PLAZI

Plazi: persistent and openly accessible digital taxonomic literature

Donat Agosti & Terry Catapano
Plazi
http://plazi.org

SIB Text Mining for Biology Meeting
June 4, 2015

PLAZI

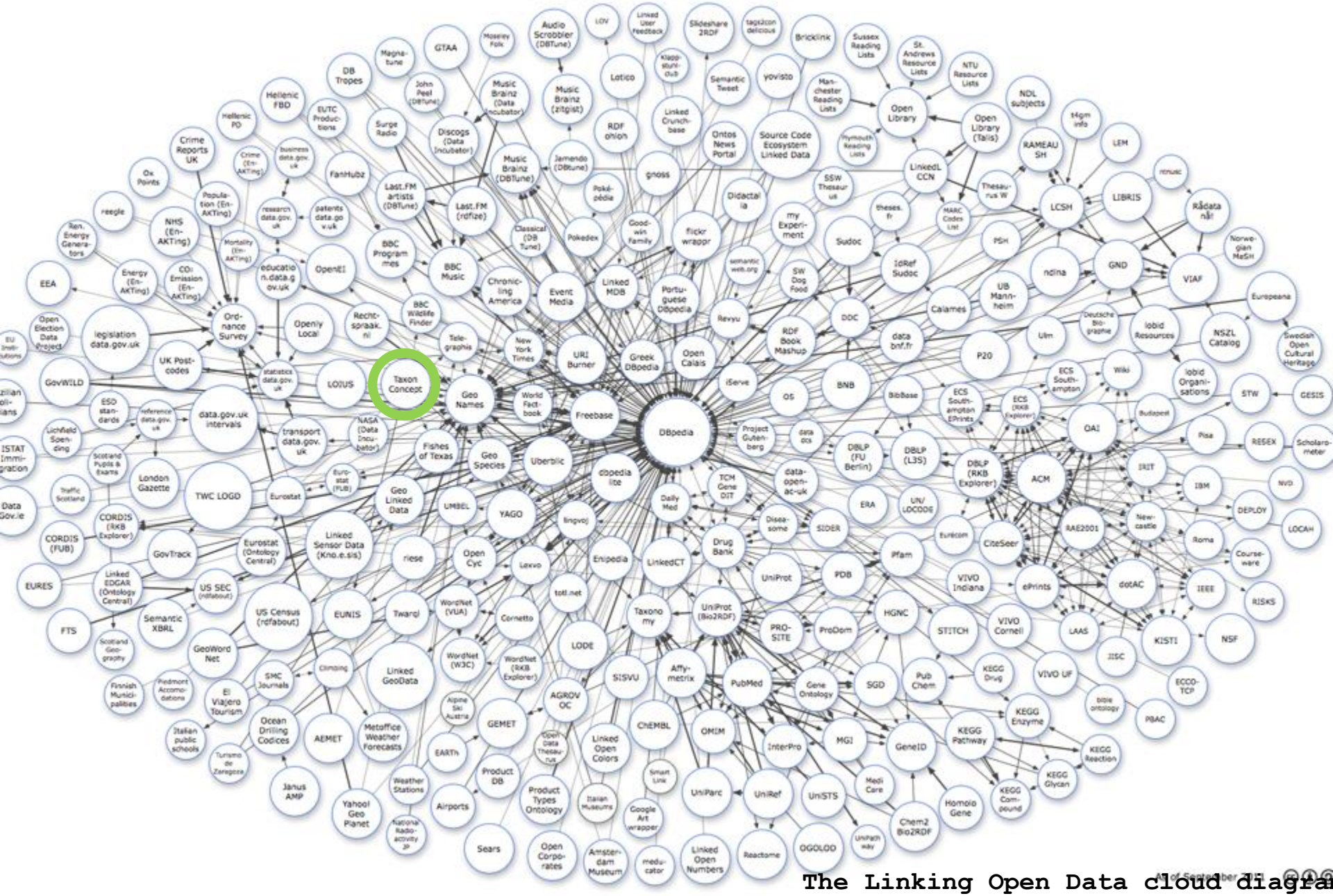
taking care of freedom

- Swiss based international NGO and SME
- Founded in 2008
- Mission to foster (Open Access) Linked Open (scientific) Data
- EU, industry, and volunteer support
- Membership global

PLAZI

taking care of freedom

Build as catalyst a TreatmentBank that includes direct access to 1 million citable treatments and data, related metadata and digital copies of the source publications



The Linking Open Data cloud diagra

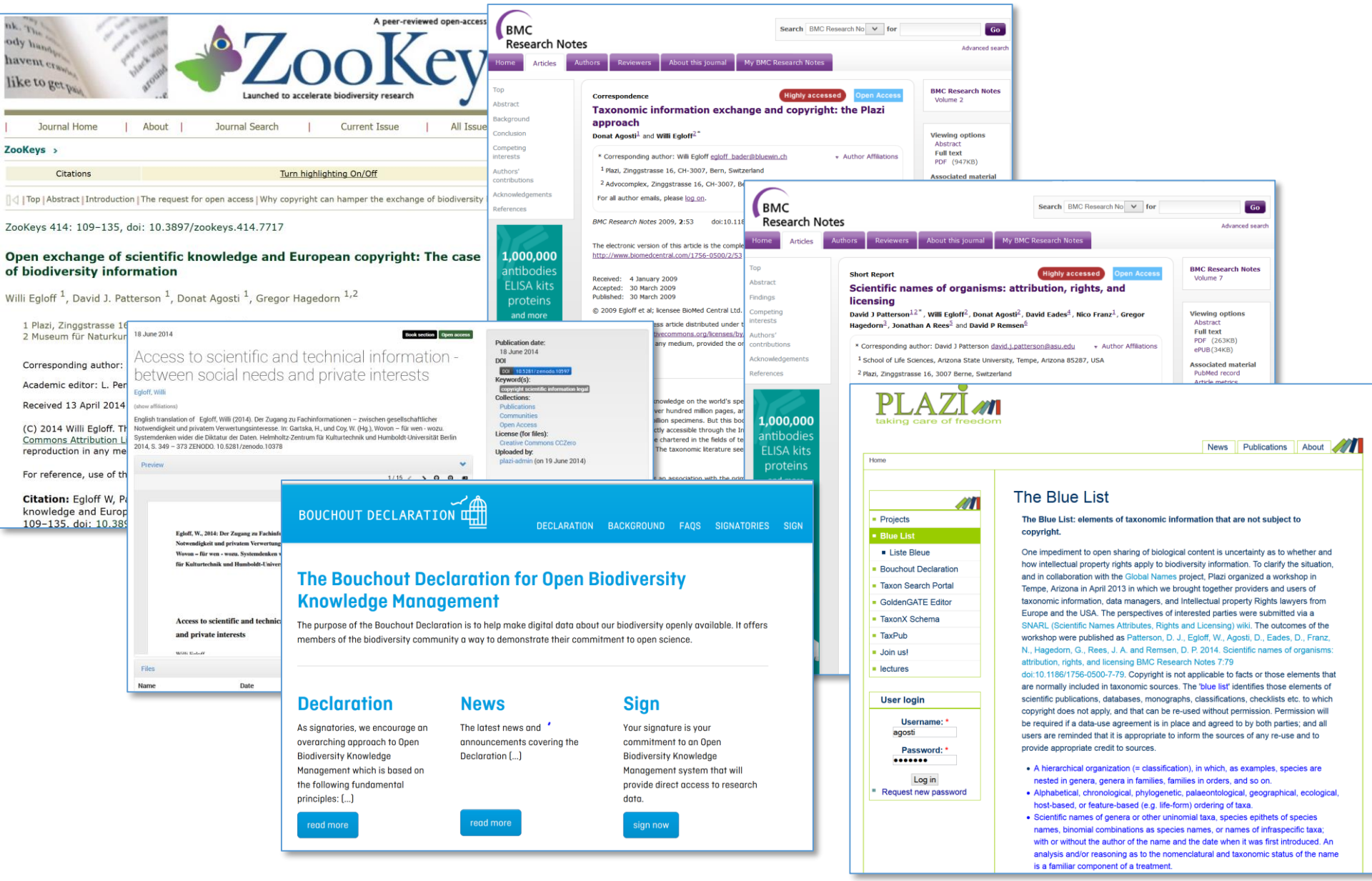
Five pillars:

- Legal advice
- technical innovations and solutions
- maintenance of a treatment repository and Biowikifarm
- consultancy and services
- advocacy

Development and maintenance of schemas

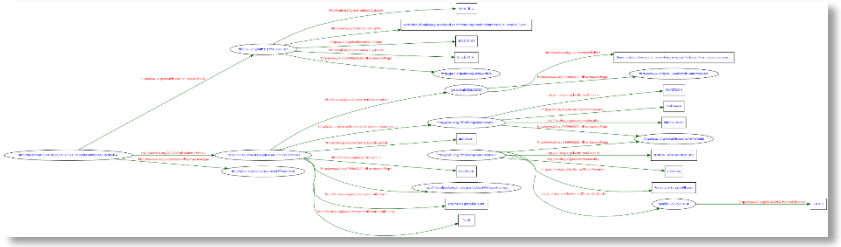
Taxpub (together with NIH/Pensoft)	TaxonX
DTD	Schema
Prospective publications	Legacy publications
Constraint	loose
Derivative of JATS	independent
Self-contained	Allows import of other schemas

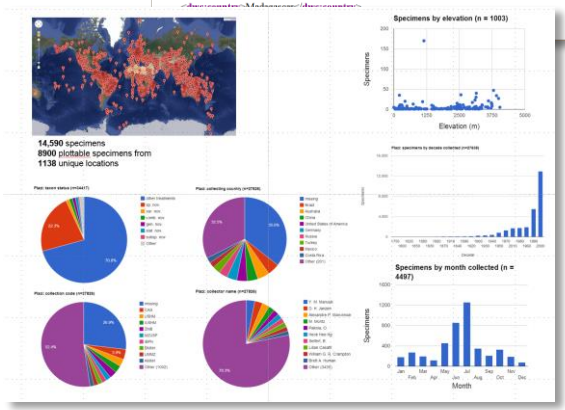
Publication, one of our footprints



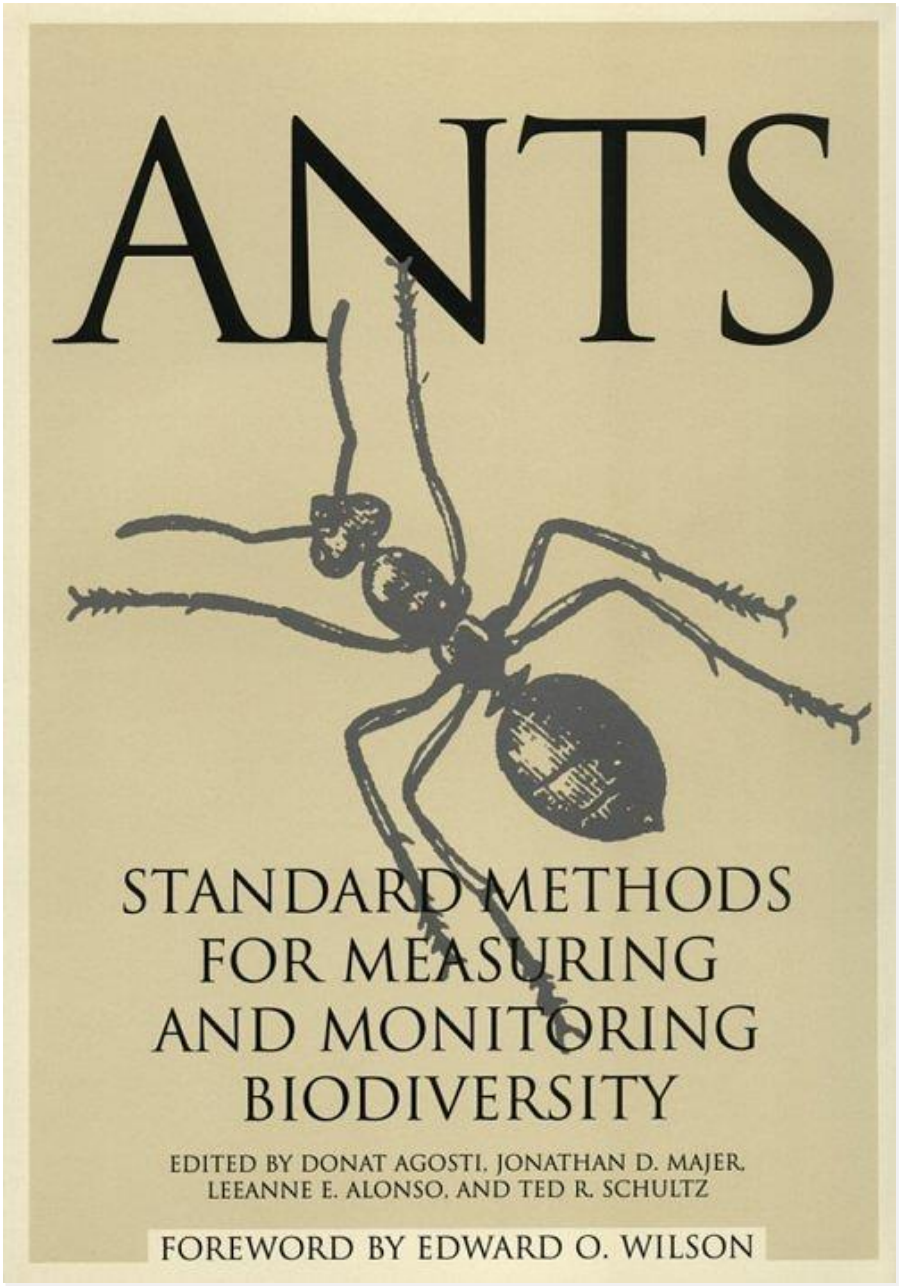
Our dream: Giant Global Species Graph

Legal
Social
Technical
Ontologies
Infrastructure





Our contribution for a better understanding of biodiversity



Compare biodiversity over space, time, collector, habitat

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

The origin

Idea

1995/2000

Our contribution for a better understanding of biodiversity

Online catalogue

comments on this story

Stories by subject

- Technology

Stories by keywords

- ant
- web site
- database
- antbase
- entomology

This article elsewhere

Blogs linking to this article

Add to Connotea

Add to Digg

Add to Furl

Add to Newsvine

Add to Del.icio.us

Add to Twitter

Published online 14 March 2002 | Nature | doi:10.1038/news020311-9

News

Ants make web

Formidable catalogue puts army of ants online.

Tom Clarke

Researchers can now crawl through data on 11,000 ant species.

© *Antbase*

After four years of cooperation and tenacity worthy of their quarry, ant experts have completed Antbase, a centralized, online information resource cataloguing all 11,000 known species of ant.

Its creators hope that Antbase will one day become the ant equivalent of GenBank, the public database of genetic sequence data, and a boon not just for ant specialists, but for entomologists and ecologists of every kind.

Ant taxonomists Donat Agosti of the American Museum of Natural History in New York and Norman Johnson of Ohio State University in Columbus built Antbase using funds from general research budgets and small institutional grants. Having compiled a full list of ant names, they hope their homespun project will now win long-term funding and a permanent home. "It needs to be institutionalized," says Agosti.

