**Species Checklist Cross-mapping and Precursor National Checklists Generation from GBIF-mediated Data**

David Remsen[1], Wouter Addink[2,3]*

[1] Marine Biological Laboratory, Woods Hole, MA, USA

[2] Species 2000, Leiden, Netherlands

[3] Naturalis Biodiversity Center, Leiden, The Netherlands

* Corresponding author

E-mail: wouter.addink@naturalis.nl

# Abstract

In a hackathon, three teams worked together to investigate the workflow needed for establishing or improving national species checklists. We investigated existing open-source tools and services, improved them where needed and created new open source solutions to establish a workflow for national species checklist creation and curation. When the components are further developed and suggested improvements are implemented, the tooling will be in place to establish an internationally standardized, common workflow.

# Introduction

Regional checklists of species provide a consolidated catalog of biodiversity related to a geographic area and are important tools for collecting and managing biodiversity data. Without complete national checklists, for example, countries cannot effectively implement national scientific biodiversity networks [1,2]. A common approach for establishing national checklists does not yet exist. Establishing a common approach would reduce duplication of work and could also introduce interoperability between national checklists in international scientific biodiversity networks . With its global and comprehensive scope, the Catalogue of Life (CoL)[2,3], when complete, would, in principle, include any species referenced in any national species checklist. With over 1.6M species already in the Catalogue, we sought to investigate how it might compare to existing national lists in coverage and how it might be used, in combination with other sources, such as the Global Biodiversity Information Facility (GBIF) [4], to support the discovery of missing species candidates to such lists [5]. The GBIF portal provides access to over 500 million biodiversity data records, from over 13,000 different source databases and many of these contain geographic information that may be linked to a country. Our goal was to explore whether such data might be used to locate species missing from existing regional/national checklists or as the basis for generating new lists. In the form of a hackathon we explored issues related to the creation, maintenance and interoperability of national species checklists, both among each other and with the Catalogue of Life.

# Materials and Methods

A hackathon was organised by Species 2000 as part of the GBIF Capacity Enhancement Support Program [6] and hosted at Naturalis in Leiden, The Netherlands, from 2-5 March, 2015. A total of 20 participants attended the event, representing 10 different countries and multiple institutions, all involved in creating national checklists, managing national checklists, or developing tools for national checklists.

This broad topic allowed us to divide the group into three teams focusing on sub-topics:

1. **Team 1 - Cross-mapping** refers to the comparison of a target list to a reference list with the goal of identifying overlapping and distinct taxa within each list [7]. This group focused on the normalization of taxon names for comparison and the development of a confidence score in matches.
2. **Team 2** - **Annotations** refers to a system that allows curators to annotate checklist cross-mappings with GBIF occurrence data to provide feedback on the presence of legitimate or suspect species occurrences within a country[8].
3. **Team 3 - Distributions** of species, sought solutions to gather evidence that a species occurs in a country to support accurate annotations.

Each team was composed of 4-5 persons with each individual assigned roles to support the objectives of the group. This included, at least, a team leader, a senior programmer and a reporter. For capacity-building reasons, people could also have a junior programmer role, to learn from the other programmers in the team. Others identified as domain-experts and moved between groups to share expertise.  Further roles included tester and data analyst.

The teams worked in separate rooms with an initial task to define the user stories that would form the basis of their work.  Work proceeded for the remainder of the meeting with regular meet-ups and discussions at the end of the each workday. Code and documentation was stored in a GitHub repository and open source licences were added.

The following user story formed the basis of work for the Cross-mapping team (team 1):

**User Story 1-1**

*As an owner of a checklist I want to know whether I am missing species for my country that are reported to occur there by other authoritative sources.*

**Goals**

1. Assess how names in national checklists correspond to names in CoL - quantity and quality of positives and negative matches resulting from current parsing and naming tools.

2. Use these examples to assess and refine the Global Names Architecture name-parsing and matching algorithms (http://gni.globalnames.org/parsers/new)

3. Provide recommendations for future checklist cross-mapping tools and interfaces

The team implemented a system in Ruby programming language that uses Global Name (GN) webservices (http://resolver.globalnames.org/data_sources.json, http://resolver.globalnames.org/data_sources) to identify matching taxon names in a target list (a national checklist) against a reference list. CoL was used as reference list containing valid taxon names and synonyms. Four candidate target lists were used to test the name-parsing and matching services (Table 1). Test results were used to improve the GN webservice.

