

Access to Biodiversity Information: From Printed to Semantic, Enhanced e-Publications

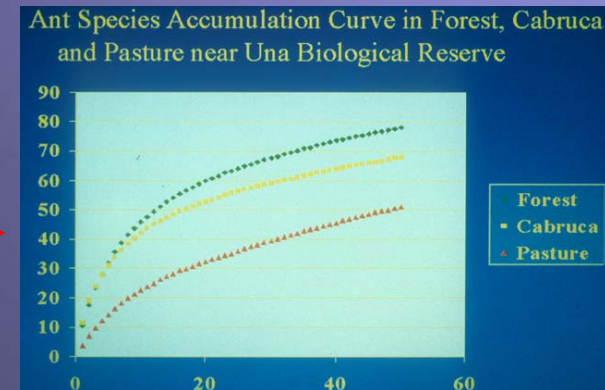
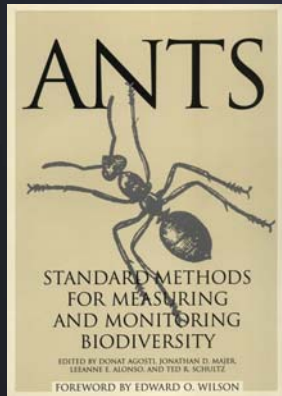
Donat Agosti

Plazi, Switzerland

March 19, 2008

NDAP International Conference, Taipei

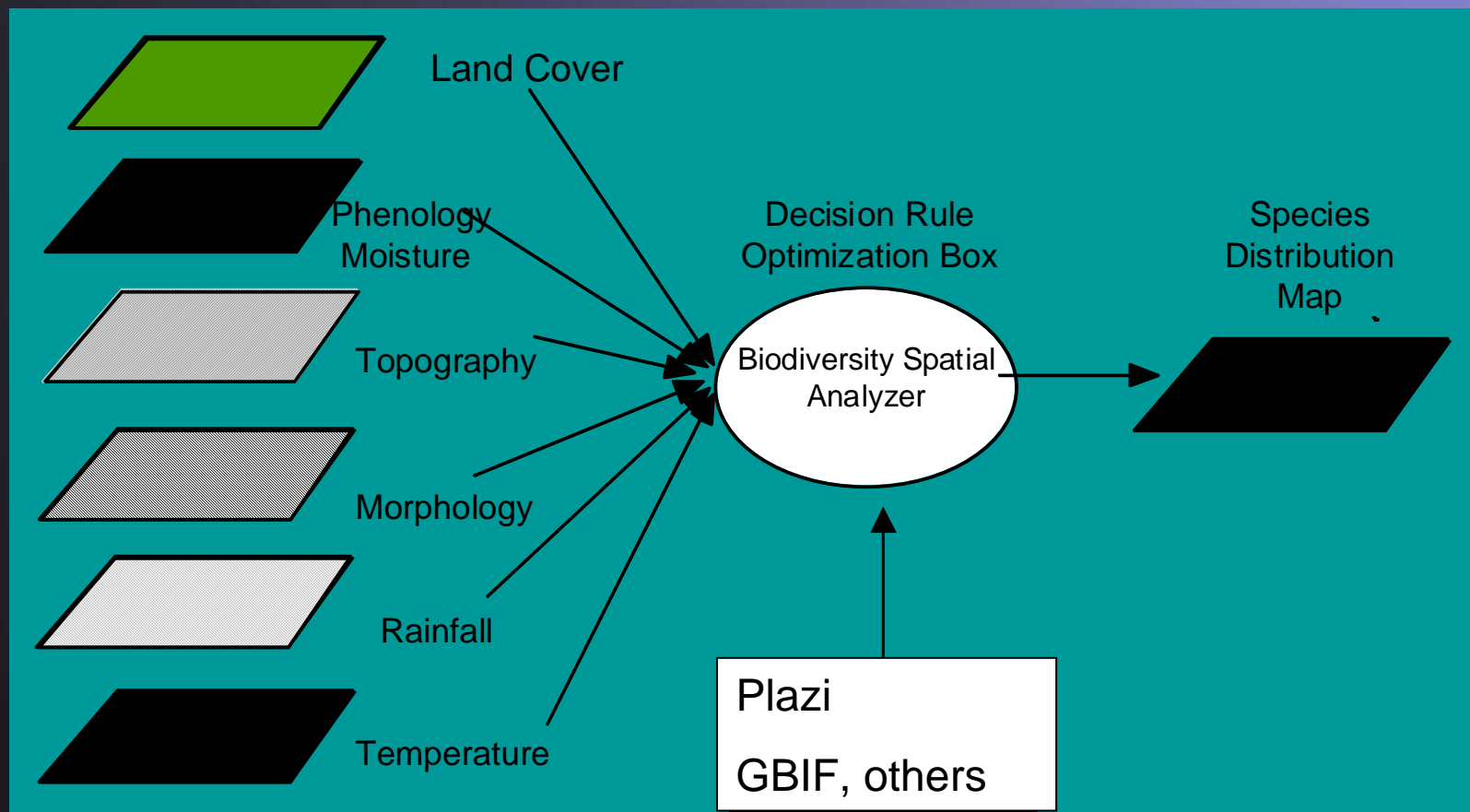
Biodiversity monitoring, or what's out there?



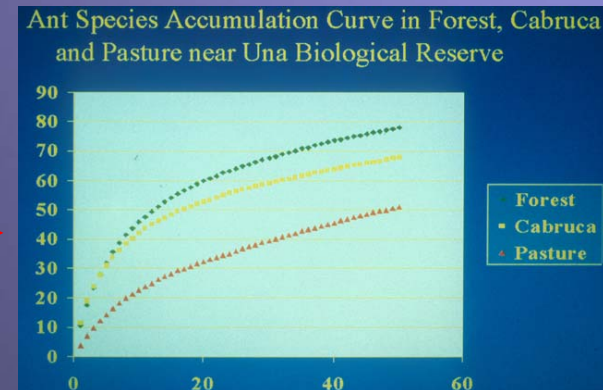
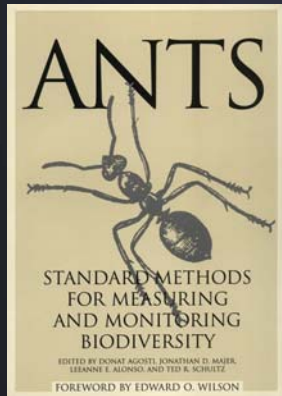
Measuring and monitoring biodiversity means standard repetitive samples:

- Access to taxonomic data is one of the main impediments to run successful surveys and to integrate survey into mainstream conservation, potentially one of the biggest user of taxonomic data

Biogeographic analysis, eg for impact of climate change



Biodiversity monitoring, or what's out there?



Measuring and monitoring biodiversity means standard repetitiv samples:

- How can we provide the fastest way this content?
- How can we enter new data, ie describe new species?
- What is doable, and what not?

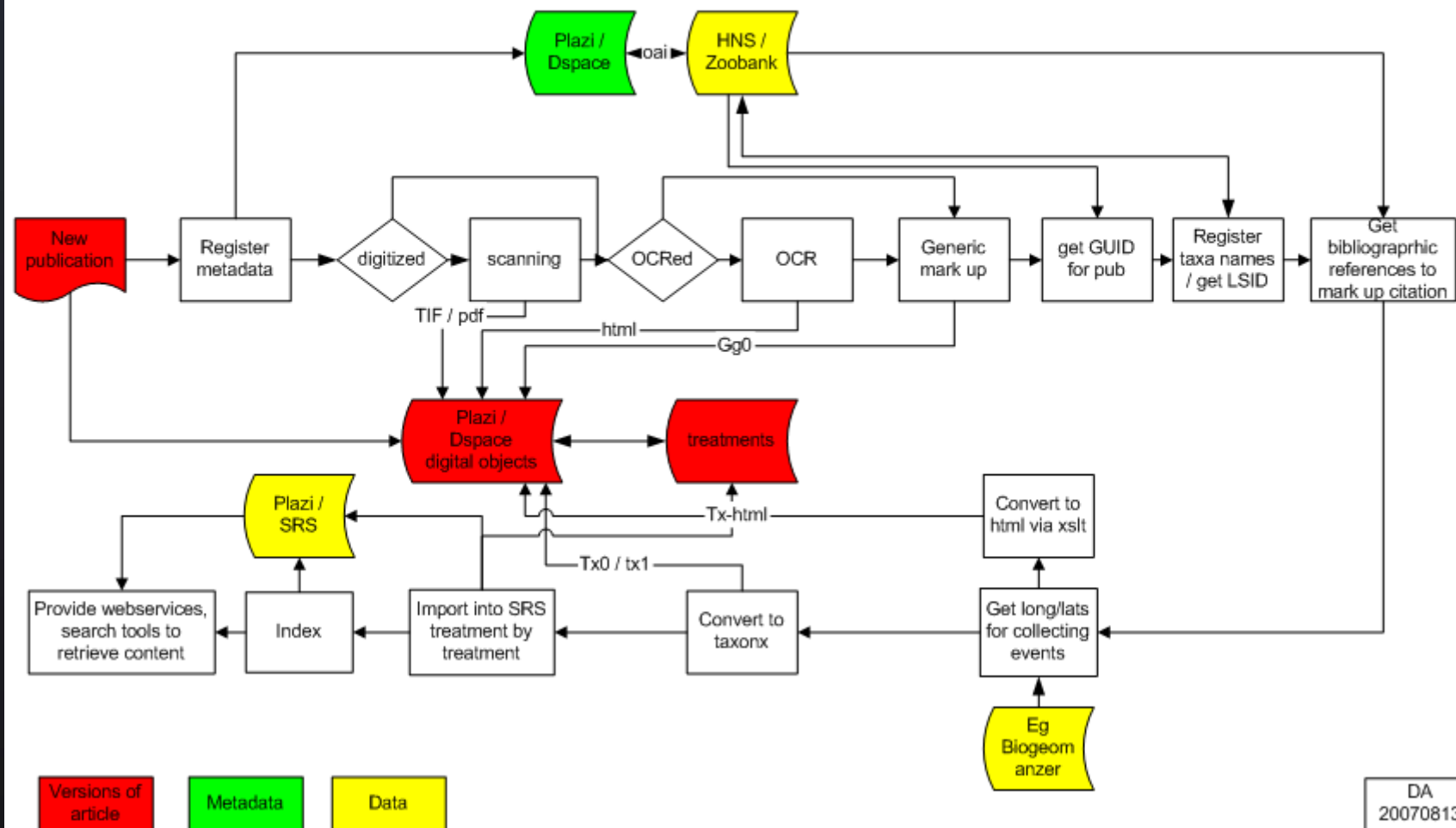
A report from a break through in a long tunnel....



For the first time, the entire production chain of ocr-ing, marking up, adding all the guids to produce a valid taxonx document is in place

Plazi is providing data/metadata of taxonomic descriptions other applications can utilize (e.g. iSpecies or EOL)

The Plazi Taxonomic Literature Archive Project

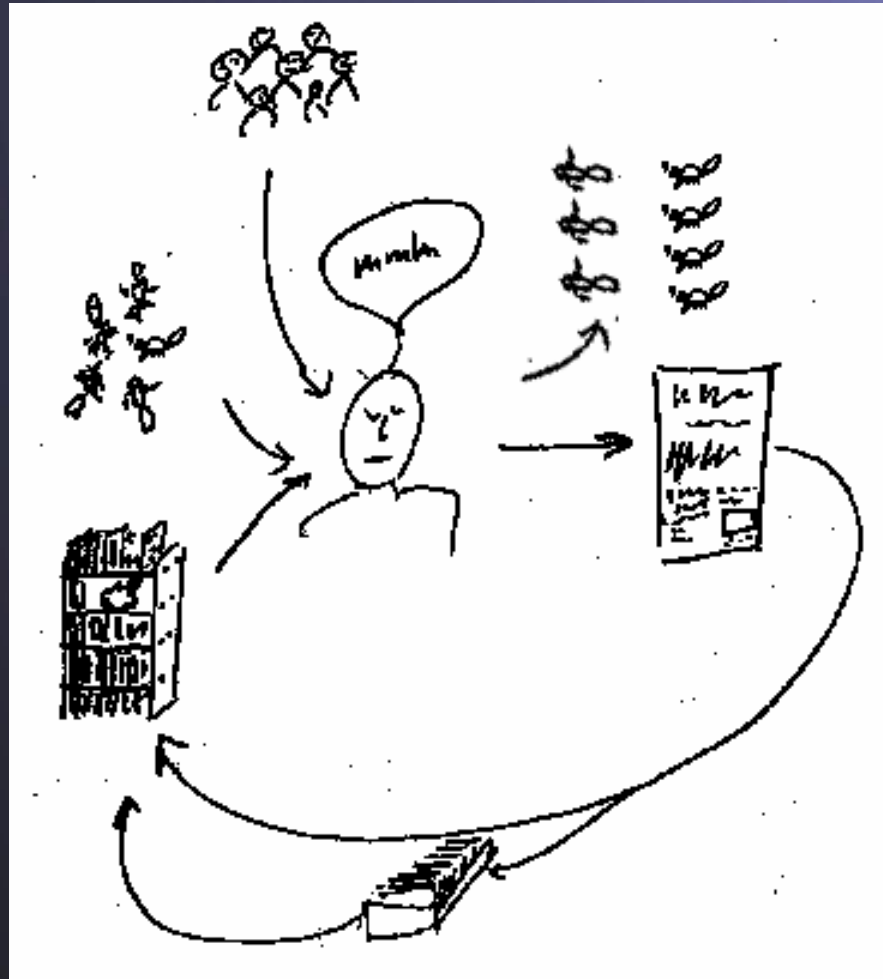


Taxonomists at work



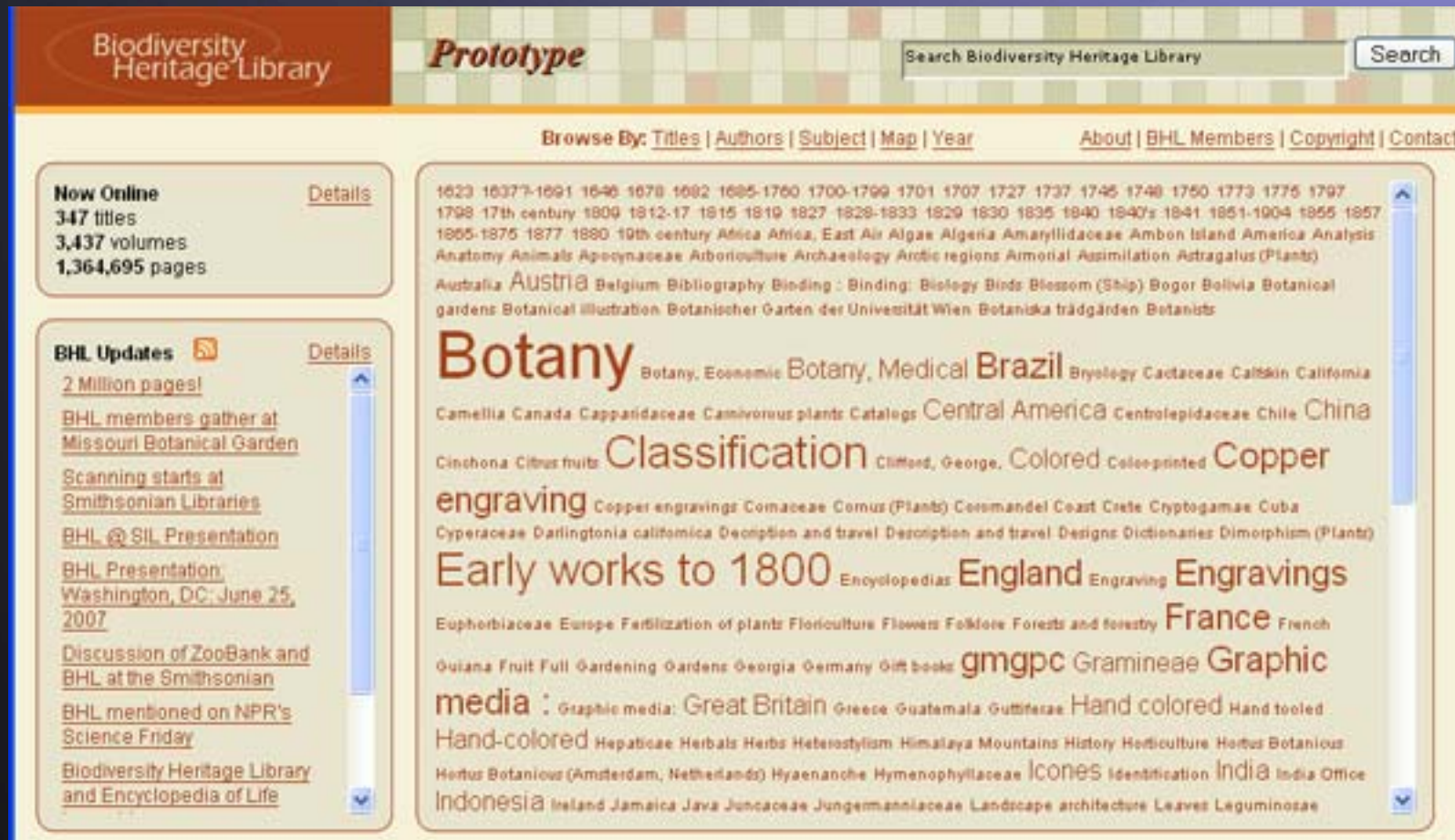
„Taxonomic impediment“, Global Taxonomy Initiative, ...