Recent activity

most recent

- Million-dollar problem cracked? 10 August 2010
- Power struggle hits Swedish institute 10 August 2010
- Upbeat oil report questioned 10 August 2010
- France digs deep for nuclear waste 10 August 2010
- Birds flock online 10 August 2010

commented

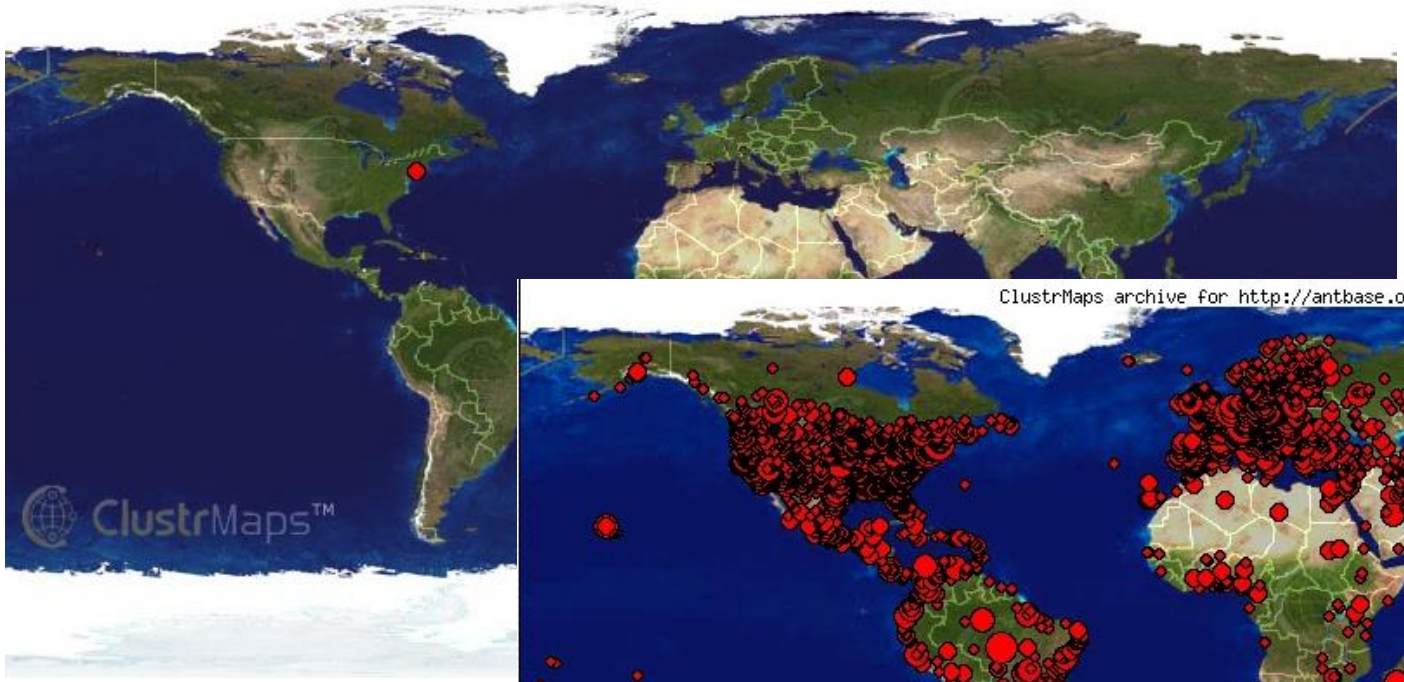
- Ecology: A world without mosquitoes • 25 comments 21 July 2010
- Physicists get political over Higgs • 13 comments 04 August 2010
- A solar salamander • 13 comments 30 July 2010
- Quantum mechanics flummoxes physicists again • 13 comments 22 July 2010
- Report maps perils of warming • 12 comments 19 July 2010

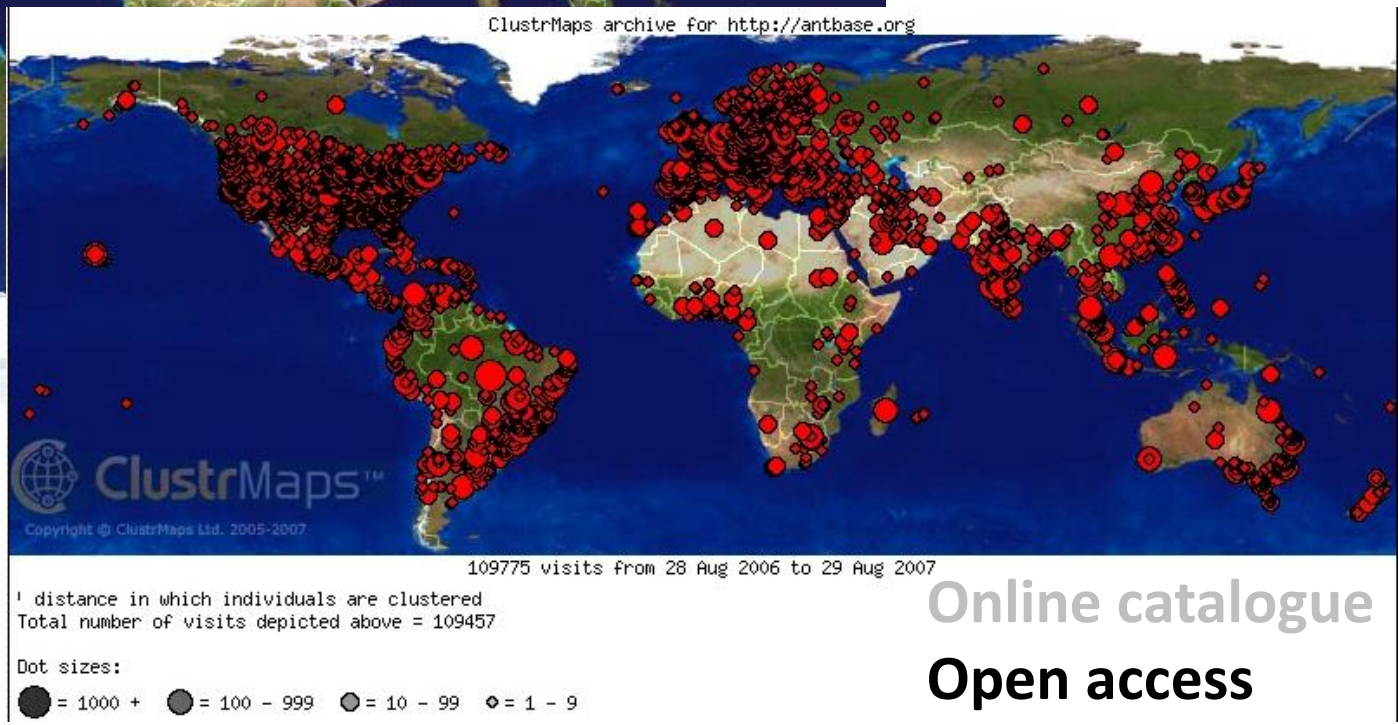
Related stories

Online catalogue 2004

Our contribution for a better understanding of biodiversity

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).

Online catalogue

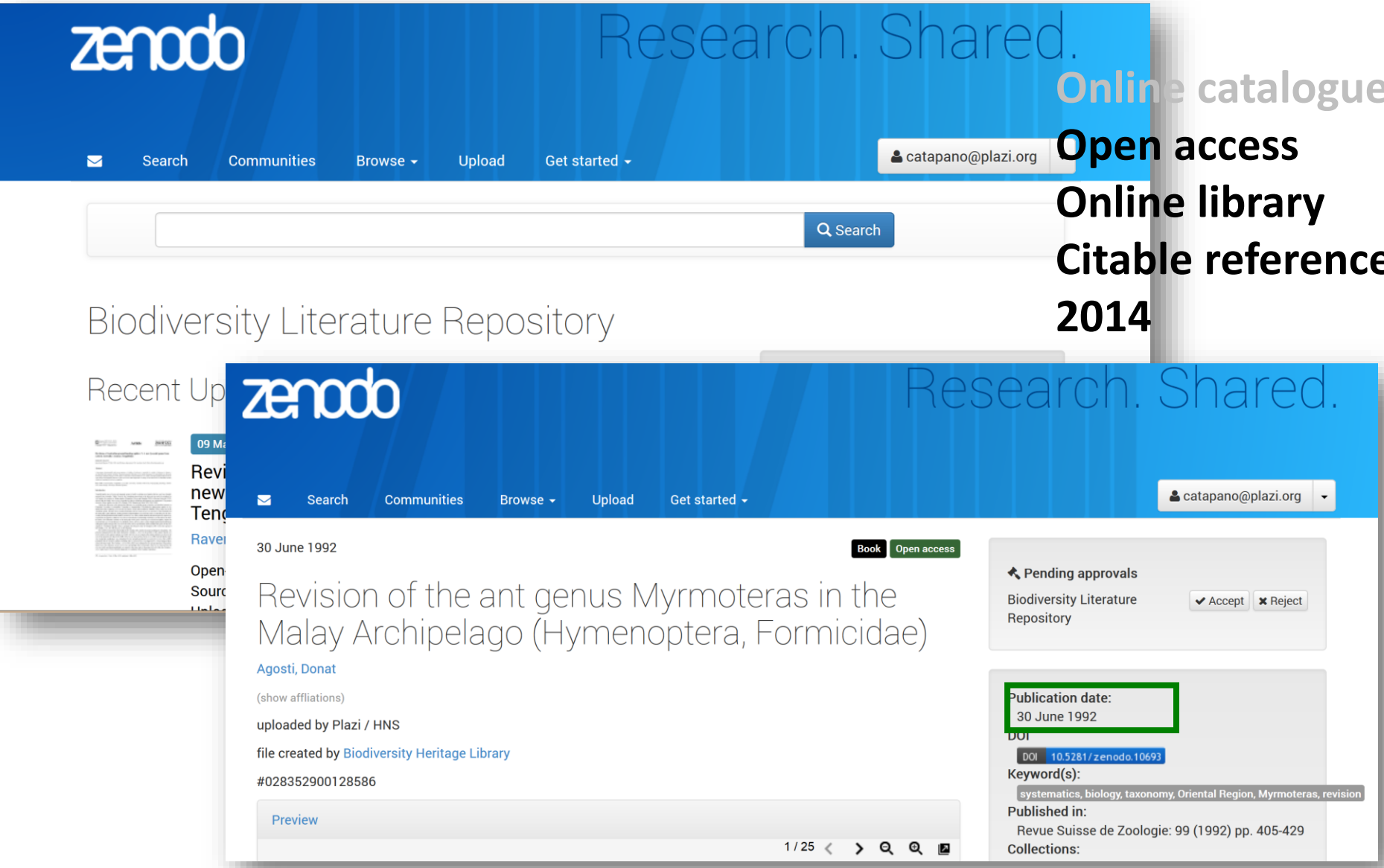
Open access

Online library

2004

Our contribution for a better understanding of biodiversity

Access, archive, DOI



Online catalogue

Open access

Online library

Citable references

2014

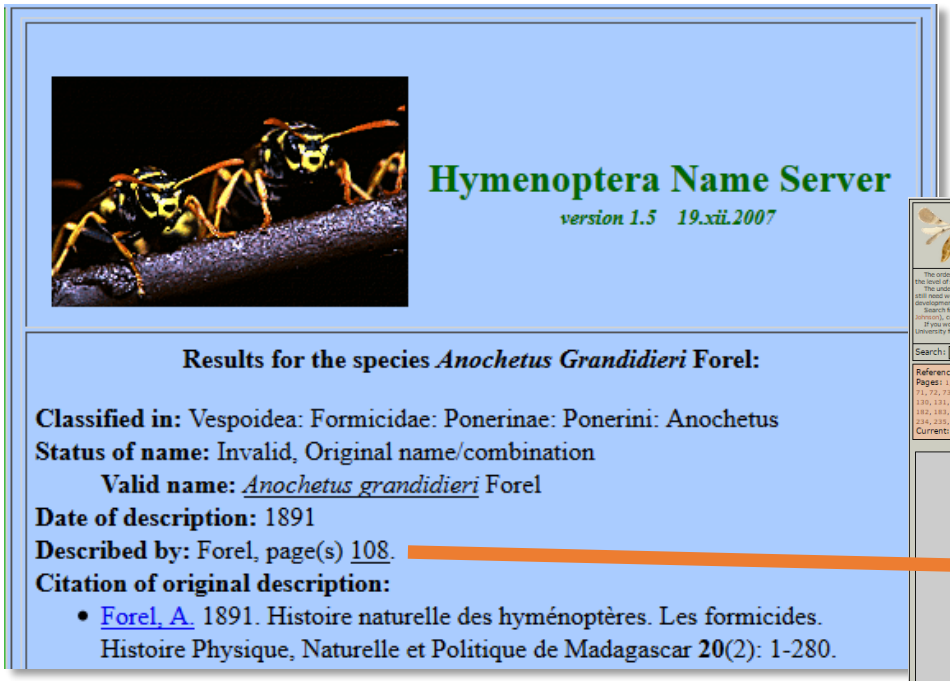
Our contribution for a better understanding of biodiversity

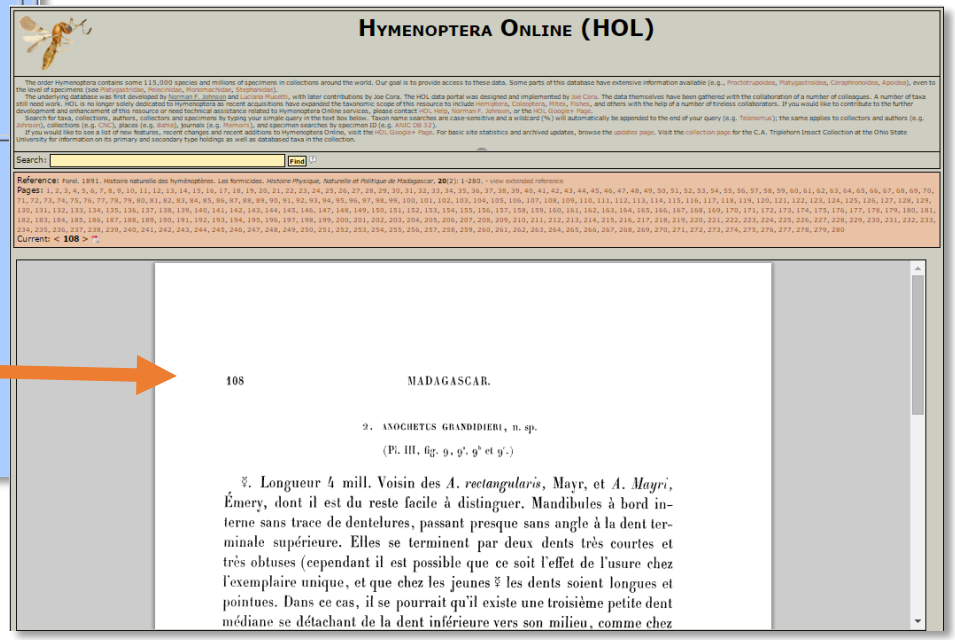
The special case taxonomic literature: The citated elements are treatments, not article

Treatment: a well defined part of an article that defines the particular usage of a scientific name by an authority at a given time (a page(s) in a publication).

From article to treatment

Online catalogue
Open access
Online library
«Linked data»
2004





Treatment citation → Treatment

Treatment

Hymenoptera Name Server

version 1.5 19.xii.2007

Results for the species *Anochetu*

Classified in: Vespoidea: Formicidae: Ponerinae
Status of name: Invalid, Original name/combin
Valid name: *Anochetus grandidieri* Fo
Date of description: 1891
Described by: Forel, page(s) 108.
Citation of original description:

- Forel, A. 1891. Histoire naturelle des hym
  Histoire Physique, Naturelle et Politique d

HYMENOPTERA ONLINE (HOL)

Hierarchy: Hymenoptera > Vespoidea > Formicidae > Ponerinae > Ponerini > Anochetus > Anochetus grandidie
(*Anochetus Grandidieri)

General Information

Synonyms (2) show fossils:

Taxon	Status	Relationship Type
Anochetus grandidieri Forel	Subsequent name/combination	Present combination
Anochetus madecassus Santschi	Original name/combination	Junior synonym

Literature (8) show synonyms: show all annotations: view type: annotation-view bibliography-view

Filter
Authors: all

- ***Anochetus Grandidieri* Forel**, 1891: 108. Original description, placed in subgenus Ponera (Ponera).
- ***Anochetus grandidieri* Forel**: Dalla Torre, 1893: 48. Cataloged.
- ***Anochetus grandidieri* Forel**: Emery, 1911: 109. Cataloged, placed in subgenus Anochetus (Anochetus).
- ***Anochetus grandidieri* Forel**: Wheeler, 1922: 1013. Cataloged, distribution.
- ***Anochetus grandidieri* Forel**: Wheeler, 1922: 791. Cataloged, distribution.
- ***Anochetus grandidieri* Forel**: Brown, 1978: 557, 564, 571, 606. Description, listed, synonymy, placed in grandidieri group, keyed, distribution.
- ***Anochetus grandidieri* Forel**: Bolton, 1995: 64. Cataloged.
- ***Anochetus grandidieri* Forel**: Fisher & Smith, 2008: 4, 8, 5c, 8b, figs .3a-3j, map6b. Description, keyed, lectotype designation, distribution.

