

Cover Page

**Part B**

Proposal Full Title

# **Distributed Dynamic Diversity Databases for Life**

Proposal Acronym

**4D4Life**

Call Identifier

**FP7 – INFRASTRUCTURES – 2008 - 2**

Type of Funding Scheme

**Combination of Collaborative Project and Coordination and  
Support Action (I3)**

**CCPCSA**

Work Programme Topics Addressed

**INFRA-2008-1.2.2: Scientific Data Infrastructure**

Name of the Coordinating Person

**Frank Bisby**

# List of Participants

<b>Participant No.</b>	<b>Participant organisation name</b>	<b>Short name</b>	<b>Country</b>
<i>Main Members of the Work Packages</i>			
1 (Coordinator)	University of Reading	UREAD	UK
2	Botanic Gardens Conservation International	BGCI	UK
3	Museum National d'Histoire Naturelle	MNHN	FR
4	Narodni Museum	NMP	CZ
5	Stichting Expertisecentrum voor Taxonomische Identificatie	ETI BioInformatics	NL
6	Cardiff University	CU	UK
7	Stichting National Natuurhistorisch Museum Naturalis	NMNH	NL
8	Species 2000	SP2000	UK
9	Vlaams Instituut Voor De Zee Vzw	VLIZ	BE
<i>Members of the GSD Network (WP3)</i>			
10	Staatliches Museum für Naturkunde Stuttgart	SMNS	DE
11	Natural History Museum	NHM	UK
12	CAB International	CABI	UK
13	Royal Botanic Gardens Kew	RBGK	UK
14	Universität Wien	UNI WIEN	AT
15	Universiteit Utrecht	UU	NL
16	National University of Ireland, Galway	NUIG	IE
17	Leibniz-Institut für Meereswissenschaften an der Universität Kiel	IMUK	DE
18	Deutsches Krebsforschungszentrum	FZK	DE
19	Botanische Staatssammlung München	BSM	DE
20	Muzeum i Instytut Zoologii PAN	MIZPAN	PL
21	Oberösterreichisches Landesmuseum	OOE	AT
22	National Museum Wales	NMGW	UK
23	Institut de Recherche pour le developpement	IRDF	FR
24	Consejo Superior de Investigaciones científicas	CSIC	ES
25	TSJ BVBA	TSJ BVBA	BE
26	Staatliche Naturwissenschaftliche Sammlungen Bayerns	SNSB, Munich	DE
27	Universita Degli Studi Di Padova	UNIPD	IT
28	Universiteit Van Amsterdam	UVA	NL
29	Institut Royal Des Sciences	RBINS	BE

	Naturelles De Belgique		
30	The Chancellor, Masters and Scholars of The University of Oxford	University of Oxford	UK
31	Wageningen Universiteit	WU	NL
32	Royal Botanic Garden, Edinburgh	RBGE	UK
<i>Members of the Multi-Hub Network (WP4)</i>			
33	Chinese Academy of Sciences, Biodiversity Committee	CAS	CN
34	Centro de Referência em Informação Ambiental	CRIA	BR
35	Integrated Taxonomic Information System	ITIS	US
36	Commonwealth Scientific and Industrial Research Organisation	CSIRO	AU
37	Landcare Research - Manaaki Whenua	Landcare Research	NZ

## Table of Contents

Cover Page .....	1
Distributed Dynamic Diversity Databases for Life .....	1
Proposal Acronym .....	1
4D4Life .....	1
List of Participants .....	2
Table of Contents .....	3
PROPOSAL .....	4
Section 1: Scientific and technical quality .....	4
1.1 Concept and objectives .....	4
1.1.1 The General Problem being addressed .....	4
1.1.2 The Present Concept .....	5
1.1.3 The Existing Programme .....	6
1.1.4 Who uses the present services? .....	7
1.1.5 What are the present services used for? .....	8
1.2 Progress beyond the state-of-the-art .....	11
1.3 Methodology to achieve the objectives of the project, in particular the provision of integrated services .....	13
1.4 Networking Activities .....	16
1.5 Service Activities .....	30
1.6 Joint Research Activities .....	41
Section 2: Implementation .....	47
2.1 Management Structure and Procedures .....	47
2.2 Individual Participants .....	50
2.3 Consortium as a whole .....	69
2.4 Resources to be committed .....	69
Section 3: Impact .....	72

3.1 Expected impacts listed in the work programme .....	72
3.2 Dissemination and/or exploitation of project results, and management of intellectual property .....	75
Section 4; Ethical Issues .....	78
4.1 Data protection issues .....	78
4.2 Issues relevant to ICT .....	79
4.3 Other issues .....	79

## **PROPOSAL**

### **Section 1: Scientific and technical quality**

#### ***1.1 Concept and objectives***

##### **1.1.1 The General Problem being addressed**

Despite 250 years of effort in the taxonomic profession, there is still, in 2008, no complete catalogue of all presently known animals, plants, fungi and micro-organisms of the world. This is a critical problem for the scientific community, and for national, regional and global organisations that organise and regulate the exchange of biotic information and materials worldwide. The set of organisms known to science is a key dimension of human knowledge concerning global biodiversity, evolution, ecology, natural resources, and biotic response to climate change. It supplies a vital set of index terms needed to access most biodiversity knowledge. There is increasing public need and expectation, focussed through the UN Convention on Biological Diversity (CBD), to complete such a catalogue of all known organisms for international uses. Many commentators are surprised that a complete catalogue does not already exist. In fact it is a non-trivial task that is too large for the individual capabilities of even the largest taxonomic institutions, due to the distributed nature of the knowledge.

This proposal is made by the Species 2000 consortium that has made substantial progress with resolving this problem. It has created, maintained and enlarged the Species 2000 Catalogue of Life to the point where it now covers 1.1 million species of plants, animals, fungi and micro-organisms, some 60% of the anticipated total of 1.8 million presently known species worldwide. It has done this by employing a radical architecture of federating global sectors of taxonomic expert knowledge from a growing array of supplier databases, and integrating these into a single taxonomic hierarchy and species checklist. The distributed system harvests taxonomic knowledge provided and maintained by a community of supplier organisations in the taxonomic profession, combining work by the major taxonomic institutions with that of smaller networks and individuals. This process was brought to production scale by the EC EuroCat project funded as a scientific infrastructure under FP 5 (2003 – 2006) and further developed since then with funding from other sources, including the EC EDIT Network of Excellence in FP 6.

Over the last two years the programme has concentrated on extending and improving the scientific content of the Catalogue of Life, which is now a unique and scientifically valuable resource. However, it has come as a bonus to see the rising and

now substantial public usage in Europe and all over the world, including by GBIF and the Encyclopedia of Life, of what is presently an incomplete service. This proposal to the Capacities Programme provides us with a timely opportunity to develop a parallel focus on services. It will enable us to enrich the variety and technical sophistication of taxonomic services that are undoubtedly possible, exploiting the taxonomic resource that we are already building. The utility of these services will secure the sustainability of the whole programme into the future.

### 1.1.2 The Present Concept

The Species 2000 Catalogue of Life (henceforward ‘the Catalogue’) has a single purpose, 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. It is already used to access data such as organism relationships, ecology, DNA sequences, protection status, invasive properties or information in any one of a myriad of other data domains. Such a Catalogue needs to be:

- i) **comprehensive:** covering all known organisms in all groups;
- ii) **global:** organisms of the whole world, in terrestrial, freshwater and marine environments;
- iii) **validated:** a responsible, modern and professional globalised taxonomic view of the classification, supported by and embedded in the profession’s activities;
- iv) **accurate:** reflecting as accurately as is practical the detail of diversity of living organisms;
- v) **accessible to all:** a clear view of the taxonomy, eventually in multi-lingual presentation;
- vi) **available to all:** widely and freely available in a variety of forms; and
- vii) **dynamic:** updated for taxonomic changes though time, either continuously or annually.

To be effective in the many applications in which it is used, the classification and the naming of species and higher taxa must be as close to ‘agreed and correct’ as is possible in taxonomy. This means for each taxon either using a consensus system, or selecting and using consistently one of the competing classifications where



Figure 1. Species page in the Catalogue

alternatives are in wide use. Because alternative classifications have been used both today and in the past, users must be able to locate species known by other names (or concepts) in the Catalogue, and discover alternative names under which to access data on the internet or in other resources. Consequently synonymy and common names must be included for each species. As much as possible should be ‘concept-based’, a precision provided by some of the databases.

The dream is simple - to create a Catalogue that contains an accurately maintained synonymic species checklist covering all known species, connected in a validated taxonomic hierarchy.

### 1.1.3 The Existing Programme

The present Catalogue of Life Programme, led by the global Species 2000 organisation based in Reading, and working with the N. American organisation ITIS, was set up as an international programme at a UK-funded (BBSRC) workshop in 2001. Bringing the programme up 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. Output is via the *Catalogue of Life Annual Checklists* on CD [2], and on the web [3], and the *Dynamic Checklist* on the web [4], both also available as web-services for electronic use.

In March 2007 an EC-funded ‘Million Species Day’ symposium was held to celebrate reaching one million species. The 2008 *Annual Checklist* now provides a quality species checklist of 1,105,589 species with unique identifiers and a hierarchy for all organisms (animals, plants, fungi, chromista, protozoa, bacteria, archaea, viruses). The estimate for the number of known extant species is currently 1.75 – 1.8 million [1]. The present Catalogue benefits from simplicity of structure incorporating minimal but standardised data for each species. These contribute to its success in providing a universal baseline needed by all biologists, and in making the project practicable. It consists of two knowledge structures, and software that enables the user to search or traverse them, and to toggle between them. i) The Species

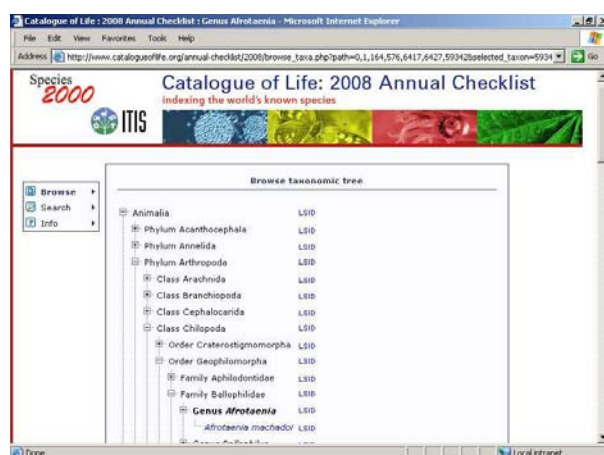


Figure 2. The Taxonomic Hierarchy

Checklist is a series of Species pages (Figure 1) that are located by name searches, with automated synonymic indexing. Each page gives the Standard data for a Species, including common names, the higher taxa it belongs to in the hierarchy, and geographical distribution. ii) The Taxonomic Hierarchy (Figure 2) is an expansible tree that can be followed down through the classification to the 1.1 million individual species. Or it can be used to navigate upwards to the higher taxa containing the one that is viewed. By clicking on a higher taxon listed on a species page, the user can transfer to the tree for that taxon, and see all its daughters. Conversely, by clicking on a species at a twig in the tree, the user can visit the relevant Species page in the Checklist.

A comprehensive checklist cannot be made simply by adding together regional or single-country lists. Different classification and naming schemes mean that a simple additive list would be massively duplicative and of little use. The current system is a successful development of the original BBSRC SPICE project. It federates the sector checklists provided by an array of distributed databases (GSDs), which are globalised checklists of a whole taxon, harvested across the Internet, and fitted together ‘end-to-end’ within a single overall classification. When enough sectors are fitted this process can eventually create a complete list. The number of GSDs contributing one or more taxonomic sectors to the Catalogue reached 52 for the 2008 edition, including 37 based in Europe, 11 in the USA, and 4 in Russia, Japan,

Australia and the Philippines. The model ensures that sectors are enhanced taxonomically by the supplier databases, and ca. 3,000 experts globally contribute to these databases.

Each GSD sector is attached at its ‘top point’ (its highest ranking taxon) in the hierarchy, and in addition to harvesting the checklist, the system also harvests branches of the tree beneath this top point for the hierarchy leading down to the species in that sector of checklist. The checklist and hierarchy created from a growing array of GSDs in this present-day architecture (‘Architecture 1’) is referred to as the ‘Global Hub’.

Despite the evident success of Architecture 1 in permitting the rapid build up of the Catalogue to its present point, its limitations have been evident for some time. The difficulty is simply that no-one anywhere in the world is creating global species databases for some of the least known taxonomic groups, so by this model these would be destined always to remain as gaps in the Catalogue. In the EC EuroCat project (2003 – 2006), we additionally experimented with making a Regional or ‘Euro-hub’ with a further set of European regional databases, and versions of SPICE that could handle multiple hubs, and the first steps towards integrating their contents using the LITCHI 2 taxonomically intelligent integrity tracking. In 2005 we started to plan an ‘Architecture 2’, in which an array of Regional Hubs might be connected to the Global Hub, this providing linkage to regional databases from many parts of the world, but also the potential for the Global Hub to harvest data or checklist sectors from Regional treatments for the species groups that were missing from the Global Hub. Good progress is being made with initiating these Regional hubs now – with the test N. American Hub now ready for connection in addition to the European Hub, and plans in WP 4 of this proposal to develop a unified concept and specification for this Multi-Hub Network working with the designated centres for China, New Zealand, Brazil, and Australia.

#### **1.1.4 Who uses the present services?**

Despite being only 60% complete, the Catalogue has become established as a significant global biodiversity resource with extensive electronic and human usage in Europe and all around the world. There is no other comparable source of this information.

*1. On the web.* Our services operating on the web [9; 10] receive ca. 40 M hits per annum, with a rising number of 30 - 40,000 unique serious users per month. But these account for only a fraction of the total usage on the web, which is by the users of other biodiversity portals that have installed a copy of the Catalogue as their taxonomic index.

*2. Globally significant biodiversity portals and biodiversity programmes.* Two globally significant biodiversity portals have installed the Catalogue as their principal taxonomic index, using both the taxonomic hierarchy and the synonymic species checklist to enable users to locate the organisms for which they seek information. The Global Biodiversity Information Facility (GBIF) has used it since the launch of its portal in 2003, and the Encyclopedia of Life (EoL) since its launch in 2008. Nearly every scientific use of those systems to locate data about an organism (and far outnumbering the 40 M hits on our own sites) makes use of the Catalogue one to many times. Neither organisation could have launched their portal in the way that it did without our Catalogue. Similar usage is made of the Catalogue by other portals (SeaLifeBase, SpeciesBase, NSF SEEK, the OpenModeller library used for climate

change studies, uBio and the BHL taxonomically intelligent tool), and related usage as an authority file by GenBank and the Red Lists of IUCN.

The Catalogue is used by BGCI (which organises information on plants conserved in botanic gardens around the world), by the Institute of Zoology in London (which works on conservation status analyses using national Red Lists and the Sampled Red List Index), and by European partners in the BHL project. Also the European Bioinformatics Institute in Cambridge has been exploring the possibility of mapping the GenBank Taxonomy to our Catalogue so that both can be used in GenBank. This has also been promoted by iBOL, the initiative by Genome Canada, and the Consortium for the Barcode of Life (CBOL) (which has adopted our Catalogue as its 'gold standard' for taxonomic annotation of DNA Barcode records that are being deposited in EBI and GenBank).

3. National biodiversity portals and programmes. About 40 national-level portals around the world have installed the Taxonomic Hierarchy and/or the Checklist: some are GBIF nodes (eg Netherlands), some national biodiversity portals (Norway, Thailand), some institute based (CRIA SpeciesLink, Brazil, Plants-of-India, Lucknow), and some NGO based.

4. Regional biodiversity portals and programmes. The Inter-American Biodiversity Information Network (IABIN) uses the Catalogue for its invasive species network and Species 2000 is involved with European initiatives - the European Distributed Institute of Taxonomy (EDIT), the Consortium of European Taxonomic Facilities, and the LifeWatch infrastructure programme.

5. Individual and small institute use of the Annual Checklist on CD. 3,500 copies of the Annual Checklist are distributed each year free of charge on CDs, enabling operation on PCs or laptops without internet access. During 2007 the Secretariat received written and emailed requests for the CD from 471 individuals in 79 countries. 300 are distributed by the National Institute for Environmental Sciences, Tsukuba, to members of the Japanese Society of Systematic Biologists. 300 are distributed by the Belgian Biodiversity Platform to focal points of the GTI and the Clearing House Mechanism from African and Lomé Convention countries. Distribution to the regional organisations of the BioNET International programme is about to start.

6. First commercial uses. A trial contract with Taylor & Francis plc provides taxonomic support to its electronic Dictionary of Natural Products. This has been started cautiously, as it involves permission from provider databases, and the possible return of some income to them as running costs.

### **1.1.5 What are the present services used for?**

What makes the Catalogue unique is the breadth of coverage of organisms, the degree of validation in the knowledge set, and the wide European and global take-up. A wide variety of interesting and significant science is being done worldwide by a range of programmes using the Catalogue as a baseline infrastructure. So, while not in itself saving the world's biodiversity, taming the effects of climate change, enabling evolutionary study, or dealing with human food and disease crises, it does increasingly contribute as an infrastructure to research on all of these.

#### *Basic Functions*

1. Species Look-ups. Vast numbers of people, both professional and amateur, use the Web to look up species of organisms, i) to verify the name they have, ii) to place a taxon in the taxonomic hierarchy, and iii) to see other basic data, such as distribution, for that species. Ideally their search will inform them of an accepted name for the species if the one used for the search is in synonymy. Unlike general search engines or knowledge stores such as Google, Wikipedia or Wikispecies, our Catalogue provides a single clear consensus taxonomy; and taxonomic consistency in classification and naming within and between groups. Many appreciate that it is provided and maintained by taxonomic experts, so that it is likely to be more current and authoritative, for example in dealing with variant spellings and misapplied names. For instance, species look-ups are the main usage by users as diverse as our Taylor & Francis customers, users of a culture collection at the Riken Institute in Japan, and undergraduates at University of Reading.

2. Sizing higher taxa. Individuals use the Catalogue to get an estimate of the size of a taxonomic group. Typically they know one member species, but have no idea whether the group is extensive or small on a world basis.

3. Synonymic indexing and synonymic amplification. Many biologists may not appreciate either the key role played by synonymy, or the extent to which the same organisms may be referred to by different scientific names in different continents or even different states in Europe. *Species checklists with automated synonymy have become the electronic organisers of biodiversity knowledge on the internet, and it is synonymy that enables the loose ends to be connected together.*

Synonymic indexing is automated in the Catalogue so that users in different countries who might search for Broad bean, or Fava Bean, or Faba bean, or *Vicia faba*, or *Faba vulgaris* will all arrive at the same species page. The page states clearly that the one species *Vicia faba* has a synonym *Faba vulgaris*, and common names, Broad Bean, Fava bean, Faba bean, all referring to the same species. The ‘unification’ or globalisation provided by accurate synonymic indexing contributes significantly to enabling international discussion and data exchange.

The reverse of this process is important – synonymic amplification. Given that one species may be referred to by many names, a person or machine searching the Internet, or a data set such as GBIF, will receive only a subset of data if they search on just one name. It is more effective to amplify the search with all synonyms of that species. There are many examples where this amplification of search strings yields sharply improved results, particularly spanning continents. The BBSRC BiodiversityWorld project introduced this as part of its data harvesting using the electronic web-service of the Catalogue, and it is now available in GBIF.

4. Taxonomic Backbone usage. GBIF has used the Catalogue as a backbone structure on which to make its own additions. It uses associative techniques to link additional species that are not presently in the Catalogue to the positions they are likely to belong. The ‘taxonomically intelligent’ tool provided by uBio to the BHL project using our hierarchy has broadly similar features. An email cited later, received as this proposal was in preparation, illustrates how the Institute of Zoology in London plans a related usage.

5. Download usage. The Annual Checklist edition on CD and the web is frequently used as a download facility to insert the data into a user’s own database, or as the database to which a user will add further data. However the present format is not ideal for this purpose. A well-structured download facility is planned as one of the novel technical services to be added in this project.

6. Global standardisation. Organisations are starting to use the Catalogue to achieve compatibility with others that already use it, for instance GBIF nodes and some national portals. *The possibility of a quantum increase in the coherence of the world's biodiversity data and analyses is beginning to emerge, simply by the process of many organisations opting to use the same Catalogue. Given the extent of society's dependence on biodiversity, this alone is a significant goal.* Compare GBIF and EoL, that have strong data compatibility and both use the same maps, with GenBank and IUCN Red Lists where the checklists are compiled independently. The GBIF/EoL compatibility arises largely because they both use our Catalogue as a taxonomic backbone. This standardisation role favours the Catalogue continuing to be generated and maintained by an independent organisation as a scientific infrastructure embedded in the taxonomic profession, rather than by a sectoral organisation. Similar issues arise in our discussions with CITES, the European Species Directive, the Global Invasive Species Programme (GISP), the CBD Biosafety Clearing House (BCH) and other regulatory frameworks.

7. Spell-checker for users of other character sets. There is anecdotal suggestion that scientists whose working languages use other character sets find Latin scientific names quite difficult, for instance in the subtle differences that denote ranks (eg. Caesalpiniaceae, Caesalpinioideae, Caesalpinia). Spell-checker usage possibly contributes to the peaks of night-time usage of the web-site, and adds to interest in the programme from Russia, China, Thailand and Japan.

#### *Compounding the basic functions*

1. Portal usage. The importation of the Catalogue by GBIF, EoL and many of the other portals provides them with all of the basic species checklist-related functions in a single package. This re-emphasizes the extent to which this is a fundamental scientific knowledge infrastructure serving these other bodies.

2. Programme usage. It is for this same reason that the CBD places the Catalogue at the core of its Global Taxonomy Initiative, and lists usages for it in six of its programmes.

3. Community player. We believe that the Catalogue creates 'value' in the biodiversity community by being a platform to which databases provide, and from which users are supplied: it is a valued knowledge platform in the biodiversity community.

#### *Contribution to higher level science and society*

1. Impacts of Climate Change. A core issue in many climate change modelling projects is how to harvest sufficient comparable world-wide occurrence records to be able to generate climatic envelope or other models for each of the species in the study. These models are then used to interact with climate change scenarios, and to generate predictions for the future. A common taxonomic framework and synonymic amplification are key tools. Our Catalogue is used as part of this process by GBIF, the NSF SEEK project, the BBSRC BiodiversityWorld Project, CRIA SpeciesLink in Brazil, and the EC INCOFish Aquamaps system.

2. Genetic Resources for Food, Disease & Biosafety Regulation, and Invasive Species. In all these international areas of biodiversity science it is important that there is integrity across species records and scientific reports between countries. Our Catalogue is used by Biodiversity International and the System-Wide Network for Genetic Resources of the international agriculture institutes, by the CBD Biosafety Clearing House (BCH) that monitors international biosafety issues, and by IABIN and GISP in international programmes relating to invasive species.

3. Identification of Biotic Materials in Foods, in Forensic Scenarios & in the Environment. Genome Canada is considering investing in excess of 50 M Canadian dollars in a global programme (iBOL) to establish usable DNA barcodes for identification of biotic materials in commercial, forensic and environmental programmes. To be effective this needs a common taxonomic backbone and synonymic integrity in their data systems. The International Consortium for the Barcode of Life (CBOL) has adopted our Catalogue as its 'gold standard' for annotation of Barcode records and we are presently discussing with GenBank and the Canadian iBOL programme how this is best implemented.

4. Conservation of Global Biodiversity. At present the family of organisations within IUCN, those linked to Conservation International and the Nature Conservancy in the Americas, and United Nations conventions use a wide variety of taxonomic catalogues, including some with known handicaps, such as taxonomic inflation of local endemics, and outdated taxon concepts enshrined in law. These difficulties provide a barrier to globalising efforts, and will not be solved rapidly. There are increasing signs that our Catalogue will be used in at least some of these contexts, for instance through the work of the Institute of Zoology and BGCI in London. The email excerpt of 24 June 2008 below (cited with permission) illustrates this.

*Dear Dr. Bisby,*

*I am writing to you to request permission to use the 2008 Annual Checklist hierarchy in my own database. At the Zoological Society of London, we are in the process of creating a master database of all National Red Lists that have been conducted worldwide. This online database will be launched at the World Conservation Congress in Barcelona this October and is intended to provide a focal point for National Red Listing. We hope that by making this data readily stored, managed and accessible, countries will be able to learn from each other in creating and updating Red Lists, and that this will help towards effective conservation planning.*

5. Uses outside Biology. Our Taylor & Francis contract is effectively for use of the Catalogue by chemists working on natural products. We are in contact with a number of wildlife image agencies, including the ARKive/WildScreen Trust in Bristol and a commercial organization in San Francisco.

