OPEN ACCESS

CrossMark

Case Study

Navigating taxonomic complexity: A use-case report on FAIR scientific name-matching service usage in ENVRI Research Infrastructures

Sharif Islam[‡,§], Dario Papale[|], Lucia Vaira[¶], Ilaria Rosati[#,¤], Johannes Peterseil[«], Christian Pichot[»]

‡ Naturalis Biodiversity Center, Leiden, Netherlands
§ DiSSCo, Leiden, Netherlands
| University of Tuscia DIBAF, Viterbo, Italy
¶ LifeWatch ERIC, Lecce, Italy
# National Research Council (CNR), Research Institute on Terrestrial Ecosystems (IRET), Lecce, Italy
¤ LifeWatch Italy, Lecce, Italy
« Umweltbundesamt GmbH, Vienna, Austria
» INRAE, Avignon, France

Corresponding author: Sharif Islam (sharif.islam@naturalis.nl)

Reviewable   v 1

Abstract

This paper presents a use-case conducted within the ENVRI FAIR project, examining challenges and opportunities in deploying FAIR-aligned (ensuring Findability, Accessibility, Interoperability and Reusability) scientific name-matching services across Environmental Research Infrastructures (RIs). Six services were tested using various name variations, revealing inconsistencies in match types, status reporting and handling of canonical forms and typos. These diversities pose challenges for RI data pipelines and interoperability. The paper underscores the importance of standardised tools, enhanced communication, training, collaboration and shared resources. Addressing these needs can facilitate more effective FAIR implementation within the ENVRI community and biodiversity research. This, in turn, will empower RIs to seamlessly integrate and leverage scientific names, unlocking the full potential of their data for research and policy implementation.

Keywords

scientific names, taxonomy, biodiversity, FAIR, ENVRI

Introduction and Background

Scientific names have served as the globally-accepted practice for identifying and describing species, bringing order to the discipline of systematics for the past 250 years (Thines et al. 2020; Hobern et al. 2021). In the era of multidisciplinary and data-driven research questions, these names are not only pivotal for scientific exploration (Patterson et al. 2010), but also hold significance in conservation, trade, biosecurity, legislation and disease management (Tedesco et al. 2014; Thompson et al. 2021). As data points, scientific names possess distinct attributes, representing dynamic hypotheses subject to classification changes as new data and knowledge about the species emerge. The taxonomy and biodiversity research community is acquainted with these dynamics (Wheeler et al. 2004; Garnett et al. 2020; Pyle et al. 2021), acknowledging that this ambiguity generates "difficulties for end users to point to single valid names referring unambiguously to single taxonomic concepts" (Grenié et al. 2022:2). These names and concepts, integral components of the data ecosystem, are essential for implementing the FAIR principles (Findable, Accessible, Interoperable and Reusable) (Wilkinson et al. 2016), particularly when linking and reusing associated data in various use-cases (Gemeinholzer et al. 2020; Thessen et al. 2021).

In the world of scientific and FAIR data, Research Infrastructures (RIs) are crucial (Borgman 2010). These are large set-ups, often collaborative and international, that provide data and services to scientists and policy-makers (Ribeiro 2021). In the environmental and biodiverisity field, RIs focus on, for instance, terrestrial ecosystems, looking at things like interactions between climate and habitat loss effects on biodiversity. Despite the inherent ambiguity around name resolution, RIs providing biodiversity and ecosystem data and services require tools, training, best practices and solutions to address challenges arising from the usage of scientific names.

This paper outlines the approach and findings of a FAIR implementation use-case exercise within the ENVRI Scientific Cluster,*[1] under the purview of the ENVRI FAIR*[2] project. The primary focus was on the utilisation of scientific names, a common element identified by various RIs within the Biodiversity and Ecosystem subdomain. The use-case targeted certain tools, chosen, based on their maturity and community acceptance, engaging in a name-matching exercise - one of the primary tasks in the data workflow -to understand current practices, tool availability and potential error scenarios. A crucial objective was to highlight implications for interoperability and linking between RIs and services.