NCBI (445)

One name → Many name(s) usages

Text

Semantically enhanced text (TaxonX)

RDF

5* 2014

Mystrium leonie B[illegible] & V[illegible]
Fig. 4–6, 11, 14

Type material. HOLOTYPE
E, 750m a.s.l., December 2

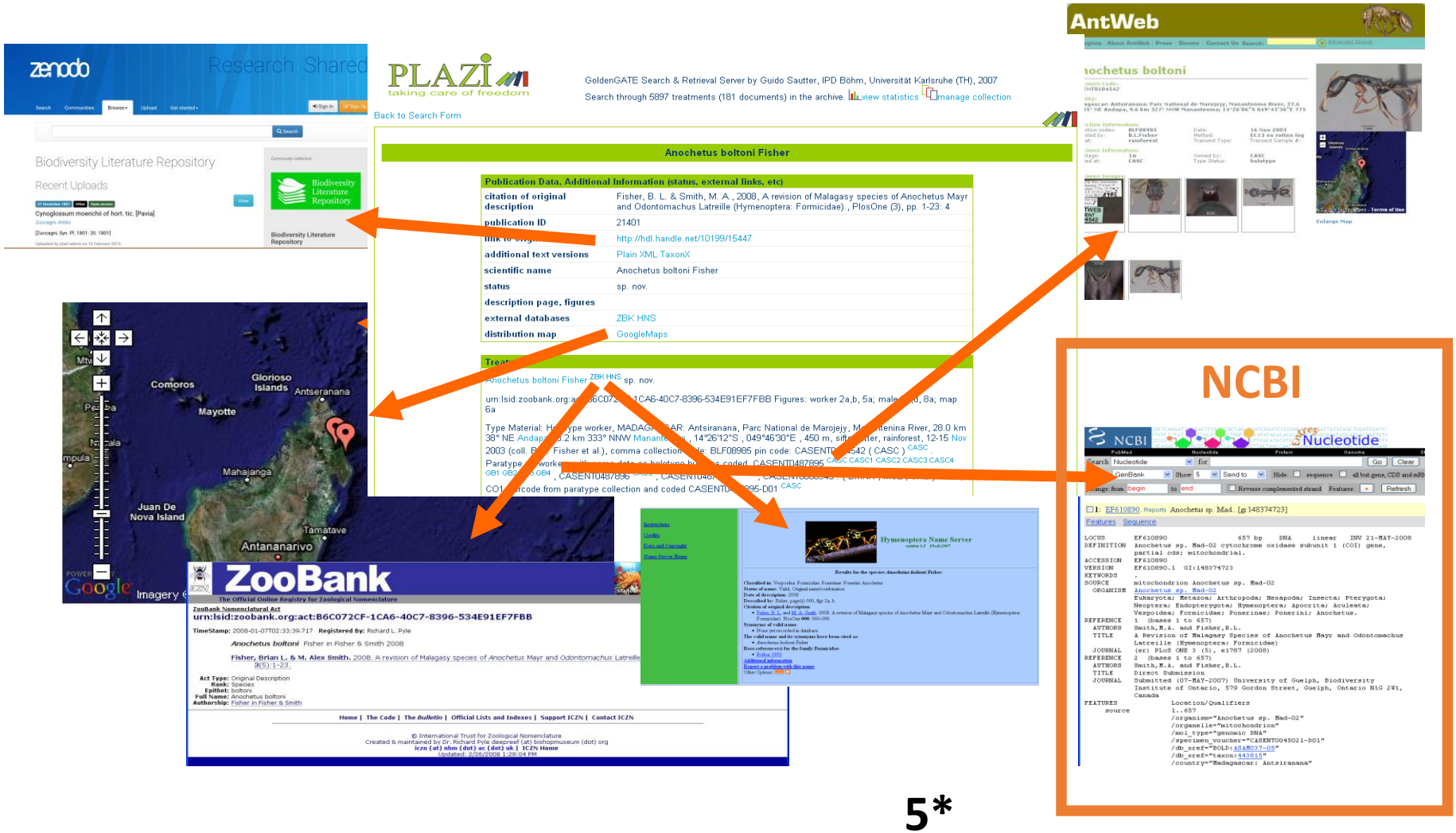
Measurements and in
2.50, PW 1.18.

Diagnosis (worker). T
the apex of each mandible
goreticulum; maxillary pal
(first) and third segment,
broadened in its distal par

```
<tax:treatment>
  <tax:nomenclature>
    <tax:name>
      <tax:xid source="HNS"
      <tax:xmldata>
        <dc:Genus>Mystrium<
        <dc:Species>leonie<
      </tax:xmldata>
      Mystrium leonie
    </tax:name>
    <tax:status>n. sp.</tax:
      Fig 1 D - F
  </tax:nomenclature>
  <tax:div type="description
    <tax:p>HOLOTYPE WORKER:
      1.30, SI 137, PW 0.73,
      to a sharp apical toot
      (Holotype with materia
      $ described below from
....
</treatment>
```

```
- <rdf:RDF>
  - <rdf:Description rdf:about="http://treatment.pla
      <rdf:type rdf:resource="http://plazi.org/vocab
      <trt:definesTaxonConcept rdf:resource="http
      <trt:publishedIn rdf:resource="http://publicat
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
      <trt:material_examined rdf:resource="http://
    </rdf:Description>
  - <rdf:Description rdf:about="http://publication.p
    - <dc:title>
        A revision of Malagasy species of Anochetus
      </dc:title>
      <dc:creator>Fisher, B. L.</dc:creator>
      <dc:creator>Smith, M. A.</dc:creator>
      <rdf:type rdf:resource="fabio:JournalArticle"
      <bibo:journal>PLoS ONE</bibo:journal>
      <dc:date>2008</dc:date>
      <bibo:volume>3</bibo:volume>
      <bibo:startPage>1</bibo:startPage>
      <bibo:endPage>23</bibo:endPage>
    </rdf:Description>
  - <rdf:Description rdf:about="http://example.org/
```

Linking of treatments to external resources



5*

2014

Different output formats

Anochetus boltoni Fisher

Publication Data, Additional Information (status, external links, etc)	
treatment citation	Fisher, B. L. & Smith, M. A., 2008, A revision of Malagasy Mayr and Odontomachus Latreille (Hymenoptera: Formicida 23: 4-6
publication ID	21401
link to original citation	http://dx.doi.org/10.1371/journal.pone.0001787
treatment provided by	Christiana
persistent identifier	http://treatment.plazi.org/id/B6C072CF-1CA6-40C7-8396-5
additional text versions	Plain XML TaxonX
scientific name	Anochetus boltoni Fisher
status	sp. nov.
external databases	ZBK HNS
distribution map	GoogleMaps

HTML

Treatment

Anochetus boltoni Fisher[ZBKHNS] sp. nov.

urn:lsid:zoobank.org:act:B6C072CF-1CA6-40C7-8396- 534E91EF7FBB Figures: worker 2a, 6a

Type Material: Holotype worker, MADAGASCAR: Antsiranana, Parc National de Marojejy, [illegible] km 38 ° NE Andapa [GoogleMaps], 8.2 km 333° NNW Manantenina [GoogleMaps], 14°26'12"S, 049°46'30"E, 450 litter, rainforest, 12-15 Nov 2003 (coll. B. L. Fisher et al.), comma collection code: BLF08985 pin code: CASENT0104542 (CASC) [GoogleMapsCASC]. Paratype. 8 workers with same data as holotype but pins code CASENT0487895 [GoogleMapsCASCCASC1CASC2CASC3CASC4GB1GB2GB3GB4], CASENT0487896 [GoogleMapsCA CASENT0487897 [GoogleMapsCASC], CASENT0006943. (BMNH, MCZ, CAS) [GoogleMapsCASC]; CO1 Barcode paratype collection and coded CASENT0487895-D01 [GoogleMapsCASC]

Figure 2. Anochetus[HNS] spp. full face and lateral view. A-B, boltoni[HNS] worker CASENT0104542. C-D, bo male CASENT0063847. E-F, goodmani[HNS] worker CASENT0104543. G-H, goodmani[HNS] ergatoid queen CASENT0454531. doi:10.1371/journal.pone.0001787.g002

Plazi internal XML


```
▼<document ENCODING="UTF-8" ID-HNS-PUB="21401" ModsDocAuthor="Fisher, B. L. & Smith, M. A." ModsDocDate="2008" ModsDocID="21401"
ModsDocOrigin="http://dx.doi.org/10.1371/journal.pone.0001787 " ModsDocTitle="A revision of Malagasy species of Anochetus Mayr
and Odontomachus Latreille (Hymenoptera: Formicidae)." checkinTime="1243372408936" checkinUser="christiana" docAuthor="Fisher,
B. L. & Smith, M. A." docDate="2008" docId="8AD0DAEF2180649D27DBA7CE08E4FF93" docLanguage="en" docName="21401" docOrigin="PLoS
ONE 3" docSource="http://dx.doi.org/10.1371/journal.pone.0001787" docTitle="Anochetus boltoni Fisher & Smith, 2008, sp. nov."
docUuid="B6C072CF-1CA6-40C7-8396-534E91EF7FBB" docUuidSource="ZooBank" lastPageId="5" lastPageNumber="6"
masterDocId="7B4CE9FDCC96F4BF0D2A5AB45EAECF74" masterDocTitle="A revision of Malagasy species of Anochetus Mayr and
Odontomachus Latreille (Hymenoptera: Formicidae)." masterLastPageNumber="23" masterPageNumber="1" pageId="3" pageNumber="4"
updateTime="1424953461880" updateUser="donat">
 ▼<mods:mods xmlns:mods="http://www.loc.gov/mods/v3">
  ▼<mods:titleInfo>
   ▼<mods:title>
      A revision of Malagasy species of Anochetus Mayr and Odontomachus Latreille (Hymenoptera: Formicidae).
    </mods:title>
  </mods:titleInfo>
  ▼<mods:name type="personal">
   ▼<mods:role>
      <mods:roleTerm>Author</mods:roleTerm>
    </mods:role>
    <mods:namePart>Fisher, B. L.</mods:namePart>
  </mods:name>
  ▼<mods:name type="personal">
   ▼<mods:rol
     <mods:r
    </mods:ro
    <mods:nam
  </mods:name
  <mods:typeC
 ▼<mods:relat
  ▼<mods:tit
     <mods:t
    </mods:ti
   ▼<mods:par
     <mods:d
    ▼<mods:d
       <mods
      </mods:
    ▼<mods:e
       <mods
       <mods
      </mods:
    </mods:pa
  </mods:rela
 ▼<mods:locat
    <mods:url
  </mods:loca
```



```
▼<rdf:RDF xmlns:dwc="http://rs.tdwg.org/dwc/terms/" xmlns:cnt="http://www.w3.org/2011/content#"
  xmlns:bibo="http://purl.org/ontology/bibo/" xmlns:trt="http://plazi.org/vocab/treatment#"
  xmlns:rdfs="http://www.w3.org/2000/01/rdf-schema#" xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#"
  xmlns:xsd="http://www.w3.org/2001/XMLSchema#" xmlns:fabio="http://purl.org/spar/fabio/" xmlns:sdd="http://tdwg.org/sdd#"
  xmlns:mods="http://www.loc.gov/mods/v3" xmlns:dc="http://purl.org/dc/elements/1.1/"
  xmlns:dwcFP="http://filteredpush.org/ontologies/oa/dwcFP#">
 ▼<rdf:Description rdf:about="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93">
    <rdf:type rdf:resource="http://plazi.org/vocab/treatment#Treatment"/>
    <trt:definesTaxonConcept rdf:resource="http://example.org/concept/id/8AD0DAEF2180649D27DBA7CE08E4FF93"/>
    <trt:publishedIn rdf:resource="http://publication.plazi.org/id/7B4CE9FDCC96F4BF0D2A5AB45EAECF74"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_1"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_2"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_3"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_4"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_5"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_6"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_7"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_8"/>
    <trt:material_examined rdf:resource="http://treatment.plazi.org/id/8AD0DAEF2180649D27DBA7CE08E4FF93#material_9"/>
  </rdf:Description>
 ▼<rdf:Description rdf:about="http://publication.plazi.org/id/7B4CE9FDCC96F4BF0D2A5AB45EAECF74">
  ▼<dc:title>
     A revision of Malagasy species of Anochetus Mayr and Odontomachus Latreille (Hymenoptera: Formicidae).
    </dc:title>
    <dc:creator>Fisher, B. L.</dc:creator>
    <dc:creator>Smith, M. A.</dc:creator>
    <rdf:type rdf:resource="fabio:JournalArticle"/>
    <bibo:journal>PLoS ONE</bibo:journal>
    <dc:date>2008</dc:date>
    <bibo:volume>3</bibo:volume>
    <bibo:startPage>1</bibo:startPage>
    <bibo:endPage>23</bibo:endPage>
  </rdf:Description>
 ▼<rdf:Description rdf:about="http://example.org/concept/id/8AD0DAEF2180649D27DBA7CE08E4FF93">
    <rdf:type rdf:resource="http://filteredpush.org/ontologies/oa/dwcFP#TaxonConcept"/>
   ▼<dwc:LSID-HNS>
      urn:lsid:biosci.ohio-state.edu:osuc_concepts:235700
    </dwc:LSID-HNS>
   ▼<dwc:LSID-ZBK>
      urn:lsid:zoobank.org:act:B6C072CF-1CA6-40C7-8396-534E91EF7FBB
    </dwc:LSID-ZBK>
    <dwc:class>Insecta</dwc:class>
    <dwc:family>Formicidae</dwc:family>
    <dwc:genus>Anochetus</dwc:genus>
    <dwc:higherTaxonomySource>Col</dwc:higherTaxonomySource>
```


RDF

Darwin Core Archive

Citing of treatments or linking of treatments to treatments

By minting persistent httpURIs for treatments, treatments can be cited like a bibliographic reference

Pseudomyrmex gracilis (Fabricius 1804)

(Fig. 6)

Formica gracilis Fabricius 1804:405. Lectotype worker, Essequibo , Guyana (ZMUC) [Examined], View Cited Treatment

Pseudomyrma bicolor Guerin 1844:427. Syntype queen (unique?), Colombia (ZSMC) [Examined] Syn. nov.

http://treatment.plazi.org/id/A9FFD1FC-4629-FFB4-968F-AD38386521BA

40. Formica gracilis

F. elongata nigra, mandibulis antennis pedibusque rufis, petiolo binodi.

Habitat in America meridionali Dom. Smidt. Mus. Dom. Lund.

Statura parva, tenuis, elongata praecedentium. Caput planum, ovatum, nigrum, antennis mandibulisque fornicatis ferrugineis. Thorax compressus, niger. Abdomen ovatum, petiolo elongato, binodi, basi interdum ferrugineo. Pedes rufi.

5*

2014

Treatment opportunities

Acacia-ant species: *Pseudomyrmex gracili*

Associated ant-acacia: *Acacia gentlei*

Treatment: redescription

Treatment

Treatments linked through citations

Pseudomyrmex *ants and* Vachellia *ant-acacias are a classic example of mutualism in biology.*

Photocredits: Alex Wild

Coremiocnemis tropix , ROBERT J RAVEN

Publication Data, Additional Information (status, external links, etc)	
treatment citation	ROBERT J RAVEN, 2005, A new tarantula species from northern Australia (Araneae, Theraphosidae), Zootaxa 1004, pp. 15-28: 18-27
publication ID	
link to original citation	http://www.mapress.com/zootaxa/2005f/zt01004p028.pdf
treatment provided by	Jeremy
persistent identifier	http://treatment.plazi.org/id/DB0A588D-4CED-9C22-4EBA-D26AD1BE604B
additional text versions	Plain XML TaxonX
scientific name	Coremiocnemis tropix
status	sp.nov.
external databases	
distribution map	GoogleMaps

Treatment

Coremiocnemis tropix sp.nov.