The traditional flux of information



...a more or less closed, intransient system

but it delivered > 90 million pages, part of which are currently being scanned and ocred through the Biodiversity Heritage Library Project, most of them out of copyright, ie older than 1925.

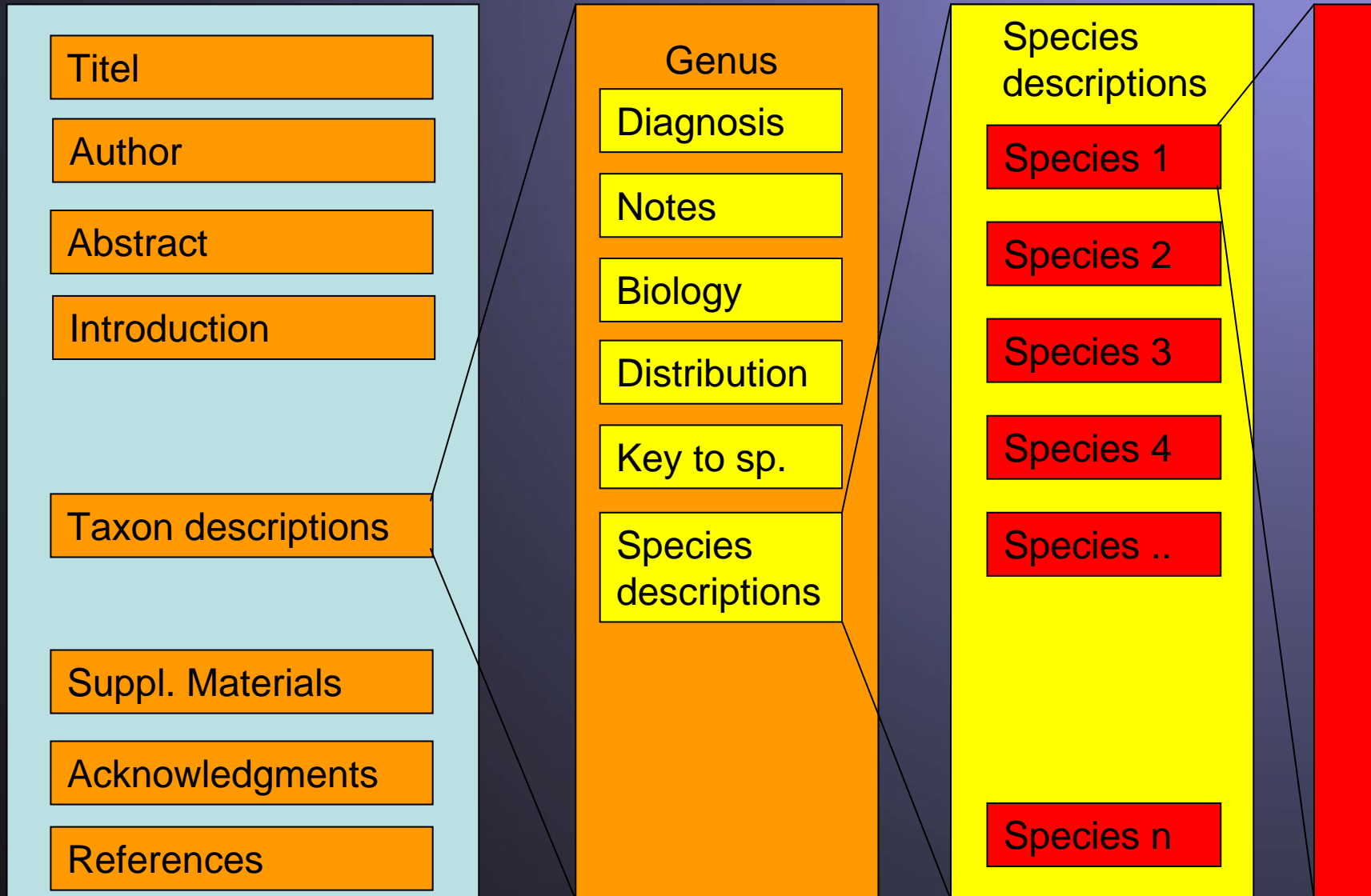


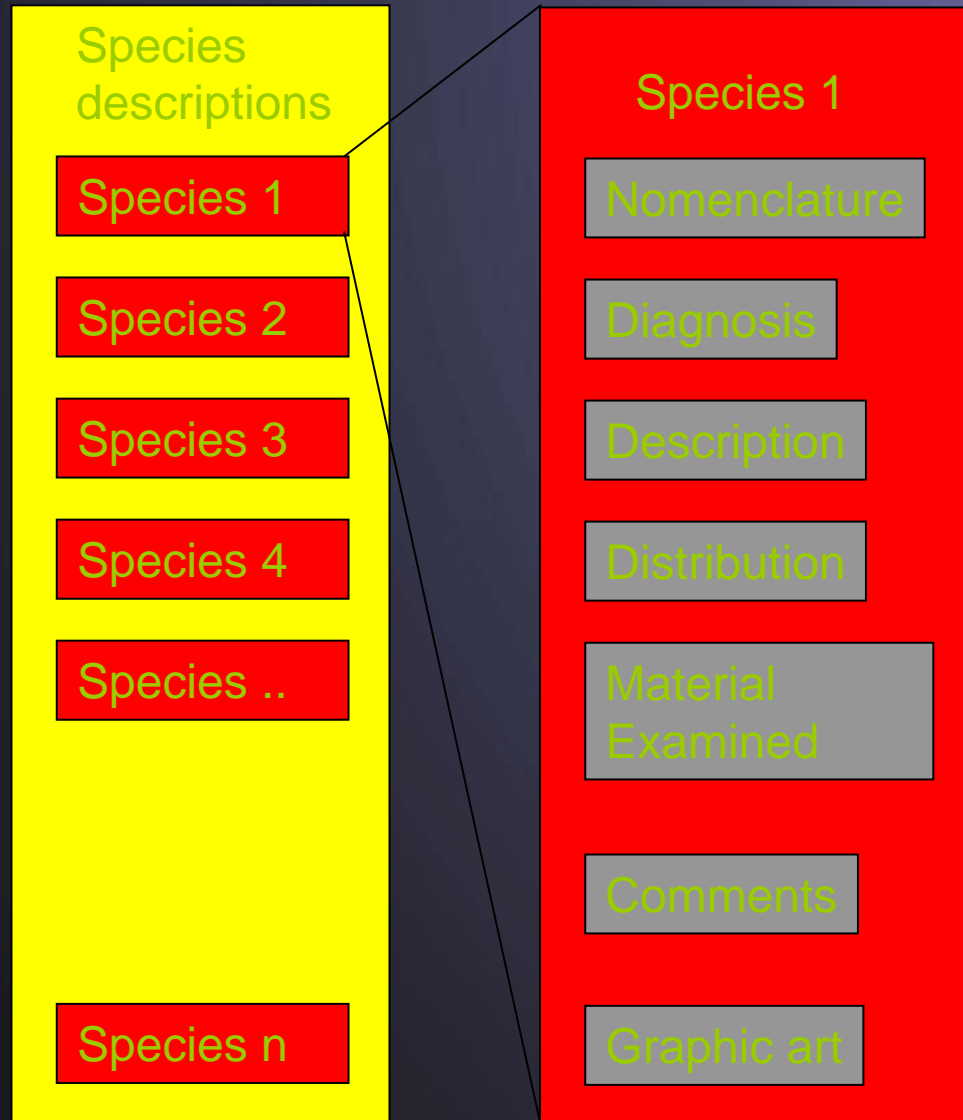
The screenshot shows the Biodiversity Heritage Library (BHL) website interface. At the top left, the logo reads "Biodiversity Heritage Library". To its right is the word "Prototype" in a stylized font. Further right is a search bar with the text "Search Biodiversity Heritage Library" and a "Search" button. Below the search bar, there are navigation links: "Browse By: Titles | Authors | Subject | Map | Year" and "About | BHL Members | Copyright | Contact".

On the left side, there are two boxes. The top one is titled "Now Online" and contains the following statistics: "347 titles", "3,437 volumes", and "1,364,695 pages". A "Details" link is next to it. The bottom box is titled "BHL Updates" and contains several links: "2 Million pages!", "BHL members gather at Missouri Botanical Garden", "Scanning starts at Smithsonian Libraries", "BHL @ SIL Presentation", "BHL Presentation, Washington, DC, June 25, 2007", "Discussion of ZooBank and BHL at the Smithsonian", "BHL mentioned on NPR's Science Friday", and "Biodiversity Heritage Library and Encyclopedia of Life".

The main content area is a large list of search results. The top result is "Botany", which is highlighted in a larger font. Below it, there are several other results, including "Brazil", "Classification", "Early works to 1800", "England", "Engravings", "France", "Graphic media", and "India". Each result is followed by a list of related terms and a vertical scrollbar on the right side of the list.

Systematics publication are highly structured, homogenous and rich in data





Species (or taxa in case of higher level revisions) descriptions can be considered the building blocks or basic data elements of taxonomic publications. They are very rich in detail. All the other elements of a publications, are inferred from the analysis and synthesis of taxon descriptions. They could be compared albeit at a higher level, to DNA-sequence, unique to a particular taxon.

The descriptions are also the ,legal' element of the publication in compliance with the ICZN.

Large homogeneous data sets are more used in science than small heterogeneous data sets.

Typical examples of such data sets are data from astronomy, particle physics or remote sensing, often produced in terabytes.

Taxonomic publications constitute a large body, but are not often used...

... because they are not accessible.



- 12,028 ant species are known today (ca 1.8 M species total)
- ca 22,000 expected ant species (ca. 10 – 50 M)
- discovery rate of >100 species per year (22,000)

The descriptions for ants are highly scattered

- 2,588 publications
- produced by 446 authors
- in 511 journals from all around the world
- and 103 books
- ca 80.000 pages
- only recently descriptions in electronic media



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.

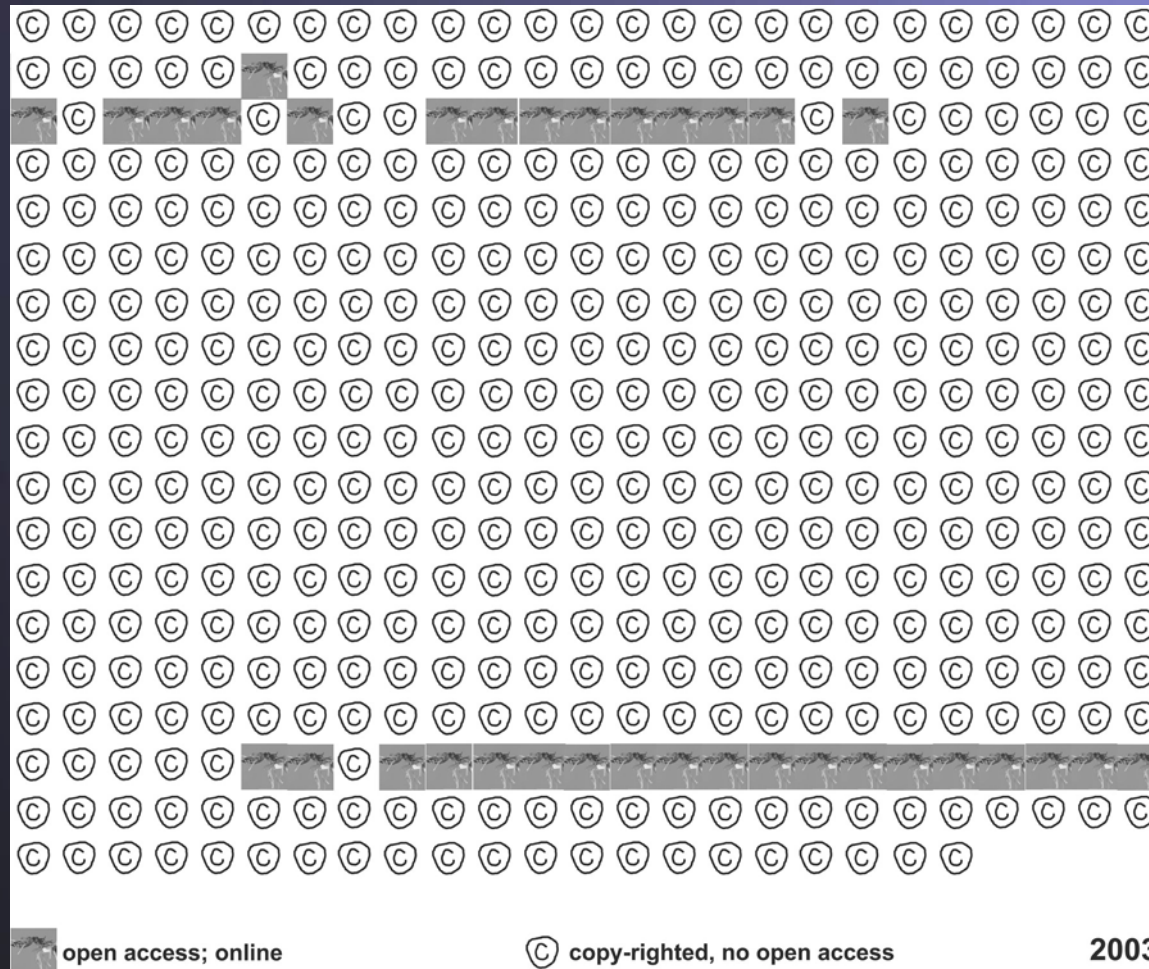


109775 visits from 28 Aug 2006 to 29 Aug 2007

⊞ distance in which individuals are clustered
 Total number of visits depicted above = 109457

Dot sizes:

● = 1000 + ● = 100 - 999 ● = 10 - 99 ◆ = 1 - 9



Access to ant taxonomic publications through antbase.org /Smithsonian Institution, including currently the entire body of non-copyrighted publications since 1758 (>4,000 publications or 85,000 pages. Source: (Agosti 2005 and antbase.org)

For a bright future, we must remove three kinds of barriers to provide access to our data and publications.

- Legal
- Technical
- Social

Legal consideration

The question should not be, **are** descriptions copyrighted, but whether they **can** be copyright.

