ARPHA Preprints

PENSOFT

Methods

Author-formatted document posted on 03/09/2026

Published in a RIO article collection by decision of the collection editors.

DOI: https://doi.org/10.3897/arphapreprints.e214296

Functional workflow for mapping local taxonomies established and documented

Anton Güntsch, **Walter G. Berendsohn,** **David Fichtmüller**

ARPHA Preprints *Author-formatted document posted on 03/09/2026.* DOI: https://doi.org/10.3897/arphapreprints.e214296

Ref. Ares(2025)6252031 - 31/07/2025
Funded by the European Union

TETTRIs

TETTRIs

Transforming European Taxonomy through Training, Research, and Innovations

Functional workflow for mapping local taxonomies established and documented

WP 2/Anton Güntsch

Authors: Walter G. Berendsohn, David Fichtmüller & Anton Güntsch

Version: 0.6
Date: 27/07/2025

TETTRIs related product

Identification	Value
Title	Functional workflow for mapping local taxonomies established and documented
Author(s)	Walter G. Berendsohn, David Fichtmüller, Anton Güntsch
Affiliation	FUBerlin
Contributors (Affiliation)	COL, FUB BGBM, Meise BG, NHMD
Publisher	CETAF
Identifier of the publisher	0058ctf29
Doc. Version	0.6
Resource	
Publication year	2025
Sensitivity	PU - Public
Date	27 July 2025
Citation	

Abstract:

Our analysis of the landscape of local taxonomies and infrastructures providing aggregated taxonomic information has identified the matching process for scientific organism names as the core element of linking taxonomic resources. Name matching is considered the primary requirement for integrating datasets containing scientific names of organisms across the Internet, and for enabling science-based semantic linking of organism-related data. We also found wide variation in user requirements as well as in the offered functionalities for this process. As a consequence, a single, universally applicable workflow does not exist, but rather a number of approaches that depend on the specific user interest. This document presents a detailed concept and a demonstrator implementation of an interactive user guide for name matching of scientific organism names, as well as background information and general recommendations. The goal is to enable local users to verify and link their usage of names to other such resources, particularly to aggregated taxonomic datasets. Local users that want to apply name matching include everybody from individual researchers with their spreadsheets (such as those used in ecological studies) to curators of national or regional taxonomic databases. The technical implementation demonstrated includes a user interface for formulating queries as well as the underlying database that holds information on publicly available taxonomic datasets, name matching services, and the data and functionality provided by their combination. The focus is on functionality accessible by simple means, but information on available technical resources (APIs, R packages, Reconciliation Interfaces) is also included.

Keywords:

organism names, name matching, data quality control, biological taxonomy, taxonomic checklist, data integration, data linking, semantic web

Funded by the European Union

Revision:

no.	Reviewer	Status	Notes
v01	Anton Güntsch	incorporated	Approval of over-all design.
v02	Andreas Müller, Anton Güntsch, Markus Döring, Quentin Groom, Thomas Pape	review incorporated	Internal review WP2
v03	Walter Berendsohn, Robert Lücking, Wolf-Henning Kusber, Roderic Page	review incorporated	Major restructuring (WB), overhaul of tables. Consulted RL regarding mycological (fungi), WHK regarding phycological (algae), and RP regarding zoological services, respectively
v04	David Fichtmüller, Walter Berendsohn	incorporated	Adding and revising implementation section, minor edits (DF); adding sections on R-packages, OpenRefine and APIs, adding further data in tables, minor edits (WB)
v05	Sofie Meeuse, Peter Warth	incorporated	Various comments on terminology, actual workflow, references, etc.
v06	Walter Berendsohn, David Fichtmüller	incorporated	Addition of the functional prototype (DF), revision of entire text with respect to terms used (added definitions), clarification of workflow, additional references, formatting (WB)
V06	Marta Leon Monedero	Approved	Approved

Funded by the European Union

Document Control Information:

Settings	Value
Document Title:	Functional workflow for mapping local taxonomies established and documented
Project Title (Acronym):	TETTRIs
Document Authors:	Walter G. Berendsohn, David Fichtmüller, Anton Güntsch
Project Coordinator	Technical Coordinator: Ana Casino, CETAF Financial Coordinator: Frederik Hendrickx , RBINS
Doc. Version:	
Sensitivity:	PU - Public
Date:	31/07/2025

Document Approver(s) and Reviewer(s):

NOTE: All Approvers are required. Records of each approver must be maintained. All Reviewers in the list are considered required unless explicitly listed as Optional.

Name	Role	Institution	Action	Date
Sofie Meus	Reviewer	MeiseBG	Reviewed	23/07/2025
Peter Warth	Reviewer	SMNS	Reviewed	23/07/205
Marta Leon Monedero	Approver	CETAF	Approved	31/07/2025

Document history:

The Document Author is authorised to make the following types of changes to the document without requiring that the document be re-approved:

- Editorial, formatting, and spelling
- Clarification

To request a change to this document, contact the Document Author or Owner. Changes to this document are summarised in the following table in reverse chronological order (latest version first).

Revision	Date	Created by	Short Description of Changes

Configuration Management: Document Location

The latest version of this controlled document is stored in
https://docs.google.com/document/d/1KmvKXRkgcCGsO3z9zERv0cKczNJ0XqvX

Funded by the European Union

TETTRIs

INDEX

ACRONYMS **7**
DEFINITIONS **7**
INTRODUCTION **8**
Background 8
Motivation for matching and linking efforts on the local side 8
Characterising the audience 8
Motivation for name matching 9
Taxonomic aggregators 9
Inherent caveats of a workflow 11
What is a “functional workflow for mapping local taxonomies”? 12
CHECKING SCIENTIFIC NAMES OF ORGANISMS - RECOMMENDATIONS **12**
Introduction 12
Scope 13
General recommendations for specific taxonomic groups 13
Checking the names of animals 13
Checking the names of plants 13
Checking the names of fungi 14
Checking the names of eukaryotic algae 14
Checking the names of Cyanobacteria (blue-green algae) 14
Checking names of Virus and Viroids 14
OpenRefine reconciliation interfaces 15
Programming tools for name matching 16
R-packages 16
Application programming interfaces (APIs) 16
Others 16
AN INTERACTIVE USER GUIDE FOR NAME MATCHING **16**
User input for identifying datasets and corresponding name matching services 17
Organism group 17
Geographic scope 17
Number of names to be checked simultaneously 17
Purpose of name matching - I want to check 17
Data linkage requirements 17
Technical requirement - I want to 18
Implementation 18
User interface 18
Caveats - work in progress 19
Underlying data 19
Descriptors for taxonomic datasets 20
Descriptors for name matching services 21
Dataset-in-service 22
Storing and updating the data 24
CONCLUSIONS **24**
ACKNOWLEDGEMENTS **25**
REFERENCES CITED **25**
APPENDIX 1: Mapping of user-interface selections to table-attributes **27**
APPENDIX 2: Related documents **30**

Funded by the European Union

ACRONYMS

CETAF	Consortium of European Taxonomic Facilities (General Secretariat)
COL	Catalogue of Life
EU	European Union
FUB-BGBM	Freie Universität Berlin - Botanischer Garten und Botanisches Museum Berlin
IF	Index Fungorum
IPNI	International Plant Name Index
PESI	Pan European Species directory Services (a.k.a. EU-Nomen)
TETTRIs	Transforming European Taxonomy through Training research and Innovations
WCVP	World Checklist of Vascular Plants
WFO	World Flora Online
WP	Work Package
WPL	Work Package Leader

DEFINITIONS

In the context of TETTRIs and this document, we define terms as follows:

Aggregator	See *Taxonomic aggregator*
Aggregator services	Online accessible services (interactive or as endpoints for computer-to-computer communication) provided by *taxonomic aggregators*. Here focussed on name-matching services.
Canonical name	A scientific name without the corresponding author or author-year citation.
Local data	A set of taxonomic names that a user wants to match using a *name matching service*
Local user	The person or team using a *name matching service*
Mapping taxonomies	The process of matching *local data* with *taxonomic datasets* and the resulting coherence of scientific names or their identifiers that enables or eases linking of taxonomic data.
Name-matching service	An *aggregator service* that allows to match name strings from *local data* to one or more *taxonomic datasets*.
Nomenclator	A *taxonomic dataset* that is restricted to information about the scientific name itself, e.g. its correct spelling or validity according to rules fixed in the international codes of nomenclature.
Taxonomic aggregator	Online accessible databases that compile taxonomic information from multiple sources into a *taxonomic dataset* or host *taxonomic datasets* from multiple sources.
Taxonomic dataset	Structured lists of scientific names that follow a single classification system. With the exception of *nomenclators*, they typically present a hierarchical, tree-like taxonomy in which each taxon represents a node. Each scientific name is either assigned as the accepted name of a taxon or treated as a synonym (except in cases where the name exists but cannot currently be resolved). This term largely corresponds to the term Taxonomic Database used by Grenié & al. (2022).
Taxonomic group	A broad term to describe a group of taxa or taxonomic ranks at which people work (Grenié & al. 2022), e.g. plants, fungi, animals.