(Figs1-37)

Material examined.

Holotype:female, Atherton ,17°16'S,145°29'E,north-eastern Queensland,Australia,25May1988,M.Jeansson,QM S6325.

Paratypes:WAM90/1959, 1male, Freshwater Creek,at Crystal Cascades,10km S.of Freshwater, 12Jul1986,M.Harvey; WAM T57516, 1male,1female, Atherton ,June1994,R.Elick. 1female, Table Mt,10km S Cape Tribulation, 16°09'S,145°26'E,rainforest,24Apr1983,G.Monteith,D.Cook,QM S10583; 1female, Gordonvale,27km SW, 17°01'S,145°45'E,rainforest,17Mar1980,N.Clyde Coleman,QM S10590; 1female, Cape Tribulation, 16°05'S,145°26'E,rainforest,29Dec1982-8Jan1983,G.Monteith,QM S10598; 1female, Freshwater Creek,Cairns, 16°58'S,145°43'E,rainforest,under stones,Aug1977,M.T.Bishop,QM S10633; 1female, Mulgrave River,27km SW Gordonvale ,16°23'S,143°59'E,19Mar1980,N.Clyde Coleman,QM S10793; 1female, Crystal Cascades, 16°58'S,145°42'E,rainforest,sheet web,24Aug1980,M.Harvey,QM S20889; 1male, Cairns ,16°55'S,145°46'E,house,5Sep1995,Stella Jeffery,QM S30136; 1male, Kuranda ,16°49'S,145°38'E,house,1Aug1991,G.Monteith,QM S25471; 1female, Cairns ,16°55'S,145°46'E,24Apr2001,A.Hunt,QM S55051; 1female, Earlville ,16°57'S,145°44'E,Nov1986,Safeway Pest Control,QM S6685; 1male,1female, Mission Beach, 17°56'S,146°05'E,May1981,E.Long,QM S10650; 1male, Clifton beach, 16°46'S,145°40'E,28May1992,QM S20388; 1male, Smithfield,Cairns, 16°51'S,145°43'E,spider bite,2May1996,P.Hawkins,QM S29732; 1male, Kurrimine Beach,30km S Innisfail, 17°47'S,146°06'E,spider bite,July1997,via R.Piper,QM S34638.All in north-eastern Queensland,Australia.

Other material examined:1juvenile, Noah Creek,Cape Tribulation, 16°08'S,145°26'E,rainforest,16Oct1980,G.Monteith,QM S10592; 1penultimate male, Cape Tribulation,2km WNW(Site2), 16°05'S,145°28'E,rainforest,23Sep– 7Oct1982,G.Monteith,D.Yeates,G.Thompson,QM S10593; 1penultimate male, Babinda ,17°21'S,145°56'E,hospital,QM S10610; 1penultimate male, Cape Tribulation, 16°05'S,145°26'E,rainforest,22Sep– 7Oct1982,Qld Naturalists Club,QM S10630; 1juvenile, Mt Hartley, 15°46'S,145°20'E,rainforest,6Nov1974,V.Davies,J.Covacevich,D.Joffe,QM S10632; 4penultimate males, Cape Tribulation, 16°05'S,145°26'E,rainforest,under logs and rocks,15-19Aug1975,W.and J.Nash,QM S10640; 3penultimate males, Crystal Cascades, 16°58'S,145°42'E,rainforest,Feb1979,N.Clyde Coleman,QM S10792; 1penultimate male, Cape Tribulation, 16°05'S,145°26'E,rainforest,26Aug1988,T.B.Churchill,QM S11216; 1juvenile, Cairns,Brinsmead, 16°55'S,145°46'E,4May1989,P.Blackman,QM S15278.All in north-eastern Queensland,Australia.

Diagnosis:Differs from both C.cunicularia (Simon1892)and C.valida Pocock 1895in the absence of the very long brushes on metatarsi IV(Figs1, 18)and in the less extensive maxillary lyra(Figs14, 23).

FIGURE1. Coremiocnemis tropix sp.nov., female,habitus.,Cape Tribulation.Photo:B.Cowell

Etymology:An arbitrary combination of letters phonetically like tropics,the origin of the spider.

Common Name:Tropix.

Holotype Female QM S6325

Description(measurements in mm):Carapace11.1long,8.7wide,chelicerae5.5.Legs(femur,patella,tibia,metatarsus,tarsus,total):I:8.0, 5.0, 6.0, 4.3, 2.8, 26.1.II:6.4, 4.2 4.5, 3.9, 2.6, 21.6.III:5.8, 3.5, 3.5, 4.2, 2.5, 19.5.IV:7.7, 4.0, 5.7, 6.6, 3.3, 27.0.Palp:5.5, 3.3, 3.5,-,3.5, 15.8.Midwidths:femora I,II,IV=1.65,III=1.90;tibia I–IV ,2.1, 1.75, 1.75, 1.65.Abdomen12.7long,7.2wide.Spinnerets:PMS,3.50long,0.55wide,1.10apart with sclerotised(but less hirsute than venter)cuticle to anal tubercle;PLS,6.0basal,4.2middle,6.0,distal;midwidths1.0, 1.0, 0.75,respectively.

Carapace:length to width1.28;uniform red brown;uniform cover of long white wavy hairs,no setae or thorns.

FIGURES2-9. Coremiocnemis tropix sp.nov., holotype female(2-4, 8-9);QM S10650(5-7).2.Carapace,dorsal view.3.abdomen,dorsal view.4.eyes,dorsal view.5.Ectal cheliceral face,ventral view showing pins.6.prolateral cheliceral face showing peg setae(inset:detail).7.cheliceral furrow,ventral view showing smaller basal teeth.8.Tarsus IV,ventral view,with dividing setae.9.Tarsus IV,dorsal view,showing tarsal weakness.Scale bar=2mm for2, 3, 8-9,all others1mm.

Eyes:ratio of group width to length2.11.PLE clearly smaller than ALE but clearly bigger than PME,AME=ALE(Figs2, 4).

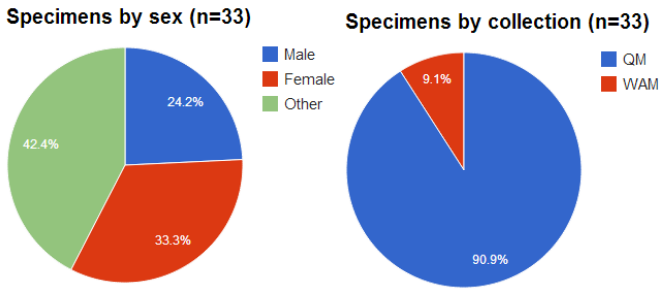
Maxillae(Fig.14):posterior ventral edge gently rounded for length;retro-face mostly pallid,smooth,glabrous.Cuspules:ca.200in anterior corner in square region(Figs11, 22).Posterior edge near heel straight or slightly concave.

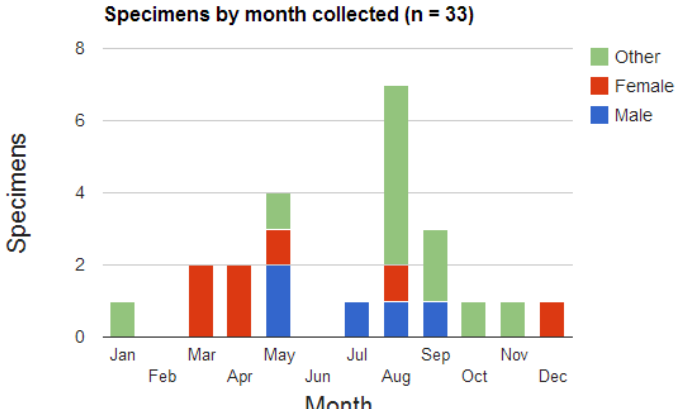
FIGURES10-19. Coremiocnemis tropix sp.nov., female,holotype(10-13, 15-19);QM S10650(14).10.Sternum,maxillae,labium and coxae,ventral view.11.Sternum,proximal maxillae and labium,ventral view.12.female genital fold,ventral view.13.labium,ventral view.14.maxilla,prolateral view,showing lyra.15.tarsus IV,lateral

Countries (Region)

Australia (Queensland)

Export species materials citations (DwC)





Etymology:An arbitrary combination of letters phonetically like tropics,the origin of the spider.

Common Name:Tropix.

Holotype Female QM S6325

Description(measurements in mm):Carapace11.1long,8.7wide,chelicerae5.5.Legs(femur,patella,tibia,metatarsus,tarsus,total):I:8.0, 5.0, 6.0, 4.3, 2.8, 26.1.II:6.4, 4.2 4.5, 3.9, 2.6, 21.6.III:5.8, 3.5, 3.5, 4.2, 2.5, 19.5.IV:7.7, 4.0, 5.7, 6.6, 3.3, 27.0.Palp:5.5, 3.3, 3.5,-,3.5, 15.8.Midwidths:femora I,II,IV=1.65,III=1.90;tibia I–IV ,2.1, 1.75, 1.75, 1.65.Abdomen12.7long,7.2wide.Spinnerets:PMS,3.50long,0.55wide,1.10apart with sclerotised(but less hirsute than venter)cuticle to anal tubercle;PLS,6.0basal,4.2middle,6.0,distal;midwidths1.0, 1.0, 0.75,respectively.

Carapace:length to width1.28;uniform red brown;uniform cover of long white wavy hairs,no setae or thorns.

FIGURES2-9. Coremiocnemis tropix sp.nov., holotype female(2-4, 8-9);QM S10650(5-7).2.Carapace,dorsal view.3.abdomen,dorsal view.4.eyes,dorsal view.5.Ectal chelicteral face,ventral view showing pins.6.prolateral chelicteral face showing peg setae(inset:detail).7.chelicteral furrow,ventral view showing smaller basal teeth.8.Tarsus IV,ventral view,with dividing setae.9.Tarsus IV,dorsal view,showing tarsal weakness.Scale bar=2mm for2, 3, 8-9,all others1mm.

Eyes:ratio of group width to length2.11.PLE clearly smaller than ALE but clearly bigger than PME,AME=ALE(Figs2, 4).

Maxillae(Fig.14):posterior ventral edge gently rounded for length;retro-face mostly pallid,smooth,glabrous.Cuspules:ca.200in anterior corner in square region(Figs11, 22).Posterior edge near heel straight or slightly concave.

FIGURES10-19. Coremiocnemis tropix sp.nov., female,holotype(10-13, 15-19);QM S10650(14).10.Sternum,maxillae,labium and coxae,ventral view.11.Sternum,proximal maxillae and labium,ventral view.12.female genital fold,ventral view.13.labium,ventral view.14.maxilla,prolateral view,showing lyra.15.tarsus IV,lateral view.16.spermathecae,ventral view.17.claws on leg IV,lateral view.18.tibia to tarsus IV,dorsal view.19.coxa I,ventral view,showing weak setae prolaterally.Scale bar=5mm for10, 2mm for11-15, 18-19;1mm for16-17.

Maxillary lyra(Fig.14):overall shape,small,ovoid,ca.0.6of mid-maxilla length;ventral edge more or less smoothly and gently convex;outer point asymmetrically rounded point;inner point bluntly truncated;large outer paddles in smoothly uniform curving line,gaps evenly spaced;overall size,central,about half length.No brush ventral of lyra.Thick paddle setae in3lines centrally then thick pointed start;lyra dorsal edge line relative to midgroove line divergent,ca.10basally diverging to15distally.

Labium(Fig.13):over200cuspules in band for two-fifths of length anteriorly;cuspules ca.similar in size and number to maxillary.Basal groove shallow,distinct.Labiosternal groove(Fig.13)not concave,flat with slight(anterior)rise,two separate large sigilla.

Chelicerae:intercheliceral spines basodorsally,ca.13larger and2-4smaller basally(Figs6, 37).Ectal lyrate area(Fig.5),groove glabrous;lyra setae,not apically convergent but parallel,in3straight lines of long thick setae,2lines of short coniform ectally with irregular band of thorns2-4deep outside that(Fig.5).

Sternum(Figs10-11):profile saddle-like,high at back and front.Posterior angle sharp but not separating coxae IV.Posterior edge easily seen,sloping gradually.Marginal thorns absent.Prostrate hair mat strong,dense,or of grey hairs.Pedicel and sternum edges form clear boundary in elevation and setation,pedicel edge pallid.

FIGURES20-27. Coremiocnemis tropix sp.nov., allotype male.20.Habitus,dorsal view.21.carapace,dorsal view.22.sternum,maxillae,labium and coxae,ventral view.23.maxilla,prolateral view,showing lyra.24.Tip of embolus,ventral view.25.bulb,ventral view.26.palp,tibia to tarsus,prolateral view.27.palp,tibia to tarsus,retrolateral view.Scale bar=10mm for20;5mm for21, 22;2mm for23, 25-27, 1mm for24.

Sigilla(Figs10-11):3pairs,posterior ca.2(1.6-3across paratypes)lengths apart and only ca.1length from margin;middle ca.half size of posterior,within ca.1length of margin;anterior clearly present.

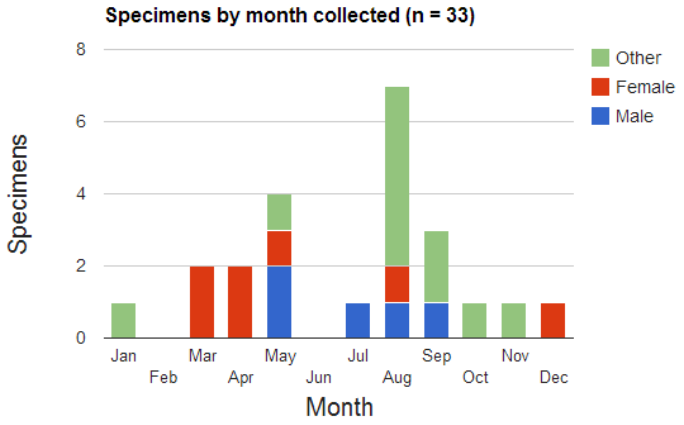
Legs:leg I clearly thicker than IV but II subequal to IV.Basifemoral thorns absent on all.Femora I–II prolaterally with long bushy fine hairs;III distinctly incrassate.Metatarsus I hardly longer than tarsus,less than twice.

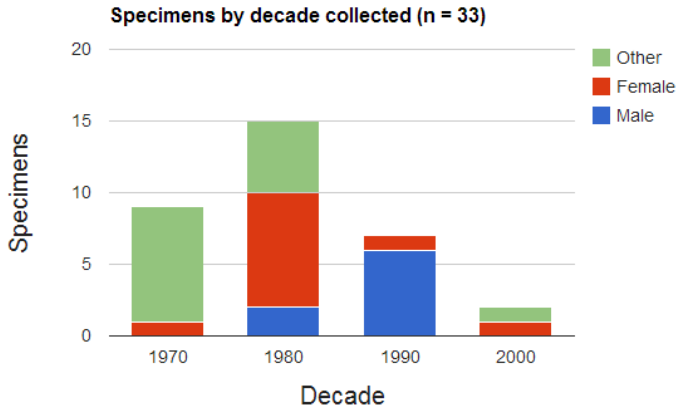
Leg pilosity(number of long hairs extending well above base layer):tibia I p=10-15,d=5-10,r=10-15;metatarsus I p=d=5-10,r=1-3,v=1-3;tibia IV p=30-40,d=20-30,r=v=40-50;metatarsus IV p=30-40,d=30-40,r=v=50-60;tarsus IV ca.20-40dorsolaterally on each side.Tuft of blunt-tipped hairs,distoventral metatarsi present at least on IV(Fig.34).Paired long sensory hairs pointing distad,long gently curved dorsolateral pair basally(0.2from base)on metatarsi I–IV ,long bent(at0.5)pair at distal0.8on metatarsi I–IV (Fig.15;possibly widespread);long curving dorsolateral pair at0.5(end of very long filiform trichobothria)on tarsi I–IV (Fig.15).Long,downwardly curved hairs(Figs15, 30)below claw tufts basally with transverse ribbing but apically with discontinuous partially spiralled ribbing(presumably thermosensory in function;viz.,Den Otter,1974).Upper tarsi IV(and presumably I–III)with two other hairs types:closed longitudinal fimbriae(Fig.35)and differentially dorsoventrally fimbriated hairs(Fig.36).