Given that many RIs collect environmental samples and species information across diverse locations, various data pipelines are established for activities, such as recording type specimen details (Sluys 2021), CO_2 flux measurement, species occurrence data, biomass and Leaf Area Index (for more examples, see Peters et al. (2014); Pastorello et al. (2020);

Géron et al. (2021)). Individuals engaged in these data pipelines, often lacking specialisation in taxonomy or biology, face challenges in navigating complexities during name-matching and resolution. To address these challenges, RIs must choose a suitable tool, deciding whether to utilise existing services or develop their own tools for specific use-cases. The use-case exercise was designed within this context. It is important to note that our exercise did not encompass an exhaustive evaluation of all tools, R packages, Python libraries, APIs and best practices related to taxonomic databases and scientific name harmonisation. For a comprehensive examination of these aspects, we refer readers to Grenié et al. (2022).

Currently, within the ENVRI cluster, the use of scientific names is fragmented, with participating RIs employing individual approaches without coordination or a shared agreement on information sources, tools and best practices. This decentralised handling poses risks, such as including the use of outdated data, missing scientific names and inconsistencies in synonym usage, leading to potential data interoperability issues (Reyserhove et al. 2020). Notably, no universally accepted solution exists across diverse user communities due to variations in taxonomic coverage, spanning from terrestrial to microbial domains. While a singular solution may prove impractical, fostering a shared consensus on tools, documentation, best practices and training has the potential to mitigate the mentioned challenges and significantly contribute to implementing the FAIR framework within the ENVRI cluster.

Our vision is to empower diverse user-groups to seamlessly utilise, link and integrate scientific names in datasets distributed by RIs, alongside data from other sources, without introducing errors and risks. A shared understanding and adoption of best practices for scientific name usage across all RIs are crucial. Ideally, taxonomic services should be interoperable, facilitating the integration of scientific name-matching and resolution within different data pipelines. In the subsequent sections, we delve into the approach used to test services related to scientific names, providing examples of results that highlight two critical aspects - match type and name status. The paper concludes with recommendations and potential plans for future work.

Approach

Various services are available for users to conduct matching and resolution of scientific names. For our use-case, we selected six services and tested the name-matching feature: Global Names Resolver (GNR), Catalogue of Life (COL) name-match tool, Global Biodiversity Information Facility (GBIF), LifeWatch taxon match services (LW), National Center for Biotechnology Information (NCBI) taxonomy database and World Flora Online (WFO). While there are numerous aspects to measure and compare, we simplified our focus to the results, specifically the name-matching status (e.g. exact, fuzzy or canonical) and the name status (accepted or synonym). From this initial list, we narrowed our discussion and collaboration to two services commonly used by the involved RIs: "LifeWatch taxon match services" and "Catalogue of Life name-match tool". In this section, we present the results of the exercise.

The use-case provides insight into challenges faced by RIs in the ENVRI Science Cluster, revealing their dependence on external services, experts, resources and training for internal service creation. While using a single service or taxonomic backbone for all ENVRI RIs may not be practical due to diverse domains, a common tool within specific taxonomic domains (terrestrial plants, for instance) might be feasible. Global taxonomic expertise and biodiversity data service providers (e.g. LifeWatch, COL, GBIF) can offer guidelines to assist RIs in selecting appropriate tools. Once a common set of tools is adopted, ensuring consistency in species names reported to RI-specific central databases becomes a challenge. The continuous influx of data from scientists and technicians with varying expertise underscores the need for an approach ensuring inter-comparability amongst different reference systems and facilitating periodic updates in case of species reclassification.

The specific requirements from RIs thus include the need for robust web services, tools and APIs for automated integration with local data pipelines for name-matching and resolution. Emphasising common matching options across different services, improved input data structuring guidelines, standardised service responses, manual and bulk options, interoperable API responses and comprehensive documentation and training are crucial for enhancing the efficiency and accuracy of scientific name validation within the ENVRI cluster.

Discussion

In this section, we highlight a few important aspects emerging from the matching exercise.