Copyright law is national, and thus very complicated to understand, and to apply in a global environment like the Internet.

In Switzerland (and in many other countries) something has to qualify as a „work“ to be protected by copyright.

“work” does not mean “text”, does not mean “data”, does not mean “information”. “Work” is something more. That kind of something more has many different definitions in the various legislations, but it is always there: It may be called originality, individuality, creation, personal expression, creative shaping or anyhow else, but it is a condition for qualifying a product as a work: “Work” is an intellectual product that is in a certain sense particular, individual, original, new.

(Willi Egloff, Kew Gardens, Feb 20, 2008)

Taxonomic publications are highly structured and homogenous, part of a global >100 million page corpus growing at a rate of ca 20,000 new species descriptions per year, not counting redescriptions. Its structure is tightly controlled by a peer review process enforcing standards, a domain specific vocabulary and not written as poem or in flowery language but scientific jargon .

It is thus hardly original work, and thus its content can not be copyrighted.

Thus the legal barrier is removed.


Technical barriers

Legacy publications are structured for human consumption, but not for machine consumption.

TaxonX (or TaXMLit) are XML schemas modeling the logic content of a publication and thus can be used to transform publications into machine readable form.

XML conversion from print to machine readable text

International Commission on Zoological Nomenclature



HOME SITE MAP SEARCH CONTACT US

Index: Pages | Taxonomy

Page [1]
Page [2]
Page [3]
Page [4]
Page [5]
Page [6]
Page [7]
Page [8]
Page [9]
Page [10]
Page [11]
Page [12]
Page [13]
Page [14]
Page [15]
Page [16]
Page [17]
Page [18]
Page [19]
Page [20]
Page [21]
Page [22]
Page [23]
Page [24]
Page [25]
Page [26]
Page [27]
Page [28]
Page [29]
Page [30]
Page [31]
Page [32]
Page [33]
Page [34]
Page [35]
Page [36]
Page [37]
Page [38]
Page [39]
Page [40]
Page [41]
Page [42]
Page [43]
Page [44]
Page [45]
Page [46]
Page [47]
Page [48]
Page [49]
Page [50]

INSECTA HYMENOPTERA. Apts. 579

Lapidaria 31. *A. hirsuta atra*, ano fulvo. *Fn. fvec.* 1015.
Prjch. inf. 9. p. 25. n. 2.
Reaum. inf. 6. t. 1. f. 1-4.
Habitat in acervis Lapidum.

Mulcorum 32. *A. hirsuta fulva*, abdomine flavo. *Fn. fvec.* 1017.
Prjch. inf. 9. p. 26. n. 8.
Reaum. inf. 6. t. 2. f. 1-3.
Habitat sub Mulco terrestri.

Hypnorum 33. *A. hirsuta fulva*, abdominis falcia nigra, ano albo.
Fn. fvec. 1018.
Reaum. inf. 6. t. 4. f. 1.
Habitat sub Mulcis.

acervorum 34. *A. hirsuta atra*. *Fn. fvec.* 1013.
Habitat sub terra.

subterranea 35. *A. hirsuta atra*, ano fulvo. *Fn. fvec.* 1014.
Habitat sub terra.

surinamensis 36. *A. hirsuta nigra*, abdomine, excepto primo segmento, flavo.
Habitat Surinam. Rolander.

aethians 37. *A. hirsuta nigra*, thorace flavo. *M. L. U.*
Habitat in calidis regionibus.

tropica 38. *A. hirsuta nigra*, abdomine postice flavo. *M. L. U.*
Habitat in calidis regionibus.

alpina 39. *A. hirsuta*, thorace nigro, abdomine luteo. *Faun. fvec.* 1016.
Habitat in Lapponiae alpidibus.

218. FORMICA. *Squamula erecta* thoraci abdominisque interjecta.
Aculeus Feminis & Neutris reconditus.
Alae Maribus & Feminis, sed Neutris nullae.

herculeana 1. *F. nigra*, abdomine ovato, femoribus ferrugineis.
Fn. fvec. 1019. *Formica magna.*

INSECTA HYMENOPTERA. Apts. 579

Lapida- 31. *A. hirsuta atra*, ano fulvo. *Fn. fvec.* 1015.
Prjch. inf. 9. p. 25. n. 2.
Reaum. inf. 6. t. 1. f. 1-4.
Habitat in acervis Lapidum.

Mulco- 32. *A. hirsuta fulva*, abdomine flavo. *Fn. fvec.* 1017.
Prjch. inf. 9. p. 26. n. 8.
Reaum. inf. 6. t. 2. f. 1-3.
Habitat sub Mulco terrestri.

Hypno- 33. *A. hirsuta fulva*, abdominis falcia nigra, ano albo.
Fn. fvec. 1018.
Reaum. inf. 6. t. 4. f. 1.
Habitat sub Mulcis.

acervo- 34. *A. hirsuta atra*. *Fn. fvec.* 1013.
Habitat sub terra.

subter- 35. *A. hirsuta atra*, ano fulvo. *Fn. fvec.* 1014.
Habitat sub terra.

subter- 36. *A. hirsuta nigra*, abdomine, excepto primo segmento, flavo.
Habitat Surinam. Rolander.

aethians 37. *A. hirsuta nigra*, thorace flavo. *M. L. U.*
Habitat in calidis regionibus.

tropica 38. *A. hirsuta nigra*, abdomine postice flavo. *M. L. U.*
Habitat in calidis regionibus.

alpina 39. *A. hirsuta*, thorace nigro, abdomine luteo. *Faun. fvec.* 1016.
Habitat in Lapponiae alpidibus.

218. FORMICA. *Squamula erecta* thoraci abdominisque interjecta.
Aculeus Feminis & Neutris reconditus.
Alae Maribus & Feminis, sed Neutris nullae.

hercule- 1. *F. nigra*, abdomine ovato, femoribus ferrugineis.
Fn. fvec. 1019. *Formica magna.*

<tax:treatment>

<tax:nomenclature>

<tax:name>

<tax:tid source="HNS" identifi

<tax:xmldata>

<dc:Genus>Formica</dc:Genu

</tax:xmldata>FORMICA </tax:na

</tax:nomenclature>

<tax:div type="description">

<tax:p>Squamula erecta thoraci abd

<tax:p>Aculeus Feminis [and] Neutr

<tax:p>Alae Maribus [and] Feminis;

</tax:div>

</tax:treatment>

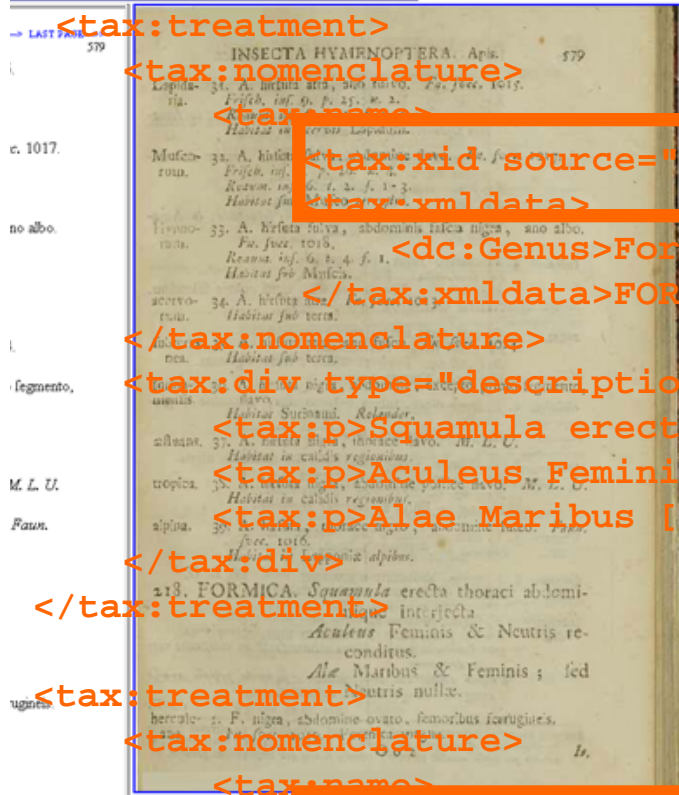
<tax:treatment>

<tax:nomenclature>

<tax:tid source="HNS" identifi

Text provided by the University Library of Cologne. Page Image Courtesy of the Biodiversity Heritage Library

... enhanced with links to external services like name servers, ontologies, specimen databases



Registration of zoological names
To be launched by January 1, 2008

<tax:xid source="LSID" identifier="urn:lsid:biosci.ohio-state.edu:osuc_concepts:13" />
<tax:xmldata>

Technical barriers

Legacy publications are structured for human, but not for machine consumption.

TaxonX is a XML schemas marking up the logic content of a publication.

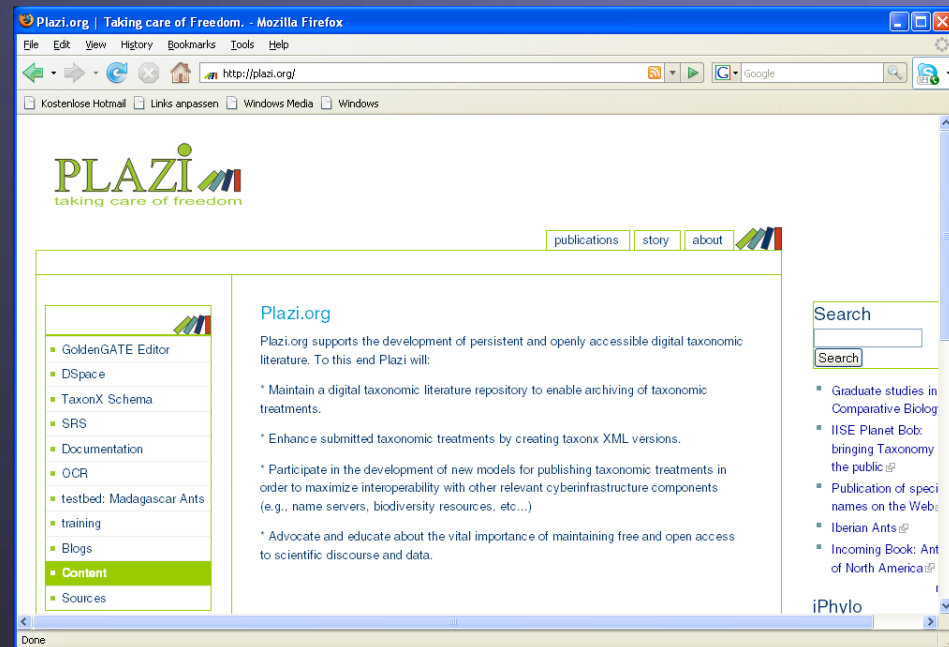
GoldenGate is an editor for semiautomatic mark up

Technical barriers (ctd.)



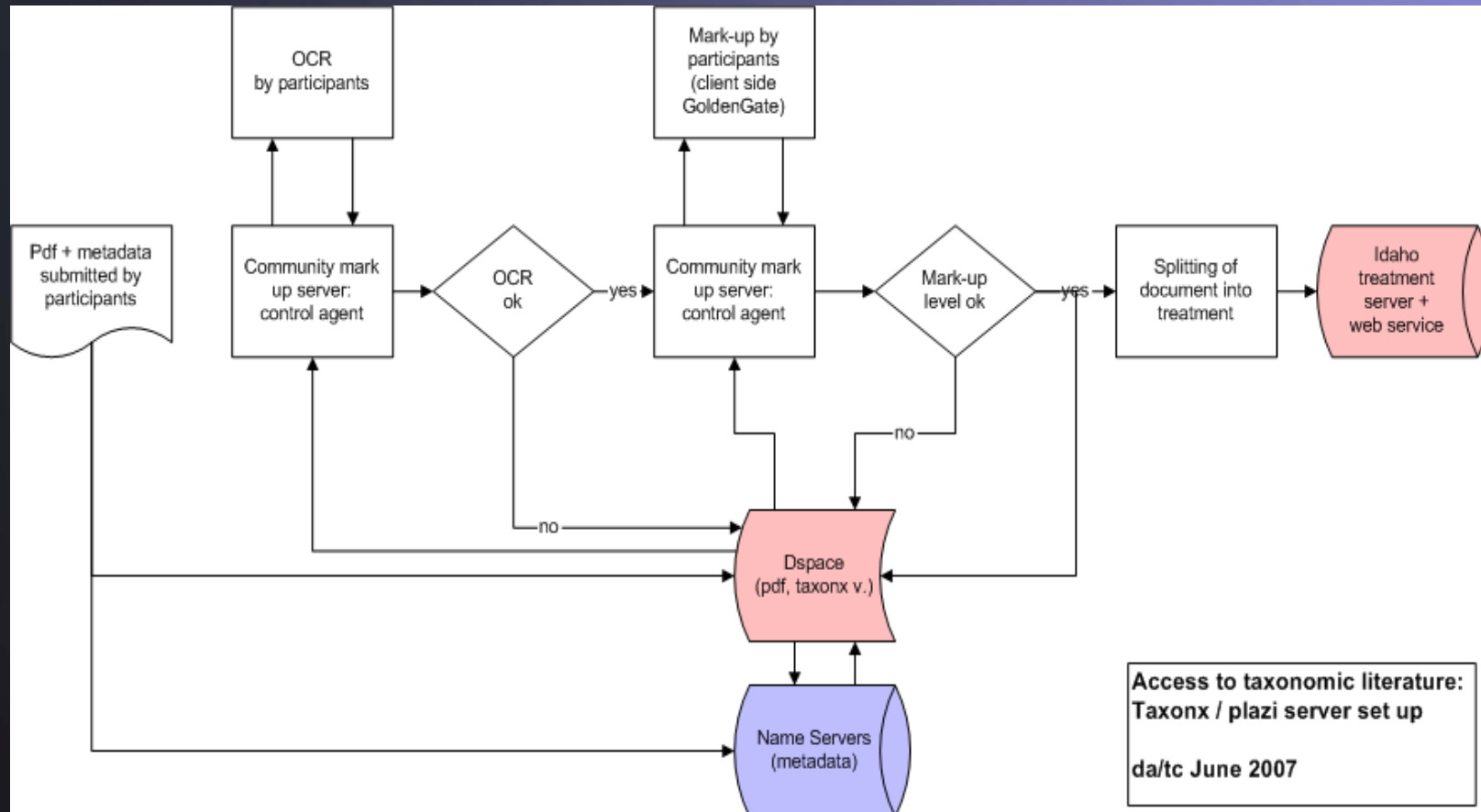
Plazi.org is a new Web site providing a service to archive marked-up taxonomic publications, and is incorporating all the tools from mark-up, to enhance and to retrieval of individual descriptions

In contrast to projects marking up floras, it is aimed at global taxonomic coverage requiring more versatility of the tools.



