

# i4Life



## Indexing for Life

### Deliverable 2.3

## Workshop 1: Policy Report on the CoL interaction with the GNA

Work package 2

Richard J. White<sup>1</sup>, Magdalena H. Sitko<sup>2</sup>, Andrew C. Jones<sup>1</sup>, Kwok Cheung<sup>2</sup>, Alastair Culham<sup>2</sup>

<sup>1</sup>School of Computer Science & Informatics, Cardiff University

<sup>2</sup>School of Biological Sciences, University of Reading

15 February 2012

Capacities Programme of Framework 7: EC e-Infrastructure Programme – Virtual Research  
Communities - INFRA-2010-2

Grant Agreement No:	261555
Project Co-ordinator:	Dr Alastair Culham
Project Homepage:	<a href="http://www.i4Life.eu">http://www.i4Life.eu</a>
Duration of Project:	36 months
Start Date:	1 November 2010
End Date:	31 October 2013



## i4Life project WP 2 Deliverable 2.3

# Workshop 1 Policy Report on the Catalogue of Life interaction with the Global Names Architecture

Richard J. White<sup>1</sup>, Magdalena H. Sitko<sup>2</sup>, Andrew C. Jones<sup>1</sup>, Kwok Cheung<sup>2</sup>, Alastair Culham<sup>2</sup>

<sup>1</sup>School of Computer Science & Informatics, Cardiff University

<sup>2</sup>School of Biological Sciences, University of Reading

## Contents

1.	Introduction .....	2
1.1.	The workshop background and goals.....	2
1.2.	Short overview of the Catalogue of Life.....	3
1.3.	Short overview of the Global Names Architecture (GNA).....	5
1.4.	Glossary.....	7
2.	Recommendations of the workshop .....	10
2.1.	Group A: wide strong support .....	10
2.2.	Group B. second tier, widely supported.....	11
2.3.	Group C. supported by individuals .....	13
2.4.	Group D. least support .....	13
3.	Additional information .....	14
3.1.	Further observations.....	14
3.2.	Existing resources, systems and components mentioned during the workshop.....	14

## 1. Introduction

Indexing for Life (i4Life) is a European e-Infrastructure project, co-funded by the European Commission's Seventh Framework Programme for Research and Technological Development. i4Life is to facilitate communication among the biodiversity informatics community of developers and users. The project was launched at the University of Reading on 1st November 2010. This project sits centrally in several areas: societal concerns for global biodiversity and the functionality of portals showing species diversity, and the provision of an electronic catalogue of organisms suitable for use in biodiversity informatics and the virtual laboratory for modelling that global biodiversity. i4Life addresses directly the ability of scientists to comprehend the scale of species diversity and to enumerate its components.

### 1.1. The workshop background and goals

For extending the taxonomic knowledge backbone, data and knowledge need to be shared and the processes that manage them need to be unified. Therefore within the scope of the i4Life project several workshops on planning workflows and protocols between various

biodiversity organisations were planned. These will include a series of design and specification workshops, and testing sessions to address policy and implementation issues on some of the scientific challenges to sharing a taxonomy between these communities.

Most types of data relating to living organisms are arranged by scientific name. In principle almost all information about species is tied to a name. One of the oldest and still the primary standard for naming species is the system of binomial nomenclature. Although taxonomy involves subjective judgements based on advancing knowledge and is therefore not stable, the procedures associated with establishing scientific names are objective and stable.

What is still needed for understanding biodiversity is a cumulative tool or platform that would gather or interlink all key information about species. As never before, modern technology is giving us an opportunity to combine and link data from many different sources. Therefore various initiatives have emerged in an attempt to provide an infrastructure for such linkage to be possible. One of those is the Global Names Architecture, the initiative that builds on emerging standards of TDWG, GBIF and EOL. The Global Names Architecture goal can be summarised as the establishment of a taxonomic framework for inventorying, reconciling and organising all information linked to species names.

The present document is the report on the output from the first of the planned i4Life workshops, which was related to the Global Names Architecture initiative. The main objective of the workshop was to establish links between the CoL and the GNA communities, and to develop recommendations for mutually beneficial collaboration, data sharing and tool development.

