

# Requirements for implementation of the supply side piping v. 2.2

---

Compiled by Viktoras Didziulis & Kwok Yin Cheung

i4Life, WP12

## Requirements:

We would like partners to provide **regular** updates of species names (*the ones that are NOT available in the latest releases of the COL*) and related nomenclatural and taxonomic information to the piping tool at Reading (WP12).

Downloads should be presented as a single table stored in comma (or tab) delimited text file using UTF-8 character encoding with file extensions set to .txt, .dat or .csv . File can be compressed (zipped or gzipped). Inclusion of DwC-A meta.xml with description of fields is optional but still recommended.

A minimum of 3 fields are compulsory (5 if data set includes infraspecies or synonyms), however other fields are also required if information is present in the provider database, therefore they are marked “compulsory where appropriate”. Please see the table below for a list of fields required by the piping tools. If in doubt, please check DwC-A definitions at DwC-A as described at [http://rs.gbif.org/core/dwc\\_taxon.xml](http://rs.gbif.org/core/dwc_taxon.xml) .

<b>Field name</b>	<b>DwC-A</b>	<b>Description</b>	<b>field status</b>
taxonID	<a href="http://rs.tdwg.org/dwc/terms/taxonID">http://rs.tdwg.org/dwc/terms/taxonID</a>	species name or taxon id as assigned by Global Biodiversity Data Provider	<b>compulsory</b>
genus	<a href="http://rs.tdwg.org/dwc/terms/genus">http://rs.tdwg.org/dwc/terms/genus</a>	First part of species binomial Latin name specifying genus (e.g. Teredo)	<b>compulsory</b>
specificEpithet	<a href="http://rs.tdwg.org/dwc/terms/specificEpithet">http://rs.tdwg.org/dwc/terms/specificEpithet</a>	Second part of species Latin binomial name (e.g. navalis)	<b>compulsory</b>
scientificNameAuthorship	<a href="http://rs.tdwg.org/dwc/terms/scientificNameAuthorship">http://rs.tdwg.org/dwc/terms/scientificNameAuthorship</a>	Author, who first described species or infraspecies- name and year (e. g. Linnaeus, 1758)	compulsory where appropriate
infraspecificEpithet	<a href="http://rs.tdwg.org/dwc/terms/infraspecificEpithet">http://rs.tdwg.org/dwc/terms/infraspecificEpithet</a>	Epithet characterising infraspecies in Latin trinomial name	compulsory where appropriate

verbatimTaxonRank	<a href="http://rs.tdwg.org/dwc/terms/verbatimTaxonRank">http://rs.tdwg.org/dwc/terms/verbatimTaxonRank</a>	var., subsp., infr. etc...	compulsory where appropriate
<b>taxonomicStatus</b>	<a href="http://rs.tdwg.org/dwc/terms/taxonomicStatus">http://rs.tdwg.org/dwc/terms/taxonomicStatus</a>	The status of the use of the scientificName as a label for a taxon. (one of: <b>accepted, synonym, unknown</b> )	<b>compulsory</b>
<b>acceptedNameUsageID</b>	<a href="http://rs.tdwg.org/dwc/terms/acceptedNameUsageID">http://rs.tdwg.org/dwc/terms/acceptedNameUsageID</a>	for synonym or infraspecies, this field should point to the <b>taxonID</b> of the accepted binomial name.	<b>compulsory for synonyms</b>
<b>parentNameUsageID</b>	<a href="http://rs.tdwg.org/dwc/terms/parentNameUsageID">http://rs.tdwg.org/dwc/terms/parentNameUsageID</a>	parent taxonID for infraspecies	<b>compulsory for infraspecies</b>
family	<a href="http://rs.tdwg.org/dwc/terms/family">http://rs.tdwg.org/dwc/terms/family</a>	The name of the family in which the scientificName is classified.	compulsory where appropriate
order	<a href="http://rs.tdwg.org/dwc/terms/index.htm#order">http://rs.tdwg.org/dwc/terms/index.htm#order</a>	The full scientific name of the order in which the taxon is classified.	compulsory where appropriate
class	<a href="http://rs.tdwg.org/dwc/terms/index.htm#class">http://rs.tdwg.org/dwc/terms/index.htm#class</a>	The full scientific name of the class in which the taxon is classified.	compulsory where appropriate
phylum	<a href="http://rs.tdwg.org/dwc/terms/index.htm#phylum">http://rs.tdwg.org/dwc/terms/index.htm#phylum</a>	The full scientific name of the phylum or division in which the taxon is classified.	compulsory where appropriate
kingdom	<a href="http://rs.tdwg.org/dwc/terms/index.htm#kingdom">http://rs.tdwg.org/dwc/terms/index.htm#kingdom</a>	The full scientific name of the kingdom in which the taxon is classified.	compulsory where appropriate
higherClassification	<a href="http://rs.tdwg.org/dwc/terms/higherClassification">http://rs.tdwg.org/dwc/terms/higherClassification</a>	For the purpose of piping this field should include full classification above species with all ranks which are present in the database, primarily for inspection by human. Data should be provided as a list of semicolon separated pairs of colon separated classification level	compulsory where appropriate

