

i4Life



INDEXING for LIFE

The Catalogue of Life e-Infrastructure project,

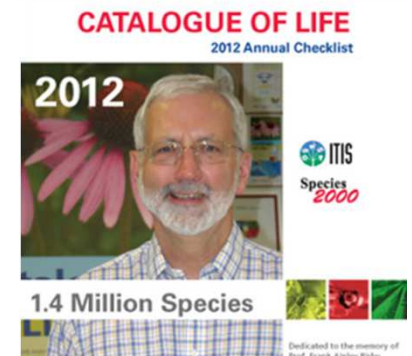
<http://www.i4life.eu>

Wouter Addink, Walter Berendsohn, Wieslaw Bogdanowicz, Guy Cochrane, Alastair Culham, Markus Döring, Viktor Didziulis, Anton Güntsch, Craig Hilton-Taylor, Andrew Jones, Cherian Mathew, Vincent Robert, Yuri Roskov, Peter Schalk, Magda Sitko, et al.



The Catalogue of Life

- A quality-assured checklist of more than 1.4 million species of plants, animals, fungi and micro-organisms, (~ 70% of all known to science).
- An uncomplicated structure, and both minimal and standardised dataset provide a sound baseline of species information for all biologists.
- Unique in its breadth of coverage of organisms, the degree of validation in the knowledge set, and its wide global take-up.
- Releases as an annual citable version and a dynamic monthly edition
- More information at: <http://www.catalogueoflife.org/>



The i4Life “Global Partners”



Large numbers of names and taxa from

- **Specimens & Observations** (Global Biodiversity Information Facility (GBIF))
- **Molecular data** (European Bioinformatics Institute EBI) and molecular barcodes (Barcode of Life, BoL)
- **Red lists** (International Union for the Conservation of Nature, IUCN)
- **Species pages** (EoL)

(+ LifeWatch - the developing European large-scale service infrastructure for biodiversity research)



i4Life project goals (1/3)

Designing and implementing a pipeline for transmitting CoL and CoL updates to the global partners platform:

Download Services

Download in DarwinCore Archive format

- Choice of latest monthly edition or earlier version
- Entire Catalogue or select taxon group and/or select fields
- Manual download or webservice

Version 1.3 currently in beta test

Import scripts for DwC-A files are provided for MS SQLServer, MySQL, PostgreSQL and SQLite3

i4Life project goals (2/3)

Designing and implementing a regular cross-mapping process between the CoL taxonomy and the global partners.

The **cross-mapping system** provides a Web interface

- accepts checklists uploaded in (i4Life-) Darwin Core Archive format
- compares the names and their arrangements in a pair of checklists, identifying *corresponds*, *overlaps*, *includes*, *included_by* and *not_found_in* relationships
- browse or download the resultant cross-map
- download full details of the additional (*not_found_in*) taxa in i4Life Darwin Core Archive format for use in the piping tool.



i4Life project goals (3/3)

Establishing the workflow by which each partner organisation can make contributions of missing taxa and names to the CoL, using the CoL piping tools



- Pulling data-sets into reception buffer of the CoL, piping to Global Species Databases
- Basically operative (as webservice)
- So far, 1.4M names collected from Global Partners have been piped to GSDs, only 67,000 names could not be assigned to a GSD
- Process monitor presently undergoing final testing, will be released together with all other components by the end of 2012.

i4Life and LifeWatch

Offer webservice for deployment in workflow environments such as Taverna and Kepler



- An EDIT platform Catalogue of Life-instance is cyclically refreshed using the i4Life download services and the associated i4Life Darwin Core Archive dialect
- The service layer works with any EDIT Platform compliant data source.
- The service layer is accessible via three service endpoints installed on servers at the Botanic Garden and Botanical Museum Berlin-Dahlem
- Tested using Taverna-based scientific workflows developed in the context of BioVeL (Developing a Biodiversity Virtual e-Laboratory)

Thank you for your attention

