Indexing for Life

Roadmap for i4Life pilot projects

Work packages 2 & 3

Alastair Culham, Peter Schalk, Yuri Roskov, Magda Sitko, Rebecca Mann and Viktoras Didziulis

9 July 2012

Grant Agreement No:	261555
Project Co-ordinator:	Dr Alastair Culham
Project Homepage:	http://www.i4Life.eu
Duration of Project:	36 months
Start Date:	November 2010
End Date:	November 2013

Contents

Introduction

There are many names and taxa in databases of the global partner programmes (GBIF, IUCN Red List, BOLD and EMBL ENA) that are not yet listed in the Catalogue of Life. These may constitute a small fraction of the total, but this makes the CoL taxonomy markedly less useful to the partner programmes. It means that even if they wished to use the CoL as their main index, they cannot do so because it would not be capable of indexing all the entries that they hold.

There is potentially a triple benefit from contributing and placing missing names and taxa into the CoL, so this is an important issue for the community:

1. It will make the CoL markedly more useful to the global partner programme from which the names and taxa came: these are among the species they need to cover.
2. It could make some large, and certainly many small improvements to the CoL, including progress towards its target of reaching 1.9 million species, and also enriching much of the CoL with a greater synonymy.
3. It would mean that work done in one of the global programmes would benefit another. It could reduce duplication of work in another organisation – providing efficiency gains in what is an expensive activity involving expert opinion.

As part of the commitment for the i4Life project, the CoL GSD-Piping Tools from Work package 12 are expected to assemble batches of name- and taxon- records not already in the Catalogue of Life (CoL) coming from the global partner programmes. These names- and taxon- records are to be partitioned into the Catalogue of Life GSD Sectors and to provide them to the appropriate GSD organisation for placement by the taxonomic experts that work with that GSD sector.

The missing names and taxa are thought to constitute:

1. additional species (and taxa of other ranks) from global partners for placement in the CoL
2. additional names that may be placed as synonyms or provisionally accepted taxa when included in the CoL
3. names of unknown quality from global partners for scrutiny and taxonomic expertise by GSD authors or new authors
4. and, also, taxonomic sectors of high quality from global partners for direct contribution to the Catalogue of Life (new/proto GSD sectors)

To make placements of missing names/taxa accurately, and over the full extent of the CoL taxonomy, is, however, a huge task involving very large numbers of taxonomic experts in the GSD network. The Catalogue of Life GSD network includes 115 contributing GSD databases, each with a set of taxonomic expert editors and contributors taking responsibility for the taxonomic sectors supplied in the Catalogue. The database organisations currently provide maintained taxonomic expertise on over 2,500 taxonomic sectors from which the CoL is composed. It is proposed that the CoL instigates a new process for GSDs to scrutinise and find accurate placement for the putatively missing names and taxa.

The i4Life programme, through workpackages 2 & 3, has responsibility to establish and implement an experimental procedure for selecting taxonomic specialists to incorporate the

new names and taxa into GSDs or proto GSDs that can in turn feed back content to Catalogue of Life. There is an allocation of approximately 200k euro plus overheads in the budget of i4Life WP5 (Species 2000) for the purpose of adding names and taxa to GSDs, for the establishment of new GSDs, and/or for editing and annotating names that are not attached to any GSD. Funds will be distributed to an array of i4Life pilot projects for placing new names/taxa in source databases and delivering them to the Catalogue of Life. Species 2000 as a partner of the i4Life project has to help develop both the written agreements and arrange payments to the GSDs that participate in the pilot project.

It is intended that GSD and protoGSD authors should maintain the taxonomic integrity and taxonomic view that they determine in the GSD with which they work. Thus, it needs to be stated clearly that names recognised as accepted species in the taxonomy of one of the global partner programmes, may sometimes be placed as synonyms of other species by the GSD authors.

This document describes the Roadmap and eligibility criteria for the i4Life pilot projects.