## **1.2.Short overview of the Catalogue of Life**

The Catalogue of Life is produced by the Species 2000 programme, in partnership with the North American organisation the Integrated Taxonomic Information System (ITIS) since June 2001. Both organisations remain separate and different in structure, however, by working together in creating a common product, the partnership has enabled them to reduce duplication of effort, make better use of resources, and to accelerate production. The goal of the Species 2000 & ITIS Catalogue of Life project (CoL) is to enable users throughout biological and biodiversity sciences, and across the many scientific and non-scientific disciplines that use organism information, to access data about all organisms by means of a species checklist and a taxonomic hierarchy

The objective of the CoL is to create a validated checklist of all the world's species (plants, animals, fungi and microbes). This is being achieved by bringing together an array of "Global Species Databases" (GSDs) covering each of the major groups of organisms. As described below, taxonomic editorial processes are involved in CoL production at three levels: within the taxonomic publications used as sources by the GSDs, within the GSDs by their own experts, and within the CoL by selecting and endorsing the GSDs that are used and by managing the integration of their separate checklists. Bringing the CoL to production scale was funded by the EC as one of its scientific infrastructures (EuroCat), with further funding by the Japanese Government, the US Government (through ITIS) and GBIF.

Rapid progress has been made recently within the 4D4Life project funded by the European Commission. Along with many developments and infrastructure improvements, there is also significant progress towards filling the gaps in the CoL and improving the quality of the databases already part of the CoL. The eleventh edition of the Annual Checklist contains 1,379,178 species. This is probably just slightly over 2/3 of the world's known species. This is a significant achievement, but this means that for many groups it continues to be deficient, and users will notice that many species are still missing from the Catalogue.

The most recent version of the Catalogue is compiled with sectors provided by 109 taxonomic databases from around the world. Many of these contain taxonomic data and opinions from extensive networks of specialists, so that the complete work contains contributions from more than 3,000 specialists from throughout the taxonomic profession. Species 2000 and ITIS teams peer-review databases, select appropriate sectors and integrate the sectors into a single coherent catalogue with a single hierarchical classification. It is planned to introduce alternative taxonomic treatments and alternative classifications, but an important feature is that for those users who wish to use it, a single preferred catalogue, based on peer reviews, will continue to be provided.

The Catalogue of Life is now a single product with several updates during a year.

The **Annual Checklist** of the Catalogue of Life is published each April online, as web-services, and on DVD. At all times the latest Annual Checklist is online at [www.catalogueoflife.org/annual-checklist](http://www.catalogueoflife.org/annual-checklist). This is a fixed edition that is published, archived and deposited in public libraries, that can be cited, and that can be used as a common catalogue for comparative purposes by many organisations. Copies are distributed free of charge on DVD, and identical copies are viewable and downloadable on the website. The archive edition for 2011 is at [www.catalogueoflife.org/annual-checklist/2011](http://www.catalogueoflife.org/annual-checklist/2011). Archived editions can be found by changing the year in the URL (e.g. /2010, /2009 etc.), and downloadable versions and web-services can be found at [www.catalogueoflife.org/services](http://www.catalogueoflife.org/services).

**The Catalogue of Life** “around the year” editions are progressively enhanced releases in what will become a dynamically developing system, made available online and as web-services. The present frequency is quarterly, with releases each year in summer, autumn, winter and on 1st April as an Annual Checklist. At all times the latest release is available for online access at [www.catalogueoflife.org/col](http://www.catalogueoflife.org/col). Archives of previous releases can be downloaded from [www.catalogueoflife.org/services](http://www.catalogueoflife.org/services).

The Annual Checklist has become well established as a cited reference used for data compilation and comparison. The Catalogue of Life is already used by the Global Biodiversity Information Facility (GBIF) and the Encyclopedia of Life (EoL) as the principal taxonomic index in their web portals; the CoL is also recognised by the CBD. Close cooperation with further global biodiversity programmes (EMBL-EBI, BOLD, IUCN Red List, LifeWatch) is now being developed within the scope of the i4Life project.

Sources of information about the CoL and related projects:

<http://www.catalogueoflife.org/col>

<http://www.sp2000.org>

<http://www.itis.gov/>

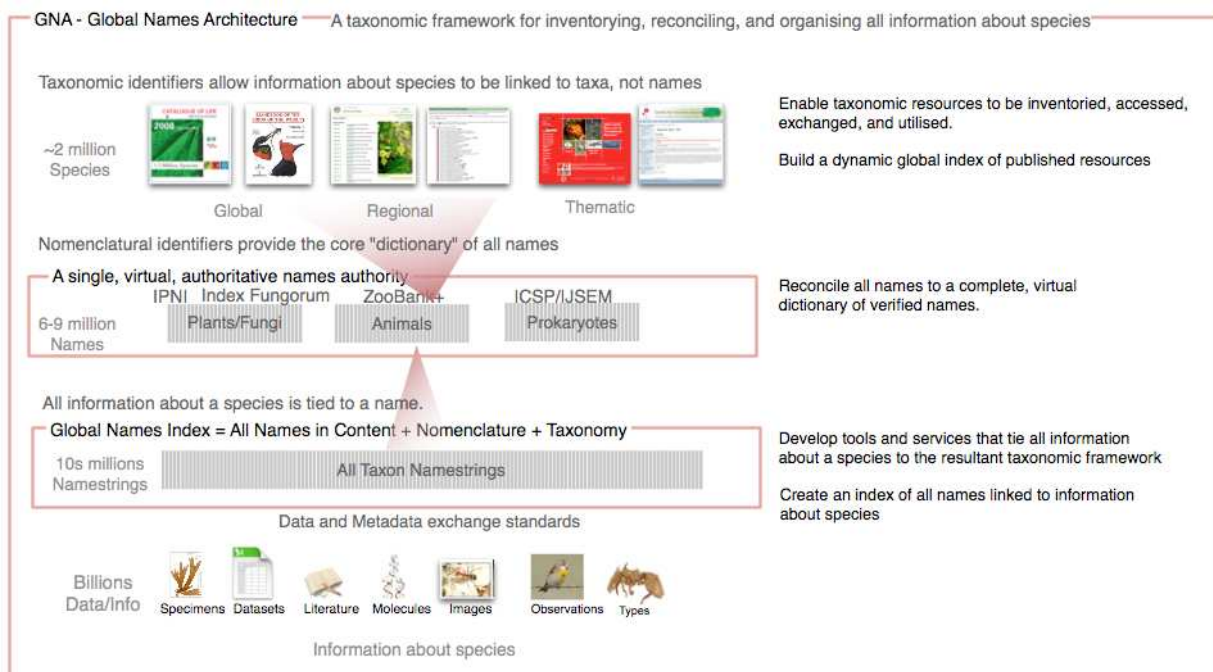
<http://www.i4life.eu/>

<http://www.4d4life.eu/>

### 1.3.Short overview of the Global Names Architecture (GNA)

Taxonomy provides the logical basis for an organisational framework for all information about a species. The GNA seeks to develop a key part of that framework. The GNA is designed to provide a framework for comprehensive integrated access to metadata relating to all names, especially scientific names, of organisms. It is planned to be a virtual data and service layer that capitalises on the almost universal use of latinised scientific names that annotate almost all information relating to organisms, allowing them to be used as pivot points to cross-link online datasets. The GNA will facilitate the use of names as a system of metadata applicable to all described species, and their arrangement in taxonomic or phylogenetic hierarchies as ontologies. The GNA framework is taxonomically neutral, i.e. it does not decide in favour of one taxonomic opinion rather than another. It proposes to make name metadata more accessible which may greatly assist taxonomic revisions and classifications. Both the GNA and CoL frameworks will facilitate the construction of systems which use computational processes to bring together information related to biodiversity.

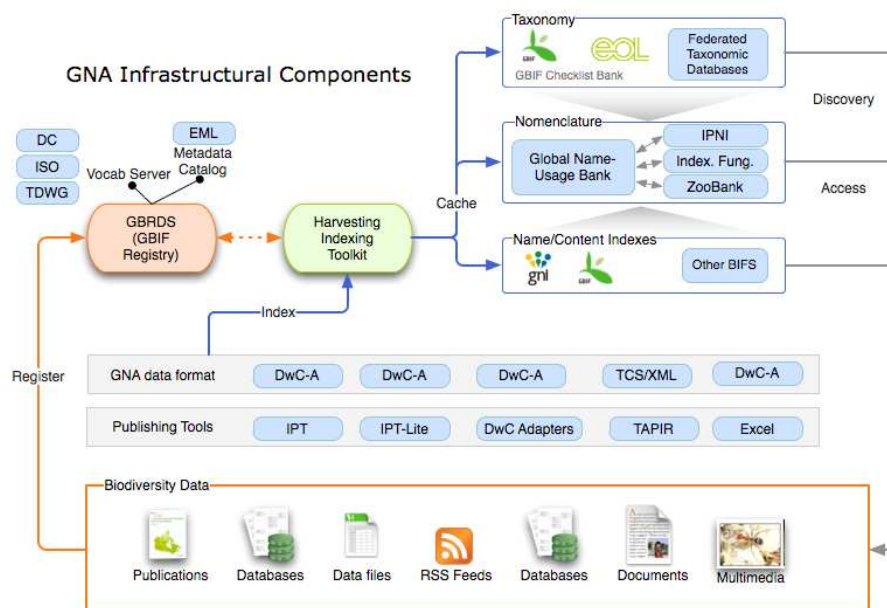
The diagram below is published at: <http://code.google.com/p/gbif-ecat/wiki/LogicalComponents>



