

The Catalogue of Life: an integrative taxonomic backbone for organizing biological data

Peter H Schalk, ETI Bioinformatics, Netherlands

Frank Bisby, University of Reading, UK

Vanderlei Canhos, CRIA, Brazil

What's in a name?

To quote Robert May (1990):

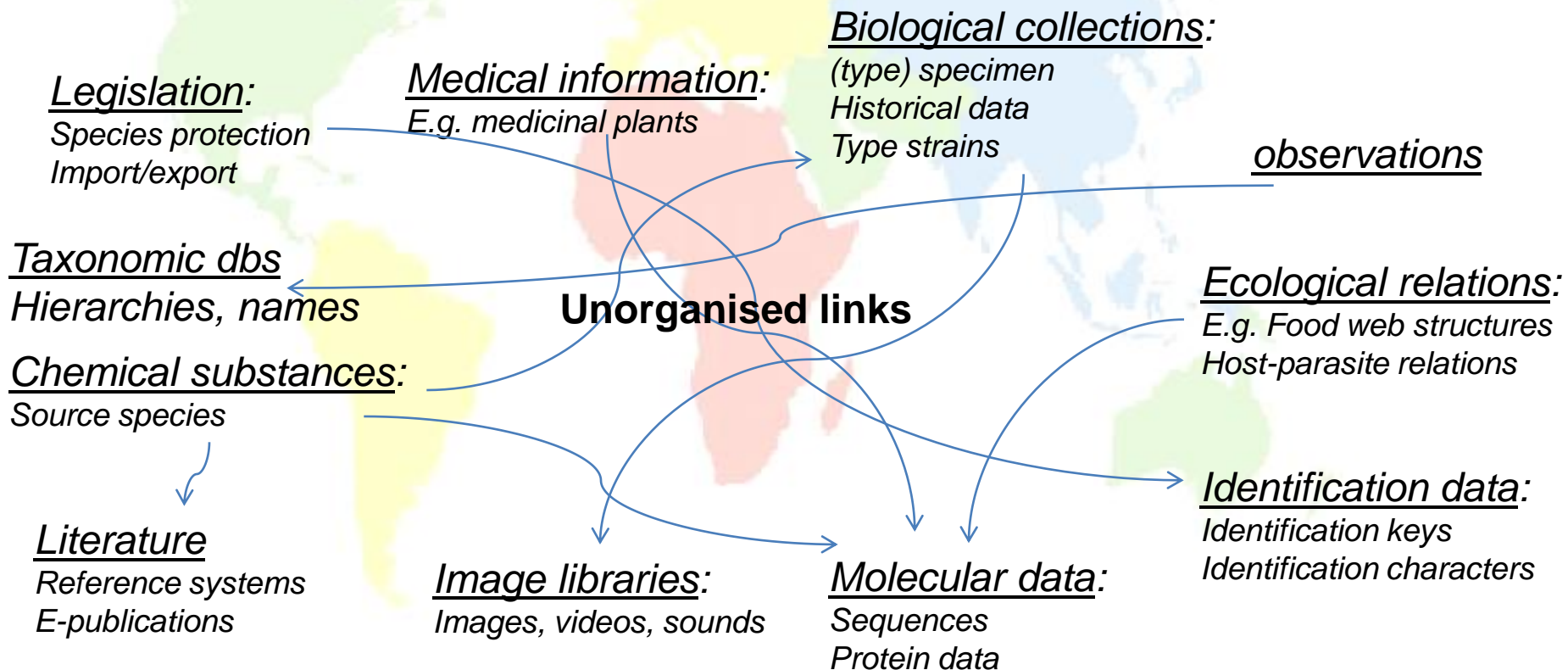
"Without taxonomy to shape the bricks, and systematics to put them together, the house of biological sciences would be a meaningless jungle"

Correct species names are increasingly important in society

Acetobactereae
Acetobacterium
Acetofilamentum
Acetogenium
Acetohalobium
Acetomicrobium
Acetonema
Acetothermus
Acholeplasma
Acholeplasmataceae
Acholeplasmatales
Achromatiaceae
Achromatium
Achromobacter
Acidaminobacter
Acidaminococcaceae
Acidaminococcus
Acidianus
Acidicaldus
Acidilobaceae
Acidilobales
Acidilobus
Acidimicrobiaceae
Acidimicrobiales
Acidimicrobiidae
Acidimicrobium
Acidiphilium
Acidiplasma
Acidisoma
Acidisphaera
Acidithiobacillaceae
Acidithiobacillales
Acidithiobacillus
Acidobacteria
Acidobacteriales
Acidobacterium
Acidocella
Acidomonas
Acidothermaceae
Acidothermus
Acidovorax
Acinetobacter
Acrocarpospora
Actibacter

Acinonyx jubatus
Acinonyx jubatus
Alces alces
Allactaga major
Alouatta seniculus
Alouatta seniculus
Alouatta seniculus
Alouatta seniculus
Alouatta seniculus
Aotus trivirgatus
Aotus trivirgatus
Aotus trivirgatus
Aotus trivirgatus
Arctictis binturong
Arctonyx collaris
Arundelconodon hottoni
Bathygenys reevesi
Bradypus variegatus
Cacajao calvus
Cacajao calvus
Cacajao calvus
Cacajao melanocephalus
Cacajao melanocephalus
Caenolestes fuliginosus
Callicebus moloch
Callicebus moloch
Callicebus moloch
Callicebus moloch
Callicebus torquatus
Callimico goeldii
Callimico goeldii
Callorhinus ursinus
Canis familiaris
Canis familiaris
Canis lupus
Canis simensis
Carollia perspicillata
Carollia perspicillata
Carollia perspicillata
Carollia perspicillata
Cebus apella
Cebus apella
Cebus apella
Cebus anella
Cebus anella

Hundred thousands databases and information systems have a taxonomic (name) component. Through the names data in different systems can be discovered.