**Table 1. National species lists used to test cross-mapping tools.**

| List | Number of Taxa |
|---|---|
| Portuguese Bryophyte | 726 species |
| Portuguese Tracheophyta | 4,458 species |
| Slovakian Checklist | 1,248 species |
| Norwegian Checklist | 10,950 species |

Team 2 focused on a system design that could be used to present assertions of a taxon occurrence within a country - to a presumed expert curator, who might then use their knowledge to assess the assertion and determine whether the taxon should or should not be added to the list. The authoritative Catalogue of Life record - linked through the cross-mapping efforts of Team 1, would then form the record-of-authority for the national list. In addition, negative matches (i.e, species asserted to occur within the country but determined to not belong there - might be linked to a comment or annotation that could serve to inform future users of the GBIF network to the nature of the suspect occurrence. This led to the articulation of the following user story.

**User Story 2-1**

> As an owner of a national checklist I want to load my checklist into a system and compare it to the list of taxa assigned to my country within the GBIF index. Matches missing from my national list may 1) represent legitimate missing taxa that should be candidates to add to my list. They may also 2) represent taxa erroneously applied to my country that should be annotated with their suspect status for future users of the record.

**Goals**

1. Can the federated GBIF portal be used to support the identification and qualification of novel species occurrence records in the development of national or regional species inventories?
2. Can annotation interfaces be used, in combination with authoritative regional or national species lists, to identify and annotate potentially erroneous species occurrences and thus inform future users of GBIF-mobilized data as to this erroneous assessment?

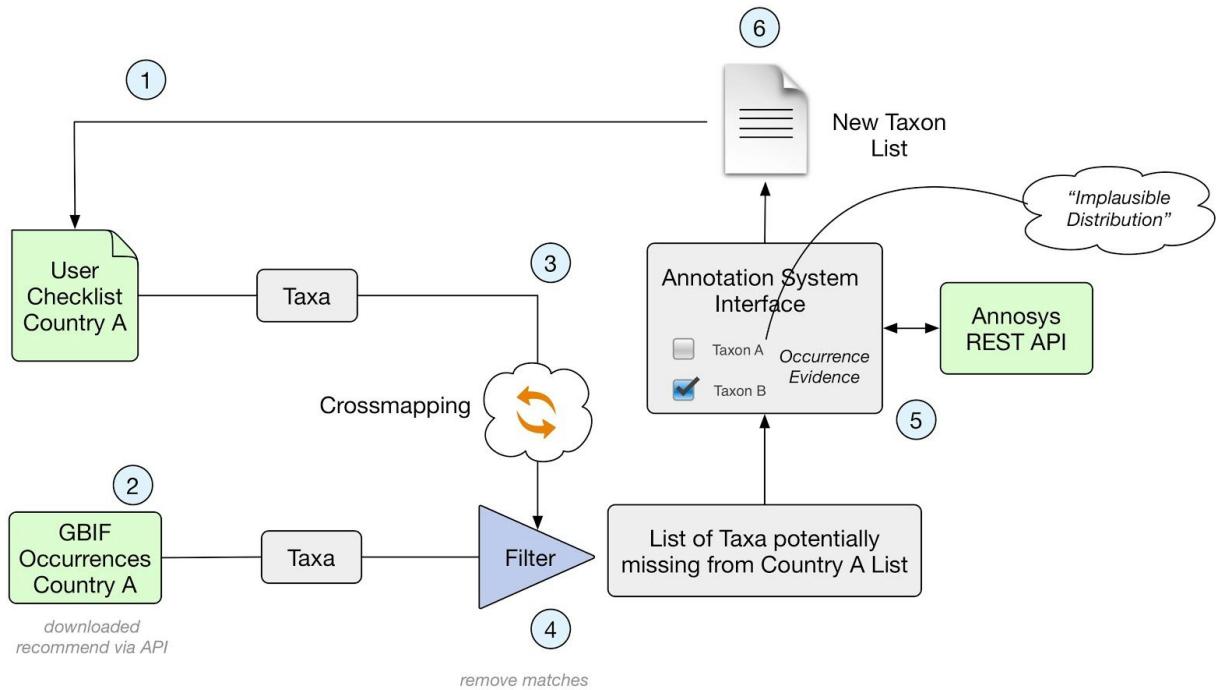A workflow with five steps was designed to investigate this (Fig. 2).



