



The Catalogue of Life: Infrastructure for Science

Thomas Orrell, Ph.D., Smithsonian Institution, National Museum of Natural History







First, a little more about ITIS...







indexing the world's known species





Create an easily accessible database with reliable information on species names and their hierarchical classification







What is ITIS?

Partnership

- federal partnership with state, private, academic, international cooperators
- taxonomic experts, compilers
- stable base funding



indexing the world's known species





ITIS Partners





















Agriculture and Agri-Food Canada Agriculture et Agroalimentaire Canada







indexing the world's known species



ITIS Key Components

- Standard taxonomic reference tool
 - dynamic standardized database of biological names & related data (over 840,000 names)
 - accepted scientific names (single "best" name) in a classification, with synonyms, and common names
 - on-line lookup, downloads
 - seven kingdom coverage
 - regional and global checklists
 - Monthly updates, full database download, services



indexing the world's known species



ITIS Key Components

- Taxonomic Serial Number (TSN) unique for each scientific name
 - facilitates machine to machine the TSN is a non-intelligent, unique and stable (more on this later!)
- Codes of Nomenclature dictate ITIS business rules
- rigorous proofing protocols and verification
- bibliographic/expert link
 - steward program insures scientific credibility
- working tools Work Bench, TaxCompare, queries



indexing the world's known species



ITIS Value to User Community

- Whom do we serve?
 - national and international
 - resource managers, scientists, biodiversity specialists, museum collections staff, librarians, park naturalists, students, editors...
 - bridge gap between systematics community and users NSF cooperation
 - the public = access to our systematic opinions
 - GGI, EOL, BHL, CBOL, NCBI, GBIF....







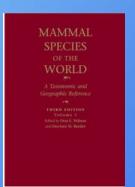
ITIS Data Sources

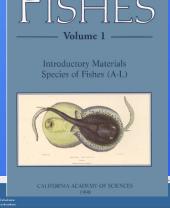












ICHTHYOLO			1998
	nstated Checklists Primary Types Imagebase Collection Staff Newsletter of Systematic Inthibution Funding Obj		
California Academy of Sci	nces > Research > Ichthyology > Catalog of Fishes		
Search the On-line Catalog		rsion, Updated April 1 ellection Abbreviations Br About the Print Version	over the Classification Glossary Instructions on Searching for Species Instruction
		THE	
	Cata	log of Fish	ies
		ON-LINE	
Select the database t	to search:		
0	PECIES © REPERENCES		
O GENERA ® S	PECIES O REFERENCES		
Search Res	Comments: William Eschmeyer		
55001	200		
Search Instructions			
A search expression or	en be specified as		
A single search	term		
9 a word	(multiple words) enclosed in double quotes - l'Dordan & Evern	menn*1	
9 a phrase		an asterisk (*) - [*Hawa	**]
o a partial	word or phrase enclosed in double quotes (*) and ending with		
• Two (or more)	search terms joined by one of three logical operators: { and	er and not)	
o a partial Two (or more): (terms can be g	search terms joined by one of three logical operators: { and rouped by using parentheses - see examples below).		
o a partial Two (or more) (terms can be g	search terms joined by one of three logical operators: { and rouped by using parentheses - see examples below), rch expressions and the number of hits when that query is issu	sed against the indicated	
o a partial Two (or more): (terms can be g Some examples of sea Example of	inarch terms joined by one of three logical operators: (and rouped by using parentheses - see examples below), inch expressions and the number of hits when that query is issues search Expression.	ued against the indicated Database	Hits
a partial Twa (or more) (terms can be p Some examples of sea Example of single word	search terms joined by one of three logical sporesters: (and I industry to using parentheses - see examples below), and expressions and the number of hits when that query is issumed to the pression of the pression.	ued against the indicated Database Genera	75
o a partial Two (or more): (terms can be g Some examples of sea Example of	search terms jained by one of three logical operators: (and I recognity of by the grantheses - see samples below), and suppressions and the number of hits when that query is issues Search Expression Salma "safe as Albula vulpes"	ped against the indicated Database Genera Species	75 1
a partial Twa (or more) (terms can be p Some examples of sea Example of single word	search terms joined by one of three logical sporesters: (and I industry to using parentheses - see examples below), and expressions and the number of hits when that query is issumed to the pression of the pression.	ped against the indicated Database Genera Species Species	75 1 21
• a partial • Two (or more) : (terms can be jo Some examples of sea Example of single word phrase	earnh terms jained by one of three logical contents; if and increasing a logical contents; if and increasing a logical contents are seen to the substitution of this when that query is issue Search Expression Salmo "sail os al Albula vulpe;" "sunonym of albula vulpes" "contents is"	ped against the indicated Database Genera Species Species Species	75 1 21 21 31
a partial Twa (or more) (terms can be p Some examples of sea Example of single word	search terms jained by one of three logical operators: ("and) rouged by using searchlasses - see counties below, when the consistence of the number of his when that query is issomether. Expressions and the number of his when that query is issomether. Expressions Search Expressions. Search Expression and the number of his when that query is issomether. Search Expression and the number of his when the number of his when the number of his when the number of his whole with the number of his number of hi	ped against the indicated Database Genera Species Species	75 1 21



