Catalogue of Life Plus (CoL+)

Project proposal version 10 July 2017

Executive project summary

All information related to species is tied to a scientific name. The Catalogue of Life has the mission to catalogue all known species: an authoritative consensus taxonomy produced by the global taxonomic community. Up to now it has completed peer-reviewed inclusion of nearly all extant species. The thus formed authoritative index allows the Catalogue of Life to serve as a means for linking biological species information. Many large institutional users (i.e. GBIF, EOL, Lifewatch) extend their copy of the catalogue with additional names and species to complete it to serve their own specific purpose. These disconnected efforts result in 'taxonomic inconsistencies' and cause confusion amongst users. The Catalogue of Life Plus (CoL+) project seeks to replace these efforts with a shared, extended catalogue and complete the reviewed name coverage without sacrificing quality. Creating an open, shared, and sustainable consensus taxonomy to serve the proper linking of data in the global biodiversity information initiatives is the ultimate goal the project contributes to. The goals for the the Catalogue of Life Plus (CoL+) project are: 1) creating both an extended and a strictly scrutinized taxonomic catalogue to replace the current GBIF Backbone Taxonomy and Catalogue of Life, 2) separating nomenclature and taxonomy with different identifiers and authorities for names and taxa for better reuse, 3) providing (infrastructural) support to the completion and strengthening of taxonomic and nomenclature content authorities, and 4) ensuring a sustainable, robust, and more dynamic IT infrastructure for maintaining the Catalogue of Life. The specific objectives are: 1) establishing a clearinghouse that covers scientific names across all life, provides a single taxonomic view grounded in the consensus classification of the Catalogue of Life along with provisional taxonomic sources, shows differences between sources, and provides an avenue for feedback to content authorities and allowing a broader community to contribute, and 2) establishing a partnership and governance for the clearinghouse and its associated components that enables continuing commitment after the project's end. The CoL+ project is a collaboration between the Catalogue of Life, the Global Biodiversity Information Facility Secretariat, Naturalis Biodiversity Center and partners with financial support by the Netherlands Biodiversity Information Facility (NLBIF) and the Netherlands Ministry of Education, Science, and Culture.

Glossary

ACEF: Annual Checklist Exchange Format

API: Application programming interface

BHL: Biodiversity Heritage Library

BOLD: Barcode of Life Data systems

COL: Catalogue of Life

DwCA: Darwin Core Archive

EOL: Encyclopedia of Life

GBIF: Global Biodiversity Information Facility

GNA: Global Names Architecture

GNUB: Global Names Usage Bank

GSD: Global species database

IPT: Integrated publishing toolkit

NLBIF: Netherlands Biodiversity Information Facility

ORCID: Open Researcher and Contributor ID

OTU: Operational taxonomic unit

TCS: Taxon concept schema

Table of contents

Background

In 2015 the global biodiversity data aggregators GBIF, CoL, EOL, BHL and BOLD took the initiative to work on the idea for building a single shared authoritative taxonomic backbone that can be used to order and connect biodiversity data across various domains. These initiatives each focus on delivery of a consistent, normalised view of available data for a particular class of biodiversity information (GBIF - specimens and occurrence records, CoL - species names and concepts, EOL - species traits and species-level information resources, BHL - biodiversity publications, BOLD - barcode sequence records). All of them depend on the use of scientific names and the associated species concepts as a fundamental axis for organising their data, and accordingly all recognise the benefits that would arise if they had access to a comprehensive and authoritative taxonomic framework for all portions of the tree of life. If such a taxonomic framework can be used by all of these data aggregators, this will facilitate increased tight linkage between available instances of all of these data classes.