Figure 2. *Workflow describing the steps needed to enable comparison between a national checklist and the taxa represented by GBIF occurrence data.*

The following methods were used for each step in the workflow:

1.  The Dutch national checklist of plants and animals [9] was used as country checklist of plants and animals of the Netherlands. The list was provided as an MS Excel file for testing purposes and can be found in the group 2 repository. For practical reasons the data was limited to only the accepted names from the checklist (n = 13 800).

2.  Occurrence data where DwC:Country = NL was downloaded manually from GBIF (download result available from doi:10.15468/dl.4edn22). To retrieve the list of taxa represented by occurrence data, the team used an SQL distinct selection on speciesKey, scientificName, genus; specificEpithet, infraspecificEpithet. This list was stored in the database (n = 43 387).

3.  This step was not been carried out in the hackathon, but could be cross-matching the checklist with the Catalogue of Life

4.  Filtering out the negative matches (the taxa missing from the original national list)
    A table was created where all taxa represented by GBIF data was inserted. This table included the following columns: taxonKey, AnnosSysUri, scientificName, blacklisted (bool), taxonStatusGBIF, existsInChecklist, checklistStatus (native, introduced etc),

occurrenceRecordCount. The fields existsInChecklist, checklistStatus were updated from the Checklist table. Extraction of the potentially missing taxa was then made by selecting which taxa in the table existsInChecklist is false.

5. Annotate the missing taxa.

   The team used AnnoSys [10] to store annotations to the taxon occurrence records. AnnoSys was originally intended to annotate biodiversity occurrence records in ABCD; an XML format. General techical documentation and documentation of the open annotation model as used by AnnoSys can be found at http://wiki.bgbm.org/annosys/index.php?title=TechnicalDocumentation To be able to annotate a taxon and not a taxon occurrence, AnnoSys was extended during the hackathon. A web interface was created in PHP to allow users to create annotations using the extended AnnoSys service.

6. The gathered evidence could be used to filter out legitimate missing species candidates for addition to the national list.

The cumulative efforts of Teams 1 and 2 established the means to compare taxa listed in a national species checklist and the aggregate list of species assigned to the same country within the GBIF index of occurrence records. Recognizing that this latter index may contain evidence for species that should be in a national list, as well as the inevitable errors, an interface was developed that allow a reviewer to provide review and annotate the GBIF records. The objective of Team 3 was to extend this process by gathering evidence for the occurrence of a candidate species within a given country (Fig. 1). Other data sources provide this evidence through various service interfaces. Sources include:

- Catalogue of Life: CoL provides information on country level species occurrence as asserted by the custodians of the taxonomic data sectors. A full review of these data was conducted and the results are available [3].
- Pan-European Species-directories Infrastructure: Provides verified occurrence information for European countries [11].
- GBIF Checklist Bank - the GBIF checklist index provides documented species checklists that may contain national occurrence information [5, 11].

- GBIF occurrence data  [12] - the GBIF occurrence data that is used to provide the summary list of species contains additional data that may provide supporting evidence for determining the status of a candidate species. This includes the following Darwin Core [13] data properties:

  * Basis of record: Observation, MaterialSample/Specimen, Living

  * Establishment means: "native", "introduced", "naturalised", "invasive", "managed"

  * Reported occurring in respective country the latest 10 years, 50 years, 100 years
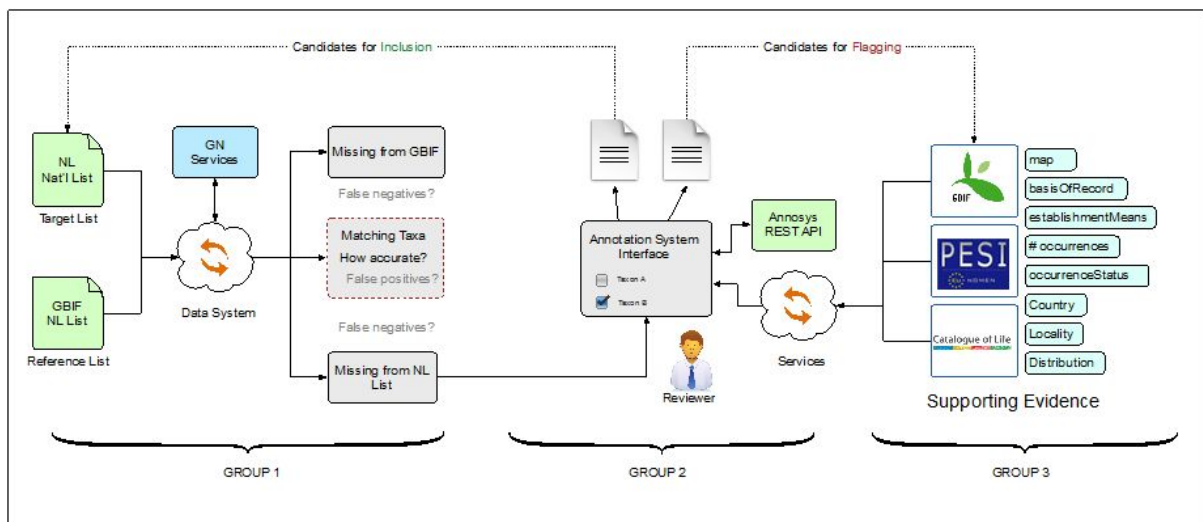