Clients wish to make such uses of our Catalogue because:

- i) it has a reputation for quality, because it is validated by experts and peer reviewed
- ii) it has a reputation for quality, because it is internally integrated by an expert team
- iii) it will cover all groups in a single system ('one-stop shopping')
- iv) users can see the full extent of a group, unlike GenBank and IUCN Red Lists
- v) no other organisation has aggregated the taxonomic opinions needed to provide accurate synonymy on this scale
- vi) it is available electronically, and as an electronic web-service
- vii) the Taxonomic Hierarchy is artificially stabilised and simplified at the higher levels as a 'management hierarchy'
- viii) it has achieved a reputation as a workable standard both in Europe and worldwide.

## **1.2 Progress beyond the state-of-the-art**

The present Species 2000 Catalogue of Life programme is a unique, significant and leading component in the landscape of international biodiversity programmes. There

are a number of different levels at which it can be variously described as unique (in a class of its own), in which it is at the state of the art (along with other leading programmes), and in which it is joining with the front runners pushing ‘beyond the state of the art’.

i) The programme has been unique from its start in 1996 in developing the concept of Global Species Databases (GSDs) and the principle that only by multiplying this process can a comprehensive and validated global catalogue be developed. This principle was subsequently accepted by the more recent GBIF and EoL programmes that have come to use our Catalogue as their taxonomic backbones, and by the very recent WoRMS and FADA programmes that we are working with to complete coverage of primarily marine and primarily freshwater taxonomic groups.

ii) The programme has been unique from its start in adopting a distributed model, and pushing to embed the responsibility for creating and maintaining the taxonomy of the many different taxonomic groups into the institutions and groups of experts within the taxonomic profession. It recognises that virtually the whole taxonomic profession is needed to complete and maintain the task, and that no single taxonomic institution has the capacity to do so. GSDs provide the vehicle by which expertise is focussed from the experts into the single aggregated view. The distributed model is tolerant of, and responsive to, a wide heterogeneity of purpose and style of activity across the taxonomic community. For instance it uses databases from large institutions, medium-sized networks, and individual expert amateurs; it uses databases across a spectrum of purposes from primarily nomenclatural to full species bank, and the databases come from natural resources, veterinary, microbiological and health organisations as well as the purely taxonomic.

iii) The programme is possibly unique and certainly distinctive in its community participation model, which differs markedly from the models used by the other major global biodiversity infrastructures. Each participating centre or individual, and particularly each supplier database custodian, is invited to become a member of the Species 2000 residual legal body that owns and governs the Species 2000 programme, and the Species 2000 component of the Catalogue of Life. Each supplier database custodian is invited to sign an Access Agreement with Species 2000 that specifies and regulates how Species 2000 may publish, credit and handle expert knowledge from the supplier databases. These two features have combined to create a solid but flexible community organisation in which the partners and contributors feel a sense of ownership. It is for this reason that the European GSD Network custodians were invited to become full members of the presently-proposed consortium and also that Species 2000, the residual legal body, is a member of the consortium, involved both in the annual production cycle, and in the moves towards sustainability.

iv) The SPICE distributed system was innovative and ahead of its time when it was first released in 2003, and was used to implement the current version of the Dynamic Checklist launched in 2005. Some of its features have since been superseded by a new generation of community standards launched by TDWG, such as TCS, but others, such as the SPICE protocol for harvesting connected checklists and taxonomic hierarchy branches have yet to be bettered in fully available protocols. It is against this background that we propose to replace the SPICE system with a fully service-based and open architecture as a main part of this proposal, and to include both

significant adjustments to altered requirements, and the use of appropriate current community standards wherever possible.

v) In 2007/8 the programme introduced Globally Unique Identifiers (using the LSID system) for the Taxa recognised in the Catalogue, putting it very much at the leading edge, and preparing the ground for the range of GUID-based tracking services proposed in this project.

vi) The programme is working with major community partners GBIF and EoL towards the establishment of a new Global Names Index that will be part of a higher level Global Names Architecture. The Global Names Index (a giant list of all known names and orthographies) and the Catalogue of Life (giving a quality species list with synonymy) will be two significant items in the plans for a seamless biodiversity informatics landscape under development through this project, through EOL, GBIF, TDWG and EDIT.

### ***1.3 Methodology to achieve the objectives of the project, in particular the provision of integrated services***

The programme already provides a simple integrated service based on the 60% complete Catalogue. This proposal is

- i) to provide a substantial enhancement and diversification of the integrated services,
- ii) to re-engineer and install an enhanced state of the art service-based architecture,
- iii) to strengthen the supplier base, the system infrastructure, and the sustainability,
- iii) to extend community participation and taxonomic coverage to other continents.

#### *1.3.1 The present process of integration, and provision of integrated services.*

The Editorial and Production Team at Reading, Los Baños and ETI operates the steady state process of integrating the Catalogue using the taxonomic sectors supplied by the array of GSDs in Architecture 1. In the Annual Checklist process, there is a 12-month cycle with the Annual Checklist brought to publication and release for 1 April each year, and this is the present product of first choice to users. The Dynamic Checklist process is operated continuously, but in the proposed project we should like to increase the frequency of caching, and complete the wrapper connections for those GSDs not yet connected.

There are four processes in the integration.

- i) Existing supplier GSD databases provide enhanced data sets. These contain additional species, and edited species records in the species checklist, and they may also contain alterations to the taxonomic hierarchy joining the species up the hierarchy to the top point (highest taxon) in the sector. They may also contain what are considered by the Editorial Team as additional or separated taxonomic sectors. In the Annual Checklist process the enhanced data set is downloaded, lightly scrutinised by the Editorial Team, and inserted to replace the prior data set. In the Dynamic Checklist the enhanced data set is cached automatically to overwrite that part of the cache, and only the part below the registered top point is utilised. (New sectors need to be notified and taken through the new sector process.) Note that a proportion of the yearly increase in the size of the catalogue comes from enhanced data sets generally including more species than in previous occasions – especially as some of the GSDs are incomplete, and their staff are adding more species.

ii) During each year new GSD databases are enlisted to the programme, and also some GSDs offer new sectors that they have added, extending the taxonomic coverage. These new GSDs and sectors go through an extensive admission gate procedure that may take several months. Most important is that the metadata questionnaire, extensive discussion with the custodian about cross-mapping the schemas, technical review of the database, and peer review of the taxonomic quality, all precede the incorporation of their material into the Catalogue. The initial work is done using the Annual Checklist download process, followed both by wrapper writing and connection to the Dynamic Checklist, and by administrative work of signing Access Agreements and initiating membership of the Species 2000 legal body.

iii) Underlying the seemingly straightforward admission work for a new GSD sector is an extremely skilled task for the head of the Editorial and Production Team – where and in what taxonomic rank to connect the incoming GSD sectors, and how to place them in relation to adjacent sectors, and how to check for overlaps. The Executive editor has two consultation groups available to assist with this task – the Species 2000 Taxonomy Group (that advises the global Species 2000 Team), and an appointed Editorial Panel, plus the structure provided by the upper levels of the Catalogue of Life taxonomic hierarchy adopted from time to time by the global Team. This is a contentious and potentially troublesome area, but the policy that Species 2000 should adopt a preferred position and structure for the higher level taxonomy has been a source of approval by user organisations, even though some mistakes have inevitably been made and then rectified.

iv) The annual cycle of the Annual Checklist production means that there is a phase in which the whole product is reviewed by the editorial panel, and double-checked by the provider database custodians. This is currently not available in the Dynamic Checklist process, but it is hoped to introduce this with the unification of the processes planned in this project.

### *1.3.2 Methodology to re-engineer and install the enhanced service-based architecture*

Significant changes are needed because the simple Architecture 1 needs to be superseded by an architecture permitting a multi-layered ecosystem of providers and harvesters. The current distributed system needs to be enhanced with a fully service-based architecture suitable for incorporation into the programmes of other organisations such as GBIF, EoL, EBI and LifeWatch.

The methodology to be used involves the full cycle of requirements analysis, specification and design from an expert team based at the Welsh e-Science Centre at Cardiff University. A prototype system will be tested with three GSDs, and a second release produced that will be handed to WP 6 for the creation of a further robust production system with management interfaces and user documentation. Further detail is described under WPs 6 and 7 that will carry out the development.

### *1.3.3 Methodology to strengthen the supplier base*

A renewed assessment and certification will be made for each GSD in the supplier array in the work plan of WP 3. This will include both the databases that joined the programme in the previous EuroCat project, and for which full certification was carried out, and the more recent recruits for which not all of the certification was completed because of shortage of resources and staff.

After the assessment each GSD will be assigned two 2-month pilot projects. In most cases at least one of these is likely to be to make and test additions/alterations to their database to increase their compliance either with the Data standard, or with the best practice guidelines developed in the EuroCat project. Although it will take time for the GSDs to start filling these additional areas, the result is expected to be a significant improvement in the uniformity and degree of fill provided by the GSD databases, and these in turn will improve the fill and quality of the integrated service content.

#### *1.3.4 Methodology to strengthen the system infrastructure*

Some key alterations to the currently used Annual Checklist and SPICE/Dynamic Checklist middleware are expected to make it possible to unify the Annual Checklist and Dynamic production routes. This will provide two improvements – that all tools in both processes can be used in a single production route, and that the difficult taxonomic positioning and insertion process can be partly automated by software that links between the metadatabase and the taxonomic hierarchy.

#### *1.3.5 Methodology to strengthen the sustainability*

Substantial work towards achieving sustainability has already been carried out. Although the programme is currently not in a sustainable position without grant support, there has been progress on several fronts, and the programme has been able to progress even through the 2006 – 2008 period without major grants.

The work needed includes discussion and financial negotiation with three of the classes of users and stakeholders – the large global programmes that depend quite heavily on the Catalogue for their own infrastructure, the many smaller portals that are in fact quite well funded at national levels, and other commercial companies that may wish to enter into small contracts such as the experimental one with Taylor & Francis. This needs to be accompanied by legal and discussion work with the supplier databases, both to gain their support for making this activity sustainable, but also for modest contributions to their running costs to be included in the financial model.

#### *1.3.6 Methodology to extend the community participation and taxonomic coverage to other continents.*

The proposed development of Regional Hubs in a Multi-Hub Network under Architecture 2, and also the devolution of active work in the programme to Centres in each continent, have the technical and institutional potential to enlarge this currently Europe-based programme to become a truly global knowledge network and scientific infrastructure.

In WP 4 it is proposed to network with the leaders of these developments in the first batch of agreed regional centres – in China, New Zealand, Australia, Brazil and N. America. At the technical level the group will both evaluate the test-bed implementation available with three regional hubs (European, N. America/ITIS and China) and draw up a proposed concept and specification for the working of the regional Hubs. At the institutional level it is proposed that they also draw up a model for how the regional centres may interact with the present global programme, with its residual legal body, its elected Directors, and its global Team.

## 1.4 Networking Activities

The Networking Activities proposed are designed to foster a culture of co-operation both between the partners in the programme (as is already developing), and between them and the different user communities – something that is new and has not been undertaken before. The intention is to generate from these activities both a stronger more professional infrastructure organisation and an enriched and more valuable set of scientific services. These activities are grouped in three work packages (WP 2, 3, and 4) each of which has the goal of both raising levels of co-operation and bringing new or enhanced service capability to the programme. These work packages are led by Sara Oldfield, Secretary General of Botanic Gardens International, an international organisation based at Richmond, UK (WP 2), Thierry Bourgoïn, Deputy Director of the Collections at the Museum National d'Histoire Naturelle, Paris (WP 3), and Jiri Kvacek, Leader of Science programmes at the National Museum in Prague (WP 4). The whole project is co-ordinated by Frank Bisby at the Species 2000 Secretariat at the University of Reading, including the consortium management (WP 1).

	Year 1					Year 2					Year 3					2012											
	2009	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug		Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	
<b>4D4Life</b>	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul		
<b>WP1 Consortium</b>	Consortium Management																										
Tasks 1.1 - 1.5																											
<b>WP2 Users</b>	User Communications																										
Task 2.1																											
Task 2.2	Analysis of Usage																										
Task 2.3	Plan New Services											Test New Services															
<b>WP3 GSD's</b>	Assessment of GSDs																										
Task 3.1																											
Task 3.2	Network Consultation																										
Task 3.3																											
Task 3.4	Pilot Projects																										
<b>WP4 Multi-Hub</b>																											
Task 4.1	Foresight Consultation																										
Task 4.2	Pilot Projects											Specification					Pilot Projects 2										
<b>WP5 Services</b>																											
Task 5.1	Enhance Present Services																										
Task 5.2	Create Novel Technical Services																										
Task 5.3	Create Novel Education Services																										
Task 5.4	Annual Production Service																										
Task 5.5	Editorial Service																										
Task 5.6	Systems Management Service																										
Task 5.7	Sustainability Engineering																										
Task 5.8	Scientific Co-ordination																										
<b>WP6 Software</b>																											
Task 6.1	a) Interface Service Enhancements - Software											b) ..... More															
Task 6.2												Novel Technical Services - Software															
Task 6.3												Novel Education Services - Software															
Task 6.4	Base Schema																										
Task 6.5	GSD Builder & Editor Software																										
Task 6.6	Wrapper Kit & Caching Software																										
Task 6.7	Cache and Linking Tools Software																										
Task 6.8																											
Task 6.9	Multi-Hub Linkages Software																										
Task 6.10	Metadata Software & Linking Tool																										
Task 6.12	Export & Conversion Tool																										
<b>WP7 e-Science</b>																											
Task 7.1	e-2 Requirements & Spec																										
Task 7.2	e-2 Design																										
Task 7.3	e-2 Protocols																										
Task 7.4	e-2 Test & Manage Tools																										
Task 7.5	e-2 Test Implementation																										
Task 7.6	2nd Version																										
<b>Meetings 1: Project Wide</b>	■																										
2. Liaison Group	X																										
3. Work Package	all																										
4. Management Committee	M																										

### WP 1: Management of the Consortium

Work Package 1 provides the consortium/financial management required by the Commission. The proposal is submitted and co-ordinated by the Species 2000 Secretariat, with support by the University of Reading Research Enterprise Services, and, if successful, will be handled in the same way as the FP5 EuroCat project.

### WP 2: Engagement with scientific users and partners

In Work Package 2 it is proposed initially to survey and identify more accurately which are the classes of users, and how they make use of the present services. Two of these classes are reasonably well-known to us – the major global biodiversity portals, and the national portals. However, the third and fourth classes, the individual scientists and citizens in Europe and worldwide who use the web-site and the distributed CD systems, and the commercial organisations that may be interested, all need to be better characterised, and processes put in place to engage with at least some of them. The needs of these four classes of users are quite likely to be rather different, so the intention is to engage separately with each of them. Surveys, bulletins and small workshops will be used to elicit needs and suggestions that can then be turned into pilot proposals for new services. In the case of the major global portals we are already talking in terms of User/Partnerships with further joint enterprises as well as our contribution to the approaching ‘seamless biodiversity informatics’ architecture, for instance with the Global Names Architecture (GNA) under discussion with EoL and GBIF. A further task associated closely with user communications is to establish a communications strategy for the programme, and establish the necessary outreach and communication programme.

Without prior knowledge as to exactly what they will be, we have broken these new services into ‘educational and popular’ services likely to address students and citizen scientists, and ‘technical services’, such as a variety of new web services built on the service-based architecture of the programme as a whole and designed to serve the technical informatics communities in biodiversity, climate change, food security and conservation. These new services will be designed and prototyped by staff working under WP 5 (Services) and supported by software innovations under WP 6 (Software Support), before being returned to the originating user groups for testing in the later part of the project. Additionally some of the suggestions may be best handled as alterations or improvements to the present services and interfaces.

In practice we already have a substantial set of suggestions in hand to which the new ones can be added, but the interactions with user groups and the service work packages are much needed – to determine which suggestions should be given high priority, and to balance these priorities against their practical and cost implications.

Initial suggestions for the two groups of new services include:

i) Educational & Popular

- Common Names in the Taxonomic Hierarchy
- Thumbnail images of representative or iconic species in the Hierarchy
- Species total tally counter on the web-site
- Recording current species totals in the hierarchy
- Catalogue of Life ‘snippets’ for use in browsers and installation on web-sites.
- Services to hand-held devices
- Downloadable subsets and displays for use in hand-held devices
- Downloadable subsets for countries and continents
- Fuzzy spelling and phonetic variant searching

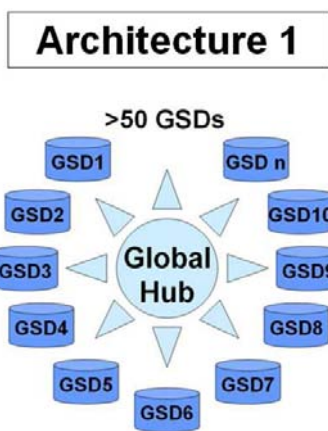
ii) Technical

- GUID-based services highlighting name changes, and circumscription changes
- GUID-based services for incremental cross-maps between annual editions
- GUID-based alert services and dynamic real time provision to certain users
- A range of download services using standard protocols and data models

A range of web services, e.g. synonymy server, geographical subset server.  
Resilient and load-sharing services able to serve certain users at high load.

### *WP 3: Strengthening the GSD Network: a Virtual Community*

In Work Package 3 we plan to build on the participatory network of supplier Global Species Databases (GSDs) already established in the previous FP5 project (“Architecture 1”), but with a number of issues to be discussed and agreed, each of which has the potential to enhance the existing services. This is the European part of the distributed array of databases on whom the project depends totally – they supply the expert taxonomic sectors that are harvested and integrated to make up the content of the Catalogue. This Network pulled together in a very co-operative way during the FP 5 project, agreeing data standards, best practice, access agreements, and linking up to enlarge the Catalogue and bring the whole integrative process to production scale.



A network of 35 organisations in a project such as this would normally be seen as a management liability. We stress that the network is already well established with an administration in place and its scientific contribution under control. Not only has it functioned continuously since the FP 5 EuroCat project, both in administrative terms (email list, access agreements and membership of the legal body) and in scientific terms (supply and integration of their data into the Catalogue yearly or continuously), but it has also continued to hold annual meetings, and another 10 databases have joined.

The issues for debate and decision among the Network are:

- i) Compliance issues: moving some of the databases towards better compliance with the agreed data standards and best practice: this will impact directly on improved quality and better uniformity in the Catalogue services.
- ii) Proposals that will come from Workpackage 8 both to agree a small increase in the data set, to cover life forms etc., and to standardise the distribution records, so that range maps can be added – both features that will enhance the services in the eyes of many users. For instance if the distribution data can be structured better, it will eventually become possible to download the complete all-taxon species list as a selected subset for each country in the world, a major step forward.
- iii) Wrapper programs and caching frequency: now that many of the GSDs are connected on-line to the SPICE cache, we should like to experiment with increased frequency of caching, so that the Dynamic Checklist can be refreshed at a higher frequency. Some new databases are still to be connected, and we will want to test the wrapper system for the new architecture when it comes on line at the end of the project. These steps will increase the currency of the Catalogue services for users.
- iv) Certification, Peer Review and Access agreements. Many of the GSDs in the network became connected in 2005/6, and will therefore need renewed Certification, Peer review, and renewed Access agreements – all of which were planned with a 3 – 5 year cycle time. In the long term it is these that secure the supply and quality of supply to the Catalogue and its services.

v) GSD alerts and services from Nomenclators. This is to discuss possibilities for a new layer in the ‘ecosystem’ of services creating the Catalogue. Traditionally the GSD custodians have gleaned and entered newly published species to their checklist sectors by hand – but now there is the possibility of gaining them electronically as alerts or feeds from the Nomenclator databases that assemble lists of names for certain sectors – Fungi, Plants etc.

vi) In addition to these material tasks, it is vitally important for the sociology of the whole programme that the GSD custodians in this Network continue to meet, communicate, and feel ownership of the programme. Most are also legal ‘members’ of the Species 2000 residual body, thus playing a formal part in the long-term governance.

Each GSD partner is funded for 4-months of work on two 2-month pilot tasks. The exact tasks undertaken will be negotiated with each GSD after their assessment in year 1. Four classes of pilot project will be possible:

- i) work on the current SPICE-based wrapper and installation
- ii) alterations to the GSD database to increase compliance with the Sp2000 data standard and best practice for management
- iii) alteration to the GSD database and trialling of the proposed new fields
- iv) trials of wrapper development for the new e-2 system in year 3.

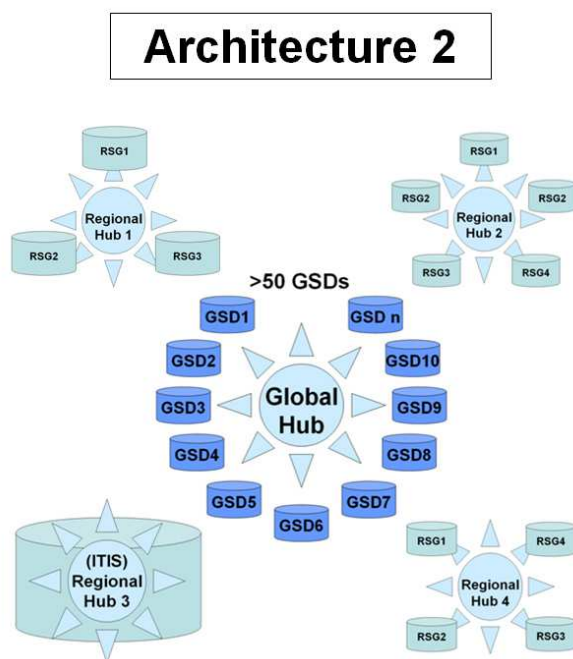
The outcome from these discussions and the array of pilot projects will be:

- a) improved harvesting performance of each GSD in the Network
- b) improved fill and uniformity of fill across the taxonomic sectors of the Catalogue – and hence an improved resource and service capability.

*WP 4: Foresight Study for Geographical Extension: the Global Multi-Hub Network*

Workpackage 4 establishes a completely new network that represents the extension of

the programme to become truly global, and to establish the Architecture 2 foreseen for future operation. It networks the leaders of the Regional Centres that will establish the ring of Regional Taxonomic Database Hubs. This development will extend both the data coverage and the organisational coverage of Species 2000 eventually to every region of the world.



data bases, these already provide to the global hub nearly complete world coverage of the species in those sectors. But for those taxonomic sectors where no GSD exists, where in the world may we find electronic records for the remaining species? To a

large measure we expect to find those records in Regional Species Databases (RSDs), particularly in the so-called mega-diverse countries: China, Brazil, New Zealand, Colombia etc. So this is one of the reasons for embarking on Architecture 2, and the installation of regional hubs. However the notion of forming a regional database hub in each of these areas has other reciprocal attractions for biodiversity managers in those regions, including benefits from linking their own taxonomic databases with the global hub of the Catalogue.

So this Workpackage is to bring together the leaders of the first batch of agreed Regional Hubs in what is now being called the 'Multi-Hub Network'. The object is to discuss a range of concepts and proposals, and to agree an overall concept and implementation strategy under Architecture 2 so that the Multi-Hub Network can be prototyped, and later implemented. The Workpackage does depend on modest funds being made available for the Multi-Hub Network partners from outside Europe to take part. In each case the major funding of the Regional Hub has already, or will shortly be made available from other sources in their own region.

For each of these five partners, we make the special case that a senior or technical staff member from that organisation will be needed for these consultations and for several meetings in Europe. The *Species 2000 China Node*, already established and funded by the Chinese Academy of Sciences, and the *Catalogo da Vida Brasil*, in development by CRIA in Campinas, working with the Brazilian Ministry of Science & Technology, are both nationally designated centres within third countries listed as International Co-operation Partner Countries. The *New Zealand Organism Register (NZOR)* nationally funded at LandCare Research, a Crown Institute; the *Integrated Taxonomic Information System (ITIS)* a US Federal Government project at the Smithsonian Institution; and the *Atlas of Living Australia (ALA)* a federally funded programme at CSIRO (Entomology) are nationally designated centres within third countries. We make the case that only nationally designated centres within these five countries can provide the technical and institutional expertise to operate these regional hubs successfully within those countries, and for the biota of those regions. In each case this expertise is not available within Europe, and is essential for the global extension of the programme.

The three Networking work packages largely engage different subsets of partners for rather different purposes, although all related to extending and enhancing the public services. Much of the time they will work separately, but at the second project workshop we plan a day-long session at which each group will be present to discuss their proposals. This is not only for information and interaction, but also against the possibility that synergies across the groups will generate yet further ideas for enhancing the services. For instance, the regional databases to be linked in WP 4 may be able to provide a richer set of common names than can be provided by the global databases in WP 3. Also the Regional centres may be able to set up mirror sites with services provided in their different languages, for example, making the Catalogue available in Chinese and Portuguese interfaces.