On this basis the Netherlands, through NLBIF, and the Naturalis Biodiversity Center, and Species 2000 made an offer at the 23rd GBIF Governing Board meeting to contribute funds, project management, informatics and IT capacity for a project to develop this concept further in close collaboration with GBIF, CoL, and other interested parties. During November 2016, GBIF and Catalogue of Life hosted a workshop in Leiden, Netherlands, under the title “Names in November”. The meeting included representatives from a number of key initiatives that are directly involved with managing nomenclatural data or are producers of taxonomic content found in the Catalogue of Life. The current project proposal is the result of a follow-up meeting with the Catalogue of Life Global Team in Woods Hole 10-14 April 2017.

The Catalogue of Life is the most comprehensive and authoritative global index of species currently available. It is constructed from more than 130 discrete taxonomic datasets to form a single integrated species checklist and taxonomic hierarchy. The Catalogue holds essential information on the names, relationships and distributions of over 1.7 million species. Several global biodiversity information initiatives, like GBIF, would like to use the Catalogue of Life as a taxonomic backbone. Although the Catalogue of Life has a broad coverage this coverage is still insufficient, making a direct use as taxonomic backbone for other purposes limited. Various institutional users enrich the CoL with their own names to ensure that data can be linked. There are various reasons for the gaps in current coverage of names. In some cases ‘accepted’ names lack complete information on synonyms. Another cause is the existence of taxonomic gaps resulting from groups which are not yet covered within the Catalogue of Life, particularly for fossil taxa. Lastly the focus on completing the catalogue at the level of species results in the omission of a large number of infraspecific taxa which are nevertheless referenced in available biodiversity information resources. The current Catalogue of Life data management infrastructure needs updating. Building the CoL still involves some manual labor, has limits on the hosting side and sustainability of its ICT solutions.The CoL+ project provides solutions to overcome these barriers.

In addition to the Catalogue of Life as a comprehensive taxonomic backbone, it is essential that the major nomenclatural databases (as stated on page 11) are reinforced and completed. Just

as GBIF, EOL, BHL and other biodiversity data infrastructures need a comprehensive Catalogue of Life and seek to entrust Catalogue of Life with responsibility for delivering and maintaining a common taxonomic backbone, Catalogue of Life (along with the other biodiversity data infrastructures) needs a comprehensive nomenclatural foundation and seeks to entrust the nomenclators with responsibility for delivering and maintaining this. A close partnership will enable all parties to work together efficiently and to relinquish some existing activities to secure a more comprehensive and secure shared solution.

Goal and objectives

Goals
Creating an open, shared, and sustainable consensus taxonomy to serve the proper linking of data in the global biodiversity information initiatives is the ultimate goal the project contributes to.

The goals for the Catalogue of Life Plus project are:

1. Creating both an extended and a strictly scrutinized taxonomic catalogue to replace the current GBIF Backbone Taxonomy and Catalogue of Life;
2. Separating nomenclature and taxonomy with different identifiers and authorities for names and taxa for better reuse;
3. Providing (infrastructural) support to the completion and strengthening of taxonomic and nomenclature content authorities; and
4. Ensuring a sustainable, robust, and more dynamic IT infrastructure for maintaining the Catalogue of Life.

Project objectives
This project has the following specific objectives:

1. Establishing a clearinghouse that covers scientific names across all life, provides a single taxonomic view grounded in the consensus classification of the Catalogue of Life along with provisional taxonomic sources, shows differences between sources, and provides an avenue for feedback to content authorities and allowing a broader community to contribute, [clearinghouse].
2. Establishing a partnership and governance for the clearinghouse and its associated components that enables continuing commitment after the project's end [partnership & governance].