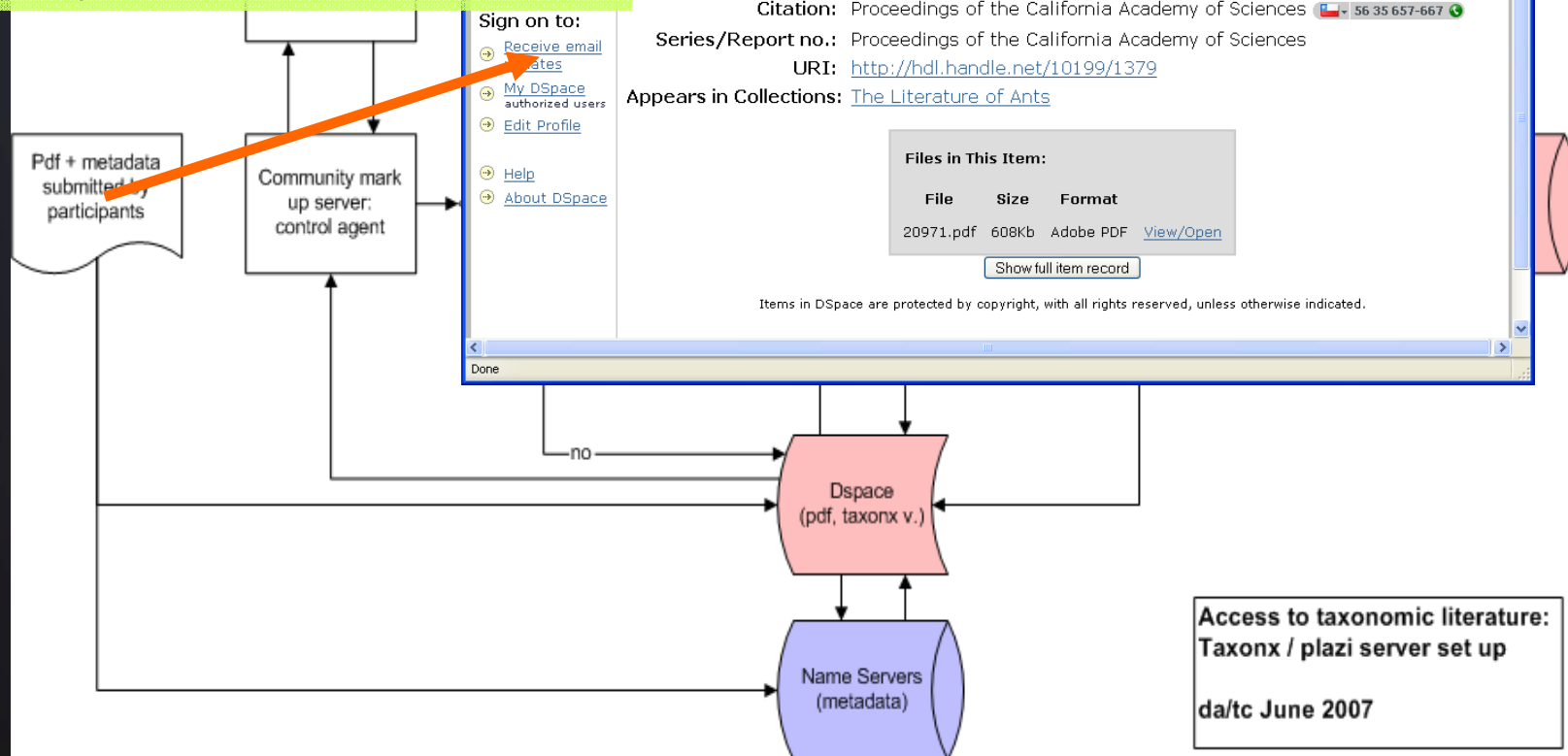
Technical barriers (ctd.)

Plazi.org: workflow



Plazi workflow

Import publications as pdf, text;
add metadata; Provide a unique
identiyer (handle)



Plazi served documents: pdf, xslt-html, xml

DSpace at Plazi.org: A new species of Discothryea Roger from Mauritius and a New species of Proceratium Roger from Madagascar

File Edit View History Bookmarks Tools Help

http://plazi.org:8080/dspace/handle/10199/1379

Kostenlose Hotmail Links anpassen Windows Media Windows

Google Docs publications to ... Docs & Spr... Creative Commons Licenses - ... DSpace at

Browse

- Communities & Collections
- Titles
- Authors
- Subjects
- By Date

Sign on to:

- Receive email updates
- My DSpace authorized users
- Edit Profile
- Help
- About DSpace

Please use this identifier to cite or link to this item
<http://hdl.handle.net/10199/1379>

Title: A new species of Discothryea Roger from Mauritius and a New species of Proceratium Roger from Madagascar (Hymenoptera: Formicidae)

Authors: Fisher, B. L.

Issue Date: 2005

Citation: Proceedings of the California Academy of Sciences

Series/Report no.: Proceedings of the California Academy of Sciences

URI: <http://hdl.handle.net/10199/1379>

Appears in Collections: [The Literature of Ants](#)

Files in This Item:

File	Size	Format	
20971.pdf	608Kb	Adobe PDF	View/Open

[Show full item record](#)

Items in DSpace are protected by copyright, with all rights reserved, unless otherwise indicated.

Pdf (original or scanned)

Zootaxa 1471: 1–14 (2007)
www.mapress.com/zootaxa
Copyright © 2007 · Magnolia Press

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5324 (online edition)

Description of two new *Temnothorax* species (Hymenoptera: Formicidae) from Italy

ANDREAS SCHULZ¹, JÜRGEN HEINZE² & KATJA PUSCH¹
¹Staatliches Museum für Naturkunde, Erbprinzenstr. 13, 70133 Karlsruhe, Germany. E-mail: schulz_ars@mus.nat.karlsruhe.de

Author Schulz, A. Author Heinze, J. Author Pusch, K. . Description of two new *Temnothorax* species (Hymenoptera: Formicidae) from Italy. Zootaxa. Zootaxa 1471: 1–14

Html via XSLT

Treatment: *Temnothorax alienus*, nov. spec.

Taxa Discussed

Temnothorax; *Temnothorax alienus*; *Temnothorax crassispinus*; *Temnothorax flavicornis*; *Temnothorax italicus*; *Temnothorax lichtensteini*; *Temnothorax luteus*; *Temnothorax nylanderi*; *Temnothorax parvulus*; *Temnothorax rabaudi*; *Temnothorax satunini*; *Temnothorax tianshanicus*;

Specimens Discussed:

[\[html\]](#) *Temnothorax alienus* nov. spec.

(Figs: 2, 3, 22, 26, 28, 29)

XML Taxonx

```

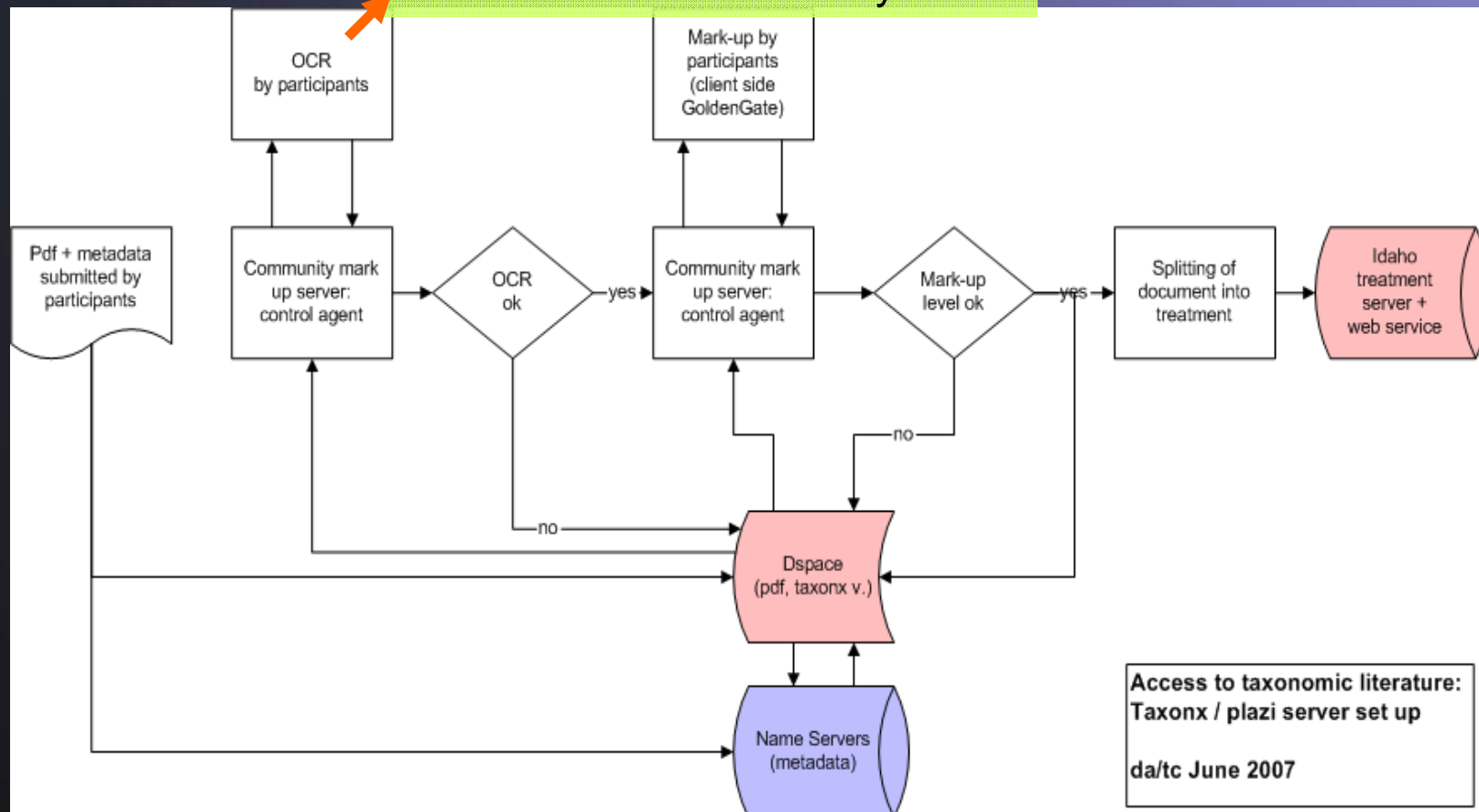
1 <?xml version="1.0" encoding="UTF-8" ?>
2 <tax:taxonx xmlns:dcc="http://digir.net/schema/conceptual/darwin/2003/1.0"
3   xmlns:mods="http://www.loc.gov/mods/v3" xmlns:tax="http://www.taxonx.org/schema/v1"
4   xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
5   xsi:schemaLocation="http://www.taxonx.org/schema/v1 http://www.taxonx.org/schema/v1/taxonx1.xsd
6   http://www.loc.gov/mods/v3 http://www.loc.gov/mods/v3/mods-3-1.xsd
7   http://digir.net/schema/conceptual/darwin/2003/1.0
8   http://digir.net/schema/conceptual/darwin/2003/1.0/darwin2.xsd">
9 <tax:taxonxHeader>
10 <mods:mods>
11 <mods:titleInfo>
12 <mods:title>Description of two new Temnothorax species (Hymenoptera: Formicidae)
13   from Italy.</mods:title>
14 </mods:titleInfo>
15 <mods:name type="personal">
16 <mods:role>
17 <mods:roleTerm>Author</mods:roleTerm>
18 </mods:role>
19 <mods:namePart>Schulz, A.</mods:namePart>
20 </mods:name>
21 <mods:name type="personal">
22 <mods:role>
23 <mods:roleTerm>Author</mods:roleTerm>
24 </mods:role>
25 <mods:namePart>Heinze, J.</mods:namePart>
26 </mods:name>
27 <mods:name type="personal">
28 <mods:role>
29 <mods:roleTerm>Author</mods:roleTerm>
30 </mods:role>
31 <mods:namePart>Pusch, K.</mods:namePart>
32 </mods:name>

```

All documents with Guides: minimally Names, mods; max. bib.refs, specimen, localities

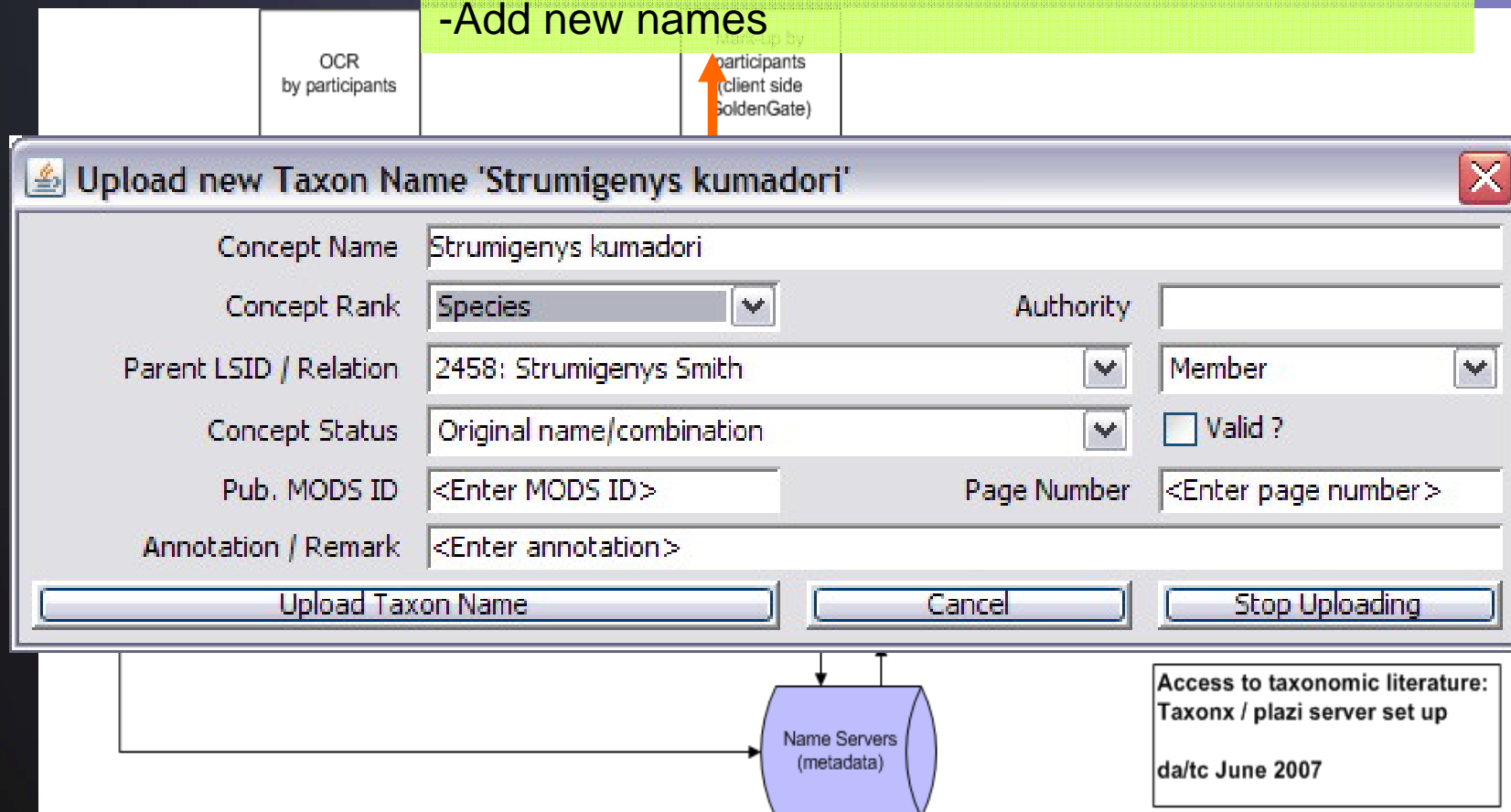
Plazi workflow

OCR-ed texts (dirty, clean)
 ABBYY training files for fonts
 ABBYY training files for journals
 ABBYY custom dictionary