**Fig. 1. The relationship between the 3 teams with the efforts of Team 3 providing evidence to validate national occurrence data mobilized through the GBIF network.**

User stories that formed the basis of work for the Distributions team:

User Story 3-1

I am the manager of a national species checklist and I am looking to verify the occurrence of new species candidates for my list.

User Story 3-2

I am the manager of a national species checklist and I am looking to verify the occurrence of existing species candidates in my list.

**Goals**

1. Analyze the existing Catalogue of Life and EU-Nomen as potential sources for authoritative assertions for regional/national occurrence for species.

2. Analyze the GBIF-mobilized data as a source for documented species candidates for a national checklist.

The following steps for providing supporting national distribution evidence of GBIF species occurrence data were carried out:

1. Obtain list of taxon names for target country (Output of Team 1: the National checklist for the Netherlands was used).

2. Retrieve taxon keys from respective working checklists -

   a. For the source national lists either via the source data or through API from checklist source.

   b. Retrieve GBIF taxon key via GBIF API

      i. http://api.gbif.org/v1/species/match?name=[taxon_name]

   c. Retrieve global checklist taxon keys for the Catalogue of Life

   d. Retrieve global checklist taxon keys for EU-Nomen

3. Retrieve country-level distribution information from the source checklists through their respective API's via taxon key identifiers

4. Retrieve occurrence counts from raw occurrence data in GBIF

   a. Example "Abax ovalis (Duftschmid, 1812)", taxonKey=5754770

   b. http://api.gbif.org/v1/occurrence/search?taxonKey=5754770 => count=588

   c. http://api.gbif.org/v1/occurrence/search?taxonKey=5754770&country=NL => count=7

5. Build a front-end to present the summarized evidence.

# Results

Team 1 compared their target national checklists against CoL with their created cross-mapping process using GN services (Fig. 1). The GN services use advanced name-parsing and matching tools to compare taxon names and is able to identify matching names that may vary slightly in their form, spelling of authorship etc. The output of this process were three sets of taxa:

1. Matching taxa are those contained in both lists

2. Taxa missing from the national list but contained in the CoL. Given the global scope of the CoL, we expect most CoL to not occur within a given country. Access to distribution data within the CoL (the work of Team 3) would help identify relevant CoL missing from the national list.

3. Taxa (or possible synonyms) missing from the CoL. This list could have potential value as candidate additions to the CoL.

When the created prototype system is run, the cross-mapping tool analyzes the two checklists and produces a tabular output showing the results of the cross-mapping. A name-parsing tool splits each scientific name into its constituent components (genus, species, authorship, year, rank, etc.) using Backus Naur grammar to establish rules regarding the composition of names. This allows the component parts to be identified for comparison using various "fuzzy" text-matching algorithms to estimate when two similar text strings may be matches. The process ends with a matching estimate that is given a match type and a confidence score.

For example, two entries from the compared lists (Table 2) form the following match record with a score of 0.75, indicated an exact match of the canonical forms of the name but non-matching (but similar) author names.

**Table 2. Match score example.**

| Name 1 | Name 2 | Score |
|---|---|---|
| Scenedesmus abundans (Kirchner) Chodat | Scenedesmus abundans (O. Kirchner) Chodat | 0.75 |

Matching score options included are shown in Table 3.