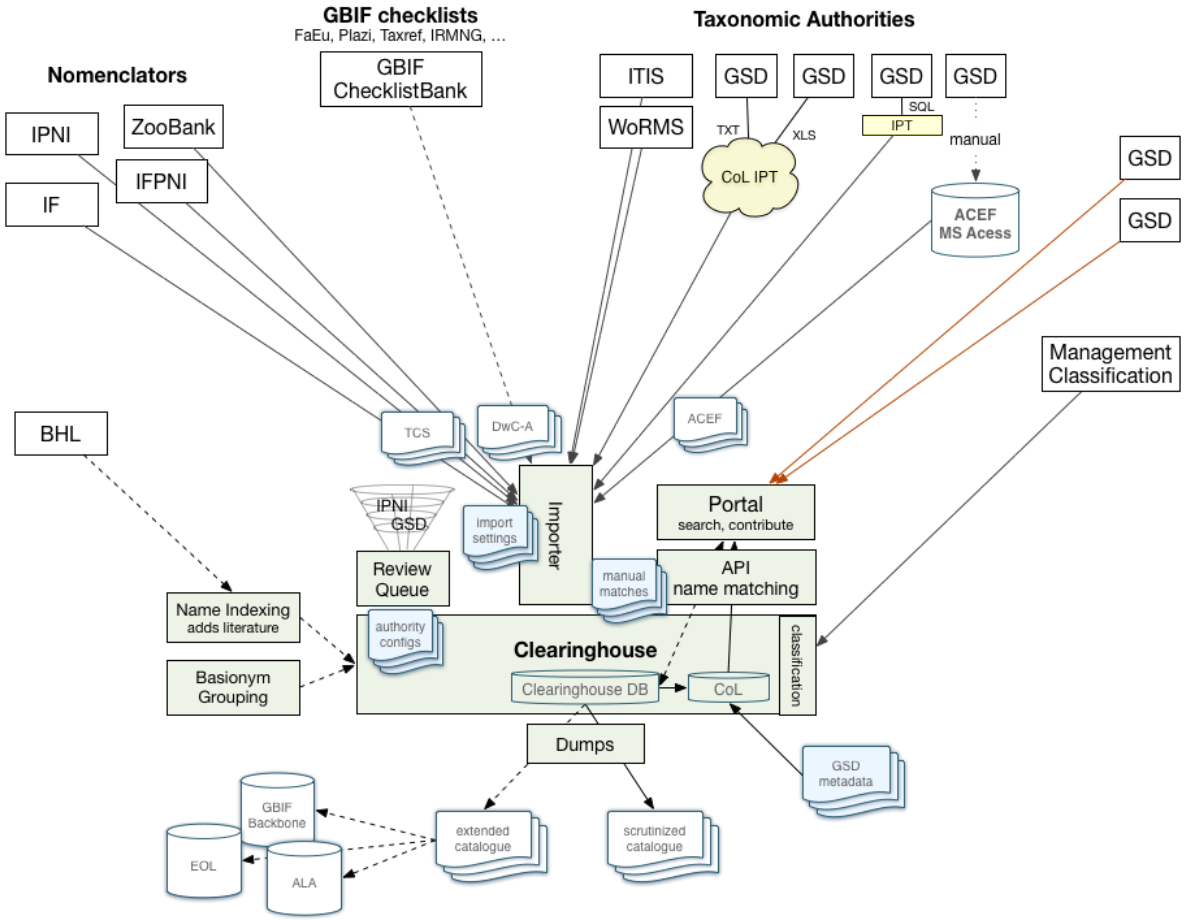

Figure 1 Green: *Clearinghouse*; Blue: *Editorial work*; Yellow: *Consolidated data publishing*
Solid lines: *scrutinized data;* Dashed lines: *provisional data;* Red lines: *remote data management*