indexing the world's known species





Solr Services



Home

About ITIS

Data Access and Tools

Get ITIS Data

Submit and Update Data

ITIS Solr Web Services

The base URL for the ITIS Solr index based web service is: http://services.itis.gov/?q=*:* where the initial que be a field in Solr and a value for it separated by a colon, such as: http://services.itis.gov/?q=tsn:182662

ITIS is currently using Apache Solr Vers. 5.4. Output from ITIS Solr is available in XML (default), CSV, JSON, JSO For example, add "&wt=ison" to the end of the query string to get JSON output returned like this http://services wt=json. The Apache Solr Reference Guide can be quite helpful in generating queries and using their output. also be helpful in elucidating the finer points of field definitions and validation rules.

We've included some ITIS Solr Web Service examples here.

Tips for Using the ITIS Solr Index

Regular Expressions. ITIS Solr supports regular expressions. For example:

- http://services.itis.gov/?g=nameWOInd:/[A-Za-z0-9]*[%20]{0.0}*/ matches only monomials
- http://services.itis.gov/?g=nameWOInd:/[A-Za-z0-9]*[%20]{1.1}[A-Za-z0-9]*/ matches only binomials
- http://services.itis.qov/?q=nameWOInd:/[A-Za-z0-9]*[%20]{1,1}[A-Za-z0-9]*[%20]{0,1}[A-Za-z0-9]*/- mal trinomials

Facets and Pivot Facets. ITIS Solr supports simple and complex facets. For example, to get a list of the numb species by kingdom use: http://services.itis.gov/?q=rank;Species&rows=0&facet=true&facet.field=kingdom If Solr will return a sample of full records along with the facet. You can also do complex facets called pivot facet full list of all valid species of birds with their common names use: http://services.itis.gov/?q=hierarchySoFar.1 rank:Species AND usage:valid&facet=true&facet.pivot=nameWInd.vernacular&facet.limit=-1&facet.mincount= facets are very computationally intensive, and therefore very slow, so please be considerate and conservative

Using JSONP. ITIS Solr supports JSONP. In addition to supporting JSON output using "wt=json" alone, ITIS S padded/wrapped JSON inside of any variable name that you choose (up to 30 characters) using "wt=json& ison.wrf=whateverNameYouWant". This means that you can call data from ITIS and display it directly in your v without having to manipulate it on the server, and without violating the cross-origin policy for javascript. Pleas for examples.

Fuzzy Matching. ITIS Solr supports customizable fuzzy matching. Use a tilde along with a value between 0 ar - <arr name="expert"> matched to indicate the level of fuzziness. For example: http://services.itis.gov/?g=nameWOInd;Liquidamber\%20stvraciflua~0.4 will match Liquidambar styraciflua.