Names and taxonomies change over time and tracing information in the jungle of information systems becomes hard or impossible.

Legislation:

Species protection
Import/export

Medical information:

E.g. medicinal plants

Biological collections:

(type) specimen
Historical data
Type strains

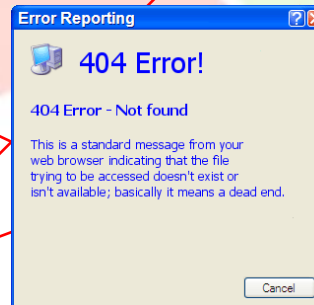
observations

Taxonomic dbs

Hierarchies, names

Chemical substances:

Source species



Ecological relations:

E.g. Food web structures
Host-parasite relations

Literature

Reference systems
E-publications

Image libraries:

Images, videos, sounds

Molecular data:

Sequences
Protein data

Identification data:

Identification keys
Identification characters

Until 1996 no single authoritative index of valid names, synonyms, common names was available. The taxonomic community created the Catalogue of Life.

Legislation:

Species protection
Import/export

Medical information:

E.g. medicinal plants

Biological collections:

(type) specimen
Historical data
Type strains

observations

Taxonomic dbs

Hierarchies, names

Chemical substances:

Source species

**CoL
NAME INDEX**

Ecological relations:

E.g. Food web structures
Host-parasite relations

Literature

Reference systems
E-publications

Image libraries:

Images, videos, sounds

Molecular data:

Sequences
Protein data

Identification data:

Identification keys
Identification characters

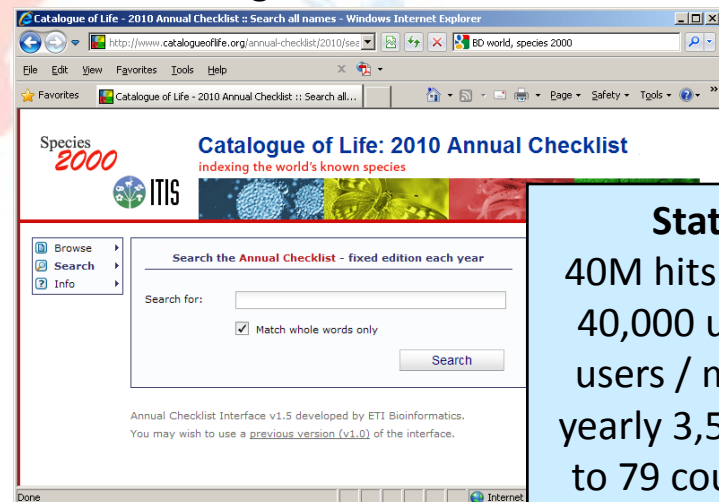
Species 2000, ITIS, and the Catalogue of Life

Species 2000 is a "federation" of database organisations working closely with users, taxonomists and sponsoring agencies. The Memorandum and Articles of Association date 11 December 1997.

Goal is to create a taxonomic community-driven validated checklist of all the world's species (plants, animals, fungi and microbes) by bringing together an array of global species databases covering each of the major groups of organisms: The Catalogue of Life.

In June 2001 a partnership was established with the Integrated Taxonomic Information system (ITIS) of North America.

The Catalogue of Life is based on a distributed model and used by a growing number of global, regional and national projects as taxonomic backbone service.



Stats:

40M hits / year;
40,000 unique
users / month;
yearly 3,500 CDs
to 79 countries

What is the Catalogue of Life?

A Resource...

- an electronic synonymic species checklist,
- a tightly integrated taxonomic hierarchy,
- intended for all 1.9 M extant known species.

....constructed by international networking

- both checklist and hierarchy constructed from sectors from many networked databases around the world
- and integrated using an international panel of experts

Catalogue of Life: Special 2010 Edition

CoL Annual Checklist 2010:

1,257,735 species

98,075 infraspecific taxa

886,882 synonyms

343,586 common names

Phyla: 111

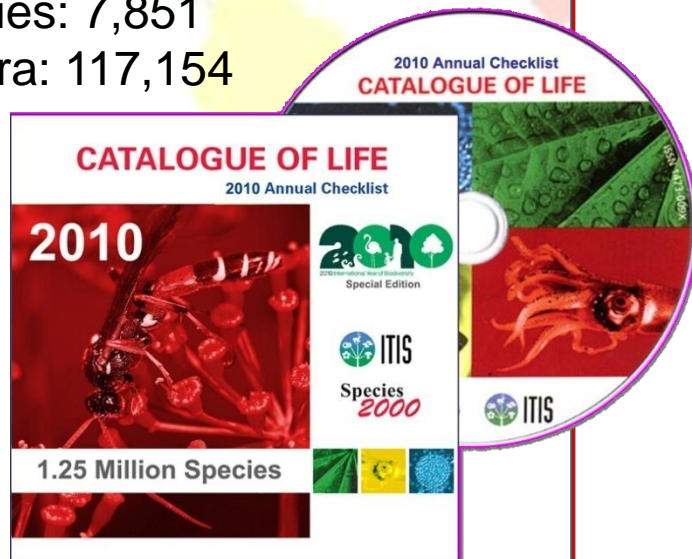
Classes: 285

Orders: 1,185

Superfamilies: 397

Families: 7,851

Genera: 117,154



Species 2000
Catalogue of Life: 2008 Annual Checklist
indexing the world's known species

ITIS

Species details

Blattodea
Species File

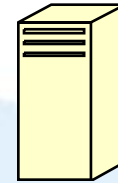
Browse Search Info

Accepted scientific name:	<i>Aeluropoda insignis</i> Butler 1882 (accepted name)
Synonyms:	<i>Aeluropoda gigantea</i> Butler 1882 (synonym) <i>Gromphadorhina hildebrandti</i> Dohrn 1887 (synonym)
Common name:	flat horned hissing cockroach English
Classification:	Animalia Phylum Arthropoda Class Insecta Order Blattodea Superfamily Blaberoidea Family Blaberidae Genus <i>Aeluropoda</i>
Distribution:	Madagascar (south-central)
Additional data:	-
Source database:	Blattodea Species File Online, 1.2/3.1, 30 July 2007
Latest taxonomic scrutiny:	Beccaloni GW, 30 July 2007
Online resource:	http://blattodea.speciesfile.org/...

