

4D4Life



Distributed Dynamic Diversity Databases for Life

Deliverable 3.1
(Status - Public report)

Report on agreed target set of improvements to the data supply by the Global Species Databases (GSDs)

Workpackage 3

David Ouvrard (MNHN)
Thierry Bourgoin (MNHN)

31 March 2010



Capacities Programme of Framework 7: EC e-Infrastructure Programme - Scientific Data Infrastructures INFRA-2008-1.2.2

Grant Agreement No:	238988
Project Co-ordinator:	Professor Frank Bisby
Project Homepage:	http://www.4D4Life.eu
Duration of Project:	36 months
Start Date:	May 2010
End Date:	May 2012

Table of contents

GSD Survey	4
Introduction	4
Objectives	9
<i>Compile data about GSDs for the improved CoL metadatabase</i>	<i>9</i>
<i>Measure the compliance of GSDs with the Sp2k SDS</i>	<i>9</i>
<i>Means of improving the compliance with the Sp2k SDS</i>	<i>9</i>
<i>New data-fields to be implemented</i>	<i>9</i>
Tools	9
<i>Questionnaire</i>	<i>9</i>
<i>CoL Annual Checklist 2009 (AC09)</i>	<i>9</i>
<i>Statistics from the CoL Annual Checklist 2010 beta (AC10β)</i>	<i>9</i>
Replies	11
Species 2000 Standard Dataset	13
Results	13
<i>Compiling Data for the New Metadatabase</i>	<i>13</i>
Peer review	13
Wrapper	15
Formal Membership to Species 2000	16
Hosting	18
Accessibility/Restrictions	19
Coverage/Completeness	19
Searchability/Download service	21
<i>Measure of compliance</i>	<i>21</i>
Accepted Scientific Name linked to References (obligatory)	21
Synonym(s) linked to Reference(s) (obligatory, as appropriate)	22
Common Name(s) linked to Reference(s) (optional)	23
Latest taxonomic scrutiny (obligatory)	24
Source Database (obligatory)	24
Additional Data (optional)	25
Family name (obligatory)	25

Classification above family, and highest taxon (obligatory, as appropriate)	25
Distribution (optional)	26
Reference(s)	26
Network Discussion	27
Extensions to the bio-data fields	27
<i>Geographical distribution</i>	27
<i>Take taxon LSIDs directly from the GSDs (if LSIDs present in the database)</i>	28
<i>Fossils</i>	28
<i>Habitat</i>	29
<i>Subgenus</i>	29
<i>Associated organisms</i>	30
<i>Reference to Nomenclators</i>	30
<i>Rejected additional fields</i>	31
Pilot Projects	32
<i>Priorities</i>	32
<i>Decisions</i>	32
Conclusion	33
Appendix	34
Appendix 1: 4D4Life Metadata Questionnaire	34

GSD Survey

1. Introduction

The Species 2000 Catalogue of Life is composed and integrated by the Species 2000 Annual Production Service working with data supplied by a large array of Global Species Databases (GSDs). The resulting Catalogue is constrained to just the fields in the present Species 2000 Standard Data Set, as used since 2005 as the common data standard. It is also constrained by the degree of compliance with this data standard and by the degree of fill for data in the optional fields. Now that more than 70 GSD databases are suppliers, the extent and patchiness of this data supply are real issues to be addressed across the ‘Species 2000 GSD Network’, and affecting the quality and uniformity of the product.

In Workpackage 3 of the 4D4Life project, there has been discussion and a survey across the GSD Network, both to understand the extent of these constraints, and to target possible improvements which are to be implemented via individual pilot-projects. These target improvements may either make small extensions to the data provided across all sectors in the Catalogue by means of new fields, or to the degree of fill and degree of uniformity in the Catalogue by improving the compliance of each database with the existing standard.

This document reports on both the survey and discussions leading up to the agreement of target improvements, and on the target improvements actually agreed. The work has been carried out by David Ouvrard and Thierry Bourgoin at MNHN Paris.

It had been planned that this report would also contain revised versions of the Species 2000 Standard Data Set, and the Species 2000 Best Practice Document. However, in the event detailed revision of these documents has been delayed, and will now be delivered in July 2010, mth. 15. The delay is occasioned by:

- i) Late signing of the Grant Agreement by the Commission (mth. 3), causing late appointment of staff to WP3, and
- ii) Agreement by the 4D4Life Management Committee (March 18 2010) that a small working group should scrutinise and agree the changes to the Standard Data Set, and that the resulting changes should be vetted by the Global Team of Species 2000, to ensure that changes proposed in Europe are acceptable to the wider global programme. That Global Team meets next on May 26/27 in New Brunswick, USA.

"Task 3.1 A survey to assess, peer review and certificate each provider GSD, its wrapper & delivery, and its formal membership of the programme, including data capture for the improved metadatabase."

"Task 3.2 Undertake prior consultation with appropriate partners (e.g. GSD custodians, nomenclator custodians, editorial services staff) and prepare a Network Discussion Document, including proposals for i) Changes to the Data Standard for existing fields (if any), ii) Changes to the Best Practice Document (if any), iii) Possible extensions to the bio-data fields, including data supply requests (if any) from the Services Team, and iv) possible internal service supply issues (e.g. name-alerts, comments from users, taxa from regional hubs, cross-checks with other lists, and unassigned name batches)."

"Task 3.3 Network consultation and workshop with GSD custodians, Nomenclator custodians and Editorial Services Staff (WP 5), to explore each of the items highlighted in the Network Discussion Document; to explain and establish priorities for the pilot project for each database; to respond to requests from the Services Team, and to provide additional training days."

M3.1 GSD Survey complete, brief listing available. Met on Schedule (Month 11)

M3.2 Network Discussion Document, including data supply requests from Services Team. Draft document about alterations, amendments and improvements of the Standard Dataset has been sent to a subgroup of the network and is currently circulating for discussion, as a result of decisions taken at the Second General Meeting in Faro ,Portugal (16-18 February 2010).

44 European Global Species Databases (GSDs) provide data for the Catalogue of Life (CoL) (table 1). These global databases supply more than 900 000 species names (table 2). These databases show a diversity of structures as well as contents, making the assessment of these GSDs necessary in regard with the CoL as a unique aggregating structure.

Other objectives will be presented, as well as the tools requested to manage and complete this global survey with success.

The involvement of GSD custodians has led to the sending of a filled in questionnaire for most of them as well as to the participation of some custodians or representatives to the two first Projects General Meetings held in Reading (15th - 17th September 2009) and in Faro (16th - 18th February 2010).

After having reminded the existing data-fields describing the structure of the CoL, the results of the survey will be given in details, in terms of data capture for the next version of the metadatabase and as a means of measurement of the compliance between the GSDs and the Species 2000 Standard Dataset (Sp2k SDS).

Architecture 1

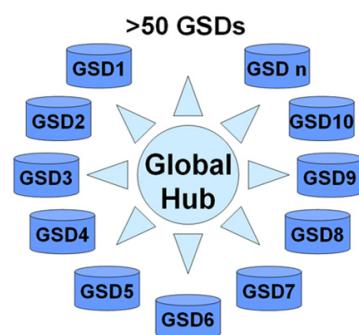


Table 1: European GSDs (GSDs marked with an asterisk are not directly involved in 4D4Life)

	Database name	Database Full Name
1	AlgaeBase	AlgaeBase
2	AnnonBase	World Species List of Annonaceae
3	BDWD	BioSystematic Database of World Diptera
4	BSF	Blattodea Species File Online
5	CCW*	Catalogue of Craneflies of the World*
6	ChiloBase	A World Catalogue of Centipedes (Chilopoda) for the Web
7	CIPA	Computer Aided Identification of Phlebotomine sandflies of Americas
8	CLEMAM	Check List of European MArine Mollusca
9	Conifer Database	Conifer Database
10	COOL	Cercopoidea Organised On Line
11	Droseraceae Database	Droseraceae Database
12	FishBase	FishBase
13	FLOW	Fulgoromorpha Lists On the WEB
14	Global Compositae Checklist	Global Compositae Checklist
15	GloBIS (GART)	Global Butterfly Information System
16	ILDIS	ILDIS World Database of Legumes
17	LepIndex	The Global Lepidoptera Names Index
18	LHD*	Lacistemataceae Holistic Database*
19	LIAS	A Global Information System for Lichenized and Non-Lichenized Ascomycetes
20	MolluscaFW	Checklist of Freshwater Mollusca
21	Nomen.eumycetozoa.com	Nomenclatural Database of Eumycetozoa (Myxomycota)
22	Odonata	Catalogue of the Odonata of the World
23	Porifera	World Porifera Database
24	PSF	Phasmida Species File
25	Psyl'list	Psyloidea database
26	RJB Geranium	Geranium Taxonomic Information System
27	Rotifera	Rotifera Database - Beta version
28	SalticidDB	Global Species Database of Salticidae (Araneae)
29	Scarabs	World Scarabaeidae Database
30	Solanaceae Source*	Solanaceae Source*
31	Species Fungorum	Species Fungorum
32	SysMyr	Systematic Myriapod Database
33	TicksBase	TicksBase
34	TIGR Reptiles	TIGR Reptile Database
35	Tineidae NHM	Global taxonomic database of Tineidae (Lepidoptera)
36	TITAN	Cerambycidae database
37	UCD	Universal Chalcidoidea Database
38	WCS	World Catalogue of Strepsiptera
39	World Checklist of Selected Plant Families	World Checklist of Selected Plant Families
40	World Gracillariidae*	Global Taxonomic Database of Gracillariidae*
41	World Umbellifer Database	World Umbellifer Database
42	WoRMS	World Register of Marine Species
43	WTaxa	Electronic Catalogue of Weevil names (Curculionoidea)
44	ZOBODAT (Vespoidea)	Zoological-Botanical Database (Vespoidea)