Objective

The pilot project objective is to allocate as many new names from global partners as possible in taxonomic checklists and, then, supply them to the Catalogue of Life during routine updates. Names, which cannot be included in GSDs or protoGSDs (for example, mistakenly allocated names from other taxa, or incomplete/incorrect names, or names with unresolved taxonomic/nomenclatural status, etc.) should be returned to i4Life buffer warehouse with appropriate comments.

Procedure

The DoW describes the procedure as follows: The 200k€ will be allocated over just $5 - 15$ Placement Projects, each allocated between 5 and 40 k. Arrangements will be made with each database member under the prior agreement of the Rotating Fund Framework. This is part of the pre-existing Species 2000 Access Agreement, already signed between provider databases and Species 2000, and in which Species 2000 contributes to the operating costs of the database organization on a not-for-profit basis. The selection of which databases are needed cannot be made until the extent and content of the new taxon and new names datasets coming from the global partners can be analyzed.

MNHN, Paris (i4Life participant under WP2) will coordinate the call with help from University of Reading (i4Life project coordinator), collecting applications and monitoring progress within funded pilot projects. An i4Life expert panel will evaluate applications and release a list of candidates selected for pilot project funding. A final list of pilot projects should be approved by the i4Life Management Committee in consultation with the Species 2000 Global Team and the Board of Directors. The Board of Directors who are the legal representatives of Species 2000 will draft written agreements and will arrange payments to the GSDs that participate in the pilot project in accordance to the principles of the Rotating Fund Framework and the regulations imposed by EU FW7 funding.

It is a European Commission requirement is that all organisations that are granted funded pilot projects must be added through a grant amendment to the project as Third Parties under

Partner 9 (Sp2000) subject to Special Clause 10 and must be eligible to comply with EC regulations (*detailed explanation of eligibility criteria is included below*).

The i4Life project Evaluation panel

In the evaluation panel should be people with taxonomic database expertise. Among others it should include one of Species2000 directors, one of i4Life Managers.

The criteria for evaluation of the pilot project proposals are:
1. Providing the best contribution in terms of numbers and impact of additional names on the CoL
2. Cost effectiveness: delivered number of names versus projected costs
3. Feasibility: likeliness of success within listed time and budget?
4. Compliance with the CoL requirements see at:
   http://www.catalogueoflife.org/content/contributing
5. Eligibility to receive funds under the EC FP7: organisations considered to become part of pilot projects must have a valid Participants Identification Code (PIC) or have applied for it.

The pilot project Roadmap

Timetable for i4Life pilot projects

Preparation phase:

i4Life/CoL placement service ready	15th July 2012
Circulate draft call text to i4Life MC	10th July 2012

Call:

Call opening date	23rd July 2012
Call closure date	31st August 2012
Proposal evaluation	1-14th September 2012
Confirmation by the i4Life MC and the CoL Global Team	15-21st September 2012
Announcement of evaluation results by Species 2000	24th September 2012
Signing of contracts with Species 2000	September/October 2012
Negotiations with EC for grant amendment (Approximately 6-8 weeks)*	October / November 2012
Start of pilot projects	1 December 2012
Delivery of intermediate report (and possibly download) of first batch of incorporated/annotated names	1 March 2013
Final date for delivery of project results to the CoL**	30th April 2013

* other processes will be delayed if this takes longer
** final delivery will consist: updated GSD plus commented names (all additional names have to be annotated) in parallel with download

The detailed description of all planned steps

Step 1. CoL Placement Service (relates to the i4Life deliverable D12.1)
In the i4Life project names/taxa are being harvested from global partners and subsequently divided between GSD sectors by the i4Life Piping Tools. Names are available for all CoL GSDs through their individual CoL Usermin accounts. Names not present in the CoL are flagged. Names in gap areas are also stored in the CoL Usermin account.
Structure of data in piping tools is described in Appendix 3.
Specification of the dataset supplied by i4Life project to GSD is described in Appendix 4.