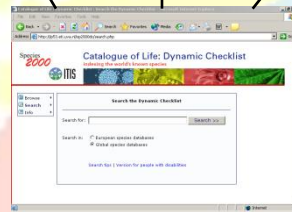
Simple data model; complicated data management!



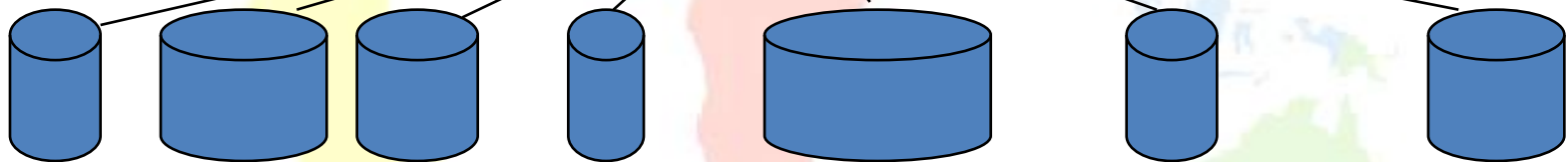
Services to users



Sp2000 Standard Data Set
absolutely minimal (7fields)



Integrated
Catalogue of Life



Array of source databases for different taxa

Species
2000



Integrated Taxonomic
Information System

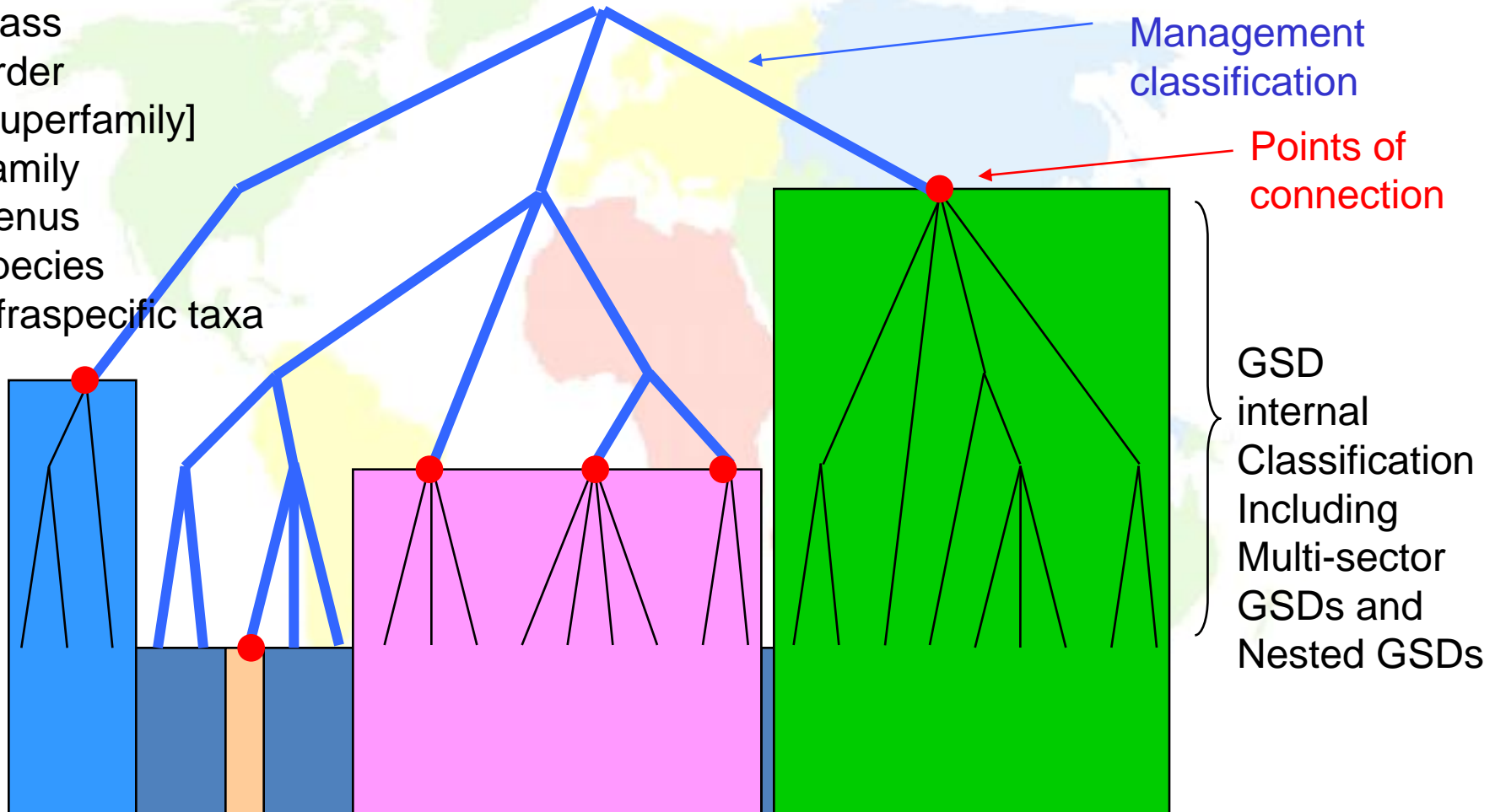
Global Species Databases for each group - GSDs

Catalogue of Life classification (> 1,000 sectors)