Different options for submitting names

A scientific name is a mandatory field in all the services used in the exercise. However, the input interfaces and matching options vary significantly. For instance, the COL cross-dataset search (Fig. 1) offers "fuzzy", "exact" and "partial" options, along with choices to restrict the search to the scientific name only. It is unclear to a non-specialist how "fuzzy", "exact" and "partial" matches differ and which of these should be considered in the initial name matching phase. The COL bulk input option lacks these choices (Fig. 2). Similarly, LifeWatch (Fig. 3) provides a bulk input service with a list of different taxon services (Fig. 4) that can be selected, running as a submitted job, but without an option to choose search filters. Some services offer both Python and R packages, while others provide only one.

Canonical vs. non-canonical match

The term "canonical name" here specifically refers to the Latinised elements. There is inconsistency in how services handle canonical versus non-canonical matches. In our test, *Pinus mugo* Smith was considered "canonical" by GNR (the API differentiates between "supplied_name_string": "Pinus mugo Smith" and "canonical_form": "Pinus mugo"), but "none" by COL. It is difficult for non-specialists to interpret this result. Additionally, the

approach of different services in handling canonical names and authorship together is not transparent. Do all services ignore the authorship and parse each component separately or do they match them as a single entity?
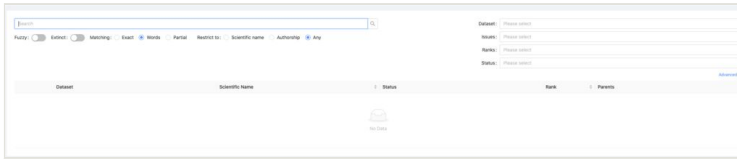

Figure 1. doi

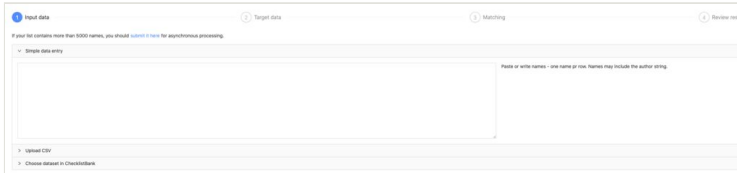COL Checklistbank cross dataset search. Screenshot captured 10 Jan 2024.

Figure 2. doi

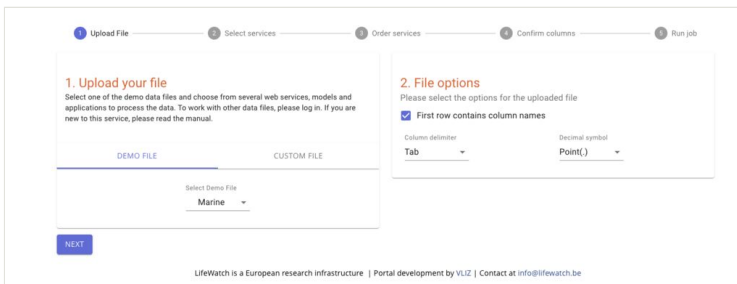COL name match tool for bulk input. Screenshot captured 10 Jan 2024.

Figure 3. doi

LifeWatch web service data upload screen. Screenshot captured 10 Jan 2024.

Flag spelling or a possible typo

In large datasets, typos and spelling errors are expected. In our test, *Pinus muco* Turra was not flagged as a typo or error.

Lack of common vocabulary for match types

As listed in Table 1 and the Juypyter notebook, there are different ways of describing the match type. For instance: "Exact match by canonical form" (GNR), "Fuzzy match by canonical form" (GNR), "fuzzy matches: more than one possibility" (LifeWatch),

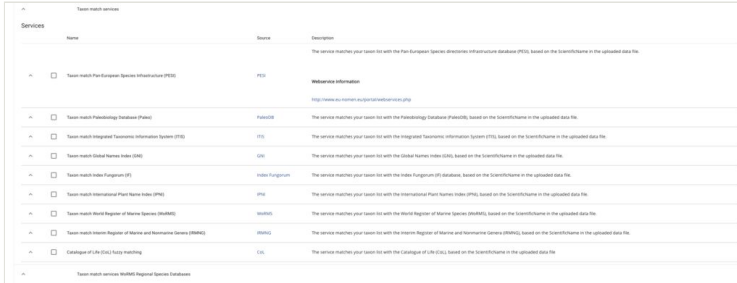"matchtype: exact" (LifeWatch), "No match found" (LifeWatch), "exact" (COL), "variant" (COL).