Step 2. Assessment of Taxonomic Sectors (relates to the i4Life milestone MS36)
The i4Life WP3 leader will manage the preliminary triage of new names allocated in GSD sectors and gap areas, and issue recommendations for GSDs/taxa of high priority. Triage will be based on:
- highest numbers of new names in GSD sectors (as well as, in gap areas)
- where possible, on "impact factors" (how many specimen or sequence records associated with each name in global partner database)

Step 3. i4Life pilot projects
The call for pilot projects will be open for prioritised GSDs/taxa and run for 6 working weeks.

GSDs will be invited to look and assess names in their CoL Usermin accounts and, then, submit application for pilot project to accommodate as many names as possible from global partners in their GSD checklists. For the names that represent gap areas in the CoL, taxonomic experts will be given the opportunity to propose a project to establish new GSDs for these sectors.
Preliminary text of call for i4Life pilot projects is given in Appendix 1.

Step 4. Delivery of placed or annotated unvetted names with GSD updates
Names/taxa placed in GSDs will be delivered to the Catalogue of Life through the existing process of updating data. Unplaced names should remain in the spreadsheet with appropriate comments. Unplaced names can be included in i4Life unresolved lists for experimental purposes (e.g. "i4Life GBIF Unvetted Names").

A preliminary list of categories/comments for unplaced names is given in Appendix 3 of this document.

Step 5. Final reports on i4Life projects
The final report from GSDs on their pilot projects should contain statistics of placed and unplaced names, feedback on categories of unplaced names and on effectiveness of used procedures, and a financial report
These reports will be later incorporated in the i4Life Wp5 final report.
Template for final report is in Appendix 6.

Eligibility criteria for i4Life pilot projects

This document outlines the eligibility criteria for those organisations that may be offered to participate in the i4Life project as part of a Pilot Project for the Placement Tasks.

European Commission requirement is that all organisations that are granted funded pilot projects must be must be added through a grand amendment to the i4Life project as Third Parties under Partner 9 (Sp2000) subject to Special Clause 10 (*see Appendix 7*) and must be eligible to comply with EC regulations (*see page 50 of the i4Life Project Annex I description of Work at: www.i4Life.eu*). The terms of the contract with the project are for guidance only and may vary.

General provisions for participants

Any company, university, research centre, organisation or individual, legally established in any country, may participate in a collaborative project (known as an indirect action) provided that the minimum conditions laid down in the Rules for Participation [PDF] in FP7 (RFP), Chapter II, Section 1, p.12 [PDF], have been met, including any additional conditions laid down by specific programmes or individual work programmes (see Article 12 of RFP). Please note:

- An individual defined as a legal entity is any natural person, or any legal person created under the national law of its place of establishment, or under Community law or international law, which has legal personality and which may, acting under its own name, exercise rights and be subject to obligations;

- In addition to the minimum conditions laid down in the RFP, specific programmes or work programmes may lay down conditions regarding the minimum number of participants, or additional conditions on the type of participant or place of establishment.

For more information see: http://cordis.europa.eu/fp7/who_en.html

Cooperation with Third Country Participants in EC Funded FP7 multi-partner research project

It is possible for organisations based outside of Europe to participate in EC FP7 Projects. Currently, organisations and researchers from more than 100 countries all of the world are already involved in various EU Research Programmes.

Third Country Participants are all those participants who are established in a non EU country, which is not associated to the Seventh Framework Programme of the European Community for Research, Technological Development and Demonstration Activities (FP7).

For a complete list of associated countries see View the Associated Countries [DOC] Version: October 2011

Third countries consist of two groups:

a. International Cooperation Partner Countries (ICPCs)

Legal entities from ICPCs may participate in FP7 activities (provided that the minimum conditions laid down in the FP7 'Rules for Participation' (RFP) have been met). They are eligible to receive EC funding.
View the list of the International Cooperation Partner Countries (ICPC) *[PDF]*

b. High-income countries

High-income countries are all the countries not included in the ICPC list and not associated to FP7. These include in particular the United States of America, Canada, Japan, the Republic of Korea, Singapore, Australia and New Zealand. Taiwan, Hong Kong and Macao are also considered as high-income territories, as well as the Vatican, San Marino, Monaco and Andorra.