Kingdom
Phylum
Class
Order
[Superfamily]
Family
Genus
Species
Intraspecific taxa

CoL
Management
classification

Points of
connection





-  Browse ▶
-  Search ▶
-  Info ▶

Species details



Accepted scientific name: *Craterostigma tasmanianus* Pocock 1902 (accepted name)

Synonyms: -

Common names: -

Classification:		Animalia	LSID ▶
	Phylum	Arthropoda	LSID ▶
	Class	Chilopoda	LSID ▶
	Order	Craterostigmomorpha	LSID ▶
	Family	Craterostigmidae	LSID ▶
	Genus	<i>Craterostigma</i>	LSID ▶

Distribution: Australia (Tasmania); New Zealand

Additional data: -

Source database: [ChiloBase](#), 1.01, May 2006

Latest taxonomic scrutiny: Minelli A., 12-Apr-2006

Online resource: <http://chilobase.bio.unipd.it/sp2000.php?sp4141>

CoL taxon LSID: urn:lsid:catalogueoflife.org:taxon:dfcda7b2-29c1-102b-9a4a-00304854f820:ac2010

Browse taxonomic tree

- + Animalia
- + Archaea
- + Bacteria
- + Chromista
- + Fungi
- + Plantae
- + Protozoa
- + Viruses

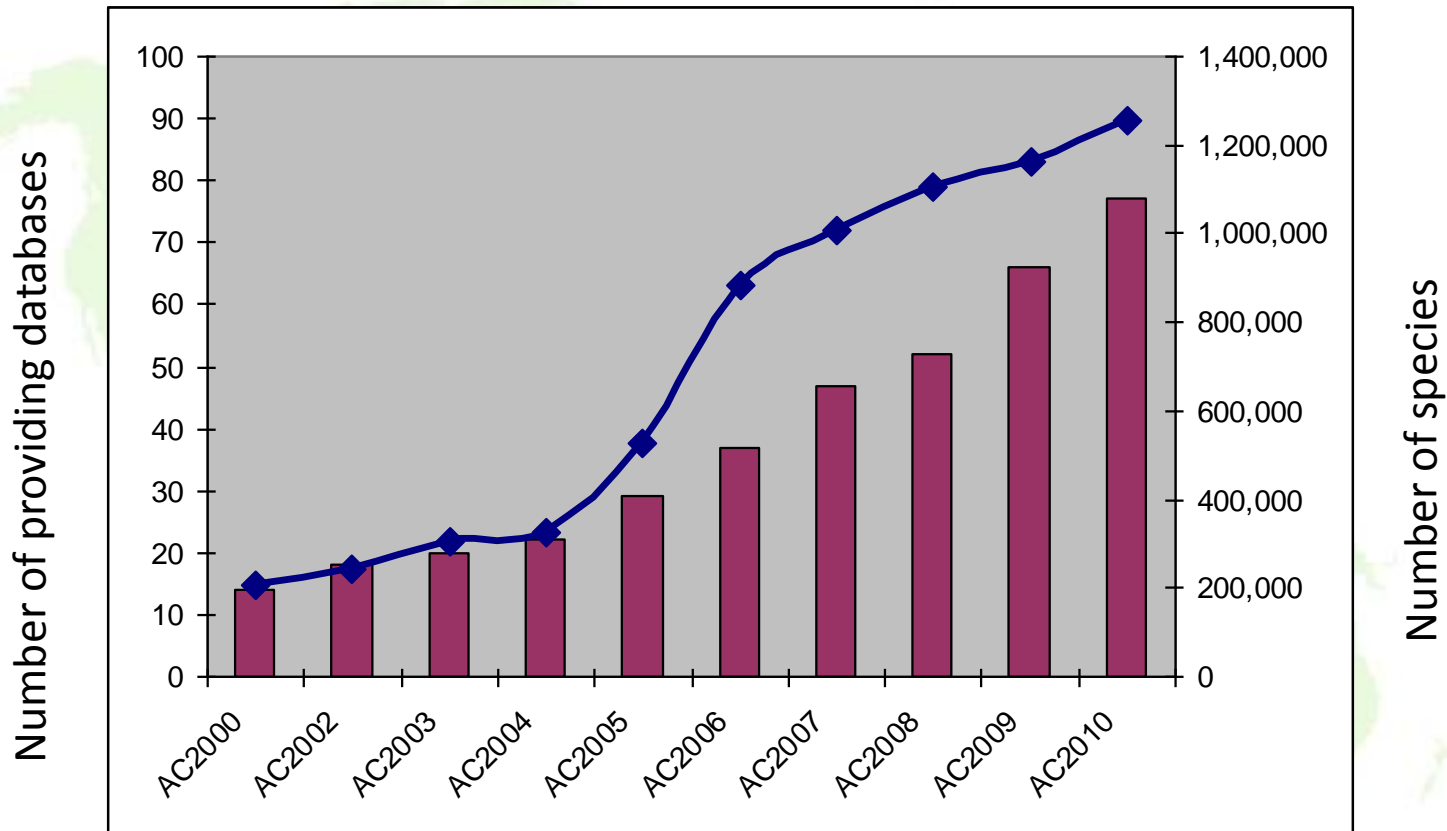
Browse taxonomic tree

- Animalia
 - + Phylum Acanthocephala
 - + Phylum Annelida
 - Phylum Arthropoda
 - + Class Arachnida
 - + Class Branchiopoda
 - + Class Cephalocarida
 - Class Chilopoda
 - + Order Craterostigmomorpha
 - Order Geophilomorpha
 - Family Aphilodontidae
 - + Genus *Aphilodon*
 - **Genus *Mecistauchenus***
 - Mecistauchenus micronyx*
 - + Genus *Mecophilus*

CoL Management
Classification

ChiloBase
Classification

CoL: a successful concept



www.catalogueoflife.org



Encyclopedia of Life

[login](#) | [create an account](#)

SHOW

Names Tags Full-text

Lathyrus pratensis L.