Figure 4. doi

LifeWatch web service taxon match list. Screenshot captured 10 Jan 2024.

Table 1.

Table1: Examples of results from different name matching services. GNR = Global Names Resolver API; COL = Catalogue of Life Checklistbank name match web tool; GBIF = Global Biodiversity Information Facility python library and API; LW = LifeWatch Taxon Match web service; NCBI = National Center for Biotechnology Information taxonomy python library, WFO = World Flora Online R package. See the companion Jupyter notebook for details.

Scientific name	Input type	Responses from different services					
		GNR	COL	GBIF	LW	NCBI	WFO
Pinus mugo Turra	Accepted name with authorship. Response expected: **Match type**: exact match with authorship. **Status**: accepted	**Match type**: Exact , provides a score, includes multiple backbones and checklists, provides status of the request ("success" or "failure"). **Status**: no indication whether the name is accepted or not.	**Match type**: Exact match, no score, **Status**: accepted.	**Match type**: Exact match, no score, **Status**: accepted.	**Match type**: Exact match, no score, includes multiple backbone and checklists, **Status**: no indication whether the name is accepted or not.	Resolving failed. NCBI only matches based on the canonical name which is composed of only the Latinised elements of a scientific name.	No **Match type**. **Status**: accepted
Pinus mugo	Accepted nameexcluding the authorship Response expected: **Match type**: Exact match by canonical form, authorship missing. **Status**: accepted	**Match type**: Exact match by canonical form, provides a score, includes multiple databases and checklists, **Status**:no indication whether the name is accepted or not.	**Match type**: variant, no score, **Status**: accepted.	**Match type**: Multiple matches based on the full name and genus, no matching status or score provided, **Status**: accepted	**Match type**: No exact match found, more than one possibility, no score, includes multiple databases and checklists, no indication whether name is accepted or not.	**Match type**: Exact match, but includes subgenus and lineage, no score, no indication whether the name is accepted or not.	No **Match type**. **Status**: accepted

Scientific name	Input type	Responses from different services					
		GNR	*COL*	*GBIF*	*LW*	*NCBI*	*WFO*
Pinus mugo Smith	Incorrect author. Response expected: **Match Type**: Exact match by canonical form, incorrect authorship. **Status**: Not accepted	**Match type**: Exact match by canonical form, provides a score, includes multiple databases and checklists, no indication whether the name is accepted or not.	**Match type**: none, no indication whether the name is accepted or not.	Empty response	No match found	Empty response	Returns multiple results (includes *Pinus mugo* Turra and synonyms)
Pinus muco Turra	Spelling mistake Response expected: **Match Type**: Suspect spelling mistake. **Status**: Not Accepted	**Match type**:Fuzzy match by canonical form, spelling mistake not flagged	**Match type**: none	Empty response	No match found	Empty response	Returns multiple results (includes *Pinus mugo* Turra and *Pinus pumilio* (Turra) Franco)
Pino mugo Turra	Incorrect genus Response expected: **Match Type**: Wrong genus. **Status**: Not Accepted.	No result. API provides API status response.	**Match type**: none	Empty response	No match found	Empty response	Returns multiple results (includes *Pinus mugo* Turra)
Pinus basso Turra	Non-existent species Response expected: **Match Type**: No matching species. **Status**: Not Accepted	**Match type**: Could only match genus, provides a score, includes multiple databases and checklists, no indication whether the name is accepted or not.	**Match type**: none	Empty response	No match found	Empty response	Returns multiple results (includes *Pinus mugo* Turra)

API and library responses

There is no common approach to providing API and software library responses after a name string are provided. Documentation quality also varies accross different providers. See Table 2 for some examples.

