

Species Index - A URI for every Taxon

This service allows you to create an HTTP URI for any named biological taxon. It works very simply.

- You follow a **simple set of rules** to create an [HTTP URI](#) (also known as a URL) based on the name of the organism.
- You use the resulting URI to mark up your data - particularly if you are using RDF.
- Anyone else who follows the same process will create the same URI for their data.
- Machines will know that the two sets of data are about the same taxon because they use the same URI. There is no more confusion around the correct way to cite a name.
- If any person or machine calls the URI (puts it in a web browser for example) the website will respond by saying the equivalent of "This URI represents a taxon with the name ... " and it will use the **simple rules** backwards to work out what the name of the organism is.
- If the name of the organism is totally unambiguous this site may add some extra **nomenclatural** data from other sources (such as Index Fungorum, IPNI and ZooBank) that isn't encoded in the URI.
- This site will **never assert anything about the taxonomy or classification of the taxon**. This is the 'unique selling point'. You can express synonymy or placement in a classification in your data and your colleagues can express something different in their data but

you can have the common ground of a shared URI.

Speciesindex.org's principle aim is to encourage people to use the same URI when they are talking about something in common.

Speciesindex.org is hardly original. People have been embedding taxon names in URLs since time began (probably 1995). This is just a refinement of the approach.

There is no data held by speciesindex.org it is just a service so until data is made available that uses the Species Index URIs the service is more or less useless. data.speciesindex.org has therefore been created to host seed data that makes use of the URIs and allows people to start creating ontologies and applications. Visit data.speciesindex.org for usage examples.

The Simple Set of Rules

The rules are really simple. Here are some examples to give you a flavour before we start.

- <http://speciesindex.org/Panthera/tigris/1758/iczn>
- That is a tiger to me maybe Shere Khan to you if you are a boy Scout.
- <http://speciesindex.org/Panthera/tigris/corbetti/1968/iczn>
- We could be more specific about it not being Shere Khan (I don't think this subspecies occurs in India.)
- <http://speciesindex.org/Rhododendron/luteum/icbn>
- The sweet smelling yellow shrub I rather like.
- <http://speciesindex.org/Solanaceae/icbn>
A family of things we like to eat even though many are poisonous.
- <http://speciesindex.org/Rhododendron/icbn/01607d9f-fb05-403d-beda-924bb0752050> A taxon that would have the name of a species of *Rhododendron* if I knew it but I don't so I am giving it a unique token for now. This is the typically "Species A" problem of

many studies. Is my species A the same as your species A - not unless you assert they have the same URI!

Here are the actual rules.

- The URI must start with "http://speciesindex.org" followed by one or more of the following separated by slashes.
- **First word of name.** Must only contain letters. Must not be the same as one of the names of the nomenclatural codes (icbn or iczn). Optional but highly recommended.
- **Second word of name.** Must only contain letters and not be a nomenclatural code name. Optional.
- **Third word of name.** Must only contain letters and not be a nomenclatural code name. Optional.
- **Year of publication.** Must be an integer greater than 1650 and equal to or less than the current year. If this is an ICZN name then this should be the year the species (epithet) was published as is commonly cited after the name. If this is an ICBN name at species or below then it is the date of the combination. Optional.
Recommended for zoological names if known. Not recommended for botanical names unless there is a known problem with homonyms in use by non-taxonomists.
- **Nomenclatural code** governing the name of the taxon. Currently this must be either 'icbn' or 'iczn'. This may be omitted if the code is unknown or not relevant. Other codes may be added to this list.
- **Qualifier** This must be a Version 4 RFC-4122 UUID. Optional. Used to generate a new independent identifier for a taxon for which the conventional name is unknown or does not exist or to indicate a particular taxon concept that bears the embedded name.
- The whole speciesindex.org URI string should be considered case

sensitive. Everything should be lower case apart from the first letter of words that are specified as having upper case in their relevant codes e.g. names at and above the rank of genus.

You can create URIs from data in your database or spreadsheet very easily simply by writing SQL queries or formulas.

Homonyms

Speciesindex.org URIs do not use authority names. Although there are standards for abbreviation of author names (notably Brummitt in botany) these are not always followed and often embellished. Furthermore it is believed that the added nomenclatural precision author names add is not worth the cost of their inclusion. If author names were included then every variation of authority string would result in a new URI implying the existence of a new taxon. This would defeat the principle goal of speciesindex.org - to get people using the same URIs for the same things. Homonyms are rare it is even rarer that they cause problems outside of taxonomy and nomenclature.