## Networking Activities: Work package list

Work package No <sup>11</sup>	Work package title	Type of activity <sup>12</sup>	Lead partic no. <sup>13</sup>	Lead partic. short name	Person-months <sup>14</sup>	Start month <sup>15</sup>	End month <sup>5</sup>
1	Management of the Consortium	MGT	1	UREAD	26	m1	m36
2	Engagement with scientific users and partners	COORD	2	BGCI	37	m1	m36
3	Strengthening the GSD Partner Network: a Virtual Community	COORD	3	MNHN	139	m1	m36
4	Foresight Study for geographical Extension: the Global Multi-Hub Network	COORD	4	NMP	22	m1	m36
<b>TOTAL</b>				<b>4</b>	<b>224</b>		

## Networking Activities: List of Deliverables

Del. no.	Deliverable name	WP no.	Nature	Dissemination level	Delivery date (proj. month)
D1.1	Consortium agreement signed	1	R	Re	m3
D2.1	Six-monthly e-bulletin	2	O	Pu	m6
D3.1	Data Standard Document	3	O	Pu	m11
D4.1	Multi-Hub Network Concept	4	O	Re	m12
D1.2	Management & monitoring in place	1	R	Re	m14
D2.2	Use Analysis Report	2	R	Re	m14
D2.3	Services Proposals Report	2	O	Re	m14
D1.3	Reporting to Commission for Year 1 done	1	R	Re	m15
D3.2	Proposal for feedback services	3	O	Re	m24
D1.4	First audit completed	1	R	Re	m24
D2.4	Public engagement strategy	2	R	Re	m25

<b>D4.2</b>	<b>Milti-Hub Pilot Report</b>	<b>4</b>	<b>O</b>	<b>Re</b>	<b>m25</b>
<b>D3.3</b>	<b>Feedback consultation document</b>	<b>3</b>	<b>O</b>	<b>Re</b>	<b>m30</b>
<b>D2.5</b>	<b>Report on new services and testing</b>	<b>2</b>	<b>R</b>	<b>Re</b>	<b>m34</b>
<b>D4.3</b>	<b>Multi-Hub Network Specification</b>	<b>4</b>	<b>O</b>	<b>Re</b>	<b>m36</b>
<b>D1.5</b>	<b>Final reports to Commission</b>	<b>1</b>	<b>O</b>	<b>Re</b>	<b>m39</b>

## Work package 1: Management of the Consortium

<b>Work package number</b>	<b>WP 1</b>	<b>Start date or starting event:</b>	<b>Month 1</b>
<b>Work package title</b>	<b>Management of the Consortium</b>		
<b>Activity type<sup>20</sup></b>	<b>MGT</b>		
<b>Participant number</b>	<b>1</b>		
<b>Participant short name</b>	<b>UREAD</b>		
<b>Person-months per participant</b>	<b>26</b>		

### Objectives

- i) To establish and manage the consortium
- ii) To manage delivery and reporting of the project tasks and deliverables for the consortium.
- iii) To administer the community financial contribution for the consortium.
- iv) To manage the audit for the consortium.

### Description of work (possibly broken down into tasks) and role of partners

Task 1.1 Agree and sign the Consortium Agreement

Task 1.2 Establish and administer a Management Committee to manage, monitor and report the progression of the project tasks, the completion of milestones and deliverables, and compliance with the grant agreement.

Task 1.3 Review the reports on tasks, and the achievement Project Milestones and Deliverables prior to transmitting them to the Commission for the Consortium.

Task 1.4 Manage the keeping of records and financial accounts, and the receipt, claiming and distribution of funds to beneficiaries at the start, each year, and at the close of the project.

Task 1.5 Manage the Audit for the Consortium.

### Deliverables (brief description) and month of delivery

D1.1 M3 Consortium agreement signed

D1.2 M14 Management Committee and full task monitoring in place

D1.3 M15 Financial and task reporting to Commission established and completed for year 1.

D1.4 M 24 First audit completed.

D1.5 M 39 Final financial and task report to Commission completed.

## Work package 2: Engagement with scientific users and partners

<b>Work package number</b>	<b>WP 2</b>	<b>Start date or starting event:</b>	<b>Month 1</b>
<b>Work package title</b>	<b>Engagement with scientific users and partners</b>		
<b>Activity type<sup>20</sup></b>	<b>COORD</b>		
<b>Participant number</b>	<b>2</b>	<b>1</b>	
<b>Participant short name</b>	<b>BGCI</b>	<b>UREAD</b>	
<b>Person-months per participant</b>	<b>36</b>	<b>1</b>	

### Objectives

- i) To engage with the major global biodiversity portals and biodiversity organisations that utilise the Species 2000 Catalogue of Life services, including GSIS, GBIF, LifeWatch, EoL, BHL, GenBank, IUCN, and CBOL to review current and potential uses and enhancements.
- ii) To locate and engage with the substantial number of national portals and programmes, NGOs, and higher education institutes that utilise the Species 2000 Catalogue of Life Services, to review current and potential uses and enhancements.
- iii) To locate and engage with particular subsets of the 40,000 individuals who make serious use of the Species 2000 Catalogue of Life Services each month to review current and potential uses and enhancements.
- iv) To locate and engage with the array of commercial organisations that either do use the Species 20000 Catalogue of Life Services, or may be considered likely to do so – for instance in natural product chemistry, in the seed and horticultural services, and the biodiversity image agency trade.
- v) To establish a communication and promotional programme that will provide ongoing communication with the classes of user and disseminate the infrastructure services even more widely than has occurred so far.
- vi) Working with the users and partner organisations, to specify, and subsequently test a range of enhanced and novel scientific services.

### Description of work

- Task 2.1 Test, establish and operate user and partner communication platforms designed to in form on progress, stimulate engagement and promote the Species 2000 Catalogue of Life, through
- a quarterly e-bulletin (that will include information on new services),
  - workshops and events (to which major institutional and potential users are invited).
- Task 2.2 A renewed analysis of:
- *current and potential users*, to analyse changes over the last 5 years, the potential for retaining them, and for engaging additional users.
  - *geographical distribution of users*, to analyse changes over 5 years, and likely change in the future,
  - *uses to which the CoL is put*, how this changed over five years, and is likely to change in the future.
- Task 2.3 Engagement with different classes of users to generate:
- proposals as to how the current user-interface and web-services might be enhanced,
  - proposals for completely new educational and popular services, including mobile device services,
  - proposals for completely new technical services, including further web-services.
- And subsequent user-testing of these new services.

### Deliverables

- D2.1 mth 6, 12 etc Six monthly E-bulletin (O, PU)
- D2.2 mth 14 Use Analysis (R, Re)
- D2.3 mth 14 Proposal for new services:  
Educational & Popular Services, & Technical & Electronic Services (O, Re)
- D2.4 mth 24? Public Engagement Strategy Proposal (R, Re)
- D2.5 mth 34 Report on new services and their testing (R, Re)

## Work package 3: Strengthening the GSD Partner Network: a Virtual Community

<b>Work package number</b>	<b>WP 3</b>		<b>Start date or starting event:</b>			<b>Month 1</b>	
<b>Work package title</b>	<b>Strengthening the GSD Partner Network: a Virtual Community</b>						
<b>Activity type<sup>20</sup></b>	<b>COORD</b>						
<b>Participant number</b>	<b>3</b>	<b>7</b>	<b>11</b>	<b>GSD</b>	<b>Network</b>		
<b>Participant short name</b>	<b>MNHN</b>	<b>NMNH</b>	<b>NHM</b>	<b>-</b>	<b>-</b>		
<b>Person-months per participant</b>	<b>36</b>	<b>7</b>	<b>6</b>	<b>4</b>	<b>each</b>		

### Objectives

- i) To strengthen and certificate the array of provider Global Species Databases (GSDs).
- ii) To move towards standardisation among GSDs, so as to make the Catalogue of Life more uniform.
- iii) To agree and pilot possible further enhancements to the GSDs, so as to enhance the Catalogue of Life Services for Users.
- iv) To roll out service-based provision to the new e-2 Architecture being developed in this project.

### Description of work (possibly broken down into tasks) and role of partners

Task 3.1 Assessment, peer review and certification of each provider GSD and existing wrapper.

Task 3.2 Network consultation and workshop with GSD custodians and Editorial Services Staff (WP 5), both to explain the assessment; to re-visit the Sp2000 Data Standard and the Sp2000 GSD Best Practice document; to consult on attitudes towards the extended bio-data fields proposed by WP 8; to explain and establish priorities for the pilot project for each database; and to provide additional training days.

Task 3.3 One Pilot Project undertaken per GSD database:

- for partially compliant GSDs to take a step up the compliance ladder (eg add fields or best practice)
- or for new GSDs (since EuroCat project), to install and test a SPICE wrapper
- or to install and test fill the new bio-data fields from WP 8.
- or a fourth class, not available until later (to test install new e-2 system wrapper)

Task 3.4 Network consultation and workshop reporting on progress with the three previous classes of Pilot Project and additional training days for the new e-2 system wrapper, with opportunity for the last set of GSDs to:

- undertake the fourth class of Pilot Project – to test install the new e-2 system wrapper.

### Deliverables (brief description) and month of delivery

D3.1 Data Standard Document (month 11)

D3.2 Proposal for feedback services (month 24)

D3.3 Feedback consultation document (month 30)

## Work package 4: Foresight study for geographical extension: the Global Multi-Hub Network

<b>Work package number</b>	<b>WP 4</b>		<b>Start date or starting event:</b>			<b>Month 1</b>	
<b>Work package title</b>	<b>Foresight study for geographical extension: the Global Multi-Hub Network</b>						
<b>Activity type<sup>20</sup></b>	<b>COORD</b>						
<b>Participant number</b>	<b>4</b>	<b>1</b>	<b>33</b>	<b>34</b>	<b>6</b>	<b>5</b>	<b>3</b>
<b>Participant short name</b>	<b>NM</b>	<b>UREAD</b>	<b>CAS</b>	<b>CRIA</b>	<b>CU</b>	<b>ETI</b>	<b>MNHN</b>
<b>Person-months per participant</b>	<b>18</b>	<b>1</b>	<b>4</b>	<b>4</b>	<b>1</b>	<b>1</b>	<b>1</b>

### Objectives

- i) To bring together for the first time leaders of the various international centres from around the world that have committed to create Regional Hubs in the Sp2000 Multi-Hub Network, so that they can work together in Networking, Discussion and Design.
- ii) Foresight Study to report the Needs, Conceptualisation, Specification and subsequently the Design both for individual Regional Hubs, and for the Multi-Hub network as a whole.
- iii) To carry out a limited amount of prototyping and testing to explore the concepts and design features agreed.
- iv) Foresight Study to point the way for other potential regional hub centres, not yet part of the Multi-Hub Network, to form Regional Hubs and join the Network in future, with a view of eventually completing coverage of all continents and oceans, and of including centres from less-advantaged parts of the world in possible future projects.

### Description of work (possibly broken down into tasks) and role of partners

Task 4.1 Network consultation and workshop discussions among regional hub organisers:

- to debate the Needs, Concepts and Specification for the Multi-Hub Network at both the taxonomic and the technical level, leading to a Preliminary Specification with outstanding issues highlighted.
- to discuss building on the Preliminary Specification and the Pilot Projects leading to a firm Concept and Specification of the Multi-Hub Network.
- to report on the Pilot Projects.
- to prepare an outline plan for the extension of the Multi-Hub Network to cover the whole world, and to include new centres, some from less advantaged parts of the world.

Task 4.2 A small number of Pilot Projects:

- for prototyping and testing based on the limited Multi-Hub Linkage tools available with the present SPICE system within Sp2000.
- for installing and testing Multi-Hub linkages using tools from the new e-2 Architecture of WP 7, and benefitting from the specification documents created in this work package.

### Deliverables (brief description) and month of delivery

D4.1 m12 Multi-Hub Network

D4.2 m25 Pilot Project Report

D4.3 m36 Multi-Hub Network Specification

### Milestones

M4.1 Pilot Project started (month 16)

M4.2 2<sup>nd</sup> Pilot Project started (month 15)

### Networking Activities: Summary of effort

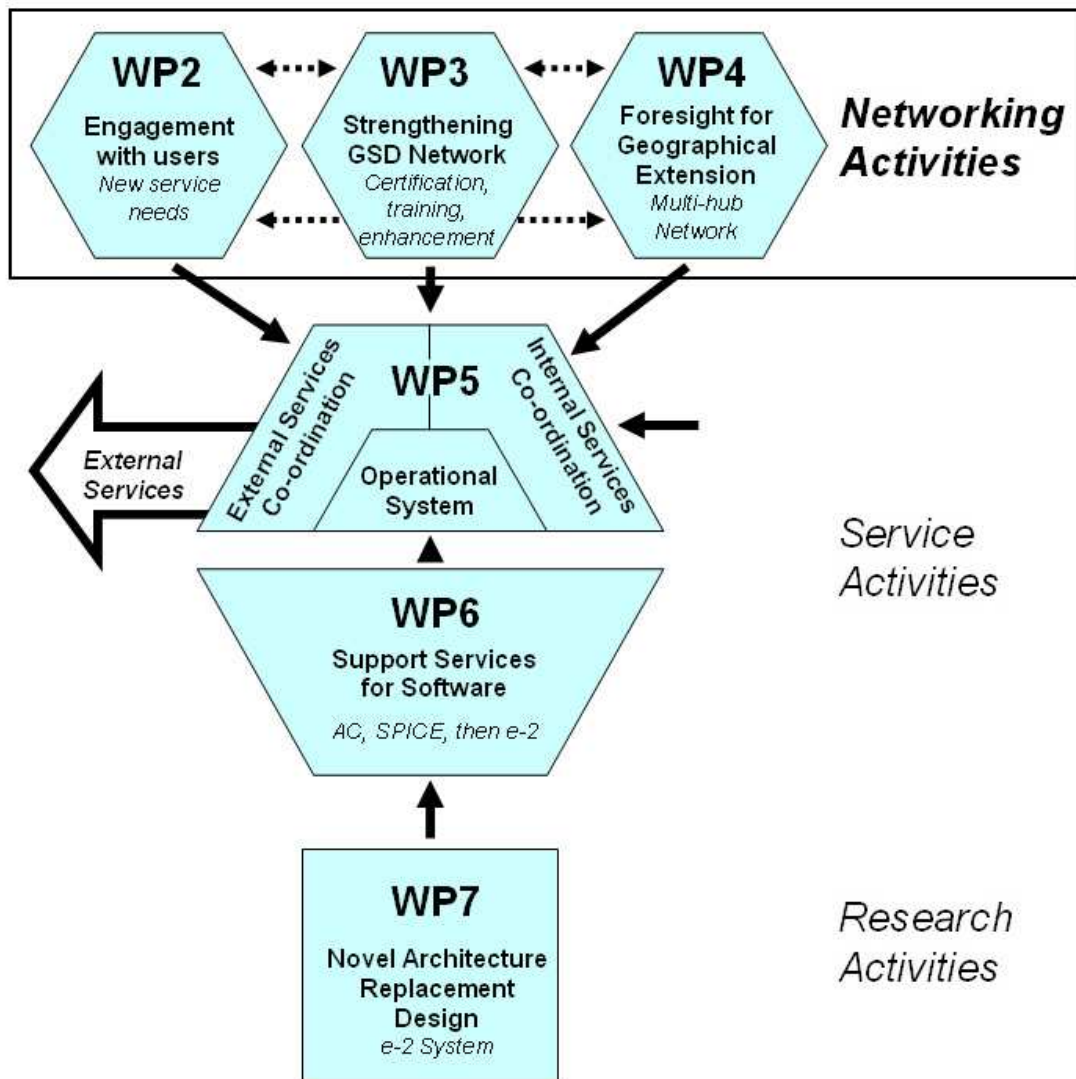
A summary of the effort is useful for the evaluators. Please indicate in the table number of person months over the whole duration of the planned work, for each work package by each participant. Identify the work-package leader for each WP by showing the relevant person-month figure **in bold**.

Partic. no.	Partic. short name	WP 1	WP 2	WP 3	WP 4	Total person months
1	UREAD	<b>26</b>	1	8	1	36
2	BGCI		<b>36</b>			36
3	MNHN			<b>14</b>	1	15
4	NMP				<b>18</b>	18
5	ETI BioInformatics				1	1
6	CU				1	1
7	NMNH			7		7
8	SP2000					0
9	VLIZ			4		4
10	SMNS			4		4
11	NHM			6		6
12	CABI			4		4
13	RBGK			4		4
14	UNI WIEN			4		4
15	UU			4		4
16	NUIG			4		4
17	IMUK			4		4
18	FZK			4		4
19	BSM			4		4
20	MIZPAN			4		4
21	OOE			4		4
22	NMGW			4		4
23	IRDF			4		4
24	CSIC			8		8
25	TSJ BVBA			4		4
26	SNSB, Munich			4		4
27	UNIPD			4		4
28	UVA			4		4
29	RBINS			4		4
30	University of Oxford			4		4
31	WU			4		4
32	RBGE			4		4
33	CAS				4	4
34	CRIA				4	4
35	ITIS					0
36	CSIRO					0
37	Landcare Research					0
<b>Total</b>		26	37	139	22	224

### **Networking Activities: Milestones**

Milestones are control points where decisions are needed with regard to the next stage of the project. For example, a milestone may occur when a major result has been achieved, if its successful attainment is a required for the next phase of work. Another example would be a point when the consortium must decide which of several technologies to adopt for further development.

<b>Milestone number</b>	<b>Milestone name</b>	<b>Work package(s) involved</b>	<b>Expected date</b>	<b>Means of verification</b>
M3.1	GSD Assessment Complete	WP3	m8	Report
M3.2	New bio-data fields agreed	WP3	m12	Report
M3.3	GSD pilot projects allocated	WP3	m15	Report
M4.1	Milti-Hub pilot projects started	WP4	m16	Report
M4.2	2 <sup>nd</sup> Milti-Hub pilot project started	WP4	m25	Report
M3.4	e-2 wrappers under test	WP3	m30	Report



Relationships between Workpackages

## 1.5 Service Activities

The whole purpose of the Species 2000 Catalogue of Life programme is to synthesize the Catalogue of Life as an ongoing integrated knowledge resource, and to use this as the basis for scientific infrastructure services to science and society. This section of the proposed project operates those two processes – the production of the integrated resource, and the generation of services using that resource: it is at the very core of the programme. Also it will be the beneficiary of ideas for new services and enhancements generated by the Networking Activities and the new e-infrastructure developed in the Joint Research Activity, so that by the end of the project these new services or enhancements will be incorporated in the main production stream leading to the public services. But between the start and finish, these two service work packages have also a very substantial role in conceptualising, creating, testing and launching those new services and features.

The work is stratified into two service layers. The taxonomic integration of the knowledge and its synthesis into the Catalogue, plus the operation of the public services are done from Reading. As seen by the public, and as seen by the participants in the project it is the Reading centre that undertakes these tasks. But the integration process and the operation of the public services both depend on having specialised software in place with which to run these internal and external services. ETI will underpin software support services for these two processes. It will take responsibility for generating and maintaining the production software, both as an internal service to participants in the programme, and for the outward facing public services themselves. The integration and synthesis of the Catalogue, and the operation of public services are led by Frank Bisby at University of Reading (WP 5), and the software support services for both the internal and external services are led by Peter Schalk at ETI in Amsterdam (WP 6).

	Year 1					Year 2					Year 3					2012																				
4D4Life	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
<b>WP1 Consortium</b>																																				
Tasks 1.1 - 1.5	Consortium Management																																			
<b>WP2 Users</b>																																				
Task 2.1	User Communications																																			
Task 2.2	Analysis of Usage																																			
Task 2.3	Plan New Services										Test New Services																									
<b>WP3 GSD's</b>																																				
Task 3.1	Assessment of GSDs																																			
Task 3.2	Network Consultation																																			
Task 3.3	Pilot Projects																																			
Task 3.4																					Further Consultation & Last Pilots															
<b>WP4 Multi-Hub</b>																																				
Task 4.1	Foresight Consultation																																			
Task 4.2											Pilot Projects										Specification															
<b>WP5 Services</b>																																				
Task 5.1	Enhance Present Services										Create Novel Technical Services										Create Novel Education Services															
Task 5.2																																				
Task 5.3																																				
Task 5.4	Annual Production Service																																			
Task 5.5	Editorial Service																																			
Task 5.6	Systems Management Service																																			
Task 5.7	Sustainability Engineering																																			
Task 5.8	Scientific Co-ordination																																			
<b>WP6 Software</b>																																				
Task 6.1	a) Interface Service Enhancements - Software										b) ..... More																									
Task 6.2											Novel Technical Services - Software										Novel Education Services - Software															
Task 6.3																																				
Task 6.4	Base Schema																																			
Task 6.5	GSD Builder & Editor Software																																			
Task 6.6	Wrapper Kit & Caching Software																																			
Task 6.7											Cache and Linking Tools Software										Export & Conversion Tool															
Task 6.8																																				
Task 6.9	Multi-Hub Linkages Software																																			
Task 6.10	Metadatabase Software & Linking Tool																																			
Task 6.12																					Production Version of e-2															
<b>WP7 e-Science</b>																																				
Task 7.1	e-2 Requirements & Spec																																			
Task 7.2	e-2 Design																																			
Task 7.3	e-2 Protocols																																			
Task 7.4	e-2 Test & Manage Tools										e-2 Test Implementation										2nd Version															
Task 7.5																																				
Task 7.6																																				
<b>Meetings 1: Project Wide</b>																																				
2. Liaison Group	X																																			
3. Work Package	all																																			
4. Management Committee	M																																			

### *WP 5 Scientific Services of the Catalogue of Life*

Work Package 5 provides the scientific co-ordination of the whole project. This includes the steady state synthesis and production of the Annual and Dynamic Checklists, the steady state Systems operation that operates both the internal and external services, and the co-ordination of the new features to be tested and introduced.

Given that the services are already widely used, it is important that they should be maintained and strengthened continuously throughout the three year project. One task that must be continued in steady state operation is the integration of the Catalogue and production of the Annual Checklist for publication each April. This involves both the introduction of new taxonomic sectors from new supplier GSD databases, and the integration of updated sectors from existing suppliers. Each year a number of new suppliers are added, and it is these that require the greatest work for the ‘admission gate’ process: recording metadata, technical scrutiny, peer review, administrative agreements, as well as test downloading data and writing and testing the wrapper program. Receiving or re-caching from the existing suppliers involves less time. Integrating the new sectors into the taxonomic Hierarchy and Checklist is a skilled taxonomic editorial task, part of the work of the Editorial and Production Team, divided between Reading and Los Baños in the Philippines.

A second task needed to maintain the services is the systems operation. Not only is there the need to operate components 24/7, but there are ongoing complexities to be managed, particularly in the area of assisting suppliers with writing and testing wrapper programs, operating the test and production hubs, and managing the availability and caching from the various databases – a complex task with 52 suppliers at present divided between Europe and world-wide.

Against this steady state background, a substantial set of changes have to be managed and intercalated during the three years of this project. These are:

- i) improvements planned for the production process in year 1, working with WP 6. For instance this includes the unification of the production process for the Annual and Dynamic Checklists, rather separate up till now, and a system giving more support from the metadatabase when new databases are connected. Implementation of these alterations will be managed by the head of the Editorial and production team. It will need to be closely co-ordinated with the annual production cycle.
- ii) Work with WP 2 and WP6 to select, design and create the new services, to test them, and then to introduce them into the operational system. A new services manager will assess the possible implementation of each suggested new service, taking into account both the supply side issues with GSD databases in WP 3, and the middleware/software enhancement issues with WP 6. Decisions as to which new services to go ahead and implement will be taken after a ‘triage’ between priorities set by the users and the practicalities of putting it in place.
- iii) Work with WP 4 to test the present Multi-Hub Network features and to agree a long-term specification.
- iv) Work with WP 7 and WP 6, to test the new distributed system, and then to roll it out in the third year.

This work package will also undertake the task of creating a second generation business plan, and of initiating the financial and legal engineering needed for the programme to approach sustainability. It will be undertaken with advice from the user community contacted in WP 2, with discussion with the GSDs partners in WP 3, as their endorsement is a prerequisite, and in partnership with the directors of Species 2000, the residual legal body through and for which financial sustainability will be achieved.

#### *WP 6 Middleware & Software Support Services*

Work Package 6 will provide a newly co-ordinated middleware and software support service within the Species 2000 community. In the past the programme has either adopted software products resulting from informatics research pilot projects, or contracted individual tasks to particular IT laboratories. Despite some extremely successful outcomes, this has left the programme dependant on more than one system, and in some cases on working prototypes that lacked the robustness, management features and documentation that are needed for what is becoming a mature public service. The middleware and software involved provide both the internal processes by which taxonomic sectors are harvested from the many databases, but also the user-facing processes by which the public services are served from the servers.

