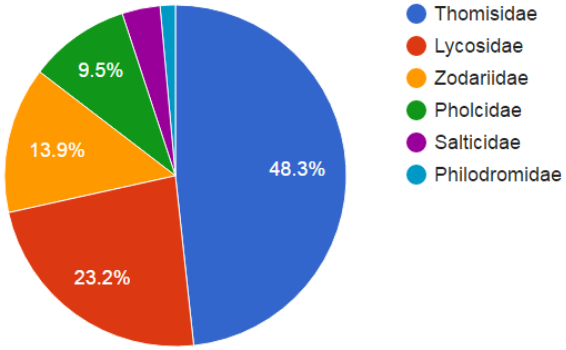


Plazi dashboard
Selected articles published in the European Journal of Taxonomy

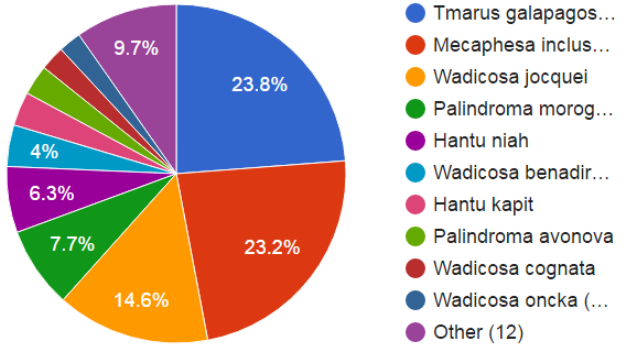
2. Treatments by taxonomic status (n=27)



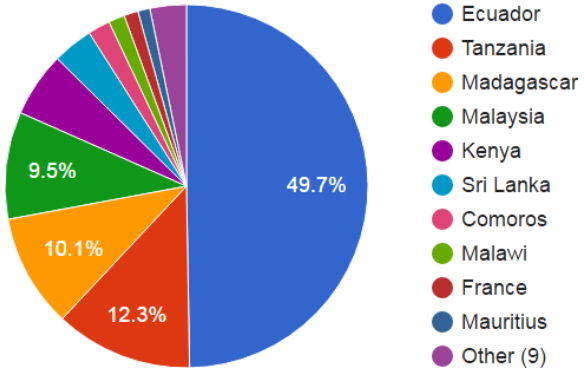
12. Specimens by family (n=555)



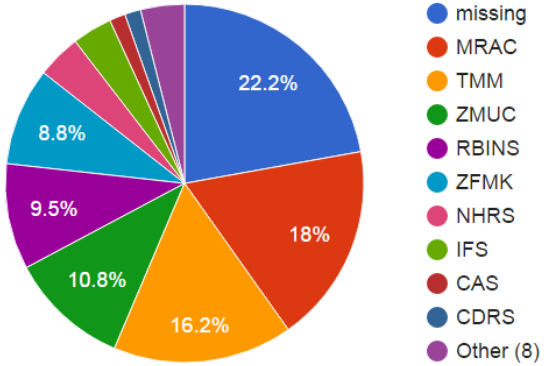
13. Specimens by species (n=555)



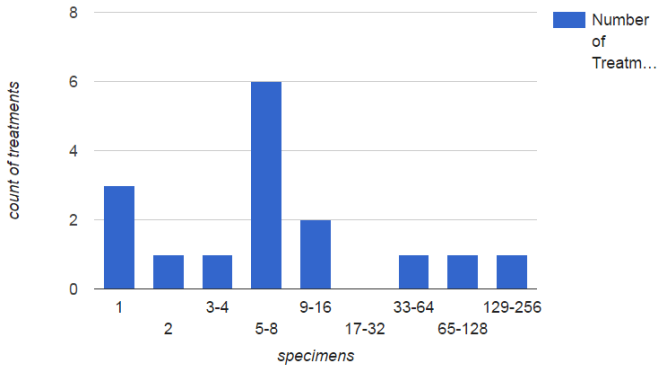
16. Specimens by collecting country (n=555)



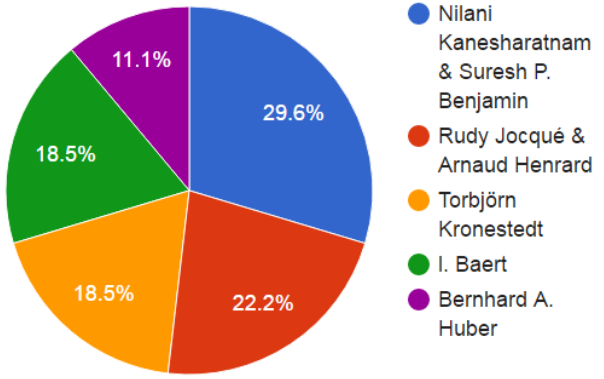
18. Specimens by collection code (n=555)



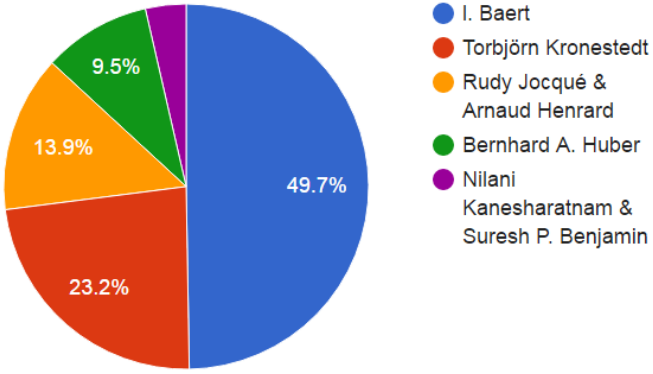
36. New species descriptions: specimens per treatment (n=16)



27. Treatments by author (n=27)



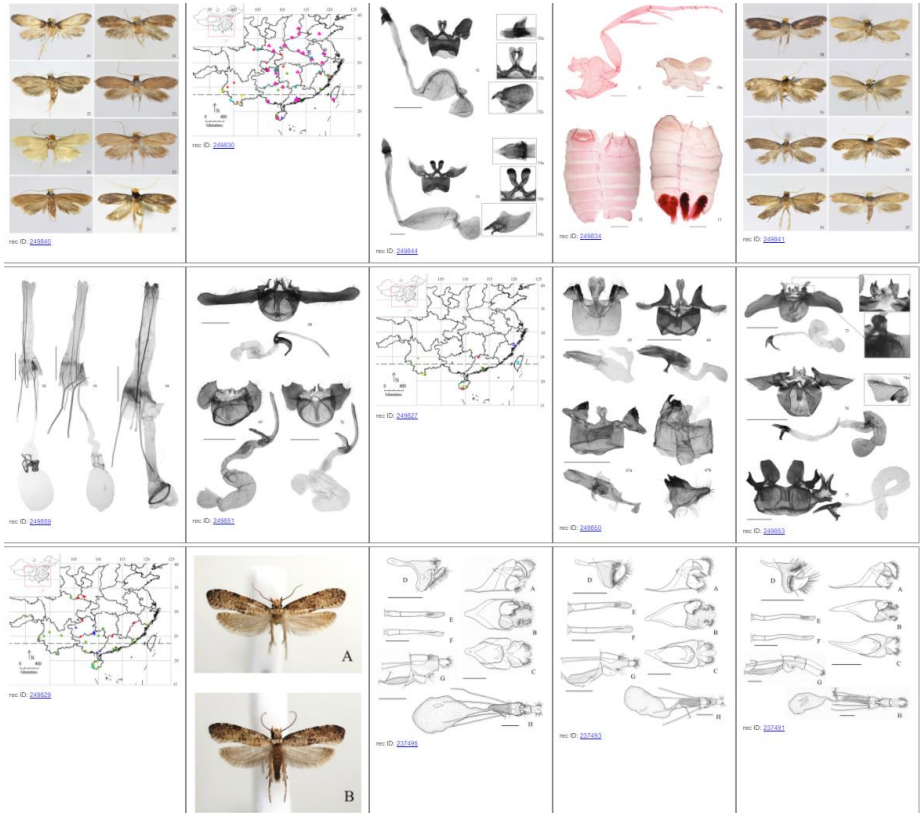
28. Specimens by author (n=555)

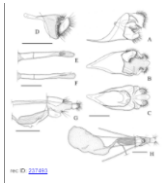


Plazi: production

The value of access to subarticle elements: figures

the value of a figure





access

http://ocellus.punkish.org/?q=tineidae&page=1

Use of published data

Biodiversity Data Journal 3: e5707
doi: 10.3897/BDJ.3.e5707

Taxonomic Paper

Utilizing online resources for taxonomy: a cybercatalog of Afrotropical apiocerid flies (Insecta: Diptera: Apioceridae)

Torsten Dikow‡, Donat Agosti§

‡ National Museum of Natural History, Smithsonian Institution, Washington, DC, United States of America
§ www.plazi.org, Bern, Switzerland

Corresponding author: Torsten Dikow (dikowt@si.edu)

Academic editor: Daniel Whitmore

Received: 21 Jul 2015 | Accepted: 30 Sep 2015 | Published: 06 Oct 2015

Citation: Dikow T, Agosti D (2015) Utilizing online resources for taxonomy: a cybercatalog of Afrotropical apiocerid flies (Insecta: Diptera: Apioceridae). Biodiversity Data Journal 3: e5707. doi: 10.3897/BDJ.3.e5707

Abstract

A cybercatalog to the Apioceridae (apiocerid flies) of the Afrotropical Region is provided. Each taxon entry includes links to open-access, online repositories such as ZooBank, BHL/BioStor/BLR, Plazi, GBIF, Morphbank, EoL, and a research web-site to access taxonomic information, digitized literature, morphological descriptions, specimen occurrence data, and images. Cybercatalogs as the one presented here will need to become the future of taxonomic catalogs taking advantage of the growing number of online repositories, linked data, and be easily updatable. Comments on the deposition of the holotype of *Apiocera braunsi* Melander, 1907 are made.

Keywords

cybertaxonomy, open-access, online repositories

Biodiversity Data Journal 3: e5063
doi: 10.3897/BDJ.3.e5063

General Article

Integrating and visualizing primary data from prospective and legacy taxonomic literature

Jeremy A. Miller‡,§, Donat Agosti§, Lyubomir Penev|, Guido Sautter¶, Teodor Georgiev#, Terry Catapano§, David Patterson¤, David King«, Serrano Pereira‡, Rutger Aldo Vos‡, Soraya Sierra‡

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ www.Plazi.org, Bern, Switzerland
| Pensoft, Sofia, Bulgaria
¶ KIT / Plazi, Karlsruhe, Germany
# Pensoft Publishers, Sofia, Bulgaria
¤ University of Sydney, Sydney, Australia
« The Open University, Milton Keynes, United Kingdom

Corresponding author: Jeremy A. Miller (jeremy.miller@naturalis.nl)

Academic editor: Ross Mounce

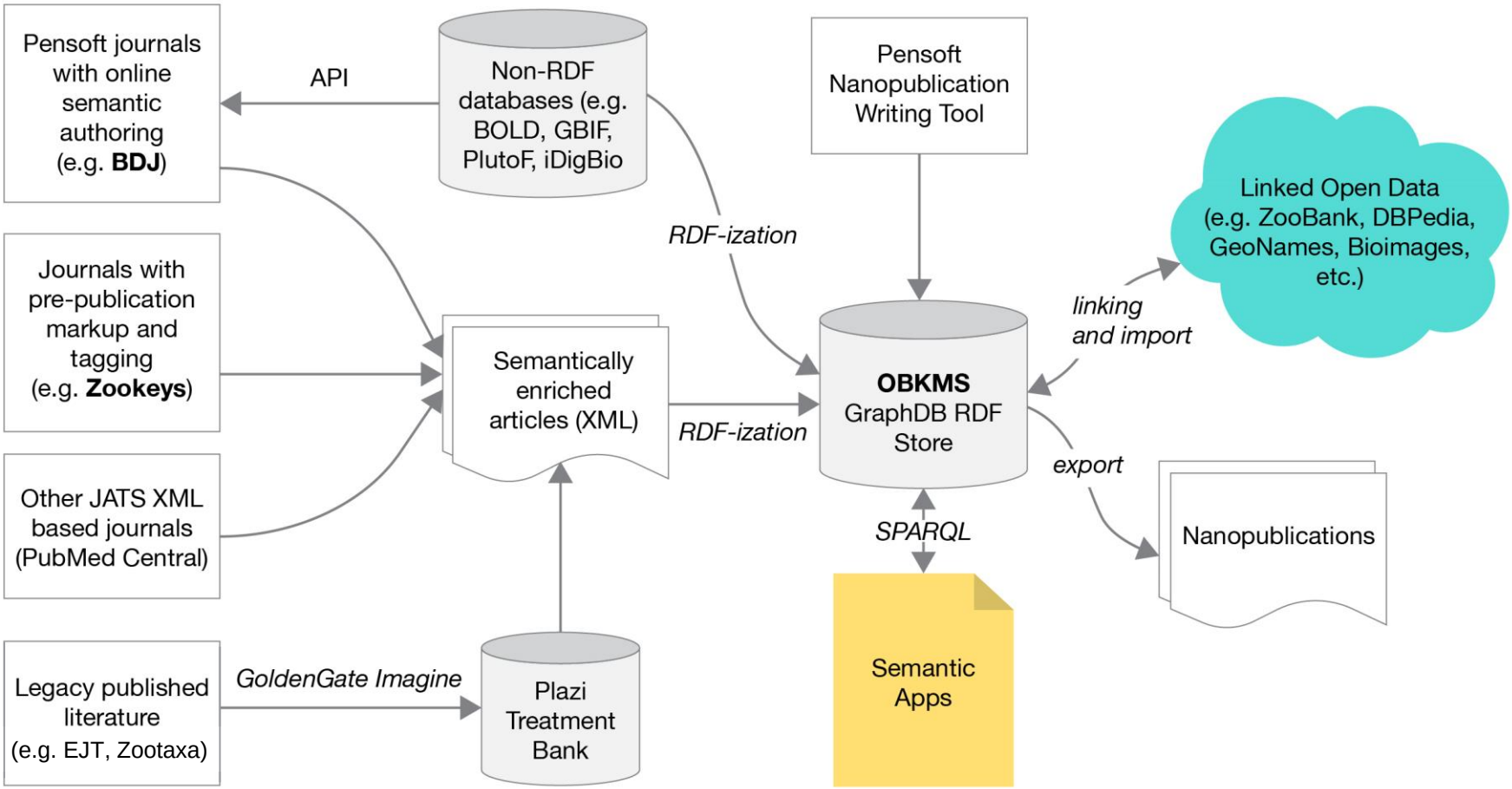
Received: 09 Apr 2015 | Accepted: 06 May 2015 | Published: 12 May 2015

Citation: Miller J, Agosti D, Penev L, Sautter G, Georgiev T, Catapano T, Patterson D, King D, Pereira S, Vos R, Sierra S (2015) Integrating and visualizing primary data from prospective and legacy taxonomic literature. Biodiversity Data Journal 3: e5063. doi: 10.3897/BDJ.3.e5063

Abstract

Specimen data in taxonomic literature are among the highest quality primary biodiversity data. Innovative cybertaxonomic journals are using workflows that maintain data structure and disseminate electronic content to aggregators and other users; such structure is lost in traditional taxonomic publishing. Legacy taxonomic literature is a vast repository of knowledge about biodiversity. Currently, access to that resource is cumbersome, especially for non-specialist data consumers. Markup is a mechanism that makes this content more accessible, and is especially suited to machine analysis. Fine-grained XML (Extensible Markup Language) markup was applied to all (37) open-access articles published in the journal Zootaxa containing treatments on spiders (Order: Araneae). The markup approach was optimized to extract primary specimen data from legacy publications. These data were combined with data from articles containing treatments on spiders published in Biodiversity Data Journal where XML structure is part of the routine publication process. A series of charts was developed to visualize the content of specimen data in XML-tagged taxonomic treatments, either singly or in aggregate. The data can be

The Open Biodiversity Knowledge Management System (OpenBiodiv)



PLAZI
TAKING CARE OF FREEDOM

http://openbiodiv.net/

PENSOFT

Removing the PDF prison

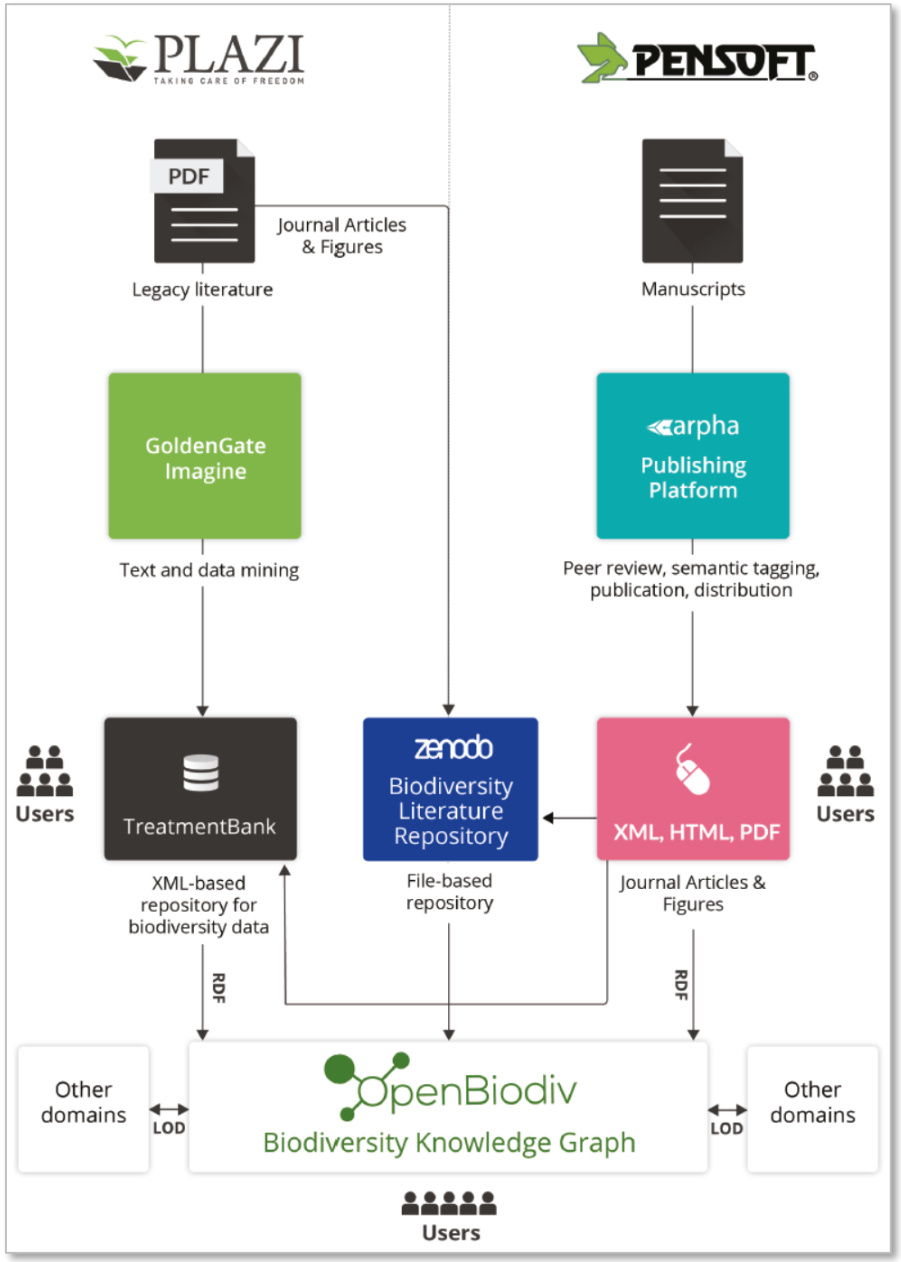
Highly **structured** research data → publishing → 

Semantically **unstructured** text → data mining → Semantically enhanced, **structured** data

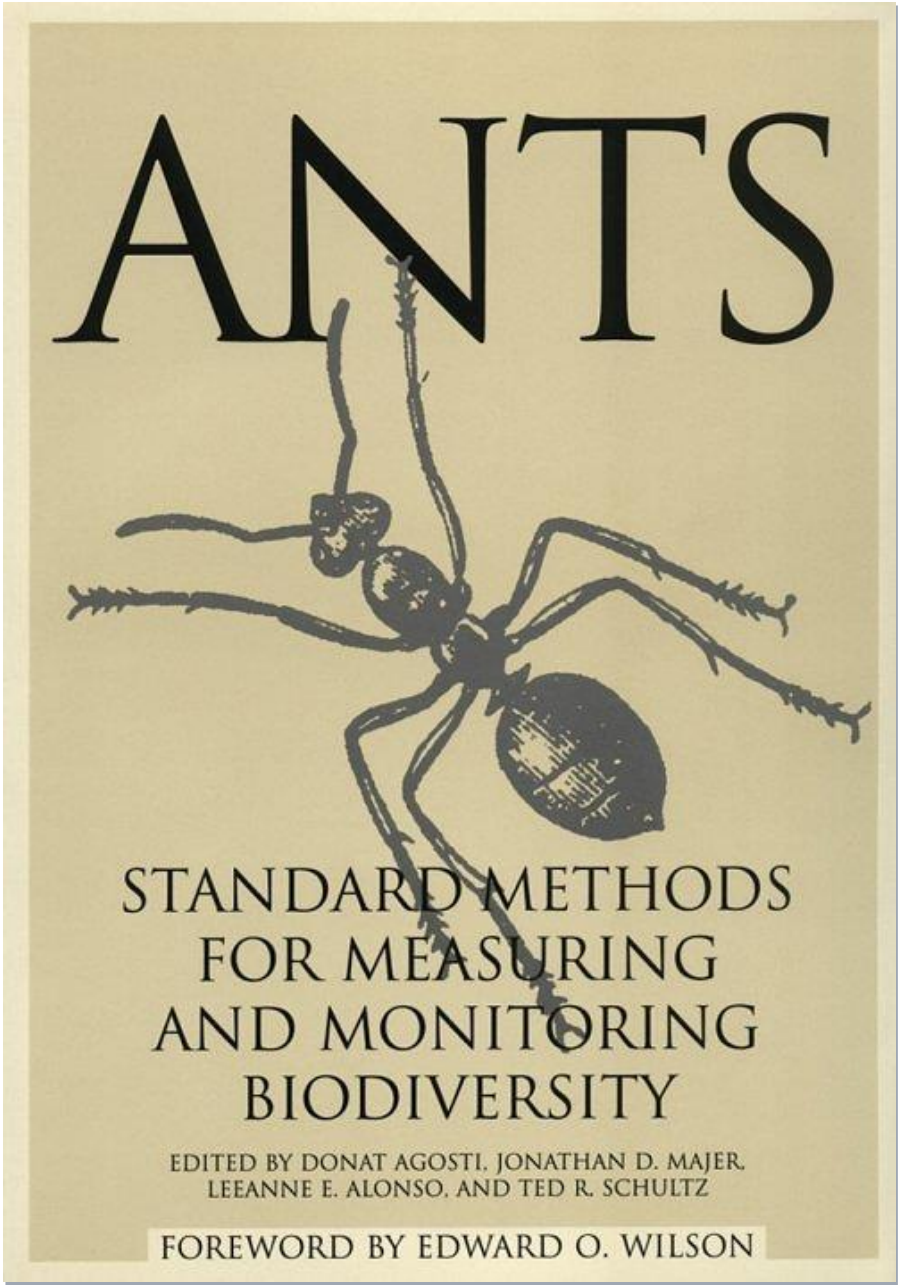
Highly **structured** research data → publishing → 

Semantically enhanced data → conversion → Semantically enhanced, **structured** data

The future will be open, semantically enhanced publishing



Our challenges or «use cases»



How many species are there on planet Earth?

