

# 4D4Life



## Year 1 Summary

(17 June 2010)

### Distributed Dynamic Diversity Databases for Life

#### *Summary*

A coherent classification and species checklist of the world's plants, animals, fungi and microbes is fundamental for accessing information about biodiversity. The existing *Catalogue of Life* has started to provide the world with a unique taxonomic service: a dynamically updated global index of validated scientific names, synonyms and common names integrated within a single taxonomic hierarchy, and working towards covering all species of organisms. The 4D4Life Project has during Year 1 enhanced the existing programme and started the process of building a state-of-the-art *Catalogue of Life* e-infrastructure. This infrastructure is to provide a stable and sustainable electronic classification framework and species checklist for use as both taxonomic backbone and species index across global, regional and national biodiversity programmes and communities. A more robust and unified workflow, a service-oriented architecture, and a new array of electronic services are being tested in Year 2 and rolled out from Year 3.

#### *Principal Objectives*

- to synthesise a significantly improved global resource, the *Catalogue of Life*.
- to facilitate enhanced information exchange within the Species 2000 networks that feed into the *Catalogue*.
- to extend the reach of these networks to both inputs and public services through regional centres across the world.
- to disseminate this synthesised knowledge in a new array of modern web-services and products developed in partnership with identified user communities.
- to renew the distributed system with improved software and an enhanced service-based architecture.
- to implement a business model that will take this e-infrastructure to a sustainable future.

## *Year 1 of the Project*

Year 1 has seen a rapid build up of activities in all three compartments of the project – networking activities, provision of services, and research & development:

### *Networking Activities*

Networking has gone ahead strongly, with interactions with the extensive user base led by Sara Oldfield at BGCI, Richmond, networking with the Species 2000 array of 41 European global species database custodians, led by Thierry Bourgoïn at MNHN, Paris, and first steps in building the Global Multi-Hub Network of connected regional taxonomy hubs led by Jiri Kvacek at Narodni, Prague. Of particular interest is the outreach strategy now adopted. Some difficulties in preparing documents for the Multi-Hub Network are being remedied early in year 2.

### *Service Activities*

Substantial improvements to the existing services have been tested and rolled out during Year 1. In particular the principal existing public service, the Annual Checklist on the web and on disc, has had its software and interface re-written. Preparations are in full process for testing the unified workflow that will bring preparation of the Annual and Dynamic Checklists into one system. Software services led by Peter Schalk and Wouter Addink at ETI, Amsterdam have been put on a full documented development cycle. Public services implemented by the Species 2000 Secretariat at Reading have been both strengthened, and prepared for future changes towards the unified workflow, while all the time maintaining the production cycle for creating and extending the Catalogue and serving and distribution of its services.

### *Research and Development Activities*

Work led by Richard White at Cardiff on designing a new service-oriented architecture for the entire Catalogue of Life community started with a full needs and requirements analysis, and is progressing to a proof of concept demonstrator. The work was delayed at the start of Year 1, and steps to catch up are being implemented during Year 2.

### *Cross-co-ordination*

With the scale of parallel activities over 38 partners in the seven work packages of this project there is an important need for co-ordination. This has been achieved by planning the work in a Services Team convened by Sara Oldfield, and a Design Team convened by Alex Hardisty at Cardiff.

## *Results during Year 1*

*Networking Activities:* There have been significant successes and achievements in the networking activities:

**1. The outreach strategy** proposed in work package 2 has been adopted and implemented, with attractive materials and a startling improvement in visibility of the service and the programme world-wide. For instance the press release for the launch of the 2010 Special Edition at SBSTTA in Nairobi led to articles in 20 press media spread all round the world, and to a visible peak in the usage statistics. The 400 DVDs laid out at SBSTTA in Nairobi, and the 100 at a Congress in Dublin had all been taken long before the end of the meeting.

**2. The Triage List** of planned new services agreed by the Services and Design Team on 17 March provides a firm base for developing new services for trial during Year 2 – triaged between the priorities expressed by users, the supply abilities of the databases, and cost and management estimates for deploying them.

**3. Small enhancements of the Standard Data set** were agreed by the global species database suppliers, partly to improve functionality (e.g. use of unique identifiers), and partly to respond to user requests (e.g. habitat types: marine, freshwater, and terrestrial).

*Service Activities:* Much of the work has been behind the scenes, preparing for the major changes in Years 2 and 3 – the unified workflow with its new base schema and quality assurance workbench software, the new public services, and work with the research activity on designing the new architecture. Nonetheless there have been significant achievements in Year 1 visible now:

**4. New software and interface for serving the Annual Checklist.** This was planned, tested and implemented in partnership between the ETI team in Amsterdam, and the Secretariat team in Reading.