Escaping spaces and percent encoding in query strings. Because you cannot always be sure how a browser will interpret strings with spaces, it is good practice to both escape and percent encode spaces in queries for strings that have spaces. For example:

http://services.itis.gov/?g=nameWOInd:Poa\%20annua.

```
<response>
- < Ist name="responseHeader">
    <int name="status">0</int>
    <int name="QTime">0</int>
  - < Ist name="params">
      <str name="q">tsn:172419</str>
    </lst>
  </lst>
- <result name="response" numFound="1" start="0">
      <str name="tsn">172419</str>
      <str name="nameWind">Thunnus alalunga</str>
      <str name="nameWOInd">Thunnus alalunga</str>
      <str name="unit1">Thunnus</str>
      <str name="unit2">alalunga</str>
      <str name="usage">valid</str>
      <str name="credibilityRating">TWG standards met</str>
      <str name="taxonAuthor">(Bonnaterre, 1788)</str>
      <str name="kingdom">Animalia</str>
      <str name="parentTSN">613007</str>
      <str name="rankID">220</str>
      <str name="rank">Species</str>
    - <arr name="hierarchySoFar">
          172419:$Animalia$Bilateria$Deuterostomia$Chordata$Vertebrata$Gnathostomata$Osteichthyes
          (Thunnus)$Thunnus alalunga$
      </arr>
    - <arr name="hierarchySoFarWRanks">
          172419:$Kingdom:Animalia$Subkingdom:Bilateria$Infrakingdom:Deuterostomia$Phylum:Chorda
          (Thunnus)$Species:Thunnus alalunga$
        </str>
      </arr>
    - <arr name="hierarchyTSN">
          $202423$914154$914156$158852$331030$914179$161030$161061$553120$161105$16608
        </str>
      </arr>
    - <arr name="synonyms">
        <str>172419:$Thunnus germo$Scomber alalunga$</str>
```

http://www.itis.gov/

- <arr name="synonymTSNs">

</arr>

<str>172419:\$172420\$202024\$</str>

Field Descriptions



indexing the world's known species



A combined effort

Species 2000 and ITIS Catalogue of Life Partnership







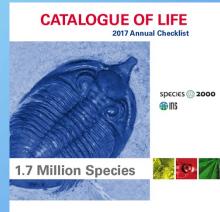
indexing the world's known species



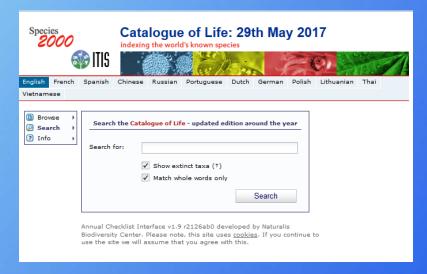


Species 2000 and ITIS Catalogue of Life

Annual CD-ROM



Dynamic Checklist ww.catalogueoflife.org/





indexing the world's known species





The Catalogue of Life is an expert (taxonomist) knowledge-base composed of global species datasets (GSDs) that cover most of described life.







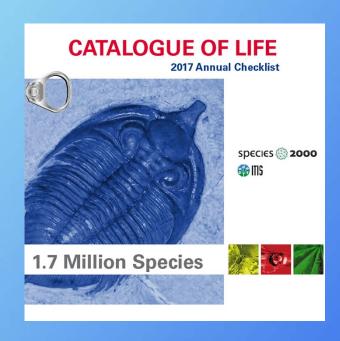






indexing the world's known species





What is the relationship between the Objective and Subjective ? What is the relationship between Nomenclature and Taxonomy?





indexing the world's known species





Nomen

Original usages/combinations

Subsequent combinations (homotypic synonyms)