Table 2: number of species provided by each European GSD

	Database name	Number of species
1	AlgaeBase	31028
2	AnnonBase	2236
3	BDWD	137991
4	BSF	4564
5	CCW*	14834
6	ChiloBase	3146
7	CIPA	402
8	CLEMAM	?
9	Conifer Database	631
10	COOL	2364
11	Droseraceae Database	182
12	FishBase	31186
13	FLOW	10672
14	Global Compositae Checklist	25952
15	GloBIS (GART)	1635
16	ILDIS	19949
17	LepIndex	243006
18	LHD*	28
19	LIAS	15000
20	MolluscaFW	?
21	Nomen.eumycetozoa.com	1034
22	Odonata	5747
23	Porifera	8215
24	PSF	2946
25	Psyl'list	2107
26	RJB Geranium	350
27	Rotifera	1996
28	SalticidDB	4845
29	Scarabs	17844
30	Solanaceae Source*	1082
31	Species Fungorum	40468
32	SysMyr	10107
33	TicksBase	869
34	TIGR Reptiles	8656
35	Tineidae NHM	2340
36	TITAN	31408
37	UCD	19847
38	WCS	600
39	World Checklist of Selected Plant Families	93954
40	World Gracillariidae*	1841
41	World Umbellifer Database	3700
42	WoRMS	12807
43	WTaxa	80728
44	ZOBODAT (Vespoidea)	5972
	TOTAL	904269

2. Objectives

2.1. Compile data about GSDs for the improved CoL metadatabase

The development of an improved metadatabase implies the collection of data allowing a precise description of the database. The aim was to compile as much and as pertinent general details about each database as possible. These informations are about database name(s), editor(s), author(s), host organisation, taxonomic coverage, URLs, sources, structure, software...

Another batch of information deals with the peer-reviewing process. Our aim was to commpile a set of 3-5 possible referees for each GSD.

2.2. Measure the compliance of GSDs with the Sp2k SDS

Species 2000 has defined ten field groups to be the standard set of data for each species (or infraspecies). Our aim was to evaluate how well the standards used in each GSD were matching this set of standards.

2.3. Means of improving the compliance with the Sp2k SDS

The analysis of this survey will be used, in concertation with each custodian, to improve the compliance of each GSD with the Sp2k SDS. Moving a few steps towards greater compliance will greatly improve the quality of the data displayed through the CoL. This implies a prioritisation of fields, depending on additional or optionnal data available from the suppliers.

2.4. New data-fields to be implemented

New services to be tested and implemented in the CoL imply the delivery of new data by the GSDs. The choice of this small increase of the current dataset results both from the survey of CoL users undergone by WP2 and from the survey of CoL providers. The GSD custodians have been asked to prioritise the fields that they would like to become additional fields in the Sp2k SDS.

3. Tools

3.1. Questionnaire

A PDF structured form has been sent to all GSDs by e-mail on 18th November 2009. The questionnaire is divided in two parts:

- Database Summary: 17 general questions
- 4D4Life taxonomic database survey: 30 questions about taxonomic standards and concepts used in each GSD

3.2. CoL Annual Checklist 2009 (AC09)

Available through the CoL website, the 2009 annual checklist provides an index of validated scientific names and synonyms (when available) for each GSD when identified as taxonomic sectors.

3.3. Statistics from the CoL Annual Checklist 2010 beta (AC10β)

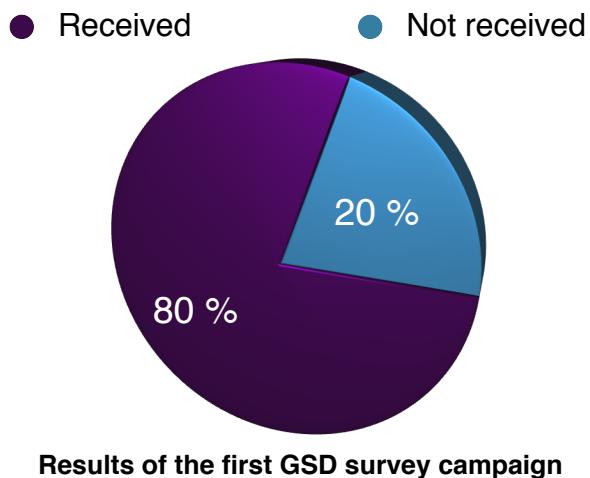
Statistics from the beta version of the 2010 Annual Checklist have been provided by the Sp2k secretariat in Reading and the production team in Los Baños, Philippines (Table 3)

Table 3: Statistics from the CoL Annual Checklist 2010 beta (Syn=Synonyms; Acc. Infrasp. Names=Accepted Infraspecific Names; Nom. Ref.=Nomenclatural Reference; Comm.to Nom. Refs.=Comments to Nomenclatural Reference; Acc. Ref.=Acceptance Reference; Ref. (null reftype)=Reference without assignment; Com. Names=Common Names; Distrib.=Distribution; LTS=Latest Taxonomic Scrutiny; Comm. to LTS=Comments to Latest Taxonomic Scrutiny; sp2k gentd.=sp2k generated)

Database Name	Accepted Names	Syn.	Acc. Infrasp. Names	Nom. Ref.	Comm.to Nom. Refs.	Acc. Ref.	References (common names)	Ref. (null reftype)	Com. Names	Distrib.	LTS	Comm. to LTS
1 AlgaeBase	31028	13814	7805	0		0	5	0	254	0	559	sp2k gentd.
2 AnnonBase	2236	1983	193	4647		0	0	43	0	0	1529	from GSD
3 BSF	4564	2455	127	697		0	0	0	105	4416	4691	
4 CCW	14834	1455	1008	16839		0	0	1	0	15290	15842	
5 ChiloBase	3146	1129	327	1208		841	0	0	0	4656	3473	
6 CIPA	402	239	21	0		259	0	0	419	418	4	sp2k gentd.
7 Conifer Database	631	1940	302	3734	sp2k gentd.	0	4	57	1346	931	933	
8 COOL	2364	1803	0	577		0	0	0	0	299	2364	
9 Droseraceae Database	182	158	52	157	sp2k gentd.	0	0	0	0	1930	234	
10 FishBase	31186	52880	467	0		6201	2842	0	272622	31462	17867	from GSD
11 FLOW	10672	6053	234	1388		0	0	0	0	2087	10906	
12 Global Compositae Checklist	25952	34635	0	0		0	0	0	0	0	25893	from GSD
13 GloBIS (GART)	1635	462	0	940		0	0	0	0	823	1635	
14 ILDIS	19949	23945	5118	422		2532	678	1417	18076	82590	5481	from GSD
15 LeplIndex	243006	156	0	0		824	0	0	0	0	0	
16 LHD	28	62	2	22		0	20	71	23	29	30	from GSD
17 Nomen.eumycetozoa.com	1034	1445	0	3177	sp2k gentd.	0	0	0	0	0	0	1034
18 Odonata	5747	1882	864	2320		0	0	22	0	0	6611	
19 PSF	2946	3407	99	1455		782	0	0	348	2981	51	from GSD
20 Psylllist	2107	790	0	553		0	0	201	0	1048	2107	
21 RJB Geranium	350	571	18	1298	sp2k gentd.	0	0	0	0	0	0	368
22 Rotifera Database	1996	607	251	0		21	0	0	0	1952	2247	
23 SalticidDB	4845	4065	9	4868		1414	0	1971	0	4769	4854	
24 Scarabs	17844	13447	1368	5854		0	0	0	0	19196	19212	
25 Solanaceae Source	1082	2101	19	3410	sp2k gentd.	0	0	0	965	0	1101	from GSD
26 Species Fungorum	40468	64959	1687	114443	sp2k gentd.	1455	912	12	918	0	42155	
27 SysMyr	10107	4016	1003	0		424	0	0	0	0	11110	from GSD
28 TicksBase	869	2473	21	15		0	114	0	571	614	890	
29 TIGR Reptile Database	8656	18444	4482	0		7109	0	16337	6307	8599	31	sp2k gentd.
30 Tineidae NHM	2340	522	0	1840		0	0	6	0	1798	2340	
31 TITAN	31408	160784	3142	0		9597	0	4497	0	67326	401	
32 UCD	19847	18567	0	6453		0	0	27	0	22331	19847	
33 World Checklist of Selected Plant Families	93954	136492	10534	281911	sp2k gentd.	0	0	30839	0	104483	104488	
34 World Gracillariidae	1841	411	7	0		1982	0	1	0	6082	1848	
35 WoRMS Brachiopoda & Phoronida	422	22	0	1		25	1	1	4	36	422	from GSD
36 WoRMS Cumacea	1286	24	0	1		25	0	0	0	1282	1286	from GSD
37 WoRMS Ophuoidea	2152	1664	60	12		68	7	0	21	318	2212	from GSD
38 WoRMS Porifera	8215	7542	299	1170		543	8	0	42	7128	8514	from GSD
39 WoRMS Proseriata & Kalyptorhynchia	732	248	13	229		143	0	0	0	682	745	from GSD
40 WTaxa	80728	0	0	12172		0	0	107	0	0	80728	from GSD
41 ZOBODAT (Vespoidea)	5972	1762	1367	0		130	0	0	0	0	7339	

4. Replies

- 35 replies out of 44 European GSDs (9 missing are AnnonBase, Blattodea Species File, CCW*, CLEMAM, Droseraceae Database, Mollusca FW, Salticidae Database, Solanaceae Source*, World Gracillariidae*) have replied to the first GSD survey campaign.