## Physical Architecture of the GNA

The diagram below is published at:

[http://code.google.com/p/gbif-ecat/wiki/GNAOverview#Data\\_Sources](http://code.google.com/p/gbif-ecat/wiki/GNAOverview#Data_Sources)



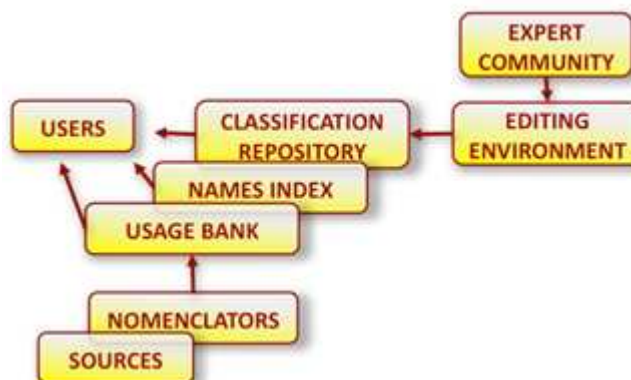
## Components of the global names architecture developed or under development:

- **Global Names Index (GNI).** The names index provides a list of all names that have been used for organisms. Within this list lie all of the nomenclaturally correct names, all of the names that are accepted as tokens for taxa, and all of the taxonomic metadata for biodiversity informaticians.

GNI is not intended for people to visit, but is set up to provide information through web services to other machines.

- **Editing environment (GNITE).** The editing environment will provide access to classifications that people are willing to share. They will be deposited in a classification and list repository.
- **Usage Bank (GNUB).** GNUB, the Global Names Usage Bank, will be the core of GNA. GNUB will be the means by which we index and organise our recorded knowledge that we have about life on earth that is available on-line in a digital format. Zoobank is a specialist prototype of GNUB that is already online at [zoobank.org](http://zoobank.org).
- **GNA CLR - Classification and list repository**
- **Technical standards and specifications**

(Taken from: <http://www.globalnames.org/Components>)



Sources of information about the GNA:

<http://www.globalnames.org>

<http://gni.globalnames.org/>

[http://code.google.com/p/gbif-ecat/wiki/GNAOverview#Data\\_Sources](http://code.google.com/p/gbif-ecat/wiki/GNAOverview#Data_Sources)

## 1.4.Glossary

### **Binomial nomenclature**

A formal system on naming species of living organisms by giving each a name consisting of two parts in Latin grammatical form. These parts are the genus and the specific epithet, both forms are italicised and the Genus name is capitalised.

### **Biodiversity Heritage Library**

The Biodiversity Heritage Library (BHL) is a consortium of natural history and botanical libraries that cooperate to digitise and make accessible the legacy literature of biodiversity held in their collections and to make that literature available for open access and responsible use as a part of a global “biodiversity commons.” BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL). [www.biodiversitylibrary.org/](http://www.biodiversitylibrary.org/)

### **Classification**

Definition of E.W. Mayr: “The arrangements of entities in a hierarchical series of nested classes, in which similar or related classes at one hierarchical level are combined comprehensively into more inclusive classes at the next higher level”.

### **DwC-A - Darwin Core Archive**

This is a draft standard that allows the transfer of information from one machine to another. It takes the form of a set of text files with a simple descriptor to provide information as to how the files are organised. Data files are arranged in a star-like manner, with one core data file (complying with the Darwin Core standard, itself an extension of Dublin Core) surrounded by any number of ‘extensions’. Each extension record (or ‘extension file row’) points to a record in the core file; in this way a DwC-A package might contain a list of species as one file, to which is attached another file with occurrence information, or further files dealing with other types of information. This is GBIF’s preferred data exchange standard and is integrated with their IPT environment.