Objective (code-governed) heterotypic synonyms

Spelling variations

Links to literature

Objective: facts based on rules (taxonomic codes) and evidence





indexing the world's known species









Catalogs of taxa

Subjective assertions of heterotypic synonyms

Higher (above genus) classifications

Phylogeny

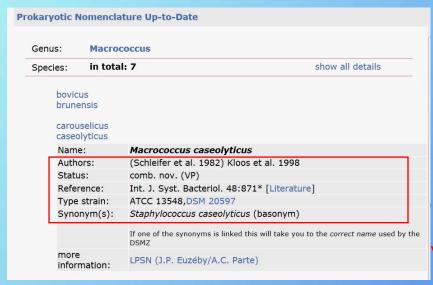
Subjective: opinions based on evidence









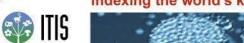


Names strings to nomenclature

Nomenclature











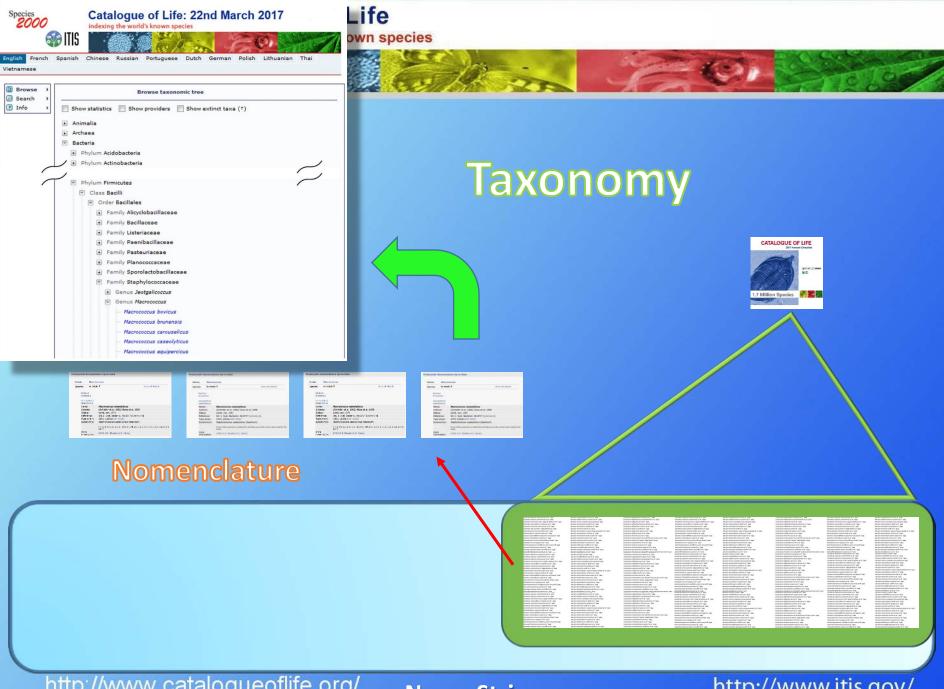


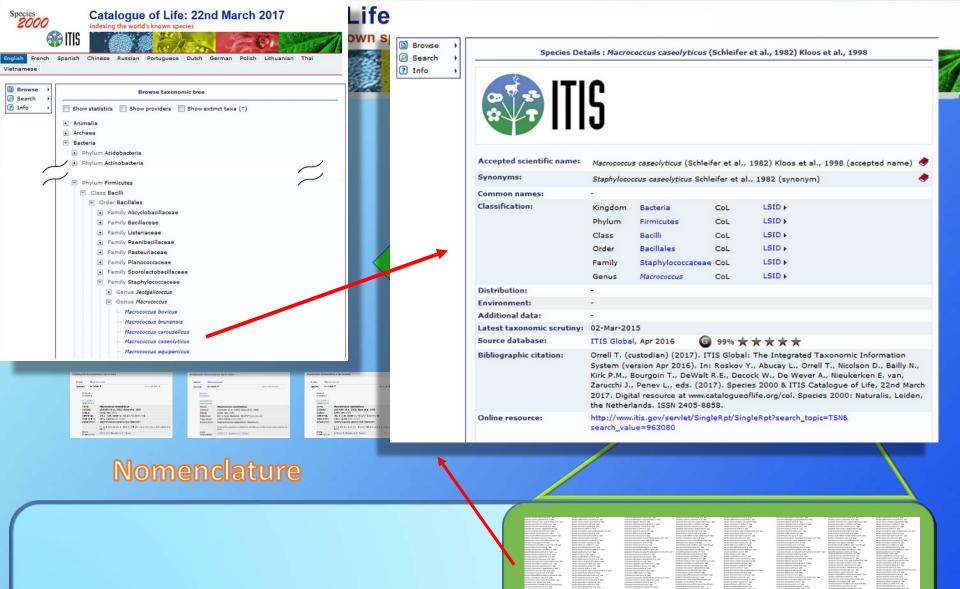




Nomenclature









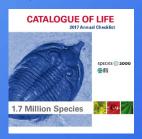
indexing the world's known species



Hey CoL, is this the correct name for this taxon?