Funded by the European Union

INTRODUCTION

Background

The central aim of the TETTRIs project is "to produce a transformative change in the role played by taxonomy in tackling biodiversity challenges by boosting taxonomic capacity and transferring critical scientific knowledge to societal actors". Within this broader context, the primary goal of Work Package 2 is the "integration of taxonomic resources to enable cross-domain data linkage", which is a central enabler for knowledge transfer.

Data linkage (or record linkage, Harron 2016) is "the aggregation of data from different sources concerning the same entity or individual." (Vaccaro & Swerts 2022).

In the context of TETTRIs, this entity is the taxon: a named group of organisms sharing defined characteristics and distinguishable from other groups. Taxa are organized hierarchically based on presumed evolutionary relationships, with species being the most relevant rank (hierarchical level) in applied contexts.

Work Package 2 focuses on improving linkages between diverse data sources that contain information related to taxa. These taxa are typically identified by scientific names, though molecular sequences are increasingly used as proxies. Traditionally, information is linked by using the taxon name as a common identifier. A practical approach to achieving this linkage for digital data is to map locally used taxon names (“local data”) to global or European taxonomic aggregators. Here, "local data" encompasses everything from individual researchers’ spreadsheets - such as those used in ecological studies - to national or regional taxonomic databases.

Aggregating taxonomic data into unified datasets facilitates knowledge transfer and enables new research opportunities (see, eg., the usage of data aggregated by the Global Biodiversity Information Facility - GBIF 2024). However, there is an inherent problem: the resulting taxonomic dataset may not be directly useful to the original providers, but the quality and usability of the aggregated data largely depends on the preparation of data at the source level. In many cases, demands imposed by research funders, governmental directives or publishers do already enforce a certain level of shareability of produced research data. Moreover, research programs may include parts that directly assist researchers to prepare their data for data sharing and provide the technical means for linkage. However, motivating data creators to produce results according to FAIR principles should be part of the scientific workflow.

To document effective and sustainable workflows we have to look first at user requirements and interest on the local side, then to document technical means to establish data linkage with the help of taxonomic aggregators, and finally, we will present the concept and a demonstrator application for an updatable and interactive guide to solutions adapted to individual use cases.

Motivation for matching and linking efforts on the local side

Name matching (Hyam 2024) requires effort on the user’s side. For those not primarily focused on taxonomic information, additional motivation may be needed to implement it. TETTRIs can play a role in lowering barriers and facilitating this process.

Characterising the audience

Categorising taxonomic or taxonomy-related data has served to work out specific aspects of the potential relationships between users and services. Criteria to categorize local usage of taxonomic names (“local data”) in datasets may include

- Purpose, e.g.
  - Aggregation of taxonomic knowledge

Funded by the European Union

    - Support for species conservation
    - Usage properties (medicinal, wood, …)
    - Governance and law (EU-Directives, CITES, …)
- Geographic scope
    - Local
    - State
    - National
    - Regional
    - Global
- Taxonomic scope
    - Revision-scale (Genus, small family)
    - Large taxonomic group (Plants, Butterflies, Mammals, Birds, …)
    - All organisms
- Taxonomic in-house expertise on the user's side (related to the taxonomic scope)
    - Rather little: Usage of taxon names in factual datasets, e.g. for ecological research; aggregators of such research results;
    - Rather much: Taxonomic revision with checklist; Taxonomic Expert Network as aggregator; Community-driven aggregator (e.g. in Botany s.l.: WFO; Euro+Med, AlgaeBase; Mycobank/Index Fungorum)
- Community scope
    - Broadly anchored and used dataset/database
    - …
    - individual contribution.

Many datasets including very small ones are hosted in ChecklistBank. The metadata for datasets there (see Data quality requirements: Checklist datasets) do not cover the above listed aspects in detail, but the 3 first bullets can be covered by the Description text field. The same is true for the dataset description in Global Names (where the focus is on large datasets; see https://verifier.globalnames.org/ under *Advanced options*, the *"i"* links under *Pick data sources*).

        - Motivation for name matching

Several use cases for name matching have been documented on the TETTRIs WP2 Wiki: https://wiki.bgbm.org/tettris/Use_cases_for_name_matching. Researchers' and data curators' motivations for applying name matching can be broadly categorized into two types: intrinsic scientific motivation and external or mandated motivation. Intrinsic scientific motivation is driven by the pursuit of high-quality scientific data. This includes ensuring correct orthography of scientific names, accurate citation of a taxon's classification (e.g., its assignment to a family), and, where taxa are the object of study, efficient retrieval and integration of all relevant information, including data associated with related taxon names. External motivation, on the other hand, arises from institutional, legal, or publishing requirements. Examples include compliance with conditions set by funding bodies (e.g., LTER research requirements in Germany), journal submission standards (e.g., Pensoft's requirement for IPNI linkage), or regulatory frameworks (e.g., the European Union's Flora Fauna Habitat Directive).

Taxonomic aggregators

In the context of TETTRIs WP2, we define *taxonomic aggregators* as online accessible databases that compile or host taxonomic information from multiple sources. Our focus is on taxonomic aggregators that offer name-matching services; however, we distinguish between the taxonomic datasets used by these services and the services themselves (see below). We have identified and documented relevant aggregator services and taxonomic datasets by means of web-searches and follow-up of references found in the literature cited below (e.g. Grenié & al. 2022, Islam & al. 2024).

Funded by the European Union

Taxonomic datasets (datasets in the following) are structured lists of scientific names that follow a single classification system. They typically present a hierarchical, tree-like taxonomy in which each taxon represents a node. Each scientific name is either assigned as the accepted name of a taxon or treated as a synonym (except in cases where the name exists but cannot currently be resolved). A special class of datasets are "nomenclators" that systematically catalog scientific names along with their authorship, publication date and references, nomenclatural status, and (sometimes) type information. These focus on nomenclatural accuracy such as on correct spelling and nomenclatural validity rather than providing taxonomic opinion or classification.

Such datasets aim to provide comprehensive coverage for particular organism groups, either globally or regionally (e.g., at a European scale). The *Catalogue of Life* monthly and annual editions are currently the only datasets attempting to unite all biota in a single global dataset. At the regional level, *PESI/EU-Nomen* (de Jong & al. 2015) serves a similar purpose for European taxa.

The sources of the data may include scientific literature, smaller taxonomic aggregators, data from publications, or, in some cases, even original unpublished data. An important consideration with respect to taxonomic datasets is in how far they are driven by and (also) used by the respective taxonomic community.

Taxonomic aggregators may be based on a single dataset or serve as repositories for multiple independent datasets. In the latter case, aggregator services can be configured to operate across all or selected datasets. When these services are restricted to a single dataset, users can generally expect a taxonomically coherent outcome.

Of course, the border between local data and the target aggregator taxonomic dataset may depend on the context. E.g,. the Euro+Med Plantbase team as a local user has extensively used World Flora Online services for name checking in the context of TETTRIs Task 2.1 *(Functional optimised EU Nomen processes,* Müller & Güntsch 2025*)*. At the same time, the Euro+Med taxonomic dataset forms part of the PESI-EU-Nomen aggregator service.

Aggregator services vary in scope and may cover diverse types of biological data. Within TETTRIs WP2, our focus is specifically on name-matching mechanisms - see https://wiki.bgbm.org/tettris/What_is_name_matching%3F for a discussion of the name matching process and https://wiki.bgbm.org/tettris/Existing_name_checking_mechanisms for some details of services. The TETTRIs satellite project Taxonomic Name Linking Service - TNLS has compiled a list of the input parameters and output fields of a number of aggregator services - see **project deliverable "TNLS: Common name matching specification".

In addition to identifying exact matches between the input name and entries in the selected taxonomic dataset (matching in the strict sense, Hyam 2024), a key functionality of name matching services is their ability to suggest potential matches for names that cannot be matched exactly. Various strategies have been developed to address this challenge. For instance, some services limit their comparison to the canonical form of a name, disregarding author citations or publication years. Another common strategy is the use of fuzzy matching algorithms, which identify similar names based on minor character differences. Grenié et al. (2022) provide an overview of fuzzy matching techniques and the associated risks.

Local users face a fragmented and sometimes confusing landscape of multiple and partially overlapping datasets and name-matching services (Grenié & al. 2022). Services differ in their features and may even give different results when applied to the same dataset. TETTRIs WP2 has investigated a number of use cases (see https://wiki.bgbm.org/tettris/Use_cases_for_name_matching), also taken from literature (Grenié & al. 2022, Islam & al. 2024), to further investigate the problems faced by local users. It became clear that an instrument linking the diverse user requirements with the features provided by datasets and aggregator services may alleviate the situation. To address this, TETTRIs has defined a list of properties for the user's requirements and created data tables

Funded by the European Union

that describe taxonomic datasets and name matching services in relation to these requirements. Building on this, tools can be built to help users in selecting datasets and services that best meet their needs. A demonstrator application is presented in this report.