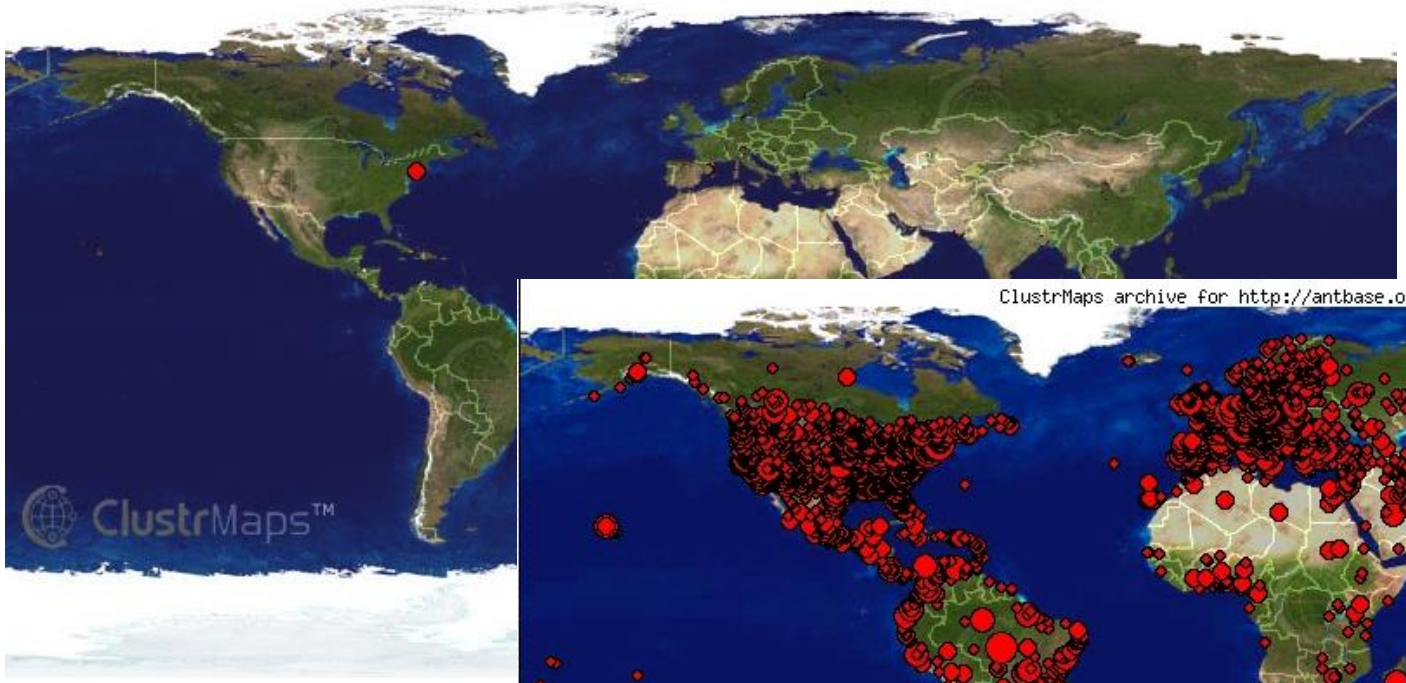
Compare biodiversity over space, time, habitat and land use change

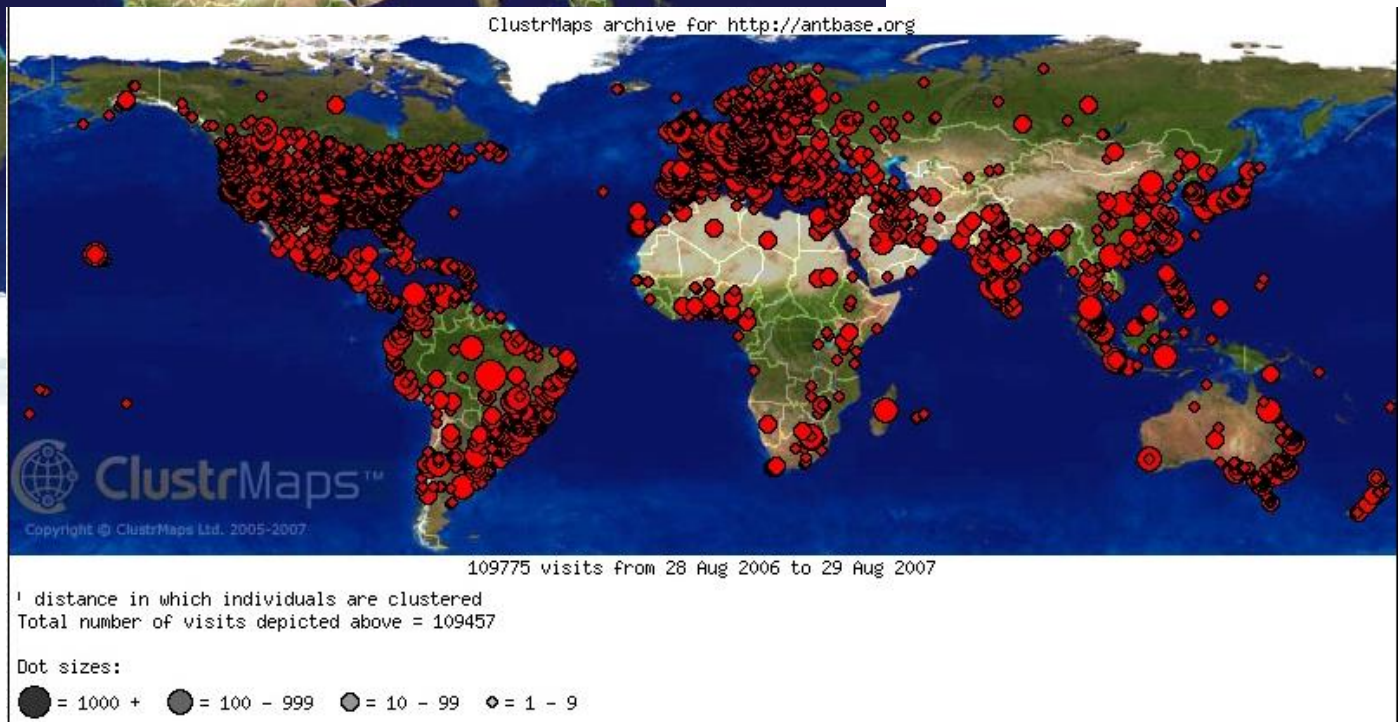
- What organism is it?
- What do I know about it?
- Where does it live?

1995: Idea
1996: NSF funded workshop
2000: SI-Press published
2018: >2,500 citations

The use of *Open Access* scholarly articles

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.





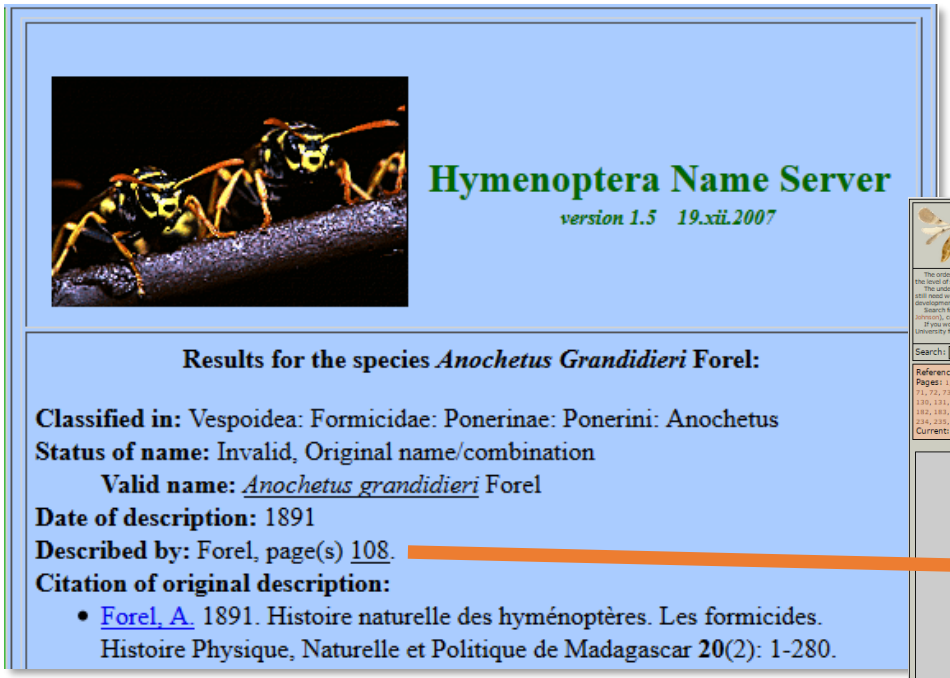
Online library, 2000 onwards

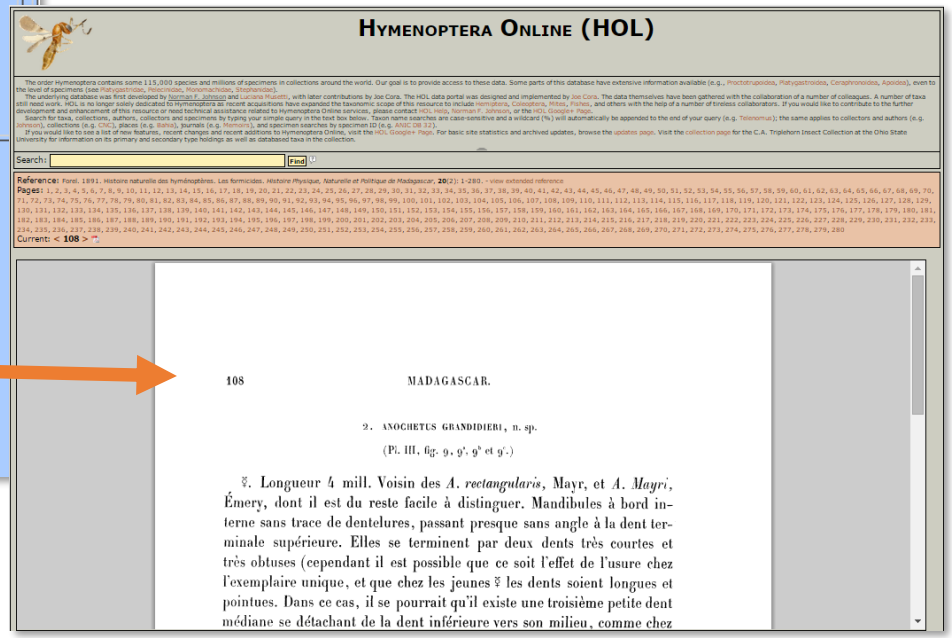
Antbase and **NMNH** through support from the **Atherton-Seidal foundation** opened up the entire ant taxonomic literature – and it has been used by a global community

Taxonomic treatment: linking to subarticle elements (and not articles)

From an article to treatments

Online catalogue
Open access
Online library
«Linked data»
2004





Treatment citation → Taxonomic treatment

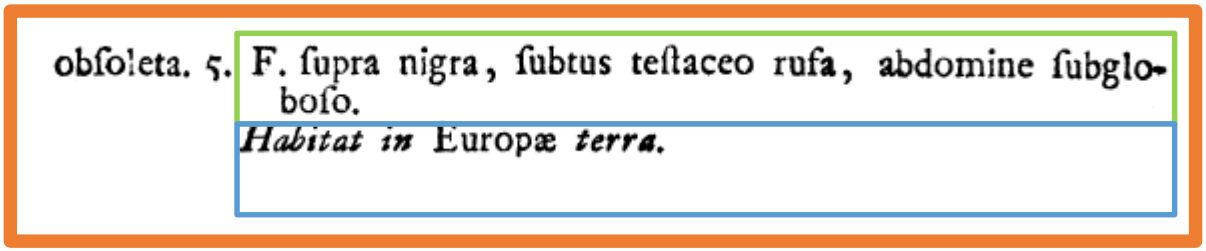
Special aspects of taxononomic publications:

Highly standardized in language and representation
Publications as container of citations and data

Taxonomic treatment

Formica obsoleta Linnaeus, 1758: 580

name


obſoleta. ς. F. ſupra nigra, ſubtus teſtaceo rufa, abdomine ſubglo-boſo.
Habitat in Europæ *terra.*

description

distribution

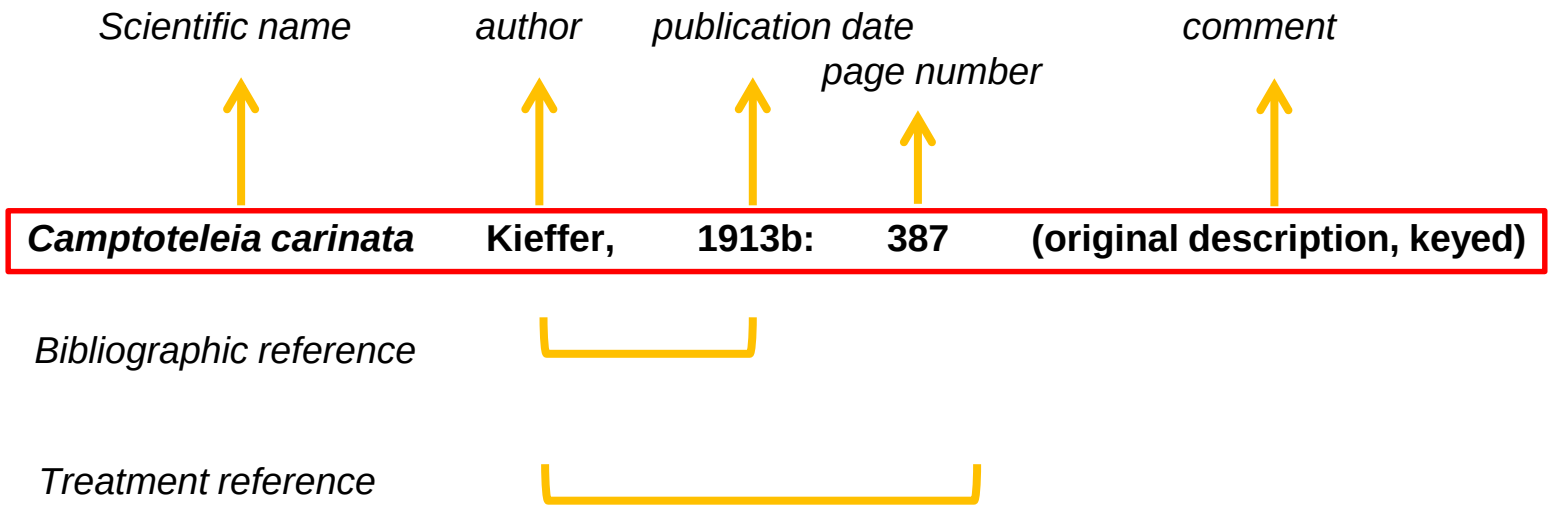
Each taxonomic name usage has it's treatment

Linnaeus has to be credited for Latin Binomen AND taxonomic treatment.

Treatment citation

Oxyscelio carinatus (Kieffer)
urn:lsid:zoobank.org:act:C91A9F90-F2C4-4391-ABB9-DCBE67993121
urn:lsid:biosci.ohio-state.edu:osuc_concepts:5009
http://species-id.net/wiki/Oxyscelio_carinatus
Figures 82–87; Morphbank [40]

Camptoteleia carinata Kieffer, 1913b: 387 (original description, keyed); Kieffer 1914: 296 (keyed); Kieffer 1916: 171 (keyed); Kieffer 1926: 380 (description, keyed).
Oxyscelio carinatus (Kieffer): Dodd 1931: 74 (generic transfer); Masner 1976: 23 (type information).
Camptoteleia kiefferi Benoit: Kelner-Pillault 1958: 150 (unnecessarily proposed replacement name, rejected by Baltazar (1966)).



Materials citation: citation of specimens

Profundiconus puillandrei sp. nov.
urn:lsid:zoobank.org:act:55807BF8-D984-42C1-B039-F3EBD8EB2C13
Figs 12A–J, 13A–D, 14

Conus ikedai – Poppe 2008: pl. 615, fig. 1a–b (non *C. ikedai* Ninomiya, 1987).
Conus smirna – Marshall 1981: 499, fig. 3j (non *C. smirna* Bartsch & Rehder, 1943).
Conus sp. C – Röckel *et al.* 1995b: 585, fig. 49.
Profundiconus n. sp. g. – Puillandre *et al.* 2014: Supplementary Material 1 (unfigured).

Etymology

This new species is dedicated to Dr. Nicolas Puillandre, mollusc curator at the MNHN. Dr. Puillandre has a long and outstanding trajectory in the study of the phylogeny of the Conoidea. The naming of the new species after him recognises his important contributions to the taxonomy of Conoidea at the molecular level.

Type material examined

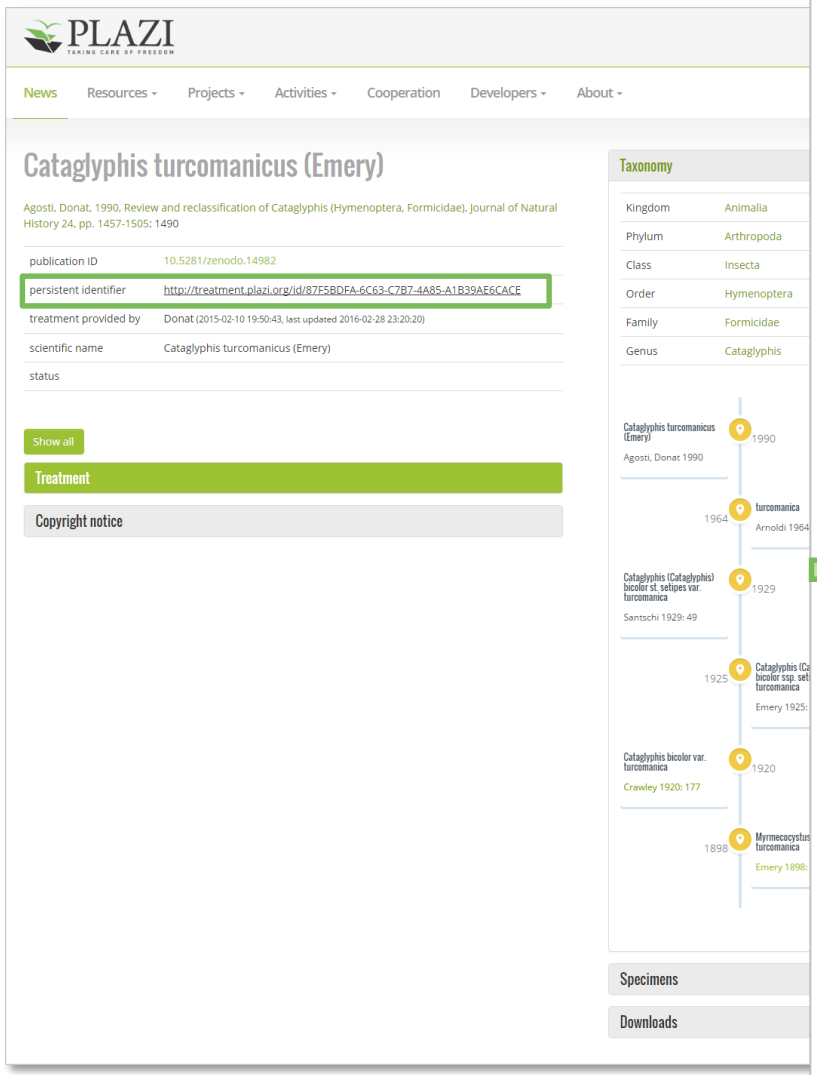
Holotype
NEW CALEDONIA: 43.2 × 18.0 mm, R/V *Alis*, NORFOLK 1 Expedition, st. DW 1707, Banc Jumeau Est, Norfolk Ridge, 23°43' S, 168°16' E, 381–493 m (MNHN IM-2000-30771; Fig. 12A).