**5. Extending the Catalogue to 1.25 million species.** Ten new databases joined the programme during the year, some quite extensive (e.g. TaxaPad – Ichneumonoidea), and including additional contributions from Brazil (2 more databases on Mites from Piracicaba, and one on Freshwater Crabs from Manaus on the Amazon). This enabled the integrated Catalogue and hierarchy to reach 1.25 million species – a major benchmark in progress towards the 1.9 million goal.

**6. Special 2010 Edition.** The DVD was prepared as a Special 2010 Edition organised with the UN Convention on Biological Diversity, used as part of the International Year of Biodiversity celebrations, and publicly launched for us in Nairobi by Dr Ahmed Djoghlaif, Executive Secretary of the Convention.

*Research & Development:* The research & development activity is necessarily a preparation for Years 2 and 3. Significant advances were these:

**7. A Comprehensive Requirements Analysis** has been completed for the entire Catalogue of Life programme and its components.

**8. A Preliminary Specification for the new design** has been proposed.

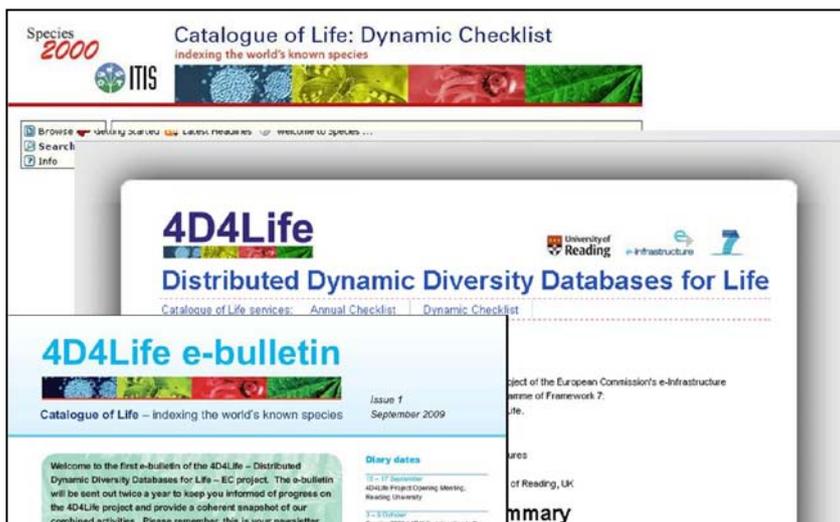
## ***Expected final results***

- i) a new array of Catalogue of Life public services,
- ii) a modern e-infrastructure with strengthened supplier base and system infrastructure,
- iii) the roll out of a state of the art service-based architecture,
- iv) extended community participation and taxonomic coverage through the Global Multi-Hub Network, around the world and in Europe,
- v) a mixed mode business model that takes the Catalogue of Life and its residual legal body, Species 2000, towards sustainable operation,
- vi) substantial progress towards completing the Catalogue of Life.

## *Expected scientific and socio-economic impact*

The Catalogue is a true infrastructure, used as much to *organise* and to *index* other biodiversity information, as to provide a source of information itself. So we do not expect all end-users to be reached through our own direct services, but rather expect to reach many indirectly through other systems that in turn use our infrastructure services to organise and index their own materials. We already see substantial usage figures through GBIF and EoL users, but it is now becoming clear that other global and regional biodiversity programmes are interested to use the Catalogue as a taxonomic backbone as well. These include not only the global programmes involved in a related new project (EBI/ELIXIR, Barcode Initiatives, IUCN Red List, LifeWatch in the i4Life project), but also regional programmes such as IABIN in South America and ACB of the ASEAN countries, and other EC projects such as KeyToNature and BHL-E/Europeana. We also have a growing confidence that there are opportunities for commercial and regulatory usage of the Catalogue as a documentation system for biotic materials world-wide. Negotiations are under way with a major multi-national company concerning documentation of materials, and the CBD has shown interest in using the Catalogue for the registry of Access & Benefit Sharing agreements to be proposed at the next Conference of the Parties in Nagoya.

From the expectations outlined above, it is clear that our impact world-wide may be twofold: in science, as the first comprehensive catalogue and taxonomic backbone for all organisms, and in e-science, as an electronic service for documenting the world's biotic resources utilised by biodiversity, commercial and regulatory organisations.



www.4D4Life.eu  
Contact: Professor Frank Bisby  
T: +44 (0)118 378 6466  
E: [f.a.bisby@reading.ac.uk](mailto:f.a.bisby@reading.ac.uk)