Additionally, we have compiled a wish list of desired properties based on use case studies, which may help to improve name-matching services: https://wiki.bgbm.org/tettris/Wish_list_for_name_matching_services.

○ Inherent caveats of a workflow

Using a scientific name as the identifier of a taxon is a pragmatic approach to cross-domain data linkage. They provide a standardised way to connect the information that pertains to, e.g., a certain species. So data from many different sources can be brought together for a certain species, for example, information on its uses (e.g. medicinal, cultural, construction), its morphology (appearance and form), its geographical distribution, its conservation status (e.g. threatened by extinction), and threats it poses (poison, invasive species, pests). Such data have been collected by humans for millennia, and put into persistent records for centuries, since the mid-17 hundreds connected to scientific names following internationally agreed rules.

However, the approach is beset with a number of caveats. First of all a potential (and frequently encountered) quality problem of the data themselves: the organism in question may have been misidentified, i.e. labelled with a name that represents a different taxon. But apart from that, scientific names present inherent caveats that may influence the quality of data linkages:

The international rules of nomenclature that govern scientific names may demand that the name of a taxon changes. These (nomenclatural) name changes may be handled in an automatic way if the aggregator has the necessary information (for example, the knowledge that two names are based on the same nomenclatural type).

A far more complex issue is that scientific names are essentially pointers to a concept of a taxon, e.g. of a species. These concepts are scientific hypotheses about the descent of the species, and with new scientific research on the evolution of species, its circumscription may change. This means that taxa that have been considered separate in the past may turn out to be a single taxon, thus some of the names become synonyms. And vice versa: careful investigation may reveal that what has been considered a single taxon in the past in reality represents more than one.

On a global level, every synonym should refer to the accepted name of a taxon. However, since (following the rules of nomenclature) names are not necessarily changed when the circumscription of the taxon changes, in some cases taxonomic expertise is needed to decide if data provided under a specific taxon name from different sources do really refer to the same currently accepted taxon. To further complicate the matter, the same name in some cases may refer to different accepted taxa, simply because the naming happened under different codes of nomenclature.

There have been attempts to handle this problem by distinguishing taxon concepts and label them with a non-nomenclatural identifier. However, if considered for all or large groups of organisms this implies an enormous amplification of the necessary knowledge base. Thus, at present, such data are restricted to a few model datasets and "*in practice, aggregators have made fundamental design choices* [...] *by excluding taxonomic conflict as an aspect of biodiversity knowledge worthy of being preserved*" (Franz & Sterner 2018).

The TETTRIs satellite project Taxonomic Name Linking Service - TNLS is working on a technical specification for a "Concept subscription service" or "Taxonomic activity notification service" that could be implemented as part of aggregator services. This implies research into the changes in the content of a taxonomic database that influence the taxonomic concept behind a scientific name (see https://wiki.bgbm.org/tettris/Database_content_changes_affecting_the_taxonomic_standing_of_a_scientific_name).

Funded by the European Union

However, a functional workflow for name matching is a prerequisite for using such a service.

What is a “functional workflow for mapping local taxonomies”?

A key finding of our analysis is that the core element of this workflow is the name matching process. However, we also observed significant variation in user requirements regarding “mapping”, i.e. the integration and linkage of local data with taxonomic aggregators. This variability extends to the range of available name matching services and underlying datasets, which may or may not meet specific user needs. As a result, there is no single, universally applicable workflow.

To address this challenge, we adopted a two-pronged approach. First, we provide a set of generalized recommendations for selecting a name matching service, tailored to user-specific factors such as taxonomic group focus or technical expertise. Second, we developed a prototype interactive tool that assists users in identifying a name matching service suited to their particular requirements.

The specific workflows employed by each name matching service are typically documented (some in more detail than others) by the service providers themselves. We provide links to these resources to support users in effectively accessing and applying the services.

The final part of the workflow is the use of the matching results by local users. This is highly dependent on the users’ underlying motivations (see section on motivation above). The action taken, inter alia, may range from simply correcting a misspelled name or erroneous classification, to incorporating a resolvable identifier from the taxonomic dataset, or to actively tracking changes in taxonomic concepts through a (yet-to-be-developed) “Concept Subscription Service.”

CHECKING SCIENTIFIC NAMES OF ORGANISMS - RECOMMENDATIONS

Introduction

Users may have one or multiple organism names they want to check against a taxonomic resource on the web.

The process generally involves two main steps. First, identify an appropriate online reference dataset against which the name could be checked. Second, select an appropriate name-checking service or interactive website.

Before using such a service, it is important to prepare the data according to the tool’s specifications. Some tools require a certain level of technical expertise, while others are more generally user-friendly and directly accessible. The choice of service will also depend on the purpose of the name check and the type of result you expect. For example:

- Do you simply want to verify the correct spelling of the name?
- Do you want to determine whether the name is currently accepted or considered a synonym?
- Are you interested in retrieving the taxonomic classification (e.g. the family to which the taxon belongs)?
- Do you need a resolvable identifier for the name or its corresponding taxon?
- Do you want to check a single name, a small list of names, or a large dataset containing thousands of names?

For uploading a list of names to be matched, a standardised name string has to be prepared for many name matching services. If this is offered by the user’s system (e.g. as DwC, COLDP, JACQ output) specific transformations are necessary to adapt to the respective name matching service. For example, to be recognised by the ChecklistBank

Funded by the European Union

tool, uploaded tables must include a column titled *scientificName*. DwC provides that as taxonName, JACQ as NameAuthorYearString, and for COLDP it should be concatenated from the scientificName and authorship columns to obtain best results. Best practice for services is to allow users to map to the respective column if it already exists under some name (e.g. WFO). Also, a specific delimiter between the column data (comma, tabulator) in the table may be required.

Scope

While some services may also provide additional information about the taxon – such as geographic distribution, descriptions, or images – searching for this kind of content is beyond the scope of the present guide.

Also, with respect to taxonomic coverage, we focus on high-level organism groups (such as plants or animals), and with respect to geographical representation we focus on global coverage, EU / European coverage and coverage of EU and associated countries (the latter only for datasets of large organism groups). A wider scope of sources is covered by Grenié & al. (2022).

Services may provide application programming interfaces (APIs) that often extend the possibilities of name matching or include further functions not provided by the interactive sites. This is also true for the R-packages listed. Undoubtedly these provide methods that can be highly useful, so we tried to document the existence of these tools. However, we refrained from testing anything that requires programming on the user's side, with the sole exception of some simple coding used in the context of OpenRefine.

General recommendations for specific taxonomic groups

Apart from the detailed information attained by the user app described below, some general recommendations can be given for specific taxonomic groups. The

Checking the names of animals

Although the zoological code of nomenclature governs the application of names of animals, the vast zoological domain has resulted in multiple subcommunities treating specific taxonomic groups, some of them with comparatively huge species numbers, some of them of specific interest to the public (e.g. groups of vertebrates). So if users are interested specifically in, e.g. fishes or birds, they are probably best served using bird or fish specific data sets. For example, name-matching via ChecklistBank is possible whenever the respective datasets have been deposited there, e.g. FishBase (Froese & Pauly 2025), HBW and BirdLife International (2024) or the IOC World Bird List (Gill & al. 2025).

Checking the names of plants

Measured by species numbers, plants are a comparatively small group. However, even here specialised datasets exist, e.g. for vascular plants, ferns and fern-allies and mosses (in the wide sense). Fungi and partly also algae, although mostly covered by the same code (in the past: the "botanical" code) of nomenclature, are not considered plants here. Fortunately, botanical institutions have formed the World Flora Online (WFO) initiative that is incorporating various of these individual datasets (as the result of Taxonomic Expert Networks). The WFO Plant List is increasingly becoming the normative source for a consensus classification of all plant names, closely collaborating with initiatives like the International Plant Name Index, the World Checklist of Vascular Plants, and Bryonames. Because it is providing stable and resolvable identifiers also for orthographic variants of names and other invalid designations used in a taxonomic context, we can recommend using that dataset for plant name matching.

Checking the names of fungi

The mycological community has been building scientific name resources for decades and since 2012 is requiring registration of new names. The registration identifier used by the

Funded by the European Union

three recognised registration centres for new names (Index Fungorum, MycoBank, Fungal Names) has also been applied to all pre-existing names, so that a consistent name identifier exists for all fungal names. The repositories continuously synchronise their name holdings so that name matching against any of the repositories gives the same results in terms of name and identifier, with the exception of potentially different instances of taxon concepts and classifications of the names. The International Commission on the Taxonomy of Fungi (ICTF) is working on a consensus classification to be adopted by the three repositories.