Legal entities from high-income countries may participate in indirect FP7 activities (provided that the minimum conditions laid down in the FP7 'Rules for Participation' [RFP] have been met). However, they are normally not eligible to receive an EC financial contribution to cover the costs related to their participation in the project. This means that participants from high-income countries will not be eligible to receive funds under i4Life pilot projects.

The Unique Registration Facility and registration process

In addition to meeting the eligibility criteria for funding, any organisation wishing to benefit from the receipt of a funded Placement Project will also need to be a registered with the European Commission and have a validated Participant Identification Code (or PIC) in order to be added as a Third Party in the Project.

For any grant agreement all participants (i.e. the institution or company for which each of the researchers works) need to be registered via the Unique Registration Facility (http://cordis.europa.eu/fp7/pp_en.html), including the third country participants. Thus, each applicant should first check if their own institution is not already registered in the URF. Registration and validation take some time. Thus, it is recommended to verify, already during proposal submission, which documents would be needed for a registration and to proceed to registration as soon as funding becomes likely, and at the latest immediately after the start of the grant negotiation phase.

It will be possible for us to consider placement projects for an organisation that do not already have a validated Participant Identification Code (or PIC). However, final acceptance of the Placement Project will only be granted on the condition that the organisation concerned in the proposal is able to complete the Registration Process with the Commission and obtain a validated PIC in order to receive the funding.

Conclusion

For consideration of Pilot Projects the criteria listed above will be taken into account. However, organisations ineligible for this funding or those whose proposal is not successful can still be provided with a list of unlisted CoL names generated by the i4Life project. This provision of names is a complementary service to these GSDs and they will not be required to complete the reporting needed from funded projects.

Appendix 1. i4Life call for pilot projects: draft document

In the global biodiversity portals of GBIF, IUCN Red List, BOLD and EMBL ENA (herein after named as "global partners"), there are many names and taxa that are not yet listed in the Catalogue of Life. Part of the i4Life EC project (www.i4life.eu) activity is to:
- discover these unlisted names
- use these names to enrich as much as possible the GSDs and Catalogue of Life
- annotate the unvetted names with expert comments (where it's possible) for future indexing use.

In the i4Life project the CoL GSD-Piping Tools will assemble batches of name- and taxon-records from the global partner programmes that do not occur in the Catalogue of Life (CoL) coming. These names- and taxon- records will be partitioned into the Catalogue of Life GSD Sectors and subsequently provided to the appropriate GSD organisation for placement by the taxonomic experts.

The names and taxa missing from the CoL are thought to constitute:
1. additional valid species (and taxa of other ranks) that need placing in the CoL
2. additional names that may be placed as synonyms or provisionally accepted taxa when included in the CoL
3. names of unknown quality offered for scrutiny by GSD authors or new authors
4. taxonomic sectors of high quality for direct contribution to the Catalogue of Life (new/proto GSD sectors)

The i4Life programme will establish and implement an experimental procedure for selecting taxonomic specialists to incorporate the new names and taxa into GSDs or proto GSDs that can in turn feed new content to Catalogue of Life. The i4Life budget commits approximately €200k for this purpose. These funds will be distributed among GSDs and new authors in the form of i4Life pilot projects using the Species 2000 Rotating Fund Framework.

All newly harvested names from global partners will be made available to the GSDs via the CoL Usermin accounts (www.catalogueoflife.org/usermin. Those who have no password are kindly requested to contact the CoL System Manager (v.didziulis@reading.ac.uk). Each species name is structured as Genus, SpeciesEpithet and AuthorString (plus, InfraspecificEpithet and Rank for infraspecific taxa). Where it's available, additional information is provided: taxonomic status of the name (i.e. accepted or synonym of other name), full classification used by global partner database, references, link to species page in global partner database for further information. *Full details of "piping dataset" are given in Appendix 3.*

Species 2000 invites all GSDs to assess names allocated in their CoL Usermin accounts and to submit a proposal for an i4Life Pilot Project, including a cost estimate for their expert work to place the additional names in their GSD.