Responsibility for co-ordinating this service is taken in this project by ETI in Amsterdam, although they will carry out some of the work, and spread other tasks among other specialist members of the work package.

The work package undertakes quite a long list of tasks that are needed for four main themes:

Theme 1: Unification and improvement of the existing Checklist creation processes. The secure editorial process of the present synthesis and production of the Annual Checklist product needs to be combined with the automated caching process of the Dynamic Checklist product, and linked more closely with the metadatabase. This will be done by securing an altered base schema that is suitable for both systems, and making adjustment so that both systems can operate on this one schema and linked to the metadatabase. Once this change is made, all the supporting tools of both systems can be brought to bear on a single production process. Other associated improvements to the SPICE 5.1 distributed system concern a better wrapper kit for the GSD suppliers, better automation of the attachment process for new GSDs, and better management interface and functionality for caching and re-caching the sectors supplied by the GSDs. Many of these features are improvements in the present middleware as seen and used by the partners within the programme.

#### Theme 2: New services

WP 5 will co-ordinate the response to suggested new services, but this work package will assist them, first to assess how much software alteration may be needed for each, and once it is decided which to go ahead with, to make those alterations or new features prior to testing, and eventual implementation. The interface and software alterations made will primarily impact on users, but to provide content for these, it may be necessary to make middleware changes back to the origin of the information.

#### Theme 3: Multi-Hub connections

Features needed for the Dynamic Checklist to link regional hubs to the Global Hub were foreseen in the EuroCat project and first steps to implement these were made at the end in 2006. They were incorporated into SPICE version 5.2, the last version of the Dynamic Checklist software in that project, and used to link the three European regional databases into the Euro-Hub, and to link the Euro-Hub with the Global Hub. It is now proposed to extend that system so that the N. American Hub, and the China Hub can be linked in year 1 on an experimental basis. That structure, with three regional hubs connected in this simple way, will then provide a background test and prototype for the discussions and long-term specifications being formulated in WP 4.

Theme 4: Introducing the new e-2 infrastructure.

Scoping, designing, and preliminary testing of the new e-2 distributed system are the separate responsibility of WP 7, although many of the altered requirements will be coordinated with opinions from this work package. But once the new system prototypes have been tested and a second prototype made available, this work package will take over the process of preparing a second, production version of the system to be rolled out across the programme.

The separation of tasks between this Work Package 6, the specification of future needs for the Multi-Hub Network in WP 4, and the prototyping of the next generation e-2 distributed system in WP 7 means that for the first time the Species 2000 Catalogue of Life programme will have proper attention to the maintenance and functionality of its current production middleware and software, at the same time as others undertake the design and prototyping for the next generation of production systems.

### Service Activities: Work package list

Work package No	Work package title	Type of activity	Lead partic no.	Lead partic. short name	Person-months	Start month	End month
5	Scientific Services of the Catalogue of Life	SVC	1	UREAD	150	m1	m36
6	Software Support Services	SVC	5	ETI	85	m1	m36
	TOTAL			2	235		

## Service Activities: List of Deliverables

<b>Del. no.</b>	<b>Deliverable name</b>	<b>WP no.</b>	<b>Nature</b>	<b>Dissemination level</b>	<b>Delivery date (proj. month)</b>
<b>D6.4</b>	<b>GSD Wrapper Kit</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m12</b>
<b>D6.5</b>	<b>Metadatabase</b>	<b>6</b>	<b>D</b>	<b>PP</b>	<b>m12</b>
<b>D5.1</b>	<b>Base schema</b>	<b>5</b>	<b>D</b>	<b>PP</b>	<b>m12</b>
<b>D6.2</b>	<b>New web-services</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m14</b>
<b>D6.6</b>	<b>Improved SPICE</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m16</b>
<b>D6.1</b>	<b>Enhanced services</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m18</b>
<b>D6.3</b>	<b>GSD builder</b>	<b>6</b>	<b>D</b>	<b>PP</b>	<b>m18</b>
<b>D6.7</b>	<b>SOICE Documentation</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m18</b>
<b>D5.2</b>	<b>New services</b>	<b>5</b>	<b>D</b>	<b>PU</b>	<b>m24</b>
<b>D5.4</b>	<b>Business plan</b>	<b>5</b>	<b>R</b>	<b>PP</b>	<b>m30</b>
<b>D6.8</b>	<b>End user documentation</b>	<b>6</b>	<b>R</b>	<b>PU</b>	<b>m34</b>
<b>D6.9</b>	<b>e-2 system</b>	<b>6</b>	<b>D</b>	<b>PU</b>	<b>m34</b>

## Work package 5: Scientific Services of the Catalogue of Life

<b>Work package number</b>	<b>WP 5</b>		<b>Start date or starting event:</b>			<b>Month 1</b>	
<b>Work package title</b>	<b>Scientific services of the Catalogue of Life</b>						
<b>Activity type<sup>20</sup></b>	<b>SVC</b>						
<b>Participant number</b>	<b>1</b>	<b>8</b>	<b>2</b>	<b>3</b>	<b>5</b>	<b>6</b>	
<b>Participant short name</b>	<b>UREAD</b>	<b>Sp2000</b>	<b>BGCI</b>	<b>MNHN</b>	<b>ETI</b>	<b>CU</b>	
<b>Person-months per participant</b>	<b>110</b>	<b>36</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	

### Objectives

- i) To strengthen and substantially enhance the public scientific infrastructure services provided by the Species 2000 Catalogue of Life programme, including an array of novel services both at the popular and educational level, and at the electronic services level. This will involve engaging closely with the User and Partner communities through the networking tasks of WP 4, as well as working with the knowledge providers in the GSD Network of WP 2, and of the new Multi-Hub Network in WP 3.
- ii) To strengthen and make secure the internal community services that take the raw taxonomic materials from an ecosystem of providers and synthesise it into the authoritative Species 2000 Catalogue of Life and its public infrastructure services. These internal services include the year on year production of the Annual Checklist; the work of the Editorial Team in taxonomic integration, editorial scrutiny and quality control; and the operational systems management needed to operate the internal and public services continuously.
- iii) To establish an outreach and promotional programme that will disseminate the infrastructure services even more widely than has occurred so far. The objective is to raise the profile of what has up to now been a purely scientific project, and to establish the Species 2000 Catalogue of Life as one of the major international biodiversity programmes that can take its place as a partner and provider at the global scale alongside GBIF, EoL, BHL, IUCN, GenBank and CBOL.
- iv) To undertake the financial and legal engineering needed to make the Species 2000 Catalogue of Life programme sustainable as a scientific infrastructure service on a European and global scale.
- v) To undertake the scientific co-ordination of the whole project whose central objective is to enhance and secure the public services for the Species 2000 Catalogue of Life, for which this work package has a central role, and the other work packages a supporting role.

### Description of work (possibly broken down into tasks) and role of partners

- Task 5.1 Design, test and implement an array of enhancements to the present public services, with WP 4, 1 & 6.
- Task 5.2 Design, test and implement novel services at the technical level, with WP 4, 1 & 6.
- Task 5.3 Design, test and implement novel services of an educational and popular nature, with WP 4, 1 & 6
- Task 5.4 Enhance and operate the annual production service for the Annual Checklist, including unification of the admission gate for the Annual and Dynamic Checklist products, with WP 6.
- Task 5.5 Enhance, secure and operate the editorial, scrutiny and quality control services for the unified production route.
- Task 5.6 Secure and operate the ongoing systems management for the internal and public services of the whole programme.
- Task 5.7 Establish an enhanced business plan for sustainability, and undertake associated financial and legal engineering.
- Task 5.8 Co-ordination of scientific and IT components of the entire project, including the interaction of this central work package with all the other work packages.

### Deliverables (brief description) and month of delivery

- D5.1 M12 Base schema and associated tools accepted for editing and attaching databases.
- D5.2 M24 Unified production of Annual & Dynamic Checklists in place
- D5.3 M24 First two new services in place
- D5.4 M30 Enhanced business plan and financial engineering for sustainability in place.

D5.5 M36 Second batch of four new services in place.

**Milestones**

M5.1 Meeting convened for Sustainability (month 12)

M5.2 Existing Service Enhancement in place (month 18)

## Work package 6: Support Services for Software Deployment

<b>Work package number</b>	<b>WP 6</b>			<b>Start date or starting event:</b>		<b>Month 1</b>	
<b>Work package title</b>	<b>Support Services for Software Deployment</b>						
<b>Activity type<sup>20</sup></b>	<b>SVC</b>						
<b>Participant number</b>	<b>5</b>	<b>3</b>	<b>9</b>	<b>6</b>	<b>34</b>	<b>1</b>	
<b>Participant short name</b>	<b>ETI</b>	<b>MHHN</b>	<b>VLIZ</b>	<b>CU</b>	<b>CRIA</b>	<b>UREAD</b>	
<b>Person-months per participant</b>	<b>68</b>	<b>5</b>	<b>2</b>	<b>1 (WP7)</b>	<b>8</b>	<b>1 (WP5)</b>	

### Objectives

To improve the usability, assembly, maintenance and sustainability of the Catalogue of Life by:

- i) Maintaining the current infrastructure but updating the software into a robust, reliable and stable suite of production software components achieved through standardized software design, coding and testing, and based on existing standards and scientific software versions, for both the internal and external services of the Species 2000 Catalogue of Life (CoL) programme.
- ii) Enhancing the assembly applications to combine and improve the quality checking processes which are currently applied separately to the Annual Checklist (AC) and Dynamic Checklist (DC), to create a single integrated quality control process and simplified data handling, without delaying the current processes.
- iii) Enhancing the maintainability of the programme's software by using a Concurrent Versions System (CVS), standardized code documentation, user documentation, and technical support (helpdesk) and a publicly accessible CoL applications repository.
- iv) Developing additional and variant applications to support enhanced and novel services, and producing middleware to support a robust and fail-safe 24/24 available Multi-Hub system.
- v) Improving and converging the various data schemas into a central base schema and preparing a production version of the new e-2 service based infrastructure designed in WP7 that can replace the current infrastructure and can support the planned growth of the CoL data network.

### Description of work (possibly broken down into tasks) and role of partners

#### *Public applications Services*

Task 6.1 Provide stability and functional enhancements to the present user interfaces, search facilities and web-services, using input from WP 5 & 4. Specification testing, blackbox usability and stability testing involving all partners.

Task 6.2 Test and provide additional or variant software for the novel technical, education and popular services, following selection and prioritization by WP5

Task 6.3 Improve documentation and version control.

#### *Middleware for Internal Services*

Task 6.4 Hold design team meeting with WP7 and WP5 to determine the middleware strategy and changes.

Task 6.5 Define a new central base schema to assist unification of all assembly and publication tools, based on current schemas in AC en DC; and produce a Global Species Database (GSD) builder tool (batch functions) and GSD editor tool (atom by atom editing) based on the new schema pre-fitted for uploading to the SPICE cache (SPICE is a distributed-computing engine that runs the Dynamic Checklist). (with Paris & VLIZ provides prototype, ETI develops production version)

Task 6.6 Prepare robust components for the SPICE 5.1 interoperability system without changing the fundamental communication protocol, enabling steps towards unification of the AC and DC production streams; produce control components for assembled data.

Task 6.7 Produce a universal GSD wrapper [kit] for data providers to link GSDs to the SPICE system (with CRIA)

Task 6.8 Alter the SPICE DC cache to adopt the new central base schema, Alter the GSD editing and inspection tools that currently work with the AC to work with the new base schema and make these available for editorial staff to edit and inspect the cache an intermediate state datasets in the base schema.

Task 6.9 Produce a stable production tool for linking GSDs to the SPICE system using experiences from task

6.8 , including a new version of the Metadatabase software and a tool for AC export & conversion from the DC cache for publication on CD.

Task 6.10 Prepare a version of the SPICE 5.1 system for supporting linkages, cross-maps and enhanced interface for use in the Multi-Hub Network.

Task 6.11 Prepare documentation and version control system for all SPICE components. (with CARDIFF)

Task 6.12 Prepare a production version of the components in the new e-2 service-based system from WP7.

**Deliverables** (brief description) and month of delivery

D 6.1 Enhanced public services and software including documentation and internationalisation (month 18)

D 6.2 new webservices, novel services defined in WP2 and mobile service available (month 14)

D 6.3 GSD builder and editor (month 18)

D 6.4 universal GSD Wrapper kit prototype (month 12)

D 6.5 Improved metadatabase (month 12)

D 6.6 Improved SPICE components (month 16)

D 6.7 Software code documentation and user documentation of middleware (Spice) components (month 18)

D 6.8 Software code documentation and user documentation of end user applications (month 34)

D 6.9 New e-2 service based system (month 34)

**Milestones**

M6.1 current CoL software system enhanced and available (month 18)

M6.2 new e-2 system in production state and available (month 36)

M6.3 Copies of all software components available in a public software repository at Reading.  
(month 36)

### Service Activities: Summary of effort

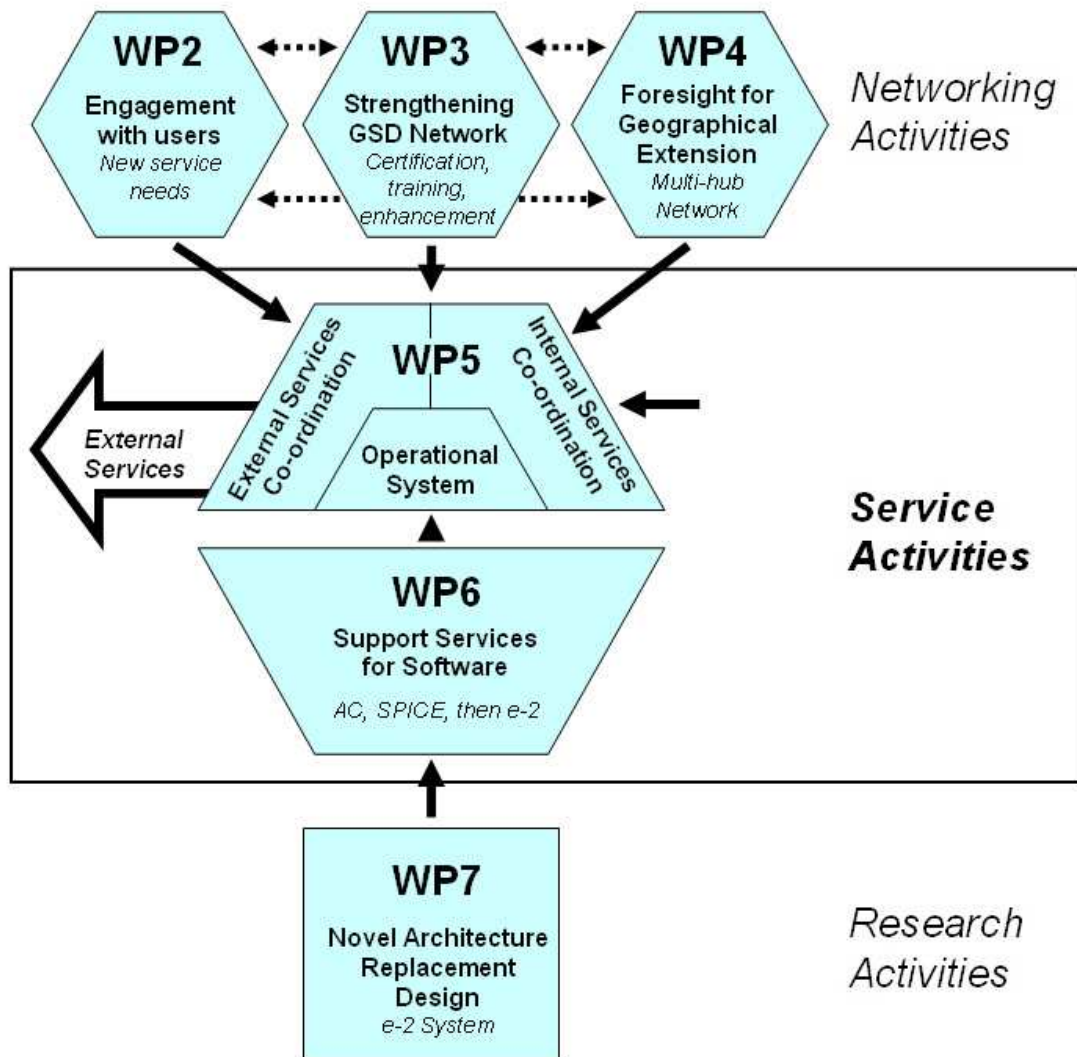
A summary of the effort is useful for the evaluators. Please indicate in the table number of person months over the whole duration of the planned work, for each work package by each participant. Identify the work-package leader for each WP by showing the relevant person-month figure **in bold**.

Partic. no.	Partic. short name	WP 5	WP 6	Total person months
<b>1</b>	<b>UREAD</b>	<b>110</b>	<b>1</b>	<b>111</b>
<b>2</b>	<b>BGCI</b>	<b>1</b>	<b>-</b>	<b>1</b>
<b>3</b>	<b>MNHN</b>	<b>1</b>	<b>5</b>	<b>6</b>
<b>5</b>	<b>ETI</b>	<b>1</b>	<b>68</b>	<b>69</b>
<b>6</b>	<b>CU</b>	<b>1</b>	<b>1</b>	<b>2</b>
<b>9</b>	<b>VLIZ</b>	<b>-</b>	<b>9</b>	<b>9</b>
<b>34</b>	<b>CRIA</b>	<b>-</b>	<b>8</b>	<b>8</b>
<b>36</b>	<b>Sp2000</b>	<b>36</b>	<b>-</b>	<b>36</b>
<b>Total</b>		<b>150</b>	<b>92</b>	<b>242</b>

### Service Activities: Milestones

Milestones are control points where decisions are needed with regard to the next stage of the project. For example, a milestone may occur when a major result has been achieved, if its successful attainment is a required for the next phase of work. Another example would be a point when the consortium must decide which of several technologies to adopt for further development.

Milestone number	Milestone name	Work package(s) involved	Expected date	Means of verification
M5.1	Sustainability	WP5	m12	Report
M6.1	CoL system	WP6	m18	Report
M5.2	Enhancements	WP5	m18	Report
M6.2	e-2 ready	WP6	m36	Report
M6.3	Repository	WP6	m36	Report



Relationships between Workpackages



standards, particularly of the TDWG organisation. We now need to make use of these more widely used generic standards.

iii) The most important reason is simply that the requirements for the architecture have changed.

- One issue is that the 'ecosystem' of providers and harvesters has changed. (Architecture 1 referred to in the introduction will here be referred to as 'knowledge architecture 1' to distinguish the knowledge architecture from the computational architecture used in the implementation – they need not be congruent.) Knowledge Architecture 1 required a single level array of GSD databases to be harvested by a single global hub. Discussions in WP 3 of this project will finalise ideas on inserting a layer of providers (the Nomenclators – databases of names) 'beneath' the layer of GSDs. These nomenclators can assist the GSDs by sending them alerts to the publication of new names, even though they cannot provide the expertise that the GSD provides, of positioning these newly discovered organisms within the classification. Similarly the discussion and specifications for the Multi-Hub Network in WP 4 will consolidate the higher level array of multiple hubs. So what started as a simple one-layer distributed array, does now resemble the idea of an ecosystem with a three level 'food chain' of providers and harvesters.

- Another issue is that the original federated model allowed for real time harvesting, as there were severe reservations about permissions to cache our partners' data. In fact the community has come to respect and trust the responsible handling of other people's data by Species 2000 both by seeing how it has behaved, and by virtue of the Access Agreements agreed with the suppliers, and their becoming members and part owners of the system and its governance. It also became clear that real-time harvesting required unrealistic performance over the network, leading in practice to poor response times, something much improved by caching. As a result the system requirement now is to operate from a cache and indeed with a buffer with potential editorial intervention, and certainly not in real time harvesting.

- A further issue is of the openness of the ecosystem. In the present system the individual GSD SPICE wrappers would be open to harvesting by any organisation using the SPICE protocols, but in reality other organisations have not done so. There is the prospect that new standard providers coming from TDWG or GBIF may make it possible for a database to serve specimen and taxon data using a common provider protocol, thus avoiding the need for different wrapper/provider middleware for different projects, and making individual service providers available for direct harvesting by others.

Several other features may be needed in the new architecture.

- i) integration of LITCHI tools for testing the integrity of checklists covering the same species in different hubs
- ii) possible 'whole biota arithmetic' and other features from the Multi-Hub Network, some requiring advanced network capacities
- iii) structures for mirroring between the regional hub sites, both for resilience against down times, and for load-sharing under heavy load.

The proposal is to make a fresh requirements analysis in year 1, and to enhance this with possibly additional requirements coming from the new services proposed in WP2, and the needs of the Multi-Hub network coming from WP 4. The new system is arbitrarily referred to as the 'e-2' system until a name or acronym can be found.

### Joint Research Activities: Work package list

Work package No	Work package title	Type of activity	Lead partic no.	Lead partic. short name	Person-months	Start month	End month
7	Novel e-Science Service-based Architecture	RTD	6	CU	56	M1	M36
	<b>TOTAL</b>			1	56		

### Joint Research Activities: List of Deliverables

Del. no.	Deliverable name	WP no.	Nature	Dissemination level	Delivery date (proj. month)
D7.1	Design of the new architecture, 'e-2' for the sustainable Catalogue of Life	7	Document	Public	12
D7.2	Implementation of a prototype system (revised/enhanced though testing) in the new architecture, 'e-2'	7	Software	Semi-public (closed user group)	24

## Work package 7: Novel e-Science Service-based Architecture

<b>Work package number</b>	<b>WP 7</b>	<b>Start date or starting event:</b>				<b>Month 1</b>
<b>Work package title</b>	<b>Novel e-Science Service-based Architecture</b>					
<b>Activity type<sup>20</sup></b>	<b>RTD</b>					
<b>Participant number</b>						
<b>Participant short name</b>	<b>CU</b>	<b>UREAD</b>	<b>ETI</b>			
<b>Person-months perparticipant</b>	<b>54</b>	<b>(WP5)</b>	<b>(WP6)</b>			

### Objectives

- i) To replace the existing distributed model with a purpose-built service-based e-Science Architecture.
- ii) To build the new Architecture around the ‘open ecosystem’ model in which all distributed components are implemented as open services, so that user or partner communities can integrate both low- and high-level services into their own e-Science communities.
- iii) To build the new Architecture to reflect the recent enlargement and raised objectives of the Catalogue of Life community. This will include replication from one distributed hub to multiple distributed hubs, for the need for the enlarged concept to be managed by an overall management system, and for the management to become a 24/7 operation involving several sites around the world.