- Checking the names of eukaryotic algae

The most comprehensive dataset for algae, AlgaeBase, is integrated into the World Register of Marine Organisms (presumably excluding limnic algae) and, in an outdated version, through GlobalNames. Algae are presently only incompletely covered by Catalogue of Life. Phycobank is the recognised repository for the registration of new names (not yet required); AlgaTerra focusses on eukaryotic microalgae; Index Nominum Algarum is a nomenclator comprising >200,000 algal names compiled from index cards; the Catalogue of Diatom Names is a database apparently not updated since 2011. Of these, only Index Nominum Algarum is currently accessible for name matching through ChecklistBank.

- Checking the names of Cyanobacteria (blue-green algae)

Cyanobacteria present a special case because they have traditionally been treated as algae and their name thus was governed by the International Code of Nomenclature for Algae, Fungi, and Plants (ICNAFP, Turland & al. 2018). However, they are prokaryotic organisms and thus today many cyanobacteria are named in accordance with the rules of the International Code of Nomenclature of Prokaryotes (ICNP Parker & al. 2019). This dual treatment may result in some names being valid under ICNAFP and not under ICNP. As a consequence, the choice of dataset (and related service) may depend on your research context. The LPSN (List of Prokaryotic names with Standing in Nomenclature, https://lpsn.dsmz.de) covers validly published names of cyanobacteria under the ICNP. This is most useful in a prokaryote-microbiological context. AlgaeBase (https://www.algaebase.org) covers cyanobacterial names published under the ICNAFP in a phycological/botanical context, relevant, e.g., for ecological research. NCBI Taxonomy (National Center for Biotechnology Information) https://www.ncbi.nlm.nih.gov/taxonomy includes under bacteria names covered by ICNP and ICNAFP but does not provide an accepted name for synonyms and does not claim to be an authoritative source for nomenclature or classification.

- Checking names of Virus and Viroids

The internationally agreed dataset is the ICTV Master Species List (ICTV 2024) that provides the current taxonomic status of viral names and their classification. The website, https://ictv.global/taxonomy allows searching for individual viral names but apparently there is no bulk upload possible for name checking. No API is offered for machine access to the database at the ICTV website, but the dataset is included in ChecklistBank.

OpenRefine reconciliation interfaces

OpenRefine is a powerful desktop application for cleaning and transforming data. A concise overview is available on Wikipedia. Mozzherin et al. (2024) present a compelling case for adopting OpenRefine reconciliation as a standard approach for name matching services.

In the context of name matching, OpenRefine's key feature is the ability to connect to external reconciliation services, i.e. APIs that match data entries against external sources. A reconciliation service can be used to match a column of scientific names in a table against the taxonomy provided by an aggregator service. The reconciliation services conform to the Reconciliation Service API, a specification developed by the W3C community.

Funded by the European Union

Users install OpenRefine locally; installation packages are available for major operating systems, including a straightforward installer for Windows. Once installed, the desired name matching service can be configured and connected to OpenRefine for repeated use.

To perform name matching, users load their tabular data into OpenRefine and initiate the reconciliation process on the column containing scientific names. In an additional new column, the service returns either exact matches or, when no exact match is found, a list of possible candidates (if any). Users can then review and select the appropriate match for each entry. Once reconciliation is complete, the enriched dataset including the new column containing the matched results can be exported for further use.

The following name-matching reconciliation services are available:

- Global Names Verifier
  - Datasets: covering all datasets that have been imported into Global Names
  - Homepage, including an excellent step-by-step documentation for OpenRefine: https://github.com/gnames/gnverifier/wiki/OpenRefine-readme
  - OpenRefine reconciliation endpoint: https://verifier.globalnames.org/api/v1/reconcile
- International Plant Name Index (IPNI)
  - Dataset: IPNI
  - Homepage: https://data1.kew.org/reconciliation/about/IpniName
  - OpenRefine reconciliation endpoint: http://data1.kew.org/reconciliation/reconcile/IpniName;jsessionid=A955B42E21
- Wikidata Reconciliation Service
  - Datasets: the entirety of Wikidata, but can be restricted to items that are instances of (P31) Taxon (Q16521) for better matches and faster processing.
  - Homepage: https://wikidata.reconci.link/
  - OpenRefine reconciliation endpoint: https://wikidata.reconci.link/en/api
- World Flora Online
  - Dataset: WFO Plant List (latest semiannual version)
  - Homepage: https://list.worldfloraonline.org/reconcile_index.php
  - OpenRefine reconciliation endpoint: https://list.worldfloraonline.org/reconcile

ChecklistBank developers recognised the need to implement a reconciliation API, but it is not yet available (see https://github.com/CatalogueOfLife/backend/issues/1265).

Programming tools for name matching

R-packages

Several of the services identified provide or are R packages, i.e. prepackaged collections of code, data and documentation that extend the functionality of the R programming language. The following packages are included in our table but were not further tested:

- icvplants (Vilela & al. undated)
- taxize (Chamberlain & Szöcs 2013)
- rTNRS (Maitner 2024, Boyle & al. 2013)
- rWCVP (Brown 2023)
- taxotools (Barvc 2025)
- Tremendous (Specker & al. 2024)
- U.Taxonstand (Cayuela & al. 2012, Zhang & Qian 2023)
- WorldFlora (Kindt 2020)

Grenié & al. (2022) cover this subject in detail, identifying 47 (60?) packages providing direct access to online taxonomic databases. They also point to an application ("taxharmonizeexplorer) that should aid R-users to select tools and datasets. The website for the app lists and graphically depicts the relationship between taxonomic datasets and R-packages (as of end of July 2025, it covers 68 packages). If this tool continues to be

Funded by the European Union

updated, it should be the primary source for R-programmers to identify useful packages for name matching, and apply these in a workflow detailed in Grenié & al.'s paper.

- Application programming interfaces (APIs)
  - ChecklistBank Name Match - see https://www.checklistbank.org/about/API
  - GBIF Species Matcher - see https://techdocs.gbif.org/en/openapi/v1/species
  - Global Names Verifier - see https://resolver.globalnames.org/api
  - Taxonomic Name Resolution Service (TNRS) - see https://github.com/ojalaquellueva/TNRSapi?tab=readme-ov-file#readme
  - World Flora Online Plant List - see https://list.worldfloraonline.org/matching_rest.php
  - WoRMS Taxon Match - see https://www.marinespecies.org/aphia.php?p=webservice
    - Others
  - Pytaxon (Proença Neto, De Sousa 2025) - see https://zenodo.org/records/14538804

AN INTERACTIVE USER GUIDE FOR NAME MATCHING

TETTRIs WP2 has implemented a prototype application of an online guide helping biodiversity data curators to identify name-matching services suited to their specific needs. It is available under https://wiki.bgbm.org/tettriswiki/uploads/tettriswiki/demo/.

The tool is based on data selected within the scope that has been detailed above. However, the scope can easily be extended, especially with regard to the geographic and taxonomic coverage. The data presented may in some cases be outdated quite rapidly because many of these resources undergo constant development and their availability may change in the future as well. In the coming months we hope to present a technical solution that facilitates community updating of the data tables and addition of content. For the time being, the current state of the data is presented in spreadsheets for download.

User input for identifying datasets and corresponding name matching services

The user will be presented with a form or a faceted search interface with the following options:

- Organism group

("multiple" or one to several organism groups, include or restrict to fossils, restrict to marine organisms)

- Multiple or unknown
- Animals
- Plants
- Algae
- Fungi
- Bluegreen Algae (Cyanobacteria)
- Bacteria
- Virus
- Include fossils (palaeontology)
- Only fossils
- Only marine species

- Geographic scope

(choose Global, Europe, or one to several countries)

- Global (includes all following)
- Europe (includes all following)

Funded by the European Union

- Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom

Number of names to be checked simultaneously

(choose one)

- up to 5000
- > 5000

Purpose of name matching - I want to check

(choose at least one)

- Orthography of canonical name (name without author / year)
- Orthography of full name (also checking author abbreviations and year)
- Acceptance (if its a taxon name, or a synonym of another name, or taxonomically unresolved)
- Classification (inclusion in higher taxa according to the dataset's hierarchy)

Data linkage requirements

(choose one)

- I am not interested in identifiers and links for the name
- I want a persistent identifier for the name in the dataset
- I want a persistent resolvable identifier for the name in the dataset
- I want a persistent resolvable identifier for the taxon (accepted name) that corresponds to my local name

Technical requirement - I want to

(choose one)

- write or paste the names into a web interface (usually limited number)
- upload a list or table and receive the result as table or tables
- use OpenRefine as the name matching tool
- download an R-package to do the name matching
- program my name matching using the on-line API of the aggregator
- install the name matching service on my server
- download the data, I'll do the matching in my own database

Implementation

User interface

The User Interface for selecting the suitable name matching services is a relatively simple filtered search page with a details panel for a single selected record. By default all of the taxonomic dataset/name-matching service combinations are listed on the page, giving a brief overview of the service, the dataset and their most important features (Fig. 1). The user can then use filters as described above to reduce the number of results presented. All of the filtering is done via JavaScript in the browser, so no dedicated backend search infrastructure is needed.

When the user clicks on one of the list entries, a more detailed page will open up, to show all of the information about the dataset/service combination (Fig. 2).