The DwC-A was adopted as a standard across the i4Life project for exchanging information between the CoL and its partners (GBIF, EMBL-EBI, IUCN RedList, Barcode Initiatives, LifeWatch). [www.gbif.org/informatics/standards-and-tools/publishing-data/data](http://www.gbif.org/informatics/standards-and-tools/publishing-data/data)

### **EDIT - The European Distributed Institute of Taxonomy**

The European Distributed Institute of Taxonomy (EDIT) is the collective answer of 29 leading European, North American and Russian institutions to a call of the European Commission, issued in 2004, for a network in “Taxonomy for Biodiversity and Ecosystem Research”. Among the members of EDIT are the premier natural history collections-based institutions worldwide, which have both the management capacity and the will to progress toward EDIT’s objectives. Their collections are global in coverage and are supported by complementary expertise. More than half of the world’s natural history specimens, which constitute the large-scale infrastructure for taxonomic research, are held in the repositories of EDIT’s membership. The inclusion in the consortium of network institutions devoted to

management of biodiversity data, and of a research organisation directly related to users of taxonomy for agriculture and environment, in addition to the links or inclusion of many partners in universities, will facilitate dissemination of EDIT taxonomic research and training toward a wide audience. [www.e-taxonomy.eu/](http://www.e-taxonomy.eu/)

### **EoL - The Encyclopedia of Life**

EoL is a free, online collaborative encyclopedia intended to document all living species known to science. It is compiled from existing databases and from contributions by experts and non-experts throughout the world. It aims to build one “infinitely expandable” page for each species, including video, sound, images, graphics, as well as text. In addition, the Encyclopedia incorporates content from the Biodiversity Heritage Library, which digitises millions of pages of printed literature from the world’s major natural history libraries. The EoL is using the Catalogue of Life as a taxonomic backbone. [www.eol.org](http://www.eol.org)

### **GBIF - the Global Biodiversity Information Facility**

GBIF is supported via governments worldwide and co-ordinated through the United Nations to be a leading clearing house of biodiversity information. GBIF enables free and open access to biodiversity data online to all and anyone, for scientific research, conservation and sustainable development. GBIF provides three core services and products:

1. An information infrastructure – an Internet-based index to a globally distributed network of interoperable databases that contain primary biodiversity data – information on museum specimens, field observations of plants and animals in nature, and results from experiments – so that data holders across the world can access and share them.
2. Community-developed tools, standards and protocols – the tools data providers need to format and share their data.
3. Capacity-building – the training, access to international experts and mentoring programmes that national and regional institutions need to become part of a decentralised network of biodiversity information facilities.

GBIF is one of the partners in the i4Life project and is using the Catalogue of Life as a taxonomic backbone. [www.gbif.org](http://www.gbif.org)

### **GNI - Global Names Index**

The Global Names Index is the first component of a semantic environment for biology called the Global Names Architecture (GNA). GNI has been developed by GBIF and the EOL. It is an index of name strings in the broadest sense (including code-compliant scientific names, vernacular names, surrogates, and erroneous forms of names), with links to environments that have the names and to the associated data or metadata associated with names. GNI is a core element of the GNA. This is a rather raw list of names which, to become useful, have to be reconciled so that alternative names for the same entity are linked together, disambiguated so that identically spelled names for different entities can be discriminated, classified for navigation and to serve the needs of taxonomists, and flagged to indicate the status of the name. <https://github.com/dimus/gni>

### **GNITE – Global Names Interface for Taxonomic Editing**

The Global Names Interface for Taxonomic Editing– an interface that will allow individual taxonomists to improve the underlying names-based infrastructure. Through this interface (under development) taxonomists will be able to review the names that are known to GNA



and add missing ones, they will be able to edit, correct and annotate names, they will be able to merge, split and edit reconciliation groups, they will be able to edit and improve disambiguation of homonyms and chresonyms, and they will be able to merge classifications, edit classifications and create new classifications.

### **GNUB - Global Names Usage Bank**

An environment to record the occurrence of names in documents, databases, notes or other records. It is the central component of the GNA. The usages will include all nomenclatural acts, making GNUB critical to nomenclators and emerging registration environments such as Zoobank. GNUB will index usages that help to clarify the meaning of each name and so contains resources that will be developed into taxonomic tools and services. <http://www.globalnames.org/GNUB>

### **Index Fungorum**

The Index Fungorum, the global fungal nomenclator coordinated and supported by the Index Fungorum Partnership, aims to include all names proposed for fungi (including yeasts, lichens, chromistan fungi, protozoan fungi and fossil forms) since 1753 at species level and below. [www.indexfungorum.org](http://www.indexfungorum.org)