### Description of work (the system to be implemented under this new Architecture is here referred to as ‘e-2’.

- Task 7.1 To assemble the requirements and specification for the new Architecture and the new implementation of the ‘e-2’ system.
- Task 7.2 To design the e-2 implementation with an open service-based architecture.
- Task 7.3 To establish common communication protocols, data model, and management framework suitable for e-2.
- Task 7.4 To implement as ‘proof-of-concept’, new e-2 architecture, test tools, and management tools.
- Task 7.5 To create and install a test implementation for e-2 with three GSDs as test-beds
- Task 7.6 To respond to testing with an enhanced prototype for e-2

### Deliverables

- D7.1. Design of the new architecture, ‘e-2’, for the sustainable Catalogue of Life (m12)
- D7.2. Implementation of a prototype (revised/enhanced through testing) system in the new architecture, ‘e-2’. (m24)

### Joint Research Activities: Summary of effort

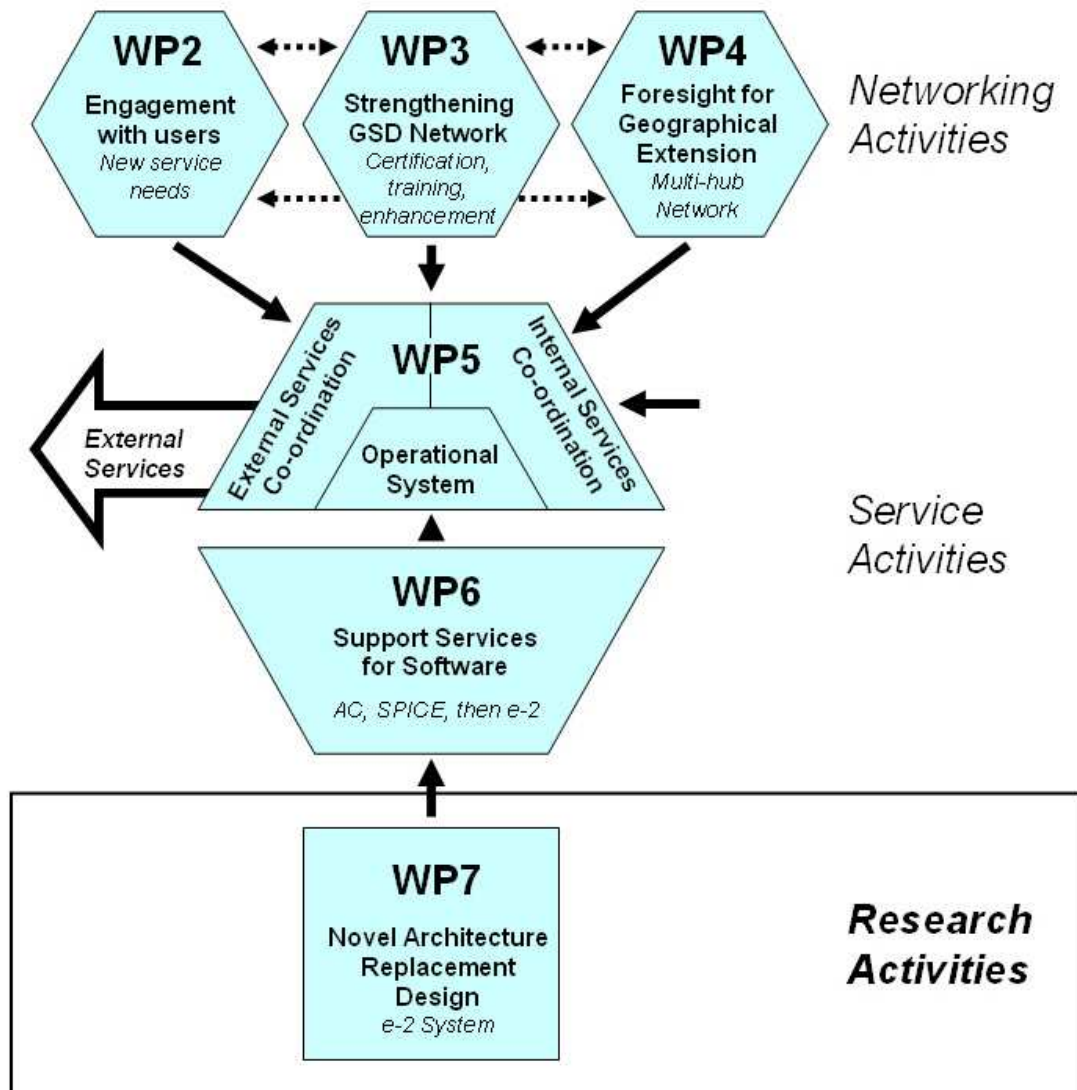
A summary of the effort is useful for the evaluators. Please indicate in the table number of person months over the whole duration of the planned work, for each work package by each participant. Identify the work-package leader for each WP by showing the relevant person-month figure **in bold**.

Partic. no.	Partic. short name	WP 7	Total person months
<b>1</b>	<b>UREAD</b>	<b>1</b>	<b>1</b>
<b>5</b>	<b>ETI</b>	<b>1</b>	<b>1</b>
<b>6</b>	<b>CU</b>	<b>54</b>	<b>54</b>
<b>Total</b>		<b>56</b>	<b>56</b>

### Joint Research Activities: Milestones

Milestones are control points where decisions are needed with regard to the next stage of the project. For example, a milestone may occur when a major result has been achieved, if its successful attainment is a required for the next phase of work. Another example would be a point when the consortium must decide which of several technologies to adopt for further development.

Milestone number	Milestone name	Work package(s) involved	Expected date	Means of verification
M1	e-2 design	WP7	m12	Report
M2	e-2 test	WP7	m18	Report
M3	e-2 2 <sup>nd</sup> ver	WP7	m24	Report



Relationships between Workpackages

## Section 2: Implementation

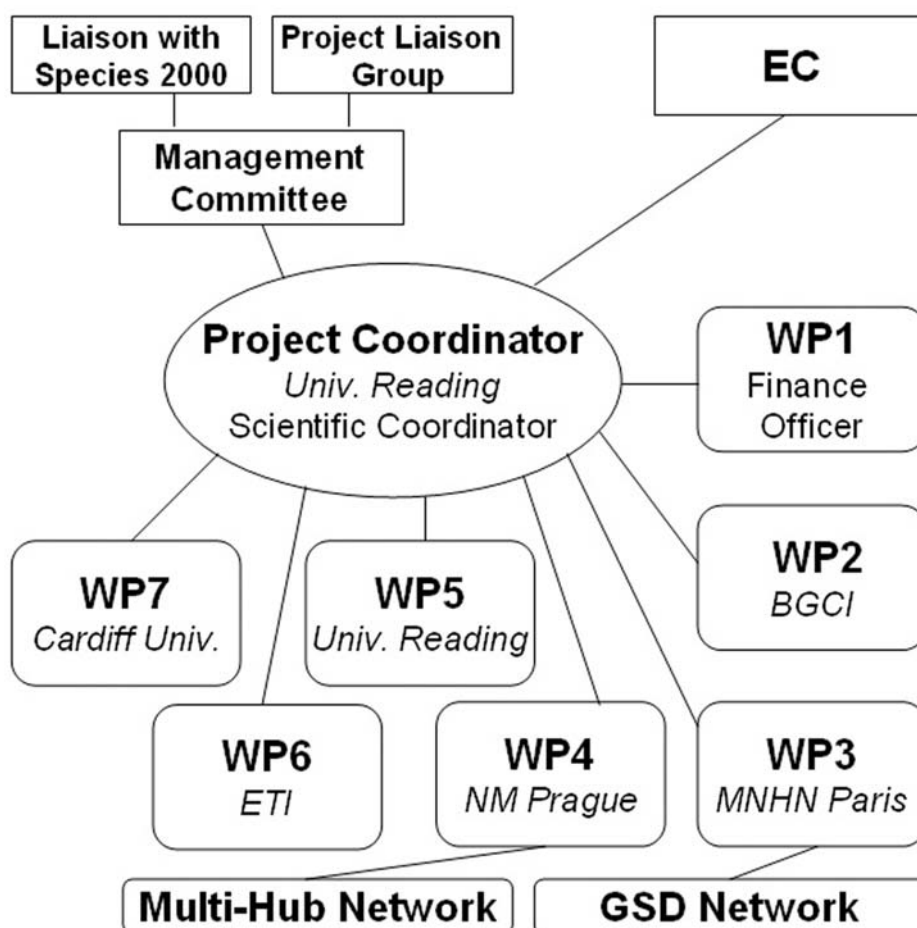
### 2.1 Management Structure and Procedures

#### Management Structure

4D4Life will be organised from the Project Coordinator's institute, the University of Reading (UREAD). The organisation of the consortium will consist of:

- i) Project Coordinator
- ii) Scientific Coordinator (full time employee of WP 5)
- iii) Finance Officer (part time employee of WP 1)
- iv) Work Package Leaders (of WP2, WP 3, WP 4, WP 6 and WP 7)
- v) Management Committee
- vi) Project Liaison Group

The management structure of 4D4Life is illustrated below. The structure is simple and straightforward, and well suited to the size and scope of the project. In addition to the main members of the work packages, the consortium contains the GSD Network members and the Regional Multi-Hub Network members.



4D4Life Management Structure

In addition to the management structures shown, there will be four project meetings at which all members of the consortium will meet to review the project and its progress:

- a) Month 1 Project Start-up Meeting with start-up meetings of work packages
- b) Month 15 Project Meeting and associated workshops of work packages
- c) Month 29 Project Meeting and associated workshops of work packages
- d) About M 36 Showcase Meeting with a wider public audience

The responsibilities of the members in the organisation are as follows:

**The Project Coordinator** (F. Bisby) has the overall responsibility for the project. He is in charge of organising and running the 4D4Life networking and coordination. The Project Coordinator oversees the partner's progress according to the work package descriptions, evaluates that progress, the meeting of milestone and deliverable dates, creation of final reports, motivates and monitors collaboration across institutions and disciplines, encourages joint publications, is responsible for the project and Liaison Group Meetings and their reporting, and organises resolution of any administrative, scientific or technical conflict that may occur. The Project Coordinator is also responsible for WP1 to manage the consortium and the financial contribution, and is responsible for project communications with users and with programmes across the biodiversity landscape through the project web-site, through the Project Liaison Group, through representing the project at European and international meetings and through the Showcase Meeting at the end of the project. The Project Coordinator has responsibility for all forms of reporting to the European Commission, for promoting gender equality in the project, and for conflict resolution should it occur.

**The Scientific Coordinator** (located at University of Reading, employed under WP5) will assist the project Coordinator with internal and external communication (including the use of a wiki, and realisation of the 4D4Life web-site), the organisation of the Project Meetings, and the recording of scientific reporting and communication with the Commission. This will occupy half that person's time, the other half being work under WP5 on designing and realising services.

**The Financial Officer** (located at University of Reading, employed part-time under WP1) is responsible for the claiming, distribution, accounting and reporting of funds distributed from the community financial contribution from the Commission.

**The Management Committee** consists of the work package leaders and is chaired by the Project Coordinator. The Management Committee follows progress of the project on a monthly basis via telephone conferencing, and normally two meetings per year, including one at each Project Meeting. Any deviations from the work plan, as well as methods for dealing with these deviations, will be discussed, and action agreed at the project level and/or in communication with the Commission. This will also be the place to resolve technical conflicts that may arise between work packages.

Agenda items will be circulated at least 7 days ahead of the meeting, decisions taken by majority vote, and minutes recorded and circulated within one month. A member of the Commission will be invited to join the management Committee and included in the circulation of agenda and minute papers.

**The Project Liaison Group** is a liaison group invited to meet with the Management Committee on just a few occasions to maintain awareness and liaison between 4D4Life and the ring of related but separate projects within the taxonomic and biodiversity informatics landscape, particularly in Europe. The following have initially been invited to join this group:

Professor Simon Tillier, EC EDIT  
Dr David Remsen, GBIF-ECAT  
Professor David Patterson, EoL Informatics  
Dr Yde de Jong, EC PESI  
Professor Stefan Claesson, CETAF  
Professor Walter Berendsohn, EC BioCase

(The Group may meet in association with certain of the 4D4Life Project Meetings such as the Showcase meeting, but also possibly at other venues including those organised, for example, by EDIT, or GBIF.)

**Liaison with Species 2000.** 4D4Life will become a significant component of the wider Species 2000 Programme, and as such will need to work closely both with the Species 2000 Global Team that coordinates scientific policy for the programme world-wide, and with other funded projects within the wider programme. This is needed because there are projects and contributors to the wider programme that are not based in Europe, and in which the partners are not part of 4D4Life. Species 2000 has indicated that it will invite the Project Coordinator (who is already a member) and one other member of the Management Committee to join the Global Team and act as regular conduit for coordination and cooperation between the wider programme and 4D4Life. Communication with the Global Team is important for this project to reach its full potential, and it is possible that two of the Global Team meetings can be arranged to coincide with two of the 4D4Life Project Meetings. An outline of the 4D4Life proposal was endorsed by the Global Team at its April 2008 meeting.

A similar arrangement worked well between the Global Team and the FP 5 Eurocat project in 2003 - 2006, with back-to-back and joint meetings held at the Malta and Stockholm project meetings of that project.

## **2.2 Individual Participants**

### **1. University of Reading, Reading (UREAD)**

The Centre for Plant Diversity and Systematics, part of the School of Biological Sciences at the University of Reading, is known world-wide for its leadership of flagship research programmes in taxonomy - Flora Europaea, the UNEP Global Biodiversity Assessment, the Families of Flowering Plants, ILDIS and most recently the Species 2000 and ITIS Catalogue of Life, BiodiversityWorld (BDWorld). The Centre has a proven track record in coordinating EC FP5 project Species 2000 europa (EuroCat) & Euro+Med, and BBSRC projects (SPICE, LITCHI, BDWorld). Participating in others EC projects: ERMS, Fauna Europaea, BioCASE, ENBI, EDIT, LifeWatch.

The Centre will provide management infrastructure for operation of two GSDs in this project: ILDIS and Conifer database.

**ILDIS World Database of Legumes** by YR Roskov, FA Bisby, JL Zarucchi, BD Schrire & RJ White (eds). Taxonomic database of world legumes: Plantae – Magnoliophyta – Magnoliopsida – Fabales – Fabaceae/Leguminosae; 19,937 spp.

**Conifer Database** by A Farjon. Taxonomic database of world conifers: Plantae – Pinophyta; 631 spp.

**Tasks:** Consortium management and finances in WP1.

Coordination of WP5 “Scientific services of the Catalogue of Life”. Enhance the public scientific infrastructure of the Catalogue of Life, including an array of novel public and electronic services. (Design, test and implement (i) enhancements to the present public services, (ii) novel services at the technical level, (iii) novel educational and popular services). Enhance and operate the annual production service for the Annual Checklist, including unification of the admission gate for the Annual and Dynamic Checklist products. Enhance, secure and operate the editorial, scrutiny and quality control services for the unified production route. Secure and operate the ongoing systems management for the internal and public services of the whole programme. Establish an enhanced business plan for sustainability, and undertake associated financial and legal engineering. To undertake the scientific co-ordination of the whole project (co-ordination of scientific and IT components of the entire project, including the interaction of this central workpackage with all the other workpackages).

ILDIS and Conifer databases will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Prof. Dr. Frank Bisby** (Project Co-ordinator, WP5 leader), played a significant role in establishing biodiversity informatics, as a lecturer and then reader at the University of Southampton, and now as Professor of Botany and Director of the Centre for Plant Diversity & Systematics and member of the Informatics Research Centre at the University of Reading. He is a founding member of the Biodiversity Information Standards organisation (TDWG), instigator of one of the first global species databases, ILDIS, and leader of the Species 2000 organisation since its launch in 1996, as well as co-ordinator of EC FP5 EuroCat project, member/workpackage leader in other EC projects such as ERMS, Euro+Med, ENBI, and EDIT. As

Executive Director of Species 2000 he works with what is now an extensive network of taxonomic database hubs around the world from China and New Zealand to Brazil and N. America, in projects to approach completion of known organism coverage in the Catalogue of Life, and to work with partner organisations on the Global Names Architecture.

**Dr Yury Roskov**, plant taxonomist, Catalogue of Life content manager.

Executive editor of the Annual Checklist editions 2006, 2007 & 2008. Co-director and responsible manager of ILDIS database.

**Dr Aljos Farjon**, world leading taxonomist of conifers, author and custodian of the Conifer database.

## 2. Botanic Gardens Conservation International (BGCI)

Botanic Gardens Conservation International (BGCI) is a UK Registered Company limited by guarantee (Company No. 04673175) and a Registered Charity (No. 1098834). We operate as a membership organisation that represents and unites the skills and expertise of botanic gardens globally, and works with partners, to conserve the world's imperilled flora. Linking more than 800 botanic gardens in 118 countries BGCI forms the world's largest plant conservation network. The HQ of BGCI is based in the UK at Royal Botanic Gardens, Kew. We have offices in South China Botanical Garden, Chicago Botanic Garden and Singapore Botanic Garden. We have a staff member based in the IUCN East Africa Office, and also leading a regional bioinformatics project for East Africa. BGCI also has national representatives in Colombia, Germany, the Netherlands, Russia and Spain.

BGCI provides the Secretariat for the Global Partnership for Plant Conservation (GPPC) mandated by the CBD to facilitate the implementation of the Global Strategy for Plant Conservation. The GPPC currently has over 30 member in a loose affiliation including IUCN, FAO, BioNet International, Bioversity International and other major providers and users of biodiversity information.

BGCI is a member of GBIF, a member of the Encyclopedia of Life Institutional Council and a member of the EoL Species Sites Group for plants. BGCI has good contacts with the corporate sector and is currently in discussion with or developing partnerships with companies in the global banking, pharmaceutical, agricultural, energy and transport industries.

### Tasks:

Coordination of WP2 "Engagement with scientific users and partners". To engage with (i) the major global biodiversity portals and biodiversity organisations, (ii) national portals and programmes, NGOs, and higher education institutes, (iii) individuals who make serious use of the Catalogue of Life Services, and review current and potential uses and enhancements in these groups. To locate and engage with the array of commercial organisations that either do use the Catalogue of Life Services, or may be considered likely to do so. To establish an outreach and promotional programme that will disseminate the infrastructure services. To establish the Catalogue of Life as one of the major international biodiversity programmes that can take its place as a partner and provider at the global scale alongside GBIF, EoL, BHL, IUCN, GenBank and CBOL. Working with the users and partner organisations, to specify, and subsequently test a range of enhanced and novel scientific services.

**Sara Oldfield**, Secretary General of BGCI with responsibility for overall management and development of the organization. She has 30 years experience of working in the biodiversity sector, including data analysis, policy development and management of a wide range of conservation initiatives..

**Suzanne Sharrock**, Global Programmes Director for BGCI. She manages BGCI's climate change and human well-being programmes. Suzanne also works on policy issues connected with the [Global Strategy for Plant Conservation](#).

**Meirion Jones**, BGCI's Head of Information Management. He maintains and develops the web-based PlantSearch and GardenSearch Databases as part of BGCI's overall information strategy.

### 3. Muséum National d'Histoire Naturelle, Paris (MNHN)

MNHN is a scientific institute dealing with biodiversity conservation. It is the scientific advisor of the French ministry for environment. Its areas of activities is research, education and training managements and enrichment of its collection (almost 80 millions of specimens), expertise and diffusion of scientific knowledge. The MNHN plays a key role in the organisation of the French information system on biodiversity and landscape, being the scientific coordinator of this national project. It also homeed the European thematic centre for biodiversity of the European environmental agency. In addition to international research collaboration at the individual level, MNHN plays also several important international roles. It has been and is heavily involved in various European programmes related to biodiversity:

- The European Topic Centre on Nature Protection and Biodiversity is homeed by MNHN.
- The French national GBIF, GTI and CBD nodes, the French Focal Point for the EU-funded BioCASE and PESI projects are located in the MNHN.
- Leading partner / work package leader in EC ParSyst, ColParSyst, Fauna Europaea, ENBI and EuroCAT, Synthesys 1, EDIT, MarBEF.

MNHN is a home organization for four GSDs:

**FLOW: Fulgoromorpha Lists On the Web** by T Bourgoïn. Taxonomic database of world planthoppers: Animalia – Arthropoda – Insecta – Hemiptera – infraorder Fulgoromorpha; 9,209 spp.

**COOL: Cercopoidea Organised On Line** by A Soulier-Perkins. Taxonomic database of world froghoppers: Animalia – Arthropoda – Insecta – Hemiptera – superfamily Cercopoidea; 2,086 spp.

**Psyl'list: Psylloidea database** by D Ouvrard. Taxonomic database of world jumping plant-lice: Animalia – Arthropoda – Insecta – Hemiptera – superfamily Psylloidea; 1,756 spp.

**CLEMAM** by P Bouchet. Taxonomic database of molluscs. Animalia – Mollusca.

#### Tasks:

Coordination of WP3 “Strengthening the GSD Partner Network: a Virtual Community”. Assesment of participating GSDs. Network consultation and workshop with GSD custodians towards Catalogue of Life standardisation. To accelerate and enhance GSD completeness and data quality. To coordinate set of Pilot Projects for enhancement of the Catalogue of Life Services for Users.

FLOW, COOL and Psyl'list databases will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Prof. Thierry Bourgoïn**, Deputy Director of the MNHN Collections, coordinator of Species 2000 europa database network (EuroCat & PESI), member of the Species 2000 Team and Taxonomy group, leading taxonomist of planthoppers, author and manager of FLOW database.

**Dr Philippe Bouchet**, taxonomist of molluscs, editor of CLEMAM database.

**Dr Adeline Soulier-Perkins**, taxonomist of froghoppers, author of COOL database.

**Dr David Ouvrard**, taxonomist of jumping plant-lice, author of Psyl'list database.

#### **4. Narodni Museum, Prague (NMP)**

The Museum is a public scientific institution which systematically enriches its collections consisting of objects of natural and historical sciences all over the world, with particular interest to the Czech Republic. It conducts research in various fields of natural and historical sciences and has a large exhibit activity. The National Museum is the most distinguished and the largest museum in the Czech Republic. It consists of five professional institutions: Natural History Museum, Historical Museum, The Náprstek Museum of Asian, African and American Cultures, Czech Museum of Music, National Museum Library. These are collections of mineralogy, palaeontology, mycology, botany, entomology, zoology, anthropology, archaeology. From the very beginning the collections have been regarded also as a treasury of the most important monuments of the Czech national history. They are located in more than 15 public museums and exhibition halls all over the country. The museum publishes 12 scientific journals and numerous other non periodical publications.

#### **Tasks:**

Coordination of WP4 “Foresight study for geographical extension: the Global Multi-Hub Network”. Network consultations and workshop discussions among regional hub organisers (to bring together for the first time leaders of the various international centres from around the world that have committed to create Regional Hubs in the Sp2000 Multi-Hub Network, so that they can work together in Networking, Discussion and Design). Pilot Projects for prototyping and testing Multi-Hub Linkage tools, installing and testing Multi-Hub linkages using tools from the new e-2 Architecture of WP 7.

**Dr Jiří Kvaček**, Head of the Department of Palaeontology, leader of Science programmes. His scientific work is focused on palaeobotany of gymnosperms and angiosperms. Since 1987 he is working for the National Museum in Prague in various functions. From 2005 to 2008 he was science director of the National Museum. He was involved in management of several scientific projects in palaeontology, a project of Scientists night and three exhibitions. In 2007 he took part in the management team responsible for the development of the new permanent exhibition of the National Museum, Prague.

## **5. Stichting Expertisecentrum voor Taxonomische Identificatie, Amsterdam (ETI BioInformatics)**

ETI BioInformatics is a project-driven organization with ample experience in biodiversity informatics and service development. It promotes access to biodiversity information and specializes in the development and implementation of professional, sustainable information infrastructures that facilitate standardization, integration, interoperability, accessibility and efficient dissemination of biodiversity information. ETI expertise concerns:

- database implementation, information management tools, user interfaces
- interoperability solutions, data merging, database wrapping and webservice
- web portals, dynamic websites, mobile applications
- computer aided identification, diagnostic systems, e-learning applications
- e-content management, product development, exploitation planning and marketing

With and for a global network of taxonomists and biodiversity specialists ETI developed tools and a mechanism for the World Biodiversity Database: a series of over 100 e-monographs published on CD-ROMs and the web. ETI is one of the founders of Species2000 and closely involved with the technical aspects of producing the CoL Annual Checklist. ETI participates in a range of EC and global projects and has an excellent track record with the timely delivery of professional ICT applications, e.g.: in Fauna Europaea, ENBI, EuroCat, LifeWatch, GBIF, MARBEF, KeyToNature. Other recent partners include: Netherlands GBIF Node, Tanzanian GBIF Node, US Dept of Agriculture, NL Ministry of Agriculture Nature Conservation and Food, National Herbarium NL, Zoological Museum Amsterdam. ETI is a member of GBIF, EMBNet and participates in TDWG.

### **Tasks:**

Coordination of WP6 “Support Services for Software Deployment”.

(i) Public applications Services. Provide stability and functional enhancements to the present user interfaces, search facilities and web-services, using input from WP 5 & 4. Specification testing, blackbox usability and stability testing involving all partners. Test and provide additional or variant software for the novel technical, education and popular services, following selection and prioritization by WP5. Improve documentation and version control.

(ii) Middleware for Internal Services. Hold design team meeting with WP7 and WP5 to determine the middleware strategy and changes. Define a new central base schema to assist unification of all assembly and publication tools, based on current schemas in AC en DC; and produce a Global Species Database (GSD) builder tool (batch functions) and GSD editor tool (atom by atom editing) based on the new schema pre-fitted for uploading to the SPICE cache (SPICE is a distributed-computing engine that runs the Dynamic Checklist). (with Paris & VLIZ provides prototype, ETI develops production version). Prepare robust components for the SPICE 5.1 interoperability system without changing the fundamental communication protocol, enabling steps towards unification of the AC and DC production streams; produce control components for assembled data. Produce a universal GSD wrapper [kit] for data providers to link GSDs to the SPICE system (with CRIA). Alter the SPICE DC cache to adopt the new central base schema, Alter the GSD editing and inspection tools that currently work with the AC to work with the new base schema and make these available for editorial staff to edit and inspect the cache an

intermediate state datasets in the base schema. Produce a stable production tool for linking GSDs to the SPICE system using experiences from task 6.8 , including a new version of the Metadatabase software and a tool for AC export & conversion from the DC cache for publication on CD. Prepare a version of the SPICE 5.1 system for supporting linkages, cross-maps and enhanced interface for use in the Multi-Hub Network. Prepare documentation and version control system for all SPICE components. (with CARDIFF). Prepare a production version of the components in the new e-2 service-based system from WP7

**Dr Peter H. Schalk**, ETI Managing Director. Specialization: HR and project management, use-user analysis, data quality, information management, knowledge product development. Chairman of the Board of Directors of Species2000, Chairman of the Board of Directors of ETI Information Services Ltd, Chairman of the Netherlands Foundation for Nature Research, Vice-chair of the GBIF Budget Committee.