- 25 custodians or representatives were present at the First Start-up Meeting in Reading (15th - 17th September 2009)
- 23 GSDs were represented by 20 custodians or representatives at the Second General Meeting in Faro, Portugal (16th - 18th February 2010)

Table 4: Attendance of GSDs to the 4D4Life General Meetings (- = absent; P = present; R = represented)

	Database name	First General Meeting (Reading, UK)	Second General Meeting (Faro, Portugal)
1	AlgaeBase	-	-
2	AnnonBase	-	P
3	BDWD	P	-
4	BSF	-	-
5	CCW*	-	-
6	ChiloBase	P	P
7	CIPA	R	R
8	CLEMAM	-	-
9	Conifer Database	P	P
10	COOL	P	R
11	Droseraceae Database	P	-
12	FishBase	P	-
13	FLOW	P	P
14	Global Compositae Checklist	P	P
15	GloBIS (GART)	-	P
16	ILDIS	P	P
17	LepIndex	P	P
18	LHD*	P	-
19	LIAS	P	P
20	MolluscaFW	-	-
21	Nomen.eumycetozoa.com	-	-
22	Odonata	P	-
23	Porifera	-	R
24	PSF	-	-
25	Psyl'list	P	P
26	RJB Geranium	-	-
27	Rotifera	P	-
28	SalticidDB	R	R
29	Scarabs	-	P
30	Solanaceae Source*	-	-
31	Species Fungorum	P	P
32	SysMyr	-	P
33	TicksBase	-	-
34	TIGR Reptiles	-	-
35	Tineidae NHM	R	R
36	TITAN	P	R
37	UCD	-	-
38	WCS	P	P
39	World Checklist of Selected Plant Families	P	P
40	World Gracillariidae*	-	-
41	World Umbellifer Database	P	-
42	WoRMS	P	-
43	WTaxa	R	R
44	ZOBODAT (Vespoidea)	-	P

5. Species 2000 Standard Dataset

The Species 2000 Standard Dataset, Version 3.2 edited in December 2004 describes ten groups of data-fields:

- Accepted Scientific Name linked to References (obligatory)
- Synonym(s) linked to Reference(s) (obligatory, as appropriate)
- Common Name(s) linked to Reference(s) (optional)
- Latest taxonomic scrutiny (obligatory)
- Source Database (obligatory)
- Additional Data (optional)
- Family name (obligatory)
- Classification above family, and highest taxon (obligatory, as appropriate)
- Distribution (optional)
- Reference(s)

6. Results

6.1. Compiling Data for the New Metadatabase

The WP 3 questionnaire will be used to fill in the new metadatabase system that will be implemented in collaboration with WP 6 (Task 6.4). One goal of the development of a new metadatabase system is to facilitate the connection of new GSDs to the general taxonomic hierarchy. The structuration of the questionnaire used for this survey as an exportable CSV file will greatly facilitate this technical issue.

6.1.1. Peer review

Peer-review of provider GSDs has contributed to the excellent reputation of CoL (as well as, basically, the validation of the data by taxonomic experts). Many European GSDs have received certification (including positive peer-review) during the EuroCat project (funded under FP5) in years 2003 to 2006. Many other "new" GSDs have joined the CoL since, but some of them have missed the peer-review process. These "new" GSDs will be peer-reviewed in priority, then the process will be re-engaged for GSDs reviewed a long time ago.

However, the peer-reviewing is known to be a long-term process, because specialists are hardly inclined to spend too much time in reviewing works of others with no special benefit, so that replies are difficult to get. The questionnaire used for the survey has been useful to get in addition names of putative referees for each GSDs. Referees details and reviews will feed the metadatabase, on a confidential basis.

Table 5: Bank of Referees

Database Name	Referees	Last Review
AlgaeBase	3 referees	Reviewed in Aug. 2004 (Ref a), Sep. 2004 (Ref b), Oct. 2004 (Ref c) & Dec. 2004 (Ref d)
AnnonBase		Reviewed in Mar. 2005 (Ref a & b) & May 2005 (Ref c)
BDWD	6 referees	N/A
BSF		Review needed
CCW*		N/A
ChiloBase	5 referees	Review needed
CIPA		Reviewed in Jan. 2004 (Ref a)
CLEMAM		Reviewed in Feb. 2005 (Ref a)
Conifer Database		Review needed
COOL	5 referees	Review needed
Droseraceae Database		Review needed
FishBase	5 referees	Reviewed in Mar. 2004 (Ref a), May. 2004 (Ref b) & Nov. 2004 (Ref c)
FLOW		N/A
Global Compositae Checklist	3 referees	Review needed
GloBIS (GART)	3 referees	Review needed
ILDIS	1 referee	Reviewed in Dec. 2003 (Ref a), Mar. 2004 (Ref b)
LepIndex		Reviewed in Dec. 2004 (Ref a) & Feb. 2005 (Ref b)
Tineidae NHM		Reviewed (Tineidae) in Apr. 2004 (Ref a) & no_date (Ref b)
LHD	1 referee	Review needed
LIAS	5 referees	Reviewed in Feb. 2004 (Ref a) & Apr. 2005 (Ref b)
MolluscaFW		
Nomen.eumycetozoa.com	4 referees	Reviewed in Jun.-Oct. 2007 (Refs. a, b, c & d)
Odonata	5 referees	Reviewed in Aug. 2005 (Ref a) & Sep. 2005 (Ref b)
Porifera		Review needed
PSF	5 referees	Review needed
Psyl'list	2 referees	Review needed
RJB Geranium		Review needed
Rotifera	5 referees	Review needed
Salticidae Database	2 referees	Reviewed in Jan. 2004 (Ref a), Feb. 2004 (Ref b) & Jul. 2004 (Ref c)
Scarabs	6 referees	Reviewed in Mar. 2004 (Ref a), Apr. 2004 (Ref b), Sep. 2005 (Ref c)
Species Fungorum	3 referees	Reviewed in Dec. 2004 (Ref a)
SysMyr	7 referees	Review needed
TicksBase	3 referees	Reviewed in Mar. 2005 (Ref a)
TIGR Reptiles	9 referees	Reviewed in Jan. 2004 (Ref a), Nov. 2004 (Ref b) & Fev. 2005 (Ref c)
TITAN	7 referees	Reviewed in Dec. 2003 (Ref a), Jan. 2004 Ref b) & May 2004 (Ref c)
UCD	6 referees	Reviewed in Jan. 2004 (Ref a & b) & Mar. 2004 (Ref c)
WCS	3 referees	Review needed
World Checklist of Selected Plant Families	5 referees	Reviewed in Feb. 2004 (Sapotaceae: Ref a; Magnoliaceae: Ref b), Mar. 2005 (Euphorbiaceae: Ref c), Apr. 2005 (Euphorbiaceae: Ref d)
World Umbellifer Database	5 referees	Review needed
WoRMS	5 referees	Review needed
WTaxa	3 referees	Reviewed in Dec. 2004 (Ref a) & Apr. 2005 (Ref b)
ZOBODAT (Vespoidea)	1 referee	Reviewed in Jan. 2004 (Ref a)

6.1.2. Wrapper

The survey of wrappers connecting dynamically GSDs to the CoL is being done by Viktoras Didziulis, Systems Manager in Reading (WP 5). The validity of old wrappers dating from the EuroCat project has to be tested. In the case the test fails, the SPICE-based wrapper will have to be reactivated. For "new GSDs", an operative wrapper will have to be written. In both cases, these tasks should be achieved through the pilot projects funded by the 4D4Life project. However, some GSDs may be involved in the trials of wrapper development for the new e-2 system during the last year of the project.

Table 6: Availability of GSD wrappers

	Database name	Wrapper exists	Queries functional
1	AlgaeBase	no	no
2	AnnonBase	no	no
3	BDWD	no	no
4	BSF	yes	yes
5	CCW	no	no
6	ChiloBase	yes	no
7	CIPA	yes	tests in progress
8	CLEMAM	yes	tests in progress
9	Conifer Database	no	no
10	COOL	yes	tests in progress
11	Droseraceae Database	no	no
12	FishBase	no	no
13	FLOW	yes	tests in progress
14	Global Compositae Checklist	no	no
15	GloBIS (GART)	no	no
16	ILDIS	yes	no
17	LepIndex	yes	no
18	LHD	no	no
19	LIAS	yes	tests in progress
20	MolluscaFW	yes	yes
21	Nomen.eumycetozoa.com	no	no
22	Odonata	yes	tests in progress
23	Porifera	no	no
24	PSF	yes	yes
25	Psyl'list	yes	tests in progress
26	RJB Geranium	no	no
27	Rotifera	no	no
28	SalticidDB	yes	tests in progress
29	Scarabs	no	no
30	Solanaceae Source	no	no
31	Species Fungorum	yes	no
32	SysMyr	yes	tests in progress
33	TicksBase	yes	no
34	TIGR Reptiles	no	no
35	Tineidae NHM	yes	no
36	TITAN	yes	tests in progress
37	UCD	yes	yes

	Database name	Wrapper exists	Queries functional
38	WCS	no	no
39	World Checklist of Selected Plant Families	yes	yes
40	World Gracillariidae	no	no
41	World Umbellifer Database	no	no
42	WoRMS	no	no
43	WTaxa	no	no
44	ZOBODAT (Vespoidea)	yes	tests in progress

6.1.3. Formal Membership to Species 2000

The current membership of each GSD to Species 2000 species will be checked in collaboration with Rebecca Mann, Administrative (WP 5) and Financial Assistant (WP 1) in Reading. The results of the WP 3 consultation through the questionnaire is given here, both for the current membership itself and the custodian's wish for the future if membership is lacking.