Sorry GBIF, I've never heard of that taxon name.



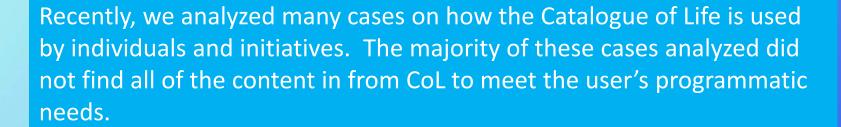
Why a Catalogue of Life Plus?





indexing the world's known specie





Why? Although the CoL is the most comprehensive compellation of Life, there remains:

- 1) Known GAPS of Taxonomic Groups
- 2) Missing content in currently covered Taxonomic Groups
 - Synonyms
 - Subsequent combinations
 - Orthographic variants



indexing the world's known specie





Key Features:

- Broaden names coverage through homotypic synonyms & expanded editorial human resources
- Enable stable names and taxon identifiers
- Annotations & feedback

Increase the efficiency of those providing GSDs by:

- Expanding the curation of facts
- Making fact checking work of the GSDs more efficient
 - Allowing taxonomists to focus on taxonomy
- Expanding the utility of their work





Plants (IPNI)

Fungi (Index Fungorum) Zoology (ICZN/

GNUB)

Examples of other sources of nomenclatural data

> Compilation (GBIF GUB)

Bacteria

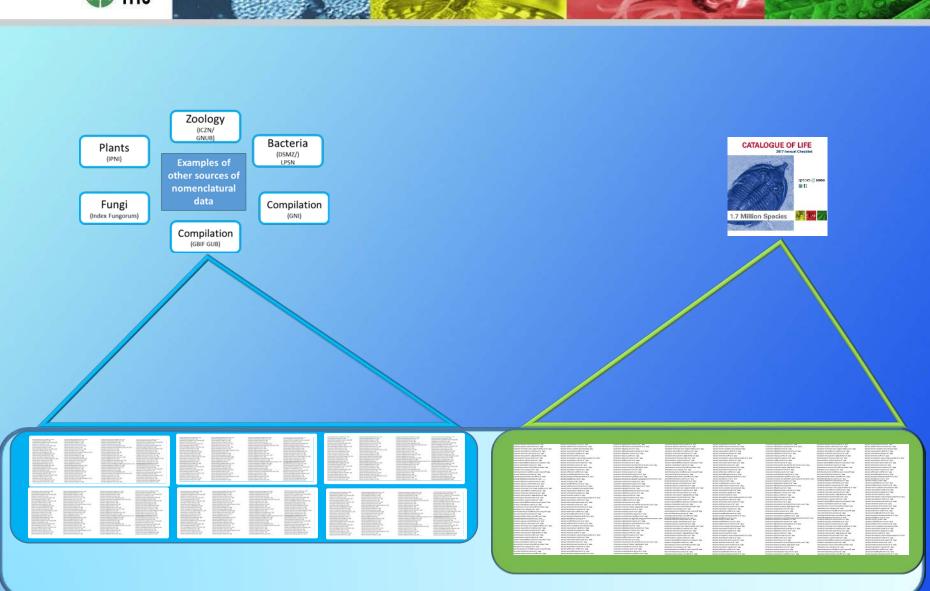
(DSMZ/) **LPSN**

Compilation (GNI)