Objective 1 Establishing a clearinghouse for taxonomy and nomenclature

The clearinghouse will cover scientific names across all life and provide a single taxonomic view grounded in the consensus classification of the Catalogue of Life along with provisional taxonomic sources. It will be based around a strict separation of nomenclature and taxonomy and allow the merging, addition and conflict resolution of data, completing information missing in other sources. Integration of data from different sources is done by relating different records to the ‘same’ name regardless of obvious spelling variations. Subsequent recombinations or the same binomial with a different author will be treated as distinct names. As a single record may ultimately be comprised of information from several sources, the provenance not only of the entire record but for each attribute will be tracked independently. A robust matching service for name strings (normally associated with some higher classification terms which serve as hints for resolution) is therefore the heart of the system and should be continuously improved to give the best results possible. Purely automated processing will never be sufficient. As part of the editorial work, bad automatic matches can be fixed manually. The clearinghouse will assign persistent identifiers to both names and taxa, where available coming from existing nomenclators. Vernacular names, species range distributions, lifezone and fossil information will also be included to build a fully scrutinized rich catalogue. Community participation is desired and any record will be open for public comments. Areas in the taxonomic tree not governed by content authorities yet are exposed for communal editing. The portal will expose metrics about contributions and allow users to customize their experience by setting preferences for specific taxonomic groups. Communal management of a freely licensed reference image will be explored. The clearinghouse and its associated components will replace the current Catalogue of Life software.

The clearinghouse will encompass (at least) the following products (or product groups):

Product 1.1 API
A thorough documentation, implementation and deployment of the API with its data model, audit trail, methods and business rules covering nomenclature & taxonomy. It covers taxa, synonymies, vernacular names, range distributions, lifezone and fossil information and includes a sophisticated name matching service. An agreed and documented set of rules explaining how taxon concepts and their identifiers are handled in the clearinghouse will be developed. All historical annual Catalogue of Life editions and the latest version should be exposed through the same installation. The essential feature set of the existing Catalogue of Life services including DwC-A downloads should be migrated. The database will enable the storing of versions of every historic Catalogue of Life assembly and enable the resolution of taxonIDs even if deleted. The API will serve external users as well as the portal, which will allow communal editing for authenticated users, preferably with ORCID accounts.

Prerequisites and dependencies

- Dependency on GBIF hosting, monitoring & central logging
- Deprecate the current CoL API at some stage and help users to migrate to the new API

Out of scope

- Cultivar names, hybrid formulas and well known OTUs (e.g. BOLD Barcode Index Numbers) will not be included in the CoL+ project.
- Alternative taxon concepts for the same name are not addressed. The extended taxonomy will also be a consensus.
- Renewing or updating the 4D4Life web services.

Product 1.2 portal

A portal for the clearinghouse build exclusively on top of the API (1.1). It acts both as a portal providing searches and includes authenticated community editing and reviewing of scientific names. The scrutinized and extended taxonomic catalogue will both be discoverable through the portal. The portal will replace the current Catalogue of Life annual checklist interface and the Drupal based CMS pages. Periodic editions (e.g. monthly and annual) of the Catalogue of Life (the scrutinized catalogue) will be available through the portal. Each annual Catalogue of Life edition will be downloadable. A mechanism to support the assembly of the annual CoL DVD will be sought.

Prerequisites and dependencies

- Dependency on a stable API (1.1)
- Dependency on GBIF hosting, monitoring & central logging
- Agreement on the migration of the current Catalogue of Life portal functionalities to the new portal
- The phasing out of the existing Catalogue of Life infrastructure within Naturalis and associated planning of maintenance of the current infrastructure.

Out of scope

- The level of internationalisation of the current Catalogue of Life portal may not be achieved during the CoL+ project. It will depend on the amount of assistance we can get from the species2000 secretariat and the wider community.

Product 1.3 importer

A software implementation that allows the ingestion of data in different standards (e.g. ACEF, DwC-A, TCS) originating from the nomenclators, Global Species Databases including data from the species2000 editorial office, and taxonomic information from the GBIF checklist bank. The importer also merges the information from different sources using the name matching API and tracks the provenance for a data record and its attributes. Stable taxon identifiers will be assigned based on objective, computable rules which will have to be determined.