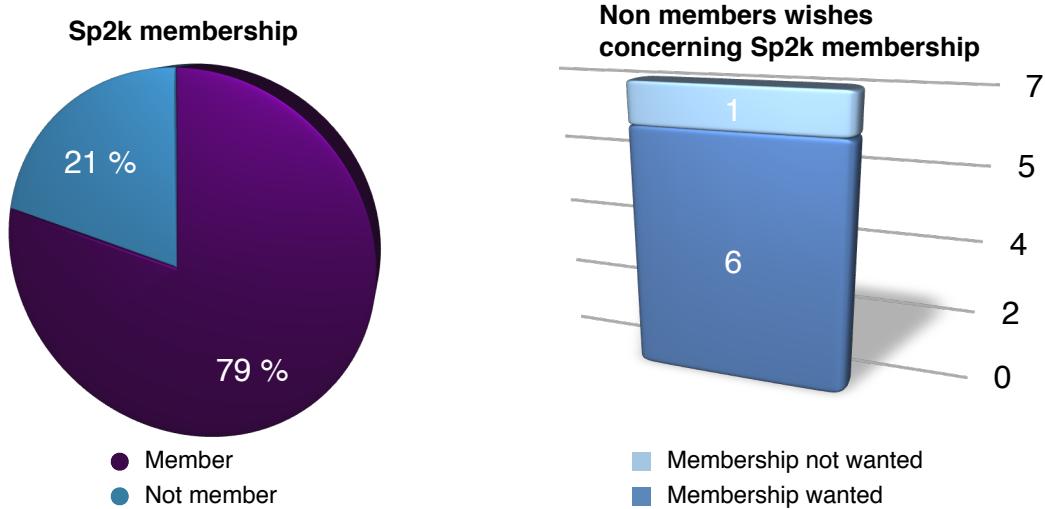


Table 7: Current membership of European GSD to Species 2000species 2000

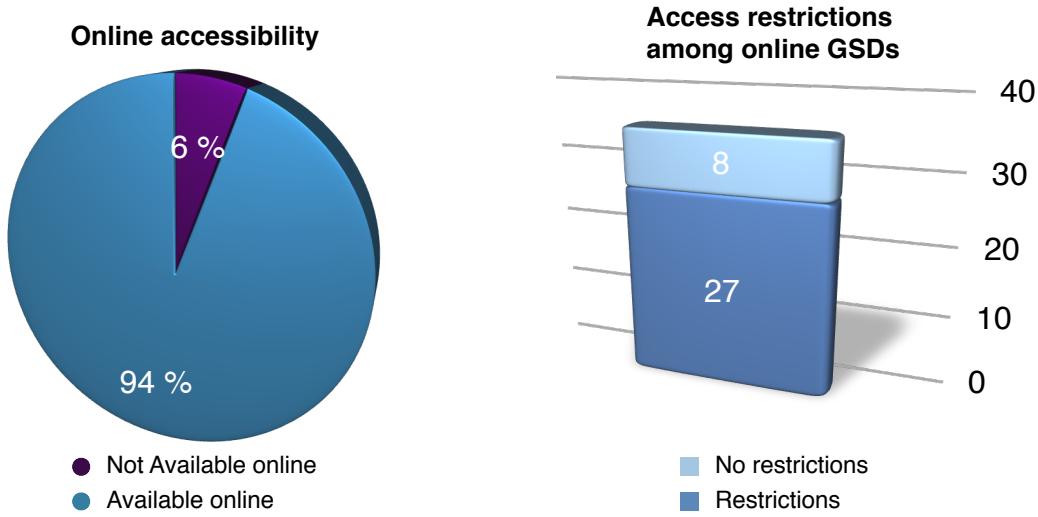
Short Database Name	Sp2k member	Sp2k membership wanted
AlgaeBase	Yes	--
BDWD	No	Yes
ChiloBase	Yes	--
CIPA	Off	Off
Conifer Database	Yes	-
COOL	Yes	-
FishBase	Yes	-
FLOW	Yes	-
Global Compositae Checklist	No	Yes
GloBIS	No	No
ILDIS	Yes	-
LepIndex	Yes	-
LHD	Yes	-
LIAS	Yes	-
Nomen.eumycetozoa.com	Yes	-
Odonata	Yes	-
Porifera	Yes	-
PSF	Yes	-
Psyl'list	Yes	-
RJB Geranium	Off	Off
Rotifera Database	Yes	-
Scarabs	Yes	-
Species Fungorum	Yes	-
Sysmyr	No	Yes
Ticksbase	Yes	-
TIGR Reptiles	Yes	-
Tineidae	Yes	-
TITAN	Yes	-
UCD	Yes	-
World Checklist of Selected Plant Families	Yes	-
World Umbelliferae Database	Yes	-
WoRMS	No	Yes
WCS	No	Yes
WTaxa	Yes	-
ZOBODAT (Vespoidea)	No	Yes

6.1.4. Hosting

Table 8: Institutions that host the GSDs

Short DatabaseName	Institution
AlgaeBase	National University of Ireland, Galway, Ireland
BDWD	Natural History Museum of Denmark, Copenhagen, Denmark
ChiloBase	Department of Biology, University of Padova, Padova, Italy
CIPA	Université Pierre et Marie Curie, Paris VI, Paris, France
Conifer Database	Hosted by the custodian
COOL	Muséum national d'Histoire naturelle, Paris, France
FishBase	FishBase Consortium
FLOW	Muséum national d'Histoire naturelle, Paris, France
Global Compositae Checklist	Wageningen University, Wageningen, The Netherlands
GloBIS	Museum für Naturkunde, Berlin, Germany
ILDIS	ILDIS project
LepIndex	Natural History Museum, London, UK
LHD	Hosted by the custodian
LIAS	Botanische Staatssammlung München, Munich, Germany
Nomen.eumycetozoa.com	Real Jardín Botánico, CSIC, Madrid, Spain
Odonata	Naturalis, Leiden, The Netherlands
Porifera	VLIZ, Flanders Marine Institute, Oostende, Belgium
PSF	Natural History Museum, London, UK
Psyl'list	Muséum national d'Histoire naturelle, Paris, France
RJB Geranium	Real Jardín Botánico, CSIC, Madrid, Spain
Rotifera Database	Royal Belgian Institute of Natural Sciences, Brussels, Belgium
Scarabs	TSJ BVBA, Belgium
Species Fungorum	CABI Europe, Egham, UK
Sysmyr	Bavarian Natural History Collections, Munich, Germany
Ticksbase	Utrecht University, The Netherlands
TIGR Reptiles	Deutsches Krebsforschungszentrum, Heidelberg, Germany
Tineidae	Natural History Museum, London
TITAN	Institut de Recherche pour le Développement, Marseille, France
UCD	Natural History Museum, London
WCS	Department of Zoology, University of Oxford, UK
World Checklist of Selected Plant Families	Royal Botanic Gardens, Kew, UK
World Umbelliferae Database	Royal Botanic Garden Edinburgh, UK
WoRMS	VLIZ, Flanders Marine Institute, Oostende, Belgium
WTaxa	Consejo Superior de Investigaciones Científicas, Madrid, Spain
ZOBODAT (Vespoidea)	Upper Austrian State Museum, Linz, Austria

6.1.5. Accessibility/Restrictions



6.1.6. Coverage/Completeness

All GSDs have a worldwide taxonomic coverage, except CIPA (restricted for the moment to South America). Completeness of species lists has also been scrutinised: 91% of the GSDs have a completeness higher than 50% and 67% higher than 90%.

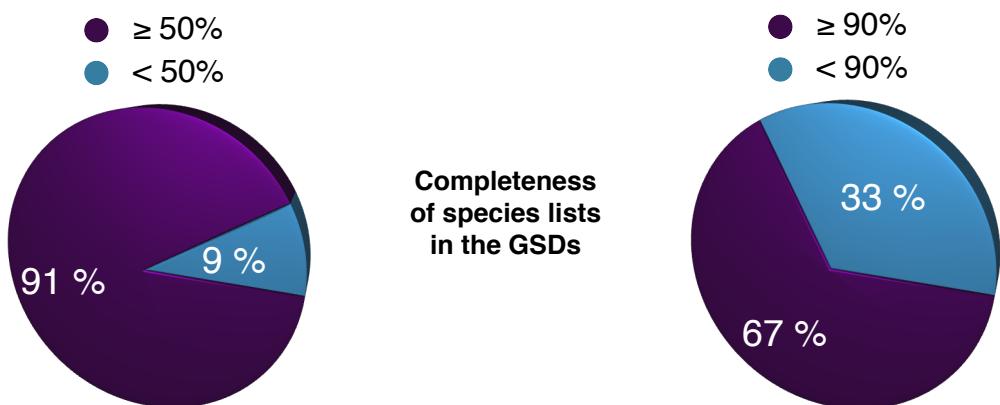
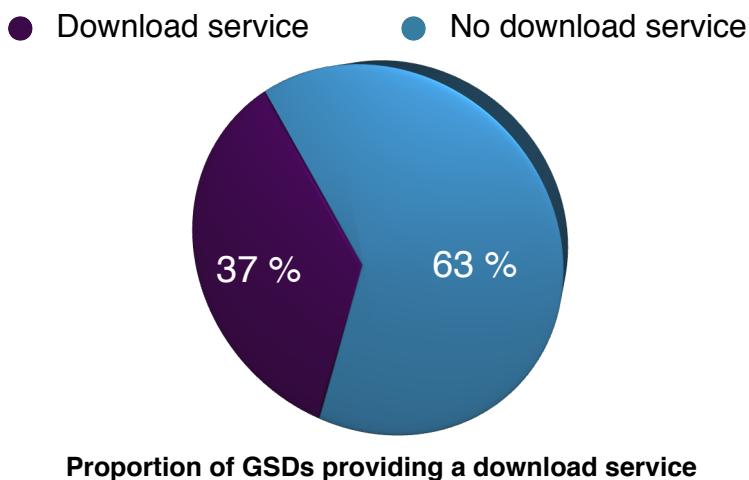


Table 9: Completeness of species lists in the GSDs

Short DatabaseName	Completeness
Conifer Database	100 %
PSF	100 %
Odonata	99,90 %
TIGR Reptiles	99,90 %
ChiloBase	99 %
FishBase	99 %
GloBIS	99 %
LHD	99 %
RJB Geranium	99 %
Rotifera Database	99 %
Ticksbase	99 %
UCD	99 %
World Umbelliferae Database	99 %
BDWD	98 %
WCS	98 %
Porifera	97 %
COOL	95 %
ILDIS	95 %
Nomen.eumycetozoa.com	95 %
Tineidae	95 %
TITAN	92 %
Sysmyr	90 %
Global Compositae Checklist	85 %
AlgaeBase	80 %
FLOW	80 %
LIAS	75 %
WTaxa	68 %
Psyl'list	65 %
WoRMS	65 %
Scarabs	50 %
CIPA	45 %
Species Fungorum	45 %
World Checklist of Selected Plant Families	33 %
ZOBODAT (Vespoidea)	N/A
LepIndex	?