A working prototype of the search interface can be found at https://wiki.bgbm.org/tettriswiki/uploads/tettriswiki/demo/.

Fig 1: The current user interface for the dataset/service selection

Funded by the European Union

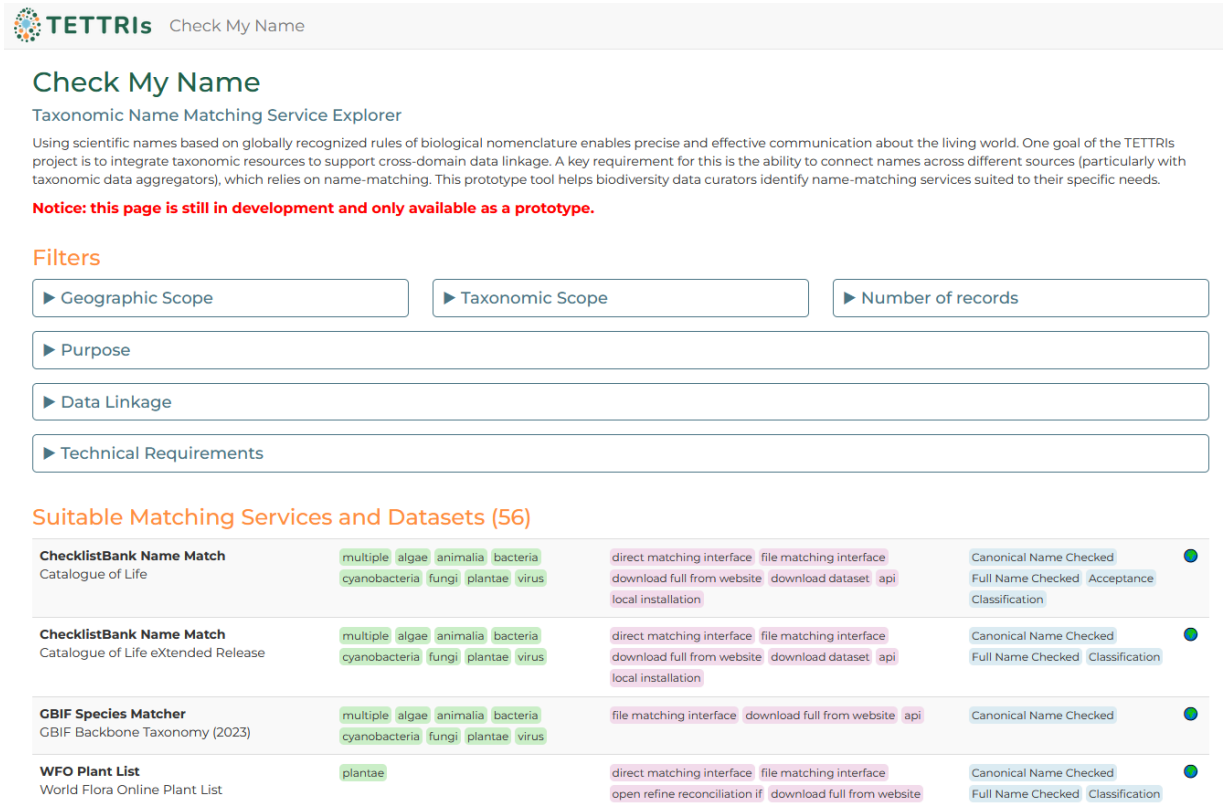
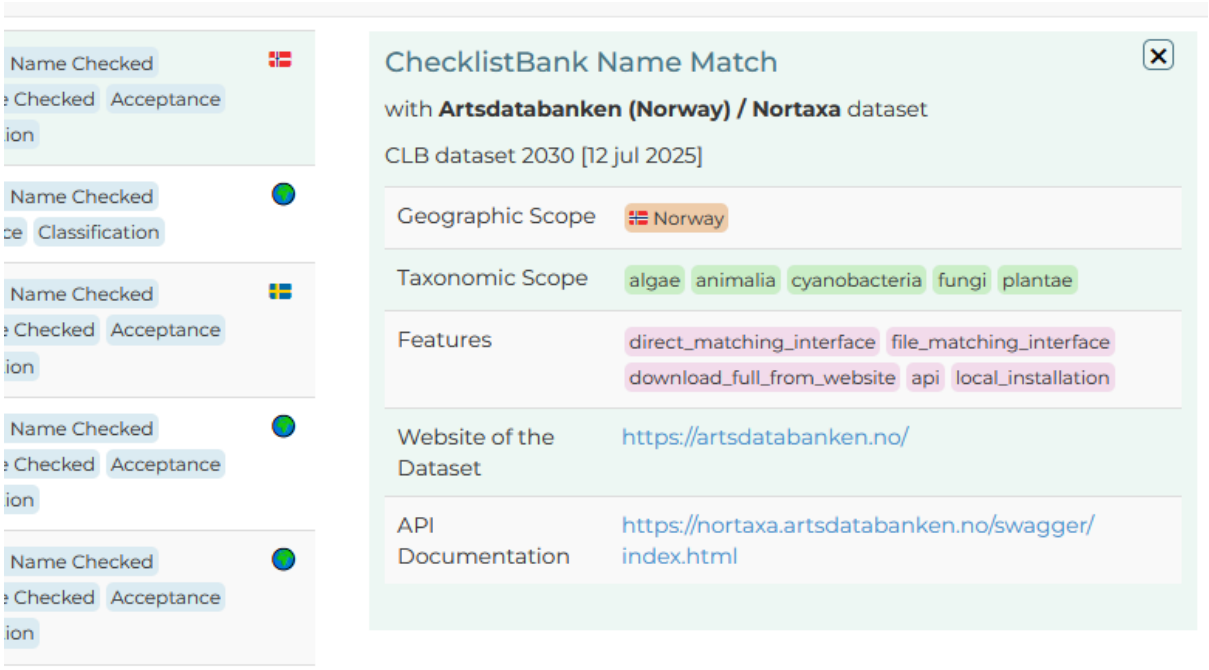


Fig. 2: The current details view for a specific dataset/service combination



- Caveats - work in progress

The implementation is still a demonstrator. The following caveats have been identified in the implementation as of end of July 2025:

- Filter *Number of records* does not yet work
- A button to reset all filters is missing

Funded by the European Union

- The details view should include more data and the structure of the display should be revised
- Links contained in the Details view do not yet work (comments need to be placed elsewhere)
- Currently the page is hosted on the TETTRIs WP2 wiki server, but it in no way depends on the wiki. It can be placed directly on github, with the additional advantage of employing the Github Issue Tracker to record and manage feedback. We are currently evaluating different options to permanently host the site. Once this has been decided, we will properly publish the URL and a redirect from the temporary URL will be put in place.
- Some errors and pending additions to the data tables have to be implemented.
- A procedure to update the underlying data tables in a seamless way has to be defined.

Underlying data

The data schema (Fig. 3) consists of three tables describing datasets, services, and the combination of dataset and service, respectively. See Appendix 1 for the mapping of user-interface selections to table-attributes.

:

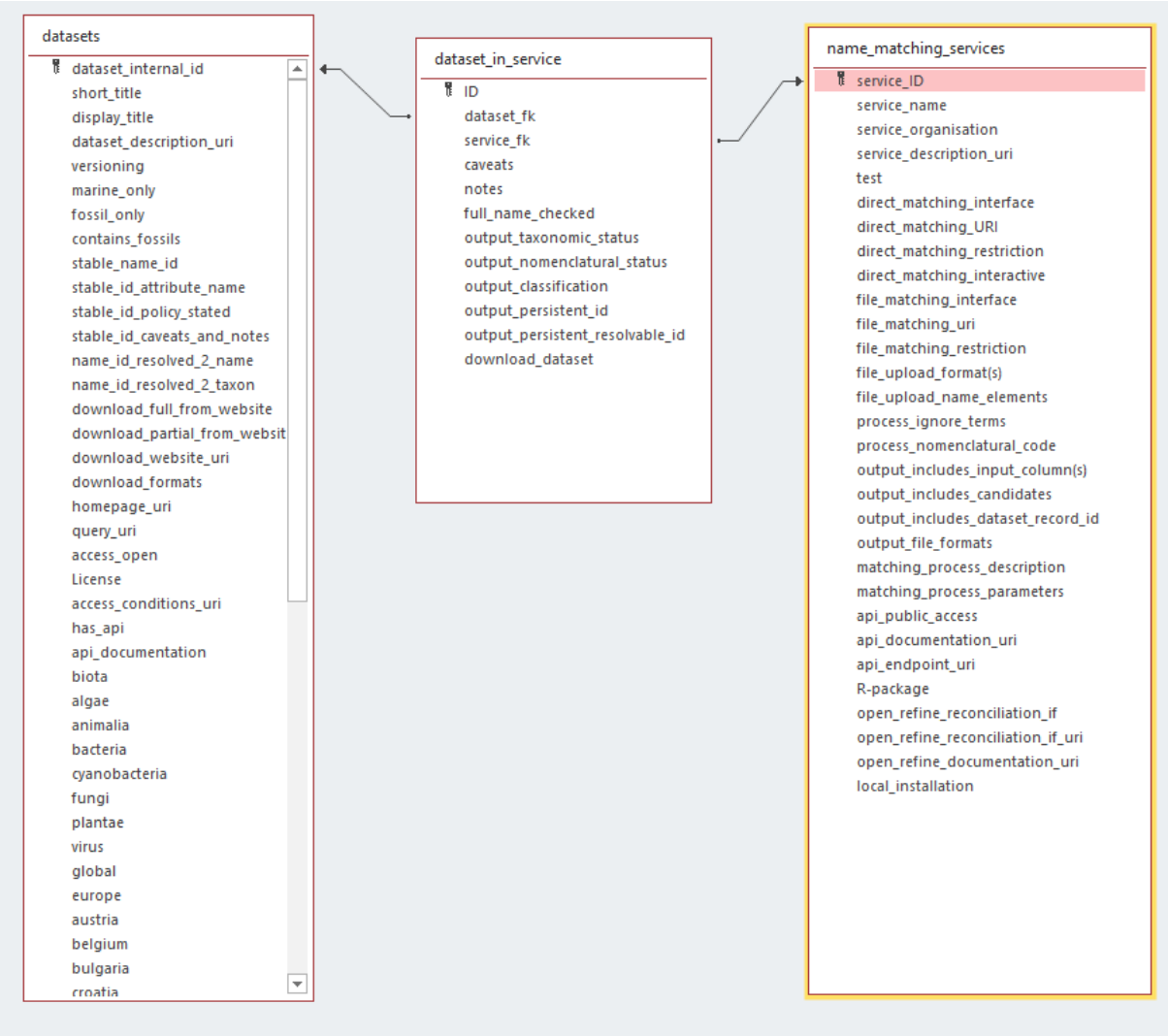


Fig. 3: Relational data structure

For user queries, the 3 data tables (datasets, services, dataset-in-service) will be denormalized (in the process, automatic filling of geographic areas and taxonomic groups will be carried out where general terms have been defined) and exported into a JSON file, which is then used by the user interface to display the services.

Descriptors for taxonomic datasets

Parameters of importance for the user's selection of a name matching workflow were recorded for 34 datasets that can be used in name matching services. Version 1.2 of the data table can be accessed here: https://docs.google.com/spreadsheets/d/1jp5us3v7YRfPdCAUXfO8DXFyfsTjv5wr/edit?usp=sharing&ouid=102259697278873139139&rtpof=true&sd=true

The structure of the table is depicted in Table 1.

Table 1: Structure of the table describing taxonomic datasets

Field Name	Data Type	Description (Optional)
dataset_internal_id	Number	An (internal) ID for the dataset record
short_title	Short Text	Short title
display_title	Short Text	Long title
dataset_description_uri	Short Text	A description given by the site providing the web interface
versioning	Short Text	Description of release policies
marine_only	Yes/No	Focus (almost entirely) on marine organisms
fossil_only	Yes/No	Focus on (almost entirely) fossil organisms
contains_fossils	Yes/No	Includes fossil names
stable_name_id	Yes/No	Stability of name ID is presumed (or stated in a policy)
stable_id_attribute_name	Short Text	Name of the attribute of the ID as output in exports / web interface
stable_id_policy_stated	Short Text	Normally a URL describing the ID policy
stable_id_caveats_and_notes	Long Text	Problems or features noted with respect to the IDs
name_id_resolved_2_name	Yes/No	Name_ID can be included in URI, leads to a name page (independent of its status)
name_id_resolved_2_taxon	Yes/No	Name_ID can be included in URI, leads to a taxon page where the name is found (as taxon or synonym)
download_full_from_website	Yes/No	ecklist data, i.e. ID(s), name, name elements and status, not necessarily also associated data
download_partial_from_webs	Yes/No	Checklist data
download_website_uri	Short Text	Download website
download_formats	Short Text	Checklist data
homepage_uri	Short Text	URI of general website of the specific dataset
query_uri	Short Text	URI of web query form for the specific dataset
access_open	Yes/No	Including -by, -sa and -nc clauses
License	Short Text	E.g. CC-by; see access_uri for access policy statement
access_conditions_uri	Short Text	A web page detailing conditions for access
has_api	Yes/No	An application programming interface is present
api_documentation	Short Text	Documentation for the application programming interface
biota	Yes/No	Taxonomic scope includes all organisms (should automatically include all individual scopes)
algae	Yes/No	Taxonomic scope includes eucaryotic algae
animalia	Yes/No	Taxonomic scope includes animals
bacteria	Yes/No	Taxonomic scope includes bacteria
cyanobacteria	Yes/No	Taxonomic scope includes cyanobacteria (blue-green algae)
fungi	Yes/No	Taxonomic scope includes fungi (including lichenes)
plantae	Yes/No	Taxonomic scope includes plants (always vascular plants, may or may not include "bryophytes")
virus	Yes/No	Taxonomic scope includes virus
global	Yes/No	Geographic scope is global (should automatically include all individual scopes)
europe	Yes/No	Geographic scope includes Europe (should automatically include all individual countries)
austria	Yes/No	
belgium	Yes/No	
bulgaria	Yes/No	

(N.B.: in the data table itself the taxonomic groups and geographical areas are abbreviated.)

Descriptors for name matching services

Parameters of importance for the user's selection of a name matching workflow were recorded for 18 name matching services. Version 1.3 of this data table can be accessed under https://docs.google.com/spreadsheets/d/17JjkDXnoI3-BvgBo8Ku6e8HBCd8MXhZZ/edit?usp=sharing&ouid=102259697278873139139&rtpof=true&sd=true

The structure of the table is given in Table 2.

Funded by the European Union

Table 2: Structure of the table describing name matching services

Field Name	Data Type	Description (Optional)
service_ID	Number	Internal ID (primary key)
service_name	Short Text	Short name of the service
service_organisation	Short Text	The organisation or institution supporting the service
service_description_uri	Short Text	A link to details about the service
test	Short Text	Test result notes
direct_matching_interface	Yes/No	Names can be directly input on the website (no file upload required)
direct_matching_URI	Short Text	Link to the matching web interface
direct_matching_restriction	Number	Limit of records that can be direcly matched
direct_matching_interactive	Yes/No	Users can manually match their unmatched records with offered fuzzy matches
file_matching_interface	Yes/No	File upload for matching possible
file_matching_uri	Short Text	The web interface for file upload
file_matching_restriction	Number	Limit of records that can be uploaded for matching
file_upload_format(s)	Short Text	Format of name-containing file that can be uploaded
file_upload_name_elements	Short Text	Allow input of name components in separate fields in uploads (e.g. genus, epithets, author(s) etc.)
process_ignore_terms	Short Text	(optional or standard) Initial processing of name can filter out terms for uncertainty (c., cf., aff., sp., prox.
process_nomenclatural_code	Short Text	Duplicate names governed by different codes may be correctly matched when the code is indicated
output_includes_input_colum	Yes/No	tching result is united with uploaded records (particularly important for local ID)
output_includes_candidates	Short Text	Exact matches and fuzzy matches are returned in one file or two separate files
output_includes_dataset_reco	Short Text	The identifier given in the source dataset is returned
output_file_formats	Short Text	Format of the output file
matching_process_description	Short Text	Description of the matching process
matching_process_parameters	Short Text	User can select, e.g., fuzzy matching
api_public_access	Yes/No	API for external users available (may require permission)
api_documentation_uri	Short Text	API textual documentation
api_endpoint_uri	Short Text	API access point
R-package	Yes/No	An R-package is provided for name matching (incl. data)
open_refine_reconciliation_if	Yes/No	OpenRefine reconciliation interface available
open_refine_reconciliation_if	Short Text	OpenRefine reconciliation interface access point
open_refine_documentation_	Short Text	How to use for OpenRefine reconciliation interface
local_installation	Yes/No	Entire service can be downloaded and installed locally

Dataset-in-service

Parameters of importance for the user's selection of a name matching workflow were recorded for 57 combinations of dataset and name matching service. See v. 1.2 of the table:

https://docs.google.com/spreadsheets/d/1DR72UewokRt_04MOFr0Ti3FrynkYIjct/edit?usp=sharing&ouid=102259697278873139139&rtpof=true&sd=true

The structure of the table is given in Table 3.