**Table 3. Match types and their scores.**

| Match Type | Score |
|---|---|
| Exact Match | 0.999 |
| Partial Canonical Form Exact Match | 0.998 |
| Canonical Form Exact Match, similar authorship | 0.995 |
| Canonical Name Only Exact Match | 0.75 |
| Genus part matches | 0.75 |

The cross-mapping results (Table 4) indicate a relatively high proportion of matches, which would be predicted if we could assume a relatively comprehensive Catalogue of Life and accurate matching capabilities of the cross-mapping tools.

**Table 4. Summary of the gross cross-mapping results with links to the full results.**

| List | Matches | % matched | unmatched species/variety/subspecies |
|------|---------|-----------|----------------------------|
| Portuguese Bryophyta | 726/**745** | 97.4% | 19/0/0 |
| Portuguese Tracheophyta | 4458/**4618** | 96.5% | 121/26/11 |
| Slovakian Checklist | 1248/**1279** | 97.6% | 26/5/0 |
| Norwegian Checklist | 10950/**11393** | 96.1% | 163/2/7 |

After the hackathon, the code was further developed into a GN cross-matching tool, which code is available here:

https://github.com/GlobalNamesArchitecture/gn_crossmap

The command-line tool allows cross-mapping against any reference list in

http://resolver.globalnames.org/data_sources.

Also a web-interface was created to do the cross-mapping against some predefined reference lists:

http://resolver.globalnames.biodinfo.org/.

After the hackathon, the Norwegian Checklist was cross-mapped against an updated version of CoL (Annual Checklist 2015) using the latest version of the developed tool, results can be found here:

https://www.dropbox.com/s/ztxie0sgt1gdbjf/norvegian_checklist_col_2015.csv?dl=0

Team 1 also created an overview of existing name-matching tools and tools for comparing hierarchies, which can be found in their report.

Team 2 extended an Annotation class of AnnoSys so it could handle information about a taxon (as opposed to a taxon occurrence).  The team further developed a new annotation model, which is also based on the W3C Open Annotation Data Model [14]. The purpose of the annotation in this case was to

express that the distribution for the taxon might, or might not, be correct. In order to express the latter, a interim RDF term (http://rs.gbif.org/terms/ao#ImplausibleDistribution) was introduced.

The validation information is then posted into AnnoSys using its REST API. The team linked the annotation to the URL representing the taxon page of GBIF i.e. http://www.gbif.org/species/taxonKey (Fig. 3). The annotation was expressed in a way that should be interpreted as: for the taxon with the taxonKey, all occurrences reported for the specified Country where the establishmentMeans do not clearly indicate non natural occurrence, should be considered as being expected errors.
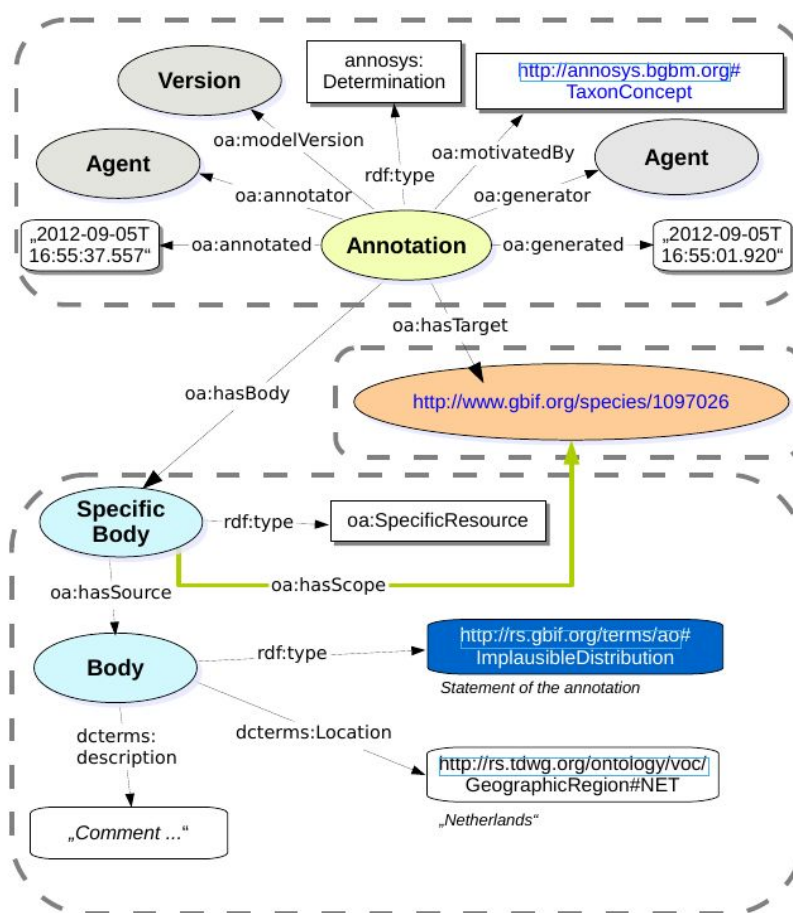