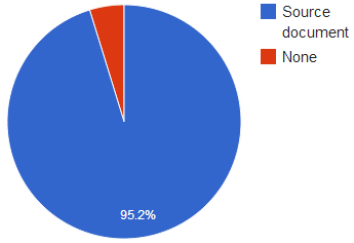
Trichobothria(Fig.33):on tarsi basal filiform field hardly wider than clavate field and merges smoothly;clavates on tarsi I only in distal2/3;long filiforms only in basal half,shorter filiforms intermixed with clavates distally.Clavate extent on II–IV cf.I,only in distal2/3;long filiforms only in basal half,shorter filforms for length.Short epitrichobothrial hair field(Figs31-33)on I shorter than clavates and uniform height for length.

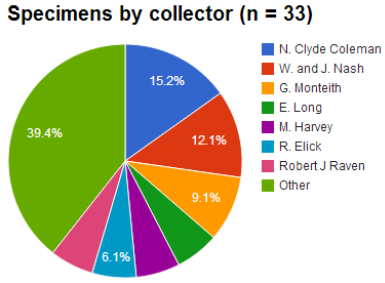
Leg coxae(Fig.10):no thorns prolaterodorsally or retrodorsally.Coxal bases dorsally easily seen from above(Fig.2).I clearly longest,about1.8times length of II;IV clearly widest and about as long as III and basally with anterior corner indistinct,edge curves dorsally,rounded at interface.III basally rectangular.Coxae ventrally with short,weak thorns only on prolateral faces of I–II (Fig.19).I–IV ventrally with many long blunt setae but most short dark and fewer long pallid;ventral surface clearly curving or sloping forward.Retrolateral setation:I–III with median narrow light brush,IV glabrous;III–IV with setose mound up from inner corner low mound with few bristles.All retrolaterally lack ventral ledge and ventrally with uniform setation(Fig.10).

Scopula:entire,dense on tarsi I and metatarsi I–II ;on tarsi II entire but with long emergent hairs in central zone,not dividing scopula;on metatarsi III for ca.0.8-9of length,but with uniformly distributed(i.e.,not a dividing line)long emergent hairs;on tarsi III entire but with long emergent hairs in central zone but not dividing scopula.For ca.0.75length of

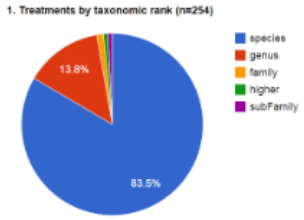


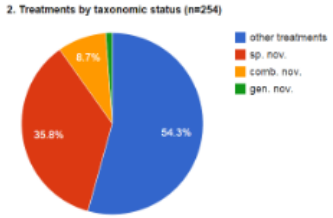


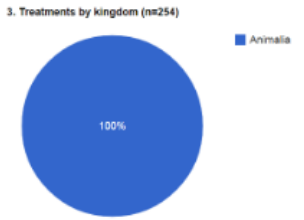


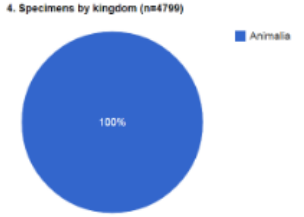


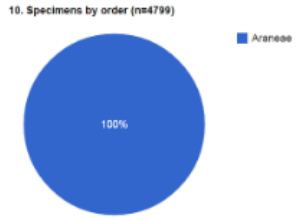
The Issue: treatment content

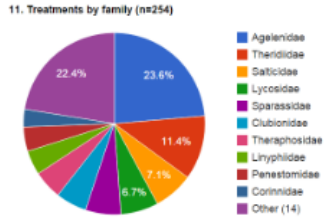


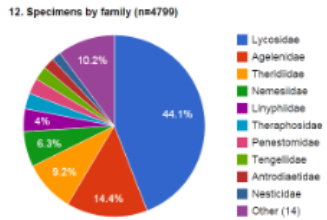


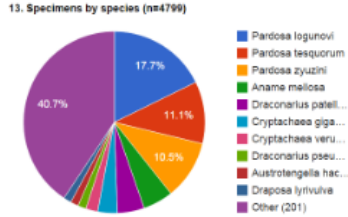


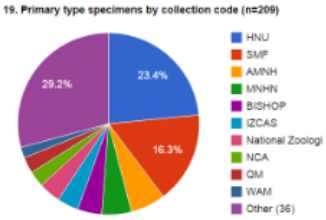


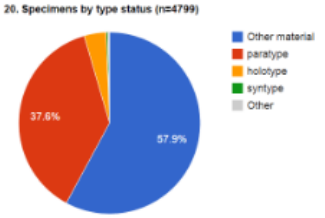


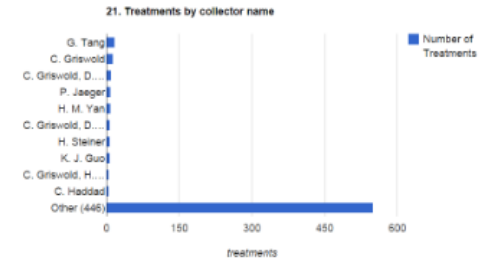


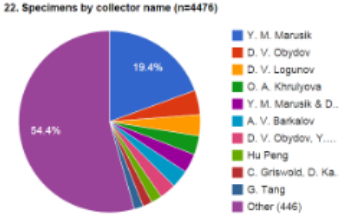


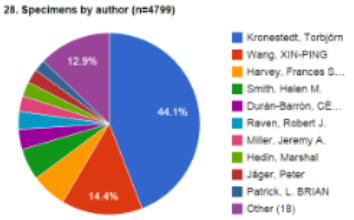


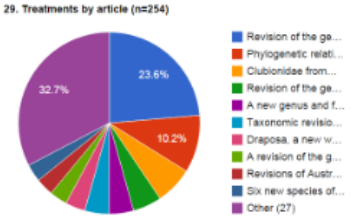


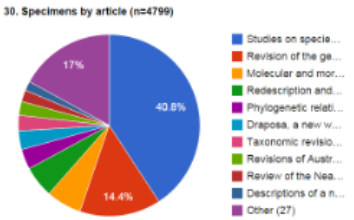


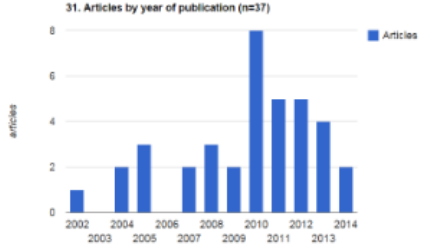




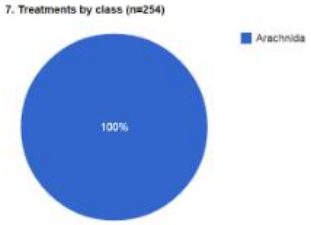


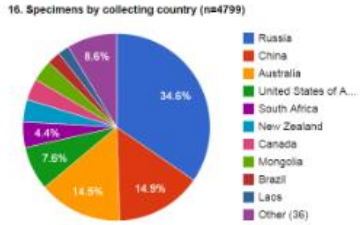


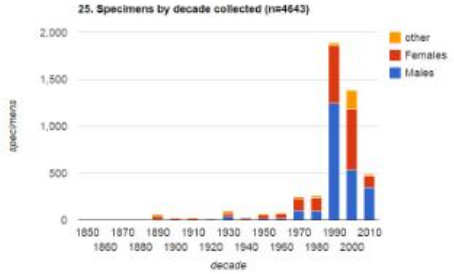


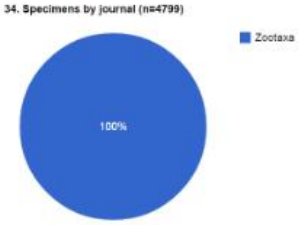


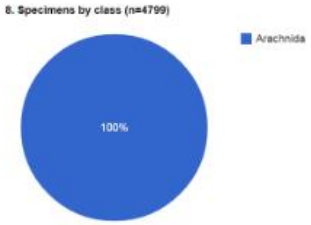
The issue: treatment content (ctd)

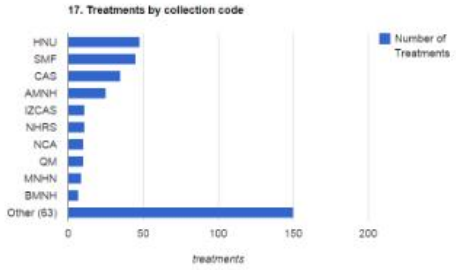


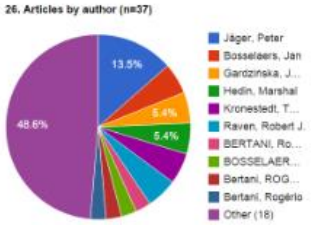


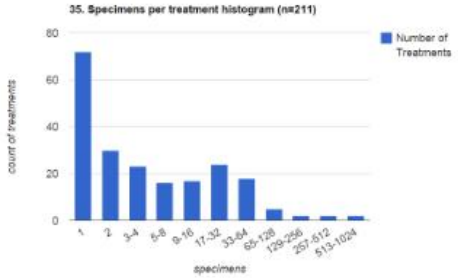


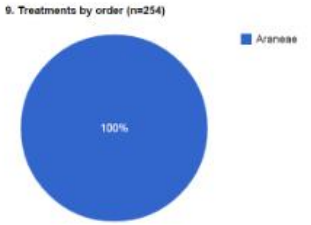


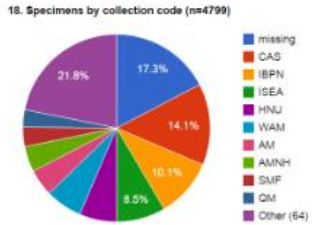


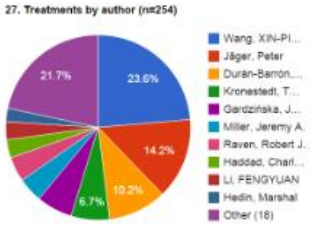


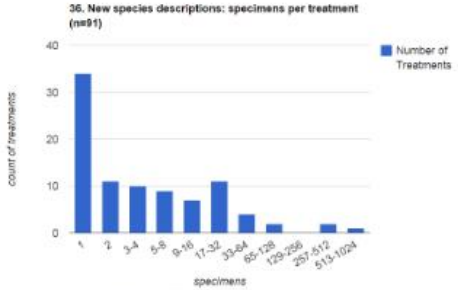










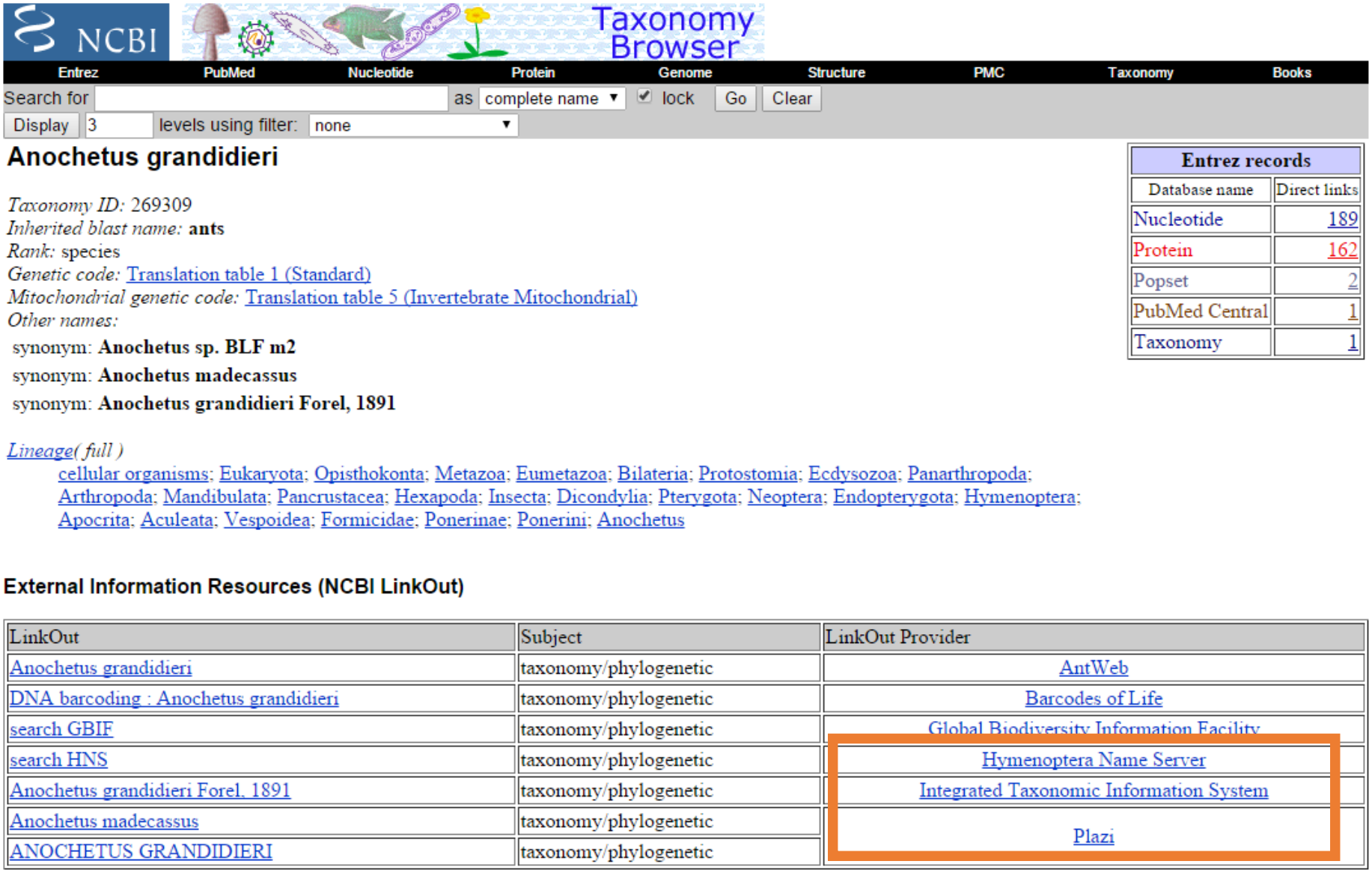


Treatments can have very rich content highly relevant to EU-BON
Treatments have a high **data quality** standard, being expert products

Tools to mine Plazi data

- API to access data
- Preset queries to extract data
  http://plazi.org/wiki/Treatment_Statistics

Plazi integration with NCBI: current



NCBI Taxonomy Browser

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search for [ ] as complete name ▾ lock Go Clear

Display 3 levels using filter: none ▾

Anochetus grandidieri

Taxonomy ID: 269309
Inherited blast name: **ants**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 5 (Invertebrate Mitochondrial)
Other names:
synonym: **Anochetus sp. BLF m2**
synonym: **Anochetus madecassus**
synonym: **Anochetus grandidieri Forel, 1891**

Lineage (full)
cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Ponerinae; Ponerini; Anochetus