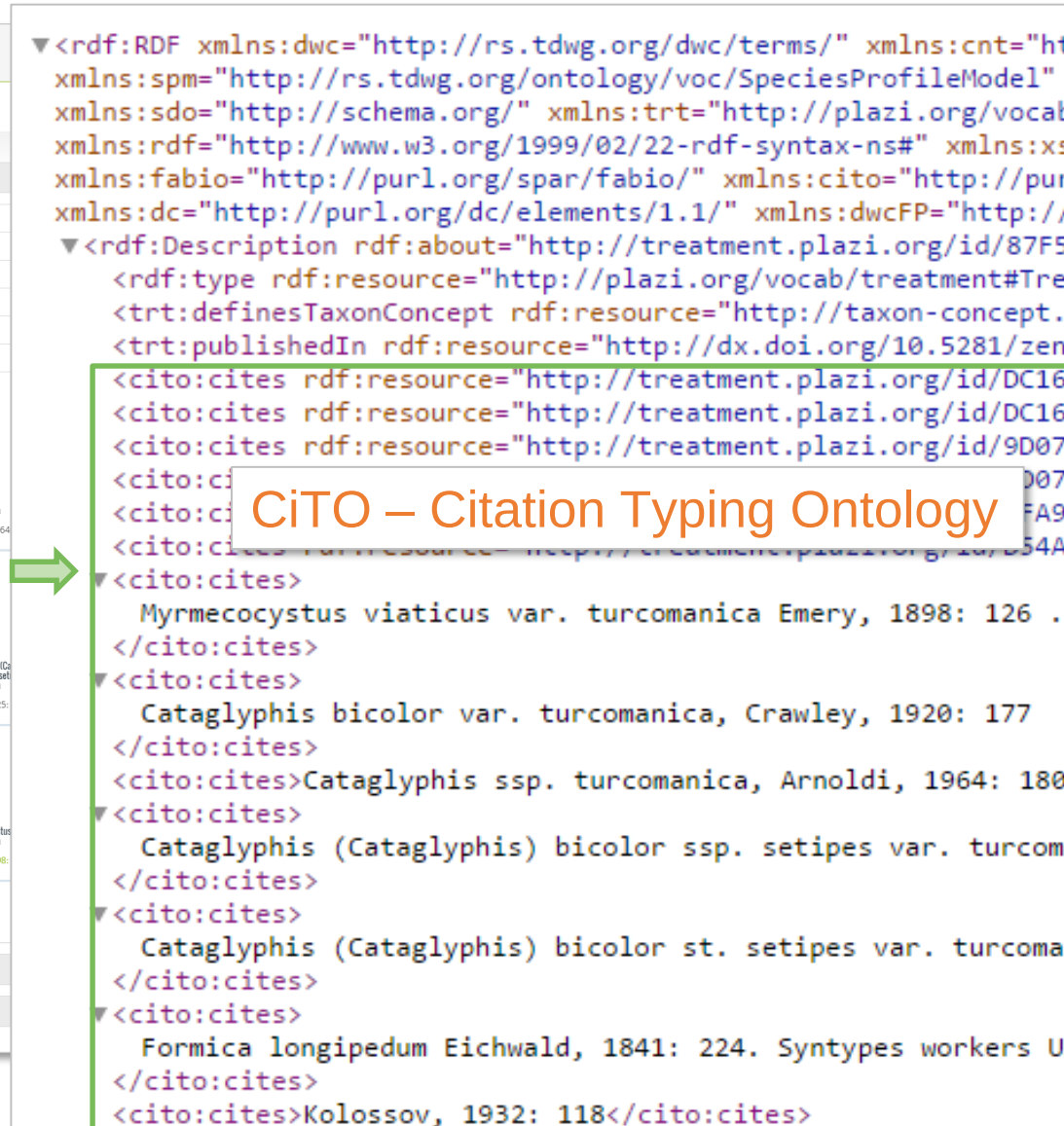
Paratypes
NEW CALEDONIA: 39.6 × 17.3 mm, R/V *Alis*, NORFOLK 2 Expedition, st. DW 2072, Banc Aramis, Norfolk Ridge, 25°21' S, 168°57' E, 1000–1006 m (MNHN IM-2000-30772; paratype 1; Fig. 12B);

From a taxonomic treatment to material citations

TreatmentBank: Citing treatments

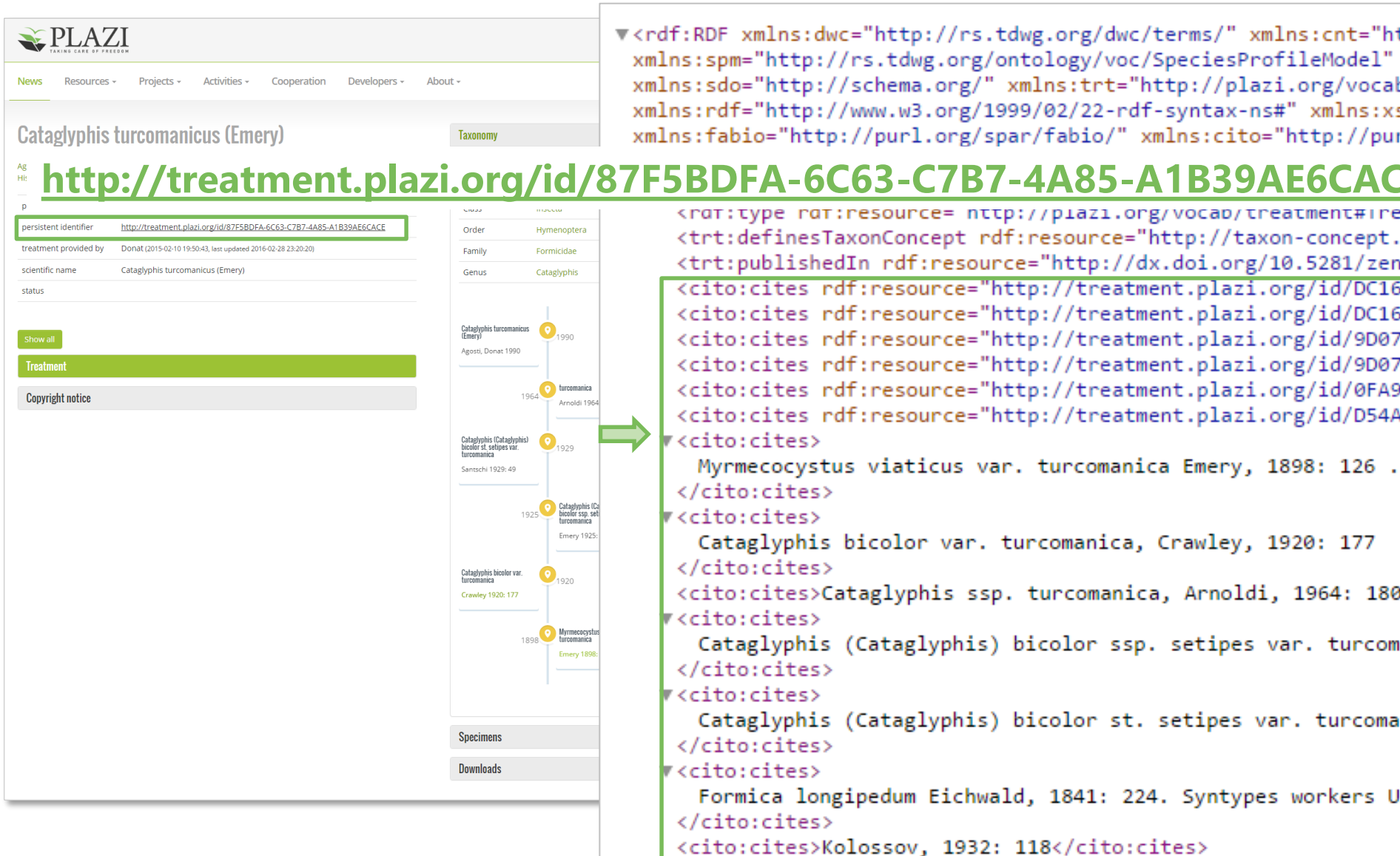
Treatment citations are similar to article citations: use of existing ontologies or vocabularies





TreatmentBank: citing treatments

Persistent Identifiers for treatments allows to cite treatments



Materials citation: linking to the digital object of the cited specimen

Profundiconus puillandrei sp. nov.
urn:lsid:zoobank.org:act:55807BF8-D984-42C1-B039-F3EBD8EB2C13
Figs 12A–J, 13A–D, 14

Conus ikedai – Poppe 2008: pl. 615, fig. 1a–b (non *C. ikedai* Ninomiya, 1987).
Conus smirna – Marshall 1981: 499, fig. 3j (non *C. smirna* Bartsch & Rehder, 1943).
Conus sp. C – Röckel *et al.* 1995b: 585, fig. 49.
Profundiconus n. sp. g. – Puillandre *et al.* 2014: Supplementary Material 1 (unfigured).

Etymology

This new species is dedicated to Dr. Nicolas Puillandre, mollusc curator at the MNHN. Dr. Puillandre has a long and outstanding trajectory in the study of the phylogeny of the Conoidea. The naming of the new species after him recognises his important contributions to the taxonomy of Conoidea at the molecular level.

Type material examined

Holotype
NEW CALEDONIA: 43.2 × 18.0 mm, R/V *Alis*, NORFOLK 1 Expedition, st. DW 1707, Banc Jumeau Est, Norfolk Ridge, 23°43' S, 168°16' E, 381–493 m (MNHN IM-2000-30771; Fig. 12A).

Paratypes
NEW CALEDONIA: 39.6 × 17.3 mm, R/V *Alis*, N
Norfolk Ridge, 25°21' S, 168°57' E, 1000–1006 m (MN

28

From a taxonomic treatment
to a material citation
to the digital object

MNHN-IM-2007-30772

MUSEUM
NATIONAL D'HISTOIRE NATURELLE

MNHN / Molluscs (IM) / 2007-30772

Conus frigidus Reeve, 1848

SPECIMEN

ALCOOL95	MNHN-IM-2007-30772
Size	37.3

MNHN-IM-2007-30772

TAXONOMY

Phylum	Mollusca
Class	Gastropoda
Order	Neogastropoda
Family	Conidae
Genus	Conus
Species	Conus frigidus
Name	*Conus frigidus* Reeve, 1848

ORIGIN

Country label	VANUATU
Locality	Baldwin Core
Expedition	SANTO 2006
Station number	VM02
Collection date	2006-09-10
Depth (meters)	0 -1
Geographic	15° 34' 51.9996" S ; 167° 2' 23.5248" E

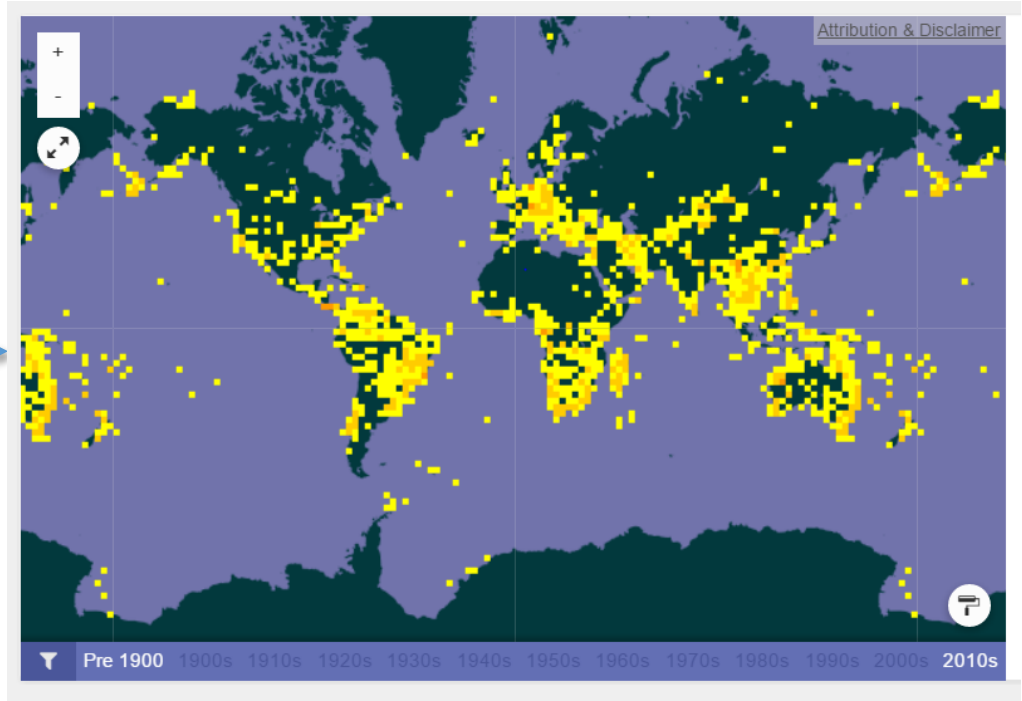
European Journal of Taxonomy:
DOI: 10.5852/ejt.2016.173
Treatment:

Data dissemination

Export of observation records into GBIF

- Observation records: 64,651
- Observation records geo-referenced: 25761
- Taxonomic names: 159,983
- Taxonomic treatments: 173,245

PLAZI
TAKING CARE OF FREEDOM

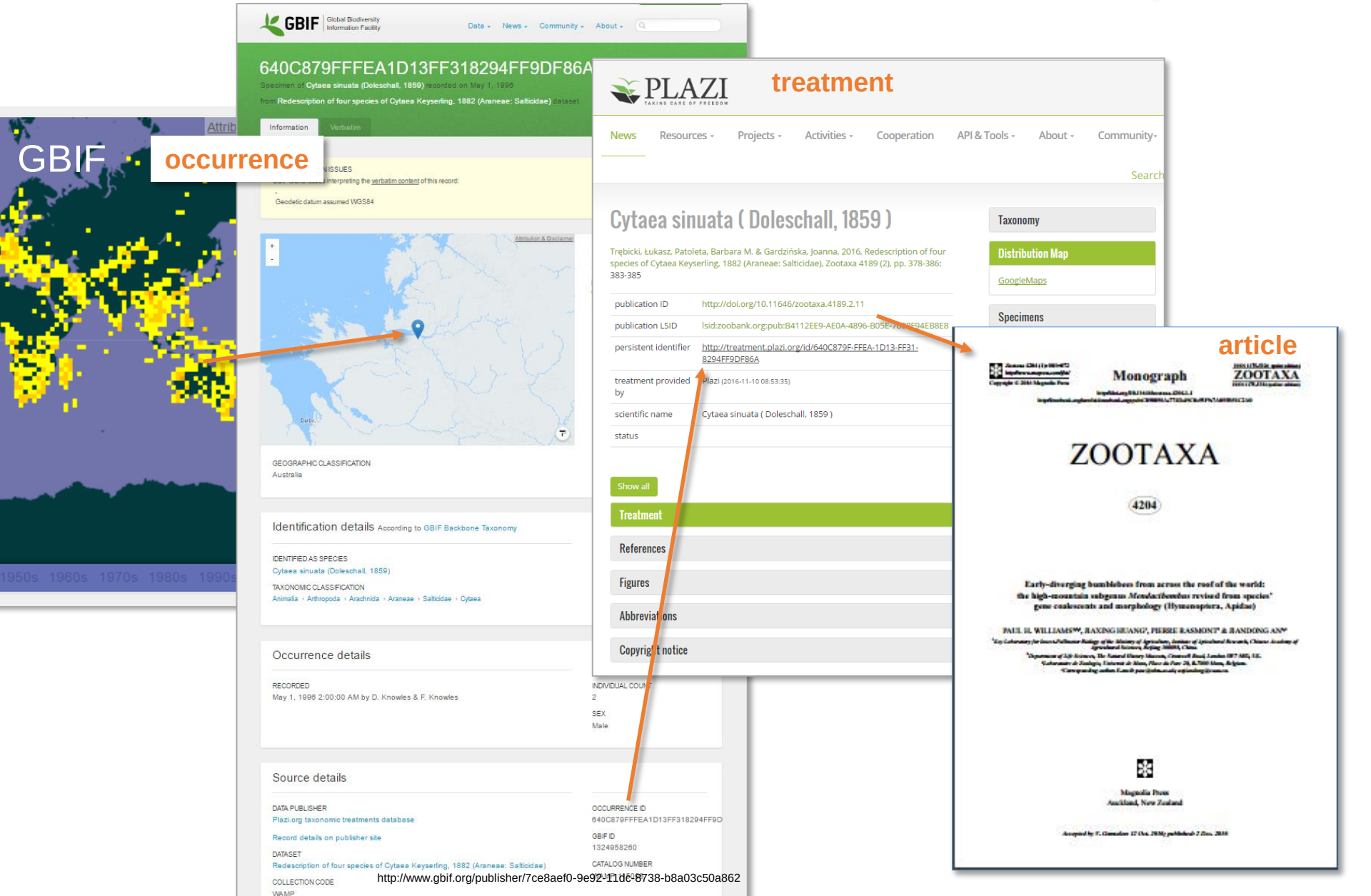


GBIF
http://www.gbif.org/publisher/7ce8aef0-9e92-11dc-8738-b8a03c50a862

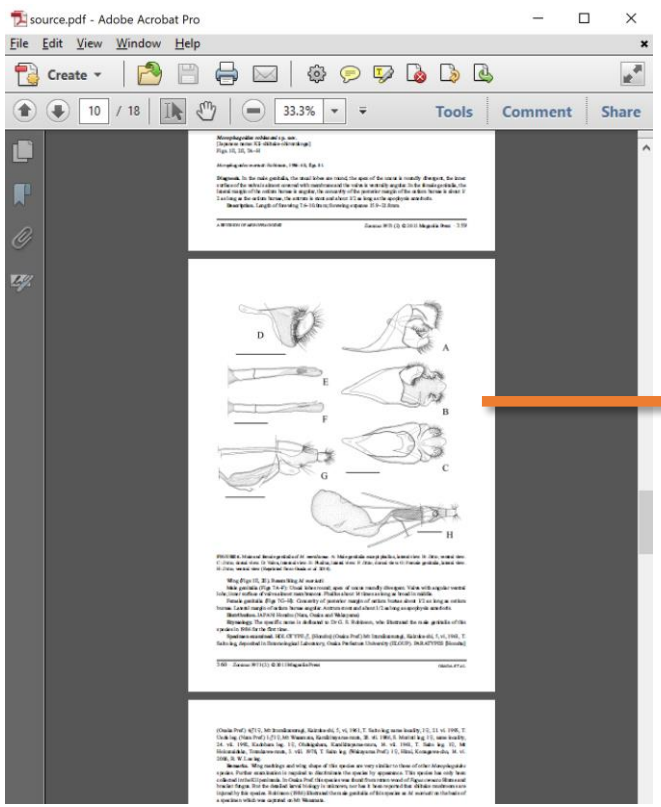
Long tail: major contribution towards little known, rare and recently discovered species

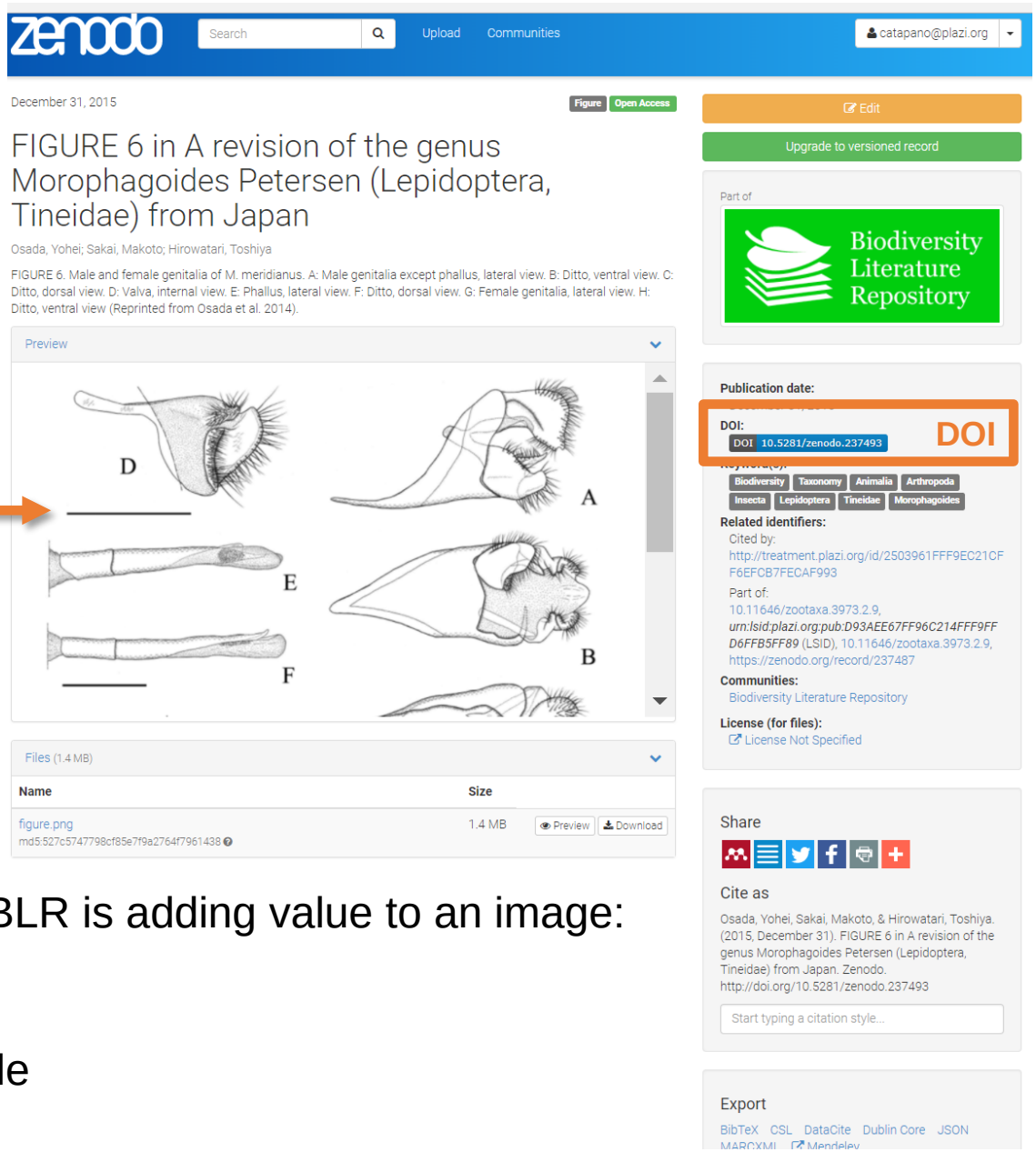
Plazi: Linking occurrences to the source publication