Concluding remarks and future work

Taxonomic data and scientific names play pivotal roles in multidisciplinary data-linking (Orr et al. 2021) and in ensuring the FAIRness of environmental and biodiversity research data (Hobern et al. 2019; Vassallo and Felicetti 2020; Sterner et al. 2021) thus, globally-accepted use of scientific names and related services remains a significant concern for RIs. The data related to the scientific practice of taxonomy are complex and multifaceted.[*3] As demonstrated, a variety of databases and services regarding name-matching and resolution are available, catering to a wide range of use-cases. Each mentioned service provider has distinct advantages and disadvantages, varying stages of maturity and differing software development plans. The heterogeneity of biodiversity and ecosystem services compounds the challenge of data integration and linking. Despite these challenges and the practical diversity of organisations with different national and regional priorities, data and services related to scientific names are essential components in the ENVRI ecosystem. To harness the vast amount of available data, RIs, stakeholders and

service providers must establish broader and long-term collaboration. Initiating from this use-case group, a training session organised in November 2022 focused on the Catalogue of Life Checklist Bank, marking a step towards such collaboration. These sessions clarified issues, provided valuable feedback for COL development and served as a channel for feedback, ensuring better and more efficient usage of taxonomic data. As stakeholders like LifeWatch, COL and GBIF develop tools, we believe the ENVRI community should actively participate in this global conversation, echoing sentiments shared in discussions within the BiCIKL project (which runs until 2024; involving DiSSCo, LifeWatch and GBIF). The issues and suggestions presented here align with findings from Grenié et al. (2022) (examining taxonomic databases) and Feng et al. (2022) (exploring the heterogeneous landscape of biodiversity databases). To conclude, we offer the following suggestions and action items for both the ENVRI community and service providers:

1.   Taxonomic experts and service providers should collaborate with the RIs to organise regular training sessions tailored for different user groups.
2.   Enhance documentation around web services and API usage for improved user understanding.
3.   Provide clarification and communicate ongoing and future development roadmaps for services, utilising platforms like GitHub or other relevant venues.
4.   Encourage RIs within ENVRI to communicate, identifying common challenges and exploring ways to share resources and expertise.
5.   Establish and adopt best practices for different use-cases.

Table 2.

Different API responses from different taxonomic services.

API Responses:

API Provider	Input String	JSON Response
GBIF	Pino mugo Turra (incorrect genus)	{ "confidence": 100, "matchType": "NONE", "synonym": false}
GNR	Pino mugo Turra (incorrect genus)	{ "id": "zbltk7bpnuh3", "url": "http://resolver.globalnames.org/name_resolvers/zbltk7bpnuh3.json", "data_sources": [], "data": [{ "supplied_name_string": "Pino mugo Turra", "is_known_name": false }] , "status": "success", "message": "Success", "parameters": { "with_context": false, "header_only": false, "with_canonical_ranks": false, "with_vernaculars": false, "best_match_only": false, "data_sources": [], "preferred_data_sources": [], "resolve_once": false }}
COL	Pino mugo Turra (incorrect genus)	Large JSON Respnse (see link) with class and family match such as "Pinopsida"

Python Libraries:

API Responses:		
API Provider	Input String	JSON Response
Library names	Input String	Response
GBIF Python Client	Pino mugo Turra	>>> species.name_suggest(q='Pino mugo Turra')[]>>> species.name_lookup(q='Pino mugo Turra') {'offset': 0, 'limit': 100, 'endOfRecords': True, 'count': 0, 'results': [], 'facets': []}
taxoniq (for NCBI query, provided by a third party)	Pino mugo Turra	KeyError: 'Pino mugo Turra'
ncbi-taxonomist	Pino mugo Turra	{"empty response": {"queryid": "SetLe7VFSwGDd464ZGw4IA==", "action": "skip"}}

These adoptions will pave the way for a more effective FAIR implementation. Building a resilient and enduring collaboration amongst taxonomists, biodiversity and ecosystem researchers, taxonomic service providers and Research Infrastructures is pivotal for success where clear roles and collaboration paths can be identified. In this context, for instance, taxonomists can concentrate on identification and classification tasks, while service providers incorporate these decisions into their offerings, ensuring seamless interoperability, linking and relationships across various tools and databases. These envisioned collaboration can empower operators and data managers working in the RIs with efficient tools to validate names against diverse databases, with service providers ensuring the necessary interlinkage with other existing databases (also see recent publications echoing similar sentiments Sandall et al. (2023) and Lien et al. (2023)). Such a framework simplifies the utilisation of scientific names within RIs and facilitates decision-making regarding various changes and updates originating from the reference taxonomic databases and services.