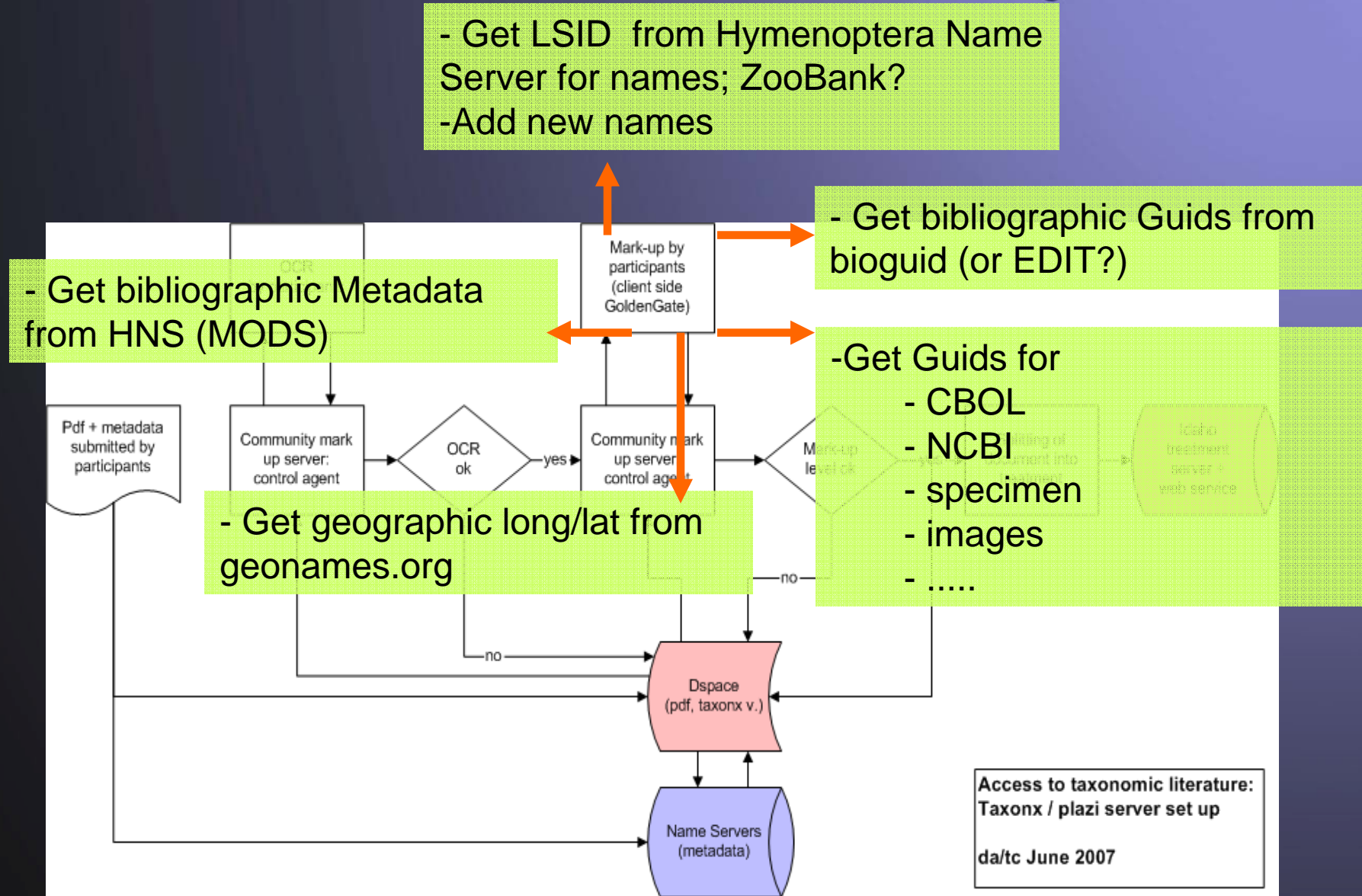


Plazi workflow

- Get LSID from Hymenoptera Name Server for names (in future from ZooBank as well, or in fact any other name server)
- Add new names



Plazi workflow



- Get LSID from Hymenoptera Name Server for names; ZooBank?
-Add new names

- Get bibliographic Metadata from HNS (MODS)

- Get bibliographic Guides from bioguid (or EDIT?)

-Get Guides for
- CBOL
- NCBI
- specimen
- images
-

- Get geographic long/lat from geonames.org

Access to taxonomic literature:
Taxonx / plazi server set up
da/tc June 2007

Technical barriers

Legacy publications are structured for human, but not for machine consumption.

TaxonX is a XML schemas marking up the logic content of a publication.

GoldenGate is an editor for semiautomatic mark up

SRS is a search and retrieval server to deposit and provide access to descriptions extracted from marked up legacy documents

Plazi.org | Taking care of Freedom. - Mozilla Firefox


File Edit View History Bookmarks Tools Help

http://plazi.org/ Google

Kostenlose Hotmail Links anpassen Windows Media Windows

PLAZI

taking care of freedom

publications story about 

- GoldenGATE Editor
- DSpace
- TaxonX Schema
- SRS
- Documentation
- OCR
- testbed: Madagascar Ants
- training
- Blogs
- Content**
- Sources

Plazi.org

Plazi.org supports the development of persistent and openly accessible digital taxonomic literature. To this end Plazi will:

- * Maintain a digital taxonomic literature repository to enable archiving of taxonomic treatments.
- * Enhance submitted taxonomic treatments by creating taxonx XML versions.
- * Participate in the development of new models for publishing taxonomic treatments in order to maximize interoperability with other relevant cyberinfrastructure components (e.g., name servers, biodiversity resources, etc...)
- * Advocate and educate about the vital importance of maintaining free and open access to scientific discourse and data.

Search

- Graduate studies in Comparative Biology
- IISE Planet Bob: bringing Taxonomy the public
- Publication of specimen names on the Web
- Iberian Ants
- Incoming Book: Ants of North America

iPhylo

Done

Use these fields to search the taxonomic name index.

Name	<input type="text"/>	Taxa Only <input checked="" type="checkbox"/>	Exact Match <input checked="" type="checkbox"/>	LSID	<input type="text"/>
Genus	<input type="text"/>	SubGenus	<input type="text"/>	Species	<input type="text"/>
SubSpecies	<input type="text"/>	Variety	<input type="text"/>		
Probolomyrmex					

Use these fields to search the location index.

Country	<input type="text"/>	Location Name	<input type="text"/>	Long/Lat Circle	<input type="text"/>	Elevation	<input type="text"/>	Elevation Circle	<input type="text"/>
Longitude	<input type="text"/>	Latitude	<input type="text"/>	1 degree				100 meters	

Use these fields to search the MODS document meta data index.

Author	<input type="text"/>	Year	<input type="text"/>	Title	<input type="text"/>
Journal / Publisher	<input type="text"/>	Page	<input type="text"/>	MODS ID	<input type="text"/>

Result Type Sub Type <Not Available>

[Back to Search Form](#)



GoldenGATE SRS Search Result [GoogleMaps](#)

Scientific Name	Status	Publication	Pages	ModsID	GoogleMaps
Probolomyrmex		Agosti, D., 1995, A revision of the South American species of the ant genus <i>Probolomyrmex</i> (Hymenoptera: Formicidae)., Journal of the New York Entomological Society (102), pp. 429-434: 432, http://research.amnh.org/entomology/social_insects/ants/publications/8080/8080.pdf	432	8080	
Probolomyrmex		Taylor, R. W., 1965, A monographic revision of the rare tropicopolitan ant genus <i>Probolomyrmex</i> Mayr (Hymenoptera: Formicidae)., Transactions of the Royal Entomological Society of London (117), pp. 345-365: 363-364, http://research.amnh.org/entomology/social_insects/ants/publications/2805/2805.pdf	363-364	2805	
Probolomyrmex tani	new species	Fisher, B. L., 2007, A new species of <i>Probolomyrmex</i> from Madagascar., Memoirs of the American Entomological Institute (80), pp. 146-152: 148-150, http://hdl.handle.net/10199/15374	148-150	21281	GoogleMaps
Probolomyrmex boliviensis		Agosti, D., 1995, A revision of the South American species of the ant genus <i>Probolomyrmex</i> (Hymenoptera: Formicidae)., Journal of the New York Entomological Society (102), pp. 429-434: 432-433, http://research.amnh.org/entomology/social_insects/ants/publications/8080/8080.pdf	432-433	8080	
Probolomyrmex brujitae	new species	Agosti, D., 1995, A revision of the South American species of the ant genus <i>Probolomyrmex</i> (Hymenoptera: Formicidae)., Journal of the New York Entomological Society (102), pp. 429-434: 433, http://research.amnh.org/entomology/social_insects/ants/publications/8080/8080.pdf	433	8080	
Probolomyrmex petiolatus		Agosti, D., 1995, A revision of the South American species of the ant genus <i>Probolomyrmex</i> (Hymenoptera: Formicidae)., Journal of the New York Entomological Society (102), pp. 429-434: 433, http://research.amnh.org/entomology/social_insects/ants/publications/8080/8080.pdf	433	8080	
Probolomyrmex		Taylor, R. W., 1965, A monographic revision of the rare tropicopolitan ant genus <i>Probolomyrmex</i> Mayr (Hymenoptera: Formicidae)., Transactions of the Royal Entomological Society of London (117), pp. 345-365: 346-351, http://research.amnh.org/entomology/social_insects/ants/publications/2805/2805.pdf	346-351	2805	
Probolomyrmex filiformis		Taylor, R. W., 1965, A monographic revision of the rare tropicopolitan ant genus <i>Probolomyrmex</i> Mayr (Hymenoptera: Formicidae)., Transactions of the Royal Entomological Society of London (117), pp. 345-365: 353, http://research.amnh.org/entomology/social_insects/ants/publications/2805/2805.pdf	353	2805	

Probolomyrmex tani , Fisher, B. L.

Publication Data, Additional Information (status, external links, etc)

citation of original description	Fisher, B. L., 2007, A new species of Probolomyrmex from Madagascar., Memoirs of the American Entomological Institute (80), pp. 146-152: 148
publication ID	21281
link to original citation	http://hdl.handle.net/10199/15374
additional text versions	Plain XML TaxonX
scientific name	Probolomyrmex tani
status	new species
description page, figures	
external databases	HNS Lookup Probolomyrmex tani
distribution map	GoogleMaps

Treatment

Probolomyrmex tani ^{HNS} , new species

Figures 1 - 7

TYPE MATERIAL

Holotype worker, Madagascar: Prov. Antsiranana, Foret d'Analabe, 30.0 km 72 ENE Daraina, 13 05 '00 " S, 049 54 ' 30 " E, 30 m, 27 Nov 2003, littoral rainforest, (coll. B. L. Fisher et al.) Collection code: BLF 9426, specimen code: CASENT 0041505, (CASC) Paratypes: 1 worker and 1 dealate queen with same data as holotype but with specimen codes CASENT 0041506 (1 dQ) (CASC) and CASENT 0041507 (1 w) (MCZC).

ETYMOLOGY

The specific name is an arbitrary combination, to be treated as a noun in apposition.

DESCRIPTION

Worker Measurements (mm): maximum and minimum based on all specimens, n = 22, (holotype), [paratype].

TL 2.3 - 2.76 (2.5) [2.5], HL 0.55 - 0.64 (0.60) [0.59], HW 0.37 - 0.43 (0.40) [0.40], LHT 0.34 - 0.46 (0.39) [0.39], CI 65 - 75 (67) [67], SL 0.36 - 0.44 (0.37) [0.37], SI 85 - 104 (91) [94], WL 0.69 - 0.87 (0.77) [0.75], PW (w) 0.28 - 0.36 (0.31) [0.31], DPW 0.17 - 0.22 (0.19) [0.19], PNI (w) 56 - 69 (62) [61], PH 0.25 - 0.32 (0.26) [0.27], PNL 0.24 - 0.32 (0.27) [0.28], LPI 86 - 104 (96) [104], PNH 0.20 - 0.23 [0.22], DPNL 0.20 - 0.25 [0.25], LPNI 96 - 117 [117].

Characters of Probolomyrmex ^{HNS} worker as described by Brown (1975: 7). In full face view, posterior margin of head

[Back to Search Form](#)

Mayriella granulata Dlussky & Radchenko

Publication Data, Additional Information (status, external links, etc)

citation of original description	Dlussky, G. M. & Radchenko, A. G., 1990, [The ants (Hymenoptera, Formicidae) of Vietnam. Subfamily Pseudomyrmecinae. Subfamily Myrmicinae (tribes Calyptomyrmecini, Meranoplini, Cataulacini).], [News of faunistics and systematics.], pp. 119-125: 123
publication ID	21374
link to original citation	http://hdl.handle.net/10199/
additional text versions	Plain XML TaxonX
scientific name	Mayriella granulata
status	sp.n.
description page, figures	
external databases	HNS Lookup Mayriella granulata
distribution map	

Treatment

Mayriella granulata ^{HNS} D l u s s k y e t R a d t s c h e n k o, sp.n.

М а т е р и а л. Рабочие, голотип и 2 паратипа: Вьетнам, арх Баятлонг, о Донгко, найдены в подстилке в т о р ч н о г о вечнозеленого ксерофитизованного тропического леса, № А - 5 7 - 8 7, 23.03. 19 8 7 (А Радченко); хранятся в И н с т и т у т е зоологии АН У С С Р (Киев).

Рабочие (рис. 2,1-3), голова трапециевидная, по длине равна ширине или немно- го больше (ИГ = 1,0-1,04), задние углы широко закруглены, затылочный край слегка вогнут, наличник с двумя маленькими зубчиками.

Грудь короткая, пронотум в профиль выпуклый, его передние углы (см. свер- ху) заострены. Ш и п ы пропододеума короткие, острые, их длина примерно равна ш и р и н е у основания (у *M. spinosior* ^{HNS} W h l. и *M.transfuga* ^{HNS} В а г.-Урб. шипы пропододеума длин- ные, в 1,5-2 раза длиннее ширины у основания). Петиоль с короткой цилиндрической частью, его передняя поверхность почти не вогнута, узелок петиоля близок к клино- видному, с узко закругленной вершиной (у *M. abstinens* ^{HNS} F o r. узелок петиоля с гори- зонтальной или слабо скошенной вершиной площадкой). Постпетиоль низкий, сверху уплощен.

Скульптура головы и груди густая, бугорчатая; наличник гладкий и блестящий; бока узелка петиоля с

Publication Data, Additional Information (status, external links, etc)

citation of original description	Wang, W., Zhou S.-Y. & Huang J.-H., 2005, A new species of the genus Vollenhovia Mayr and a new record species of the genus Myrmica Latreille from China (Hymenoptera: Formicidae). [In Chinese with English summary], Acta Zootaxonomica Sinica (30(4)), pp. 835-836: 835
publication ID	21054
link to original citation	http://hdl.handle.net/10199/15445
additional text versions	Plain XML TaxonX
scientific name	Vollenhovia lucimandibula
status	sp. nov.
description page, figures	
external databases	HNS Lookup Vollenhovia lucimandibula
distribution map	

Treatment

壳?扁胸切叶?, 新? *Vollenhovia lucimandibula* ^{HNS} sp. nov. (? 1 ~ 2)

正模工?: TL 3 1 2, HL 0 1 61, HW 0 1 54, CI 89,SL 0 1 41,SI 76, PW 0 1 46,AL 0 1 83, ED 0 1 13, PL 0 1 22, PH 0 1 24, DPW 0 1 20。

?近矩形?大于? ??近平直,后??中央凹陷。?眼中等大小,位于???中部稍前?。?角 12?,粗?。?角柄?末端伸??角?至后???距离的 3/4?。?鞭?端部 3?形成?角棒,?角棒的?度大于第 2 ~ 9?之和?,端??度大于其前 2?之和,第 3 ~ 9??大于?。?唇基具?隆?,向前分 歧,其中部明?凹陷。?唇基前?中央凹陷。上??三角形,咀嚼?具 7?,端部 4??大,端?尖?,?端部?第 3??相??的?距明??大,基部 3??小,排列?密。背面?,前胸背板??,背板扁平,由前向后逐??窄。?面?,前胸?中胸?接?密,前中胸背板?不明?,后胸??断,且明?凹陷。?胸腹?基面?斜面近相等,基面向后?弧形?渡到斜面,斜面??具突起的脊?。?腹柄?前面近垂直;背面短于前面,向后?斜,后面?坡形,??背面等?。?腹柄下突??,近半?形。后腹柄?背面?形,?大于?。后腹部?卵形。

Probolomyrmex tani, Fisher, B. L.