For names in gap areas Species 2000 welcomes new authors. The requirement in this case is to establish new GSDs and to become a member of Species 2000.
For more details about required procedure please visit Species 2000 website at:
http://www.sp2000.org/index.php?option=com_content&task=view&id=43&Itemid=75

The target is to include as many "good" names in your GSD as possible and to annotate the other names, which cannot be included in your taxonomic checklist for various reasons, in a spreadsheet (*list of standard annotation categories given in Appendix 5*). We expect experts to follow taxonomic & nomenclatural rules and the classification accepted in the GSD when adding new names in the checklist. If experts decide to add "good" species in the checklist, compliance with the CoL standard dataset (i.e. synonyms, references, distribution, life zone, and common names) will be appreciated. A potentially "good" name in your taxonomic group with unresolved taxonomic status may get the CoL "Provisionally accepted name" status.

Names placed in the GSDs should be delivered to the Catalogue of Life through routine update. Unplaced names should be delivered to i4Life WP2/WP12 in annotated spreadsheet. The final report should include statistics of placed and unplaced names, comments on unplaced names and feedback on effectiveness of used procedures.

Call:

Call opening date	23rd July 2012
Call closure date	31st August 2012
Proposal evaluation	1-14th September 2012
Confirmation by the i4Life MC and the CoL Global Team	15-21st September 2012
Announcement of evaluation results by Species 2000	24th September 2012
Signing of contracts with Species 2000	September/October 2012
Negotiations with EC for grant amendment	
(Approximately 6-8 weeks)*	October / November
2012Start of pilot projects	1 December 2012
Delivery of intermediate report (and possibly download)	
of first batch of incorporated/annotated names	1 March 2013
Final date for delivery of project results to the CoL***	30th April 2013

* other processes will be delayed if this takes longer
** final delivery will consist: updated GSD plus commented names (all additional names have to be annotated) in parallel with download

Proposals must
- be written in English,
- include a workplan and budget calculated in Euro.
- comply with the template given below.

Proposals have to be submitted electronically, using MS Word DOC, RTF or PDF format, to: Aurélie Delavaud (aurelie.delavaud@mnhn.fr), the pilot project call coordinator employed under i4Life WP2. *Required format of the proposal is given in Appendix 2.*

Appendix 2. Required format of the proposal

Page 1

Title for Proposed Project:
i4Life *[please add GSD short name]* pilot project for allocation of new names from global partners

Taxonomic sector covered by GSD:

Number of new names available in CoL Usermin account:

Required budget in euro:

Contact Information for Principal Investigator:
> Name:
> Address:
> Telephone:
> FAX:
> Email:

Contact information for Managing Organisation
> Organisation Name:
> Address:
> Organisation Contact Person
>> Name:
>> Telephone:
>> Fax:
>> Email:

Participant Identification Code (PIC):
> *(or information that the process of applying for it has already started)*

Bank details *(required for money transfer):*

Date of submission

Page 2(-3)

Project Description**:** a description (no more than two pages) of the project management strategy, which is to include the following items:
- Total cost of project;
- The number of species contained in the product before the start of the proposed project;
- How your project will deal with the i4Life names (covering both names included in GSD and left in the spreadsheet as unplaced names with annotations);
- The proposed results, which is due no more than six months after receipt of the seed money;

- Project milestones and timeline;
- The human and institutional infrastructure of the project (expertise base);
- Institutional support and potential partnerships if applicable;
- The technical and human resources necessary to complete the project;
- Risk assessment.