Kibana no Renri so

Species recognized by YR Roskov, FA Bisby, JL Zarucchi, BD Schrire & RJ White (eds), [ILDIS World Database of Legumes](#) in [Catalogue of Life](#)

IUCN RED LIST STATUS: **NOT EVALUATED**

Image

... free and open access to biodiversity data

GLOBAL BIODIVERSITY INFORMATION FACILITY

Search HOME SPECIES



Species: *Lathyrus pratensis* L.

Lesavy Lyadzei (Bel)

»Kingdom: [Plantae](#) »Phylum: [Spermatophyta](#) »Class: [Dicotyledones](#) »Order: [Lamiales](#) »Genus: [Fabaceae](#) »Genus: [Lathyrus](#) »Species

Actions for *Lathyrus pratensis*

- Explore:** [Occurrences](#) [Names and classification](#)
- List:** [Countries with occurrences](#) [Datasets with occurrences](#)
- Download:** [Darwin Core records](#) [One-degree cell density overlay for Google Earth](#) [Placemarks for Google Earth](#)

Names and classification

According to [Catalogue of Life: 2007 Annual Checklist: ILDIS World Database of Legumes](#)

Name [Lathyrus pratensis](#) L.

Classification »Kingdom: [Plantae](#) »Phylum: [Magnoliophyta](#) »Class: [Magnoliopsida](#) »Order: [Fabales](#) »Family: [Fabaceae](#)

Status [Accepted name](#)

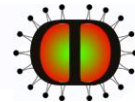


Australian Government
Department of the Environment and Heritage

AFD



WBD



Droseraceae Database

WoRMS
World Register of Marine Species
Brachiopoda & Phoronida

CCW

Catalogue of Craneflies



Blattodea
Species File



algaeBASE



BDWD

naturalis



FishBase



Aphid
Species File



CIPA

Bacteriology Insight
BIOS
Orienting System



Glomeromycota

Brassicaceae



GloBIS/GART



Plecoptera



SCARABS



ICTV

LHP



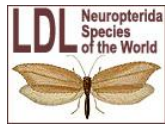
MOST



LepIndex



Species File



LDL
Neuropterida
Species of the World



LecyPages



Mantodea
Species File



OSF
ONLINE



RJB Geranium
CSIC



Zoological Institute PAS
PARHOST
World Database of Parasites

Psyl'list



Phyllachorales

Rotifera Database

eumycetozoa.com



NZOR
New Zealand Organisms Register



Rhytismatales



ScaleNet



Solanaceae
Source

SpidCat



SysMyr



Species
Fungorum



NATURAL
HISTORY
MUSEUM

Tineidae



ICTD

TicksBase



TIGR
Reptiles

Tetramycetes

Africa
TERVUREN



AnnonBase



Kew
PLANTS PEOPLE
POSSIBILITIES

WoRMS
World Register of Marine Species
Cumacea



TITAN

World Register of Marine Organisms



Xylariaceae

WoRMS
World Register of Marine Species
Ophiuroidea

WoRMS
World Register of Marine Species
Porifera



Zygomycetes

Oberösterreichisches

Landes
Museum
Biologiezentrum



NATURAL
HISTORY
MUSEUM

UCD



ETI
Bioinformatics

WoRMS
World Register of Marine Species
Proseriata
Kalyptorhynchia



World Information
Network on Weebels

WTaxa

WoRMS
World Register of Marine Species
Porifera

Governance

members

Board

Secr

CoL Team

TaxGr

ISGr

4D4Life

i4Life

various

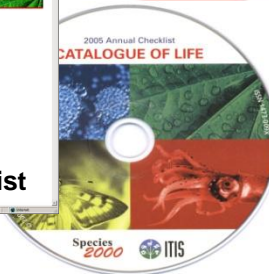
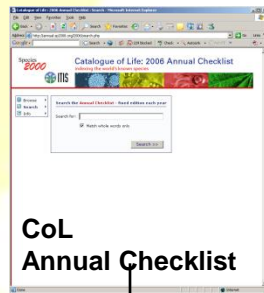
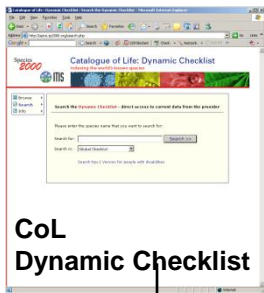
ownership

organization

action programs

main output or products
(realized in partnership with ITIS)

specific services



Open Membership

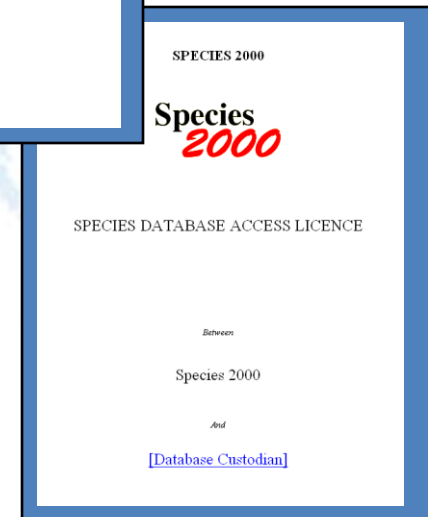
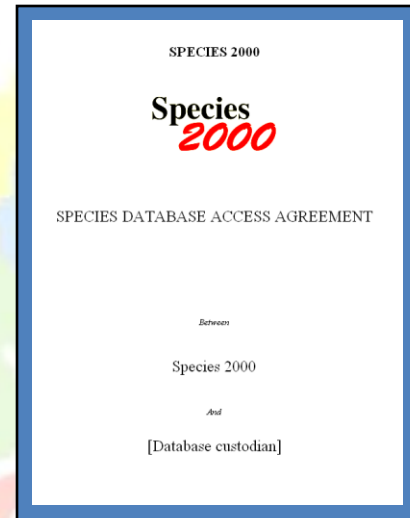
Species 2000 is interested in hearing from any individual or organisation which has a database which (or is intended) to cover the world's species within one particular group – *a global species database*.