Occurrence → treatment → article



Plazi infrastructure: Biodiversty Literature Repository

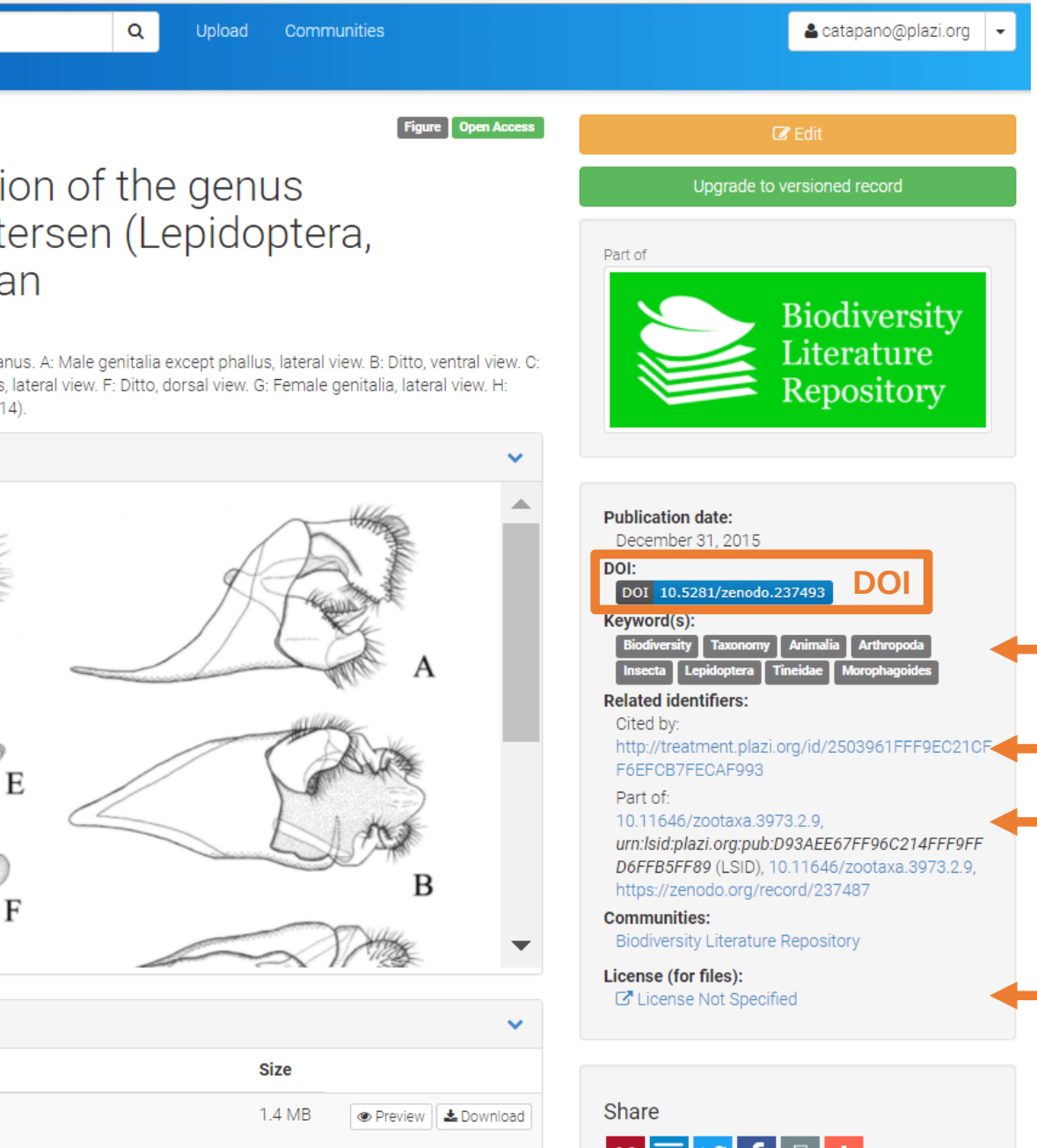




Depositing in BLR is adding value to an image:

- findable
- accessible
- interoperable
- reusable

Biodiversity Literature Repository: enhanced illustration



Keywords / vocabularies

Cited by

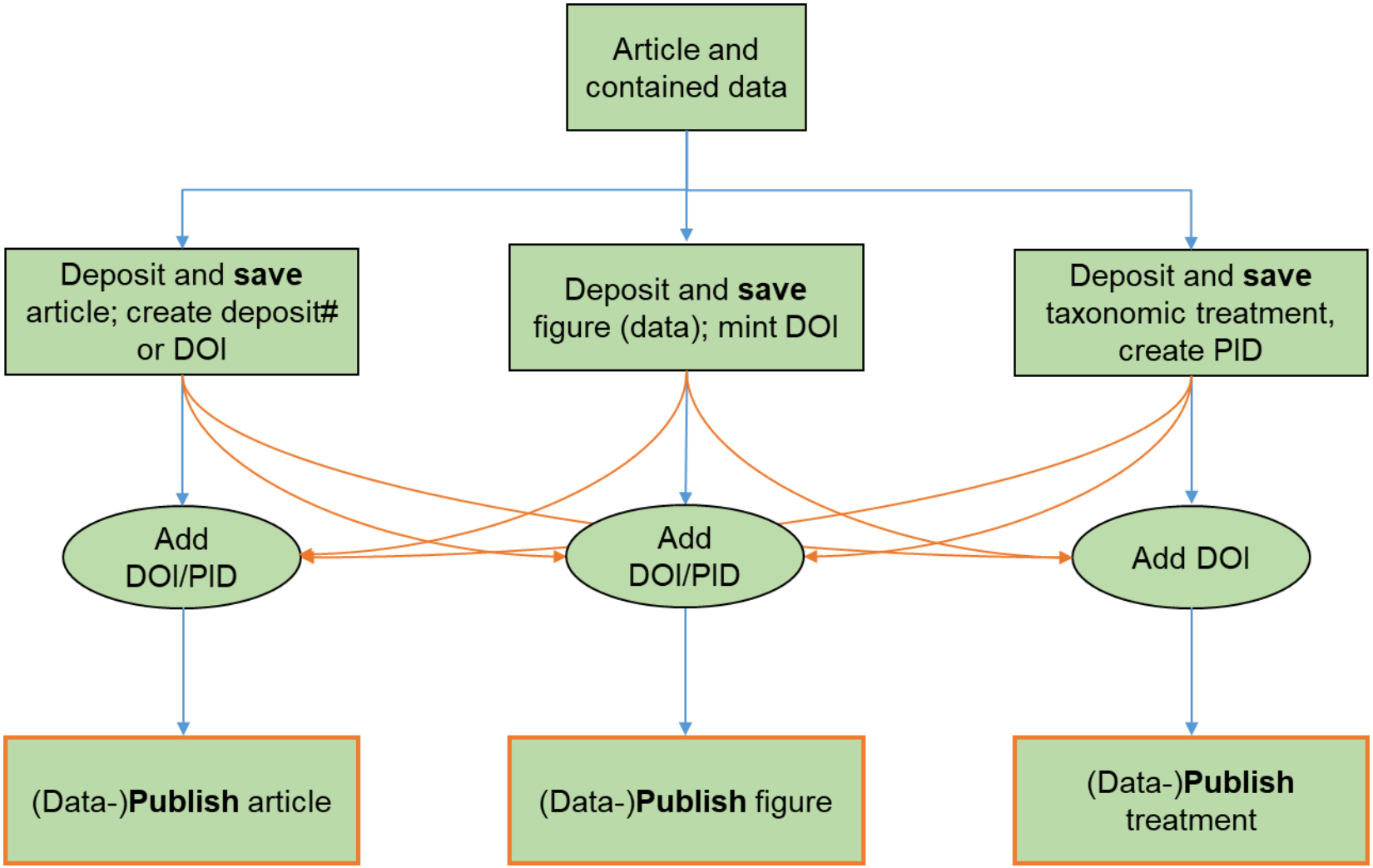
Part of

Licence

Biodiversity Literature Repository: articles

https://zenodo.org/record/1154504

Deposition of an article and its data on BLR/Zenodo



Biodiversity Literature Repository

A community in *Zenodo* to share publications and published data related to bio-systematics.

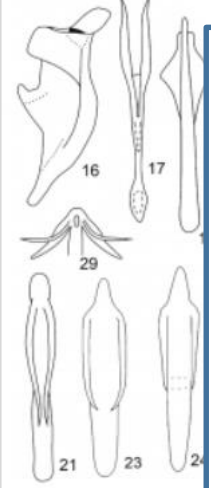
- All scientific, published results
- DOI for all deposits
- All deposits linked to source
- All deposits tagged, linked
- Daily input
- Collaboration of Plazi, Pensoft and Zenodo/CERN
- Support from Arcadia Fund and EU ICEDIG

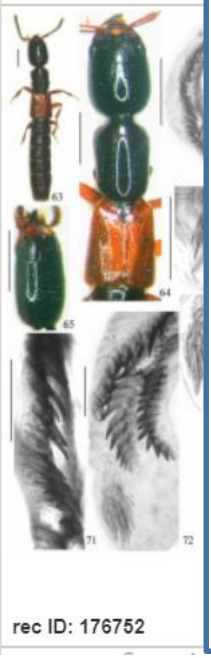
- 205,373 deposits total
- 32,012 articles
- 171,401 images

Source: BLR
https://ocellus.punkish.org/?q=aedeagus&page=1

Biolitrepo.org

Biodiversity Literature Repository





Plazi: Linking and citing treatments at antweb

PLAZI
TAKING CARE OF FREEDOM

News | Resources | Projects | Activities | Cooperation | API & Tools | Ab

Phrynoponera pulchella Bolton & Fisher , Bolton, B. & Fisher, B. L., 2008

Bolton, B. & Fisher, B. L., 2008, The Afrotropical ponerine ant genus Phrynoponera Wheeler (Hymenoptera: Formicidae)., Zootaxa 1892, pp. 35-52: 47-49

publication ID	22424
persistent identifier	http://treatment.plazi.org/id/857407D3-73CD-1CFE-0B98-B4011B9750FD
treatment provided by	Christiana (2009-08-06 00:05:20)
scientific name	Phrynoponera pulchella Bolton & Fisher
status	sp. n.

Show all

Treatment

Phrynoponera pulchella Bolton & Fisher HNS Edit sp. n. Edit

(Figures 5a-d) Holotype worker. TL 11.1, HL 2.48, HW 2.56, CI 103, SL 2.04, SI 80, PW 2.04, WL 3.76, maximum diameter of eye 0.56, OI 22. Edit

Mandible smooth with scattered setal pits, with 5 teeth, the basal tooth no more than a broad angle. Median portion of anterior clypeal margin with a small and extremely shallow median indentation which rounds broadly and evenly into the clypeal margin on each side; without a prominent tooth on each side of the midline. Dorsum of head near midline very finely and very densely rugulose, the narrow rugulae predominantly longitudinal but with some fine anastomoses; spaces between rugulae extremely finely reticulate-punctate to shagreenate. More laterally on cephalic dorsum, and behind the eyes, the reticulae are wider though no more strongly defined. Dorsum of mesosoma coarsely, irregularly reticulate-rugose, spaces between reticulae shagreenate. Metanotal groove in dorsal view visible laterally but obliterated towards central area of dorsum. Median strip of propodeal dorsum with pubescence that is much more dense than anywhere else on dorsum of mesosoma. Propodeal teeth in dorsal view short, broad at base and blunt apically. Petiole broad, width at base of spines ca. 2.00, surmounted by 5 spines that are all short and stout. Outer pair of petiolar spines the longest, broadly triangular and weakly divergent. Median spine only slightly longer than the flanking pair. Gastral tergites 1-3 shagreenate to extremely finely and densely reticulate-punctate, with scattered larger setal pits; tergites also densely clothed with subappressed to appressed pubescence. Setae on first gastral tergite suberect to subdecumbent, short, maximum length ca. 0.22. Edit

Paratype workers. TL 11.5-12.0, HL 2.40-2.48, HW 2.40-2.62, CI 100-106, SL 1.92-2.04, SI 78-80, PW 1.88-1.96, WL 3.44-3.68, maximum diameter of eye 0.56-0.58, OI 22-23 (3 measured). As holotype but mandible with 3-5 teeth. In one specimen the left mandible has 5 teeth, the right mandible 3. Edit

Holotype worker. Kenya: Kora Res., Tana River, 22.vii.1983, 0-100 m., no. 5c, Acacia-Commiphora scrub (N.M. Collins & M. Ritchie) (BMNH). Edit

Paratypes: 9 workers (2 dissected, on same pin), with the same data (BMNH, CASC). Edit

AntWeb v6.10.6 About | Participate | AntBlog | AntShare | Press | Contact | AntCat | API |

Browse: Global | Georegions | Bioregions | Projects | Museums | Favorites Search for: Go Advanced Searc

Current View: All Antweb

Cite this page

Specimen: CASENT0900674 Phrynoponera pulchella

Overview | Images

Classification: Order: Hymenoptera Family: Formicidae Subfamily: Ponerinae Genus: Phrynoponera Species: pulchella Specimen: CASENT0900674

Persistent Identifier:
https://www.antweb.org/specimen/CASENT0900674

Locality Information:

Biogeographic Region:	Afrotropical
Locality:	Kenya: Kora Reserve, Tana River, no 5c
Latitude:	0.0
Longitude:	0.0
Lat/Long Max Error:	
Elevation	50 m
Elevation Max Error:	
Locality Notes:	

Collection Information:

Collection code:	ANTC19776
Collected by:	NM. Collins, M. Ritchie
Date collected:	22 Jul 1983
Date collected end:	
Habitat:	
Method:	
Microhabitat:	
Collection notes:	

Specimen Information:

Life stage:	w
Located at:	BMNH
Owned by:	BMNH, London, U.K.
Determined by:	B. Bolton
Date determined:	31 Dec 2008
Type status:	T Holotype of Phrynoponera pulchella

Plazi: Taxonomic treatments reused

Species: Phrynoponera pulchella Bolton & Fisher, 2008

Overview | Specimens | Images | Map

View in AntCat

Classification: Order: Hymenoptera Family: Formicidae Subfamily: Ponerinae Genus: Phrynoponera Species: pulchella

Compare Images | Download Data

Taxonomic History (provided by Barry Bolton, 2016)

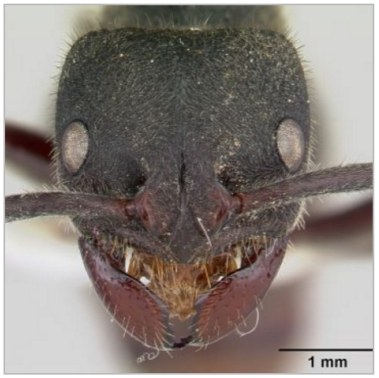
Phrynoponera pulchella Bolton & Fisher, 2008B PDF: 47, figs. 5a-d (w.) KENYA. Afrotropic. AntCat AntWiki HOL

Distribution:

Afrotropical Region: Kenya, Tanzania

Distribution Notes:

specimen CASENT0178203 and CASENT0178204 on AntWeb



Enlarge Map

Taxonomic Treatment (provided by Plazi)

Treatment Citation: Bolton, B. & Fisher, B. L., 2008, The Afrotropical ponerine ant genus Phrynoponera Wheeler (Hymenoptera: Formicidae)., Zootaxa 1892, pp. 35-52

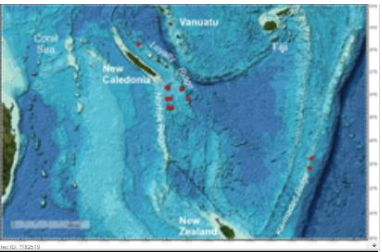
Phrynoponera pulchella Bolton & Fisher[HNS] sp. n.

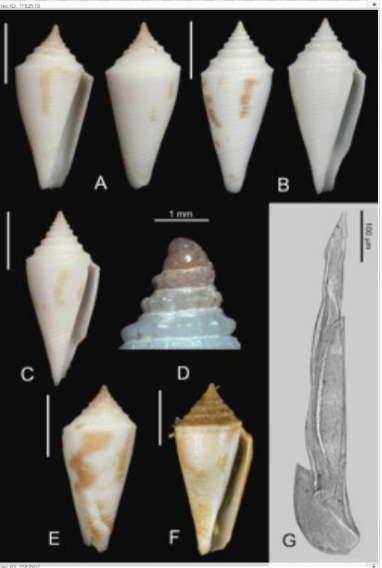
(Figures 5a-d) Holotype worker. TL 11.1, HL 2.48, HW 2.56, CI 103, SL 2.04, SI 80, PW 2.04, WL 3.76, maximum diameter of eye 0.56, OI 22.

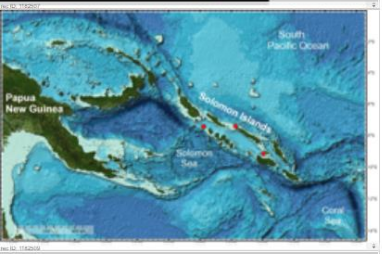
FIGURE 5. Phrynoponera pulchella[HNS] paratype worker CASENT0178204: A, dorsal view of body; B, dorsal view of petiole; C, lateral view of body; D, full-face view of head.

Mandible smooth with scattered setal pits, with 5 teeth, the basal tooth no more than a broad angle. Median portion of anterior clypeal margin with a small and extremely shallow median indentation which rounds broadly and evenly into the clypeal margin on each side; without a prominent tooth on each side of the midline. Dorsum of head near midline very finely and very densely rugulose, the narrow rugulae predominantly longitudinal but with some fine anastomoses; spaces between rugulae extremely finely reticulate-punctate to shagreenate. More laterally on cephalic dorsum, and behind the eyes, the reticulae are wider though no more strongly defined. Dorsum of mesosoma coarsely, irregularly reticulate-rugose, spaces between reticulae shagreenate. Metanotal groove in dorsal view visible laterally but

Context


Vanuatu
Coral Sea
New Caledonia
Fiji
New Zealand

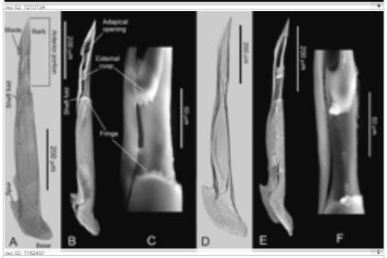

A
B
1 mm
C
D
E
F
G

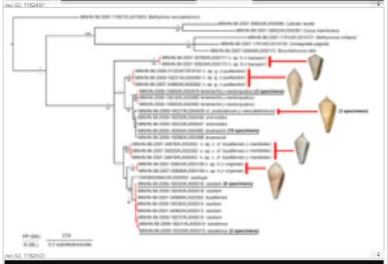

South Pacific Ocean
Papua New Guinea
Solomon Islands
Solomon Sea
Coral Sea

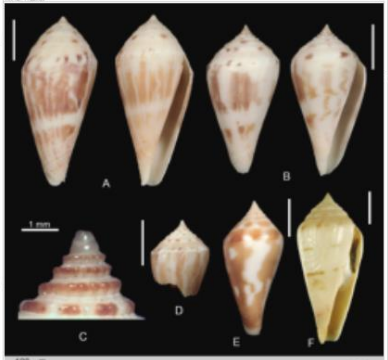

100 µm
A


100 µm
B

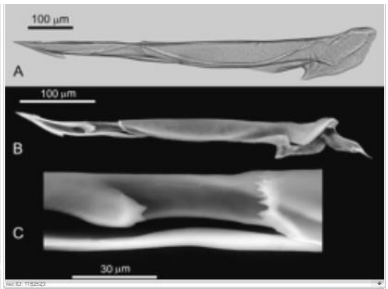

© MNHN Paris, France

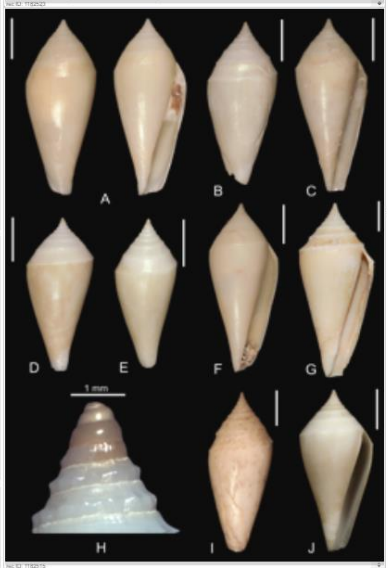

Adapical opening
Fringe
A
B
C
D
E
F

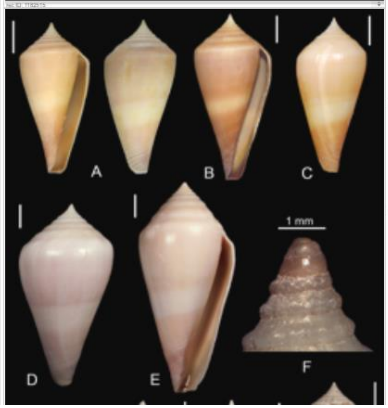


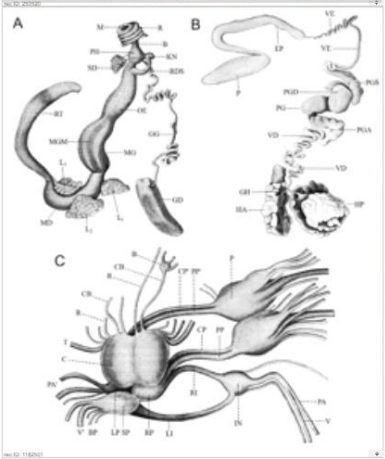

A
B
1 mm
C
D
E
F

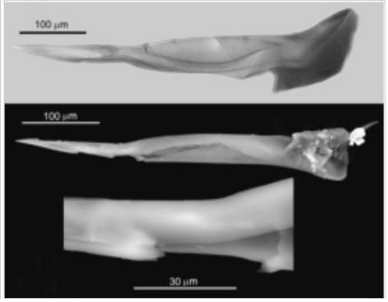

100 µm
G
100 µm

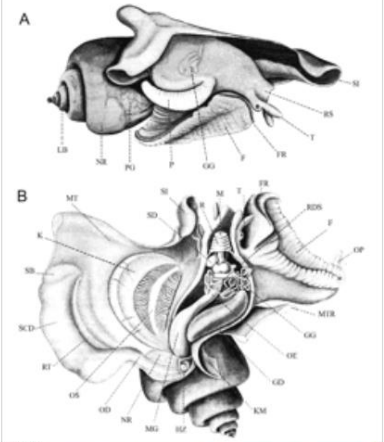

100 µm
A
100 µm
B
C
30 µm

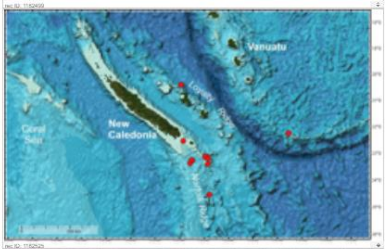

A
B
C
D
E
F
G
1 mm
H
I
J


A
B
C
1 mm
D
E
F

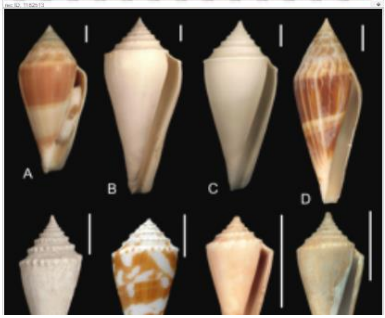

A
B
C


100 µm
100 µm
30 µm


A
B


Vanuatu
New Caledonia
Coral Sea


Vanuatu
New Caledonia
Coral Sea

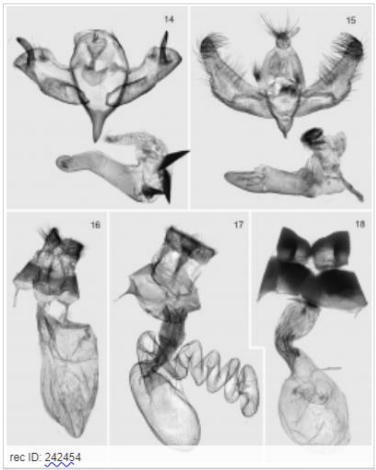

A
B
C
D

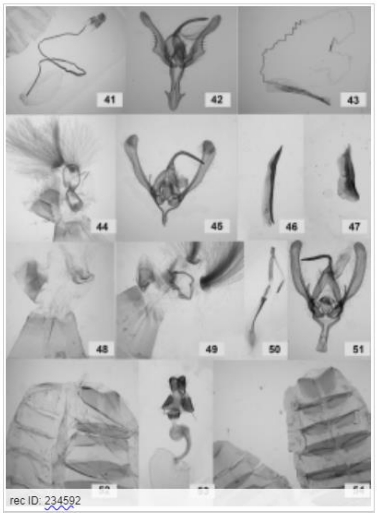
Novel applications, novel access to publications