Publication Data, Additional Information (status, external links, etc)	
citation of original description	Fisher, B. L., 2007, A new species of Probolomyrmex from Madagascar, Memoirs of the American Entomological Institute (80), pp. 146-152: 148
publication ID	21281
link to original citation	http://hdl.handle.net/10199/15374
additional text versions	Plain XML Taxon
scientific name	Probolomyrmex tani
status	new species
description page, figures	
external databases	HNS Lookup Probolomyrmex tani
distribution map	GoogleMaps

Treatment

Probolomyrmex tani^{HNS}, new species

Figures 1 - 7

TYPE MATERIAL

Holotype worker, Madagascar: Prov. Antsiranana, Foret d'Analabe, 30.0 km 72 ENE Daraina, 12° 30' E, 30 m, 27 Nov 2003, littoral rainforest, (coll. B. L. Fisher et al.) Collection code: BLF 9426, CASENT 0041505, (CASC) Paratypes: 1 worker and 1 dealate queen with same data as holotype codes CASENT 0041506 (1 dQ) (CASC) and CASENT 0041507 (1 w) (MCZO).

ETYMOLOGY

The specific name is an arbitrary combination, to be treated as a noun in apposition.

DESCRIPTION

Worker Measurements (mm): maximum and minimum based on all specimens, n = 22, (holotype)

TL 2.3 - 2.76 (2.5) [2.5], HL 0.55 - 0.64 (0.60) [0.59], HW 0.37 - 0.43 (0.40) [0.40], LHT 0.34 - 0.46 (0.37) [0.37], SL 0.36 - 0.44 (0.37) [0.37], SI 85 - 104 (91) [94], WL 0.69 - 0.87 (0.77) [0.75], PW (w) [0.31], DPW 0.17 - 0.22 (0.19) [0.19], PNI (w) 56 - 69 (62) [61], PH 0.25 - 0.32 (0.28) [0.27], PNL (w) [0.26], LPI 66 - 104 (96) [104], PNH 0.20 - 0.23 [0.22], DPNL 0.20 - 0.25 [0.25], LPNI 96 - 117 [111]

Characters of *Probolomyrmex*^{HNS} worker as described by Brown (1975: 7). In full face view, poste

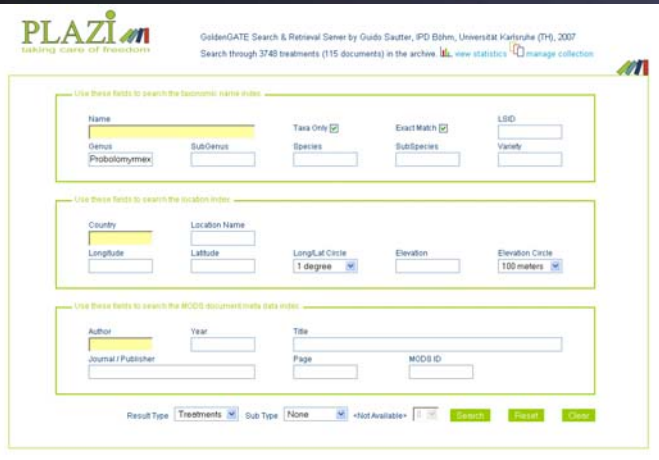
```

<tax:taxonomy xsi:schemaLocation="http://www.taxon.org/schema/v1 http://www.taxon.org
http://dgr.net/schema/conceptual/darwin/2003/1.0 http://dgr.net/schema/conceptual/darwin/20
- <tax:taxonomyHeader>
- <tax:taxonomyHeader>
- <mods:mods>
- <mods:titleInfo>
  <mods:title>A new species of Probolomyrmex from Madagascar.</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:typeOfResource>text</mods:typeOfResource>
- <mods:relatedItem type="host">
- <mods:titleInfo>
  <mods:title>Memoirs of the American Entomological Institute</mods:title>
</mods:titleInfo>
- <mods:part>
  <mods:detail type="volume">
    <mods:number>80</mods:number>
  </mods:detail>
- <mods:extent unit="page">
  <mods:start>146</mods:start>
  <mods:end>152</mods:end>
</mods:extent>
  <mods:date>2007</mods:date>
</mods:part>
</mods:relatedItem>
<mods:identifier type="HNS-PUB">21281</mods:identifier>
- <mods:location>
  <mods:url>http://hdl.handle.net/10199/15374</mods:url>
</mods:location>
  
```



The underlying marked-up descriptions and XSLT engine allows export in any particular format. The degree of mark up is the limiting factor

Search and Retrieval Server: Access to data



TAPIR

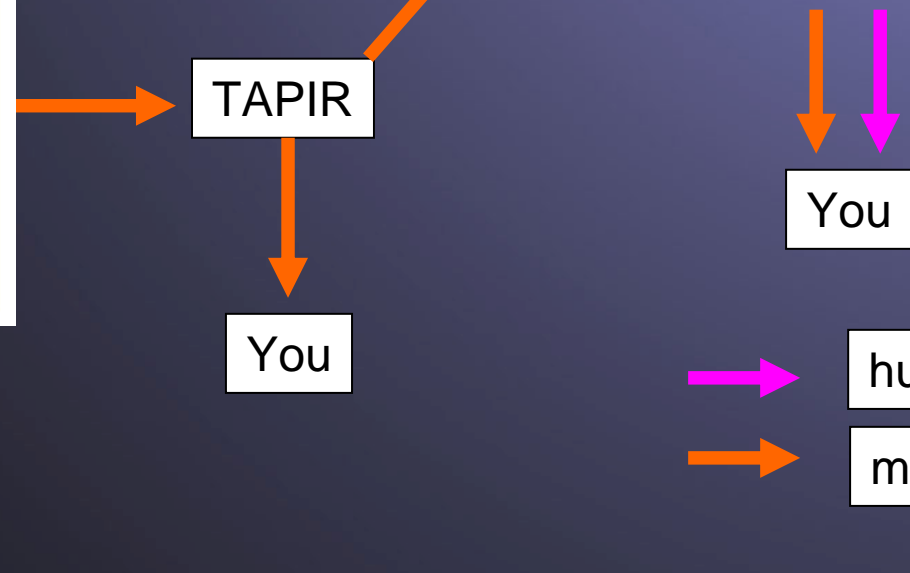
You

You

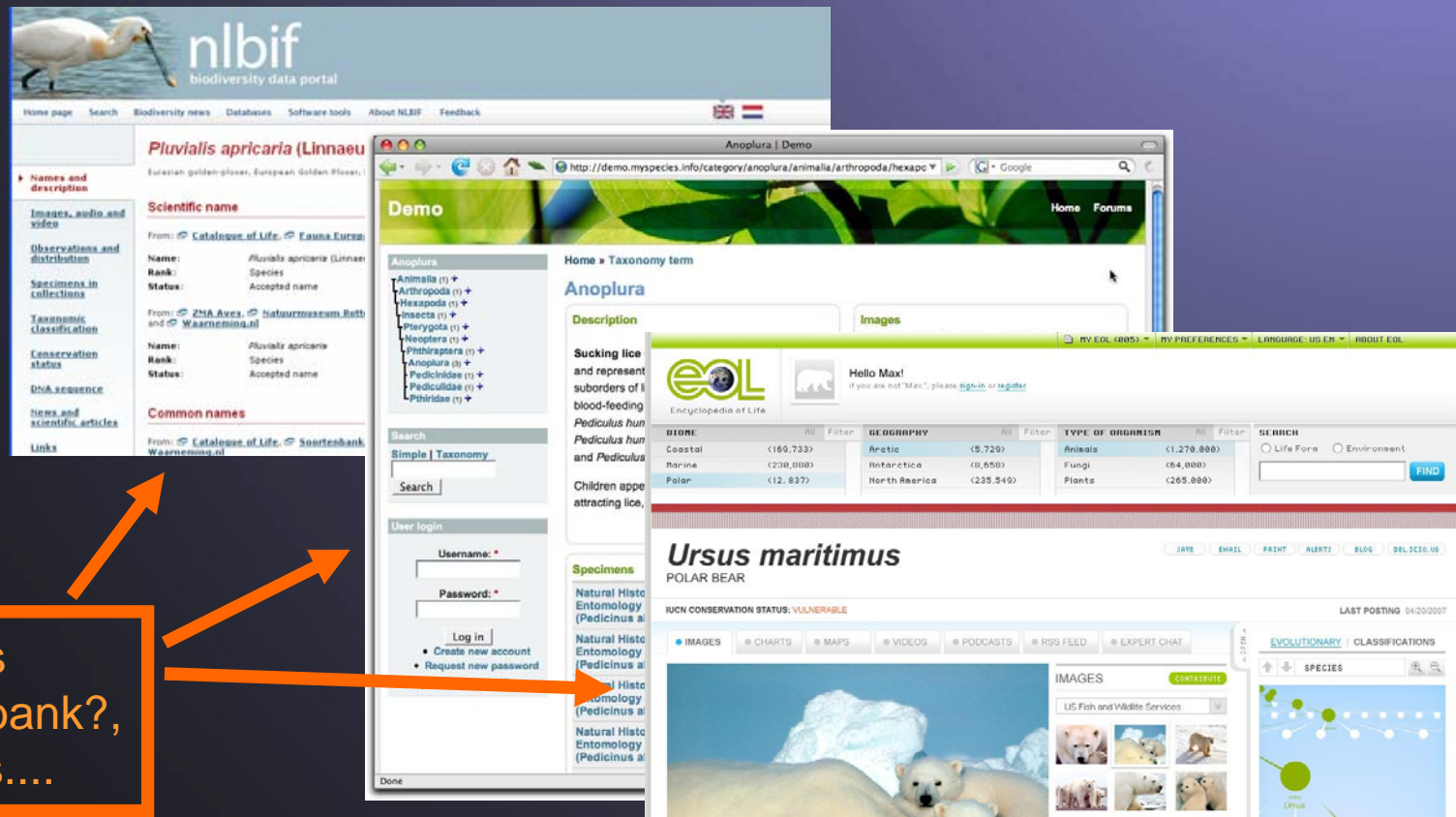
You

human

machine



The estimated well over 10 M treatments of species are the obvious feed for mashups or a yet untapped source for data mining, extraction



nlbif
biodiversity data portal

Home page Search Biodiversity news Databases Software tools About NLBIF Feedback

Pluvialis apricaria (Linnaeus)
Eurasian golden-plover, European Golden Plover, ...

Scientific name
From: Catalogue of Life Fauna Europaea

Name: *Pluvialis apricaria* (Linnaeus)
Rank: Species
Status: Accepted name

From: ZNA Aves Natuurmuseum Rotterdam Waarneming.nl

Name: *Pluvialis apricaria*
Rank: Species
Status: Accepted name

Common names
From: Catalogue of Life Soortbank Waarneming.nl

Names and description
Images, audio and video
Observations and distribution
Specimens in collections
Taxonomic classification
Conservation status
DNA sequence
News and scientific articles
Links

Anoplura | Demo
http://demo.myspecies.info/category/anoplura/animalia/arthropoda/hexapoda

Demo Home Forums

Anoplura Home » Taxonomy term

Description Images

Sucking lice and represent suborders of blood-feeding *Pediculus humanus* and *Pediculus humanus*
Children appear attracting lice.

Specimens
Natural History Entomology (Pediculus a...
Natural History Entomology (Pediculus a...
Natural History Entomology (Pediculus a...
Natural History Entomology (Pediculus a...

User login
Username: *
Password: *
Log in
• Create new account
• Request new password

Encyclopedia of Life
Hello Max!
If you are not "Max", please [sign in](#) or [register](#).

BIOME	All	Filter	GEOGRAPHY	All	Filter	TYPE OF ORGANISM	All	Filter	SEARCH
Coastal	(160,733)		Arctic	(5,720)		Animals	(1,270,800)		<input type="radio"/> Life Form <input type="radio"/> Environment
Marine	(230,000)		Antarctica	(0,650)		Fungi	(64,000)		<input type="text"/> <input type="button" value="FIND"/>
Polar	(12,837)		North America	(235,540)		Plants	(265,000)		

Ursus maritimus
POLAR BEAR

IUCN CONSERVATION STATUS: VULNERABLE

LAST POSTING: 04/20/2007

IMAGES CHARTS MAPS VIDEOS PODCASTS RSS FEED EXPERT CHAT

EVOLUTIONARY CLASSIFICATIONS

US Fish and Wildlife Services

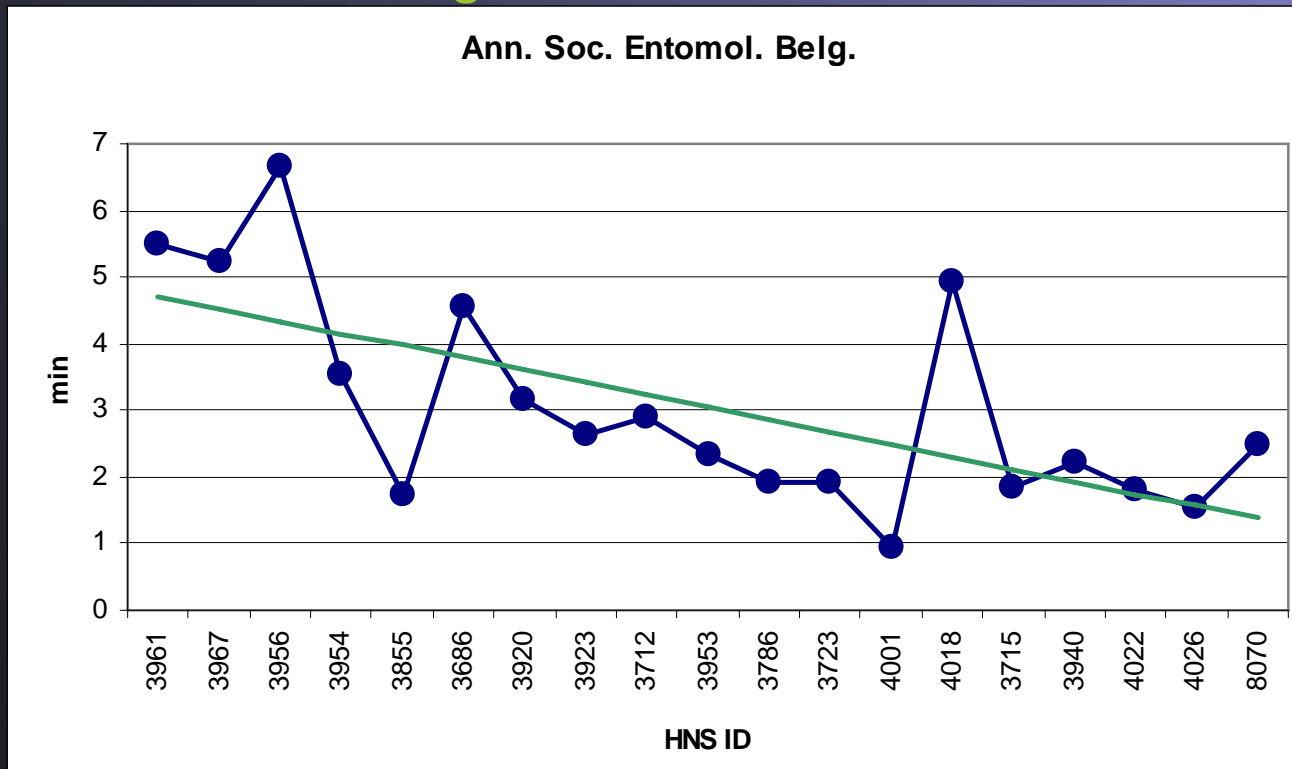
US Fish and Wildlife Services

Treatments
(plazi, zoobank?,
other feeds....

Does the mark-up process scale up to the millions of page needed to be processed?