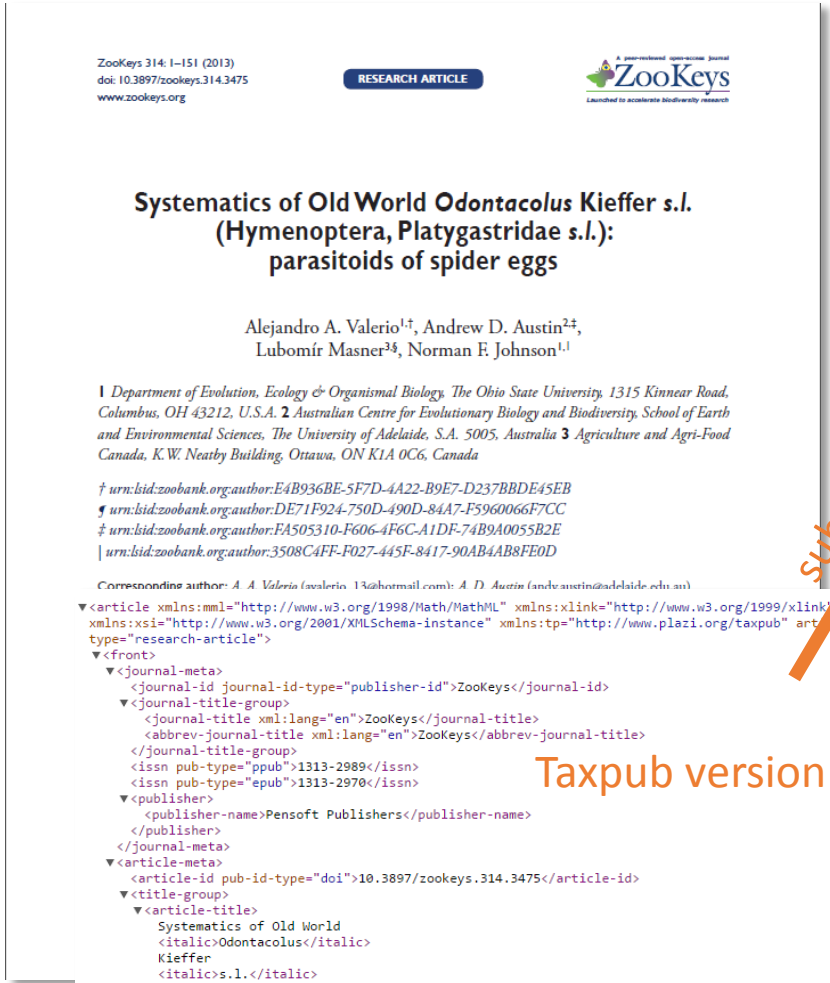
Entrez records

Database name	Direct links
Nucleotide	189
Protein	162
Popset	2
PubMed Central	1
Taxonomy	1

External Information Resources (NCBI LinkOut)

LinkOut	Subject	LinkOut Provider
Anochetus grandidieri	taxonomy/phylogenetic	AntWeb
DNA barcoding : Anochetus grandidieri	taxonomy/phylogenetic	Barcodes of Life
search GBIF	taxonomy/phylogenetic	Global Biodiversity Information Facility
search HNS	taxonomy/phylogenetic	Hymenoptera Name Server
Anochetus grandidieri Forel, 1891	taxonomy/phylogenetic	Integrated Taxonomic Information System
Anochetus madecassus	taxonomy/phylogenetic	Plazi
ANOCHETUS GRANDIDIERI	taxonomy/phylogenetic	

Treatment: From article to treatment

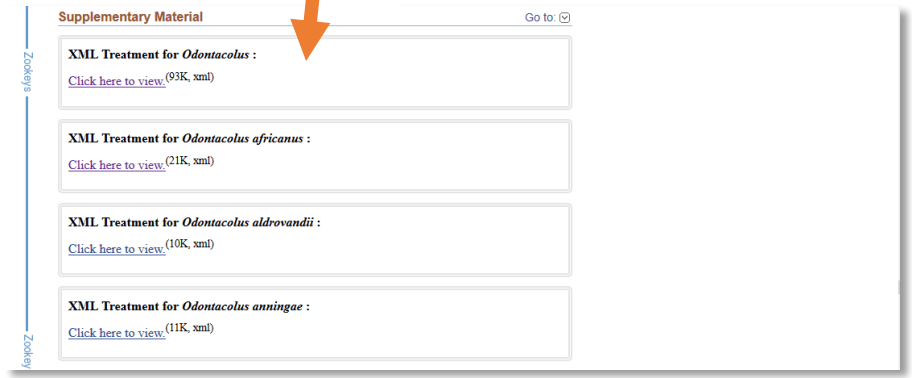



```
▼<article xmlns:mml="http://www.w3.org/1998/Math/MathML" xmlns:xlink="http://www.w3.org/1999/xlink"
 xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance" xmlns:tp="http://www.plazi.org/taxpub" article-
 type="research-article">
 ▼<front>
  ▼<journal-meta>
    <journal-id journal-id-type="publisher-id">ZooKeys</journal-id>
   ▼<journal-title-group>
      <journal-title xml:lang="en">ZooKeys</journal-title>
      <abbrev-journal-title xml:lang="en">ZooKeys</abbrev-journal-title>
    </journal-title-group>
    <issn pub-type="ppub">1313-2989</issn>
    <issn pub-type="epub">1313-2970</issn>
   ▼<publisher>
      <publisher-name>Pensoft Publishers</publisher-name>
    </publisher>
  </journal-meta>
 ▼<article-meta>
    <article-id pub-id-type="doi">10.3897/zookeys.314.3475</article-id>
   ▼<title-group>
    ▼<article-title>
       Systematics of Old World
       <italic>Odontacolus</italic>
       Kieffer
       <italic>s.l.</italic>
       (Hymenoptera, Platygastridae
       <italic>s.l.</italic>
       ): parasitoids of spider eggs
     </article-title>
   </title-group>
  ▼<contrib-group>
   ▼<contrib contrib-type="author" xlink:type="simple">
    ▼<name name-style="western">
       <surname>A.Valerio</surname>
       <given-names>Alejandro</given-names>
     </name>
     <xref ref-type="aff" rid="A1">1</xref>
   </contrib>
   ▼<contrib contrib-type="author" xlink:type="simple">
    ▼<name name-style="western">
       <surname>Austin</surname>
```


Taxpub version

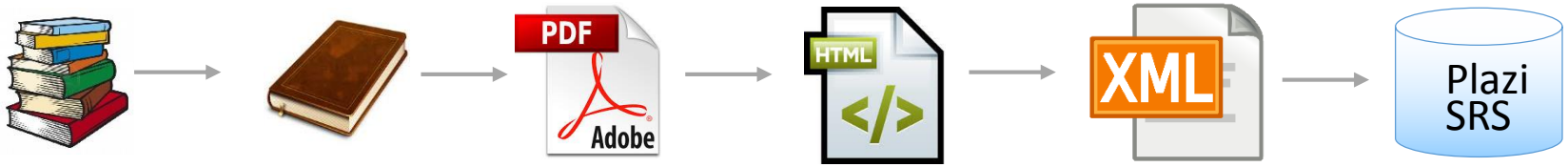
submission

extraction



Plazi content

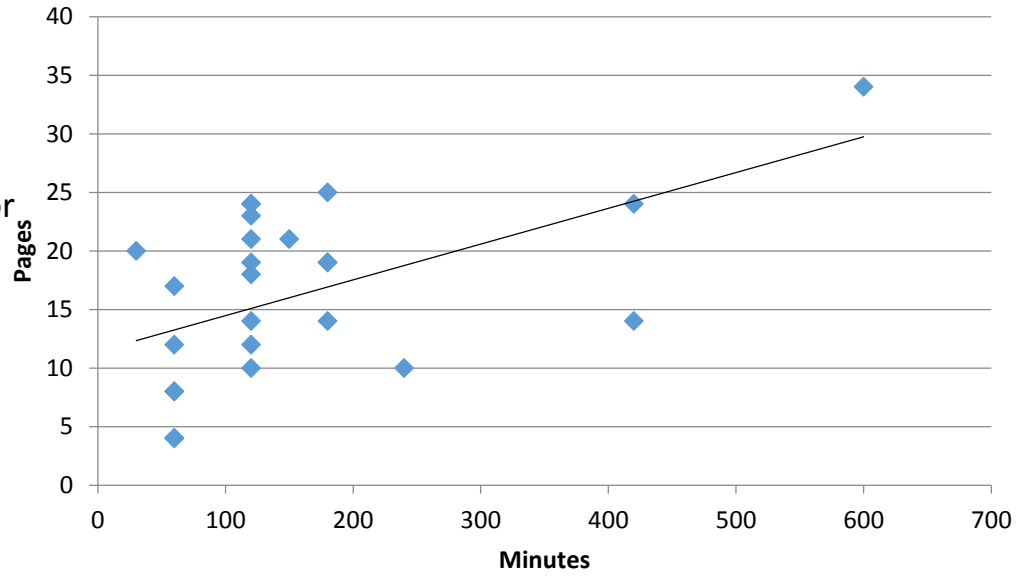
Mark up costs for markup including materials citations



find scan «OCR» markup store

Average:
6 min / page complete

OCR: 0.80 EUR /page vendor



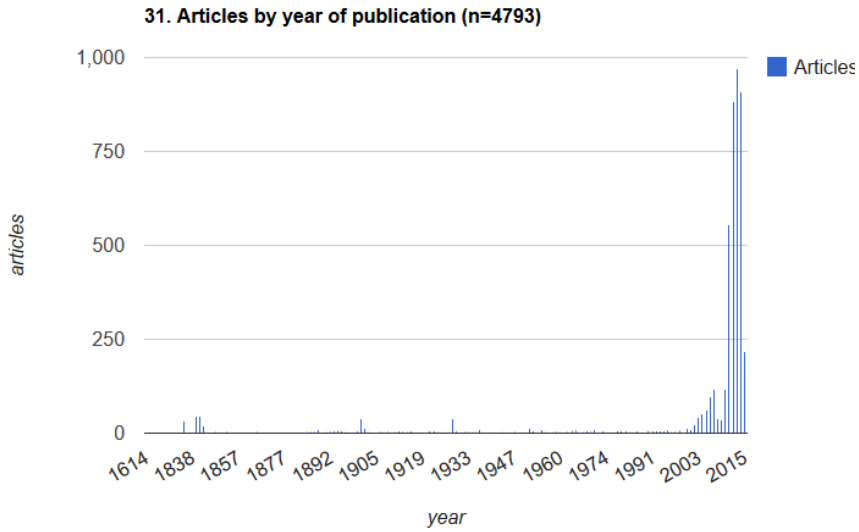
Source: Spider Pilot, Jeremy Miller

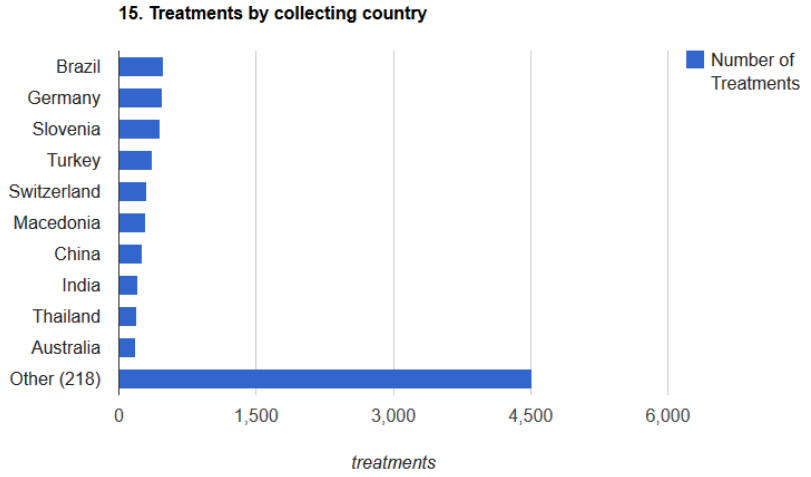
Mark up costs can be lowered by improvied tools and parallel and pre-processing

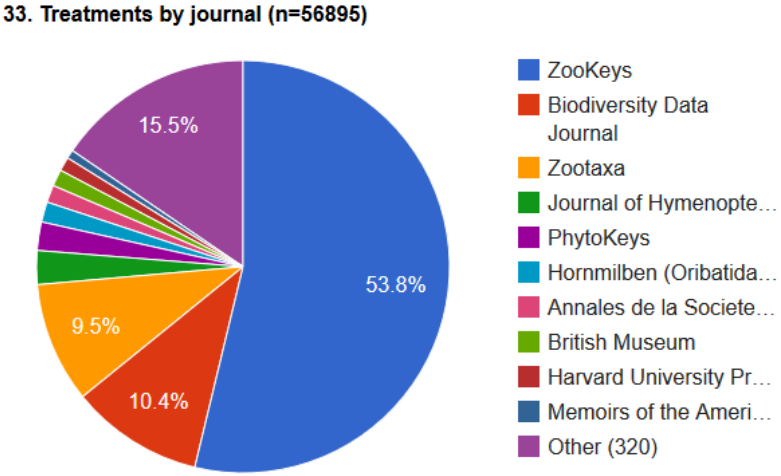
Actual Plazi development and content

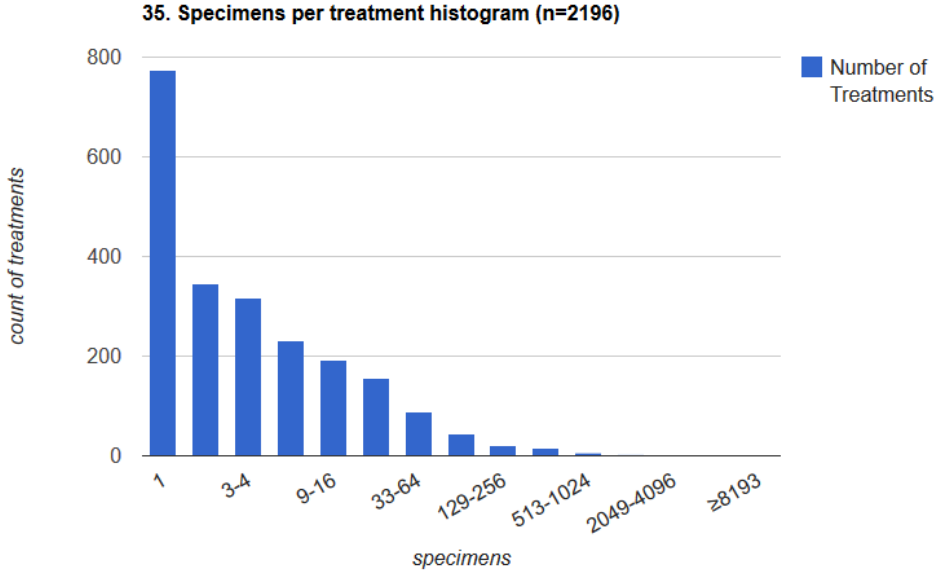
- Betaversion of Imagine markup tool to convert PDF
- Import facility to import Taxpub based journals (eg. Pensoft journals)
- TreatmentBank with 126,000 taxa, 56,895 with treatments
- Biodiversity Literature Repository in collaboration with CERN to store taxonomic articles and issue DOIs
- Functional Darwin Core Archive based API to input data into GBIF/EU-BON
    - 28,000 observation records of 3249 taxa (=treatments) that match the GBIF taxonomic back bone.
    - 28% with geo-coordinates

Plazi development and content









Treatment mining proposal

Combine the data inherent in the genomic and taxonomic world by linking genomic data with organims names with the (phenotypic) knowledge in the taxonomic domain.

Build the necessary infrastructure to provide a text mining workflow.

Treatments are like abstracts in biomed literature: very concise description of the current information and data on a specific topic.

Organism names have in most cases a long history, and thus multiple treatments for a currently accepted name exist.

Together, this constitutes a huge untaped resource for trait extraction of particular organisms.