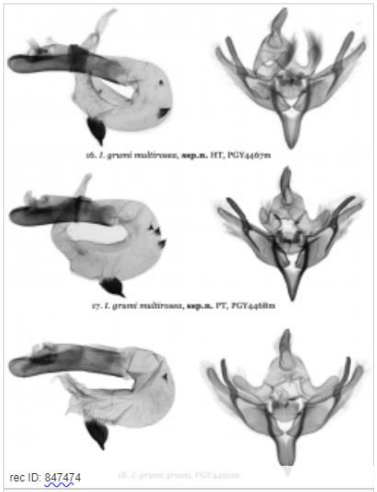
OCELLUS

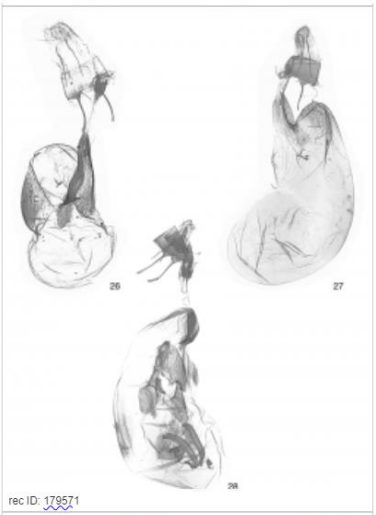
noctuidae+genitalia GO

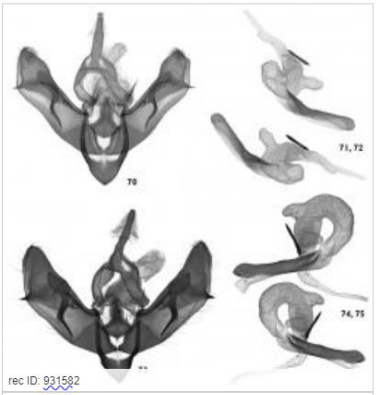
13111 records found

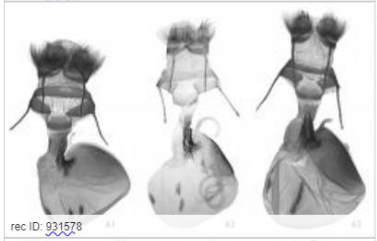


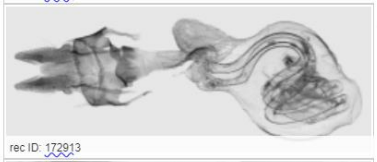


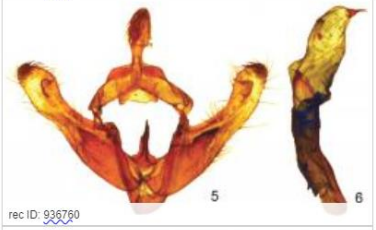


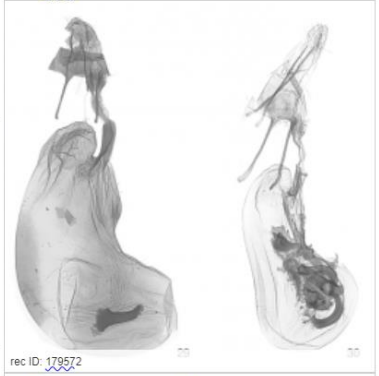


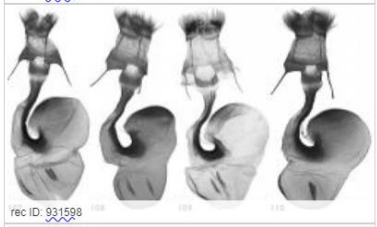


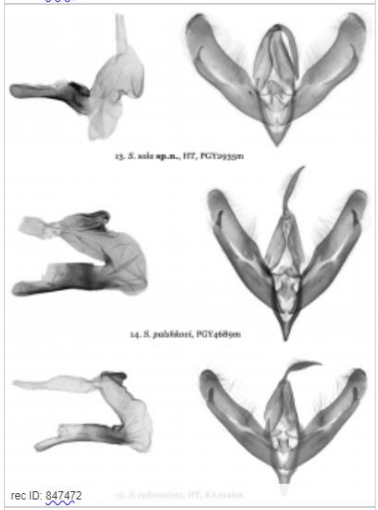


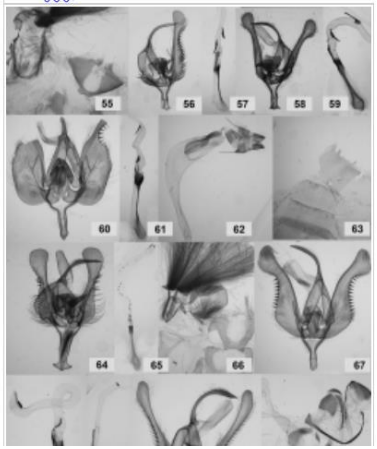


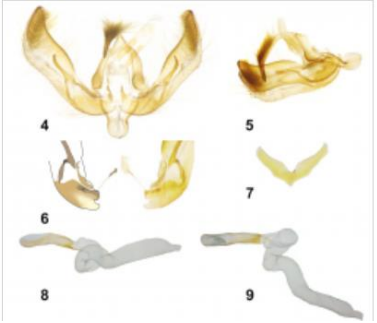




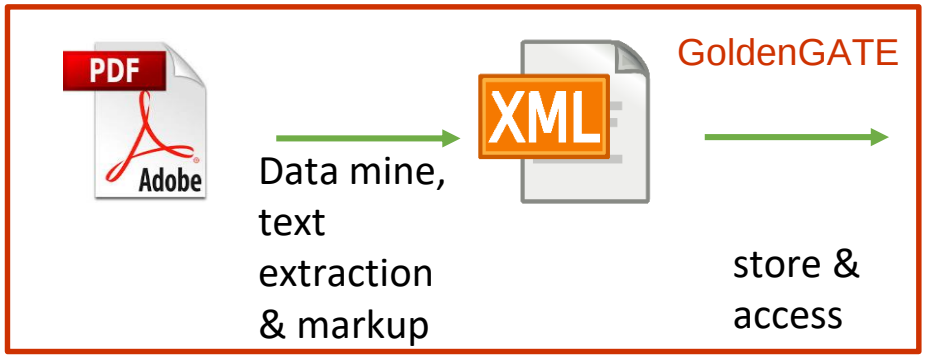




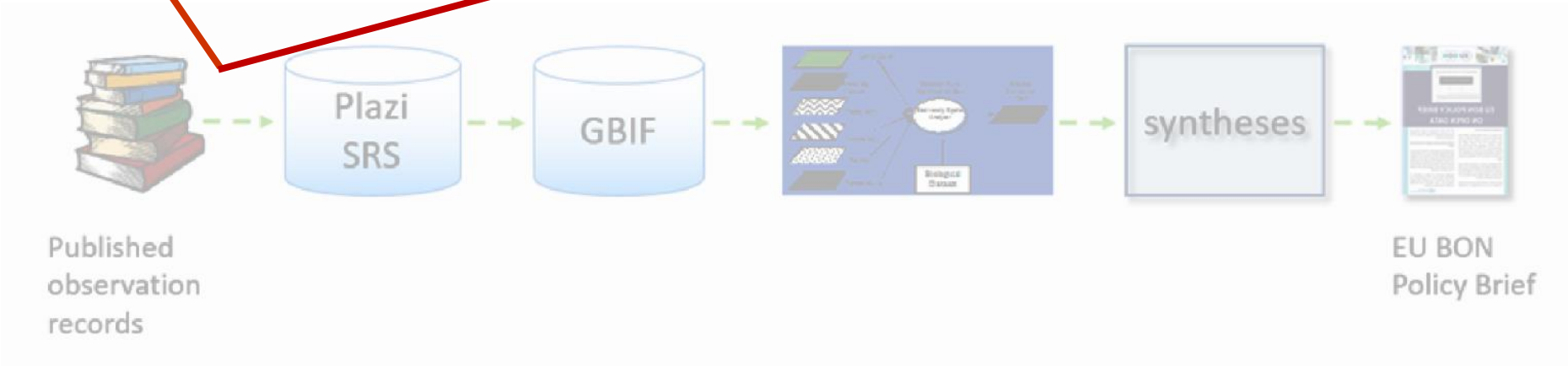




Plazi workflow from article to citable data



GoldenGATE
- Native version
- Customized versions
- Batch mode



Alternatively, direct import of XML-based arti

Improved science communication

TDM comes at a high cost, that ought to be omitted to discover data objects in text.

This can be done using semantically enhanced publishing.

An example is Pensoft's use of the Journal Archival Tag Suit (JATS) extension Taxpub which includs all the semantic elements describing the data in taxonomy.

Pensoft publications are automatically imported, without a TDM step, into TreatmentBank.

JATS based publications can automatically be imported in PubMed Central

Expeditions to discover the known biodiversity

Make use and promote existing tools to liberate taxon related data from publications

Build corpora of knowledge centred around taxa, institutions, locations, authors or collectors, making use of the citations

Make all new taxonomic treatments and its data immediately accessible an imperative

Create the catalogue of life by machine

Expeditions to discover the known biodiversity

It’s a big task: Let’s move together!

PLAZI
TAKING CARE OF FREEDOM

Biodiversity
Literature
Repository

zenodo

CERN

Thank you!

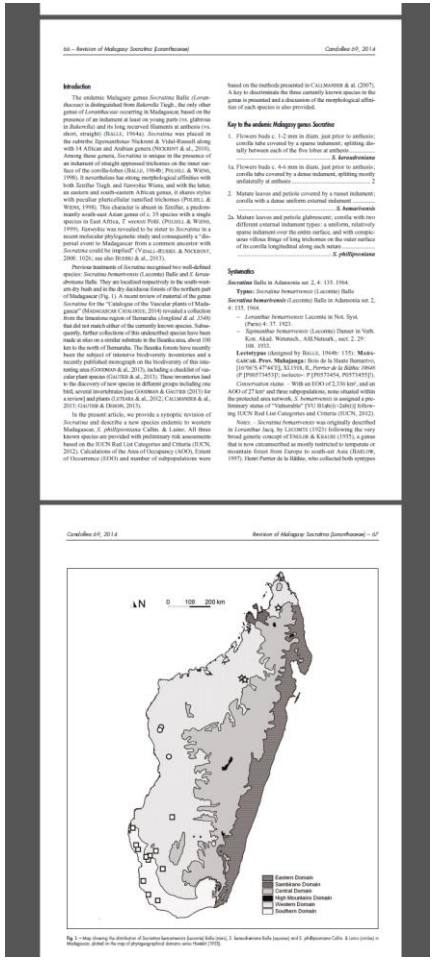
Donat Agosti

agosti@plazi.org

Appendix: workflow

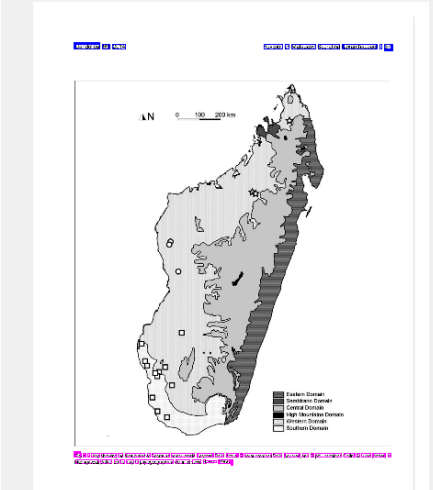
Plazi: conversion: extracting text streams and other elements

Re-creating a semantically enhanced publication

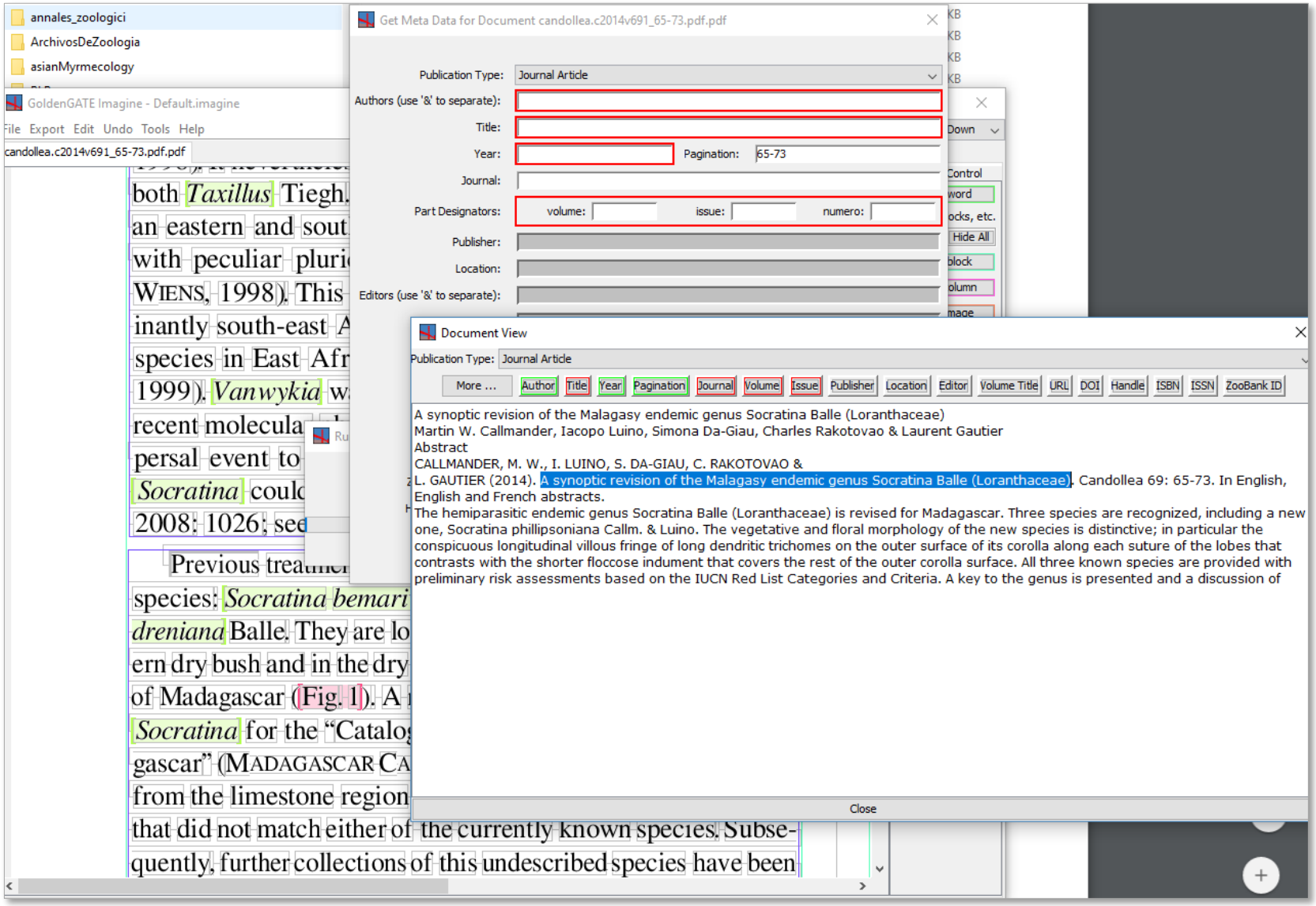


preparation

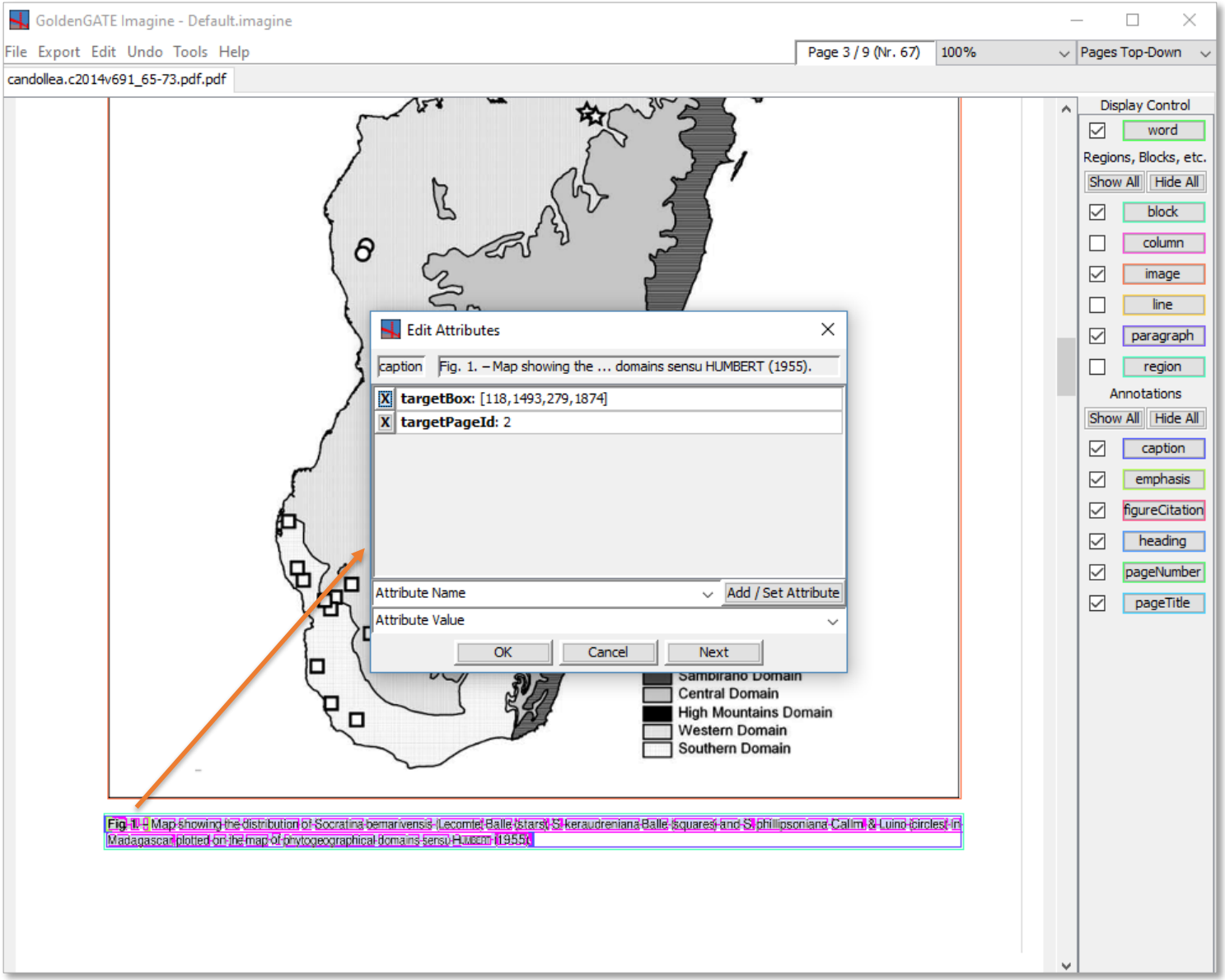
Extraction of text
text stream
various digital elements