**Figure 3.  Example PUT request to the AnnoSys to create a new taxon annotation.**

The resulting web interface is shown in figure 4. It supports annotation of all listed taxa which according to GBIF occurrence data is likely to candidates to add on to the existing national checklist. For each row a taxon-specific link to the distribution map generated by GBIF is provided in order to help the user evaluate the underlying data and judge whether or not the taxon is likely to exist in the country.
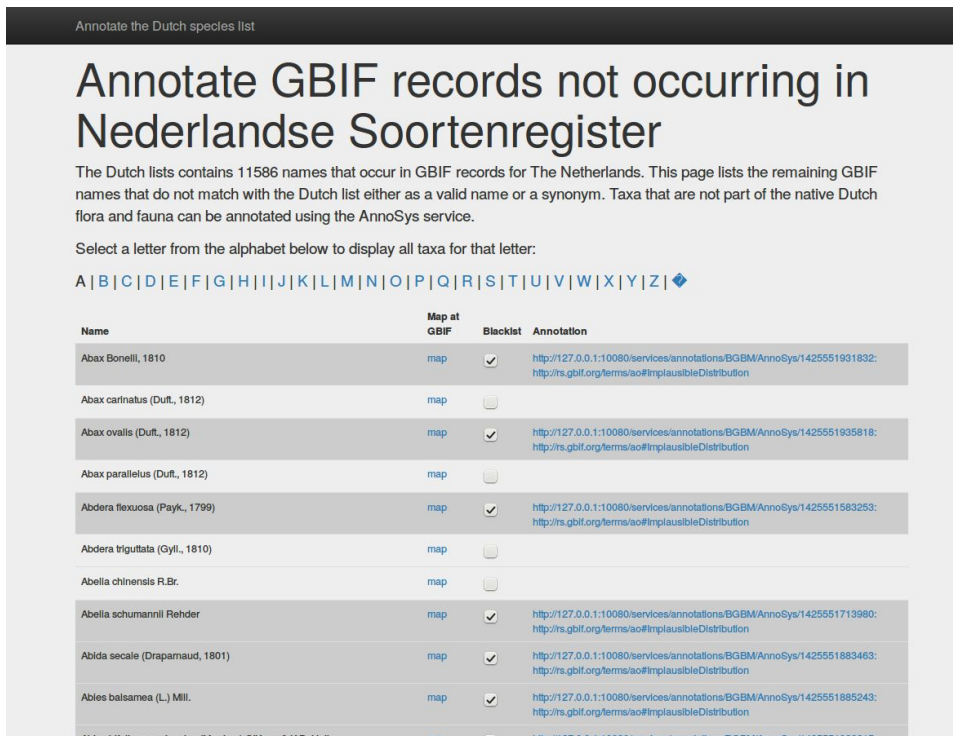
**Figure 4. Screen shot of the Annotation web interface with the list taxa potentially missing in the checklist of the Netherlands.**

To annotate a taxon the user has to use the checkbox. When checked, a taxon-specific annotation is posted at the AnnoSys repository. When this is done the application posts the response with the URL to the new annotation in the system database table.

Team 3 investigated Catalogue of Life and EU-Nomen as potential sources for authoritative assertions for regional/national occurrence for species.  The results were:

- Given a taxon identifier, CoL was checked for national distribution data in a consistent and standardized format. Current CoL distribution data is inconsistent across taxonomic sectors and heterogeneous in format. A detailed analysis of CoL distribution data was conducted during the hackathon. See the resulting report, Distribution Data in the Catalogue of Life and the associated summary of distribution data across Global Species datasets.

- Access to PESI, via LSID, appears to be inconsistent across source datasets. To be precise: ERMS, Fauna Europaea and the Index Fungorum provide an LSID and the Euro+Med Plantbase does not, it provides a GUID instead. This makes it difficult to use as source.

GBIF-mobilized data was also analyzed as a source for documented species candidates for a national checklist. The team created a web interface where a checklist can be uploaded and distributions can be matched against GBIF data: http://134.213.149.111/group_3/hackathon_group3/web/index.php, see also the screenshot (Fig. 5). The GBIF API was implemented to collect the data.