Prerequisites and dependencies

- External information sources exposing their data in the appropriate data standards
- Dependency on a stable API (1.1)
- Dependency on GBIF hosting, monitoring & central logging

Out of scope

- Information from other biodiversity sources not registered with GBIF

Product 1.4 review tools
The clearinghouse will also implement nomenclatural and taxonomic rules to flag issues for review. It will experiment with tools for reviewing and highlight potential problems such as homonyms, orthographic variants, missing basionyms, conflicting synonymies or missing brackets in the authorship of recombinations. It will allow reviewing all sources for the same name in a single view. A review queue of issues could be brought under the attention of existing nomenclators and Global Species Databases or provide a starting point for community input.

Prerequisites and dependencies
- Ability to absorb information from the review queue and responsiveness of content authorities and community
- Some outreach to content authorities and the wider community is needed

Out of scope
- Automatic synchronisation with the content authorities

Product 1.5 content bots
The clearinghouse will be assisted by "bots" that augment missing data as much as possible, e.g. by deriving basionym relations and linking names to exact pages of their original description in BHL. Especially creating links to existing data can often be discovered by software and proposed to the system for later, manual review by experts. Examples of the information that can be discovered are:

- links & literature metadata for the original name publication by
  - scanning the BHL Name Index
  - querying additional sources, e.g. GNUB, BioNames
- links to type specimens in GBIF or JSTOR
- the basionym of a recombination by looking at shared terminal epithets from the same original author within a family.

Prerequisites and dependencies
- Dependency on a stable API (1.1)

Product 1.6 releases
From the clearinghouse a scrutinized and an extended catalogue (consisting of scrutinized and provisional data) should be released - either on editorial request or in a regular weekly or even daily interval. The extended taxonomic catalogue will be the basis for a taxonomic backbone for GBIF and other biodiversity information aggregators. A release will be dumped as a file and archived in a database.

Prerequisites and dependencies
- Dependency on a stable API (1.1)

Objective 2 partnership & governance

For the clearinghouse and its associated components a partnership and governance is required. Service level agreements between different partners are needed to ensure the system is developed and sustainably maintained. The governance model should ensure the responsibility for improving content remains with the respected content authorities (e.g. nomenclators and global species databases/Catalogue of Life), as well as with the editors of the scrutinized portions of the catalogue (i.e. the Catalogue of Life) and of the extended portions. A governance should also ensure others can contribute to gaps in content covered within the clearinghouse and or join in the development and expansion of the clearinghouse. The clearinghouse should be developed through stated and prioritized user requirements. An engagement plan describing the main stakeholders and how interaction with these stakeholders takes place, is needed.

The partnership and governance of the clearinghouse should develop or use existing rules and agreements on the following elements:

Editorial work
A scrutinized catalogue (Catalogue of Life) relies not only on the work of global species databases and nomenclators, but also on an overall editorial work that verifies data, checks consistencies and asserts that species from different taxonomic groups do not overlap. Apart from isolated direct changes to names and taxa the clearinghouse needs editorial work persistently stored for the following aspects:

- Manual name string matches to override wrong automated matches OR to confirm validation of automated matches
- Various source specific settings for importing. This will likely include filters to exclude certain names and taxonomic groups, custom mappings of ranks and mappings of families and higher taxa to the CoL classification
- Authority configurations defining the part of the tree that is governed by an authority
- Metadata for sources covering the existing CoL GSD metadata[1]

By flagging issues that can be detected by rules the system will assist in the evaluation of data quality. This includes interpretation of controlled vocabularies like rank, country or language.

For the provisional taxonomic data an editor or a controlled group of editors should also ensure one taxonomic consensus view, in the circumstance the Catalogue of Life is not providing this view for particular taxonomic groups. GBIF and its ChecklistBank offers many good taxonomic sources[2] that can be used to fill gaps in the current Catalogue of Life with provisional data. In

[1] http://www.catalogueoflife.org/annual-checklist/2017/details/database/id/83
[2] http://www.gbif.org/dataset/search?type=CHECKLIST

particular, the resources used to build the GBIF Backbone[3] have proven useful and should be used to fill gaps.