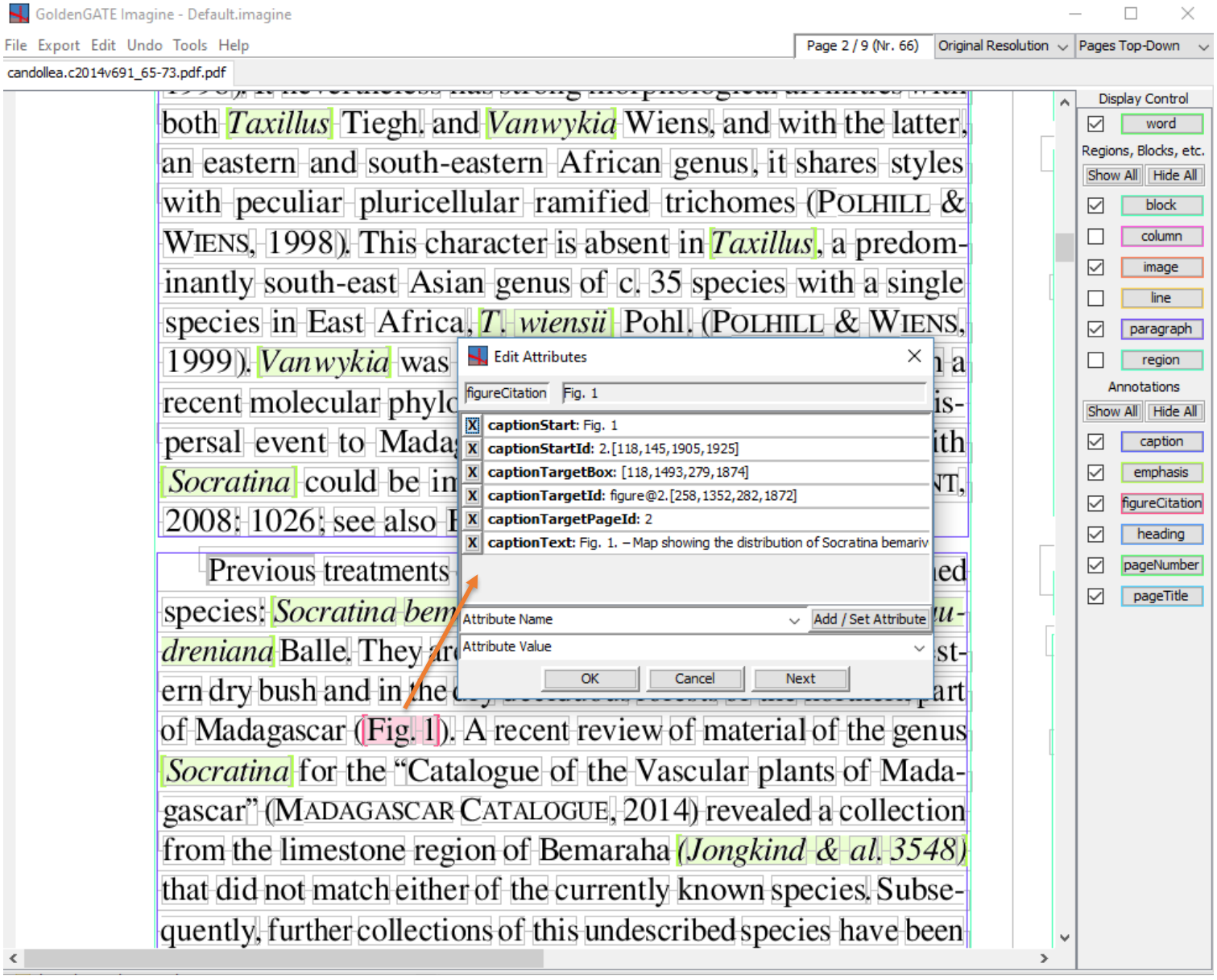
Plazi: conversion: keep the provenance



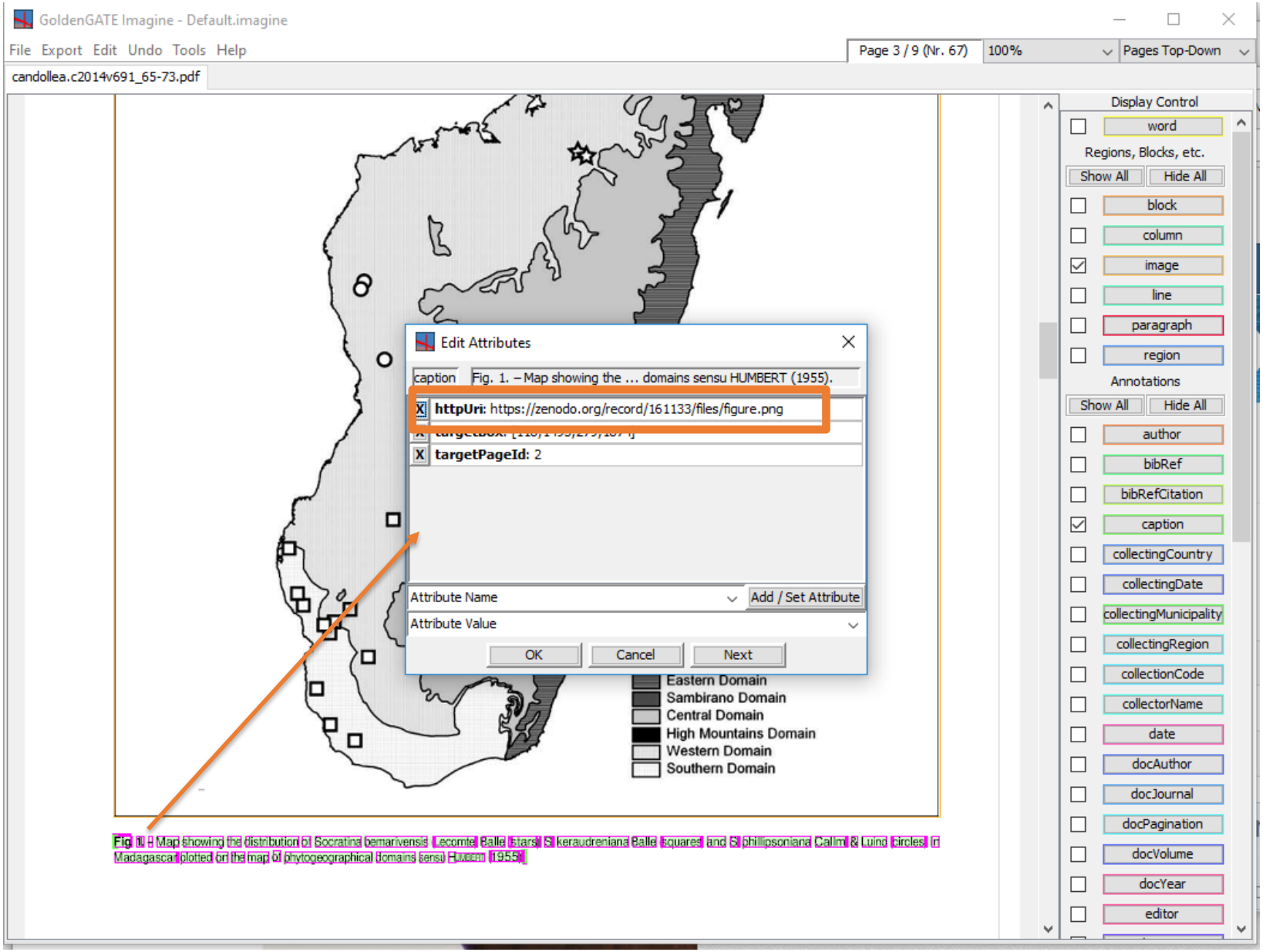
Plazi: conversion: discovering images



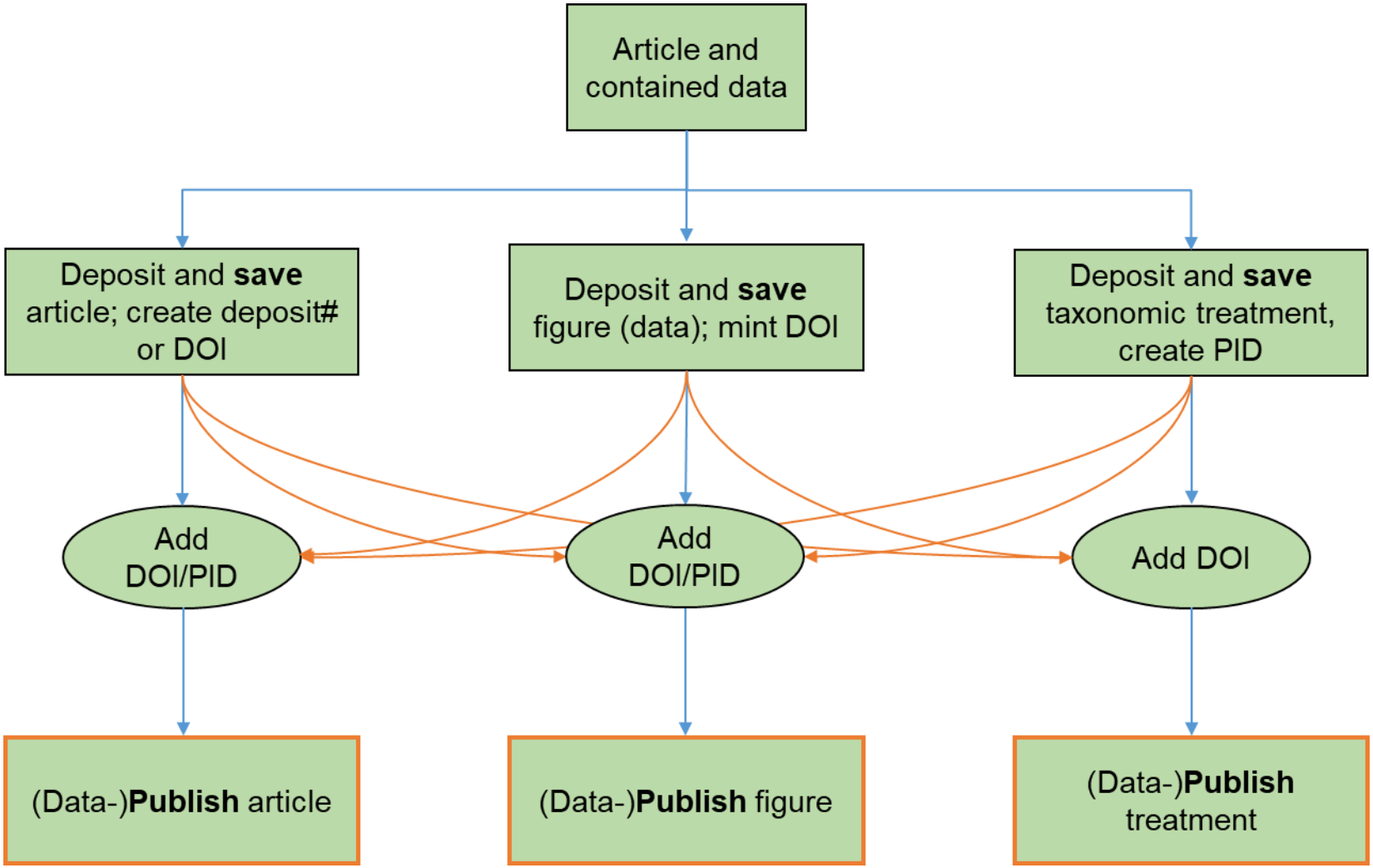
Plazi: conversion: discovering images



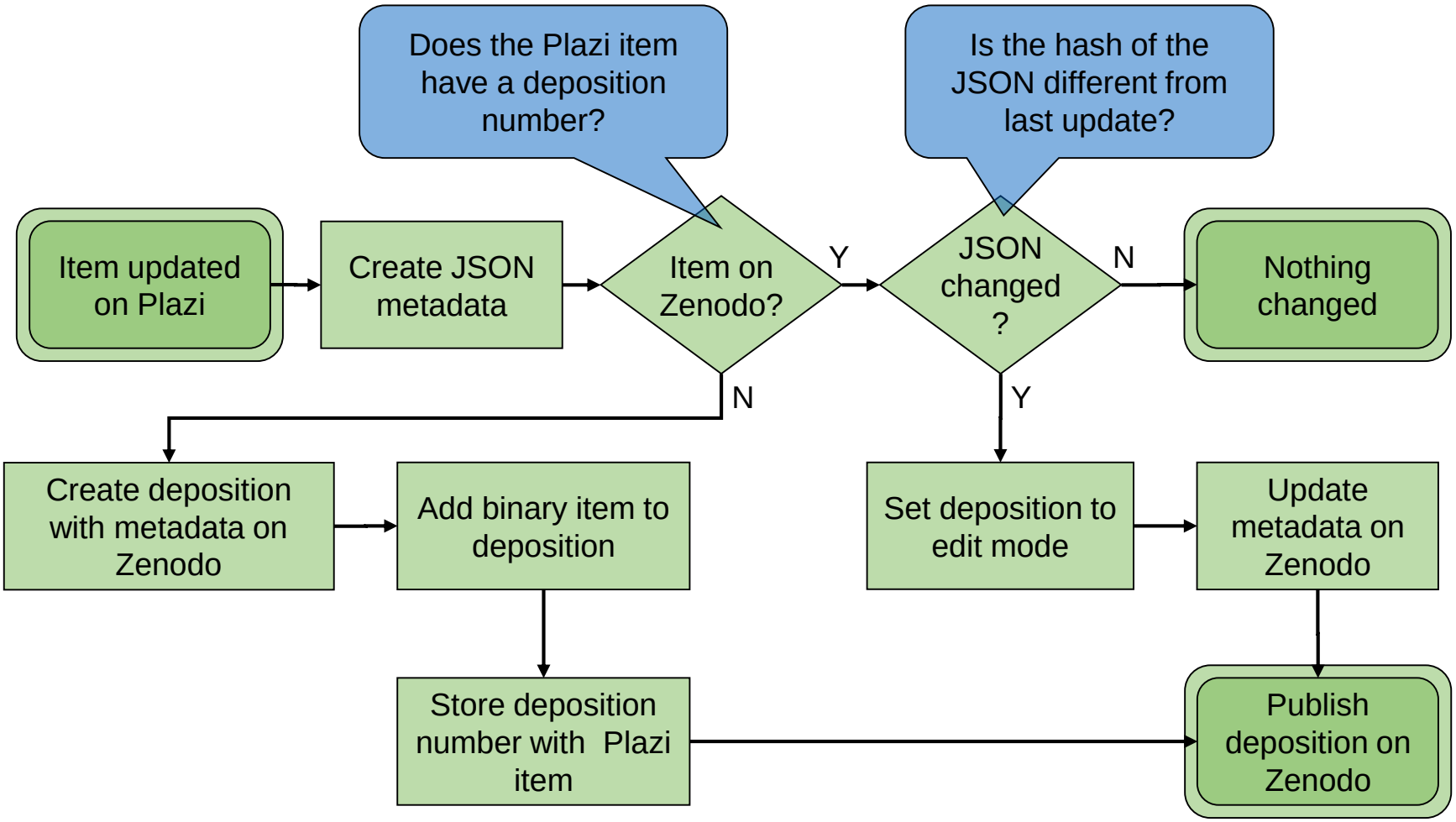
Plazi: conversion: discovering images



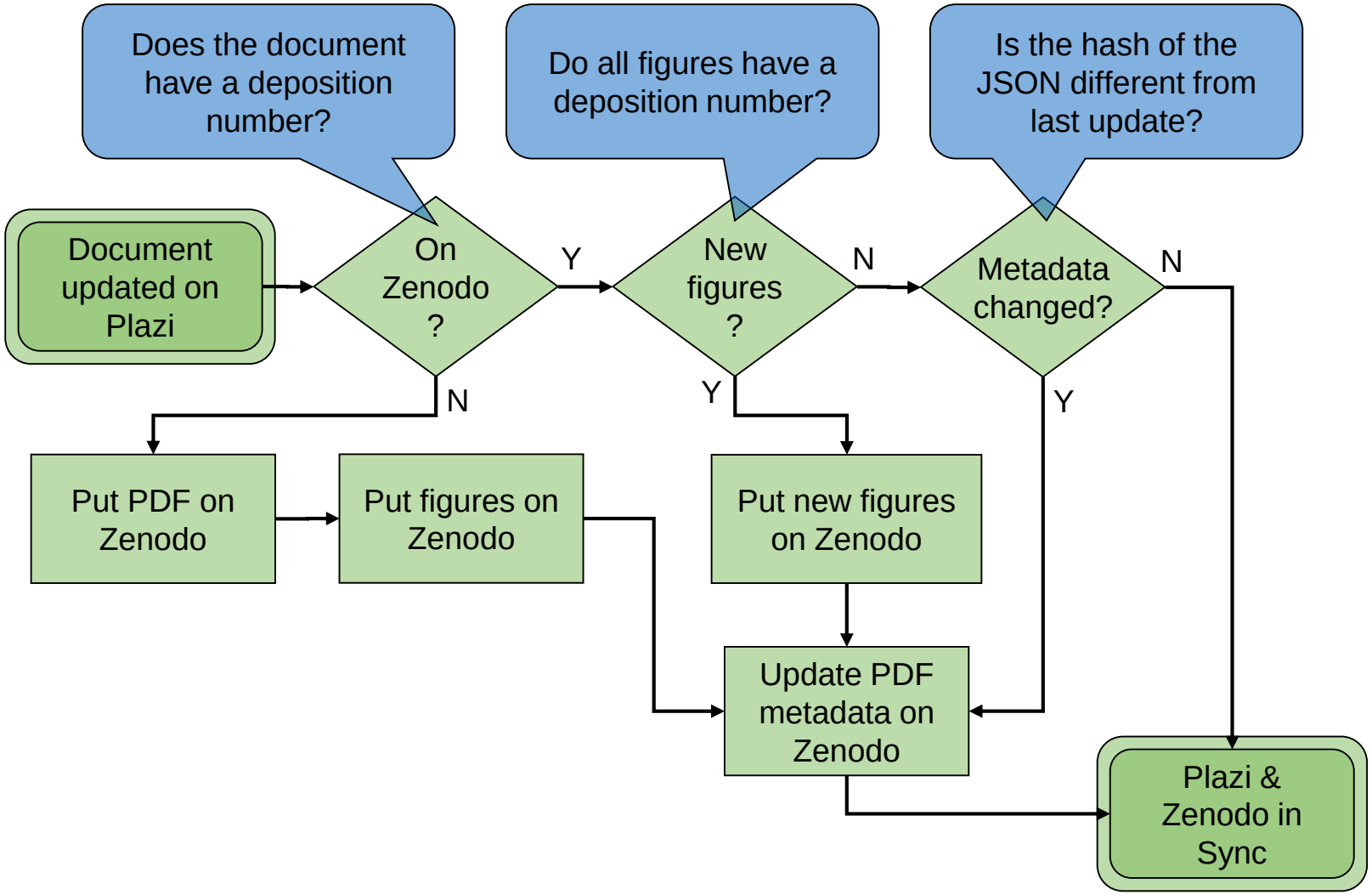
Deposition of an article and its data on BLR/Zenodo