Acknowledgements

Conflicts of interest

The authors have declared that no competing interests exist.

References

- Borgman CL (2010) Scholarship in the digital age: Information, infrastructure, and the Internet. MIT press
- Feng X, Enquist BJ, Park DS, Boyle B, Breshears DD, Gallagher RV, Lien A, Newman EA, Burger JR, Maitner BS, Merow C (2022) A review of the heterogeneous landscape of biodiversity databases: opportunities and challenges for a synthesized biodiversity knowledge base. Global Ecology and Biogeography https://doi.org/10.1111/geb.13497
- Garnett S, Christidis L, Conix S, Costello M, Zachos F, Bánki O, Bao Y, Barik S, Buckeridge J, Hobern D, Lien A, Montgomery N, Nikolaeva S, Pyle R, Thomson S, van Dijk PP, Whalen A, Zhang Z, Thiele K (2020) Principles for creating a single authoritative list of the world's species. PLOS Biology 18 (7). https://doi.org/10.1371/journal.pbio.3000736
- Gemeinholzer B, Vences M, Beszteri B, Bruy T, Felden J, Kostadinov I, Miralles A, Nattkemper TW, Printzen C, Renz J, Rybalka N (2020) Data storage and data re-use in taxonomy-the need for improved storage and accessibility of heterogeneous data. Organisms Diversity & Evolution 20 (1): 1-8. https://doi.org/10.1007/s13127-019-00428-w
- Géron C, Lembrechts J, Borgelt J, Lenoir J, Hamdi R, Mahy G, Nijs I, Monty A (2021) Urban alien plants in temperate oceanic regions of Europe originate from warmer native ranges. Biological Invasions 23 (6): 1765-1779. https://doi.org/10.1007/s10530-021-02469-9
- Grenié M, Berti E, Carvajal-Quintero J, Dädlow GM, Sagouis A, Winter M (2022) Harmonizing taxon names in biodiversity data: A review of tools, databases and best practices. Methods in Ecology and Evolution https://doi.org/10.1111/2041-210X.13802
- Hobern D, Baptiste B, Copas K, Guralnick R, Hahn A, van Huis E, Kim ES, McGeoch M, Naicker I, Navarro L, Noesgaard D (2019) Connecting data and expertise: a new alliance for biodiversity knowledge. Biodiversity data journal 7.
- Hobern D, Barik SK, Christidis L, T Garnett S, Kirk P, Orrell TM, Pape T, Pyle RL, Thiele KR, Zachos FE, Bánki O (2021) Towards a global list of accepted species VI: The Catalogue of Life checklist. Organisms Diversity & Evolution1-14. https://doi.org/10.1007/s13127-021-00516-w
- Lien A, Banki O, Barik S, Buckeridge J, Christidis L, Cigliano MM, Conix S, Costello MJ, Hobern D, Kirk P, Kroh A, Montgomery N, Nikolaeva S, Orrell T, Pyle R, Raz L, Thiele K, Thomson S, van Dijk PP, Wambiji N, Whalen A, Zachos F, Zhang Z, Garnett S (2023) Widespread support for a global species list with a formal governance system. Proceedings of the National Academy of Sciences 120 (45). https://doi.org/10.1073/pnas.2306899120
- Orr MC, Ferrari RR, Hughes AC, Chen J, Ascher JS, Yan YH, Williams PH, Zhou X, Bai M, Rudoy A, Zhang F (2021) Taxonomy must engage with new technologies and evolve to face future challenges. Nature Ecology & Evolution 5 (1): 3-4. https://doi.org/10.1038/s41559-020-01360-5
- Pastorello G, Trotta C, Canfora E, Chu H, Christianson D, Cheah YW, Poindexter C, Chen J, Elbashandy A, Humphrey M, Isaac P (2020) The FLUXNET2015 dataset and the ONEFlux processing pipeline for eddy covariance data. Scientific data 7 (1): 1-27.