6.2. Searchability/Download service

All databases available online are searchable through their web portal, but only 37% of surveyed GSDs provide a download service of their data.

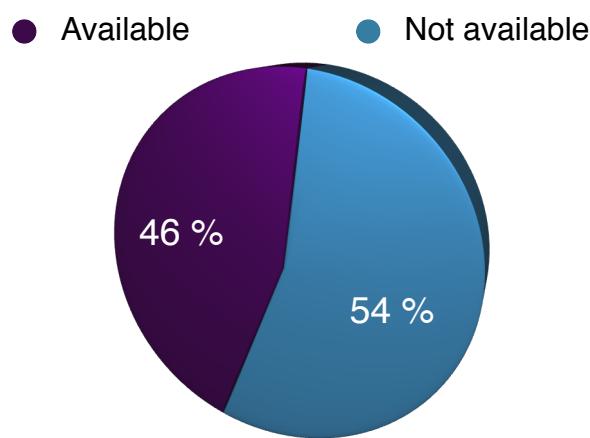


6.3. Measure of compliance

6.3.1. Accepted Scientific Name linked to References (obligatory)

Provided by all GSDs. For each accepted name, all GSDs give the full genus name, the specific epithet, and the author string.

Less than half of the GSDs (46%) have the possibility to provide names with the "provisionnally accepted" status.



Information about the provisionnal acceptance of a name among all GSDs

No statistics are available for the moment about the proportion of provisionally accepted *versus* accepted names in the databases. Three (LepIndex, Scarabs, TITAN) out of the 35 surveyed GSDs give no information at all about the status of the provided names.

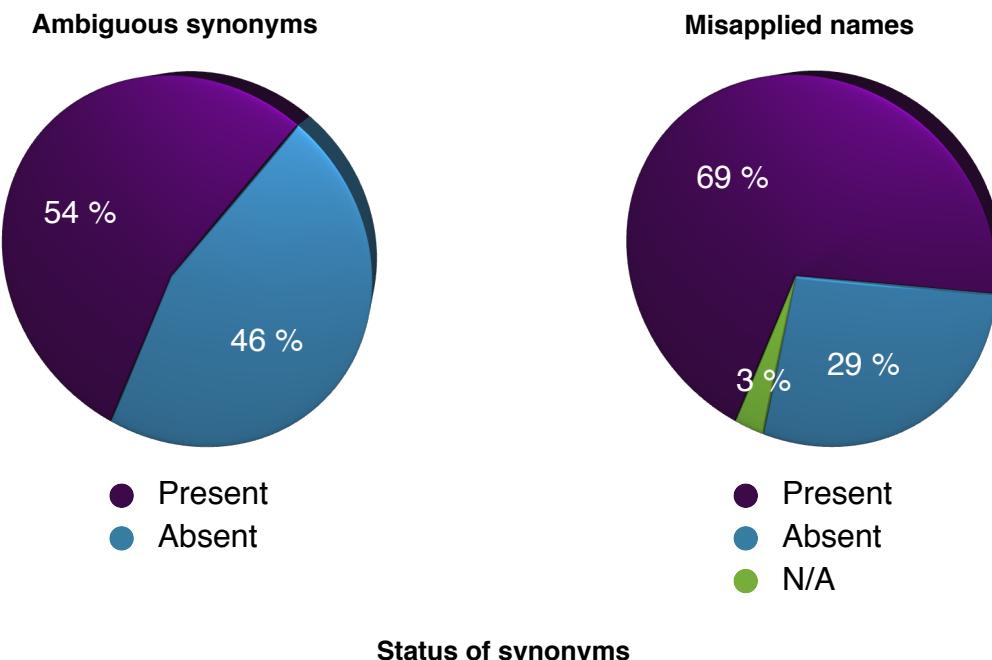
All other GSDs provide Accepted Scientific Names with a wide range of statuses (most of them involve synonyms instead of valid names): valid, previous combination, synonym, nomen novum, nomen praecoccupatum, homonym, incorrect original spelling, incorrect subsequent spelling, junior synonym, accepted, unaccepted, nomen nudum, alternate representation, nomen dubium, species inquirenda, temporary name, uncertain, unchecked, provisionally accepted, invalid, doubtfull, pro parte, nomen protectum, nomen oblitum, previous combination, nomen revivisco.

It appears that the statuses requested by the Species 2000/species 2000 Standard Dataset (accepted/provisionnaly accepted) for names are not nomenclatural statuses. In order to avoid any confusion between a taxonomic opinion made by the custodian (from provisionnaly to accepted) and the various nomenclatural statuses available from the GSDs for each name, the definition of NameStatus (in the sense of CoL) must be spread largely among the providers community.

6.3.2. Synonym(s) linked to Reference(s) (obligatory, as appropriate)

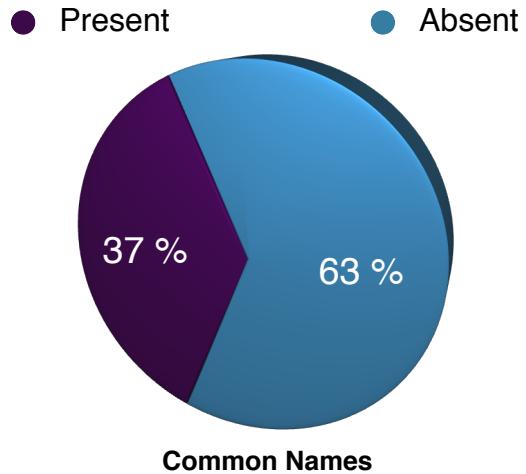
Again, all surveyed GSDs store synonyms, all (except LepIndex, badly structured) providing full genus name, specific epithet, author string and name status (except one for this last information: Scarabs). In AC10 β , WTaxa displays no synonyms.

The various statuses associated with names in Synonyms were also given by custodians in the part dealing with Accepted Scientific Names (see 6.2.1 above).

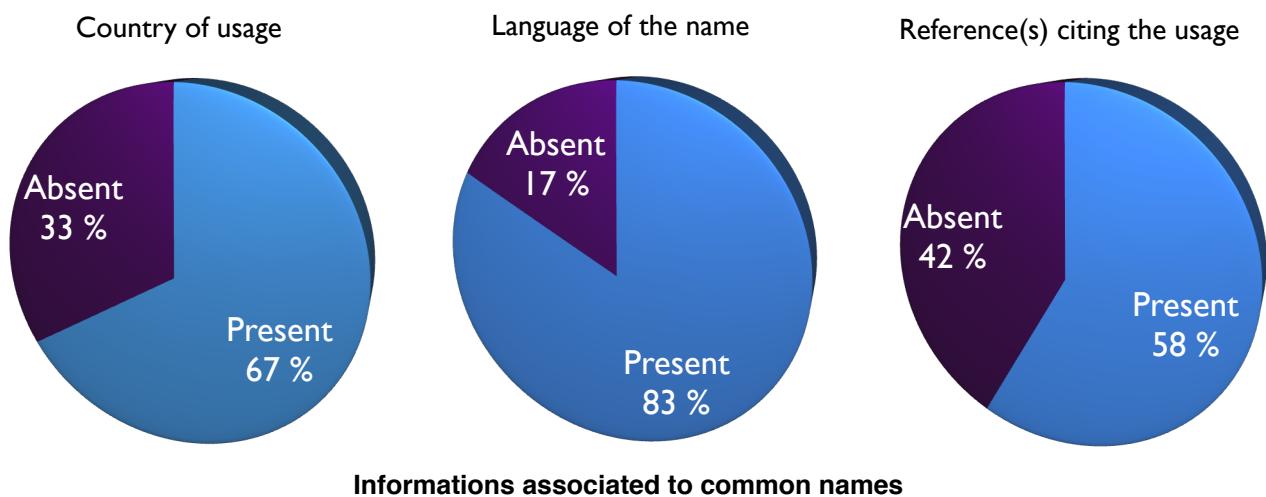


6.3.3. Common Name(s) linked to Reference(s) (optional)

Common names are present in only 37% of all surveyed GSDs.