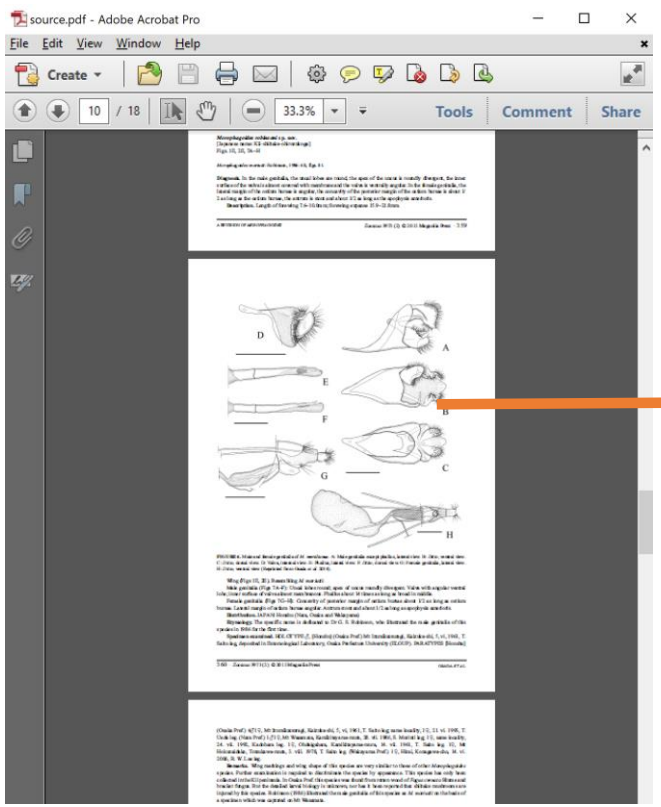
Up-Sync Process – Overall: Illustration

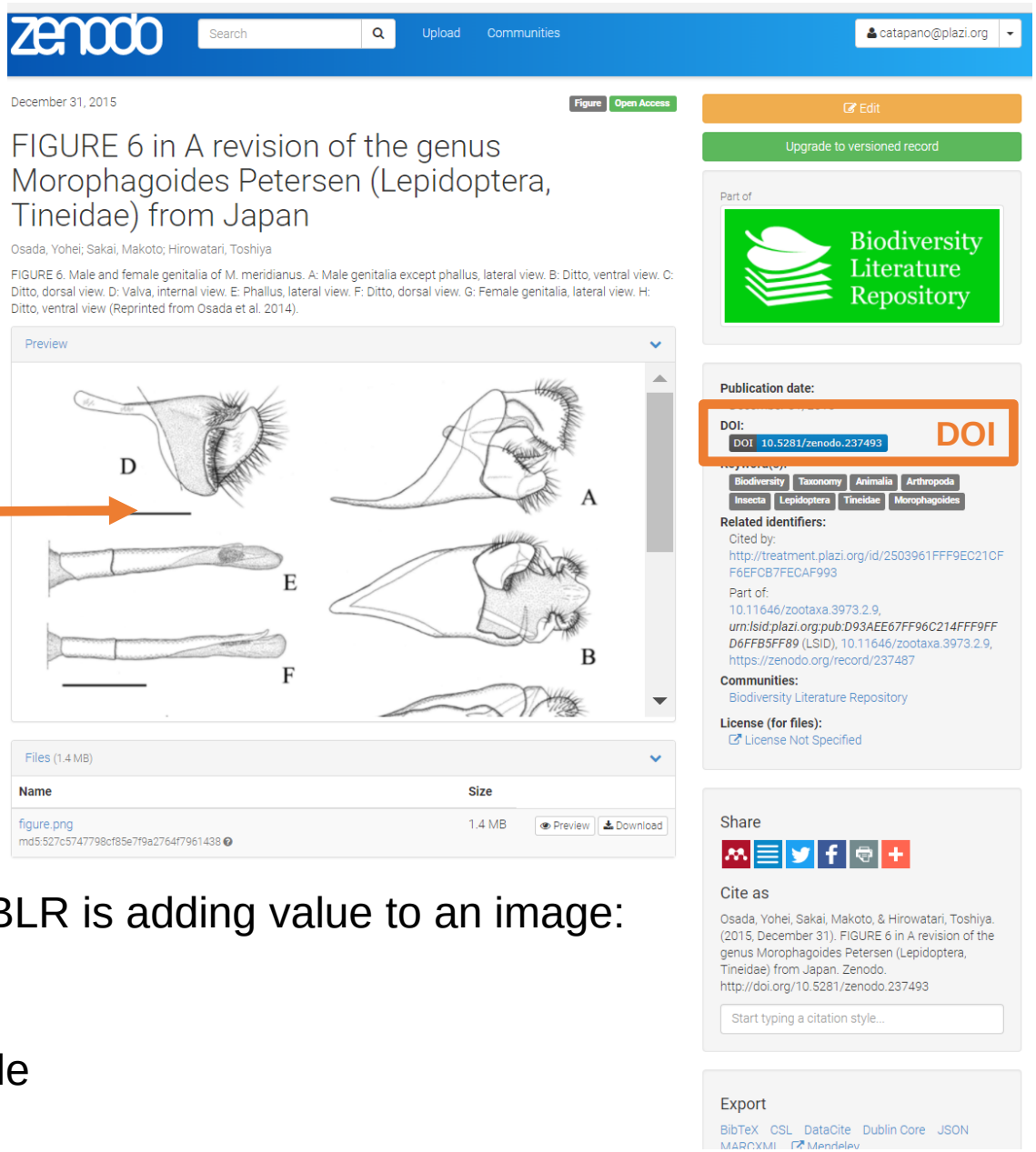


Up-Sync Process – Overall: Article



Plazi infrastructure: Biodiversty Literature Repositor

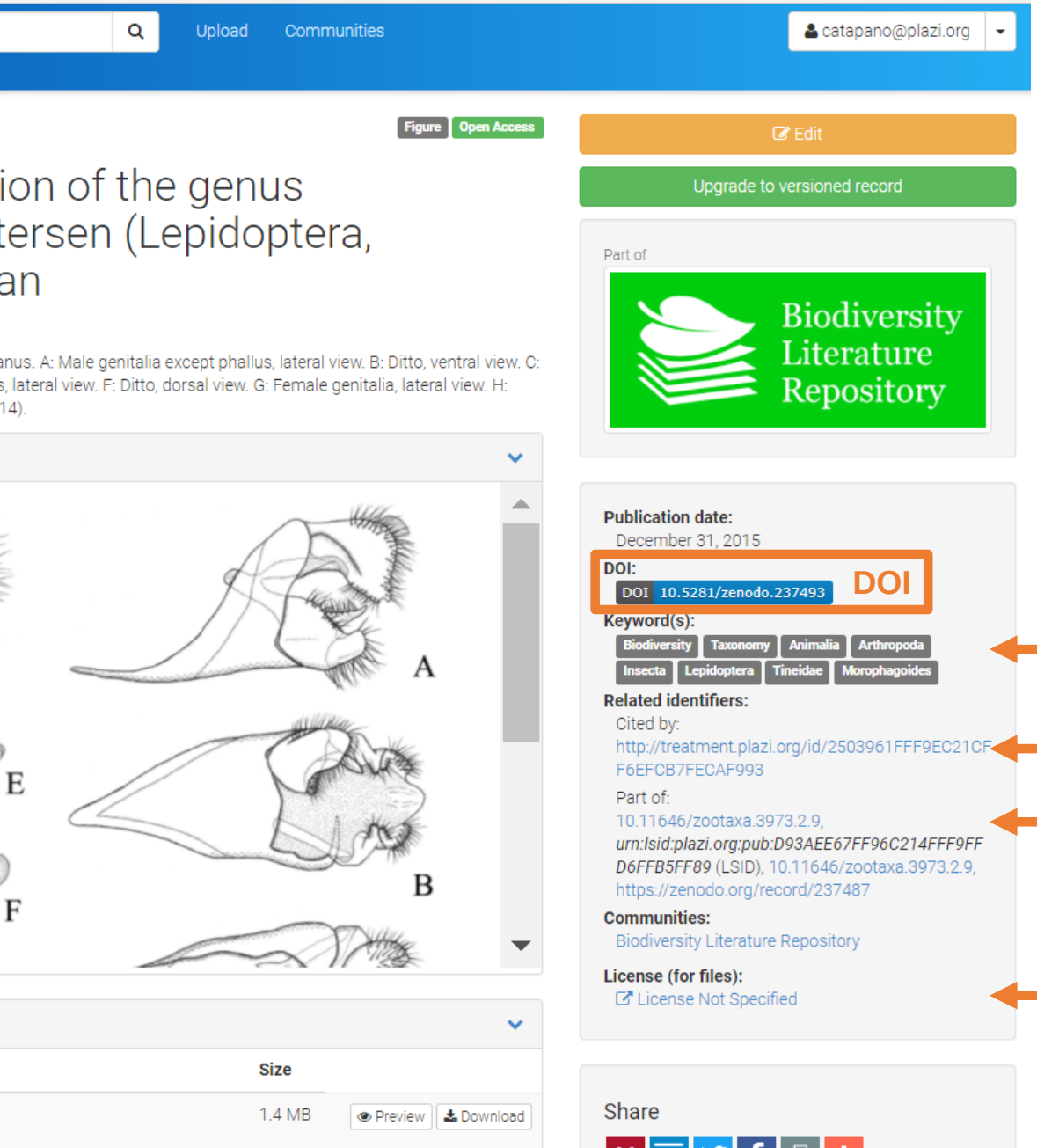




Depositing in BLR is adding value to an image:

- findable
- accessible
- interoperable
- reusable

Biodiversity Literature Repository: enhanced illustration



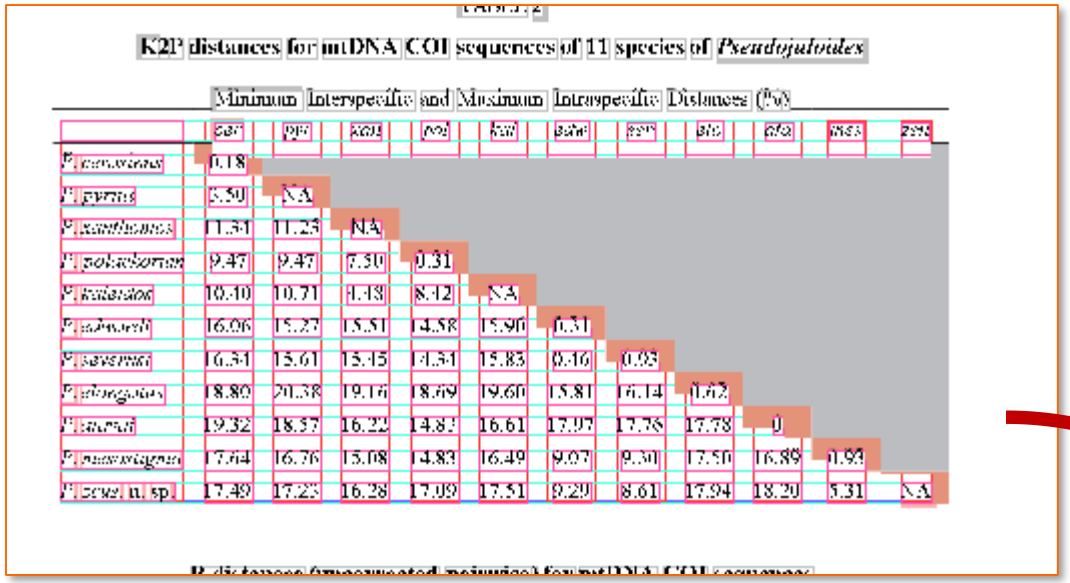
Keywords / vocabularies

Cited by

Part of

Licence

Plazi: GoldenGATE: Table extraction



K2P distances for mtDNA COI sequences of 11 species of *Pseudojulodes*

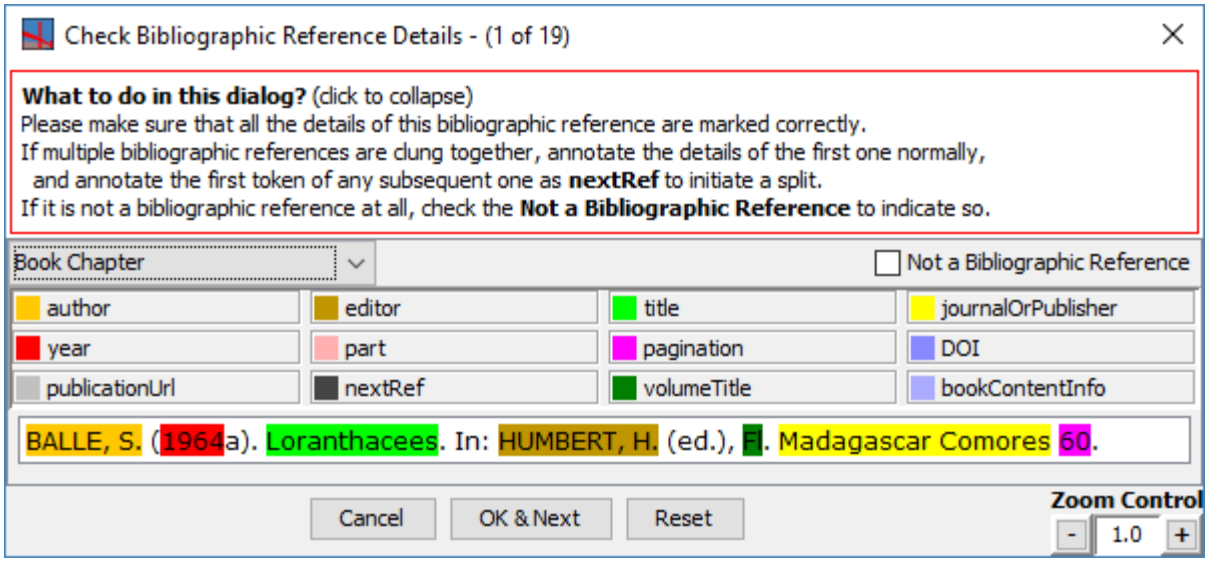
Minimum Interspecific and Maximum Intraspecific Distances (%)

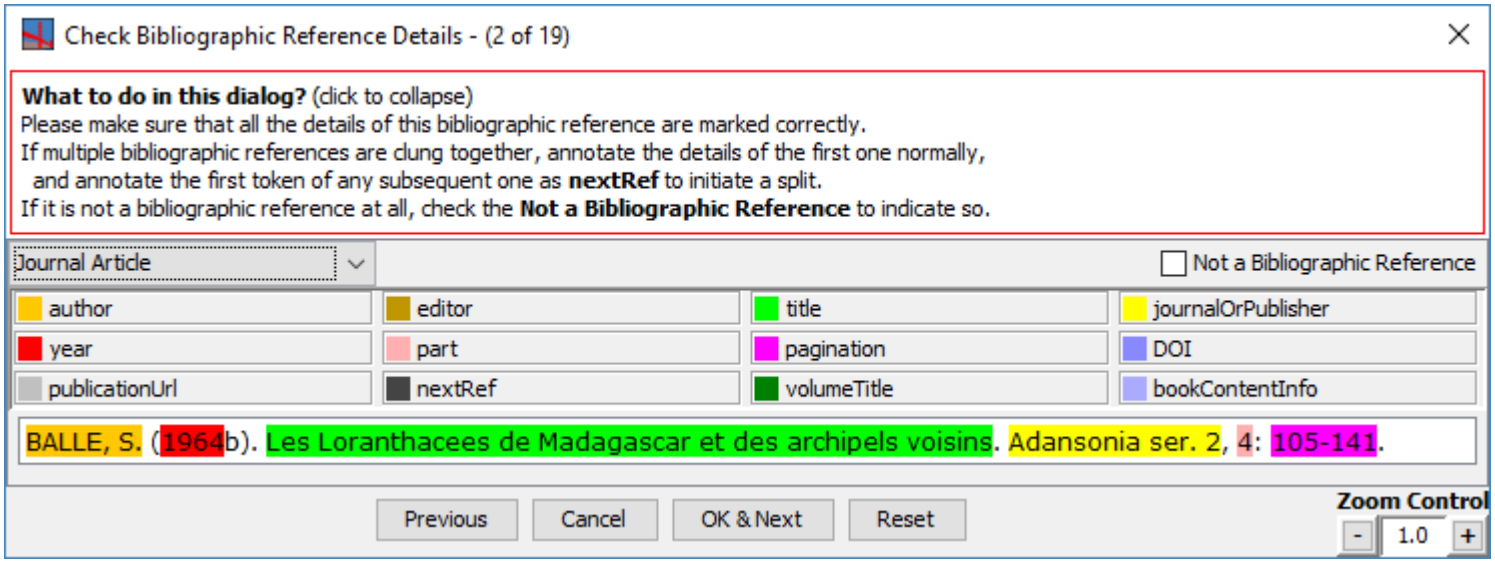
	cer	pyr	xan	pol	kal	edw	sev	elo	ata	mes	zeu
P. cerasinus	0.18										
P. pyrius	3.50	NA									
P. xanthomos	11.34	11.23	NA								
P. polackorum	9.47	9.47	7.50	0.31							
P. kaleidos	10.40	10.71	4.48	8.42	NA						
P. edwardi	16.06	15.27	15.51	14.58	15.90	0.31					
P. severnsi	16.34	15.61	15.45	14.34	15.83	0.46	0.93				
P. elongatus	18.80	20.38	19.16	18.69	19.60	15.81	16.14	0.62			
P. atavai	19.32	18.57	16.22	14.83	16.61	17.97	17.76	17.78	0		
P. mesostigma	17.64	16.76	15.08	14.83	16.49	9.07	9.30	17.50	16.89	0.93	
P. zeus, n. sp.	17.49	17.23	16.28	17.09	17.51	9.29	8.61	17.94	18.20	5.31	NA

Mark, copy and paste
(csv, text, excel)

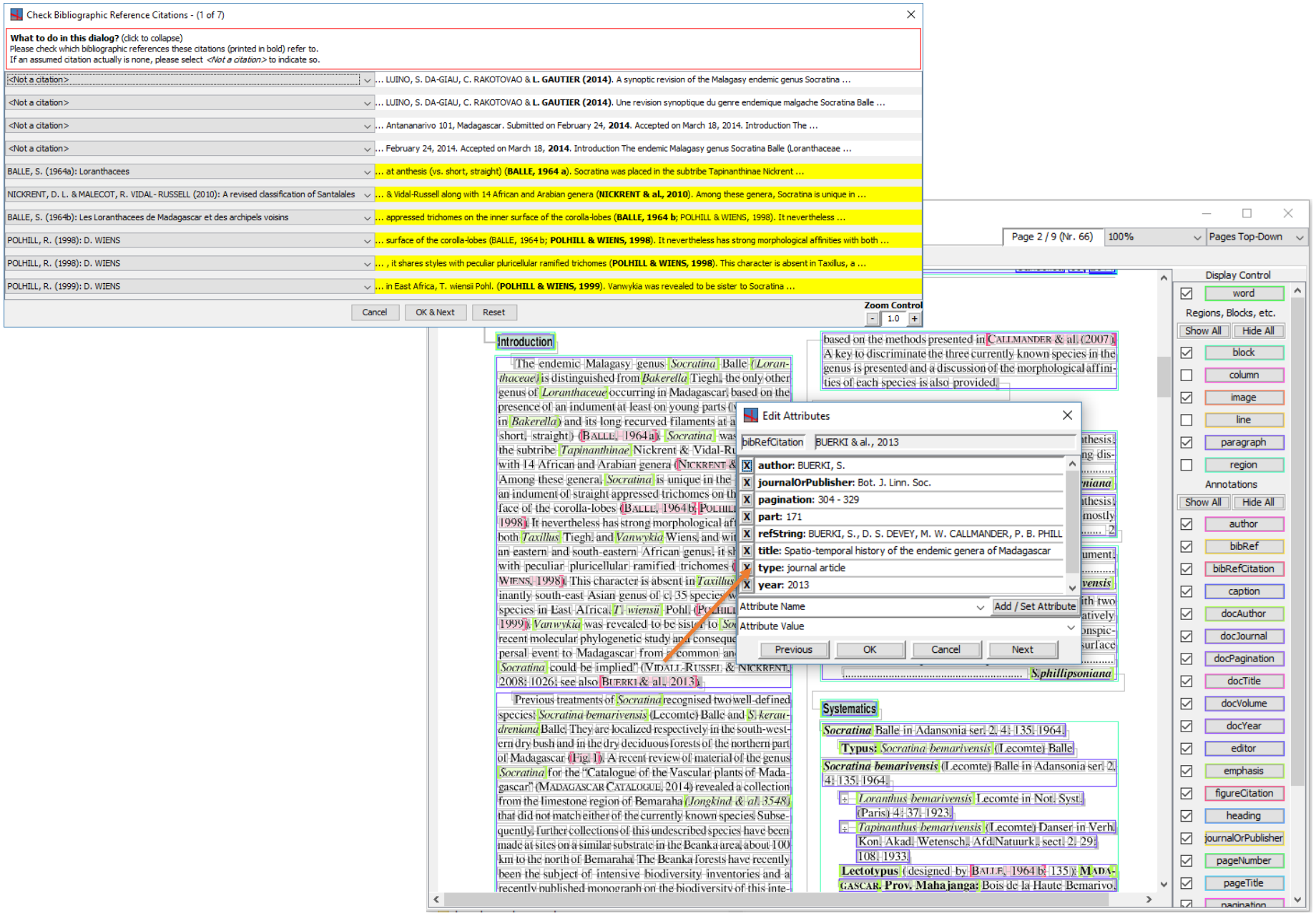
	cer	pyr	xan	pol	kal	edw	sev	elo	ata	mes	zeu
P. cerasinus	0.18										
P. pyrius	3.50	NA									
P. xanthomos	11.34	11.23	NA								
P. polackorum	9.47	9.47	7.50	0.31							
P. kaleidos	10.40	10.71	4.48	8.42	NA						
P. edwardi	16.06	15.27	15.51	14.58	15.90	0.31					
P. severnsi	16.34	15.61	15.45	14.34	15.83	0.46	0.93				
P. elongatus	18.80	20.38	19.16	18.69	19.60	15.81	16.14	0.62			
P. atavai	19.32	18.57	16.22	14.83	16.61	17.97	17.76	17.78	0		
P. mesostigma	17.64	16.76	15.08	14.83	16.49	9.07	9.30	17.50	16.89	0.93	
P. zeus, n. sp.	17.49	17.23	16.28	17.09	17.51	9.29	8.61	17.94	18.20	5.31	NA