Table 3: Structure of the table describing dataset/service combinations

Field Name	Data Type	Description (Optional)
ID	AutoNumber	
dataset_fk	Number	Dataset ID
service_fk	Number	Service ID
caveats	Short Text	Potential problems in the linkage
notes	Short Text	Notes
full_name_checked	Yes/No	False if only canonical name check
output_taxonomic_status	Yes/No	Taxonomic status (e.g. accepted / synonym / unresolved) is included in output
output_nomenclatural_status	Yes/No	Nomenclatural status is included in output
output_classification	Yes/No	Classification (higher taxon, accepted taxon) is included in the output
output_persistent_id	Yes/No	Service provides a persistent identifier for the records in the dataset
output_persistent_resolvable_	Yes/No	rvice provides a persistent resolvable identifier (may resolve to the source or to the service provider)
download_dataset	Yes/No	Dataset can be downloaded from service site

Funded by the European Union

Table 4 lists the combinations of datasets and services presently recorded. With the exception of the two "services" that actually represent R-packages (lcvplants, rWCVP), the services were tested with a constructed "local" name dataset containing names from different taxonomic groups:

Table 4: Combinations of datasets and services

Dataset	Service
Algaebase	ChecklistBank Name Match
Algaebase	Global Names Verifier
Artsdatabanken (Norway) / Nortaxa	ChecklistBank Name Match
Barcode of Life	ChecklistBank Name Match
Catalogue of Life	ChecklistBank Name Match
Catalogue of Life	Global Names Verifier
Catalogue of Life eXtended Release	ChecklistBank Name Match
Dyntaxa (Sweden)	ChecklistBank Name Match
Encyclopedia of Life	Global Names Verifier
Euro+Med PlantBase	PESI - eu-nomen
FinBIF taxon database	ChecklistBank Name Match
Fungal Names	Global Names Verifier
GBIF Backbone Taxonomy (2023)	GBIF Species Matcher
GBIF Backbone Taxonomy (2023)	Global Names Verifier
GBIF Backbone Taxonomy (2023)	ChecklistBank Name Match
iNaturalist Taxonomy	Global Names Verifier
iNaturalist Taxonomy	ChecklistBank Name Match
Index Fungorum	ChecklistBank Name Match
Index Fungorum	Global Names Verifier
Index to Organism Names (ION)	Global Names Verifier
Integrated Taxonomic Information System	ChecklistBank Name Match
Integrated Taxonomic Information System	Global Names Verifier
International Committee on Taxonomy of Viruses (ICTV) Virus Taxonomy: ICTV Master Species List	ChecklistBank Name Match
International Plant Name Index	Kew Names Matching Service
International Plant Name Index	Global Names Verifier
International Plant Name Index	ChecklistBank Name Match
Le Référentiel Taxonomique Taxref	Global Names Verifier
Le Référentiel Taxonomique Taxref	ChecklistBank Name Match

Funded by the European Union

Leipzig Catalogue of Vascular Plants	ChecklistBank Name Match
Leipzig Catalogue of Vascular Plants	lcvplants
List of Prokaryotic names with Standing in Nomenclature (LPSN)	Global Names Verifier
Mycobank	Global Names Verifier
Mycobank	ChecklistBank Name Match
NCBI Taxonomy	ChecklistBank Name Match
NCBI Taxonomy	Global Names Verifier
Nomenclator Zoologicus (genera and subgenera)	ChecklistBank Name Match
Nomenclator Zoologicus (genera and subgenera)	Global Names Verifier
Paleobiology Database	ChecklistBank Name Match
Paleobiology Database	Global Names Verifier
Pan European Species Information Infrastructure EU - Nomen	PESI - eu-nomen
Species Fungorum [Plus]	ChecklistBank Name Match
Tropicos	ChecklistBank Name Match
Tropicos	Global Names Verifier
UK Species Inventory / Species Dictionary	ChecklistBank Name Match
World Checklist of Vascular Plants	Kew Names Matching Service
World Checklist of Vascular Plants	ChecklistBank Name Match
World Checklist of Vascular Plants	rWCVP
World Checklist of Vascular Plants	Taxonomic Name Resolution Service (TNRS)
World Checklist of Vascular Plants	Global Names Verifier
World Flora Online Plant List	Global Names Verifier
World Flora Online Plant List	ChecklistBank Name Match
World Flora Online Plant List	WFO Plant List
World Flora Online Plant List	Taxonomic Name Resolution Service (TNRS)
World Register of Marine Species	WoRMS Taxon match
World Register of Marine Species	Global Names Verifier
World Register of Marine Species	ChecklistBank Name Match

- Storing and updating the data

At the moment the data tables are stored in online spreadsheets. However if updates are done to the services or the documentation is extended, this transformation needs to be run again. Possible alternative ways to store the data in a way to allow for easy future updates

Funded by the European Union

and seamless integration with the transformation/export mechanism using openly accessible tools are being evaluated.

CONCLUSIONS

The workflow for mapping local taxonomies consists of five main steps: (i) defining the purpose of the mapping, (ii) selecting the target aggregator (including the dataset and associated name matching service), (iii) preparing the names to be matched (e.g., formatting, column titles), (iv) performing the actual name matching, and (v) processing the results. The results of our analysis primarily address the initial steps of the workflow, offering background for users, highlighting common pitfalls, and suggesting potential improvements to aggregator tools.

We propose a simple set of user-oriented criteria to support step (i), and our interactive tool links these criteria to relevant datasets and services in step (ii). The requirements for data preparation (step iii) and details of the name matching process (step iv) are documented by the respective dataset and service providers. The final step (v), processing the results, is inherently context-specific and depends entirely on the user's purpose and motivation, ranging from simple validation to complex integration tasks within local taxonomic infrastructures.

Beyond ongoing technical refinements, the most significant remaining challenge is the reliable updating and expansion of the underlying metadata. Improving dataset coverage will be a focus of TETTRIs Work Package 2 (WP2) in the coming months, in collaboration with both data and service providers. Additional efforts will be made to promote the tool across different user communities. Maintaining up-to-date content requires minimal but continuous editorial effort, which also involves coordination with data providers. WP2 is currently exploring sustainable solutions that may involve institutional commitments from CETAF member organizations.

ACKNOWLEDGEMENTS

The discussions with the TETTRIs satellite project Taxonomic Name Linking Service - TNLS (Roger Hyam, Liesbeth Lyssens, Bart Vanhoorne, with Andreas Müller and Walter Berendsohn as TETTRIs liaison) have provided valuable input. Many thanks to Sofie Meeus for her extensive review of an earlier version of this paper.

REFERENCES CITED

Barvc, V (2025). Package 'taxotools'. Taxonomic list processing. https://doi.org/10.32614/CRAN.package.taxotools]

Boyle B, Hopkins N, Lu Z, Raygoza Garay JA, Mozzherin D, Rees T, Matasci N, Narro ML, Piel WH, Mckay SJ, Lowry S, Freeland C, Peet RK, Enquist BJ (2013). The taxonomic name resolution service: an online tool for automated standardization of plant names. BMC Bioinformatics 14, 16 (2013). https://doi.org/10.1186/1471-2105-14-16

Brown M (2023). Matching names to WCVP with rWCVP. [Accessed 27 jul 2025 at https://matildabrown.github.io/rWCVP/articles/redlist-name-matching.html]

Cayuela L, de la Cerda ÍG, Albuquerque FS, Golicher DJ (2012). TAXONSTAND: an R package for species names standardisation in vegetation databases. Methods Ecol. Evol. 2012;3:1078–1083. https://doi.org/10.1111/j.2041-210X.2012.00232.x

Funded by the European Union

Chamberlain SA, Szöcs E (2013). taxize: taxonomic search and retrieval in R. F1000Research 2013, 2:191. [Version 2] https://doi.org/10.12688/f1000research.2-191.v2

de Jong Y, Kouwenberg J, Boumans L, Hussey C, Hyam R, Nicolson N, Kirk P, Paton A, Michel E, Guiry MD, Boegh PS, Pedersen HÆ, Enghoff H, von Raab-Straube E, Güntsch A, Geoffroy M, Müller A, Kohlbecker A, Berendsohn W, Appeltans W, Arvanitidis C, Vanhoorne B, Declerck J, Vandepitte L, Hernandez F, Nash R, Costello MJ, Ouvrard D, Bezard-Falgas P, Bourgoin T, Wetzel FT, Glöckler F, Korb G, Ring C, Hagedorn G, Häuser C, Aktaç N, Asan A, Ardelean A, Borges PAV, Dhora D, Khachatryan H, Malicky M, Ibrahimov S, Tuzikov A, De Wever A, Moncheva S, Spassov N, Chobot K, Popov A, Boršić I, Sfenthourakis S, Kõljalg U, Uotila P, Gargominy O, Dauvin J-C, Tarkhnishvili D, Chaladze G, Tuerkay M, Legakis A, Peregovits L, Gudmundsson G, Ólafsson E, Lysaght L, Galil BS, Raimondo FM, Domina G, Stoch F, Minelli A, Spungis V, Budrys E, Olenin S, Turpel A, Walisch T, Krpach V, Gambin MT, Ungureanu L, Karaman G, Kleukers RM.J.C, Stur E, Aagaard K, Valland N, Moen TL, Bogdanowicz W, Tykarski P, Węsławski JM, Kędra M, M. de Frias Martins A, Abreu AD, Silva R, Medvedev S, Ryss A, Šimić S, Marhold K, Stloukal E, Tome D, Ramos MA, Valdés B, Pina F, Kullander S, Telenius A, Gonseth Y, Tschudin P, Sergeyeva O, Vladymyrov V, Rizun VB, Raper C, Lear D, Stoev P, Penev L, Rubio AC, Backeljau T, Saarenmaa H, Ulenberg S (2015). PESI - a taxonomic backbone for Europe. Biodiversity Data Journal 3: e5848. https://doi.org/10.3897/BDJ.3.e5848

Franz N, Sterner BW (2017) To increase trust, change the social design behind aggregated biodiversity data. *Database*, Volume 2018, 2018, bax100. https://doi.org/10.1093/database/bax100]

Froese R, Pauly D ed. (2025) FishBase. World Wide Web electronic publication. [Accessed 27 jul 2025 at www.fishbase.org].