Membership of Species 2000 is **open** to any individual, project or institution.

Please contact the Secretariat:

sp2000@sp2000.org.

- Currently 42 members
- There are now 77 participating databases
- Potential is well over 100 databases and partners
- Aim is to increase number of members and databases



Organization units

Organization units

- CoL Project Team (science policy, think tank)
 - Taxonomy Group (Team members + advisors)
 - Information Technology Group (Team members + advisors)
 - User Forum (through projects)
- Secretariat (Reading) & Executive Director
 - Supporting activities, administrative tasks, CoL production
- Board of Directors
 - legal framework, connection to international efforts, politics

Open management structure

- Board of Directors (5-8 directors). Term 4 years, renewable for 4 years. Gradual replacement scheme. Continuous outlook for potential candidates. BoD selected, AGM appointed.
- The CoL Team is continuously 'rejuvenated' (12-18 members). Term is 2 years, with a 2 years extension for active members. BoD appointed (Team suggests).
- Secretariat (hosting agreement). Subject to bidding procedure. Agreement (minimum 5 year) up for renewal: November 2012
- Executive Director. Term 5 years, renewable for 5 years. BoD appointed. End of term: November 2013

Note: Hosting of Secretariat and appointment of ED are independent!

Development Through (Funded) Action Programs

- Early history: 1998-2003 small funds from multiple sources (IUCN, UNEP, CODATA, ITIS, ETI, RIKEN, BIOSIS, a.o.)
 - Anchoring in taxonomic community; international network
 - Stable organization structure: Team, Secretariat, Board
 - Simple infrastructure and split approach to AC/DC products
- From 2004-2007 large projects / funding: EC EUROCAT (Species 2000 Europa), EC ENBI, GBIF (CoL Partnership) resulted in big steps forwards
 - Corporate identity (logo's, banner, user interface, website)
 - Strengthened legal basis: membership agreements, IPR agreements
 - Enhanced CoL product development; extended content
 - Basis for Regional en Global Hub structures
- From 2008-2009: under own steam in minimum staff configuration
 - Contributions from EoL; financial support from ETI, ITIS, FISHBASE a.o.
 - Critical mass achieved (over 1 million species) attracting non-academic users
- From 2009 – 2013: Significant EC FW7 project funding (towards completion)
 - Novel e-infrastructure (**4D4Life**) and virtual research community (**i4Life**)
 - Targeted services for non-academic users (supporting running costs)

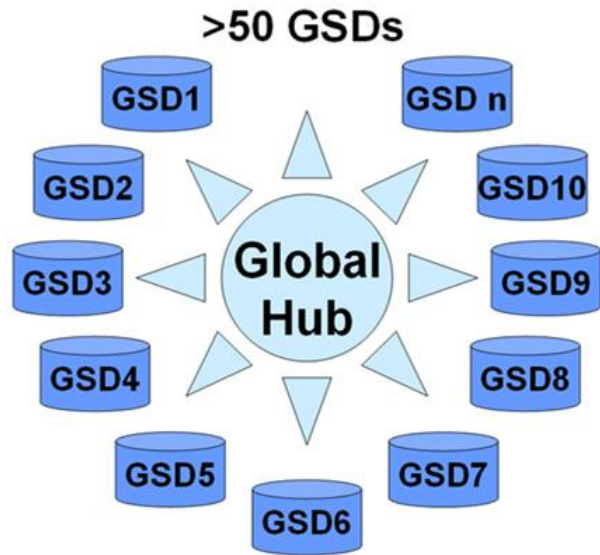
The EC funded 4D4Life Project (2009 – 2012)

3.3 M€ builds a state-of-the art e-infrastructure that:

- modernized information management tools
- partly automated tasks; more efficient CoL production
- structured information exchange in the project networks
- synthesises a globally significant resource for science
- dissemination in an array of web-services and products
- multi-hub structure for regional deployment