		and name, e.g.: genus:Cacoblatta;family:Blaberidae;superfamily:Blaberoidea;order:Blattodea;superorder:Dictyoptera;cohort:Polyneoptera;infraclass:Neoptera;class:Insecta;phylum:Arthropoda;kingdom:Animalia	
namePublishedIn	<a href="http://rs.tdwg.org/dwc/terms/namePublishedIn">http://rs.tdwg.org/dwc/terms/namePublishedIn</a>	Author name, year and title of publication describing the species (Nomenclatural publication).	compulsory where appropriate
taxonRemarks	<a href="http://rs.tdwg.org/dwc/terms/taxonRemarks">http://rs.tdwg.org/dwc/terms/taxonRemarks</a>	Any additional information data provider wishes to pass to GSD	compulsory where appropriate
source	<a href="http://purl.org/dc/terms/source">http://purl.org/dc/terms/source</a>	Any url pointing to origin of species information	compulsory where appropriate
relationshipRemarks	<a href="http://rs.tdwg.org/dwc/terms/relationshipRemarks">http://rs.tdwg.org/dwc/terms/relationshipRemarks</a>	taxonID is: a) specific to the dataset, b) globally unique name id, c) globally unique taxon id	compulsory where appropriate

### Proposed steps for implementation:

- 1) We would like to ask the partners to provide a web folder on their web servers, so that piping tools can access its contents over http. URL of this folder should be forwarded to Viktoras Didziulis ([v.didziulis@reading.ac.uk](mailto:v.didziulis@reading.ac.uk)) and Kwok Yin Cheung ([k.y.cheung@reading.ac.uk](mailto:k.y.cheung@reading.ac.uk))
- 2) Data providers would regularly provide a file for download naming it according to date of the last update in format DDmmmYYYY.zip. Eventually the listing of a web folder contents might look like this: 20jun2011.zip, 05aug2011.zip, 12aug2011.zip
- 3) An alternative to step 2) : data provider creates either a static web page, text file or a dynamic webservice pointing to delimited text downloads as a result of database query or any other data retrieval process on data provider's web server. The names of the links should follow same format: DDmmmYYYY . Eventually the web page or output of webservice would list URLs pointing to relevant updates:  
<http://www.dataprovider.org/piping/20jun2011.zip>  
<http://www.dataprovider.org/piping/05aug2011.zip>

<http://www.dataprovider.org/piping/12aug2011.zip>

- 4) We are also considering a slightly different approach (which finally might prove to be a preferred way of updates) where the supply-side partner creates a regular snapshot of the whole database in the format described above and puts them for download always under the same file name (no timestamp in file name is required). We would regularly download the whole data set.