Appendix 3. Specification of the dataset supplied by i4Life global partners

Field name	DwC-A description	Description	Field status (for global partners)
taxonID	http://rs.tdwg.org/dwc/terms/taxonID	Species name or taxon id as assigned by global partners	**compulsory**
genus	http://rs.tdwg.org/dwc/terms/genus	Genus name (e.g. Teredo)	**compulsory**
specificEpithet	http://rs.tdwg.org/dwc/terms/specificEpithet	Species epithet (e.g. navalis)	**compulsory**
scientificNameAuthorship	http://rs.tdwg.org/dwc/terms/scientificNameAuthorship	Author string (e. g. Linnaeus, 1758)	compulsory where appropriate
infraspecificEpithet	http://rs.tdwg.org/dwc/terms/infraspecificEpithet	Infraspecific epithet	compulsory where appropriate
verbatimTaxonRank	http://rs.tdwg.org/dwc/terms/verbatimTaxonRank	var., subsp., form etc.	compulsory where appropriate
taxonomicStatus	http://rs.tdwg.org/dwc/terms/taxonomicStatus	Name status in global partner database (**accepted, synonym, unknown**)	**compulsory**
acceptedNameUsageID	http://rs.tdwg.org/dwc/terms/acceptedNameUsageID	Accepted name ID for synonym (this field should point to the **taxonID** of the accepted binomial name)	**compulsory for synonyms**
parentNameUsageID	http://rs.tdwg.org/dwc/terms/parentNameUsageID	Parent name ID for accepted infraspecies	**compulsory for accepted infraspecies**
family	http://rs.tdwg.org/dwc/terms/family	Family name as it is in global partner	compulsory where

		database	appropriate
order	http://rs.tdwg.org/dwc/terms/index.htm#order	Order name as it is in global partner database	compulsory where appropriate
class	http://rs.tdwg.org/dwc/terms/index.htm#class	Class name as it is in global partner database	compulsory where appropriate
phylum	http://rs.tdwg.org/dwc/terms/index.htm#phylum	Phylum name as it is in global partner database	compulsory where appropriate
kingdom	http://rs.tdwg.org/dwc/terms/index.htm#kingdom	Kingdom name as it is in global partner database	compulsory where appropriate
higherClassification	http://rs.tdwg.org/dwc/terms/higherClassification	Full classification with all ranks as it is present in global partner database.	compulsory where appropriate
namePublishedIn	http://rs.tdwg.org/dwc/terms/namePublishedIn	Reference associated with the name in global partner database.	compulsory where appropriate
taxonRemarks	http://rs.tdwg.org/dwc/terms/taxonRemarks	Any additional information data provider wishes to pass to GSD	compulsory where appropriate
source	http://purl.org/dc/terms/source	URL pointing to species information in global partner database.	compulsory where appropriate
relationshipRemarks	http://rs.tdwg.org/dwc/terms/relationshipRemarks	taxonID is: a) specific to the dataset, b) globally unique name id, c) globally unique taxon id	compulsory where appropriate

Description is taken from the document: "Requirements for implementation of the supply side piping, v. 2.2" by Viktoras Didziulis & Kwok Yin Cheung (i4Life, WP12).

Appendix 4. Specification of the dataset supplied by i4Life project to GSD

Field name	*Description*	*field status*
taxonID	species name or taxon id as assigned by Global Biodiversity Data Provider	**compulsory**
genus	First part of species binomial Latin name specifying genus (e.g. Teredo)	**compulsory**
specificEpithet	Second part of species Latin binomial name (e.g. navalis)	**compulsory**
scientificNameAuthorship	Author, who first described species or infraspecies- name and year (e. g. Linnaeus, 1758)	compulsory where appropriate
infraspecificEpithet	Epithet characterising infraspecies in Latin trinomial name	compulsory where appropriate
verbatimTaxonRank	var., subsp., infr. etc...	compulsory where appropriate
taxonomicStatus	The status of the use of the scientificName as a label for a taxon. (one of: **accepted, synonym, unknown**)	**compulsory**
acceptedNameUsageID	for synonym or infraspecies, this field should point to the **taxonID** of the accepted binomial name.	**compulsory for synonyms**
parentNameUsageID	parent taxonID for infraspecies	**compulsory for infraspecies**
family	The name of the family in which the scientificName is classified.	compulsory where appropriate
order	The full scientific name of the order in which the taxon is classified.	compulsory where appropriate