GBIF Secretariat (2024). GBIF Science Review No. 11. https://doi.org/10.35035/d9pk-1162

Gill F, Donsker D, Rasmussen P eds. (2025). IOC World Bird List (v15.1). [Accessed 27 jun 2025 at https://www.worldbirdnames.org].

Grenié M, Berti E, Carvajal-Quintero J, Dädlow GML, Sagouis A, Winter M 2022: Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. Methods in Ecology and Evolution 14 (1), 12-25. https://doi.org/10.1111/2041-210X.13802

HBW and BirdLife International (2024). Handbook of the Birds of the World and BirdLife International digital checklist of the birds of the world. Version 9. [Accessed 27 jul 2025 at: http://datazone.birdlife.org/userfiles/file/Species/Taxonomy/HBW-BirdLife_Checklist_v9_Oct24.zip].

Harron K (2016). An introduction to data linkage. ADRN Publications, University of Essex, Wivenhoe Park, UK. [Accessed 27 jul 2025 at https://eprints.ncrm.ac.uk/id/eprint/4282/1/Guide%20to%20data%20linkage.pdf].

Hyam R 2024. What is name matching? TETTRIs WP2 wiki. [Accessed 26 jul 2025 at https://wiki.bgbm.org/tettris/What_is_name_matching%3F]

ICTV (2024). International Committee on Taxonomy of VIruses Master Species List. [Accessed 27 jul 2025 at https://ictv.global/msl/current]

Islam S., Papale D, Vaira L, Rosati I, Peterseil J, Pichot C (2024). Navigating taxonomic complexity: A use-case report on FAIR scientific name-matching service usage in ENVRI Research Infrastructures. Research Ideas and Outcomes 10: e121871. https://doi.org/10.3897/rio.10.e121871

Kindt R. (2020) WorldFlora: An R package for exact and fuzzy matching of plant names against the World Flora Online taxonomic backbone data. Appl Plant Sci. 2020 Sep 25;8(9):e11388. https://doi.org/10.1002/aps3.11388

Maitner B (2024). EnquistLab/RTNRS: CRAN release 0.3.6. https://doi.org/10.5281/zenodo.11186237

Funded by the European Union

Mozzherin D, Paul D, Whitmire A (2024). Can we standardize name reconciliation via OpenRefine? Biodiversity Information Science and Standards 8: e134910. https://doi.org/10.3897/biss.8.134910

Müller A, Güntsch A (2025). Functional optimised EU-Nomen processes and services up-and-running. TETTRIs deliverable 2.1

Parker CT, Tindall BJ, Garrity GM (Eds.) (2019). International Code of Nomenclature of Prokaryotes. Internat. J. Syst. Evol. Microbiol. 69(1A): S1-S111. https://doi.org/10.1099/ijsem.0.000778

Proença Neto MA, De Sousa MPA. Pytaxon: A Python software for resolving and correcting taxonomic names in biodiversity data. Biodivers Data J. 2025 Jan 8;13:e138257. https://doi.org/10.3897/bdj.13.e138257

Specker F, Paz A, Crowther TW, Maynard DS (2024). Treemendous: an R package for integrating taxonomic information across backbones. PeerJ 12:e16896. https://doi.org/10.7717/peerj.16896

Turland NJ, Wiersema JH, Barrie FR, Greuter W, Hawksworth DL, Herendeen PS, Knapp S, Kusber W-H, Li D-Z, Marhold K, May TW, McNeill J, Monro AM, Prado J, Price MJ, Smith GF [Eds] (2018). International Code of Nomenclature for algae, fungi, and plants (Shenzhen Code) adopted by the Nineteenth International Botanical Congress Shenzhen, China, July 2017. Regnum Vegetabile 159. Koeltz Botanical Books, Glashütten. https://doi.org/10.12705/Code.2018

Vaccaro, G. & Swerts, E. (2022). Data Linkage. FORS Guide No. 18, Version 1.0. Lausanne: Swiss Centre of Expertise in the Social Sciences FORS. https://doi.org/10.24449/FG-2022-00018

Vilela B, Gentile A, Freiberg M, Winter M (undated). Taxonomic resolution using lcvplants. [Accessed 27 jul 2025 at https://idiv-biodiversity.github.io/lcvplants/articles/taxonomic_resolution_using_lcvplants.html].

Zhang J, Qian H. (2022). U.Taxonstand: An R package for standardizing scientific names of plants and animals. Plant Divers. 2022 Sep 8;45(1):1-5. https://doi.org/10.1016/j.pld.2022.09.001

APPENDIX 1: Mapping of user-interface selections to table-attributes

Organism group (choose multiple or one or several of the following; mark fossil / marine)

	Attribute	Type	Notes
Multiple or unknown	datasets.biota	boolean	(markes the following 7 groups upon matrix input)
Animals	datasets.animalia	boolean	
Plants	datasets.plantae	boolean	
Algae	datasets.algae	boolean	
Fungi	datasets.fungi	boolean	
Bluegreen algae	datasets.cyanobacteria	boolean	
Bacteria	datasets.bacteria	boolean	
Viruses	datasets.virus	boolean	
Include fossils	datasets. contains_fossils	boolean	
Only fossils	datasets. fossile_only	boolean	

Funded by the European Union

Only marine species	datasets. marine_only	boolean	

Geographic scope (choose Global, Europe, or one to several countries)

	Attribute	Type	Notes
Global	datasets.global	boolean	(marks all the following upon input)
Europe	datasets.europe	boolean	marks all contries upon input
[country name]	datasets.[country name]	boolean	upon input replaced by iso country code

Countries: Austria, Belgium, Bulgaria, Croatia, Republic of Cyprus, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovakia, Slovenia, Spain, Sweden, Switzerland, Ukraine, United Kingdom

Data size
(Number of names to be checked simultaneously)

	Attribute	Type	Notes
.. cut and paste IF	services.direct_matching_restriction	numeric	useful classes: < 5000, >=5000
.. file upload	services.file_matching_restriction	numeric	useful classes: < 5000, >=5000

Purpose of name matching (I want to check ... - choose at least one)

orthography of canonical name (name without author / year)	*		covered by all datasets
Orthography of full name (also checking author abbreviations and year)	dataset_in_service.full_name_checked	boolean	
Acceptance (if its a taxon name, or a synonym of another name, or taxonomically unresolved)	dataset_in_service.output_taxonomic_status	boolean	
Classification (inclusion in higher taxa according to the dataset's hierarchy)	dataset_in_service.output_classification	boolean	

Data linkage requirements (choose one)

	Attribute	Type	Notes
I am not interested in identifiers and links for the name	*		
I want a persistent identifier for the name in the dataset that corresponds to my local name	dataset_in_service. output_persistent_id	boolean	

Funded by the European Union

I want a persistent resolvable identifier for the name in the dataset that corresponds to my local name	dataset_in_service.output_persistent_resolvable_id	boolean	
I want a persistent resolvable identifier for the taxon (accepted name) that corresponds to my local name	dataset.name_id_resolved_2_taxon	boolean	

Technical requirement (choose one)

I want to	Attribute	Type	Notes
write or paste the names into a web interface (usually limited number)	services.direct_matching_interface	boolean	(may include file pasting)
upload a list or table and receive the result as table or tables	services.file_matching_interface	boolean	
use OpenRefine as the name matching tool	services.open_refine_reconciliation_if	boolean	
download the data, I'll do the matching in my own database	datasets.download_full_from_website OR dataset_in_service.download_dataset	boolean, boolean	
program my name matching using the on-line API of the aggregator	services.api_public_access	boolean	(available, may need permission or login)
use an R-package that can be downloaded	services.R_package	boolean	(documentation: field for API-documentation)
install the name matching service on my server	services.local_installation	boolean	(usually a Docker package)

Funded by the European Union

- APPENDIX 2: Related documents

ID	Reference or Related Document	Source or Link/Location
1	https://tettris.eu/wp-content/uploads/2024/07/TETTRIs-MS4-Agreed-strategy-for-mapping-local-taxonomies-.docx.pdf	LINK
2		LINK
3		LINK
4		LINK
5		LINK
6		LINK
7		LINK

Funded by the European Union