**Wouter Addink**, ICT project-leader, system developer.

**Dr Marc Brugman**, database programmer, project coordinator.

**Dr Maarten Schermer**, system developer.

**Dr Ruud Altenburg, Dr Edwin van Spronsen, Gideon Gijswijt** application programmers.

## 6. Cardiff University, Cardiff (CU)

The Cardiff School of Computer Science (COMSC) undertakes internationally ranked research in areas related to the focus of the present proposal through its Knowledge and Information Systems (KIS) research group. The Welsh e-Science Centre (WeSC) within COMSC complements and works with KIS. It is one of the UK's established e-Science centres, set up in 2001 as part of the national programme for e-Science/Grid/e-Infrastructure. The School brings relevant knowledge and experience in the areas indicated below through its participation in international and national funded projects. Funders include: EC, UK Research Councils, and JISC.

- ***e-Science for Biodiversity Informatics:*** Building an e-Science platform for biodiversity research, addressing relevant research problems in workflow system integration and usability, security, interoperability and integration of heterogeneous, distributed and legacy systems and databases, in projects such as BiodiversityWorld, GRAB, BioDA, ASMIMA and Ark 2010. TDWG sponsored SpiceTIP project researched and prototyped the deployment of LSIDs. Executive partner for technical planning in the EC funded project for preparing the LifeWatchinfrastructure.
- ***Biodiversity Informatics:*** Building distributed information systems, including the Catalogue of Life, addressing research problems in architectures for interoperability, quality control and standards for data and software and how to work with different quality levels, in projects such as ENBI, EuroCat, SPICE and LITCHI.
- ***e-Science:*** Partner operating node in the UK National Grid Service; Configuration and management of Grid middleware, build and operation of Service-oriented (SOA) Grid test-beds, in projects such as RTGrid, GECEM and Grid4Wales. International Grid and e-Science networking activities; links to Coregrid, Gridlab and EGEE. Addressing research problems in Virtual Organisations, Problem Solving Environments (Portals), use of varying quality data, workflows and mechanisms for provenance, in projects such as CONOISE, COVITE, GECEM; PROVENANCE, PASOA, Triana, WoSE, Gridlab and GridOneD.

### Tasks:

Coordination of WP7 “Novel e-Science Service-based Architecture”. To assemble the requirements and specification for the new Architecture and the new implementation of the ‘e-2’ system. To design the e-2 implementation with an open service-based architecture. To establish common communication protocols, data model, and management framework suitable for e-2. To implement e-2 test tools, and e-2 management tools. To create and install a test implementation for e-2. To respond to testing with an enhanced prototype for e-2

**Alex Hardisty**, Grid Centre Manager of the Welsh e-Science Centre and a technical planner for EC FP7 LifeWatch. Chair of the Steering Committee for the Community Engagement and Support strand of the JISC e-Infrastructure programme.

**Richard White**, Lecturer in Bioinformatics, the convenor of the Species 2000 Information Systems Group.

**Andrew Jones**, Senior Lecturer, a member the Species 2000 IS group.

## 7. Stichting National Natuurhistorisch Museum Naturalis, Leiden (NMNH)

Naturalis is a home organisation for Odonata database. Member of Species 2000. Participated in EC FP5 EuroCat project.

**Odonata: Catalogue of the Odonata of the World** by J van Tol. Taxonomic database of world dragonflies and damselflies: Animalia – Arthropoda – Insecta – Odonata; 5,321 spp.

**Tasks:** Odonata database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Prof. Jan van Tol**, leading taxonomist of dragonflies and author of Odonata database.

## 8. Species 2000 (Sp2000)

Species 2000 is a Network organisation that creates an index of the world's known organisms. It is a UK Registered Company limited by guarantee (Company No. 3479405). Species 2000 distributed model synthesises the index from sectors supplied by taxonomic databases across Europe and around the world, many from the major European institutions of CETAF. The programme reached production scale as an EC scientific infrastructure under the FP5 EuroCat project, and as a member of EDIT celebrated coverage of one million species in 2007. Its Catalogue of Life is a global service ([www.catalogueoflife.org](http://www.catalogueoflife.org)) recognised by the UN Convention on Biological Diversity, and presently comprising a synonymic species checklist of 1.1 million plants, animals, fungi and micro-organisms, about 2.5 million names, and a comprehensive taxonomic hierarchy. It provides a taxonomic backbone for global biodiversity portals, such as EoL and GBIF, for about 40 national portals worldwide, and is used by scientists in 79 countries. It contributes content and the taxonomic hierarchy used by the uBio taxonomically intelligent tool in the BHL programme.

**Tasks:** Participation in WP2, WP3, WP4 and WP5 activities.

**Prof. Dr. Frank Bisby**, Executive Director of Species 2000, works with what is now an extensive network of taxonomic database hubs around the world from China and New Zealand to Brazil and N. America, in projects to approach completion of known organism coverage in the Catalogue of Life, and to work with partner organisations on the Global Names Architecture.

**Dr Yuri Roskov**, Dr Yuri Roskov, Species 2000 Content Manager.

**Rebeca Mann**, Species 2000 Secretariat.

## 9. Vlaams Instituut voor de Zee, Ostend (VLIZ)

VLIZ is a home organisation for WoRMS database.

**WoRMS: World Register of Marine Species**, compile by the consortium of marine biologists. Taxonomic database of marine organisms of the world, 122,500 spp.

**Tasks:** WoRMS database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality. IT specialists of

VLIZ will take active part in software development in WP 6 (GSD builder and editor based on the new schema).

**Ward Appeltans**, WoRMS data manager.

**Bart Vanhoorne**, Aphia system manager and programmer.

## 10. Staatliches Museum für Naturkunde, Stuttgart (SMNS)

Museum is a home organisation for GloBIS database. Participated in EC FP5 EuroCat project.

**GloBIS/GART: Global Butterfly Information System** by C Häuser, J Holstein & A Steiner (eds). Taxonomic database of world swallowtail butterflies: Animalia – Arthropoda – Insecta – Lepidoptera – Papilionidae; 553 spp.

**Tasks:** GloBIS database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Christoph Häuser & A Steiner**, specialists in taxonomy of Lepidoptera and authors of GloBIS database.

## 11. Natural History Museum, London (NHM)

NHM is a home organisation for 5 databases: LepIndex, UCD, Tineidea NHM, BSF and PSF. Member of Species 2000. Participated in EC FP5 EuroCat project.

**BSF: Blattodea Species File Online** by GW Beccaloni. Taxonomic database of world cockroaches: Animalia – Arthropoda – Insecta – Blattodea; 4,430 spp.

**LepIndex: The Global Lepidoptera Names Index** by GW Beccaloni, MJ Scoble, GS Robinson & B Pitkin. Digitised version of the index card archive to the scientific names of the living and fossil butterflies and moths of the world at the Natural History Museum: Animalia – Arthropoda – Insecta – Lepidoptera; 250,203 spp.

**PSF: Phasmida Species File** by PD Brock. Taxonomic database of world stick and leaf insects: Animalia – Arthropoda – Insecta – Phasmida; 2,821 spp.

**Tineidae NHM: Global Taxonomic Database of Tineidae** by GS Robinson. Taxonomic database of world clothes-moths: Animalia – Arthropoda – Insecta – Lepidoptera – Tineoidea – Tineidae; 2,340 spp.

**UCD: Universal Chalcidoidea Database** by J Noyes & A Polaszek. Taxonomic database of world chalcidoid wasps: Animalia – Arthropoda – Insecta – Hymenoptera – superfamilies Chalcidoidea & Serphitoidea; 19,847 spp.

**Tasks:** LepIndex, UCD, Tineidea NHM, BSF and PSF databases will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Christopher Lyal**, Department of Entomology, leading taxonomist of weevils, co-author of WTaxa database.

**Dr George Beccaloni**, taxonomist of cockroaches, content manager of LepIndex and author of BSF database.

**Dr Andy Polaszek**, taxonomist of chalcidoid wasps and manager of UCD database.

## 12. CAB International (CABI)

CABI is a home organisation for Species Fungorum database. Member of Species 2000. Participated in EC FP5 EuroCat project.

**Species Fungorum** by PM Kirk. Taxonomic database of various fungal taxa: Fungi (diverse taxa), Chromista – Hyphochytriomycota, Labyrinthulomycota, Oomycota; Protozoa – Acrasiomycota, Dictyosteliomycota, Plasmodiophoromycota; 30,085 spp.

**Tasks:** Species Fungorum database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Paul Kirk**, editor of Species Fungorum database.

## 13. Royal Botanic Gardens Kew (RBGK)

RBG Kew is a home organisation for World Checklist of Selected Plant Families database. Member of Species 2000. Participated in EC FP5 EuroCat project.

**World Checklist of Selected Plant Families** by R Govaerts (ed). Taxonomic database of 94 plant families: Plantae – Ginkgophyta, Gnetophyta, Magnoliophyta – Liliopsida (except Iridaceae, Juncaceae, Poaceae & Potamogetonaceae) & Magnoliopsida – Araliaceae, Betulaceae, Euphorbiaceae, Fagaceae, Garryaceae, Magnoliaceae, Nothofagaceae, Pandaceae, Phyllanthaceae, Picrodendraceae, Putranjivaceae, Sapotaceae & Ticodendraceae; 83,298 spp.

**Tasks:** Kew database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Alan Paton**, Assistant Keeper of the Herbarium, official representative of RBG Kew in Species 2000.

**Dr Rafael Govaerts**, editor of Kew database.

**Dr Robert Allkin**, Information Projects Manager.

## 14. Universitaet Wien (UNI WIEN)

The University is one of the home organisation for AnnonBase.

**AnnonBase: World Species List of Annonaceae** by H Rainer & LW Chatrou (eds). Taxonomic database of world custard apples: Plantae – Magnoliophyta – Magnoliopsida – Magnoliales – Annonaceae; 2,235 spp.

**Tasks:** AnnonBase will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Heimo Rainer**, editor and IT manager of AnnonBase.

## 15. Universiteit Utrecht (UU)

The University is a home organisation for TicksBase. Participated in EC FP5 EuroCat project.

**TicksBase** by AM Nijhof, AA Guglielmono & IG Horak. Taxonomic database of world ticks: Animalia – Arthropoda – Arachnida – Parasitiformes (Acari – suborder Ixodida); 868 spp.

**Tasks:** TicksBase will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Frans Jongejan & Dr Ard Nijhof**, authors of TicksBase.

## **16. National University of Ireland, Galway (NUIG)**

The University is a home organisation for AlgaeBase. Member of Species 2000. Participated in EC FP5 EuroCat project.

**AlgaeBase** by M Guiry . Biodiversity database of world algae: Bacteria – Cyanobacteria; Chromista – Cryptophyta, Haptophyta, Sagenista; Protozoa – Acritarcha, Cercozoa, Dinophyta, Euglenozoa, Flagellata, Parabasalia; Plantae – Bacillariophyta, Chlorophyta, Cyanidiophyta, Glaucophyta, Prasinophyta, Rhodophyta; 30,487 spp.

**Tasks:** AlgaeBase will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Prof. Micahel Guiry**, taxonomists of seaweeds, author of AlgaeBase.

## **17. Leibniz-Institut Fuer Meereswissenschaften an der Universitaet Kiel (IMUK)**

IfM-GEOMAR is a home organisation for FishBase. Member of Species 2000. Participated in EC FP5 EuroCat project.

**FishBase** by R Froese & D Pauly (eds). Biodiversity database of world fishes: Animalia – Chordata – Actinopterygii, Cephalaspidomorphi, Elasmobranchii, Holocephali, Myxini, Sarcopterygii; 29,879 spp.

**Tasks:** FishBase will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Prof. Rainer Froese**, editor and leading manager of FishBase.

## **18. Deutsches Krebsforschungszentrum (FZK)**

FZK will provide administrative and technical facilities for Reptile Database.

**Reptile Database** by P Uetz. Taxonomic database of world reptiles: Animalia – Chordata – Reptilia; 8,624 spp.

**Tasks:** Reptile Database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Peter Uetz**, taxonomist of reptiles, author and custodian of Reptile Database. Member of Species 2000. Participated in EC FP5 EuroCat project.

## 19. Botanische Staatssammlung München (BSM)

BSM is a home organisation for LIAS database. Member of Species 2000. Participated in EC FP5 EuroCat project.

**LIAS: A Global Information System for Lichenized and Non-Lichenized Ascomycetes** by D Triebel & al. Biodiversity database of world lichens: Fungi – Ascomycota – Ascomycetes.

**Dr Dagmar Triebel**, editor and manager of LIAS database.

## 20. Muzeum i Instytut Zoologii PAN, Warsaw (MIZPAN)

The Institute is a home organisation for Salticidae Database. Participated in EC FP5 EuroCat project.

**Salticidae Database: Global Species Database of Salticidae (Araneae)** by Jerzy Proszynski. Taxonomic database of world jumping spiders: Animalia – Arthropoda – Arachnida – Araneae – Salticidae; 5,000 spp.

**Tasks:** Salticidae Database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Wieslaw Bogdanowicz**, Director of the Institute.

**Dr Jerzy Proszynski**, taxonomists of spiders and author of Salticidae Database.

## 21. Oberösterreichisches Landesmuseum, Biologiezentrum, Linz (OOE)

The Museum is a home organisation for ZOBODAT database. Member of Species 2000. Participated in EC FP5 EuroCat project.

**ZOBODAT: Zoological-Botanical Database (Vespoidea)** by J Gusenleitner. Taxonomic database of world vespid wasps: Animalia – Arthropoda – Insecta – Hymenoptera – superfamily Vespoidea; 5,925 spp.

**Tasks:** Salticidae Database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Gerhard Aubrecht**, zoology of vertebrates.

**Michael Malicky**, ZOBODAT database system manager.

## 22. National Museums and Galleries of Wales, Cardiff (NMGW)

The Museum is a home organisation for MolluscaFW database. Participated in EC FP5 EuroCat project.

**MolluscaFW: Checklist of Freshwater Mollusca** by M Seddon & al. Taxonomic database of world freshwater mollusks

**Tasks:** MolluscaFW database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot project to enhance GSD quality.

**Dr Mary Seddon**, MolluscaFW database editor.

**Ben Rowson**, MolluscaFW database manager.

### 23. Institut de Recherche pour le Développement, Paris (IRDF)

IRDF is a home organisation for TITAN database. Participated in EC FP5 EuroCat project.

**TITAN: Cerambycidae database** by G Tavakilian. Taxonomic database of world longhorn and timber beetles: Animalia – Arthropoda – Insecta – Coleoptera – Cerambycidae; 27,221 spp.

**Tasks:** TITAN database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Gerard Tavakilian**, leading taxonomist of large family of timber beetles, author of TITAN database.

### 24. Consejo Superior de Investigaciones Cientificas, Madrid (CSIC)

The Museo Nacional de Ciencias Naturales CSIC & Real Jardín Botánico de Madrid CSIC are home organisations for 3 databases. Member of Species 2000.

**WTaxa: Electronic Catalogue of Weevil Names** by M Alonso-Zarazaga & C Lyal. Taxonomic database of world weevils: Animalia – Arthropoda – Insecta – Coleoptera – superfamily Curculionoidea; 80,728 spp.

**Eumycetozoa.com: Nomenclatural Database of Eumycetozoa (Myxomycota)** by C Lado. Taxonomic database of world slime moulds: Protozoa – Mycetozoa; 1,074 spp.

**RJB Geranium: Geranium Taxonomic Information System** by C Aedo. Taxonomic database of world cranesbills: Plantae – Magnoliophyta – Magnoliopsida – Geraniales – Geraniaceae – genus *Geranium*; 409 spp.

**Tasks:** all 3 database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Marian Ramos**, Museo Nacional de Ciencias Naturales.

**Dr Miguel Alonso-Zarazaga**, taxonomist of weevils and co-author of WTaxa database.

**Dr Carlos Lado**, taxonomists of slime moulds and author of Eumycetozoa.com data portal.

**Dr Carlos Aedo**, taxonomists of flowering plants and author of RJB Geranium database/

### 25. TSJ BVBA

TSJ BVBA is a home organisation for Scarabs database.

**Scarabs: World Scarabaeidae Database** by P Schoolmeesters. Taxonomic database of world scarab beetles: Animalia – Arthropoda – Insecta – Coleoptera – superfamily Scarabaeoidea; 17,444 spp.

**Tasks:** Scarabs database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Paul Schoolmeesters**, author and custodian of Scarabs database. Member of Species 2000. Participated in EC FP5 EuroCat project.

## **26. Staatliche Naturwissenschaftliche Sammlungen Bayerns, München (SNSB, Munich)**

ZSM is a home organisation for SysMyr database.

**SysMyr: Systematic Myriapod Database** by J Spelda. Taxonomic database of world diplopods, pauropods and symphylans: Animalia – Arthropoda – Diplopoda, Pauropoda & Symphyla; 10,344 spp.

**Tasks:** SysMyr database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Joerg Spelda**, taxonomist, author of SysMyr database.

## **27. Universita Degli Studi Di Padova (UNIPD)**

UoP Department of Biology is a home organisation for ChiloBase.

**ChiloBase: A World Catalogue of Centipedes (Chilopoda) for the Web** by A Minelli (ed). Taxonomic database of world centipedes: Animalia – Arthropoda – Chilopoda; 3,139 spp.

**Tasks:** ChiloBase database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Prof. Alessandro Minelli**, leading taxonomist of centipedes and editor of ChiloBase.

## **28. Universiteit Van Amsterdam (UVA)**

The Museum is a home organisation for Porifera database.

**Porifera: World Porifera Database** by R van Soest (ed). Taxonomic database of world sponges: Animalia – Porifera; 8,052 spp.

**Tasks:** Porifera database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Rob van Soest**, taxonomist of sponges and editor of Porifera database.

## **29. Institut Royal Des Sciences Naturelles De Belgique, Brussels (RBINS)**

The Institute is a home organisation for Rotifera database.

**Rotifera database** by H Segers. Taxonomic database of world rotifers: Animalia – Rotifera; 1,983 spp.

**Tasks:** Rotifera database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Hendrik Segers**, taxonomist of rotifers and author of Rotifera database.

### **30. The Chancellor, Masters and Scholars of the University of Oxford (Oxford)**

Department of Zoology is a home organisation for WCS database.

**WCS: World Catalogue of Strepsiptera** by J Kathirithamby. Taxonomic database of world Strepsiptera (in progress): Animalia - Arthropoda - Insecta – Strepsiptera; 600 spp.

**Tasks:** Rotifera database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Jeyaraney Kathirithamby**, taxonomist of Strepsiptera and author of WCS database.

### **31. Wageningen Universiteit (WU)**

WU Department of Plant Sciences is a home organisation for Compositae Checklist database. Participated in EC FP5 EuroCat project.

**Global Compositae Checklist** by C Flann (ed). Taxonomic database of world Compositae (in progress): Plantae – Magnoliophyta – Magnoliopsida – Asterales – Asteraceae.

**Tasks:** Compositae Checklist database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Christina Flann**, plant taxonomist and editor of Global Compositae Checklist.

### **32. Royal Botanic Garden, Edinburgh (RBGE)**

RBGE is a home organisation for World Umbellifer Database.

**World Umbellifer Database** by M Watson (ed). Taxonomic database of world umbellifers (in progress): Plantae – Magnoliophyta – Magnoliopsida – Apiales – Apiaceae.

**Tasks:** World Umbellifer Database will participate in WP3 GSD Network and respond to GSD assessment, network consultations and workshops, acceleration of completeness and data quality, pilot projects to enhance GSD quality.

**Dr Mary Gibby**, plant taxonomist.

**Dr Mark Watson**, plant taxonomist and editor of World Umbellifer Database.

### **Multi-Hub Network:**

### **33. Chinese Academy of Science, Biodiversity Committee, Beijing, China (CAS)**

Institute of Botany and Institute of Zoology CAS are non-for-profit organisations, which belongs to and powered by the Chinese Academy of Science. It is national leading institution for research and education in biodiversity. Since October 2006 IBCAS become a home organisation for “Species 2000 China Node”, a regional

Catalogue of Life hub in China. In present, two key organisations contribute to the Node activity: Institute of Botany and Institute of Zoology CAS. The Biodiversity Committee of Chinese Academy of Sciences supports and supervises the establishment of the Node. The Node issued two editions of the Catalogue of Life China database (2007, 2008).

**Tasks:** CAS will represent Species 2000 China Node and collaborate with WP 3 “Foresight study for geographical extension: the Global Multi-Hub Network”. It will take significant role in Multi-Hub Network discussion, design and prototype project, which will link Catalogue of Life China to the Second Architecture of the Catalogue of Life.

**Prof. Keping Ma**, Director of the Institute of Botany, Co-ordinator of Species 2000 China Node.

**Dr Haining Qin**, plant taxonomist, content manager of China checklist of higher plants.

**Prof. Liqiang Ji**; co-ordinator of Chinese Animal Information System.

**Huijie Qiao**, system manager of Chinese Animal Information System.

**Xuehong Xu**, Secretary for Species 2000 China Node.

#### **34. The Centro de Referência em Informação Ambiental, Campinas, Brazil (CRIA)**

CRIA (Reference Center on Environmental Information) is a not-for-profit, non-government organization. Its aim is to contribute towards a more sustainable use of Brazilian biodiversity through the dissemination of high quality scientific information to support policy and decision making and to promote education for sustainable development. CRIA’s activities are focused on the development of information systems for the dynamic integration and visualization of biodiversity data from distributed information resources.

**Tasks:** CRIA will collaborate with WP 3 “Foresight study for geographical extension: the Global Multi-Hub Network”. It will take significant role in Multi-Hub Network discussion, design and prototype project, which will link Brazilian data to the Second Architecture of the Catalogue of Life.

**Dr Vanderlei Perez Canhos**, President Director of CRIA, member of the Species 2000 Board of Directors.

**Sidnei de Souza**, Director of Information Systems at CRIA.

#### **35. Integrated Taxonomic Information System at Smithsonian Institution, Washington, USA (ITIS)**

Smithsonian Institution is an educational and research institute and associated museum complex, administered and funded by the US government . Smithsonian Institution is a home for the Integrated Taxonomic Information System (ITIS). This is a partnership designed to provide consistent and reliable information on the taxonomy of biological species. It is an international body, with Canadian and Mexican

government agencies participating. The primary focus of ITIS is North American species, but many groups are worldwide. ITIS is a partner of Species 2000 with goal to create the Catalogue of Life.

**Tasks:** ITIS will collaborate with WP 3 “Foresight study for geographical extension: the Global Multi-Hub Network”. It will take significant role in Multi-Hub Network discussion, design and prototype project, which will link ITIS North America to the Second Architecture of the Catalogue of Life.

**Dr Thomas Orrell**, Acting Director of ITIS.

**David Nicolson**, ITIS Data Development Coordinator.

### **36. Commonwealth Scientific and Industrial Research Organisation, Canberra, Australia (CSRO)**

The Australian Biological Resources Study is part of the Parks Australia Division of the Commonwealth Department of the Environment, Water, Heritage and the Arts. ABRS was established to coordinate research in taxonomy and document the flora and fauna of Australia. ABRS is Australia’s national focal point for taxonomy. The Atlas of Living Australia (ALA) is a partnership of Australian scientific institutions and organisations, funded under the Australian Government. The ALA is developing a data management system to catalogue and organise information relevant to the study of Australia’s biodiversity. The project will bring together information from a wide variety of sources, including many of the country’s most significant natural history collections and herbaria, ecological and observational data sets, images, online literature, diagnostic tools and molecular data.

**Tasks:** ALA will collaborate with WP 3 “Foresight study for geographical extension: the Global Multi-Hub Network”. It will take significant role in Multi-Hub Network discussion, design and prototype project, which will link Atlas of Living Australia to the Second Architecture of the Catalogue of Life.

**Cameron Slatyer**, (Director, ABRS) with responsibility for the direction and function of ABRS and the ABRS National Taxonomy Research Grant Program.

**Donald Hobern**, ALA Director, coordinating and planning the development of the ALA as a set of coordinated data services.