Content authorities
Various nomenclators already exist that will act as authorities for names in their specific group. Any additions or changes to the names within such group found in the clearinghouse will be communicated back for final review. It should be explored how to prioritize names within the review queue, e.g. by the frequency of names used in the name matching service.

Presently CoL+ seeks collaboration with the following nomenclators:

- The International Plant Names Index (seed plants, ferns and lycophytes)
- Index Fungorum (fungi)
- The International Fossil Plant Names Index and/or The Fossil Plant Index (fossil plants)
- PhycoBank (algae)
- Index Nominum Hepaticarum and/or Tropicos (mosses).
- ZooBank (animals)
- Prokariotic Nomenclature Up-to-date (bacteria)
- ICTV (viruses)

All nomenclatural codes have started initiatives for registration of new names with most of the above nomenclators also being official registrars. It is desirable to see newly registered names showing up in the clearinghouse as quick as possible. For the range of ongoing registration initiatives within the framework of ICN see Barkworth et al. 2016[4].

The Catalogue of Life annual checklist 2017 is created from 156 source databases[5] which act as taxonomic authorities for their group. These sources currently provide information about both the nomenclature and taxonomy. For large areas in zoology where there is no existing nomenclator covering scientific names, the source databases of the Catalogue of Life will act as the *de facto* authorities also for names. In other well curated groups like higher plants data from the nomenclator will be preferred.

Data contributors
Ideally all sources expose their data in a standard exchange format. Currently, global species databases manage their data in all kinds of custom ways (including Word documents). Consolidating the data publishing could be a project on its own, and is not something to which

[3] https://github.com/gbif/checklistbank/blob/master/checklistbank-nub/nub-sources.tsv
[4] https://doi.org/10.12705/653.43
[5] http://www.catalogueoflife.org/annual-checklist/2017/info/databases

CoL+ can aspire at this point. Nevertheless CoL+ tries to show some path forward and offers various ways to contribute to the clearinghouse:

- DwC archives with extensions for vernacular names, distributions and nomenclatural relations.
- CoL data submission format, allowing the continuation of manual data entry based on the ACEF format (Annual Checklist Exchange Format)
- TCS for sharing rich nomenclatural data (support by nomenclators provided)

A shared CoL IPT[6] will be hosted for GSDs to upload spreadsheets and text files and convert them into supported formats.

Manual entry of GSD data by the Catalogue of Life into an ACEF compliant assembly database will be continued as needed by GSDs.

Community contributions
Direct contributions by any authenticated user, regardless of institutional affiliation, will be permitted in groups not yet well governed by authorities, both for nomenclature and taxonomy. Establishing active links to literature sources allows objective information to be verified. This approach is well suited to communal contribution and review in areas lacking authorities, and will allow much better scaling to complete the coverage of all original and subsequently recombined names and tie them to their publication.

As with external sources every change applied to the clearinghouse by a contributor is tracked in the provenance of the data. In well managed taxonomic groups access will be controlled and only simple commenting or suggestions will be supported. The exact governance model needs to be established during the project.

Wider stakeholders
The following main groups of stakeholders could initially be discerned apart from the CoL+ project partners:

- Global biodiversity databases
- Nomenclator initiatives
- Biodiversity information initiatives at global, international, national, and or regional scale (including biodiversity management authorities, NGOs, GBIF national nodes, the Convention on Biological Diversity)
- Current users of the Catalogue of Life
- Taxonomist, biodiversity conservation specialists, scientists
- The wider biodiversity informatics community, including research infrastructure initiatives (e.g. LifeWatch, DiSSCo)

[6] http://www.gbif.org/ipt

During the CoL+ project it is important to capture user requirements from these wider stakeholders, map and prioritize these requirements to the development of the CoL+ clearinghouse.