class	The full scientific name of the class in which the taxon is classified.	compulsory where appropriate
phylum	The full scientific name of the phylum or division in which the taxon is classified.	compulsory where appropriate
kingdom	The full scientific name of the kingdom in which the taxon is classified.	compulsory where appropriate
higherClassification	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 and name, e.g.: genus:Cacoblatta;family:Blaberidae;superfamily:Blaberoidea;order:Blattodea;superorder:Dictyoptera;cohort:Polyneoptera;infraclass:Neoptera;class:Insecta;phylum:Arthropoda;kingdom:Animalia	compulsory where appropriate
bibliographicReference	Author name, year and title of publication describing the species.	compulsory where appropriate
taxonRemarks	Any additional information data provider wishes to pass to GSD	compulsory where appropriate
source	Any url pointing to origin of species information	compulsory where appropriate
updated	Date the name was submitted to the piping tools (last record update timestamp)	
provider	Name of a global partner data come from (GBIF, IUCN etc.)	
in_col	state that the name is inside CoL already.	
gsd_comments	where GSDs leave their free-text notes	

gsd_comments_predefined	Where GSDs select one-off comments from a set of vocabulary	
gsd_short_name	lists of all GSDs the name has been piped to	
gsd_status	either approved or rejected (reasons for rejection may be given in gsd_comments)	
history_status	simply counts how many times same name has been submitted to buffer database by global partners (e.g. is a number 1, 2, 3, … 1000)	
history_comments	we may wish to put any remark on re-appearance of a name	
matched_by	indicates rank which was used to pipe a name to GSD (e.g. matched by family, or order, or class, or phylum, or genus)	
relationshipRemarks	taxonID is: a)

specific to the dataset, b) globally unique name id, c) globally unique

taxon id | compulsory where appropriate |

Specification of the distribution side piping v. 2.3 , compiled by Viktoras Didziulis & Kwok Yin Cheung (i4Life, WP12).

Appendix 5. A list of standard annotation categories for unvetted names

Category	Description	Comment
Incomplete name	All sorts of Latin names, with incomplete or abbreviated genus name, species or infraspecific epithets, names without authorstrig (e.g. *S. aguabonita Jordan, Salmo a. Jordan, Salmo aguabonita*).	If, by the chance, you know complete name, please, input it to your GSD and give it in the spreadsheet.
Chresonym	Scientific name with not validated authorship, which refers to published usage of the name rather then to the true author of the name. Please visit http://en.wikipedia.org/wiki/Chresonym for more details. Example: ***Actinopus crassipes (Keyserling, 1891)*** *Pachyloscelis crassipes Keyserling, 1891: 3, pl. 1, f. 1* *Actinopus crassipes Strand, 1916b: 81* - chresonym *Actinopus crassipes Mello-Leitão, 1923a: 18, f. 128* - chresonym *Actinopus crassipes Bücherl, 1957: 384, f. 5* - chresonym *Actinopus crassipes Schiapelli & Gerschman, 1962b: 72, pl. II, f. 3* - chresonym *Actinopus crassipes Lucas & Bücherl, 1965: 89, f. 1-18* - chresonym	If, by the chance, you know original author of the name, please, put full binomial/trinomial with correct authorship to your GSD and give it in the spreadsheet.
Name with unresolved nomenclatural status	All sorts of not validly published names, or names with the type of unknown location, etc., etc. – Latin names which you are not taking in your taxonomic checklist.	Keep these names in the spreadsheet only.
Hybrid formula	You may find hybrid formulas as "*Dianthus caryophyllus × D. plumarius*" in the spreadsheet.	Keep these hybrid formula names in the spreadsheet only, but add both names of parents in your GSD. If hybrid has binomial