**Pam Beesley** and **Alice Wells** manage the *Australian Faunal Directory* database.

### **37. Landcare Research – Manaaki Whenua, New Zealand (Landcare Research)**

Landcare Research is one of nine independent Crown Research Institutes founded in 1992 from a reorganisation of Government funded research in New Zealand. New Zealand’s leading provider of solutions and advice for sustainable development and the management of land-based natural resources. Research focuses on three key outcome areas: Sustaining & restoring biodiversity, Sustaining land environments, Sustainable business & living. About 380 staff, in nine locations: at Lincoln, Palmerston North, Hamilton, Auckland, Gisborne, Nelson, Dunedin and Alexandra. It

is home for Species 2000 Asia-Oceania initiative which aims to create database of New Zealand and adjacent marine areas biota.

**Tasks:** Landcare Research will collaborate with WP 3 “Foresight study for geographical extension: the Global Multi-Hub Network”. It will take significant role in Multi-Hub Network discussion, design and prototype project, which will link database of New Zealand biota to the Second Architecture of the Catalogue of Life.

**Jerry Cooper**, member of Landcare Research staff in Lincoln, system manager and coordinator of NZ biodiversity database.

## **2.3 Consortium as a whole**

A proportion of the partners in the Consortium have worked together before, primarily in the FP5 EuroCat project that established the Species 2000 Catalogue of Life as a European scientific infrastructure, but also in other related activities such as EC ENBI and the EC EDIT NoE. This is important because the large GSD Network is already constituted and its members play a vital continuing role in providing the content that is refreshed and integrated to make the Catalogue.

The Consortium is well-balanced in that it builds where needed on parts of the existing programme, and brings in new partners with particular skills related to the new developments. Two of the new partners have joined the Consortium to lead its new developments. BGCI joins because of its skills in networking and outreach in the wider biodiversity community, and Narodny Museum Prague joins because of its experience with long-distance international partnerships.

Each of the GSD Network Partners is self-selecting on the basis of its prior taxonomic expertise and long-term database coverage – all that we know of in Europe have been included. In several cases the partner hosts more than one GSD, and five of the main partners also have GSDs at their institutions.

Partner 8, the Species 2000 legal body, brings a rather different role to the Consortium. It was established by the Species 2000 Catalogue of Life Programme to provide both a holding company that could hold IPR and Access Agreements on its behalf, and as a residual body that could persist between funded projects. It is registered in England as a not-for-profit scientific organisation (technically a company limited by guarantee), and is owned and governed by the participating organisations and individuals in the Species 2000 Catalogue of Life Programme. In 4D4Life it provides access to the prior achievements of the programme; it is the vehicle for any legal or financial sustainability; and it will again provide the residual body carrying the programme forward when the project is complete. Consortium partners in 4D4Life that are not already members will be invited to become members, so that they too have an interest in the residual body.

## **2.4 Resources to be committed**

4D4Life requests approx. 3.4 M € contribution from the Commission for expenditure as follows:

(NB All partners, with exception of ITIS N.America, CSIRO/ALA Australia, and NZOR New Zealand receive minimum travel funds to attend all four Project Meetings)

### *WP 1 Management of the Consortium*

University of Reading requests one part-time clerical assistant to deal with the large consortium, including the GDS Network, and the associated financial administration, reporting and audit, plus 5% of Project Coordinator.

### *WP 2 Engagement with scientific users and partners*

Botanic Gardens Conservation International requests one full time equivalent position to organise the networking events, organise the e-bulletin and communications strategy and platform, and both to elicit from user groups the prioritised proposals for new technical and new educational and popular services, and to get the new services trialled and tested when they become available..

Travel, workshop, and promotional materials costs are requested.

Additional travel for Project Liaison Group handled through this WP.

#### *WP 3 Strengthening the GSD Network: a Virtual Community*

MNHN Paris requests one full time position to lead and organise networking and funded pilot projects with all the custodians of databases in the GSD Network.

Funds allocated to members of the GSD network are a standardised allocation of 16,000€ for up to 4 months work on the pilot projects, and 2800€ travel to project meetings.

Travel, minor equipment and minor consumables are requested.

NMNH Naturalis requests 3 PM for additional task on Nomenclator services to GSDs, in addition to its standard GSD allocation.

#### *WP 4 Foresight Study for geographical extension: Global Multi-Hub Network*

NM Prague requests one full time position to lead and organise networking with all partners in the Multi-Hub Network, plus pilot projects by CRIA Brazil and CAS China, plus create, maintain, and publish significant concept, design and prototype test documents.

Travel and subsistence for attendance at meetings in Europe for CRIA Brazil and CAS China. (ITIS N. America, CSIRO/ALA Australia and NZOR self-funded)

#### *WP 5 Scientific Services of the Catalogue of Life*

UREAD requests three full time staff plus 5% of the Project Coordinator:

1) Scientific Coordinator for the entire project, also leading on specifying, creating and testing each of the new or enhanced services, working with WP2, WP3 and WP6, and the production & systems team at Reading.

2) Taxonomic Editorial Assistant to lead on the continued and developing production cycle of the Annual Checklist and Dynamic Checklist production of an integrated Catalogue of Life.

3) Systems Manager to install, upgrade and maintain the internal and external service systems 24/7, to manage caching and refreshing processes, to manage the Annual Checklist test cycles for suppliers to view, and to lead on installation, development, and management of wrapper programmes across the entire project, as well as project web-sites and web-services.

Travel requested for Project Coordinator & staff to Project Meetings, staff visits to partners, Coordinator visits to partners, staff and coordinator presentations to international meetings, including outside the EU.

Equipment, materials and legal requested for Servers & rack, server services, promotion materials, legal fees for Access Agreements and sustainability issues, and consumables.

Sp2000 requests full time time assistant for Annual Checklist and CD-ROM compilation and production cycle, at low rate in Sp2000 Philippines Office.

Staff travel to Project Meetings and Sp2000 secretariat in Reading.

*WP 6 Software Support Services*

ETI requests 2 full-time equivalent systems and programmer staff for the multiple programming and design tasks in WP6.

Test server, and travel to Project Meetings and to partners requested.

VLIZ requests programmer time at 4 months.

CRIA Brazil requests programmer time at 6 months.

MNHN Paris requests programmer time at 3 months.

*WP 7 Novel e-Science Service-based Architecture*

CU requests 2 full time staff, 24 months of system designer, 18 months of programmer, plus 10% of e-science designer/manager.

Test server and minor consumables requested.

## Section 3: Impact

### 3.1 Expected impacts listed in the work programme

4D4Life will increase the scale and stratification of federation within the Catalogue of Life community, and promote yet further interoperation of digital repositories, including across the world. The project has a strategy for enhancing the content and curation of the contributing repositories, and this, tied with middleware enhancements will increase the robustness of the data infrastructure. Also, by adopting a common management strategy to unify the presently separate integration of our two products (the Annual Checklist and the Dynamic Checklist) we anticipate reduced costs and more robust outcomes.

4D4Life does require a European level approach, that is it is neither national/local, nor yet truly global in activity and focus of expertise. The size of the task, as well as the technical architecture needed to complete it, does require major large scale cooperation between taxonomists and taxonomic institutions on a continental and global scale. However it is also a fact that even at the global and continental scale, the main concentration of taxonomists, expertise and activities is in Europe. This has been established in the EDIT NoE for the profession as a whole, and those statistics are echoed by the distribution of GSD activity seen in the Species 2000 Catalogue of Life programme. 38 of the 52 GSDs currently supplying data are european based, compared with 11 in the USA, and 3 from elsewhere.

<b>Leveraged value of existing databases connected</b>			
Lower bound mean (extremes excluded) £ 78 per species			
Upper bound mean (including extremes) £ 131 per species			
	Databases	Lower bound	Upper bound
USA	11	£ 29,600k	£ 49,680k
Europe	38	£ 58,870k	£184,084k
Japan	1	£ 520k	£ 860k
Russia	1	£ 160k	£ 270k
Australia	1	£ 74k	£ 124k
<b>TOTAL</b>	<b>52</b>	<b>£ 86,224k</b>	<b>£144,704k</b>
Invested cost of GSDs leveraged by central expenditure in GBP, Y. Roskov, Secretariat			

An important issue within the landscape of international scale biodiversity and taxonomic programmes is a perception possibly raised by the exceptional publicity attending the start-up of GBIF, and the recent launch of EoL. Both of those programmes are deeply involved in the delivery of taxonomic information, and in the

use of taxonomic structures to link and index other information. But it is now absolutely clear that neither GBIF nor EoL is attempting the same core task as we are carrying out in 4D4Life and the wider Species 2000 Catalogue of Life programme. In recent months both those organisations have confirmed, in public and in private, that they continue to use the Catalogue of Life as taxonomic backbone for their activities. Enthusiastic letters of support are attached for both organisations – indeed two from EoL.

4D4Life is projected to make significant impact in five areas:

- i) Opening up a range of electronic e-taxonomy services as infrastructure that can be built into the main international biodiversity information portals: GBIF, EoL, CBOL, iBOL, EBI/GenBank, IUCN.
- ii) Opening up the same range of electronic e-taxonomy services as infrastructure that can become part of the seamless architecture of biodiversity informatics, and thus available to be built into the much wider range national local and NGO portals in Europe and around the world, and into novel experimental systems that are the subject of a growing research and development community.
- iii) Opening up the same range, plus additional specially tailored e-taxonomy services to certain scientific communities, such as the biodiversity and climate change modelling community, the oceanographic community, or the genomics community.
- iv) Launching a completely new range of taxonomic products based on the Catalogue of Life, and to be tailored and downloaded onto laptops and hand-held devices. These will provide educational and popular products that will impact students, fieldworkers, and amateurs.
- v) and finally, simply by enhancing our present services we anticipate some additional impact on the wider user community.

### *3.1.1 International Biodiversity Information Portals*

The new range of electronic e-taxonomy services will have an immediate and beneficial effect on our technical relations with the major international biodiversity portals that use our Taxonomic Hierarchy and Species Checklist as their taxonomic backbone. These include EoL, GBIF, uBio, IUCN, iBOL, CBOL. Indeed, an eagerness for this project to move to enhanced electronic services can be clearly seen in their letters of support appended to this Part B document.

The services outlined in the description of Work Package 2 in section 1.4 Networking Activities fall into two classes: i) those that will be used for uploading to the users portal once only the entire or incremental content of the Hierarchy or Checklist by those who hold a full copy; and ii) those who plan to use our service repeatedly in real time, and not to hold a full copy. In the first class are several services to notify users of incremental changes to the Catalogue including species name changes, species circumscription changes, and higher taxon classification changes. These will be based on the taxon Globally Unique Identifiers (GUIDS) recently introduced using the Life Science Identifiers (LSID) system. These will be of interest especially to GBIF, EoL and uBio. They will allow GBIF, and uBio to track taxonomic concepts between

annual editions, and to link external data with high accuracy. In the second class are “Find this name”, and “Synonymy server” services, and these will be of interest to iBOL, EoL and IUCN.

### *3.1.2 Services to the general Biodiversity Informatics and Biodiversity Portals community*

In the longer term, the same types of electronic service as already mentioned in 3.1.1 are likely to have a more significant impact on the biodiversity community as they become available as standard components of the emerging seamless architecture for biodiversity informatics. This will mean that an even wider range of national, local and NGO biodiversity portals will incorporate our Catalogue or make use of our services in real time. It also means that our Catalogue and services become available to the growing community of innovative researchers in biodiversity informatics. This relatively new community is starting to realise the exciting potential and to experiment with novel ideas, structures and connections. Many of these may soon lead to completely new products and portals in what is a comparatively open field for future development. 4D4Life expects to be part of these new developments.

### *3.1.3 E-taxonomy electronic services to particular scientific communities.*

The Catalogue of Life is already used in a number of scientific communities that have particular and different requirements. Examples include:

#### i) Climate Change Studies.

Using the catalogue as an index for harvesting specimen and observation records – for instance over the Internet or from GBIF or SpeciesLink. Such harvested datasets are frequently used in creating range maps to show distribution, or for use in species envelope modelling to predict distribution changes in climate change studies. To be accurate, and to yield the maximum harvest of records, the process needs to use both full synonymic expansion (to locate records of a species under all its names), but countered with alerts or blanking for ambiguities caused by homonyms, misapplied names and pro-parte synonyms (cases where the name alone cannot give accurate searches).

#### ii) Collections data cleansing

Data cleansing processes for herbarium, botanic garden and museum specimen record databases is another area where special issues are involved. The tools being developed are particularly good at alerting the users to errors caused either by small spelling variants, by regular spelling flips related to latin (phytos vs phyllos etc), and by confusions with phonetically close words. But at present we do not provide any facility to sort the species checklist and synonyms by these error-associations.

#### iii) Accessing DNA Barcode and Nucleotide data

The Programme has been approached by EBI and GenBank to discuss the idea of maintaining a cross-map between the Catalogue and the NCBI taxonomy used by the nucleotide databases and GenBank. Although the NCBI taxonomy is used in the established database, there is a need to provide indexing using the catalogue of life, because of its increasing use with DNA BarCode records being placed in Genbank by the CBOL and iBOL initiatives that cite the Catalogue of Life as a gold standard for annotation.

These examples illustrate the way in which the Catalogue of Life will provide special infrastructure services to particular disciplines, and how user communities wish to build our services into their own systems for daily use. These are examples of true infrastructure applications that we believe will play an increasing part in the impact of our public services – to professional communities rather than directly to the public.

#### *3.1.4 Educational and popular products*

This is a new area for the Catalogue of Life. A new generation of field biologists, students on field courses, expert amateurs and others expect to make heavy use of portable and hand-held electronic devices in the field and at home, no longer thinking of books as the principle medium of expert knowledge delivery.

Our new products may include components of the hierarchy and species list, attractive presentation with icons or illustrations of iconic species, and be both tailorable by taxonomic group and geographical area, and downloadable to laptops or hand held devices. Students attending fieldcourses, field study centres, field survey workers, amateurs travelling abroad, trainee field staff may all be interested in better tools to overcome the “taxonomic impediment”, in a world where basic taxonomic knowledge is no longer taught to every biology undergraduate.

#### *3.1.5 Simple improvements to our existing products*

The Species 2000 Catalogue of Life programme has developed up to now as a primarily academic programme in taxonomy to complete as much as possible, responding to the issue of reaching critical mass, the need to establish the resource in the taxonomic community, and the need to convince critics that it can be created and can be maintained. The result is that development of a range of services has not been a high priority. It is clear from what users tell us that we can make a larger impact, and increase the regular public usage simply *by enhancing our present services, interfaces, web-services and web-sites*. The focus on services that this project will provide, will enable us to add to our impact by this relatively simple change of emphasis, in addition to the novelties discussed above.

### **3.2 Dissemination and/or exploitation of project results, and management of intellectual property**

#### *3.2.1 Dissemination*

4D4Life will become the dissemination arm of the wider Species 2000 Programme. The programme is already tuned to the concept of developing as a scientific infrastructure, and has a clear and unique role in the landscape of biodiversity platforms and services. What it has lacked up to now is a project outside academic taxonomy to develop its service role. 4D4Life provides the opportunity to realise and give momentum to the new services that we know are wanted by the user communities.

The present services are branded as “Species 2000 & ITIS Catalogue of Life”, and are made available in four editions:

Annual Checklist 2008 (and preceding years to 2000) on the web:  
<http://www.catalogueoflife.org/annual-checklist/2008/search.php>

Annual Checklist 2008 (and preceding years back to 2000) on CD-ROM (ISSN 1473-009X).

Annual Checklist 2008 (and preceding years to 2005) web-service:  
<http://webservice.catalogueoflife.org/annual-checklist>

Dynamic Checklist on the web, launched October 2005:  
<http://www.catalogueoflife.org/dynamic-checklist/search.php>

Dynamic Checklist web-service  
<http://webservice.catalogueoflife.org/dynamic-checklist/CASWebService>

LSID resolver service (requires LSID plug-in for users browser)

The present public services are operated from servers at University of Reading, accessed via the web-sites of:

Species 2000 & ITIS Catalogue of Life - [www.catalogueoflife.org](http://www.catalogueoflife.org)  
Species 2000 – [www.sp2000.org](http://www.sp2000.org)  
ITIS – [www.itis.gov](http://www.itis.gov)

The CD-ROM is distributed at conferences and mailed out, or it can be either ordered or downloaded as a CD-ROM image from the Species 2000 web-site

The Annual and Dynamic Checklists attract more than 40 million hits per year, with usage by serious unique visitors rising in 2007 from 20,000 per month to 40,000 per month. Other much larger usage is achieved when GBIF and EoL users make use of the catalogue or traverse between species on the GBIF and EoL portals.

The 3,500 CD-ROMS are distributed each year, including 600 to GTI and BioNET International focal points in developing countries, and including mail requests from 79 countries.

In addition to the dissemination of the Catalogue products, 4D4Life represents a scientifically interesting development that includes novelties, and it will be the policy of the project both to encourage partners to publish research articles about the project, but also to encourage all partners to present the developments of the project or its outcomes at relevant scientific and professional organisational events.

### *3.2.2 Management of Intellectual Property*

It is proposed that 4D4Life will adopt the existing IPR framework set in place by the EC EuroCat project in FP5, and work in liaison with Species 2000 if it appears that changes are needed.

The key components in the existing framework are these:

i) 4D4Life and previous components of the wider Species 2000 programme operate an integrated service based on an integrated resource composed with contributions harvested from supplier databases. Standard Access Agreements are made with the supplier databases, and held by Species 2000 the residual legal body. These Access

Agreements make clear that ownership, copyright and responsibility for taxonomic enhancement and maintenance of each database remain with the original custodian. They also place a responsibility on Species 2000 to provide the three level credits in the published catalogue, and to make maintaining visibility of those credits a condition of third party licenses.

The entire Catalogue of Life is copyrighted as a taxonomic work by Species 2000 on behalf of its suppliers, and on behalf of ITIS, the North American partner. This enables Species 2000 and ITIS:

- i) to enforce that third party licencees maintain the three level credit system that is important both to supplier databases, and to individual taxonomists named in the “latest taxonomic scrutiny” field.
- ii) and to attempt to maintain control of the Catalogue as a single product in the community, that is without ‘speciation’ of variant forms.

The joint policy of the Species 2000 programme and of ITIS has been that the Catalogue is available to all individuals free of charge at point of use.

During 2007/2008 Species 2000 and ITIS have embarked on a trial license to a commercial company, where the service offered by the commercial company is quite outside the area of our own services. The model employed establishes an annual financial contribution, and includes the possible return of contributions to running costs from Species 2000 to the supplier databases. A letter of support from Taylor & Francis Ltd. for this 4D4Life proposal is appended to this Form B.

## Section 4; Ethical Issues

### 4.1 Data protection issues

#### 4.1.1 Metadata

4D4Life will use and further develop the Species 2000 Metadatabase that contains, or will contain, technical metadata concerning the structure, management and content of each participant database. This metadatabase contains, or will contain, metadata from participants in all previous and current components of the Species 2000 programme, in addition to those joining the programme through this 4D4Life Project.

At the start of the 4D4Life Project the Directors of Species 2000 (partner 8 in the Consortium) will give permission for the 4D4Life Project to access and work on the metadata as an integral part of the wider Species 2000 Programme. At the end of the 4D4Life Project, the custodianship of the metadatabase will remain with Species 2000, the residual legal body for this and other components of the wider Species 2000 Programme.

The Metadatabase contains information supplied by the custodians of participating databases. The initial data for each database is taken from a questionnaire response provided by the custodian, and at the end of the questionnaire the custodian is required to give their permission for this data to be used and published by Species 2000. For databases actively contributing to the Catalogue, the data is refreshed each year, and some components of this are explicitly intended for publication or use in the credit fields of the Catalogue.

Public display and publication of some parts of the metadata (including logos and the names of experts marked in the “latest taxonomic scrutiny” field) is also regulated by the Access Agreement as a compulsory component of material published by Species 2000, and a compulsory component of datasets licensed to third parties. The Access Agreement is made between the database custodian and the Species 2000 legal body.

Partners in 4D4Life and their employees working on 4D4Life will be required to observe the confidentiality of certain items that relate to metadata, such as for instance the content of peer reviews. The content of peer reviews is confidential to the wider Species 2000 Programme, and shared only with the custodian of the database and staff and appointed officers of the wider Species 2000 Programme.

#### 4.1.2 Data protection – personal data

Personal data, such as names, addresses, and contact details of database custodians and technical staff are held in the metadatabase and at the Species 2000 Secretariat, and are subject to the UK Data Protection Act. These are and will be kept to the minimum needed for the adequate functioning of the 4D4Life and wider Species 2000 Programme.

## 4.2 Issues relevant to ICT

As already listed above ethical and legal issues relating to intellectual property, and relating to data protection have been adequately addressed within the Species 2000 Programme, and will be adopted in this 4D4Life project. These arrangements are compliant with both the Charter of Fundamental Rights of the European Union and the opinions of the European Group on Ethics in Science and New Technologies (EGE).

The 4D4Life Project does not make use of:

- i) Data gathered on human volunteers
- ii) Animals used in ICT Research
- iii) ICT implants or wearable computing
- iv) Personal data on health or genetics
- v) Bio/Nano-electronics

## 4.3 Other issues

Within 4D4Life there are no issues relating to:

- i) Informed Consent
- ii) Use of animals
- iii) Human embryonic stem cells

Ethical Issues table

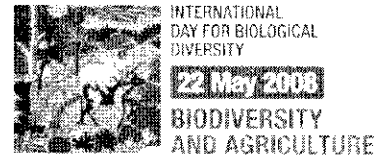
	YES	PAGE
<b>Informed Consent</b>		
• Does the proposal involve Children?		
• Does the proposal involve patients or persons not able to give consent?		
• Does the proposal involve adult healthy volunteers?		
• Does the proposal involve Human Genetic Material?		
• Does the proposal involve Human biological samples?		
• Does the proposal involve Human data collection?		
<b>Research on Human embryo/foetus</b>		
• Does the proposal involve Human Embryos?		
• Does the proposal involve Human Foetal Tissue / Cells?		
• Does the proposal involve Human Embryonic Stem Cells?		
<b>Privacy</b>		
• Does the proposal involve processing of genetic information or personal data (eg. health, sexual lifestyle, ethnicity, political opinion, religious or philosophical conviction)		
• Does the proposal involve tracking the location or observation of people?		
<b>Research on Animals</b>		
• Does the proposal involve research on animals?		
• Are those animals transgenic small laboratory		

animals?		
• Are those animals transgenic farm animals?		
• Are those animals cloned farm animals?		
• Are those animals non-human primates?		
<b>Research Involving Developing Countries</b>		
• Use of local resources (genetic, animal, plant etc)		
• Impact on local community		
<b>Dual Use</b>		
• Research having direct military application		
• Research having the potential for terrorist abuse		
<b>ICT Implants</b>		
• Does the proposal involve clinical trials of ICT implants?		
<b>I CONFIRM THAT NONE OF THE ABOVE ISSUES APPLY TO MY PROPOSAL</b>	<b>YES</b>	

End of proposal



# Secretariat of the Convention on Biological Diversity



Ref.: SCBD/STTM/RH/JS/aw/64756

9 September 2008

Subject: **Letter of Support for the Distributed Dynamic Diversity Databases for Life (4D4Life) project of the Catalogue of Life**


Dear Professor Bisby,

The Catalogue of Life has been instrumental in advancing scientific and taxonomic progress, essential for the understanding of life on Earth. The Conference of the Parties to the United Nations Convention on Biological Diversity (CBD) has acknowledged in several decisions that the Catalogue of Life is a major contributor to the improved management of biodiversity information, and is especially significant for the Global Taxonomy Initiative, which seeks to address the obstacles hindering faster progress in filling knowledge gaps in taxonomy and in improved management of biodiversity data. The Distributed Dynamic Diversity Databases for Life (4D4Life) on the enhanced architecture may further improve the accessibility of biodiversity data and facilitates the use of biodiversity data for global-scale analyses in conservation, climate change, invasive species, molecular biodiversity and regulatory domains. The service component of the 4D4Life could also make a contribution to achieving and assessing progress towards the target of significant reduction the rate of loss of biodiversity by 2010.

The Catalogue of Life has made significant efforts towards the expansion of biodiversity knowledge, infrastructure and electronic services. The CBD fully supports the efforts of the Catalogue of Life to further develop its scientific data infrastructure through its 'Distributed Dynamic Diversity Databases for Life' (4D4Life) project.

We believe that the future progress of the Catalogue of Life is of great value to the scientific community and we wish you great success in the 4D4Life project.