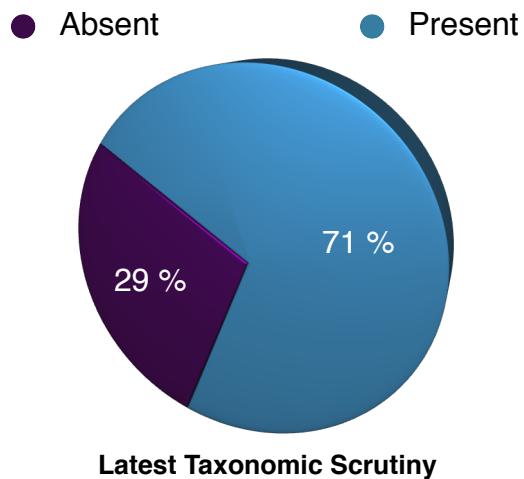


Among these 37%, the following statistics give details about associated informations (country of usage, language, reference):



6.3.4. Latest taxonomic scrutiny (obligatory)

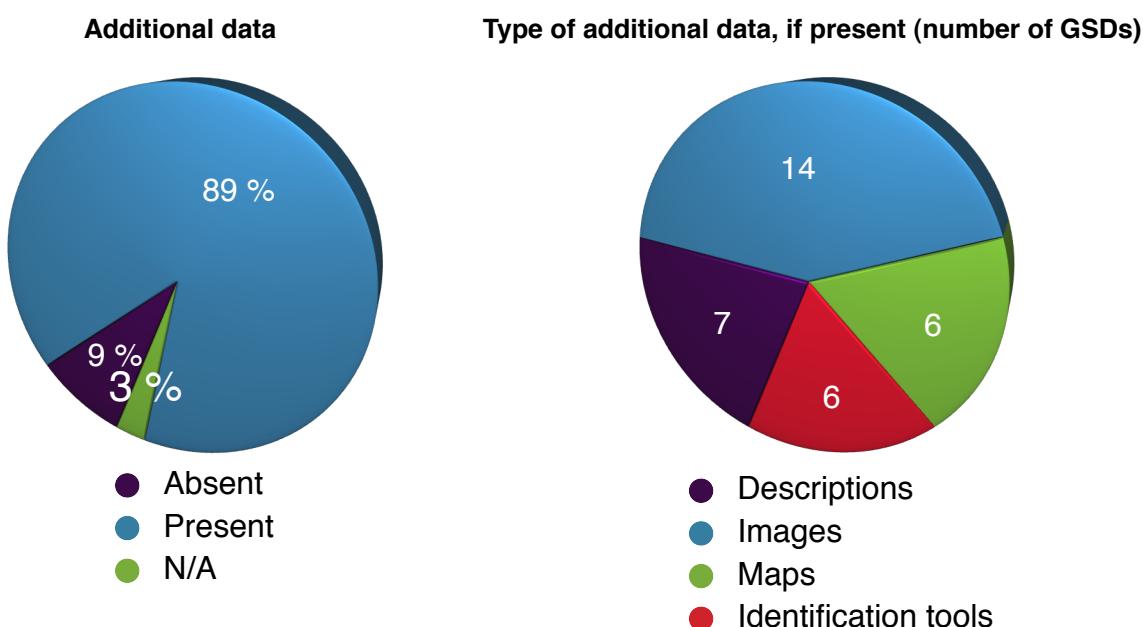
This information is provided by 71% of the surveyed GSDs. It is automatically generated by Species 2000/species 2000 with the name of the custodian and the last update for the other 29%.



The survey has shown that the definition of this field is not filled in with the same information by all GSDs. Here also, the definition has to be clarified and spread among the GSD Network. The “record of the latest taxonomic scrutiny of this species record in the source database” is sometimes interpreted inside the community of zoologists as the last author (with date) who validated the name, even if this information comes from a very old publication.

6.3.5. Source Database (obligatory)

No particular issue.



6.3.6. Additional Data (optional)

Table 10: other additional data present in GSDs

Literature	Parasites	Conservation status
Distribution	Phoretics	Geo-referenced specimens
Common names	Predators	Habitats
Type depository	Mimetics	Aquaculture techniques
Host-plants	Visited flowers	Uses
Legal aspects	Reproduction	Pathogens transmitted
Glossary	Trophic ecology	Type localities
Digitized literature (OCR)	Genetics	Life span
Host records	Human use	Free text comment
Population dynamics	Etymology	

6.3.7. Family name (obligatory)

No particular issue from the survey. A discussion occurred during the WP3 workshop in Faro (Second General Meeting) about the lack of family assignment for some names. A clear recommendation has to be done on Family Assignment:

- if really not known : “not assigned”,
- if expert choice: a sentence to be found to document this clearly.

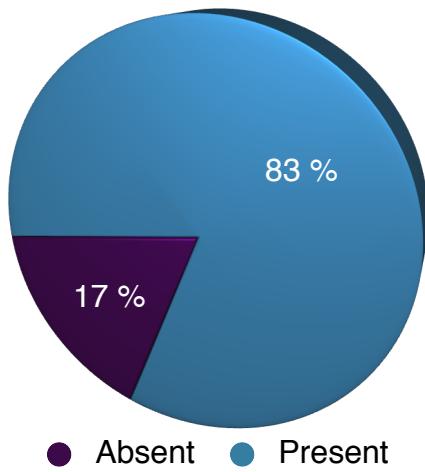
Tracing back knowledge as much as possible is recommended and in any case, the CoL Annual Checklist should become at least, the reference for this assignment.

6.3.8. Classification above family, and highest taxon (obligatory, as appropriate)

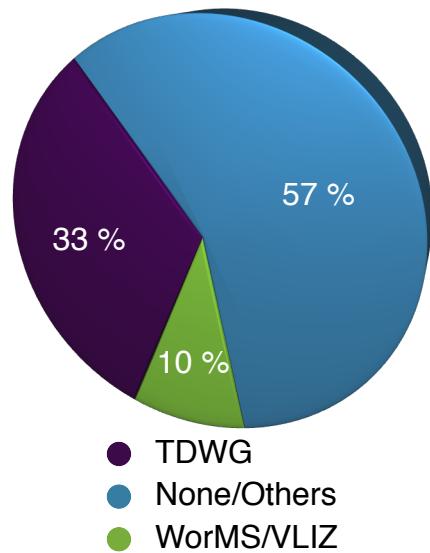
No particular issue.

6.3.9. Distribution (optional)

Information about distribution



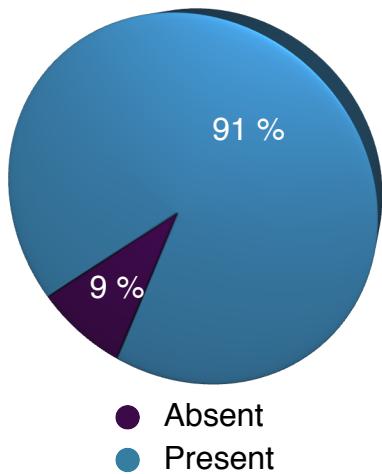
Standards used



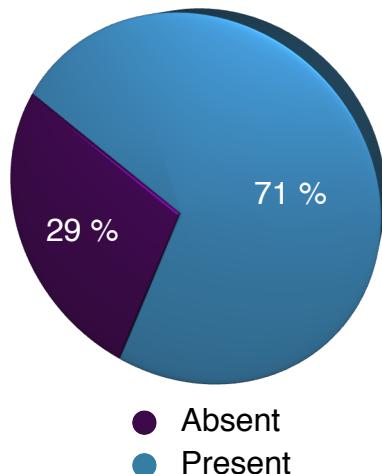
6.3.10. Reference(s)

Assignment by the provider of references to one reference type or another is not a trivial issue, as shown by the high number of references unassigned in AC10 β (94154 references in 21 databases).

Nomenclatural Reference



Acceptance Reference



Network Discussion

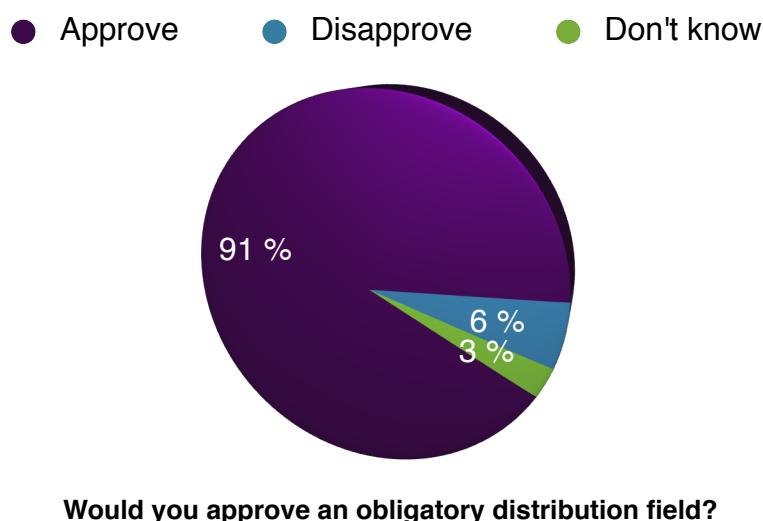
1. Extensions to the bio-data fields

GSD custodians have been consulted in the process of extending the current Sp2k Standard Dataset with new data fields. Custodians were asked which data fields they would approve to become additional fields in the Species 2000 Standard Dataset, among following ones:

- Geographical distribution
- Take taxon LSIDs directly from the GSDs (if LSIDs present in your database)
- Fossils
- Known habitat (whether the species is a marine, freshwater, estuary, terrestrial... one)
- Subgenus
- Host-relationships/Associated organisms
- Reference to Nomenclators (via LSIDs)
- Maximum size of the organism
- Conservation/legislation status
- Hybrids: names should appear in CoL
- Hybrids: should be linked to both "parents" taxa if present in any database

1.1. Geographical distribution

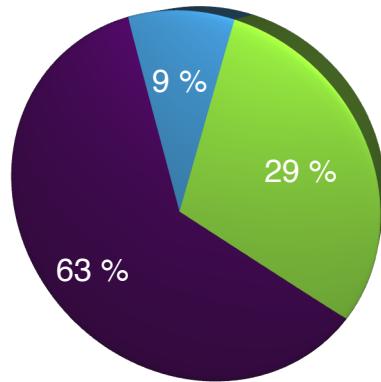
A large majority of the GSDs approved an obligatory geographical distribution field in the future version of the Sp2k SD. Custodians also expressed their interest for a well-structured field, following TDWG-Level4 standard. This is in full concordance with potential new services asked by external users in WP2.



1.2. Take taxon LSIDs directly from the GSDs (if LSIDs present in the database)

This option has been approved by 62% of the surveyed custodians.