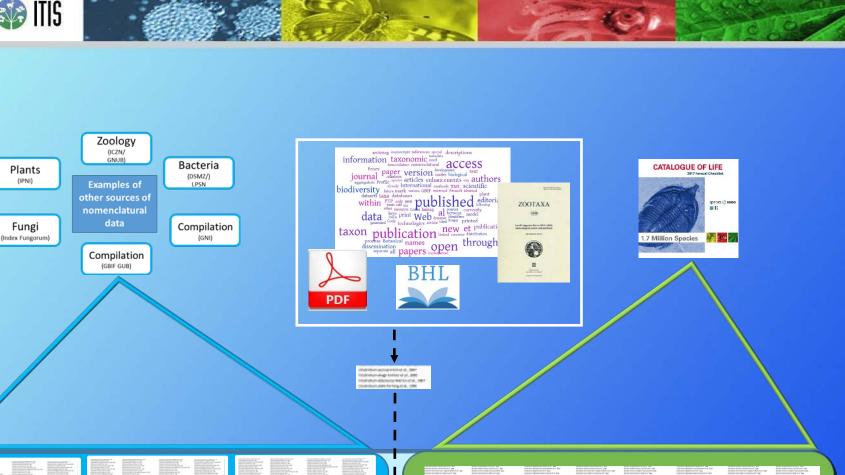




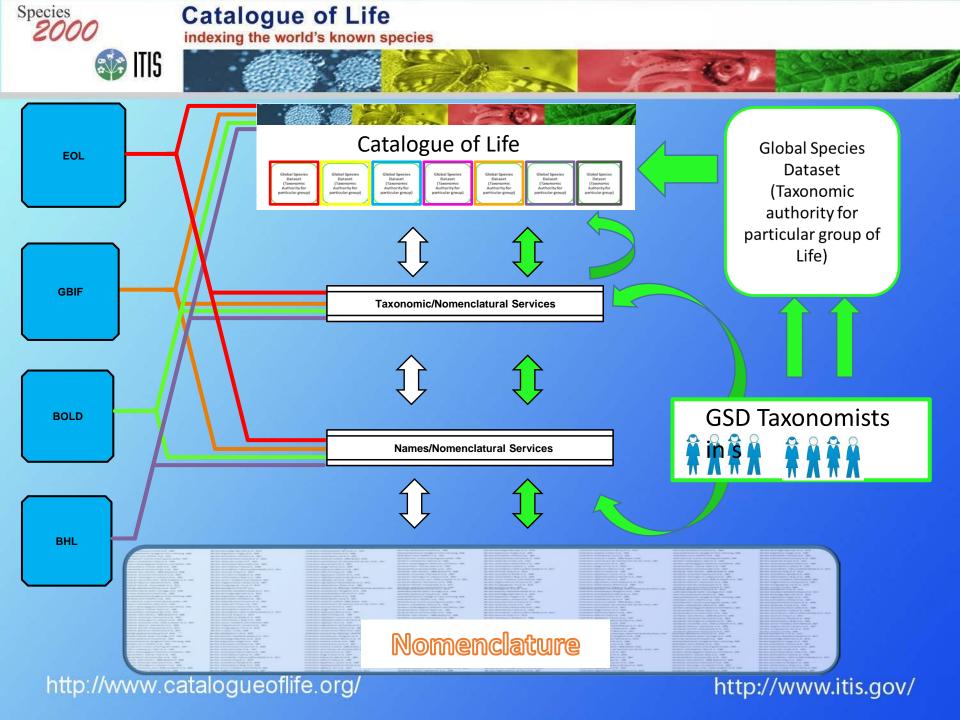


Plants

Fungi





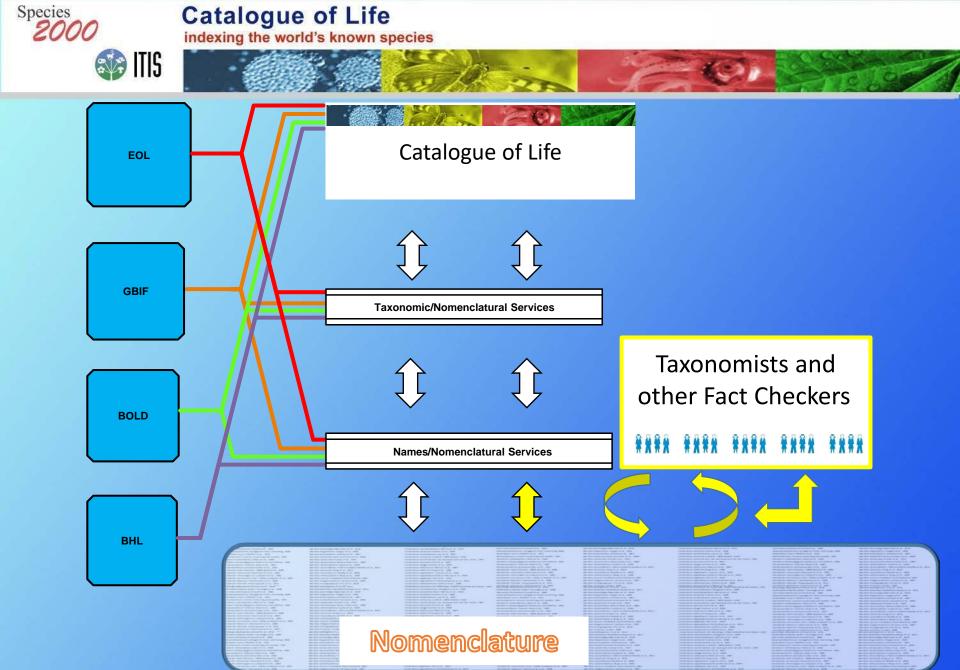


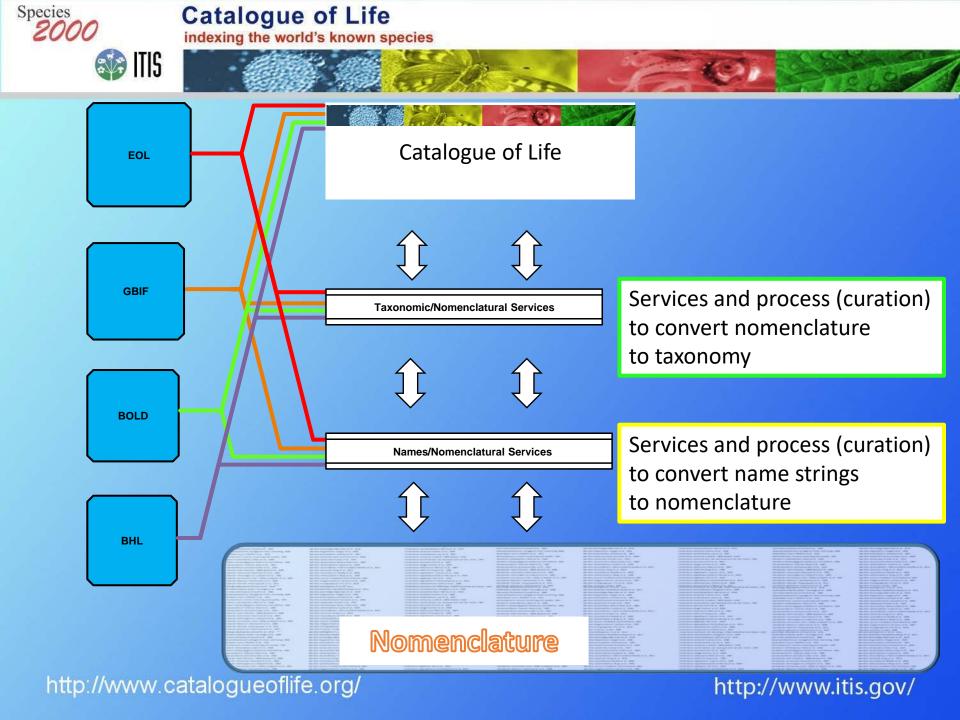
Species 2000 Catalogue of Life indexing the world's known species SITIS Catalogue of Life **EOL GSD Taxonomists GBIF Taxonomic/Nomenclatural Services BOLD** Names/Nomenclatural Services **BHL**

Nomenclature

http://www.catalogueoflife.org/

http://www.itis.gov/







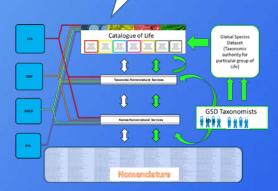
indexing the world's known species



Hey CoL, is this correct name for this taxon?



No, that is not the current name for that taxon, please use...



Why a Catalogue of Life Plus?



Why a Catalogue of Life Plus?

Without taxonomy, it isn't biology

A robust (more complete), Catalogue of Life is infrastructure for science because:

- It increases precision of searches and organization of information
- It increases the exchange of knowledge
- It is a link between the past and the future