### **ITIS - Integrated Taxonomic Information System**

The ITIS mission is to create a scientifically credible database of taxonomic information on plants, animals, fungi, and microbes, placing primary focus on taxa of interest to North America, with world treatments included, as available. Since 2001 ITIS and Species 2000 are partners developing together the Catalogue of Life. [www.itis.gov](http://www.itis.gov)

### **LifeDesks**

LifeDesks is a content management system for biodiversity that allows taxonomists to create web sites for biodiversity without incurring the cost of developing the software. Hosted content is made semantically ready so that users can chose to make information openly available to other semanticised web-based biodiversity projects, such as the Encyclopedia of Life. LifeDesks is being developed by EOL in concert with the EU EDIT Scratchpads. <http://scratchpads.eu/about>. [www.lifedesks.org](http://www.lifedesks.org)

### **Taxonomy**

The science of identifying and naming species, and arranging them into a classification.

### **TDWG - Biodiversity Information Standards**

This is the new name for the TDWG organisation that was originally referred to as the Taxonomic Databases Working Group – the source of many standards for biodiversity information transfer. TDWG is a not-for-profit scientific and educational association, affiliated to the International Union of Biological Sciences, formed to establish international collaboration among biological database projects so as to promote the wider and more effective dissemination of information about the world's heritage of biological organisms for the benefit of the world at large. Its goals are:

- To develop, adopt and promote standards and guidelines for the recording and exchange of data about organisms
- To promote the use of standards through the most appropriate and effective means and

- To act as a forum for discussion through holding meetings and through publications.

[www.tdwg.org](http://www.tdwg.org)

### **ZooBank**

ZooBank is intended as the official registry of Zoological Nomenclature, according to the International Commission on Zoological Nomenclature (ICZN). <http://zoobank.org/>

## **2. Recommendations of the workshop**

Two and a half days of deliberations on 8-10 November 2011 at the Catalogue of Life – Global Names Architecture workshop in Paris were distilled into thirty-one potential recommendations. The number and diversity of these recommendations reflect the range of complementary interests and expertise of the workshop delegates, and for that reason we have not attempted to reduce their number or filter them; instead we subjected them to a ranking procedure.

The recommendations were then circulated to the members of the group. These potential recommendations were prioritised by voting after the meeting. Many of the suggestions had been put forward as strong recommendations by individuals, many of whom also supplied further comments along with their votes.

All recommendations are listed in decreasing order of the combined priority calculated from the priority preferences expressed by each voting member of the workshop. We have divided them into four groups A to D in order to describe the priority assigned to them. Groups A and B received wide support from many individuals. Group C received strong support from a minority of members. Group D did not receive strong support from anyone, or received a third place from one member and no support from anyone else; the last recommendation received no votes (this could be an artefact, because members were asked not to vote for more than ten recommendations; it had been received favourably at the meeting and has merit).

### **2.1. Group A: wide strong support**

The following recommendations were assigned the highest priority. They were given a high priority by six to nine members, and all of these recommendations have a priority which is higher than any recommendations in group B.

1. The development of “proto-GSDs” to fill gaps in the CoL from nomenclators and other sources should be fast-tracked: prioritise the gaps to be filled, with fast entry of high-value missing names, based on the needs of priority users (e.g. IUCN), impact and search frequency; cooperate and consolidate proto-GSD generation to avoid too much fragmentation; develop a mechanism for the incorporation of proto-GSDs into the CoL (annotated as not yet taxonomically vetted); stress the urgency at the December 2011 CoL Global Team meeting (given an agreed mechanism, there are strong possibilities to fill gaps in the fungi during April/May/June 2012 and perhaps also in arthropods).