● Approve ● Disapprove ● Don't know

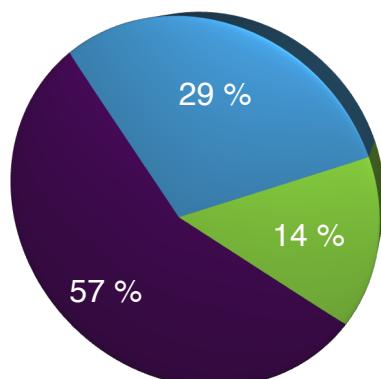


Would you approve the use by CoL of LSIDs provided directly by the GSD?

1.3. Fossils

The presence of a field dedicated to fossil taxa in CoL has been approved by 57% of the surveyed GSDs. Due to difficulties in manipulating "form taxa" (as they appear in palaeobotany) from a taxonomic point of view, and without conclusive tools to reconstruct the "real taxon", it has been agreed that the presence of a marker in the GSD as a quoted field (*i.e.* it is or not a fossil) was necessary, but that the CoL would not provide these fossil or extinct taxa's names directly. However, entering availability of such information in the GSD in the metadatabase, in order to find a link to the database where this fossil data occurs, would satisfy custodians wishes.

● Approve ● Disapprove ● Don't know

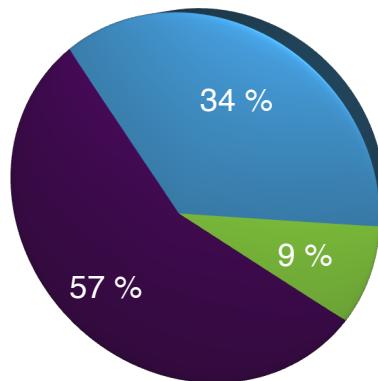


Would you approve a field for fossil taxa?

1.4. Habitat

A field describing the habitats of the taxa in CoL has been approved by 57% of the surveyed GSDs. The standards have still to be defined in the Network Discussion Document.

● Approve ● Disapprove ● Don't know

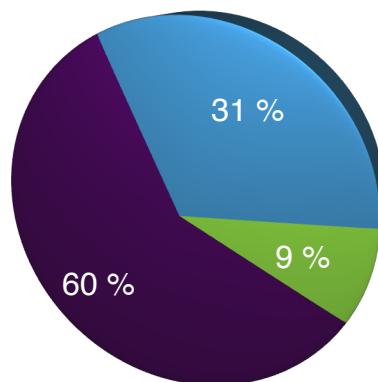


Would you approve a field describing habitat?

1.5. Subgenus

This rank is wanted to be added to the Sp2k SD, in addition to Genus, Species and Infraspecies levels.

● Approve ● Disapprove ● Don't know

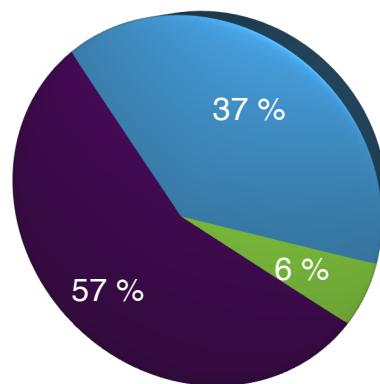


Would you approve a field for the rank "Subgenus"?

1.6. Associated organisms

57% of the GSDs would approve a field advising the associated organisms or the hosts of a CoL taxon. A recommendation will be made for promoting an external or satellite project to establish datasets of relations between 2 species in the CoL.

● Approve ● Disapprove ● Don't know

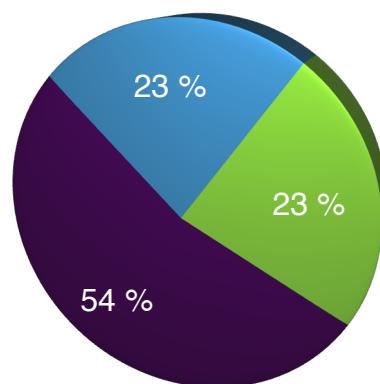


Would you approve a field for the other organisms associated to a taxon?

1.7. Reference to Nomenclators

54% of the GSDs would approve a link between CoL names and Nomenclators. This internal service would greatly help providers to maintain their data up-to-date through an alert service.

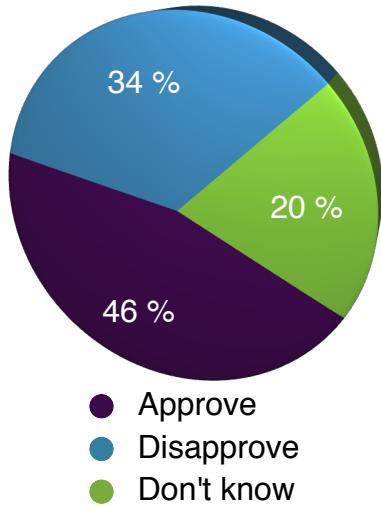
● Approve ● Disapprove ● Don't know



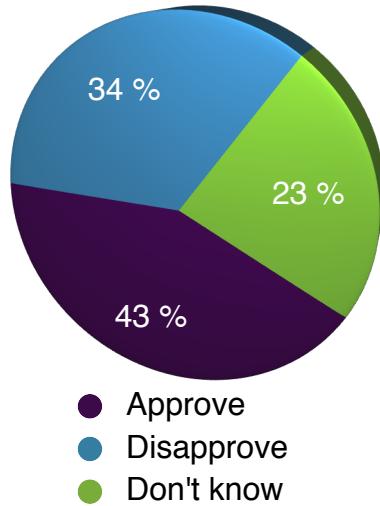
Would you approve links to Nomenclators (via LSIDs)?

1.8. Rejected additional fields

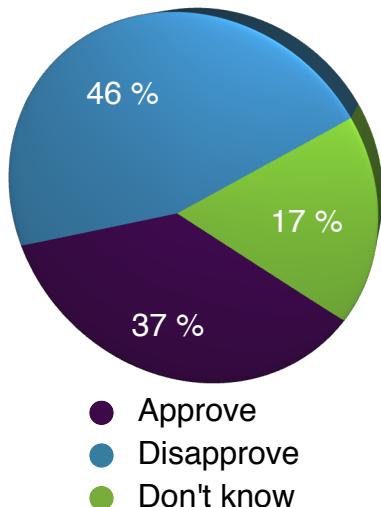
Hybrids: Names in Col



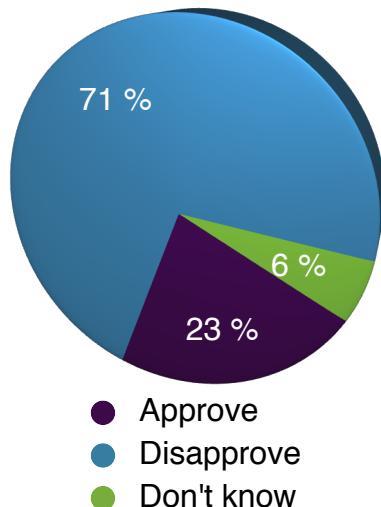
Hybrids: Links betw. taxa



Conservation Status



Maximum Size



Rejected additional fields

2. Pilot Projects

2.1. Priorities

The analysis of the data provided by the GSDs to the CoL and the conclusions of the users survey undertaken by Work Package 2 has led to a listing of actions, which will improve the compliance of GSDs data with regards to the CoL standards.

These priorities are:

- Synonymy: clarify the names for synonyms and clean the data
- Geography: structure the data and use TDWG-Level4 standards
- References: structure the data
- Common Names: give common names at least for important (economic, social...) taxonomic groups. Structure the data in providing also country and language of usage as well as references
- Latest Taxonomic Scrutiny: structure the data

This increase of data compliance as well as data coverage will be undertaken by each GSD through two 2-month pilot projects.

2.2. Decisions

GSDs represented at the Second General Meeting in Faro, Portugal (16-18 February 2010) have been assigned their pilot projects after discussion about data compliance (following survey's conclusions) and agreement.

Concerning GSDs for which no representatives were present in Faro at the Second General Meeting for the WP3 Workshop (AlgaeBase, BDWD, BSF, CLEMAM, Droseraceae Database, FishBase, MolluscaFW, Nomen.eumycetozoa.com, Odonata, PSF, RJB Geranium, Rotifera, Salticidae Database, TicksBase, TIGR Reptiles, UCD, World Umbellifer Database), meetings with custodians will be setup before July 2010 in order to take decisions about pilot projects.

Table 11: Decisions for pilot projects

Database Name	Pilot project 1	Pilot project 2
AnnonBase	common names	distribution
ChiloBase	TDWG-L4	common names
CIPA	data increase / become global	references
Conifer Database	full references	a few common names + ref to common names
COOL	references	common names
FLOW	references	common names
Global Compositae Checklist	infraspecific taxa	distribution
GloBIS (GART)	TDWG-L4 + common names	wrapper (after migration --> new environment)

Database Name	Pilot project 1	Pilot project 2
ILDIS	find software (tests)	nomenclatural references + common names
LepIndex & Tineidae NHM	migration to a sustainable system (easier for Tineidae/discussion needed for LepIndex)	
LIAS	solve SpFung/LIAS overlap + adapt LIAS to CoL	
Porifera	wrapper	uniformisation of the distribution (TDWG-L4)
Psyl'list	common names	references
Scarabs	common names	data (Melolonthinae)
Species Fungorum	clean up nom_references	distribution (spp of economic importance)
SysMyr	synonymy + common names + TDWG-L4 standards	wrapper
TITAN	move to DBTNT schema	references
WCS	migrate Word-->XML format	add taxonomy with editor tool
World Checklist of Selected Plant Families	Name LSIDs	new wrapper (2nd batch)
WTaxa	clean data (prov. acc. to acc. or syn.)	wrapper + a few common names
ZOBODAT (Vespoidea)	references + distribution + common names	add new GSD (Ciliates)

Conclusion

The work carried out by MNHN in Workpackage 3 has supplied an extensive survey of the extent of data supply, and of data-standard compliance by the GSDs. It has also provided opinions and agreement on a small set of additional fields proposed by 4D4Life for addition to the Species 2000 Standard Data Set.