Hosting and maintenance of the clearinghouse system
The clearinghouse infrastructure will be hosted through the GBIF Secretariat. All technologies chosen to build the clearinghouse and associated components should match the GBIF Secretariat hosting environment and allow better maintenance in the future. A consortium formed during the CoL+ project, starting with the Catalogue of Life, GBIF Secretariat, and Naturalis will develop and maintain the clearinghouse. This consortium will be open for other parties to join the development.

Data license
All data in the clearinghouse should be open and freely reusable by anyone. All contributions are expected to be compliant with at least CC-BY, but ideally follow CC0 Creative Commons data licensing.

Risk assessment

Potential risks and mitigation
(1 = low, 3 or 9 are high)

	Risk + consequence	**Likeli-hood (L, 1-3)**	**Impact (I, 1-3)**	**Risk score (L*I, 1-9)**	**Mitigation strategy**
1.	Risk: The clearinghouse as presented within the current proposal represents a very integrated view. This view may be too integrated and may not sufficiently respond to user requirements by the main project partners. Consequence: Prioritized user requirements may be insufficiently met.	2	3	6 (moderately high)	The development of additional components, like an API and/or portal might be considered in the case priority user requirements are insufficiently addressed by the CoL+ proposal. To avoid resourcing issues,an attempt to broaden the consortium for developing the CoL+ clearinghouse and its associated components will be made.
2.	Risk: The clearinghouse is not ready during the project period to replace the current Catalogue of Life infrastructure. Consequence: The current CoL infrastructure would have to	2	3	6 (moderately high)	It is the intent to strive for migrating the current functionality in the CoL infrastructure as best as possible within the CoL+ project. A plan for phasing out the current CoL infrastructure should be made. It is wise to plan to keep the current CoL infrastructure in parallel to the new system for a while, especially to allow

	be run in parallel to the new infrastructure for a period in time after the CoI+ project has ended.				sufficient time for users to switch over to the new API.

The project steering committee assesses the potential risks and mitigation on a regular basis. Re-evaluation of the risks occurs after a period of developing and testing the data model of the clearinghouse and associated components.

Project organisation

Steering committee
The project sponsor will be Peter Schalk, executive secretary Species 2000, as representing the governance of the Catalogue of Life. The role of specialist will be covered by Donald Hobern, executive secretary of the Global Biodiversity Information Facility, and Jeroen Snijders Chief Information Officer Naturalis Biodiversity Center. The steering committee could be extended on a needs basis; for example when the consortium of contributing partners for developing the clearinghouse grows.

The secretary for the CoL+ project will be he Naturalis Biodiversity Center.

Project manager
The project manager is Olaf Bánki working at the Naturalis Biodiversity Center.

Project team
The project team will consist of staff from the Naturalis Biodiversity Center and the Global Biodiversity Information Facility as well as from other project partners.

The following team will be responsible for the achievement of products:

- Markus Döring will have the technical lead for developing the clearinghouse and associated components (including architectural design);
- Products 1.1 API, 1.3 Importer, 1.4 Review tools, 1.5 content bots, and 1.6 releases will be developed by Markus Döring, Ayco Holleman, and possibly others
- The execution of product 1.5 may involve Dmitry Mozzerhin
- Front-end design and development of product 1.2 portal will be done by GBIF Secretariat and/or Naturalis staff or possibly others. This will be determined after the data model of the clearinghouse is determined;
- The execution of objective 2, including capturing user requirements, will be done by Olaf Bánki, David Remsen together with the project steering committee and others. Yuri Roskov and Luisa Abucay will be involved to ensure the Catalogue of Life editorial work can happen properly within the clearinghouse

The following persons could augment the project team by performing particular tasks:

- Interaction with the Catalogue of Life executive editor Yuri Roskov and database manager Luisa Abucay
- Connections with existing CoL infrastructure: Wouter Addink, Ruud Altenburg, Hugo van Duijn, Wilfred Gerritzen