2. Names in the Catalogue of Life should be matched to names in nomenclators and be labelled with their name identifiers; to the extent that related names in nomenclators are grouped, this will reduce reliance on fuzzy matching when cross-mapping checklists. This cross-mapping can also identify names in nomenclators that relate to taxa which are not yet in the CoL, and thus assist in the creation of “Proto-GSDs”.
3. A programme for the further development of cross-mapping tools should consider options for their design and operation, including the ability for taxon specialists to refine the resulting cross-maps and the use of the tools and cross-maps in other software including the tools for handling multiple checklists being developed by the Species 2000 China node.
4. Draw up a roadmap of tasks achievable in the near future (and later, with funding): identify what the GNA can provide to the CoL *via* i4Life and 4D4Life and *vice versa*; develop a strategy for progressing with complementary and overlapping tools; identify workflows and commonalities between projects; consider funding possibilities and the appropriate time-scales and requirements for deliverables.
5. An agreed common approach to track taxonomic and nomenclatural changes in the CoL and other checklists is needed: the CoL should track name and taxon concept identifiers over time (including both those it generates and any it receives) and provide services to answer questions such as “what has changed in taxon (group) X since date Y?”, “has any data linked to this set of identifiers changed since date Y?”; to complement this, alert systems and synchronisation strategies for GSD environments are required.
6. Introduction of persistent resolvable identifiers for names and taxa: establishment of a shared id system for all elements; agreement on a standard format & schema for resolver output.
7. Feedback and communications channels between nomenclators and GSDs need to be designed and established. Decisions are needed on the direction of communication channels and which organisations manage them; their roles may include the transfer (“piping”) of new names and associated identifiers between nomenclators and registration systems (GNUB) to GSDs, ideally with the CoL as a broker. The use of suitable identifiers will allow information about homotypic groups of names to be passed but will require changes to GSDs’ data sets and CoL’s data schema.
8. It is important to maintain provenance metadata to represent the chain of providers through which the data has passed, as is being modelled in the GBIF Registry; credit for at least the first and last steps in the chain was suggested (but the CoL already uses a three-step model where the first step in the chain is the expert performing taxonomic scrutiny); there are potential uses for this metadata for data quality assessment.

## 2.2.Group B. second tier, widely supported

The following recommendations were given a high priority by four or five members (or in one case three members), and all of them have a priority lower than Group A and higher than group C.

9. Measures to assist improvement of established GSDs: request feedback from GSDs about how to fulfil their needs; identify potential data resources & software tools to help GSDs to fill

- gaps and to use, propagate and if necessary create name and concept identifiers.
10. Clarify and encourage the use of data sharing formats and protocols (e.g. Darwin Core extensions with metadata, etc.): we need data-sharing tools to use these formats, and to build name “reconciliation groups” (e.g. to assist cross-mapping where names referring to the same concepts may vary in presentation).
  11. Additional name status values are desirable for GSDs to “grade” the status of names: status values are needed to indicate dubious names, names of taxa that don’t belong in the GSD to which they appear to relate, and names that have not yet been assessed by the relevant GSD specialists.
  12. Multiple classifications for higher taxa should be made visible in the CoL: discussion and collaboration is required between CoL and GBIF; this issue may need to be resolved in more than one way because there are multiple user groups; different classifications from various sources need to be reconciled and unified where they do not disagree (perhaps using the Checklist Bank); users should be able to choose their preferred hierarchy, or accept a recommended default.
  13. Techniques for improving proto-GSDs need to be investigated: fuzzy matching to store associations between nomenclator IDs in an “almost-same-as” repository; need better animal nomenclators; possibly make higher taxon lists (perhaps generated by GN/IRMNG) available on-line & invite users to annotate with links to nomenclators; mechanisms to compare proto-GSDs with other available resources such as national and regional checklists.
  14. There is a need to take the lead in advocating the adoption of usable standards in the biodiversity community: this will require clarifying, refining and consolidating existing standards; common Web Service operations for common tasks; extended vocabularies for vernacular name status and other categories; a unified architecture and framework for data, especially the core data which needs to be shared for constructing proto-GSDs.
  15. Taxonomic tree tools are required to work with alternative classifications: for example, to build consensus views of the classification hierarchy and merge identical sub-trees.
  16. Taxon look-up services for user-supplied sets of name strings: for example, services which return the taxa the supplied names match, with confidence levels and persistent identifiers, or which return the taxonomic tree subset covered by the names (in a given existing classification, i.e. all the relevant higher taxa).
  17. Include fossils in the Catalogue of Life
  18. Name look-up services for user-supplied sets of name strings: for example, services which return the matched name records (including all nomenclatural data and citations) with confidence levels and persistent identifiers; also a service to create identifiers for name strings that are not already listed.
  19. Increase the availability of common names: sources include GSDs, regional hubs, GBIF, OpenUp; align common names to GSD taxa in the Catalogue of Life; frequently used cultivar

and trade names can be treated as common names.

### **2.3.Group C. supported by individuals**

The following recommendations were given a high priority by only one or two members, but each recommendation was given the highest priority (or in one case the second highest priority) by one member.