New e-infrastructure: an ecosystem of services

Architecture 1

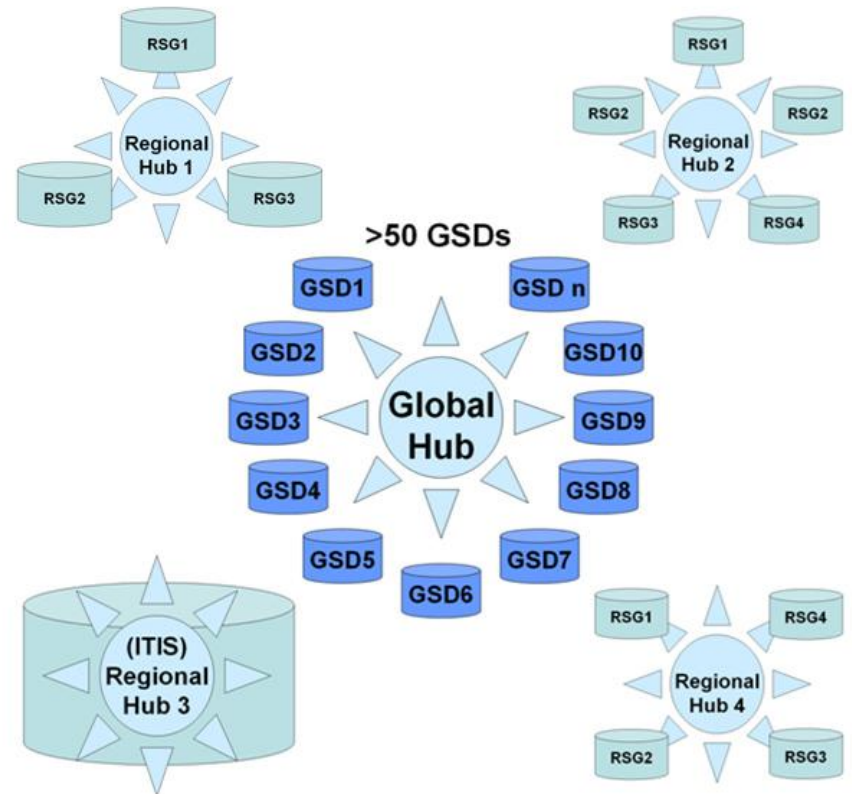


GSDs for each group needed

4D4Life



Architecture 2



GSDs plus infill and harvest additional information from regional systems

World-wide Multi-Hub Network, with Regional Hubs

1. Species 2000 China Node (BioD. Com. CAS)

Keping Ma & Liqiang Ji

2. Australian Hub (ABRS with ALA/CSIRO)

Cameron Slatyer & Donald Hobern

3. New Zealand Hub (NZOR)

Jerry Cooper

4. Catalogo da Vida Brasil (CRIA & partners)

Vanderlei Canhos

5. ITIS N. America (Smithsonian NMNH)

Tom Orrell

6. Sp2000 Euro-Hub (PESI/ Pan European Species List)

Thierry Bourgoïn & Yde de Jong

4D4Life

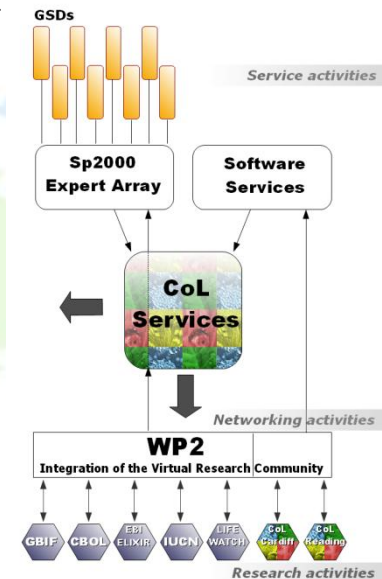
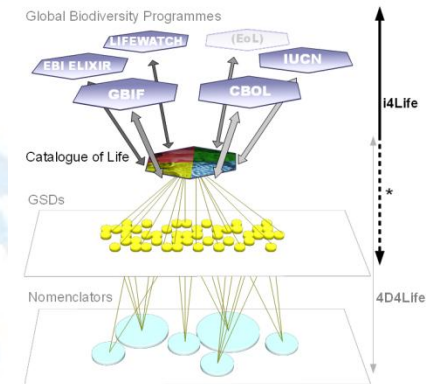


i4Life: Indexing for Life

The i4Life project (2010 – 2013) : 2.3 M€

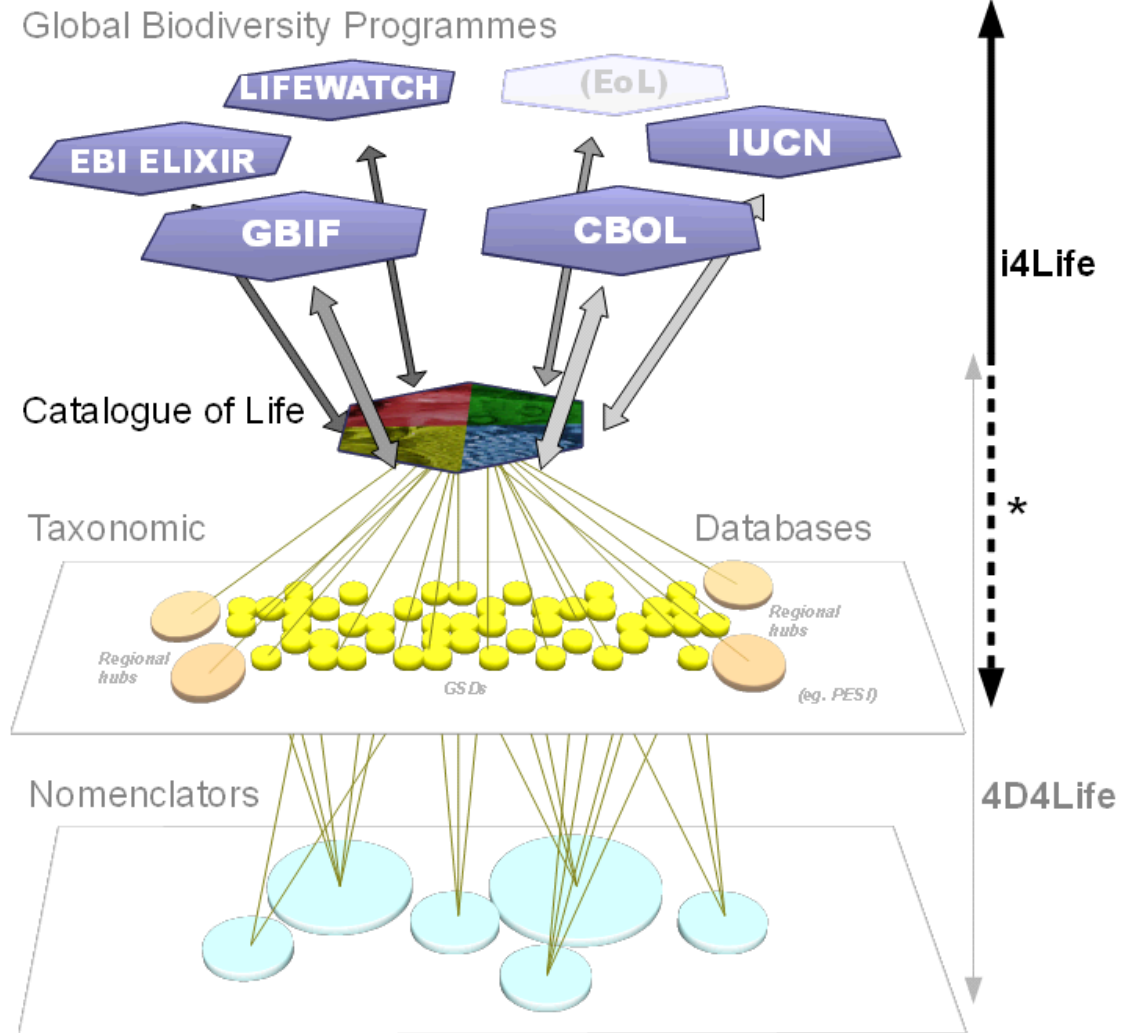
Will establish a Virtual Research Community to integrate the indexing of biological species between the world's major biodiversity programmes.