- Domain experts: e.g. Nicolas Bailly, Thierry Bourgoin, Tom Orrell, Rod Page, Rich Pyle, David Remsen
- Architectural design: Tim Robertson
- GBIF helpdesk for CoL+ cloud based IPT installation
- A community of testers will be formed to test the product developments of the clearinghouse and its associated components at various stages in their development

The Catalogue of Life governance is represented by Peter Schalk in the steering committee of the CoL+ project. At appropriate moments the Catalogue of Life global team will act as an advisory board for the CoL+ project. Some members of the global team can pick-up tasks as domain experts for the CoL+ project team. In discussion with the project management smaller working groups could be formulated to resolve particular issues. In principle, David Remsen is responsible for ensuring connections and avoid overlap with activities of the information systems group of the Catalogue of Life global team. Nicolas Bailly is responsible for ensuring connections and avoid overlap with activities by the taxonomy group of the Catalogue of Life global team.Implementation strategy and planning

Approach
The development of products within the CoL+ project will be based on the principles of agile development as best as possible with an associated periodic (daily) scrum (involving the lead developers, project manager, and others where needed). This means short reiterated cycles of specifications, development, and testing will take place for each of the components of products. The focus will be at first on defining the API of the clearinghouse before continuing the development of other clearinghouse components. A main CoL+ repository is set up in Github. If necessary, other channels for communication could be setup to involve other non-technical experts.

Planning
In broad terms the CoL+ project actually started at the 23rd meeting of the GBIF Governing Board in October 2016. The NLBIF co-funded project phase will run from the 1st of May 2017 to 30th of April 2019. Markus Döring started formally working on the project by the 1st of April 2017. Olaf Bánki started formally working on the project by the 1st of May 2017. In terms of planning the establishment of the project team is a priority. After the project team is established a more detailed planning covering short periods of time could be established.

An initial sequence in the development of the listed products is given below, with the first items being the first to be undertaken:

- API definition (1.1)
- API Implementation (1.1)
- Importer (1.3)
- Releases (1.6)
- Portal (1.2)
- Content bots (1.5)
- Review tools (1.4)

Defining the API of the clearinghouse is the first activity that will take place, and work has already started. The API definitions should be progressed by the fall of 2017 so these could be presented to a wider community (e.g. TWDG). After that the development of the different CoL+ clearinghouse components could start. The project team will strive for deploying the portal (1.2) at the end of 2018.

Reporting and branding

Progress on the CoL+ project could be reported at the GBIF 24th and 25th Governing Board meetings. A mid term content and financial report covering the entire project (both in-kind and in-cash efforts) will be delivered to NLBIF in April 2018 and a final report in April 2019. All products will be branded with the logo's of the main project partners and funders.

Budget

Delivering the clearinghouse and associated components requires additional funding beyond the current capacity of the partners developing the infrastructure. Funds from NLBIF will be instrumental to ensure proper project management, engagement with key stakeholders, and to ensure key developers and domain experts can work on getting towards the deliverables. To show longer term commitment by project partners substantial in-kind commitments are provided as well to ensure the best delivery of the clearinghouse infrastructure.

Item	**Project partner**	**In kind budget K euro project partners**	**Budget in K euro 2017 - 2019 (NLBIF)**
Project management	Naturalis		60
User requirement capture, development of examples and documentation	MBL		20
Development activities for CoL+, development of scoping document, architectural design etc.	GBIF Secretariat	120	40
Development activities for CoL+	Naturalis	120	80
Maintenance, development and systems design of the current Catalogue of Life infrastructure	Catalogue of Life	120	
Additional developer capacity	to be recruited		80
Travel and meetings	all		50
Application Hosting and inkind services like build server, a central logging server and monitoring	GBIF Secretariat	8	12
Miscellaneous costs (e.g. communications costs)			18
	Total	**368**	**360**
Overall project budget in K euro	**728**		