Yours sincerely,



Ahmed Djoghlat  
Executive Secretary

Prof. F.A. Bisby  
Species 2000 Secretariat  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK  
E-mail: [f.a.bisby@reading.ac.uk](mailto:f.a.bisby@reading.ac.uk)



United Nations  
Environment Programme

413 Saint Jacques Street, Suite 800  
Montreal, QC H2Y 1N9, Canada

Tel: +1 514 288 2220  
Fax: +1 514 288 6588

<http://www.cbd.int>  
[secretariat@cbd.int](mailto:secretariat@cbd.int)



David E. Schindel, Executive Secretary  
Consortium for the Barcode of Life  
Smithsonian Institution  
National Museum of Natural History  
P.O. Box 37012, MRC-105  
Washington, DC 20013-7012

8 September 2008

Prof. F.A. Bisby  
Species 2000 Secretariat  
Plant Science Laboratories  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

Dear Frank,

It is a pleasure to write in support of your proposal to the European Commission for support of the Catalogue of Life, and particularly your plan to develop and implement network services. The Consortium for the Barcode of Life (CBOL) represents one sector of the biodiversity informatics field, and like the other sectors, we rely on the Catalogue of Life as a critical resource. One of CBOL's principal goals is to create interoperability among specimen data, gene sequence data repositories such as EMBL, and taxonomic authorities such as the Catalogue of Life. Your proposed plan to mobilize Catalogue of Life data through new web services would help to overcome one of the most important obstacles to this high level of integration and interoperability.

In 2005, CBOL proposed the BARCODE Data Standard to GenBank which then proposed it to EMBL and the DNA Data Bank of Japan, the other members of the International Nucleotide Sequence Database Collaboration. The BARCODE standard was the first to create hyperlinks to connect DNA sequences with voucher specimens and taxonomic names. The Catalogue of Life provides the most reliable index of reviewed species names, and for this reason we have urged GenBank to use it in validating which GenBank records are awarded the keyword BARCODE. Several barcoding projects, such as the global Fish Barcode of Life Campaign (FISH-BOL) use the Catalogue of Life species list as the authority file for preparing their data records for submission to GenBank. This and other projects rely on the Catalogue of Life as an independent check on the validity of species names, so the plan to develop and implement web services would have immediate and widespread impact.

On behalf of CBOL's Executive Committee, thank you for the support that the Catalogue of Life provides to the DNA Barcoding Initiative. Our joint work is improving taxonomic research and its impact on society. We wish you every success in your application to the European Commission.

Sincerely yours,

David E. Schindel, Executive Secretary  
Consortium for the Barcode of Life



# CODATA

**CODATA Secretariat**  
5 Auguste Vacquerie, 75016 Paris, France  
Telephone: +33 1 45250496 - Fax: +33 1 42881466  
Email: [codata@dial.oleane.com](mailto:codata@dial.oleane.com)  
World Wide Web: <http://www.codata.org/>

---

*Committee on Data for Science and Technology of the International Council for Science*

Prof. F.A. Bisby  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

9 September 2008

**Re: Distributed Dynamic Diversity Databases for Life' (4D4Life).**

Dear Professor Bisby,

CODATA, the Committee on data for Science and Technology, is very pleased to lend its support to the Species 2000 consortium to the European Commission to develop its Scientific Data Infrastructure in a proposal titled 'Distributed Dynamic Diversity Databases for Life' (4D4Life).

CODATA is a international non governmental organization based in Paris in France. Our mission is to strengthen international science for the benefit of society by promoting improved scientific and technical data management and use. It is a resource that provides scientists and engineers with access to international data activities for increased awareness, direct cooperation and new knowledge. CODATA's members are comprised of 24 National Members countries, 15 International Scientific Union Members and 4 Co-opted Members. For more information please see <http://www.codata.org>

CODATA recognizes Species 2000 & ITIS Catalogue of Life, as an important component in the international biodiversity knowledge infrastructure. It will launch new electronic services, undertake a series of enhancements to its already established services, and move to a second generation federated architecture. This in turn benefits the wider scientific community. Indeed CODATA, as you are aware, Professor Bisby, was one of the original sponsors of SPECIES 2000 when it was launched and we are delighted to see how much it has developed and grown from strength to strength over the years

We are very pleased to support the application.

If you require any additional information please do not hesitate to contact me.

Kind regards,

Kathleen Cass  
Executive Director  
CODATA

CC: Krishan Lal, India, President CODATA; Shuichi Iwata, Japan, Past-President CODATA; Robert Chen, USA, Secretary General CODATA; Steve Rossouw, South Africa, Vice-President, CODATA; Gordon Wood, Canada, Vice-President CODATA; Jean Jacques Royer, France, Treasurer CODATA

---

Executive Committee:  
President: Krishan Lal, India  
Vice-Presidents: Steve Rossouw, South Africa, and Gordon Wood, Canada  
Secretary-General: Robert Chen, U.S.A.  
Treasurer: Jean-Jacques Royer, France  
Past-President: Shuichi Iwata, Japan

Elected Members:  
Jean Garnier, IUPAB - Sara Graves, U.S.A. - GUO Huadong, Chinese Academy of Sciences - Fedor Kuznetsov, Russia - Ray Norris, IAU - Antoni Nowakowski, Poland - Michel Sabourin, Canada - Mikhail Zgurovsky, Ukraine

Executive Director: Kathleen Cass



**European Distributed Institute of Taxonomy**  
**EDIT**

*Muséum National d'Histoire Naturelle*  
57, rue Cuvier, CP43, 75231 Paris cedex 05  
(0)1 40 79 53 86/32 75 - Fax (0)1 40 79 53 99



Prof. F.A. Bisby  
Species 2000 secretariat  
Plant Sciences Laboratories  
School of Biological Sciences  
University of Reading  
Reading RG6 6AS  
UK

Paris, September 10th, 2008

**Letter of support for the FP7 proposal from Species2000 for completing the Catalogue of Life**

As a gathering of the most important taxonomic institutions in the world, the European Network of Excellence EDIT participates in and supports the Catalogue of Life. Indeed to reach its objective of integration of taxonomic research in Europe and increase the field's contribution to ecosystem and biodiversity research, the EDIT strategy relies on developing the IT tools which will make taxonomic research interactive in real time and give simultaneous access to taxonomic data, knowledge and references. To implement this strategy through the EDIT Internet Platform for Cybertaxonomy, the Catalogue of Life is central in order to maintain the relation between the various sources of taxonomic information and make them interoperable.

Beyond EDIT the completion of this programme is crucial for access to biodiversity information and sustainable management of biodiversity, which requires completion and integration of taxonomic indices. The Catalogue of Life is therefore central in international programmes aiming at providing a taxonomic information e-gateway, such as the GBIF, TDWG, PESI, Life Watch and EoL.

Professor Simon TILLIER  
EDIT Project Leader



EMBL- EBI  
Wellcome Trust Genome Campus  
Hinxton  
Cambridge  
CB10 1SD  
UK  
T +44 (0) 1223 494467  
F +44 (0) 1223 494496  
<http://www.ebi.ac.uk>

8<sup>th</sup> September 2008

Dear Professor Bisby,

The European Bioinformatics Institute supports the proposal by the Species 2000 consortium to the European Commission to develop its Scientific Data Infrastructure in a proposal entitled 'Distributed Dynamic Diversity Databases for Life' (4D4Life).

The European Bioinformatics Institute is a centre for research and services in bioinformatics. The Institute manages databases of biological data including nucleic acid and protein sequences, and macromolecular structures. Data resources are provided under a genomic model in which biomolecular information is organised in the context of the sequenced genome of the organism.

In our data repositories (most notably in the nucleotide sequence database) we have, a broad but shallow sampling of the diversity of species, where a potential 400,000 species have some representation in biomolecular records, most with relatively sparse information. Where sequenced genomes are not available, we focus on the sequenced genomes of close evolutionary relatives to ensure integration with the main body of biomolecular data.

We appreciate that a wealth of information is available for diverse species within biodiversity resources, such as those of the Global Biodiversity Information Facility, the Encyclopaedia of Life and the Biodiversity Heritage Library. We plan to develop technology that will connect our records, in the 'biomolecular domain', with information in the 'biodiversity domain'. In order to do this, it is imperative that there is a robustly maintained taxonomic backbone to biodiversity information, as will be provided under 4D4Life.

Specifically, based on 4D4Life infrastructure, we expect to propose projects that will develop technology and content to permit mapping between the Catalogue of Life and records at EBI. Approaches such as this, greatly facilitated by 4D4Life, will make available biodiversity information to biomolecular users and biomolecular information to biodiversity users.

Yours sincerely,

Graham Cameron

Guy Cochrane

European Molecular  
Biology Laboratory

Laboratoire Européen  
de Biologie Moléculaire

Europäisches Laboratorium  
für Molekularbiologie

September 9, 2008

Prof. F. A. Bisby  
Species 2000 Secretariat  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

Dear Professor Bisby,

The Encyclopedia of Life (EOL) is pleased to provide a letter of support for your proposal titled "Distributed Dynamic Diversity Databases for Life (4D4Life)", being submitted to the FP7 Scientific Data Infrastructure competition.

EOL ([www.eol.org](http://www.eol.org)) is an ambitious project devoted to assembling web pages documenting everything that is known about all species on Earth. These web pages are mash-ups of information from a variety of sources, including the Catalogue of Life. Begun in May 2007, EOL is now serving about 40,000 pages with at least moderate amounts of content, as well as an additional 1 million placeholder pages waiting to be populated with information.

However, the scientific community is still not sure how many species have been validly described to date (the consensus estimate is about 1.8 million), nor how many other species still wait to be described (estimates range from 10 to more than 100 million). For some groups of organisms, there is also dispute about what are the proper scientific names for the species. The Catalogue of Life is the only project currently taking a comprehensive approach to building a consensus taxonomy for already-known species, and to date has consensus names for more than 1 million species. For this reason, from its inception EOL has used the Catalogue of Life as its default taxonomy.

The activities being proposed by 4D4Life are significant, badly needed enhancements to Europe's (and the world's) taxonomic infrastructure. The proposed second-generation federated data architecture, taking advantage of the most modern IT developments, is absolutely essential to the further growth of the Catalogue of Life and, by extension, of EOL. In addition, the networking components, which will engage new partners and set up a global multi-hub network, are the best approach to meeting the Catalogue of Life's essential goal of completing a checklist of all the world's plants, animals, fungi and micro-organisms by 2011. EOL therefore gives the 4D4Life proposal its strongest possible support and looks forward to collaborating with the Catalogue of Life to ensure the proposal's success.

Sincerely,



James L. Edwards  
Executive Director

**MBL**

Biological Discovery in Woods Hole

 THE ENCYCLOPEDIA OF LIFE

**David Patterson**

*Senior scientist*

*Co-Principal Investigator, Encyclopedia of Life*

**MBL**

7 MBL Street  
Woods Hole, MA 02543  
USA

tel: 508.289.7260

fax: 508.457.4727

dpatterson@mbl.edu

Prof. F.A. Bisby  
Species 2000 Secretariat  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

Dear Frank

I write in support of your Distributed Dynamic Diversity Databases for Life (4D4Life) proposal to the FP7-INFRASTRUCTURES-2008 program.

The Encyclopedia of Life is a project that aims to deliver a web site for all species over the next 10 years. Our approach is to index and repurpose information already available through the internet. Our informatics strategy depends on a names-based infrastructure, and in particular relies heavily on the existing Species 2000 ITIS Catalog of Life structure for navigation and indexing of content. We welcome a further expansion of our interactions. The evolution of more electronic taxonomy services will ensure that the content is current. We are building an open and integrated cyberinfrastructure and are keen to work with you to realize your vision of a distributed computing system.

I believe these are exciting times for biodiversity informatics, and I trust that you are successful in your efforts.



September 5, 2008



Prof. Frank A. Bisby  
Species 2000 Secretariat  
Plant Science Laboratories  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

11 September 2008

Dear Dr. Bisby,

The Global Biodiversity Information Facility supports the ambitions of the Distributed Dynamic Diversity Databases for Life (4D4Life) proposal. I look forward to coordinating our own activities in developing a comprehensive infrastructure for taxon names with this project. To facilitate this, GBIF accepts your invitation to be a member of your Project Liaison Group and I will represent GBIF in this appointment.

Best regards,

David Remsen

Senior Programme Officer  
Electronic Catalogue of Names of Known Organisms  
Global Biodiversity Information Facility  
Universitetsparken 15  
DK2100, Copenhagen, Denmark

Global Biodiversity  
Information Facility  
Secretariat

Universitetsparken 15  
DK-2100 Copenhagen Ø  
Denmark

Tel.: +45 35 32 14 72  
Fax: +45 35 32 14 80  
Email: [dremsen@gbif.org](mailto:dremsen@gbif.org)  
Web: [www.gbif.org](http://www.gbif.org)

September 7, 2007

Professor Frank Bisby  
Executive Director, Species 2000  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

Dear Dr. Bisby,

I am pleased to write on behalf of the International Barcode of Life (iBOL) Project to endorse the proposal, *Distributed Dynamic Diversity Databases for Life (4D4Life)*, submitted by the Species 2000 consortium to the European Commission to develop its Scientific Data Infrastructure. The work described in this proposal represents a vital extension of the international biodiversity knowledge infrastructure. For iBOL, your work will provide an extremely important resource that is critical for achieving global taxon coverage and for linking barcode sequence profiles with valid taxonomic names. The objectives of our respective programs, to encompass all known organisms, are co-linear.

Perhaps it is worth noting that the iBOL research alliance spans 25 nations and will involve a projected investment of \$150M over the next 5 years. These funds will allow the development of a DNA barcode library that will provide a global resource for species identification. There are very many technical and scientific areas where this library will see application – in monitoring pathogens, invasive species, forensic issues, trade in biological materials and biotic surveys of all kinds. Ultimately, iBOL will form a central scientific component of the global biodiversity and climate change research infrastructure. Although iBOL has strong ambitions to advance biodiversity science, it cannot properly carry out its work without the important taxonomic validation studies undertaken by the Species 2000 Consortium. Certainly, your plans to launch new electronic services, combined with a move to a second-generation federated architecture under 4D4Life, will enable the development of a real-time connectivity between our respective informatics platforms. This, in turn, will yield important synergies for those working to protect and manage biodiversity.

In summary, my colleagues and I enthusiastically support the 4D4Life scientific data infrastructure proposal as a critical and creative extension to the advancement of global knowledge of biodiversity.

Yours sincerely,



Paul Hebert  
PhD, FRSC  
Director, Biodiversity Institute of Ontario  
Guelph, Ontario, N1G 2W1, Canada



IUCN  
Rue Mauverney 28  
1196 Gland  
Switzerland  
Tel. +41 22 999 0000  
Fax +41 22 999 0002  
mail@iucn.org  
www.iucn.org

Professor F.A. Bisby  
Species 2000 Secretariat  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading RG6 6AS  
UNITED KINGDOM

8 September, 2008

Dear Frank,

**Subject:- 4D4Life**

I am pleased to support the proposal by the Species 2000 Catalogue of Life consortium to the European Commission entitled *Distributed Dynamic Diversity Databases for Life (4D4Life)*.

As Head of the Species Programme of IUCN, I view the Species 2000 Catalogue of Life as an essential tool in the international biodiversity knowledge infrastructure. I am aware that many of the 7,000 or so expert volunteers of the IUCN Species Survival Commissions' Specialist Groups greatly value the services provided by the Species 2000 Catalogue of Life as do IUCN's members and partners.

We all need access to good quality and reliable taxonomic data to underpin our biodiversity conservation work and I am excited by the prospects of the enhanced services, developed through a consultative process, and the advanced federated architecture of the underlying system offered by the 4D4Life proposal. I do hope that the work goes ahead as outlined and look forward to involvement in the consultative process proposed.

With my best wishes

Yours sincerely,

A handwritten signature in black ink, appearing to read 'Jane Smart', written over a horizontal line.

Dr Jane Smart  
Head, Species Programme



FACULTY OF SCIENCE

*Zoological Museum Amsterdam*

Prof. dr Frank A. Bisby  
Species 2000 Secretariat  
Plant Science Laboratories  
School of Biological Sciences  
University of Reading  
READING, RG6 6AS  
UK

*Mauritskade 61  
P.O. Box 94766  
1090 GT Amsterdam  
The Netherlands  
Tel: +31 20-525 7191  
Fax: +31 20-525 5402*

**Subject: Letter of Support from PESI for Species2000 FP7 grant proposal**

Amsterdam, 9 September 2008

Dear Frank,

PESI<sup>1</sup> is please to see the Species2000 FP7 grant proposal and we much appreciate your lead in pushing forward the important programme on completing 'the Catalogue of Life for all known organisms'. It is clear that a proper development and implementation of environmental management systems is urgent because of global problems related to conservation and sustainable use of biological resources. Access to reliable biodiversity information is, however, hindered by the incompleteness of authoritative taxonomic meta-data sets. The Catalogue of Life is a vitally global programme for completing and integrating taxonomic indices, which is an important contribution to international efforts on establishing a common taxonomic information e-gateway in collaboration with associated principal initiatives, like GBIF, TDWG, PESI, Life Watch and EoL.

As you know the PESI project is the next step in integrating and securing taxonomically authoritative species name registers that underpin the management of biodiversity in Europe. PESI will integrate the three main all-taxon registers in Europe, namely the *European Register of Marine Species*, *Fauna Europaea*, and *Euro+Med PlantBase* in coordination with EU based nomenclators and the network of EU based Global Species Databases (the latter in close collaboration with *Species2000 Europe*).

Your activities complement PESI goals on establishing taxonomic standards for Europe. Therefore I am pleased to see that rebuilding the structure of regional hubs is part of your plan and that you envision a future PESI contribution to the *Catalogue of Life* as the so-called 'EuroHub', including an associated partnership.

---

<sup>1</sup> [www.eu-nomen.eu/pesi](http://www.eu-nomen.eu/pesi)



FACULTY OF SCIENCE

*Zoological Museum Amsterdam*

Good luck with completing the FP7 proposal and we very much hope that we will be able to congratulate you and your team in the near future with a positive outcome of your efforts.

Sincerely yours,

Yde de Jong

Project coordinator PESI

Project manager Fauna Europaea



OFFICE SeaLifeBase Project, WorldFish Center, Khush Hall, IRRI College, Los Baños, Laguna, 4031, Philippines  
° PHONE (+63-2) 580 5659 Loc 2827, (+63-49) 536 0202 ° FAX (+63-2) 580 5699 ° EMAIL  
www.sealifebase.org

9 September 2008

Subject: Support Letter to "4D4Life"

To Whom It May Concern:

The SeaLifeBase Project is a joint project between the Sea Around Us project of the Fisheries Centre, University of British Columbia (Vancouver, Canada) and The WorldFish Center (also known as ICLARM – International Center for Living Aquatic Resources Management), Philippines Office. The project is funded by the Oak Foundation (Geneva, Switzerland). SeaLifeBase, patterned after the successful model of FishBase, is an online Biodiversity Information System (BIS) that aims to provide general knowledge on the geographic distribution, biology, ecology, conservation and fishery biology of non-fish marine metazoans (see [www.sealifebase.org](http://www.sealifebase.org)).

SeaLifeBase, whose aims and objectives exclude the development of structure of taxonomy and nomenclature, acknowledges the significant global contribution of the 4D4Life project proposal, whose aim is to improve this structure to users like SeaLifeBase. Since its inception in 2005, SeaLifeBase has applied the Catalogue of Life as its taxonomic and nomenclature backbone. As our work relies on the data provided by the Catalogue of Life, we believe that the data comprehensiveness and service development and improvements targeted by 4D4Life is in no doubt of considerable importance, not only to us but also the whole biodiversity informatics community, and other BIS, e.g. Encyclopedia of Life (EoL). We, therefore, highly encourage its use and as a pledge of support, have made small contributions to the operating costs of Catalogue of Life.

As heavy users of the Catalogue of Life, it is thus not surprising that we highly recommend the completion of the 4D4Life. In addition, we look forward to continued support for Catalogue of Life from our biodiversity informatics colleagues.

Sincerely yours,

Dr. Maria Lourdes Palomares  
SeaLifeBase Project Coordinator  
c/o The WorldFish Center - Philippine Office  
Khush Hall, IRRI, College  
Los Baños, Laguna Philippines



**Taylor & Francis Books, Inc.**

Taylor & Francis Group plc

---

Albert House, 4<sup>th</sup> floor  
1-4 Singer St  
London  
EC2A 4BQ

Prof. Frank Bisby  
School of Biological Sciences  
University of Reading  
Whiteknights  
Reading  
Berks  
RG6 6AJ

6<sup>th</sup> September 2008

Dear Sir

I write to support the proposed further development of the Species 2000 Catalogue of Life services in the project 'Distributed Dynamic Diversity Databases for Life' (4D4Life) being submitted to the European Commission.

Taylor & Francis appears to be the first commercial user of the Species 2000 Catalogue of Life, and on the basis of our first year's contract, we are impressed with the growth and verified quality of the content which is extremely useful to us in enhancing our flagship product, the Dictionary of Natural Products on DVD. We are pleased with the online service provided from the Catalogue of Life web-site that has been available and functional throughout the period of the contract.

We look forward to the Species 2000 Catalogue of Life reaching its goal of covering all the world's taxa and to continue to use this very important facility.

Yours faithfully

Dr Steven Walford  
Managing Editor  
Taylor & Francis Group



CATHERINE N. NORTON  
Director  
MBL/WHOI Library

MARINE BIOLOGICAL LABORATORY  
WOODS HOLE OCEANOGRAPHIC INSTITUTION  
LIBRARY



7 MBL STREET • WOODS HOLE • MASSACHUSETTS • 02543-1015 • (508) 548-3705

September, 08,2008

Dear Frank:

I am delighted to speak for the uBio project at the MBLWHOI Library in Woods Hole, MA. US and to offer full endorsement for your proposal to the European Commission. uBio fully supports the Species 2000 consortium development of its Scientific infrastructure aptly titled the 'Distributed Dynamic Diversity Databases for Life' (4D4Life).

The Universal Biological Index and Organizer began in 2002 and is a networked information service for biological information offering through the Taxonomic Names Server a thesaurus of taxonomic names. This service is freely available to all to use and has many services that utilized taxonomic intelligence to manage information about organisms.

uBio relies on Species 2000, ITIS, and the Catalog of Life for its “gold standard” of names that is a fundamental resource in the international biodiversity knowledge infrastructure. What is proposed in this solicitation would be extremely useful to the biodiversity community as a whole and would enhance the current services that we use.

I wholeheartedly endorse this proposal.

Regards,

Cathy Norton  
PI, uBio  
Director, MBLWHOI Library  
Deputy Directory, Biodiversity Heritage Library  
Vice Chair, Biodiversity Heritage Library, Institutional Council

DIRECT TELEPHONE:  
(508) 289-7341  
(508) 289-2441

INTERNET:  
[cnorton@mbl.edu](mailto:cnorton@mbl.edu)  
[cnorton@whoi.edu](mailto:cnorton@whoi.edu)



Informatics Department  
1345 Jayhawk Boulevard  
Lawrence, KS 66045

Professor Frank Bisby  
Species 2000 Secretariat  
The Harborne Building  
School of Biological Sciences  
University of Reading  
Reading, RG6 6AS, UK

7 September 2008

Dear Professor Bisby,

I am enthused to support your initiative entitled: 'Distributed Dynamic Diversity Databases for Life' and pledge to collaborate with my department's staff and resources to integrate the web services you propose into our software applications. For 10 years, we have hoped for such an architecture, to be able to integrate access to on-demand, vetted, taxon names with our Specify software application for cataloging and internet publishing of museum holding information. We have used the ITIS Catalog of Life for this purpose in recent years, batch loading data into Specify, and it has been the single most useful source of taxon information that spans all of the biological collections disciplines. We now support over 220 biological collections with Specify ([www.specifysoftware.org](http://www.specifysoftware.org)) with 21% of that total being outside of the U.S. With the release of the open source, Java, cross-platform Specify 6 this fall, we anticipate supporting 350-400 collections in two years time. We are eager to enable each of those collections to use the web services you propose to validate and complete the taxon data needed for collection documentation. Specify 6 is specifically designed for that level of web-service integration.

Our other cornerstone project, Lifemapper ([www.lifemapper.org](http://www.lifemapper.org)) a geospatial archive of environmental niche models for the world's terrestrial species, is stymied for taxonomic indexing and higher-level taxon categorization, including browsing and searching functions, on account of the lack of a web-accessible taxon data source. Your proposed architecture would be a huge step forward for the provision of niche models for forecasting the impacts of global climate change and addressing other global-scale, multi-species research questions, if we could resolve the museum occurrence data in Lifemapper (derived from the GBIF cache), so that species names would be consistently applied. This would be an immensely valuable service for us and for the global biodiversity cyberinfrastructure.

Sincerely,

A handwritten signature in black ink, appearing to read 'James H. Beach'.

James H. Beach, Ph. D.  
Assistant Director for Informatics