- It will use the existing Catalogue of Life as a yardstick with which to compare and unify what is known.
- It will for the first time provide a summary of all species known on earth.
- It will create a global standard for taxonomic data integration in electronic infrastructures world-wide.



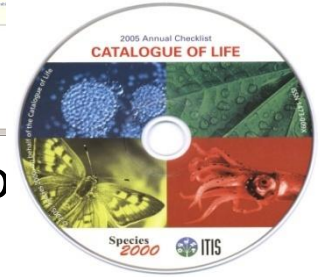
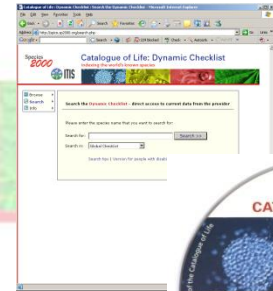
i4Life: integrated approach to completing the CoL

The i4Life component provides an index to the large biodiversity projects and generates feedback data (names) to the underlying databases. Aim: completion!



i4Life

Sustainability



How can we make such services sustainable? Many information systems set up with public funds collapse when funding stops.

The CoL generation costs amounted to 11 million Euros over 10 years (network, infrastructure, innovation, product development) **not** including costs at data providers side!

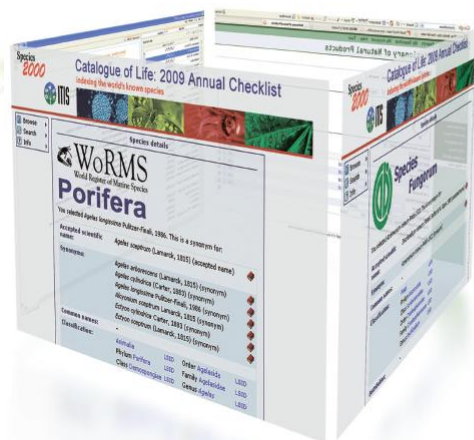
CoL operation (sustainable) running costs are 450 kEuro per annum (in part carried by host institute).

The CoL now has sufficient critical mass. It is a 'mature' product aimed at professional market. It is believed that the CoL production facility can be made sustainable.

Large users in the community and non-academic users will contribute to running costs. Surpluses flow back to data providing databases.

Deployment outside the academic realm

What is the Catalogue of Life?

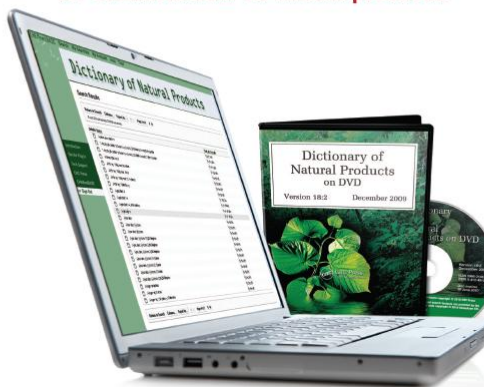


Catalogue of Life (CoL) is the ultimate global catalogue. Used by the Global Biodiversity Facility (GBIF) and the UN Convention on Biological Diversity (CBD), the CoL documents more than 1 million species, more than half of the known species on earth. It contains three elements:

- ❖ A synonymic catalogue of the scientific names of species
- ❖ A small set of data about each species (such as synonyms, common names, distribution, and literature citations)
- ❖ A taxonomic tree and classification depicting relationships between the groups of species

The Web version is now PC and MAC compatible
The DVD version is now compatible with MAC and Windows Vista

Save hours in the library with the most comprehensive source available for chemical data on natural products.



Annual Subscriptions Ensure Current Data

CRC Press continually works to improve and expand its databases and make this the most useful and useable resource for all fields of chemistry. Your annual subscription to *Dictionary of Natural Products* ensures that you always have the latest updates – published every six months – with new compounds and the most recent enhancements to its functions and interface.

Special pricing packages are available for all DVD and Web products. Please call for more details. Ask about our:

- ❖ Multi-year contract for site licenses
- ❖ Licensing for products on client servers for corporate use, including periodic updates
- ❖ Special academic consortia pricing
- ❖ Online subscriptions

In the Americas
Tel: 1-888-318-2367
(Outside Continental U.S.) 1-561-998-2505
e-mail: e-reference@taylorandfrancis.com

Rest of the World
Tel: +44 (0) 20 7017 6062
e-mail: online.sales@tandf.co.uk

Continuously Updated!
10,000 new compounds added each year

Search Results

Version 18:2 December 2009

“If you are interested in natural products, you will want access to this dictionary.”
—Natural Product Reports

CRC Press
Taylor & Francis Group

Now available in two convenient formats

Sustainability: Financial support for CoL running costs through custom products & services



More information?

For more information on Species 2000, the Catalogue of Life please visit the website: www.catalogueoflife.org or www.species2000.org

Individual academic use of the CoL online service is free of charge.

If you prefer to receive the Annual Checklist on (free) DVD-ROM please email the Species 2000 Secretariat: sp2000@sp2000.org

For joining the CoL Team or queries on submitting databases feel free to contact: sp2000@sp2000.org

Questions about the governance or joining Species 2000 as a member: pschalk@eti.uva.nl (Chairman Board of Directors)



WFCC

WORLD FEDERATION FOR CULTURE COLLECTIONS

ICCC-12 Conference 2010

Biological Resource Centers: gateway to biodiversity and services for innovation in biotechnology

Thank you for your attention!