		as "*Dianthus* × *allwoodii*", please add it to your database, trying to avoid cross symbol in species epithet field.
Misspelled name	All kind of published orthographic variants or typos, both in paper and digital media.	Please input correct name to your GSD and give it in the spreadsheet.
Fossil name	Latin name given to fossilised specimen. Names of fossils are out of subject area of the Catalogue of Life.	Keep these names in the spreadsheet only.
Unidentified specimen	All sorts of names used for unidentified specimens/samples (e.g. *Salmon sp.1*, *Gagea aff. lutea* - aff. for "affinis", cf. for "confer")	If, by the chance, you are able to identify specimen/sample and provide correct scientific name, please, input this name to your GSD and give it in the spreadsheet.
Non scientific name	You may find strange names delivered by our partners, which are not Latin (e.g. *Lathyrus with white flowers from Argentina*)	Keep these names in the spreadsheet only.
Name from other taxon	A name which does not belong to your GSD sector and was placed in the list by mistake	If, by the chance, you know correct taxon to which this name belongs, please, give it in the spreadsheet.
Others		Please, add appropriate comment in the spreadsheet

Appendix 6. Template for final report on i4Life pilot project

Page 1

Title of the Project (*as in application form*):
i4Life [please add GSD short name] pilot project for allocation of new names from global partners

Contact Information for Principal Investigator
 Name:
 Address:
 Telephone:
 FAX:
 Email:

Contact information for Managing Organisation
 Organisation Name:
 Address:
 Organisation Contact Person
  Name:
  Telephone:
  Fax:
  Email:

Date of report

Page 2:

Taxonomic sector covered by GSD/planned proto GSD

Number of new names available in Usermin account at the beginning of the project

Requested budget in euro

Number of names, which have been included in GSD

Number of unincorporated (annotated) names left in spreadsheet

Feed back on procedures and recommendations for a "best practice"

Appendix 7. Special Clause 10 – Addition of Third Parties

The wording for Special Clause 10 is provided below (full document can be found at: ftp://ftp.cordis.europa.eu/pub/fp7/docs/fp7-ga-clauses-v8_en.pdf)

"The following third parties are linked to Species 2000 (*"beneficiary n° 9"*)

2. This *beneficiary* may charge costs incurred by the above-mentioned third parties in carrying out the *project*, in accordance with the provisions of the *grant agreement*. These contributions shall not be considered as receipts of the *project*.

The third parties shall identify the costs to the *project* mutatis mutandis in accordance with the provisions of part B of Annex II* of the *grant agreement*. Each third party shall charge its eligible costs in accordance with the principles established in Articles II.14 and II.15. The *beneficiary* shall provide to the *Commission*:

- an individual financial statement from each third party in the format specified in Form C. These costs shall not be included in the *beneficiary*'s Form C

- certificates on the financial statements and/or on the methodology from each third party in accordance with the relevant provisions of this *grant agreement*.

- a summary financial report consolidating the sum of the eligible costs borne by the third parties and the *beneficiary*, as stated in their individual financial statements, shall be appended to the *beneficiary*'s Form C.

When submitting reports referred to in Article II.4, the *consortium* shall identify work performed and resources deployed by each third party linking it to the corresponding *beneficiary*.

3. The eligibility of the third parties' costs charged by the *beneficiary* is subject to controls and audits of the third parties, in accordance with Articles II.22 and 23.

4. The *beneficiary* shall retain sole responsibility towards *the Union* and the other beneficiaries for the third parties linked to it. The *beneficiary* shall ensure that the third parties abide by the provisions of the *grant agreement*.

*Article II referred to within the wording of the Special Clause 10, refers to the Grant Agreement - Annex II General Conditions (see at: ftp.cordis.europa.eu/pub/fp7/docs/fp7-ga-annex2-v5_en.pdf).