Plazi: conversion

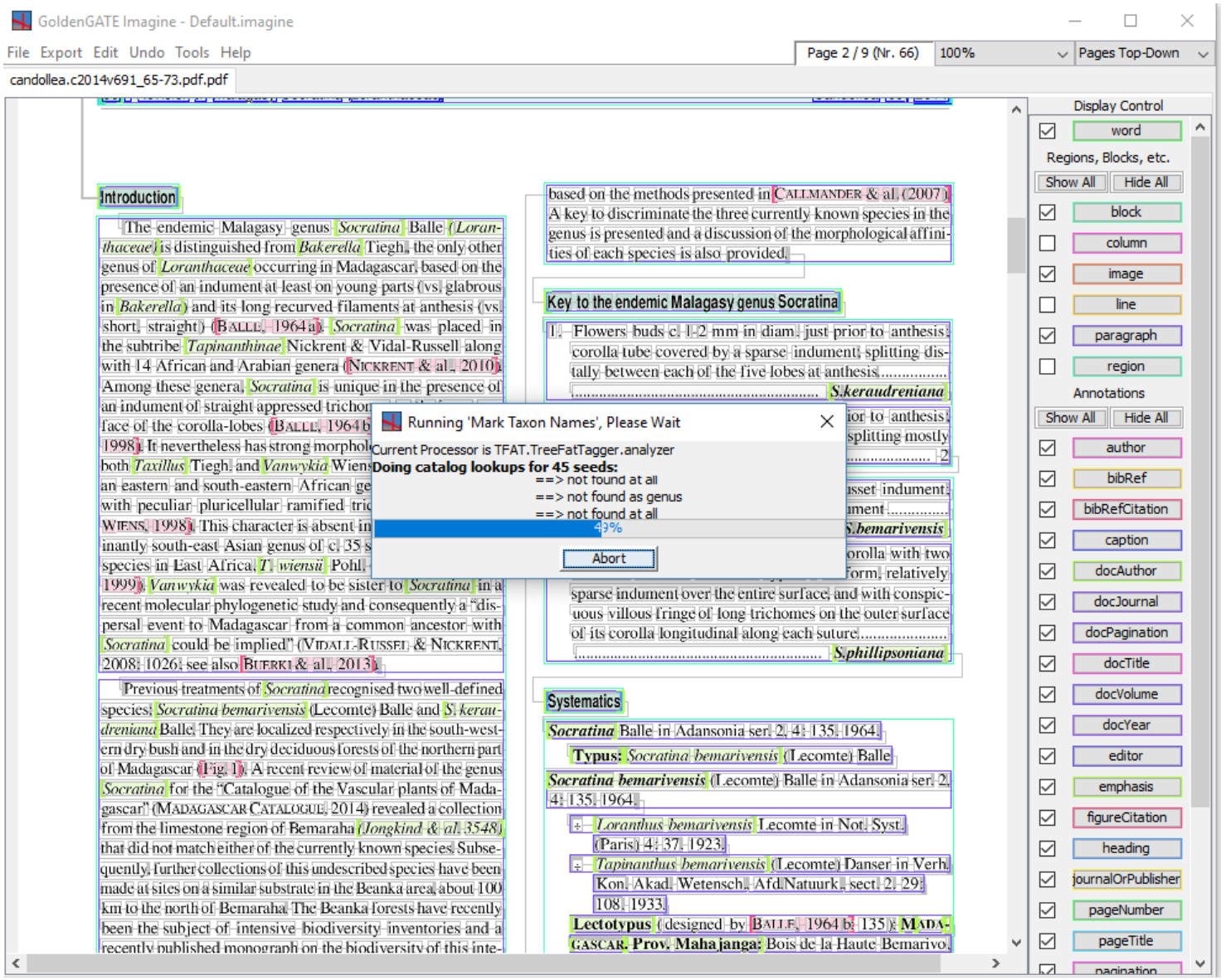




Plazi: conversion

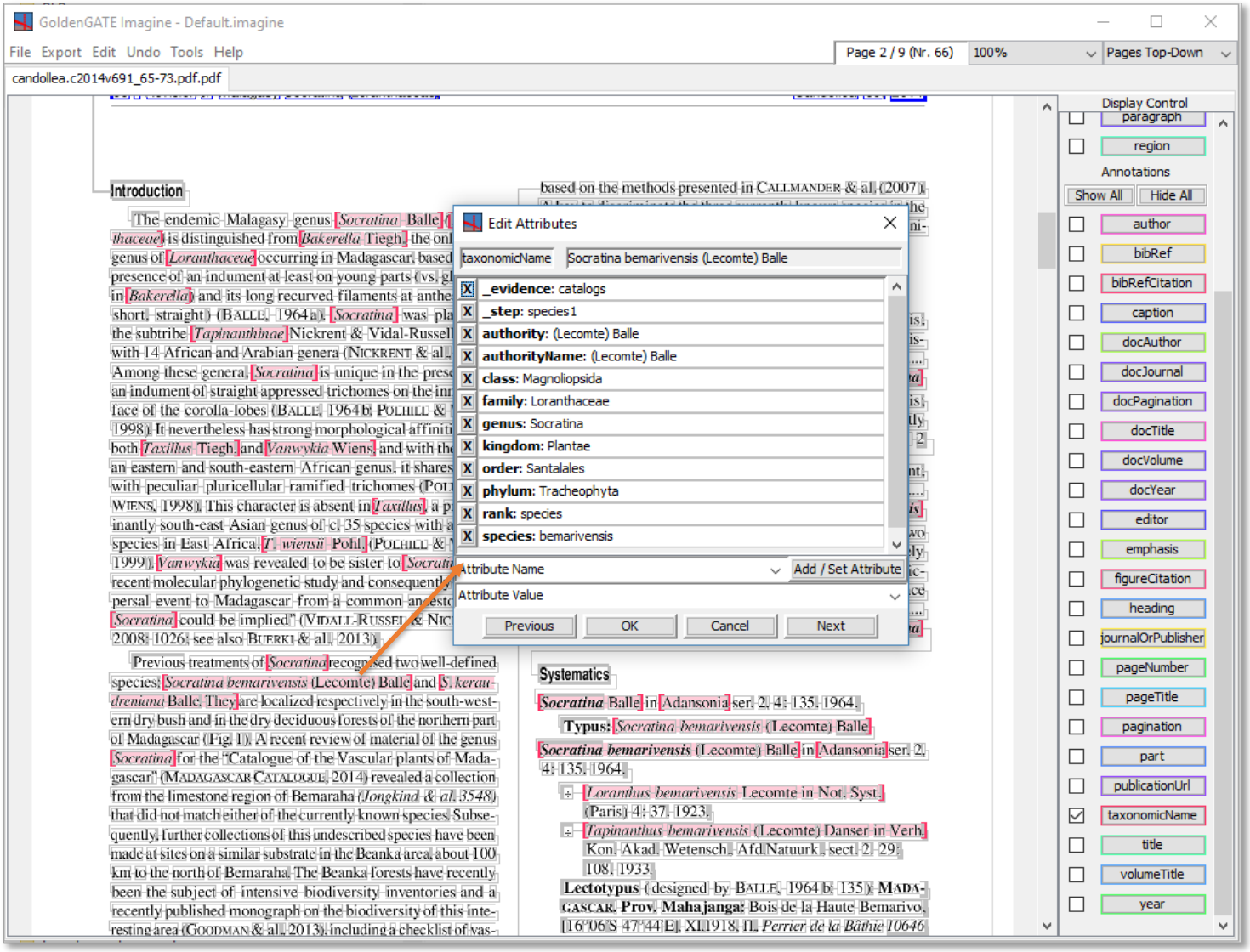


Plazi: conversion

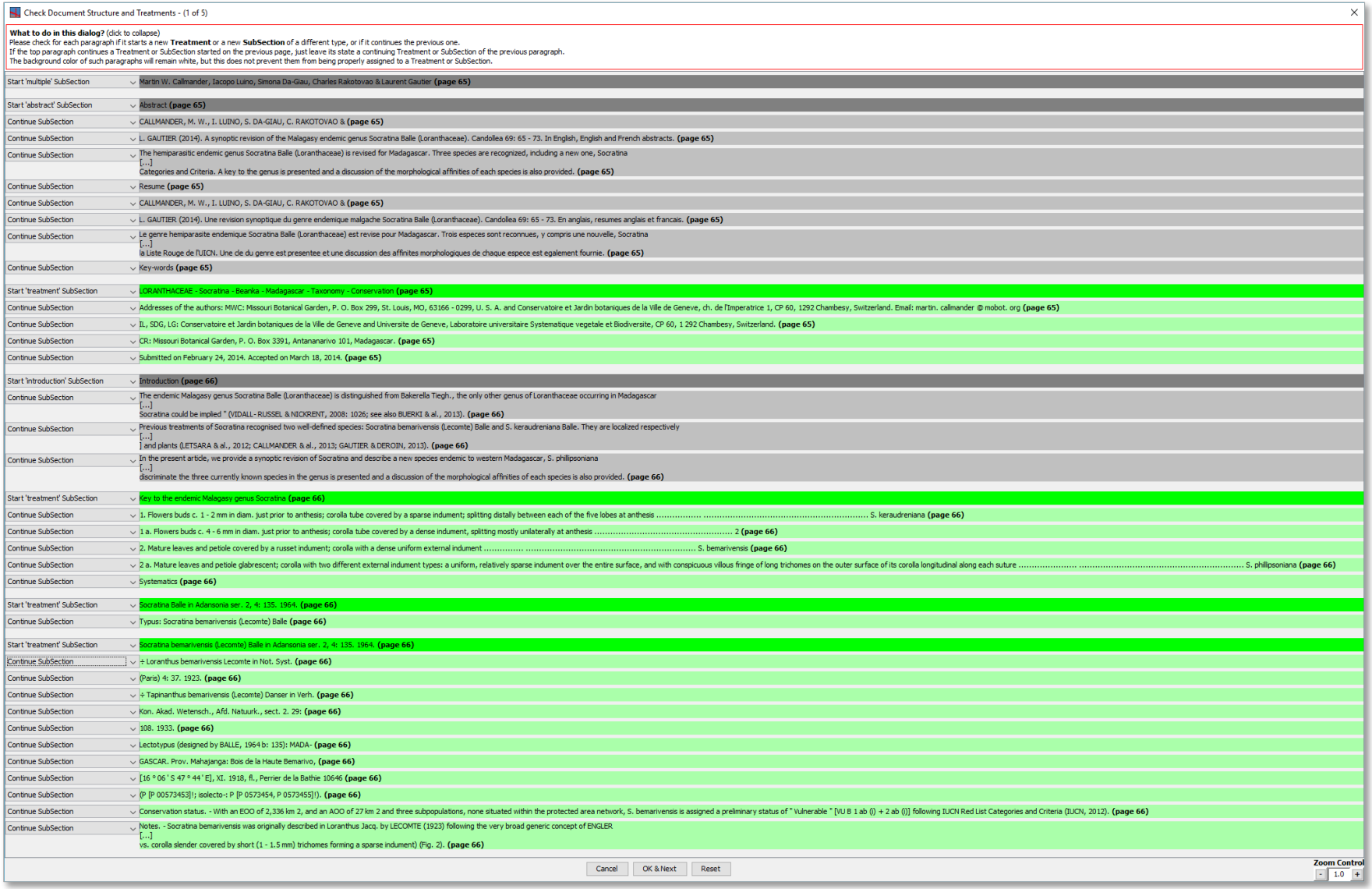


Plazi: conversion

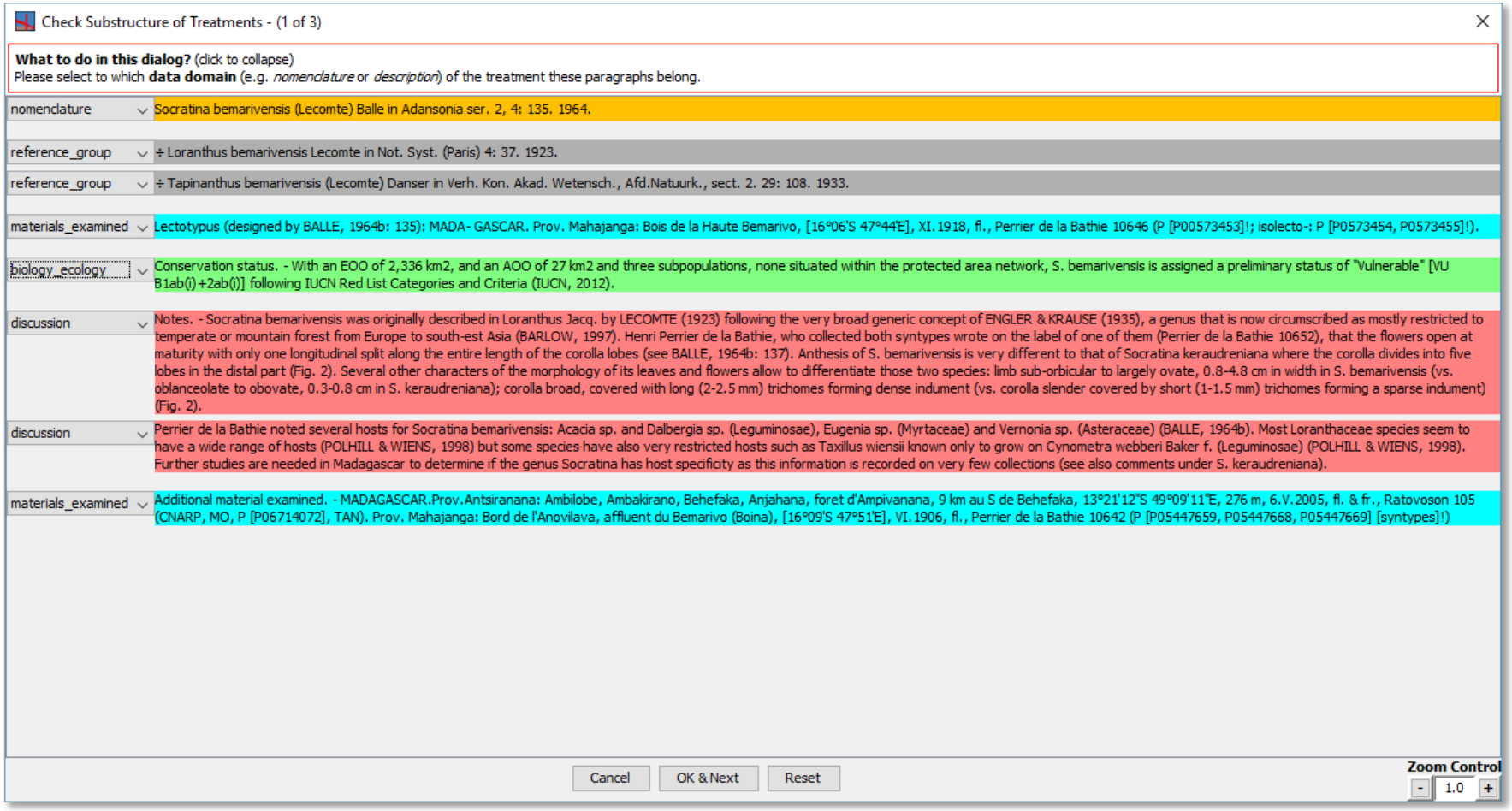
Retrieving higher ranks from *GBIF* and *Catalogue of Life* (external resources)



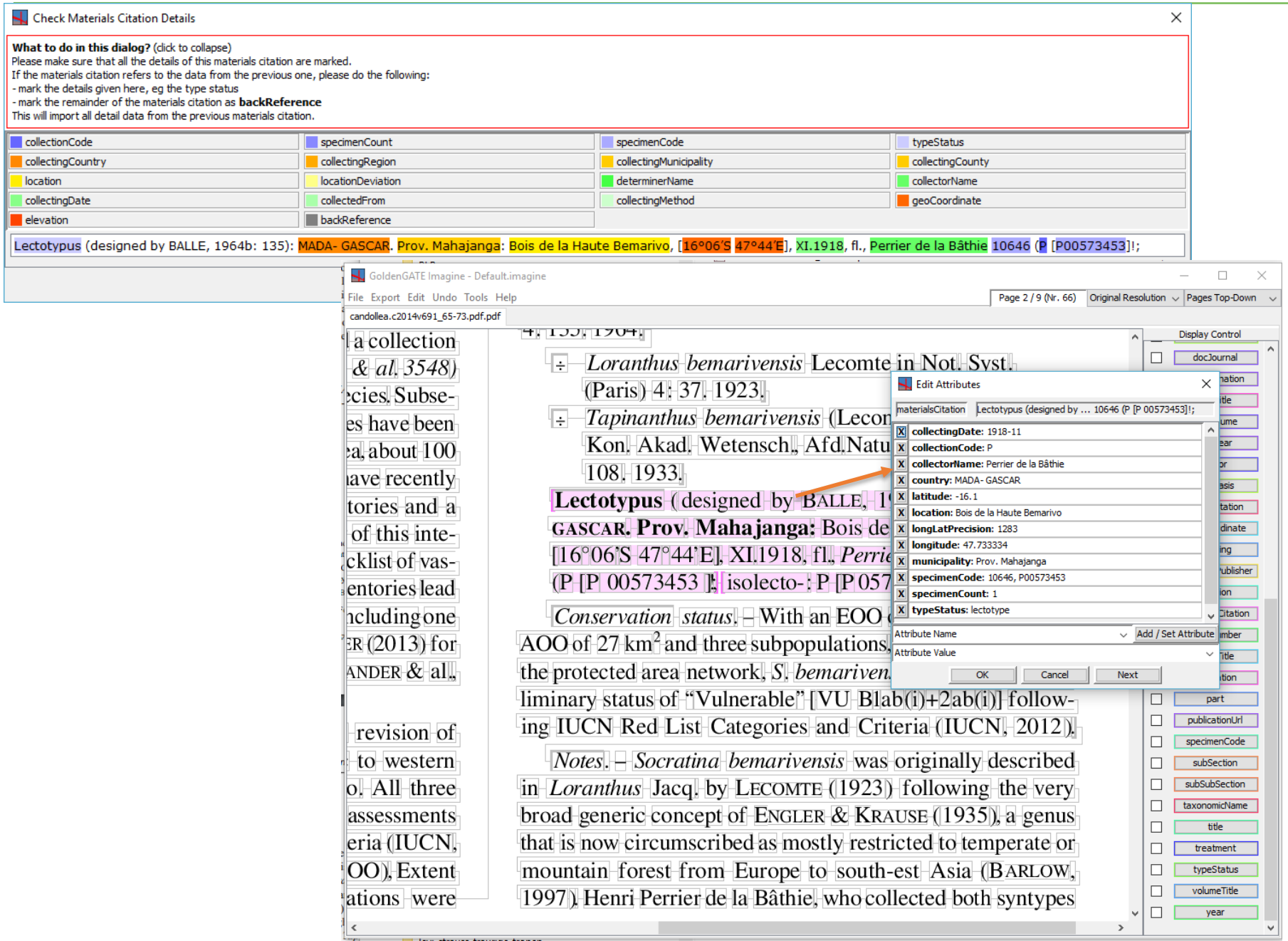
Plazi: conversion

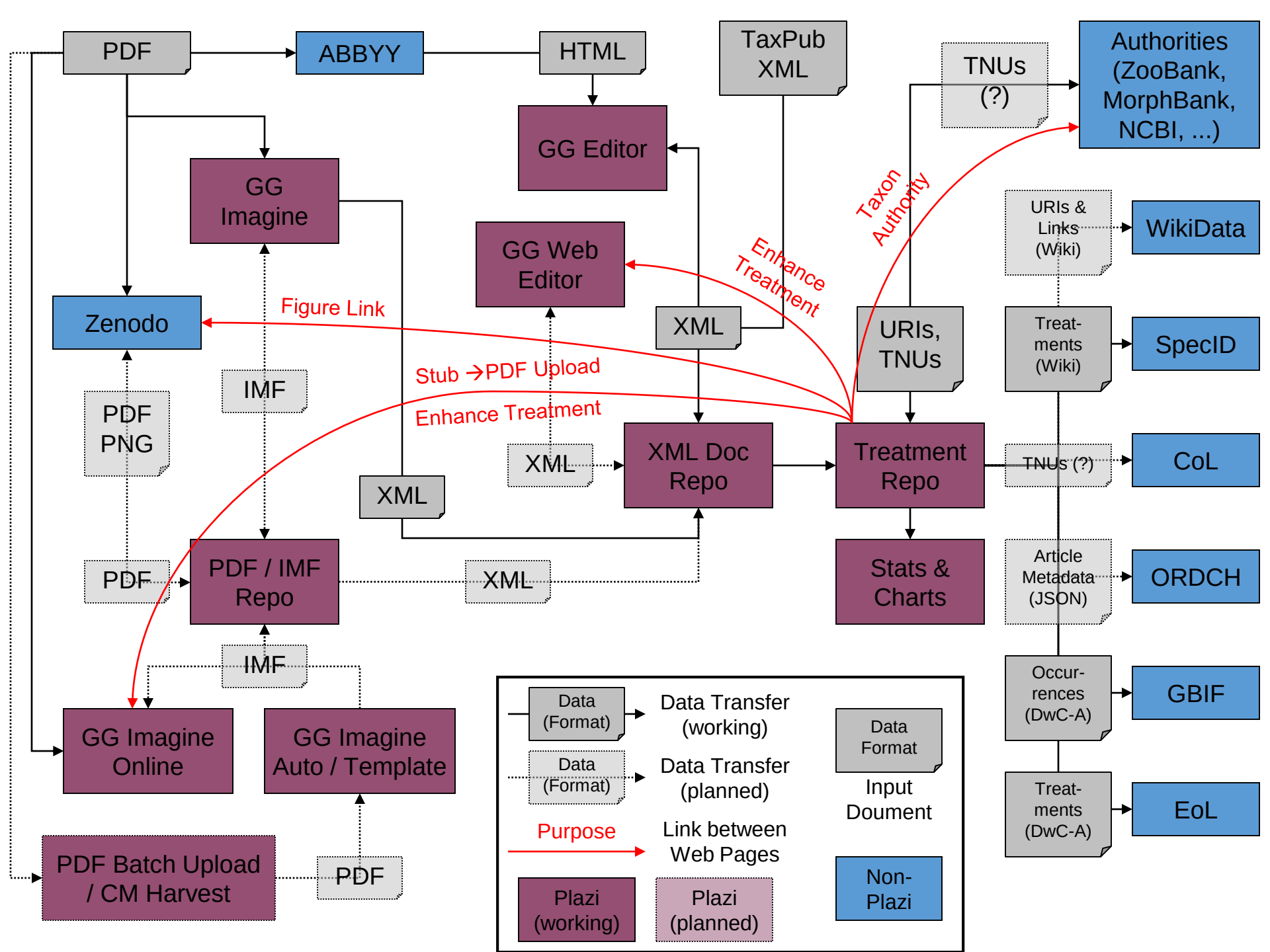


Plazi: conversion



Plazi: conversion




PDF
ABBYY
HTML
TaxPub XML
TNUs (?)
Authorities (ZooBank, MorphBank, NCBI, ...)
GG Editor
GG Imagine
GG Web Editor
Taxon Authority
URIs & Links (Wiki)
WikiData
Enhance Treatment
Figure Link
Zenodo
XML
URIs, TNUs
Treat-ments (Wiki)
SpecID
Stub →PDF Upload
IMF
PDF PNG
Enhance Treatment
XML
XML Doc Repo
Treatment Repo
TNUs (?)
CoL
XML
PDF
PDF / IMF Repo
XML
Stats & Charts
Article Metadata (JSON)
ORDCH
IMF
Occur-rences (DwC-A)
GBIF
GG Imagine Online
GG Imagine Auto / Template
Treat-ments (DwC-A)
EoL
PDF Batch Upload / CM Harvest
PDF
Data (Format)
Data Transfer (working)
Data Format
Data (Format)
Data Transfer (planned)
Input Doument
Purpose
Link between Web Pages
Plazi (working)
Plazi (planned)
Non-Plazi

Plazi workflow is based on Swiss copyright

- Copying of excerpts for internal use is a permitted act
- Individual articles are considered excerpts of journals
- Distribution of copies in institutions is a permitted act
- Data (text, figures, images) ≠ copyright protected work

→ Standortvorteil for Swiss science

(Agosti&Egloff, 2009).