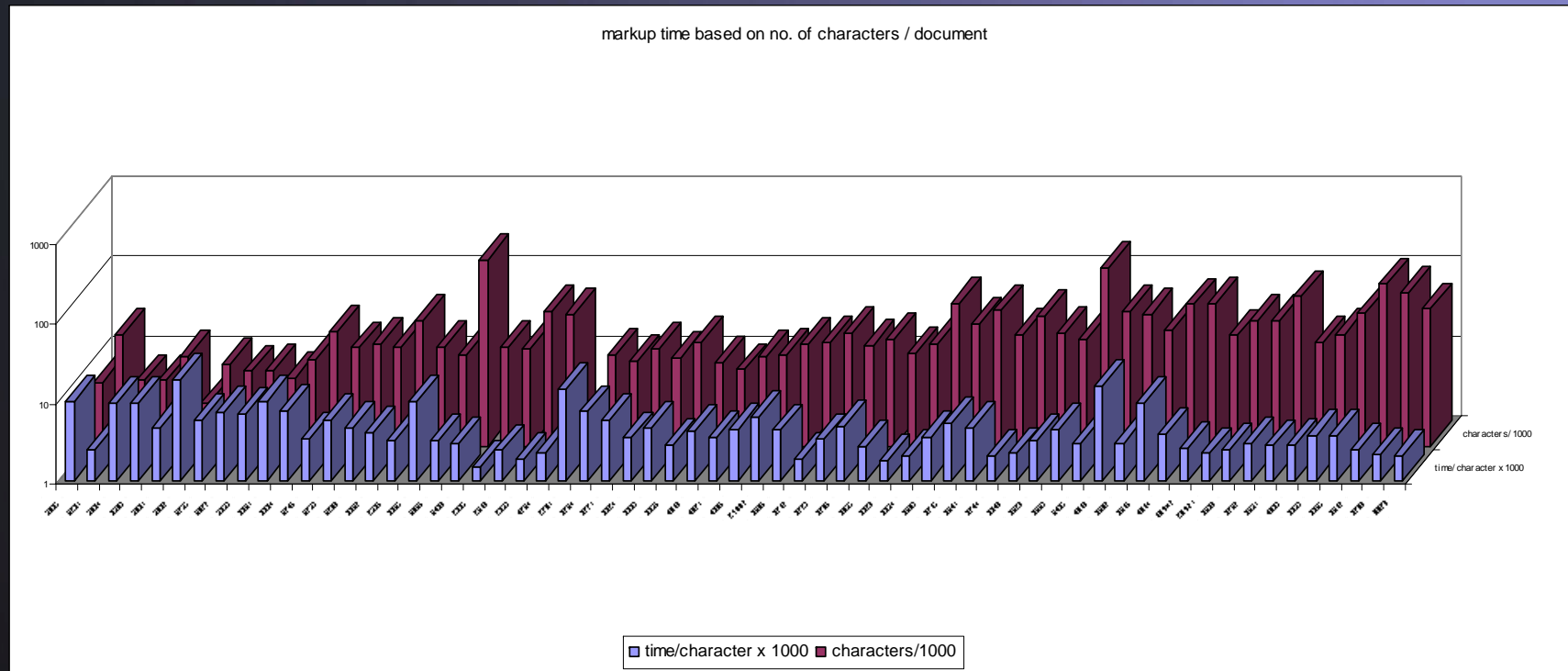
Is this the future?

Metrics to measure effort and compare various approaches and algorithm



Time per minute to produce clean OCR using ABBYY;
publications in chronological order

Metrics of the effort and and to compare various approaches and alogrithm to mark up documents



Does the mark-up process scale up to the millions of page needed to be processed?

Is this the future?

Only part of it: Mark up takes about
5min/page: For 100 M pages = 700 man
years (but it is only a first tool...)

Is this the future?

Only part of it: Mark up takes about 5min/page: For
100 M pages = 700 man years

However, the data is highly structured and ready to
use in contrast to the deliverables of BHL
(names, unclean OCR) that need be checked
after the production at a yet unknown cost.


What is the future?

Mark up upfront and production of XML output that includes taxonomic specific elements.

A NLM/taxonx publishing XML module is in production, and PLOS-One, BMC and Zootaxa are working on implementing taxonomic specific mark-up into their production XML.

Because of technical constraints, taxon-xml might be added as additional material

A recent example: *Chromis* (A fish)

 *Zootaxa* 1671: 3–31 (2008)
www.mapress.com/zootaxa/
Copyright © 2008 Magnolia Press

ISSN 1175-5326 (print edition)
ZOOTAXA
ISSN 1175-5334 (online edition)

Five new species of the damselfish genus *Chromis* (Perciformes: Labroidei: Pomacentridae) from deep coral reefs in the tropical western Pacific

RICHARD L. PYLE*, JOHN L. EARLE† & BRIAN D. GREENE‡

Department of Natural Sciences, Bishop Museum, 1525 Bernice Street, Honolulu, Hawaii 96817-2704, USA.¹

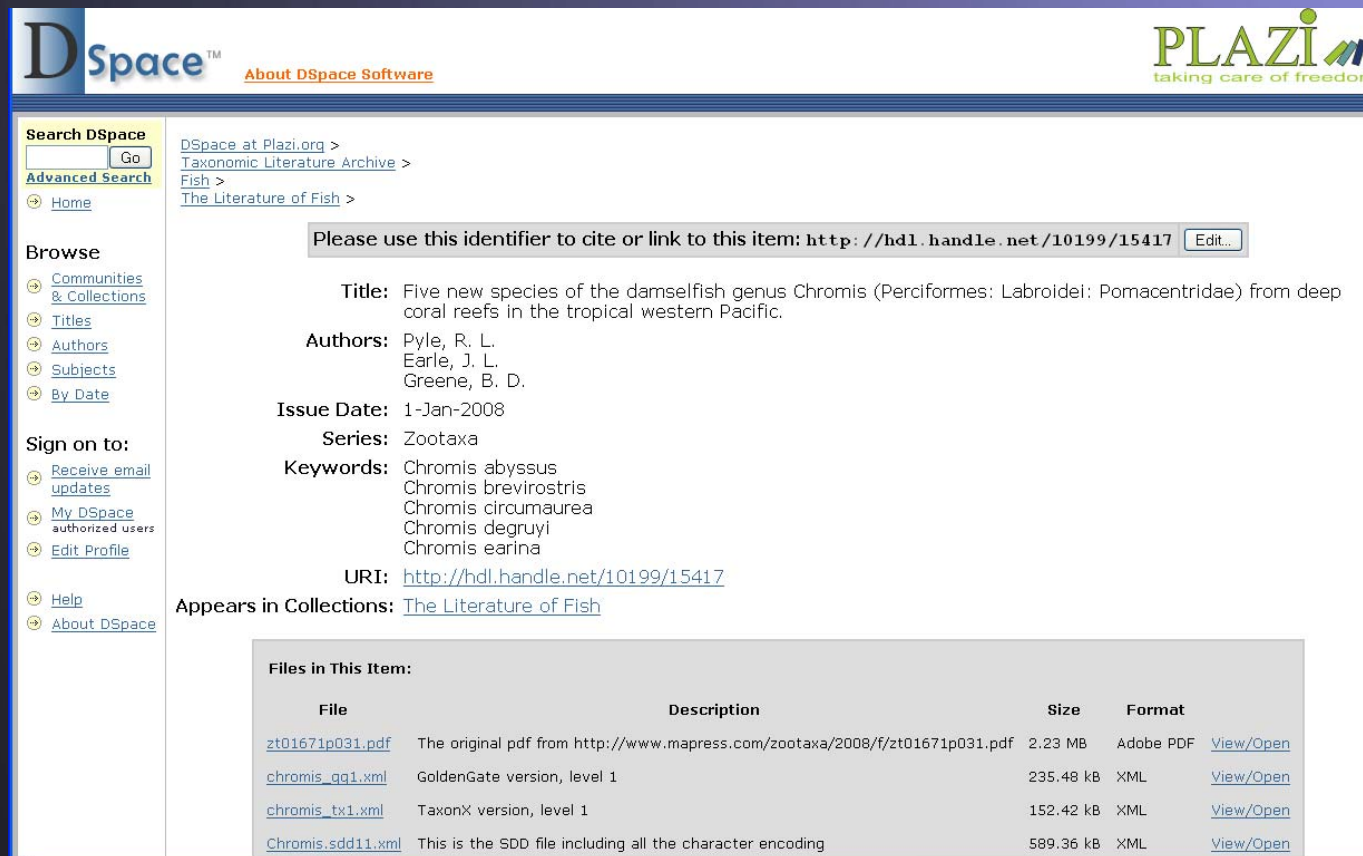
*E-mail: *deepreef@bishopmuseum.org, †earlej001@hawaii.rr.com, ‡bgreene@hawaii.edu.*

Table of contents

Abstract	3
Introduction	3
Material and methods	4
<i>Chromis abyssus</i> , new species	6
<i>Chromis brevirostris</i> , new species	10
<i>Chromis circumaurea</i> , new species	15

PDF with 196 links embedded to ZooBank using LSIDs and GUIDs for names, literature, CBOL, imagery and videos, and external marked up versions....

A recent example: Chromis



The screenshot shows a DSpace repository page for a publication. The page includes a search bar, navigation links, and a list of files associated with the item. The publication details are as follows:

Title: Five new species of the damselfish genus *Chromis* (Perciformes: Labroidae: Pomacentridae) from deep coral reefs in the tropical western Pacific.

Authors: Pyle, R. L., Earle, J. L., Greene, B. D.

Issue Date: 1-Jan-2008

Series: Zootaxa

Keywords: *Chromis abyssus*, *Chromis brevirostris*, *Chromis circumaurea*, *Chromis degruyi*, *Chromis earina*

URI: <http://hdl.handle.net/10199/15417>

Appears in Collections: [The Literature of Fish](#)

Files in This Item:

File	Description	Size	Format	View/Open
zt01671p031.pdf	The original pdf from http://www.mapress.com/zootaxa/2008/f/zt01671p031.pdf	2.23 MB	Adobe PDF	View/Open
chromis_gg1.xml	GoldenGate version, level 1	235.48 kB	XML	View/Open
chromis_tx1.xml	TaxonX version, level 1	152.42 kB	XML	View/Open
Chromis_sdd11.xml	This is the SDD file including all the character encoding	589.36 kB	XML	View/Open

.... In this case stored on plazi and including an SDD and Taxonx version, as well as access to treatments through SRS

Is this the future?

The problem of manual mark up persists:

What is needed are journal production systems with underlying XML templates and access to the various name servers (eg Zoobank in the future, bioguid for bibliographic references, collections for specimen LSIDs, CBOL...)

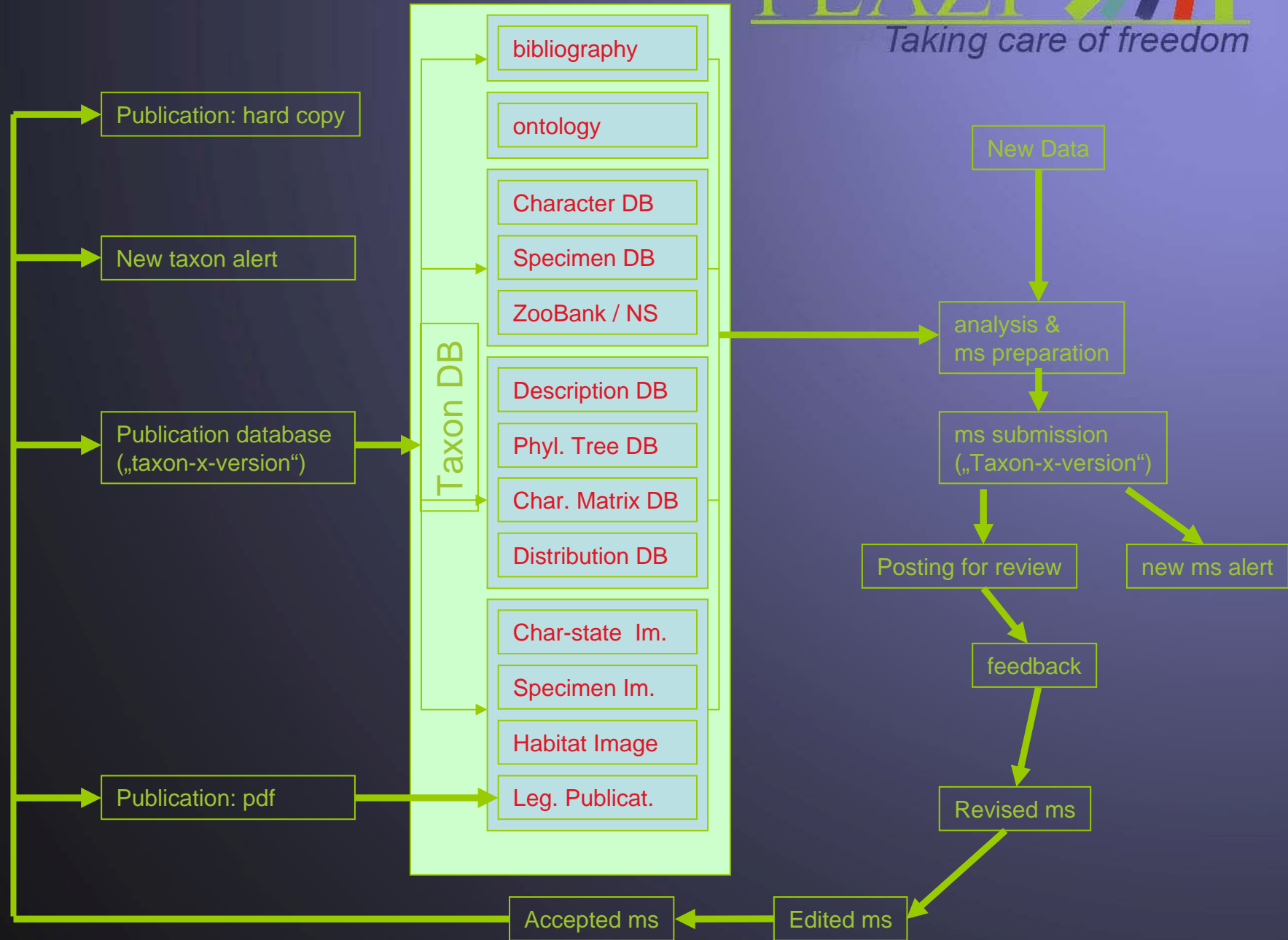
Is this the future?

Only the near term. The future will be a complete change of the function of publications. They will become a version control vehicle of what synthesis has been derived based on the existing data bases, plus a human readable insight into what data is accessible.

The publications then will be the „quality checked“ information that could be displayed in systems like EOL or Ispecies

Publications will become seeds in systems like Scratchpads where people can annotate, add revised content, before a next revision is being submitted to a peer review process

The future of publications: The publication as a version control instrument



Change of paradigm

Main emphasis on building, populating and maintaining databases

This means developments of:

- Standards (eg TDWG)
- Ontologies
- Exchange protocols (eg Digir, Tapir)
- Technical infrastructure
- Collaborative Networks
- Access
- Adequate credit system (i.e citation record)

Such an Open Access infrastructure is what (at least with public money funded) science is about: disseminating scientific result as widely as possible, the logic return to all taxpayers

It is a simple answer to Access and Benefit Sharing, the taxonomic Impediment in CBD..

Open Access is the logic dissemination strategy. It not only widens the user base of data, but allows a much more refined credit system then the coarse journal based citation indices.