Consider the following classification of confidence limits from International Panel on Climate Change (taken from [here](#))

- virtually certain - more than 99%
- extremely likely - more than 95%
- very likely - more than 90%
- likely - more than 60%
- more likely than not - more than 50%
- unlikely - less than 33%
- very unlikely - less than 10%
- extremely unlikely - less than 5%

Now consider the estimate in Paton *et al* (2008) Taxon 57:602-611 that 4.1% of plant names have homonyms i.e. it is "extremely unlikely" that any one name

is a homonym. Also consider the following list of kinds of homonyms:

- **Nomenclatural Artefacts** These occur where the same taxon is published multiple times. Perhaps the same publication comes out in two languages or is published a second time with a slightly different title and set of authors. For all intent and purposes these do not matter as the names are intended to refer to the same taxon.
- **Competitive Publication** New material is found. Two authors publish accounts based on it using the same names. The taxa are substantively the same.
- **Quickly Synonymised.** An author publishes new species only for someone to quickly realise that this is a homonym and publish the fact. Subsequent publications place it in synonymy and it is never widely used. The name in circulation will almost always refer to the correct taxon but the homonym will be kept in circulation due to always being mentioned as being a homonym in monographs, floras and faunas. Modern indexing will exasperate this situation.
- **Back From The Dead** Everyone is happy using a junior (or later) homonym without knowing it when a taxonomist finds a publication containing the senior (earlier) homonym and overturns the nomenclatural apple cart. The rules of nomenclature say that the taxon now needs a new name even if the senior homonym is not currently the name of an accepted taxon. There is a case for nomenclatural conservation of the junior homonym or rejecting the senior homonym. Either way the original usage of the name is the most common.
- **Problematic Homonyms** The same name string is widely used for multiple taxon concepts. This is

rarer in terms of nomenclatural homonyms (where different names have actually been published) than it is where authors have simply used the same name in different senses (taxon concepts and/or misapplied names). This is particularly common with European names being used for the "wrong" taxa in the New World. Author strings are of no help here as the nomenclature is correct only the usage incorrect. A full-blown taxon concept based approach is needed to handle these situations.

Speciesindex.org takes the premise that names specified to nomenclatural code, spelling and, in the case of zoological names, year are "virtually certain" to be referring to the same general taxon.

Rank

An early version of the speciesindex.org approach made use of rank as one of the elements of the URI. After consultation this was dropped as the additional resolution of homonyms was low compared to the added level of complexity. Most codes have ending conventions for names at particular ranks (e.g. aceae for botanical families). Capitalisation differentiates between subgenera and species binomials. There could be some confusion with trinomials in botany but we figure we can live with this.

Taxon Concepts

Names are not always used in the same sense. If we can never be sure that two studies have used the same name in the same sense then it isn't clear whether we can combine their results. This is discussed in [Taxa, Taxon Names and Globally Unique Identifiers in Perspective](#) which is something of a summary paper on taxon concept issue.

Although the taxon concept issue has been actively discussed for over fifteen years there are very few systems that track actual taxon circumscriptions. Most people simply use names when they mean taxa. speciesindex.org is an acknowledgement of this and an attempt to clarify what is meant when a name is used.

The safest interpretation of the taxa identified by speciesindex.org URIs is as *senus lato* taxa. i.e. as representing the sum of all taxa for which this is the accepted (in botany) or valid (in zoology) name. Practically these taxa may be interpreted as the community accepted circumscriptions of the taxa in the context in which they are found. Different sources may make conflicting assertions about the taxa but it is up to the data consumer to keep track of which assertion came from which source.

UUID qualifiers can be used as part of speciesindex.org UUIDs to represent specific named taxon concepts, whether or not these names have valid/accepted names under either of the codes. Data publishers can produce their own URIs for their own taxon concepts and link them to speciesindex.org taxa either via names or taxon-taxon relationships such as 'includes' or 'isIncludedIn'.

If a data publisher produces their own URIs for their own concepts then it is **highly recommended** that the Dublin Core [dc:isVersionOf](#) property is used to link the taxon to the speciesindex.org taxon. This means that a consumer can 'ingest' data from multiple sources that are talking about multiple versions of the same taxon and detect this fact without having to manually assert the relationships on the basis of name parsing.

It is hoped that by giving a mechanism for people to use taxon concepts correctly when they only have (and possibly only care about) name strings

then in the longer term the use of more tightly defined taxon concepts will become established where it is required.

Expansion of Nomenclatural Data

Purely nomenclatural data is purportedly objective. There are databases of nomenclatural data available.

Speciesindex.org can use these databases to expand on the nomenclatural data provided in the URI.

An initial implementation of this approach has been made to test the concept. Speciesindex.org currently uses the Index Kewensis subset of IPNI to try and expand names that have the code 'icbn' and a two part name. If the name is unambiguous it will use the IPNI URI as the URI for the **name** object it returns within the taxon. It will also add some authorship and publication details to the name. The URI

<http://speciesindex.org/Rhododendron/ponticum/icbn>

is an example of this behaviour. If the species name is ambiguous according to IPNI (i.e. there is more than one record for the binomial) then speciesindex.org will create its own URI for the **name** and add a link to each of the potential IPNI records using the dcterms:relation property. This would allow a user (person or machine) to navigate to the possible name records without implying a strong semantic relationship. The URI <http://speciesindex.org/Rhododendron/luteum/icbn> is an example of this behaviour as it is a homonym.

This limited demonstration could be expanded in collaboration with nomenclatural data providers should it be found desirable.

Another possibility for nomenclatural expansion is to link directly to citations in [BHL](#) if the original place of publication can be identified.

Suggestions, Comments, Criticisms and Condemnations

The best place to discuss this site is probably the tdwg-content@lists.tdwg.org mailing list. You can join [here](#) if you are not a member already.

Please contact me ([Roger Hyam](#)) directly if there is something about the site that isn't working.

Frequently Asked Questions

to be asked...