**Figure 5. Web Interface showing the results of national distribution evidence from GBIF species occurrence data.**

The results show that evidence can be collected  from GBIF data, which can be used by an expert to assess whether a species should be added to a national checklist or not. The team also found however that some essential data like Basis of Record (BoR) is not supplied by most GBIF providers. The team

also identified several services which are currently not present in the GBIF API that, when present, could make this task of collecting evidence easier and more straightforward.

# Discussion

Sources of biodiversity occurrence data, such as catalogued and indexed by GBIF, may serve as a means to both verify or extend the list of taxa found in national species checklists [15]. They might also serve as the means to start a de-novo national species list [16]

The primary goal of the matching exercise by Team 1 focused on assessing the capability of the name-matching system to accurately match names so that the curated CoL taxa can be used, via reference, as the core concepts in one or more national lists, hence promoting interoperability. A detailed look at the matching results fom team 1 (Table 4) reveals some degree of ambiguity in the relevance of the matches. Relevance in matching can be divided into two measures: precision and recall. The cross-mapping tool provides very good recall (97%), but the precision (the proportion of false positive matches to true positive) is a measure that should be more thoroughly explored.

The confidence score provides a starting point in assessing precision but a thorough analysis of these scores was not performed during the workshop. A general review would indicate that high confidence scores over 0.995 were likely to be positive matches. Scores less than this could be either true or false positives. The example in Table 5 illustrates a positive match with a lower confidence due to the variation in the authorship, which a domain expert would identify as referring to the same authorship instance.

**Table 5. Positive match example.**

| Name 1 | Name 2 | Score |
|--------|--------|-------|
| Anacolia webbii (Mont.) Schimp. | Anacolia webbii W. P. Schimper, 1876 | 0.75 |

On the other hand, an equivalent score is given to the example in Table 6. This example is clearly not an exact canonical match but a more general match between a species and its parent genus.

**Table 6. Example of a general match between a species and its parent species.**

| Name 1 | Name 2 | Score |
|--------|--------|-------|
| Campylium stellatum (Hedw.) Lange & C.E.O.Jensen | Campylium | 0.75 |

Another example of an ambiguous match is related to what appears to be non-standardized use of authorship when citing a scientific name (Table 7).

**Table 7. Example of a non-standardized use of authorship.**

| Name 1 | Name 2 | Score |
|--------|--------|-------|
| Cratoneuron filicinum (Hedw.) Spruce | Cratoneuron filicinum (Fiorini-Mazzanti) Latzel, 1931 | 0.75 |

These names could refer to homonyms or to one or more *chresonyms*, where the authorship represents a citation of a particular publication that is not one of the code-regulated citations used to generate a code-compliant nomenclatural act. In these cases, the ambiguity is entirely related to the content and cannot be evaluated by parser rules alone.

Team 2 attempted to annotate GBIF records with AnnoSys in such a way that for a taxon with a GBIF taxon URL, the occurrences can be flagged for a certain country to be errors (unless indicated as non natural occurrence). This can be seen as a kind of 'blacklisting' by an expert, so other users do not need to investigate these records again. The team had to alter Annosys for this, as it was only supporting annotations of XML records (ABCD records), not URLs and proved that it could be done with a working system. The implemented system is fairly limited however, and does not allow to undo an annotation. To

be used as a GBIF service for real usage, it should be further developed. The team came up with a new User Story that should be further explored:

As a curator of GBIF data and harvest processes, I want to be able to look up annotations indicating that imports of occurrence data with a specified combination Country and Taxon may be incorrect in order to handle these records in an appropriate way (e.g., records of lions occurring within the Netherlands and not indicating zoo specimens). They may lead to more accurate assessment of species composition within countries.

In order to make it more feasible to implement a system supporting this User Story at least two major changes are needed in existing web services:

1. GBIF occurrence service

   A method that can utilize the same set of input parameters for querying occurrence data as the existing that delivers the result as a list of represented (observed) taxa with all relevant DwC terms plus the various GBIF taxonKeys (identifiers for taxa in the GBIF Backbone Taxonomy). Relevant could be adding a column for number of occurrence records per taxon listed.

2. AnnoSys API

   Adding functionality that supports a more generic solution for annotations of information.

Team 3 found that it takes two API calls to obtain national distribution information for a taxon.