Social Changes

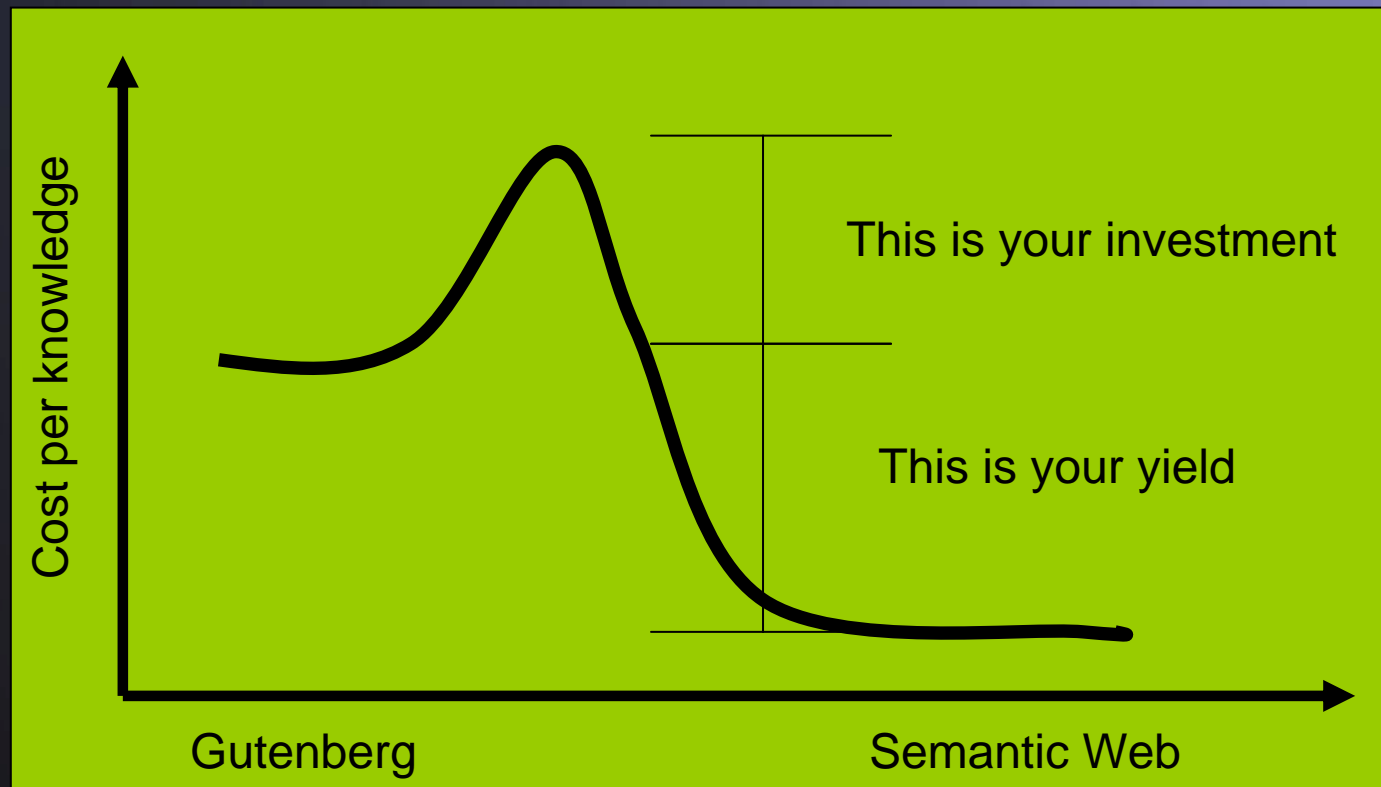
Adoption of Codes allowing e-publications

Social Changes

Adoption of Codes allowing e-publications

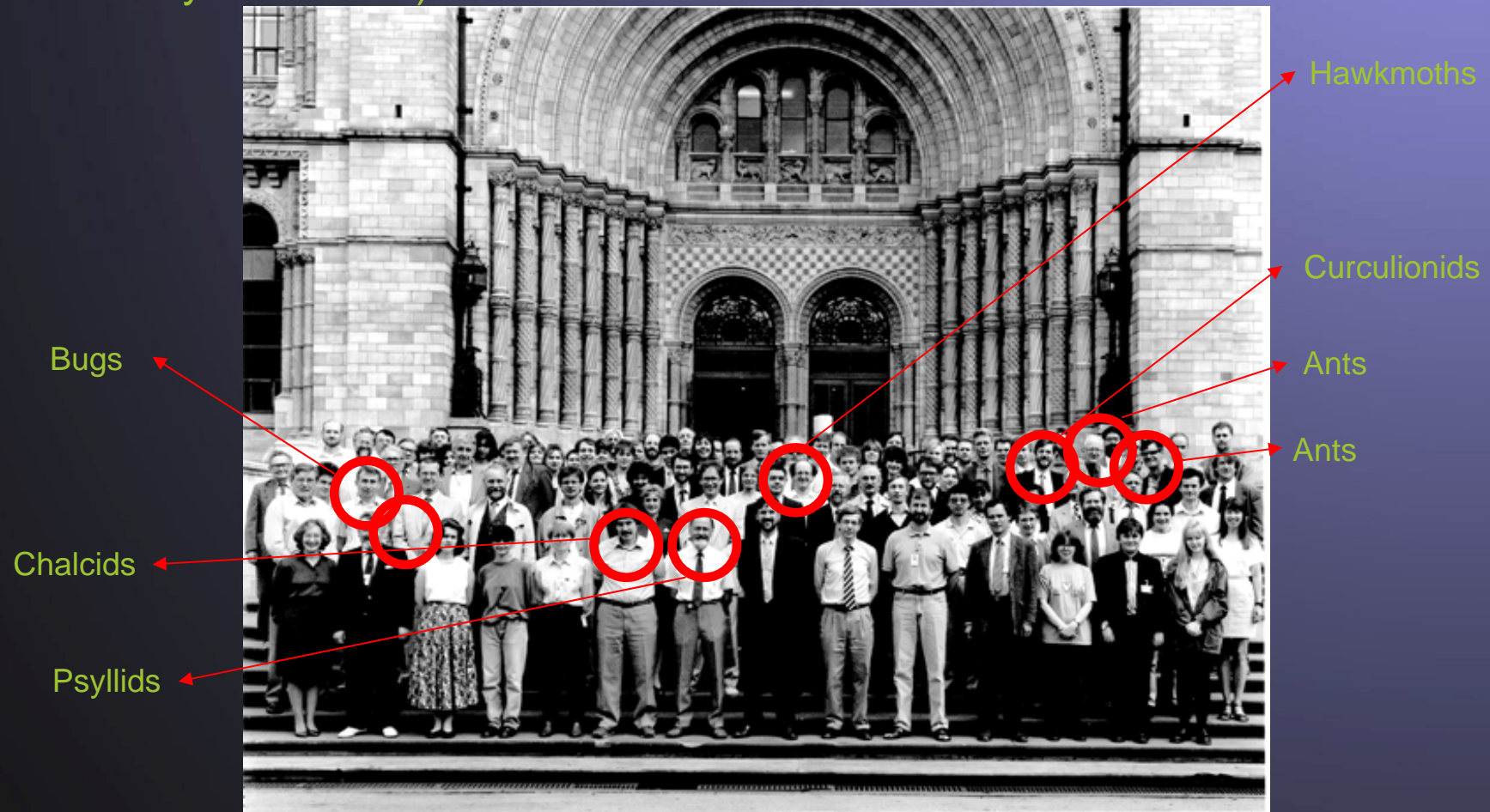
Social Networks

Data conversion seen as an activation energy from a current state of knowledge to a one in a semantic Web environment.



From individual approach to

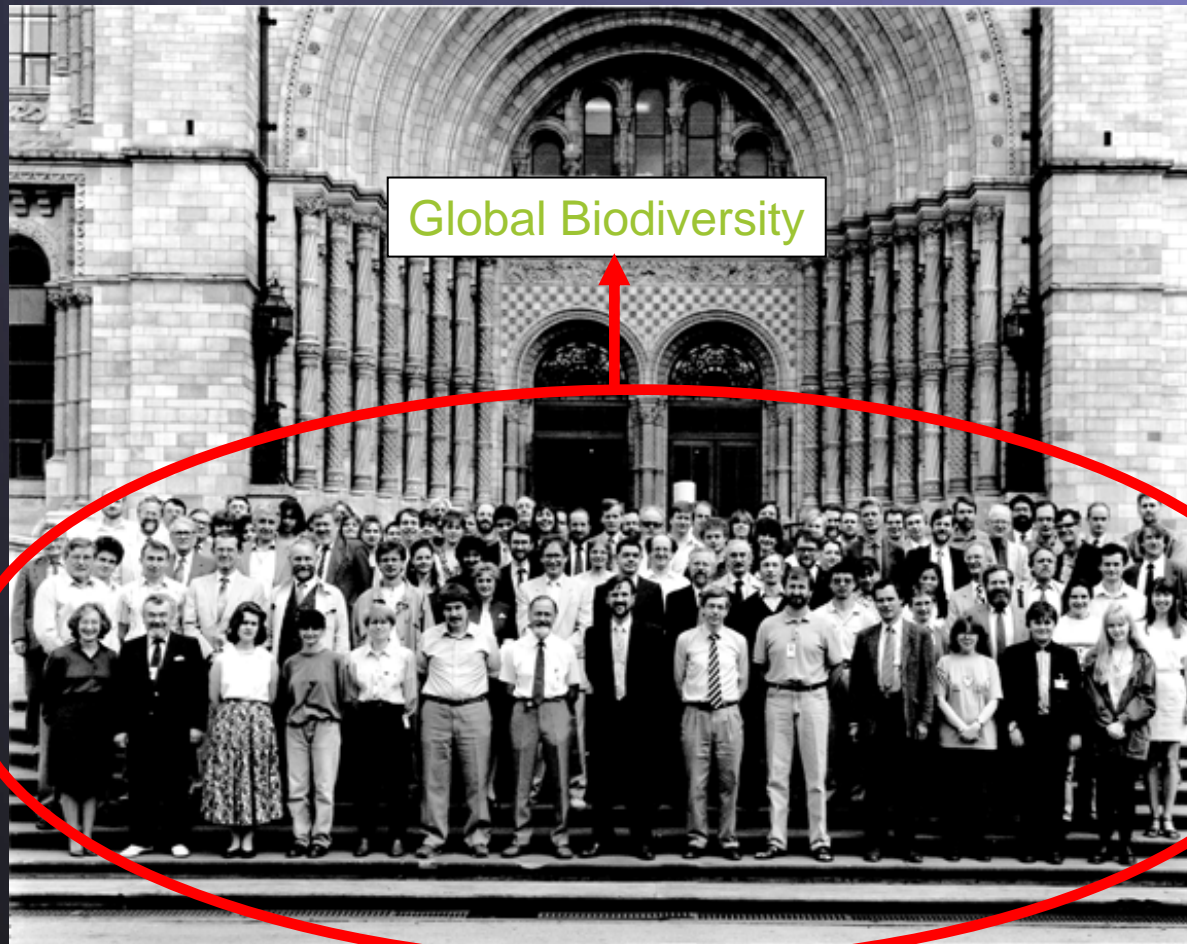
> 6,000 taxonomists world wide, major institutions (Herbaria, Natural History Museums)



The staff of Entomology at The Natural History Museum, London, 1993

...operate as big science.

> 6,000 taxonomists world wide, major institutions (Herbaria, Natural History Museums)



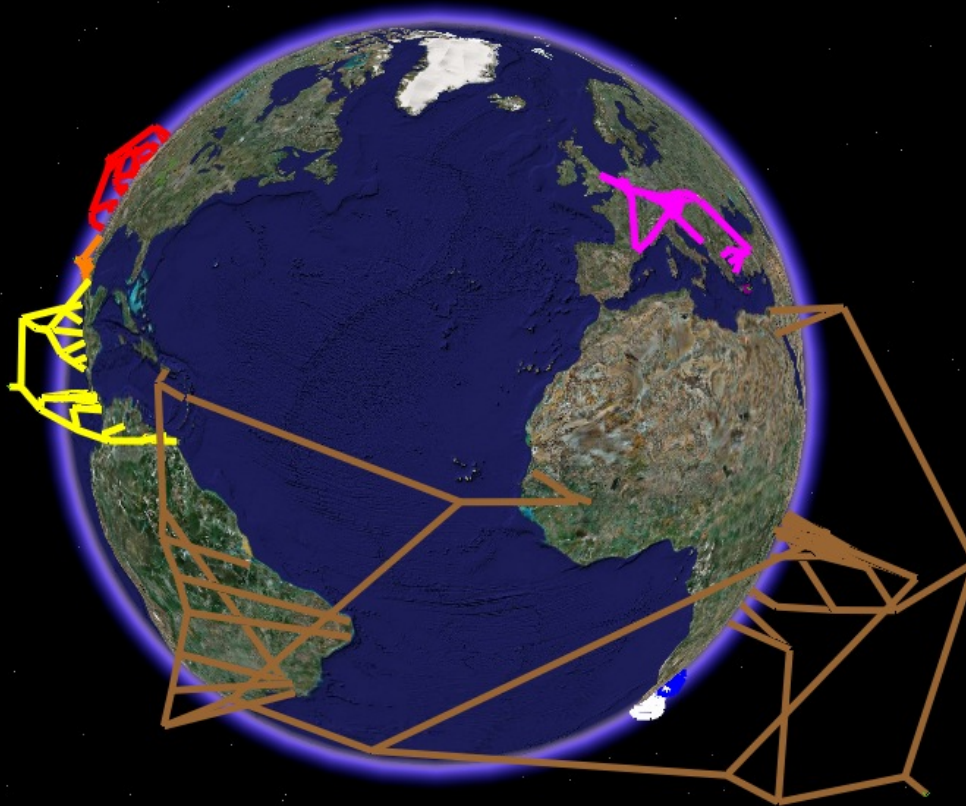
The staff of Entomology at The Natural History Museum, London, 1993

Social Changes

Adoption of Codes allowing e-publications

Social Networks

Demonstrate the power of the new access to avoid
the GBIF-syndrome



Phylogenetic trees and
Google Earth



Rod Page



Image © 2006 NASA
Image © 2006 TerraMetrics

©2006 Google™

Pointer 24° 21'49.41"N 28° 36'41.27"W

Streaming ||||| 100%

Eye alt 15043.38 km

Social Changes

Adoption of Codes of Nomenclature allowing e-publications

Social Networks

Make use of these new chances: Science

Demonstrate the power of the new access to avoid the GBIF-syndrome

Work towards a sustainable model for the new cyberinfrastructure (at local to political (up to EU- and eg \hat{O} ECD) level)

Conclusions and recommendations (legal side):

Individual level

Assure that all you do is open access

- Understand copyright – be not afraid of copyright
- Self archive (the Green Road)
- Don't sign any contracts giving away rights
- Urge your institution to adopt and build a repository for your research
- Talk to your scientific societies and museum to adopt a policy to at least allow self archiving
- Demonstrate the power of access through inovative research projects and data: Research will be the only motivation to change law and build up infrastructure

Conclusions and recommendations (legal side):

Community (eg EDIT) level

Assure that all you do is open access

- Understand, adopt and propagate an adequate copyright policy
- Urge your institution to adopt and build a repository for your research
- Talk to your scientific societies and museum to adopt a policy to at least allow self archiving
- Support the emergence of standards and transfer protocols
- Lobby at the respective national or EU level

Conclusions and recommendations (legal side):

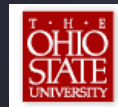
Our (plazi) own

Assure, that from now on all ant descriptions are archived and online accessible through plazi, and find ways to allow others to partizipate

Participating organization and support



Universität Karlsruhe (TH)
Forschungsuniversität • gegründet 1825



antbase.org

Main support by US-NSF, German DFG, GBIF

Thank You!

agosti@amnh.org

Plazi

The purpose of the Association is to support the development of persistent and openly accessible taxonomic literature.

To this end Plazi will maintain a digital taxonomic repository and participate in the development of new models for publishing taxonomic treatments in order to maximize interoperability with other relevant cyberinfrastructure components.