As a result of these network discussions it has further been agreed what pilot-projects should be undertaken for each of the GSDs to move a few steps closer to full compliance. Just a subset of the GSDs remain to agree their pilot-projects before July 2010, the technical start date for this Task 3.4.

The greater compliance with the Standard Data, and the addition of these new fields is expected to make a significant contribution to the perceived quality and uniformity of the Catalogue of Life service.

Appendix

Appendix 1: 4D4Life Metadata Questionnaire

Questionnaire part E:

Database summary

Taxonomic Databases Survey

Full Database Name: _____

Short Database Name
(nickname or acronym): _____

Custodian: _____

1. Institution that host your database

2. Author(s) of the database content (could be individual or collective):

3. Editor(s) of the database content (if applicable):

4. Which major taxa does the Database cover (please, give both Scientific and English names of the group)?

Latin name & rank: _____

English Name: _____

First 'parent taxon' above the database scope (e.g. Order, Class, Phylum or Kingdom, and its name): _____

5. Are there any restrictions/access rights to your database? Yes / No

6. How your database should be cited?

7. How your data sources should be cited?

8. Does the Database cover the taxon worldwide? Yes / No

If not, which geographical areas does it incorporate?

9. Is your database available online? Yes / No

If yes, URL of the Database (title page): _____

10. Is your database human searchable (through a portal)? Yes / No

If yes, URL of the search page: _____

1 / 2

Questionnaire part E

Taxonomic Databases Survey

Database summary

11. Is your database machine searchable (web server/server to server link)?

Yes / No

12. Do you provide a download service?

Yes / No

13. What are the sources of the taxonomic data and opinions (*Literature, Expert data, Collection data [specimens], Survey data, On-line sources, others - explain*)?

14. In what form is the Dataset? (*Relational db, Excel table, HTML system, Flat text, Card index, Monograph, etc.*)

15. What is the Database Software?

Software of 'master database':

Software of web-based database:

16. Is the Database species- or name- based or both (i.e. is it a synonymous checklist or a name list)?

17. Presence of additional data & Comment field:

Does the Database have additional data? Yes / No

Descriptions Yes / No

Identification tools Yes / No

Images Yes / No

Maps Yes / No

Others (i.e. Habitats, Conservation Status, etc.; please, explain)

Yes / No

2 / 2

Questionnaire part 2: 4D4Life/Species2000

Taxonomic Databases Survey



Full Database Name:

Short Database Name
(nickname or acronym):

Custodian:

1. Version of the database (number or code):

2. If known, release date (or date of download for the Catalogue of Life):

3. Search URL of the Database (search page):

4. URL of Project or Institution Logo (if logo available on a sever):

5. Name of any wider project (if the Database is linked to, or part of any other database, project or programme):

6. How many species does it contain at present? (Please, give the precise number of species for the present version):

7. How many EXTANT species does it contain?

8. How many EXTINCT species does it contain, if any? If extinct species present, how are they spotted or marked in your database?

9. What is the approximate total number of names (species+synonyms)?

10. What is the estimated number of known species when database will be completed?

11. What is the estimated total number of names (species+synonyms) when database will be completed?

Questionnaire part 2:

4D4Life/Species2000

Taxonomic Databases Survey



12.What is the completeness of species list in the Database (in percentage)?

13.How many taxonomists are involved in maintenance and data quality verification?

14.Does the Database contain non-standard characters (i.e. diacritics etc.)?

 Yes / No

If 'yes', where do they occur in the data (i.e. scientific names, author string, common names, bibliographic references, person names, distribution data, comment fields, etc.)?

15.What is the approximate frequency of updating the Database (daily, weekly, monthly, annually, sporadically, fixed version)?

16.Which ranks of higher taxa above the Family does the Database contain (Suborder, Order, Class..)?

17.What is the name of this highest taxon covered by the database (e.g. Super-Family Puyoidea, Family Amanaceae, Phylum Porifera...)?

18.What is the name of the very next higher taxon among these five basic ranks -Kingdom, Division (Phylum), Class, Order, Family- on which the attachment can be made in Species 2000 hierarchy (e.g., with link to examples given in question 17: Order Hemiptera, Order Magnoliales, Kingdom Animalia...)?

19.Which medium is the database available in? (HDD, CD, DVD, local server, Web, etc.)

20.Does the Database have infraspecific taxa with accepted names?

 Yes / No

If 'yes', do you keep this infraspecific information in a separate field?

 Yes / No

21.Does the Database have infraspecific taxa in synonymy?

 Yes / No

If 'yes', which Infraspecific Ranks are used in the Database (i.e. Subspecies, Varieties, Forms, Strains):

22.Does the database have Infrageneric groupings?

 Yes / No

If 'yes', of which kind? (Subgenus, Section, species group..)

Questionnaire part 2:

4D4Life/Species2000

Taxonomic Databases Survey



23. Accepted, Valid or Correct scientific name:

Is this information present for all species?

Full Genus Name

 Yes / No

Specific Epithet

 Yes / No

Author String

 Yes / No

Name Status (i.e. Accepted, Provisionally accepted, others - please, give a list of statuses)

 Yes / No

Can species names in your database be classified as:

Accepted species

 Yes / No

Provisionally accepted species

 Yes / No Yes / No

Do species have a Nomenclatural Reference (bibliographic source which contains the original [validating] publication of taxon name or new name combination)

 Yes / No

Do species have Acceptance Status Reference(s) (bibliographic sources which accept this species in the same taxonomic status, and with the same name)

 Yes / No

24. Synonyms

Does the Database have synonyms?

 Yes / No

Is this following information present for all synonyms?

Full Genus Name

 Yes / No

Specific Epithet

 Yes / No

Author String

 Yes / No

Name Status (i.e. Synonym, Ambiguous Synonym, Misapplied Name, others - please, give a list of statuses)

 Yes / No

Can synonyms in your database be classified as:

Synonyms (names which point unambiguously at one species)

 Yes / No

Ambiguous synonyms (names which are ambiguous because they point at the current species and one or more others e.g. homonyms, pro parte synonyms)

 Yes / No

Misapplied names (names that have been wrongly applied to the current species and may also be correctly applied to another species)

 Yes / No

Questionnaire part 2:

4D4Life/Species2000

Taxonomic Databases Survey



Do synonyms have a Nominalatural Reference (bibliographic source which contains the original [validating] publication of taxon name or new name combination)

 Yes / No

Do synonyms have Status Reference(s) (bibliographic sources which accept this species in the same taxonomic status, and with the same name)

 Yes / No

25. Common Names

Does the Database have Common Names?

 Yes / No

If 'yes', does it give for each common name:

The Country of usage

 Yes / No

The Language of the name

 Yes / No

Reference(s) citing the usage

 Yes / No

26. Latest taxonomic scrutiny

Does the Database have a taxonomic scrutiny field (name of taxonomic expert who is responsible for the taxonomic concept and data quality of each species record)?

 Yes / No

If 'yes', does the Database show:

Name of Expert

 Yes / No

Date of latest scrutiny

 Yes / No

27. Family name

Does the Database contain information about which Family the species belongs to?

 Yes / No

28. Distribution

Does the Database have information about the distribution of each species?

 Yes / No

If 'yes', does the Database use TDWG standard for Countries level (Biodiversity Information Standards www.tdwg.org) or the WoRMS/VLIMAR system (<http://www.vliz.be/vmdc/datalim/vlimar/vlimar.php?p=details&id=1901>)

TDWG

If the Database contains distribution information in some other form, please, explain.

29. Establishment/Occurrence Status

Does the Database have information about the establishment status of each species (Native, Introduced, Naturalised, Invasive, etc.)?

 Yes / No

Questionnaire part 2:

4D4Life/Species2000

Taxonomic Databases Survey



30. Among following data fields, which one(s) would you approve to become additional field(s) in the Species 2000 Standard Dataset (<http://documents.ap2000.org/Documentation/standarddatasetv3.2dec2004.doc>)?

- | | |
|--|---|
| 30.1. Known habitat (whether the species is a marine, freshwater, estuary, terrestrial... one) | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.2. Maximum size of the organism | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.3. Reference to Nomenclatura (via LSIDs) | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.4. Host-relationships/Associated organisms | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.5. Fossils | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.6. Geographical distribution | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.7. Take taxon LSIDs directly from the GSDs (if LSIDs present in your database) | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.8. Conservation/legal status | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.9. Hybrids: names should appear in CoL | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.10. Hybrids: should be linked to both 'parent' taxa if present in any database | <input type="radio"/> Approval <input type="radio"/> Disapprove |
| 30.11. Infragenetic ranks | <input type="radio"/> Approval <input type="radio"/> Disapprove |

Feel free to express your wishes about additional data fields, which could improve Species 2000 Standard Dataset:

Questionnaire part 2:

4D4Life/Species2000

Taxonomic Databases Survey

**9. Responsible Officer (Senior Person, Principal Investigator):**

Title:

First Name:

Family Name:

Address:

Tel.:

Fax:

Email:

WWW:

10. Taxonomic Manager:

Title:

First Name:

Family Name:

Address:

Tel.:

Fax:

Email:

WWW:

11. System Manager:

Title:

First Name:

Family Name:

Address:

Tel.:

Fax:

Email:

WWW:

12. Sp2000 Membership:

Are you a member of Species 2000?

 Yes / No

If 'not' would you agree to become a member of
Species 2000? (visit http://www.sp2000.org/index.php?option=com_content&task=view&id=84&Itemid=49)

 Yes / No

Questionnaire part 2: 4D4Life/Species2000

Taxonomic Databases Survey



Species 2000 peer review

Please suggest names and contact e-mails of three taxonomists who could be invited as Species 2000 reviewers for your database.