Treatment mining proposal

NCBI Taxonomy Browser

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search for [ ] as [complete name] [lock] [Go] [Clear]

Display [3] levels using filter: [none]

Anochetus grandidieri

Taxonomy ID: 269309
Inherited blast name: **ants**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 5 (Invertebrate Mitochondrial)
Other names:
synonym: **Anochetus sp. BLF m2**
synonym: **Anochetus madecassus**
synonym: **Anochetus grandidieri Forel, 1891**

Entrez records	
Database name	Direct links
Nucleotide	189
Protein	162
Popset	2
PubMed Central	1
Taxonomy	1

Lineage (full)

cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Ponerinae; Ponerini; Anochetus

External Information Resources (NCBI LinkOut)

LinkOut	Subject	LinkOut Provider
Anochetus grandidieri	taxonomy/phylogenetic	AntWeb
DNA barcoding : Anochetus grandidieri	taxonomy/phylogenetic	Barcodes of Life
search GBIF	taxonomy/phylogenetic	Global Biodiversity Information Facility
search HNS	taxonomy/phylogenetic	Hymenoptera Name Server
Anochetus grandidieri Forel, 1891	taxonomy/phylogenetic	Integrated Taxonomic Information System
Anochetus madecassus	taxonomy/phylogenetic	Plazi
ANOCHETUS GRANDIDIERI	taxonomy/phylogenetic	

Treatment mining proposal

Treatment mining proposal

NCBI Taxonomy Browser

Entrez | PubMed | Nucleotide | Protein | Genome | Structure | PMC | Taxonomy | Books

Search for [ ] as [complete name] [lock] [Go] [Clear]

Display [3] levels using filter: [none]

Anochetus grandidieri

Taxonomy ID: 269309
Inherited blast name: **ants**
Rank: species
Genetic code: Translation table 1 (Standard)
Mitochondrial genetic code: Translation table 5 (Invertebrate Mitochondrial)
Other names:
synonym: **Anochetus sp. BLF m2**
synonym: **Anochetus madecassus**
synonym: **Anochetus grandidieri Forel, 1891**

Entrez records	
Database name	Direct links
Nucleotide	189
Protein	162
Popset	2
PubMed Central	1
Taxonomy	1

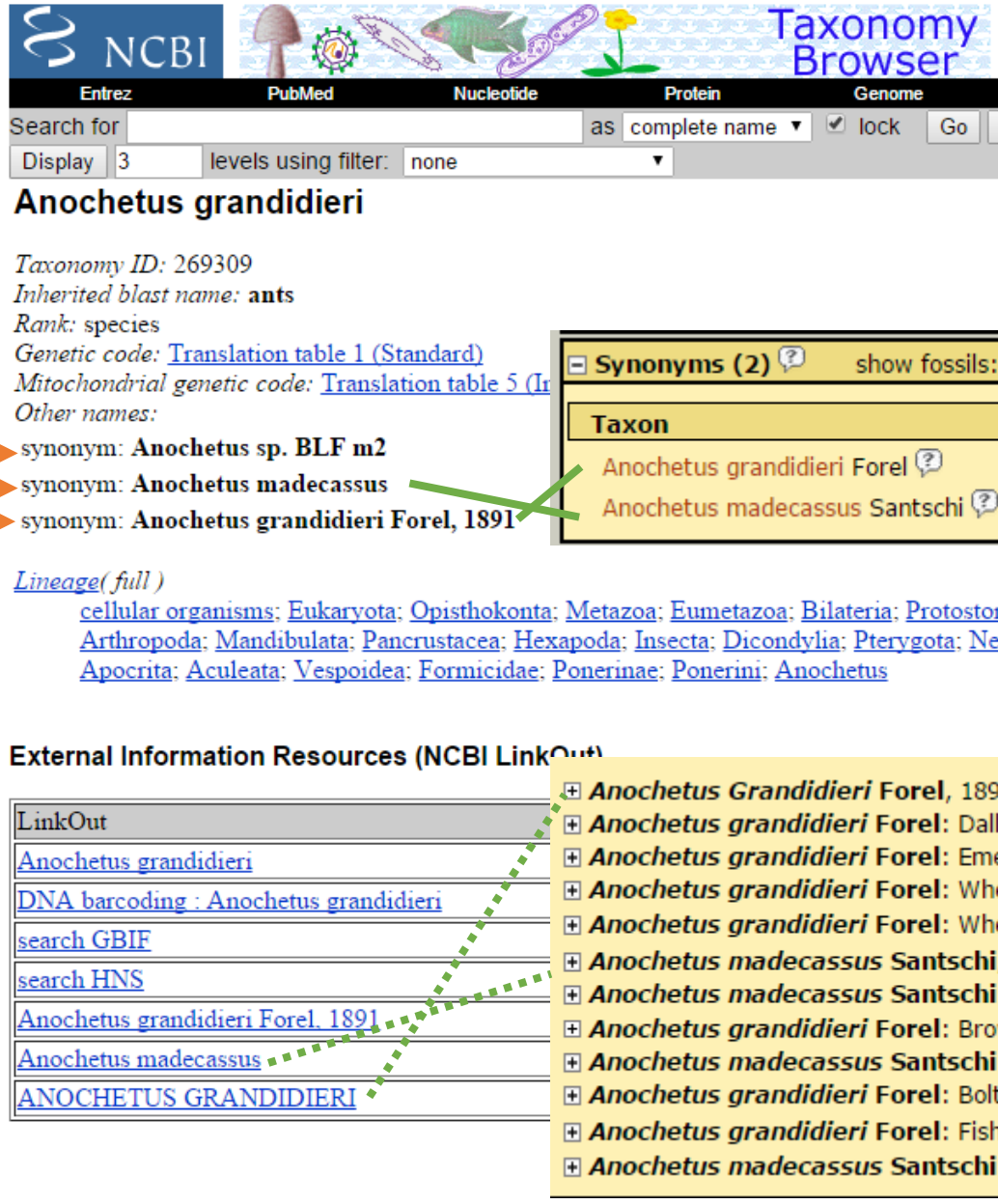
Lineage (full)

cellular organisms; Eukaryota; Opisthokonta; Metazoa; Eumetazoa; Bilateria; Protostomia; Ecdysozoa; Panarthropoda; Arthropoda; Mandibulata; Pancrustacea; Hexapoda; Insecta; Dicondylia; Pterygota; Neoptera; Endopterygota; Hymenoptera; Apocrita; Aculeata; Vespoidea; Formicidae; Ponerinae; Ponerini; Anochetus

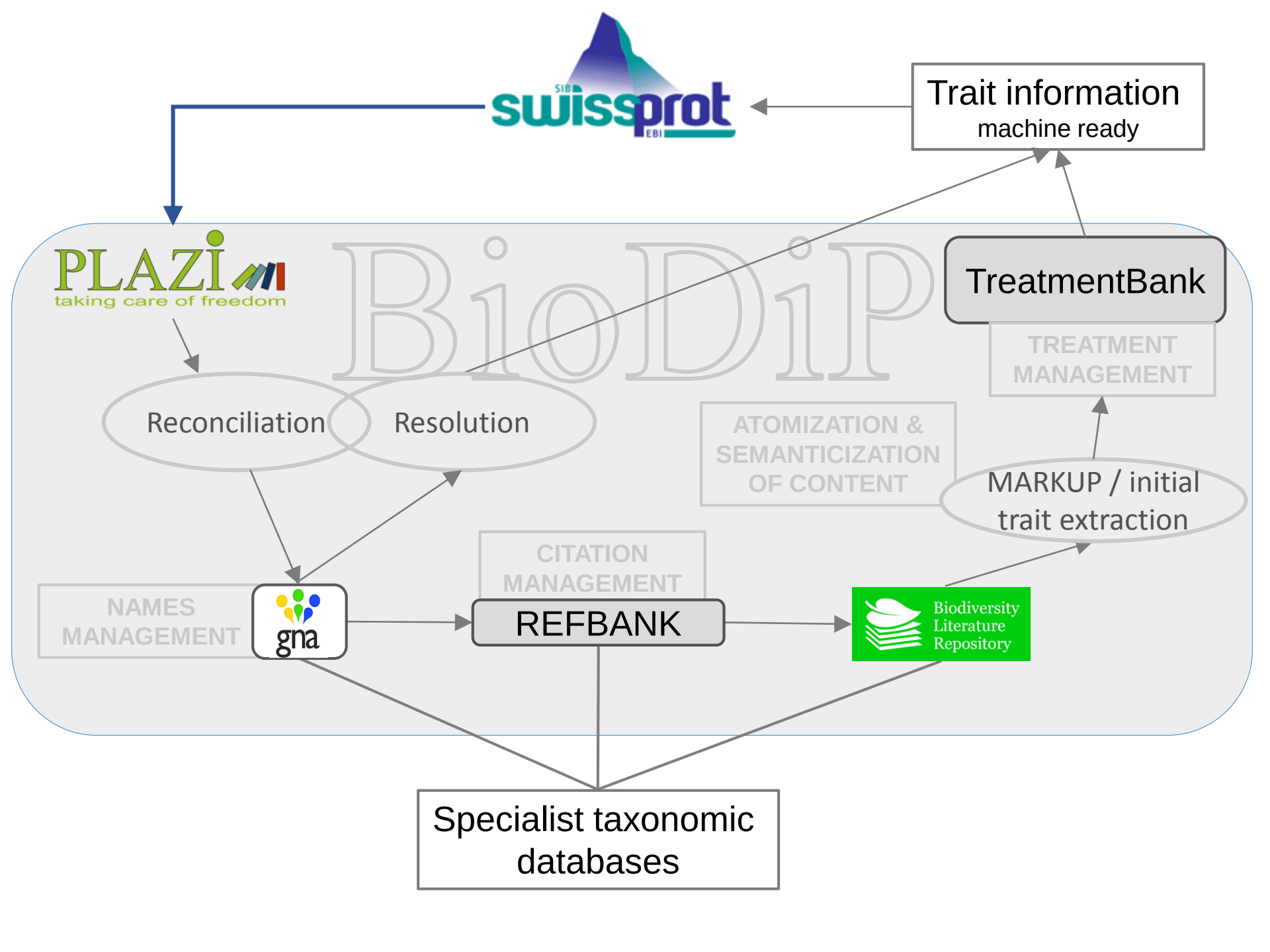
External Information Resources (NCBI LinkOut)

LinkOut	Subject	LinkOut Provider
Anochetus grandidieri	taxonomy/phylogenetic	AntWeb
DNA barcoding : Anochetus grandidieri	taxonomy/phylogenetic	Barcodes of Life
search GBIF	taxonomy/phylogenetic	Global Biodiversity Information Facility
search HNS	taxonomy/phylogenetic	Hymenoptera Name Server
Anochetus grandidieri Forel, 1891	taxonomy/phylogenetic	Integrated Taxonomic Information System
Anochetus madecassus	taxonomy/phylogenetic	Plazi
ANOCHETUS GRANDIDIERI	taxonomy/phylogenetic	

Treatment mining proposal



Treatment mining proposal


swissprot
Trait information
machine ready
PLAZI
taking care of freedom
BioDiP
TreatmentBank
TREATMENT MANAGEMENT
Reconciliation
Resolution
ATOMIZATION & SEMANTICIZATION OF CONTENT
MARKUP / initial trait extraction
CITATION MANAGEMENT
NAMES MANAGEMENT
gna
REFBANK
Biodiversity Literature Repository
Specialist taxonomic databases

Treatment mining proposal

Taxonomic coverage:

Drosophilds:

4279 species level taxa (6633 available species level names)
16,000+ bibliographic references and PDF copies

Ants:

13,000 species level taxa (60,000 taxonomic name usages)
ca 5000 bibliographic references and PDF copies

Spiders:

43,000 species level taxa
ca 12000 bibliographic references and PDF copies

All taxa are completely covered and the data is or will be available, and the bibliographic refernces and PDFs will be hosted on BLR

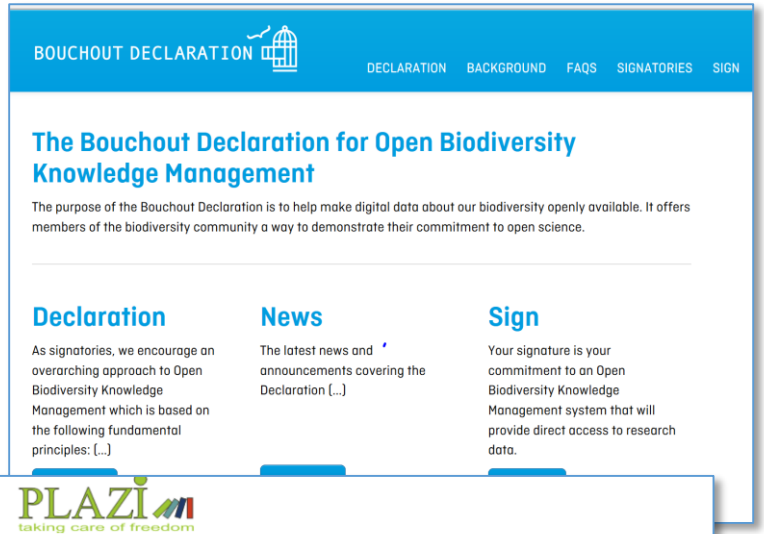
How do we do this?