20. Recommendations to improve data availability: a preference for registered holders of name and concept data to make their data freely available, with appropriate handling of IPR issues; specific interchange activities; methods to embed taxonomic information within biodiversity informatics frameworks and biodiversity information systems.
21. Other user services proposed to meet user needs, which were not specified in detail.
22. Inventory of data resources and software tools: identify and develop registries for existing and planned data resources, software tools and frameworks; a resource discovery mechanism (e.g. GBIF registry), resource-level metadata and vocabularies for access and description of resources.
23. A review by CoL of its priorities and scope for sources of names and how to obtain and present them: its primary role should be the taxonomic validity of names; supply taxonomic data (relationships of new names) to GNA; understand the CoL user base & requirements; prioritise resource allocation accordingly; make the CoL more of a process than a product, open it up for new concepts.
24. Use the CoL more extensively as a link-out point to other sources of information.

### **2.4.Group D. least support**

The following recommendations were not given a priority higher than third by any member and most were supported by only one member.

25. Services for name relationships (in taxa): discovery services for synonymy relationships; transregional homonym disambiguation.
26. Open up data for user curation ("crowd sourcing"): editing environments and their functionality; how to handle edited versions and the implications for GSDs.
27. Competition and overlap between GSDs: public review process of GSD selection; an inclusive policy – "offering seats at the table" for experts; soften the authority of any individual GSD.
28. Three-level user interface: user-supplied names are searched for (i) in the CoL (GSDs, proto-GSDs), (ii) in nomenclators, (iii) in the GNI; investigate how this would be presented in products; implemented as a separate portal; make dirty name data available through the Catalogue, perhaps by third parties rather than by the CoL].
29. "Same-as" services for name and taxon IDs, for users or systems to discover alternative IDs for the same entities, perhaps provided by nomenclators or third parties rather than by the CoL].
30. Users to be able to add or edit comments to provide feedback on taxa in the CoL.

31. Images and links to images and other media resources in CoL: use EoL as an aggregator for rich media.

### 3. Additional information

#### 3.1. Further observations

Workshop participants made a number of observations both during and shortly after the workshop, which should be borne in mind when considering their recommendations. The participants at the meeting felt that the group should continue to exist in some form to act as a forum for the discussion and coordination of name-related issues.

One participant felt that the guiding principles for future actions should be to “move towards completion of the list of all species, and [to] be as inclusive of players and consumers as we can”, expanding this to say that we should “minimise the differences, improve dynamic interactions (content synchronisation services, co-operative software development). We need to understand duplications, protect redundancy where we think that is valuable, but working in tight collaboration where we feel duplication has little benefit.” It was also pointed out that there are some present or potential future overlaps between GNA and i4Life (and other CoL-related projects), including GNITE, Multi-hub and Cross-mapping tools, and the Chinese tools, including TTT (Taxonomic Tree Tools).

An important point perhaps not clearly spelled out in the Recommendations is that no name or taxon concept or relevant metadata should ever be deleted after it has been incorporated into an electronic system, for that would “change history” or create broken links. Subsequent actions

David Eades brought the draft document to the attention of the Catalogue of Life Global Team at their meeting in Los Baños, Philippines (6-8th December 2011), where it was given preliminary consideration. Further discussion will take place at the CoL Team meeting in Reading (25-26th March 2012). We trust that David Patterson and others will convey it appropriately to the Global Names US project members. Nicky Nicolson contributed an earlier draft of this document to a discussion of priorities at the Royal Botanic Gardens, Kew. A meeting of the US GN project participants in Hawaii is planned for March 2012 and will be attended by some CoL Global Team members.

#### 3.2. Existing resources, systems and components mentioned during the workshop

This is not intended as a complete list, but may be useful.

- ZooBank
- CiteBank
- IRMNG
- Index Fungorum, Dictionary of the Fungi (10th edition, 2008)
- Kew (plant) 'Families' book, list of Genera (Dick Brummit), IPNI
- Checklists PESI, WoRMS, etc.
- uBio Name Bank and Classification Bank (no longer curated)

- GBIF IPT, Checklist Bank, Registry
- GNA
  - GNI
  - GNUB
  - GN-US
    - Classification List Repository
    - GNRD (Recognition & Discovery)
    - Union (interim strictly holophyletic classification)
- CoL China node pilot project and tools
  - EOL China
  - Species 2000 China node
  - Checklist warehouse
  - Species Data Management Environment
  - Taxonomic tree tools
- Content construction platforms
  - ScratchPads
  - EDIT CyberPlatform
  - APHIA
  - NZOR
  - C-INT
  - Species File software
  - GNITE