1. Input a species name to obtain an identifier that can be used to obtain distributions (referred to as a usageKey)

   a. Example "Abax ovalis (Duftschmid, 1812)":

   b. http://api.gbif.org/v1/species/match?name=Abax%20ovalis%20%28Duftschmid,%201812%29

   c. => usageKey=5754770 (aka GBIF taxonKey)

2. Utilize the usage key to obtain a list of distributions.

   a. http://api.gbif.org/v1/species/5754770/distributions => list of "localities" (aka countries)

Current output only shows the source taxon key. It would be useful to include a source dataset key with the distribution data. It would also be useful to add a datasetKey as a search parameter e.g., http://api.gbif.org/v1/species/5754770/distributions?datasetKey=90d9e8a6-0ce1-472d-b682-3451095dbc5a. The GBIF API does not provide a service to add search parameter country (as interpreted by GBIF) and to get a TRUE or FALSE response only:

> http://api.gbif.org/v1/species/5754770/distributions?country=NL => T/F

This would make retrieving a dataset for a country much easier.


Given a national checklist with taxa occurring in a country, it would be worthwile to check the GBIF occurrences if there are additional taxa observed as occurring in the given country. For this operation we would propose an inventory service from the GBIF occurrence API:

http://www.gbif.org/developer/occurrence#inventories For instance a  request

http://api.gbif.org/v1/occurrence/counts/species/?country=NL to return a list of taxonKey(s) with counts.

# Acknowledgments

# References

1.  Mace GM. 2004. "The Role of Taxonomy in Species Conservation." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359 (1444): 711–19.

2.  Lughadha EN. 2004. "Towards a Working List of All Known Plant Species." *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 359 (1444): 681–87.

3.  Roskov Y, Abucay L, Orrell T, Nicolson D, Kunze T, Culham A, et al. 2015. "Species 2000 & ITIS Catalogue of Life, 2015 Annual Checklist."

4.  Edwards JL. 2000. "Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop." *Science* 289 (5488): 2312–14.

5.  Vaidya G. 2013. "Validating Scientific Names with the GBIF Checklist Bank." *VertNet*. July 22. http://blog.vertnet.org/post/56169017224/taxonomic-validation-vaidya.

6.  GBIF Secretariat. 2015. "Capacity Enhancement Support Programme." *Global Biodiversity Information Facility*. Accessed July 30. http://www.gbif.org/capacity-enhancement/support-programme.

7.  White R. 2013. "Cross Mapping Tools." *i4Life*. November 27. http://www.i4life.eu/i4lifewebsite/cross-mapping-tools/.

8.  Tschöpe O, Suhrbier L, Güntsch A, Berendsohn W. 2013. "AnnoSys, a Generic Annotation System for Biodiversity Data." *Global Biodiversity Information Facility*. http://www.gbif.org/resource/81487.

9.  Naturalis Biodiversity Center. 2015. "The Dutch National Species Register." *Nederlands Soortenregister*. Accessed August 28. http://www.nederlandsesoorten.nl.

10. Tschöpe O, Macklin JA, Morris RA, Suhrbier L, Berendsohn WG. 2013. "Annotating Biodiversity Data via the Internet." *Taxon* 62 (6): 1248–58.

11. De Jong Y. 2010. "A Pan-European Species- Directories Infrastructure (PESI)." In *Tools for Identifying Biodiversity: Progress and Problems*, edited by Nimis PL and Vignes Lebbe R, 49–51.

12. Chapman AD. 2005. *Uses of Primary Species-Occurrence Data, Version 1.0. Report for the*

*Global Biodiversity Information Facility.* Edited by Global Biodiversity Information Facility. Vol. 1.0. Copenhagen: Global Biodiversity Information Facility.

13. Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Renato Giovanni, et al. 2012. "Darwin Core: An Evolving Community-Developed Biodiversity Data Standard." *PloS One* 7 (1): e29715.

14. Sanderson, Robert, Paolo Ciccarese, Herbert Van de Sompel. 2013. *Open Annotation Draft Data Model*.

15. Hamer, M, J Victor, G F Smith. 2012. *Best Practice Guide for Compiling, Maintaining, Disseminating National Species Checklists*. Edited by David Remsen. Copenhagen: Global Biodiversity Information Faclity.

16. Broad, Gavin, and Broad Gavin. 2014. "Checklist of British and Irish Hymenoptera - Introduction." *Biodiversity Data Journal* 2: e1113.