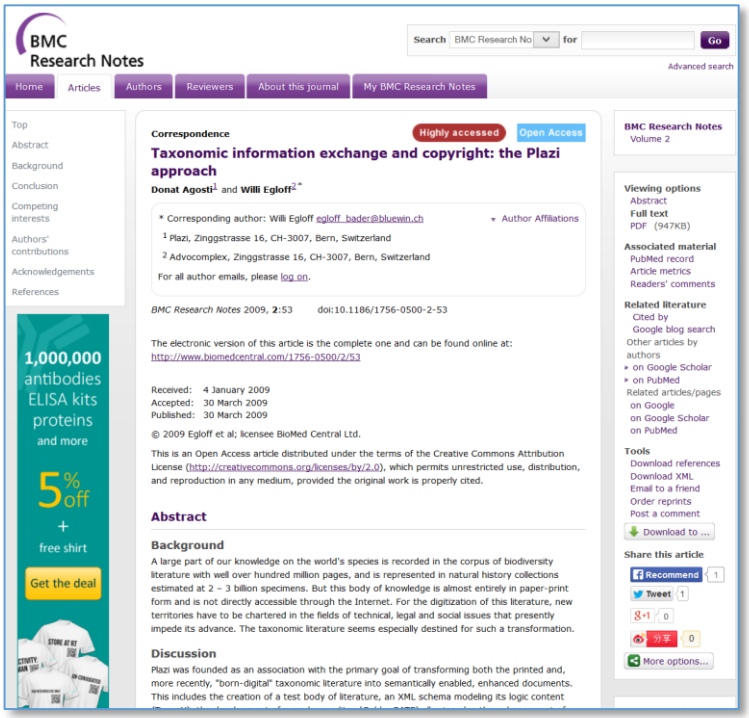
Copyright issues

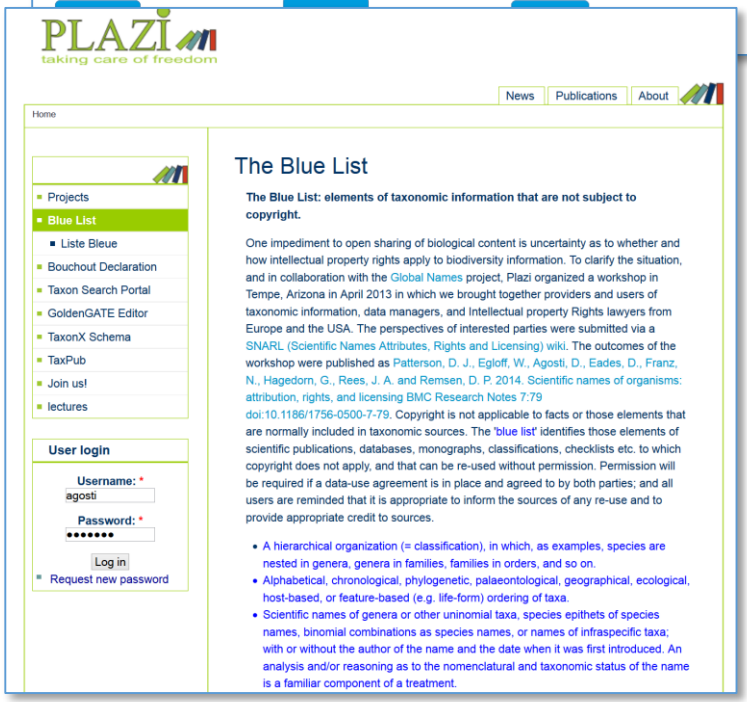
Bouchout Declaration

Blue List

Plazi workflow

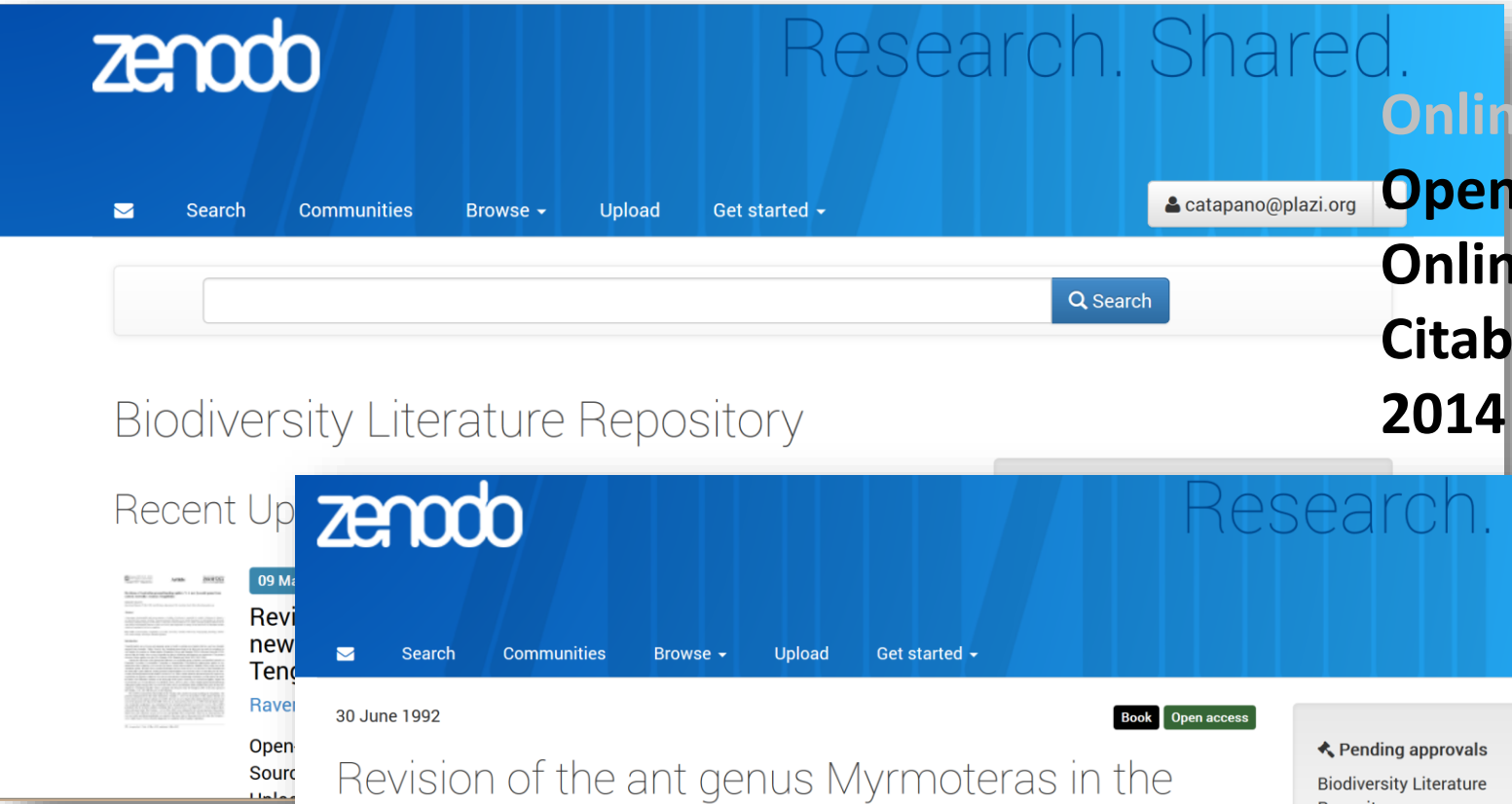






Literature archive

Access, archive, DOI



Online catalogue

Open access

Online library

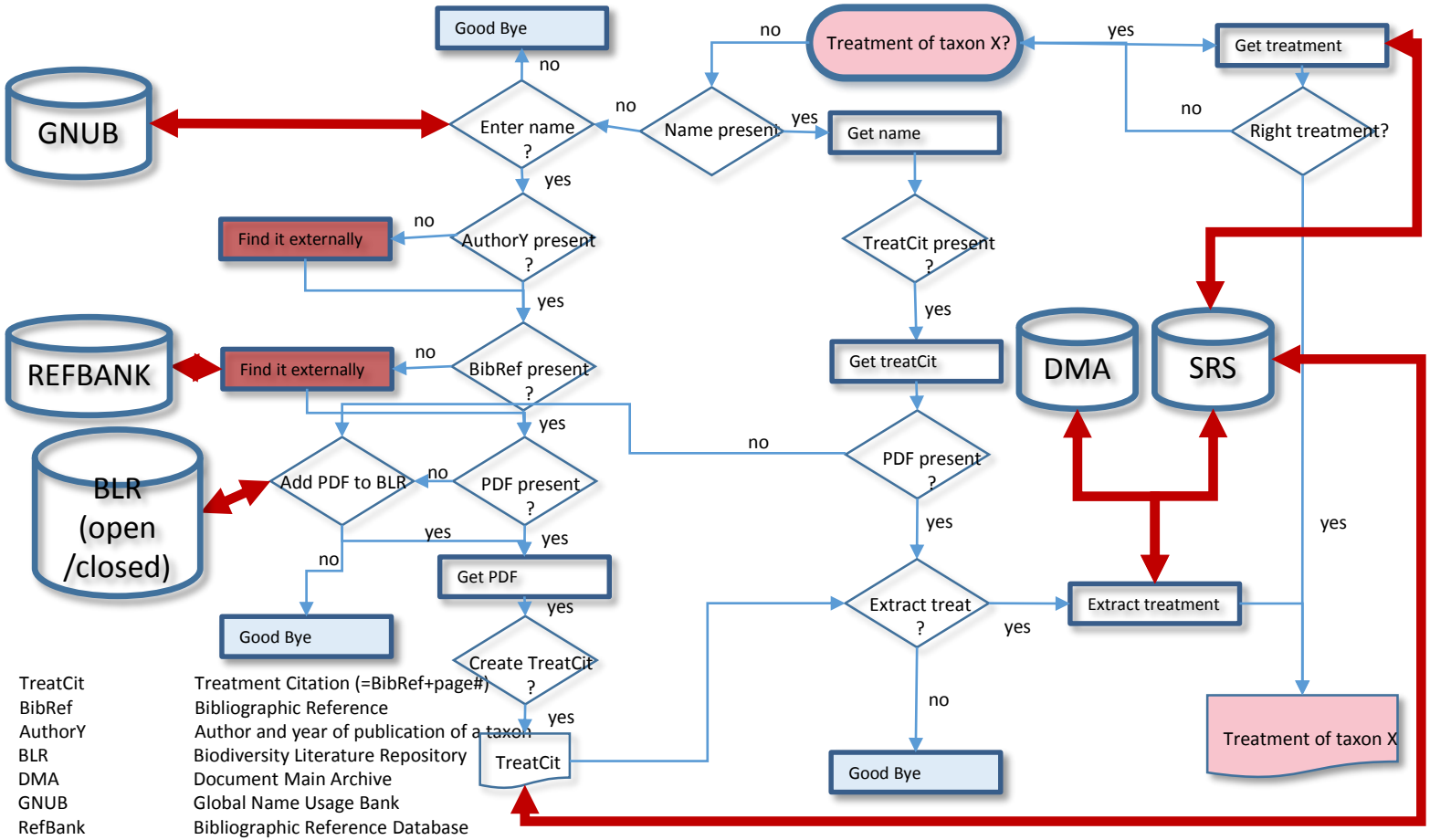
Citable references

2014

[See Grimaldi and images on BLR]

Publication conversion workflow

«please give me the treatment of taxon X»

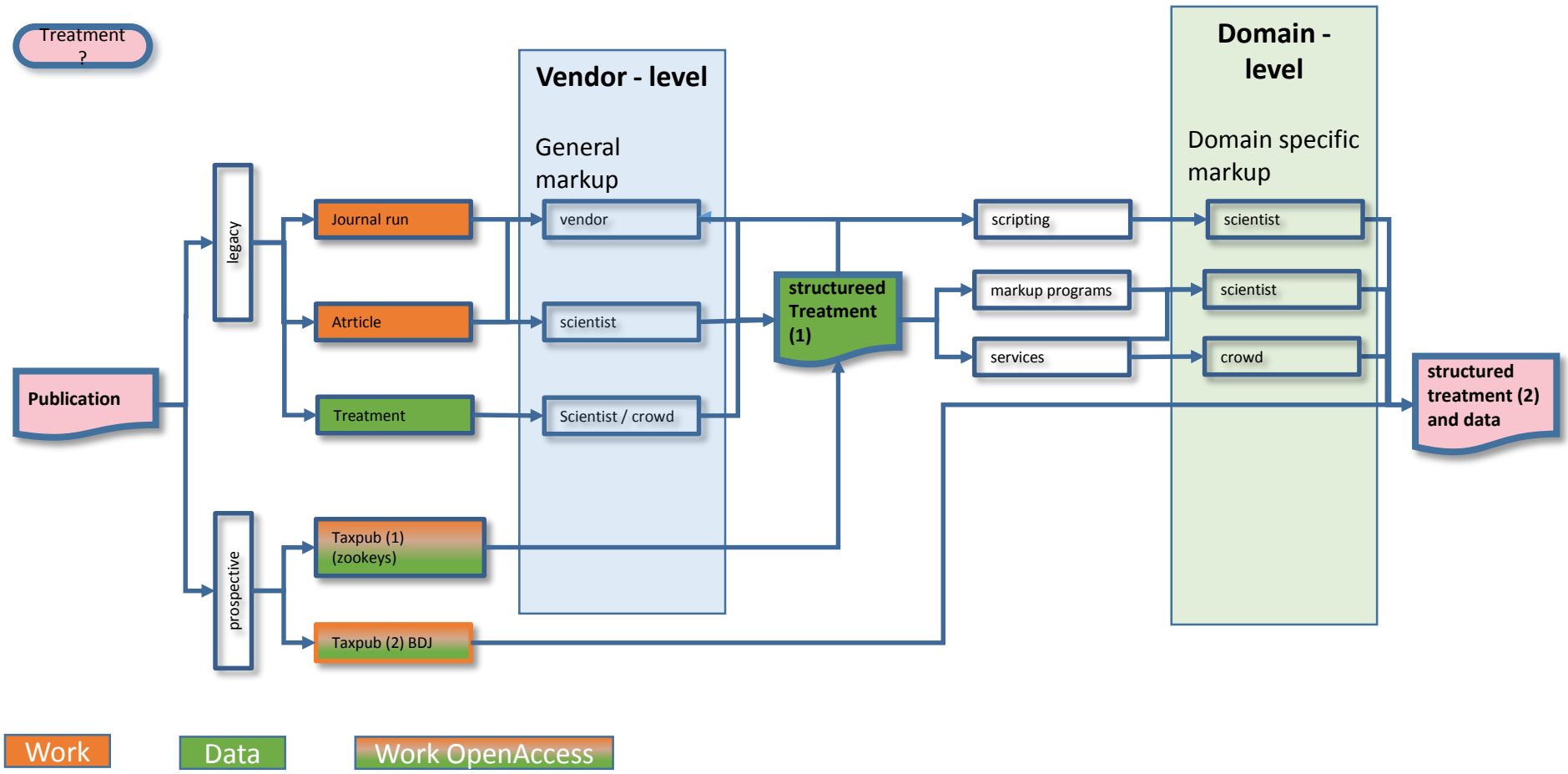


TreatCit	Treatment Citation (=BibRef+page#)
BibRef	Bibliographic Reference
AuthorY	Author and year of publication of a taxon
BLR	Biodiversity Literature Repository
DMA	Document Main Archive
GNUB	Global Name Usage Bank
RefBank	Bibliographic Reference Database
SRS	Treatment Server

Plazi workflow

Treatment

1. Publication conversion workflow



Markup

Markup

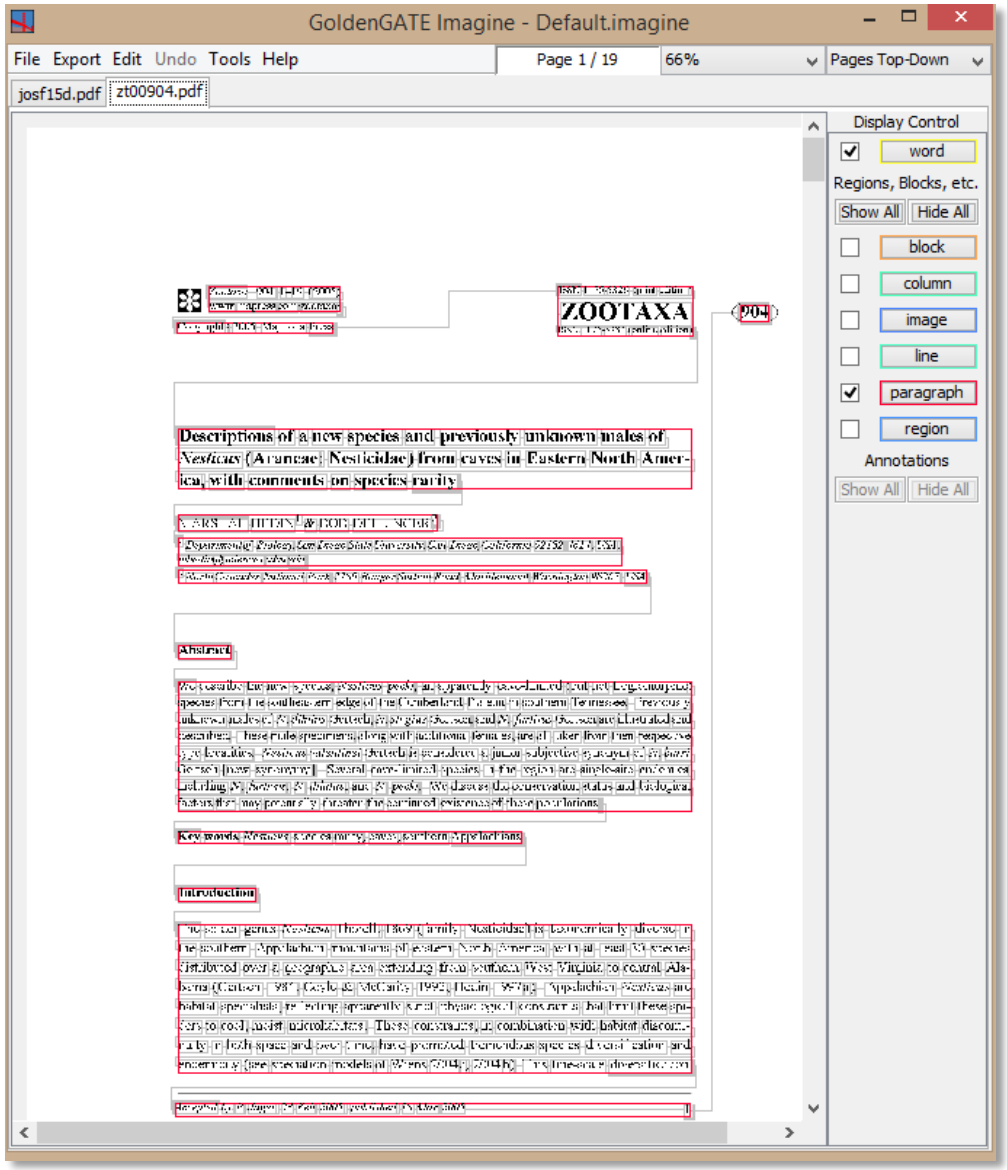
Goal: high quality data based on high quality extracted unstructured text from images to digital born PDFs.

Emphasis on extraction of treatments and substructure, scientific names, , various codes, treatment citations, bibliographic references, observation records, traits, figure legends and figures, tables, and linking to external resources.

Markup

Solution: After extensive testing decision to build our own, modular, open source extraction tool *GoldenGate Imagine*.

Demonstration of Imagine



Markup

Status quo: Functional prototype desktop application in daily use.

Next steps: Build a server side application that will be integrated as automatic step after upload of new documents on BLR.
Improve UI.

Conclusions

Avoid create legacy data….

Thanks!

Donat Agosti
agosti@plazi.org

Terry Catapano
catapanoth@gmail.com