- Patterson DJ, Cooper J, Kirk PM, Pyle RL, Remsen DP (2010) Names are key to the big new biology. Trends in ecology & evolution 25 (12): 686-691. https://doi.org/10.1016/j.tree.2010.09.004
- Peters DP, Loescher HW, SanClements MD, Havstad KM (2014) Taking the pulse of a continent : expanding site-based research infrastructure for regional-to continental-scale ecology. Ecosphere 5 (3): 1-23. https://doi.org/10.1890/ES13-00295.1
- Pyle RL, Barik SK, Christidis L, Conix S, Costello MJ, van Dijk PP, Garnett ST, Hobern D, Kirk PM, Lien AM, Orrell TM (2021) Towards a global list of accepted species V. The devil is in the detail. Organisms Diversity & Evolution 21 (4): 657-675. https://doi.org/10.1007/s13127-021-00504-0
- Reyserhove L, Desmet P, Oldoni D, Adriaens T, Strubbe D, Davis AJ, Vanderhoeven S, Verloove F, Groom Q (2020) A checklist recipe: making species data open and FAIR. Database 2020 https://doi.org/10.1093/database/baaa084
- Ribeiro M (2021) Towards a Sustainable European Research Infrastructures Ecosystem The Economics of Big Science. In: Beck HP, Charitos P (Eds) The Economics of Big Science. Science Policy Reports. Science Policy Reports. Springer https://doi.org/10.1007/978-3-030-52391-6_2
- Sandall E, Maureaud A, Guralnick R, McGeoch M, Sica Y, Rogan M, Booher D, Edwards R, Franz N, Ingenloff K, Lucas M, Marsh C, McGowan J, Pinkert S, Ranipeta A, Uetz P, Wieczorek J, Jetz W (2023) A globally integrated structure of taxonomy to support biodiversity science and conservation. Trends in Ecology & Evolution 38 (12): 1143-1153. https://doi.org/10.1016/j.tree.2023.08.004
- Sluys R (2021) Attaching Names to Biological Species: The Use and Value of Type Specimens in Systematic Zoology and Natural History Collections. Biol Theory 16: 49-61. https://doi.org/10.1007/s13752-020-00366-3
- Sterner B, Upham N, Gupta P, Powell C, Franz N (2021) Wanted: Standards for FAIR taxonomic concept representations and relationships. Biodiversity Information Science and Standards 5 https://doi.org/10.3897/biss.5.75587
- Tedesco PA, Bigorne R, Bogan AE, Giam X, Jézéquel C, Hugueny B (2014) Estimating how many undescribed species have gone extinct. Conservation Biology 28 (5): 1360-1370. https://doi.org/10.1111/cobi.12285
- Thessen AE, Bogdan P, Patterson DJ, Casey TM, Hinojo-Hinojo C, De Lange O, Haendel MA (2021) From Reductionism to Reintegration: Solving society's most pressing problems requires building bridges between data types across the life sciences. PLoS Biology 19 (3). https://doi.org/10.1371/journal.pbio.3001129
- Thines M, Aoki T, Crous PW, Hyde KD, Lücking R, Malosso E, May TW, Miller AN, Redhead SA, Yurkov AM, Hawksworth DL (2020) Setting scientific names at all taxonomic ranks in italics facilitates their quick recognition in scientific papers. IMA fungus 11 (1): 1-5. https://doi.org/10.1186/s43008-020-00048-6
- Thompson C, Phelps K, Allard M, Cook J, Dunnum J, Ferguson A, Gelang M, Khan FAA, Paul D, Reeder D, Simmons N, Vanhove MM, Webala P, Weksler M, Kilpatrick CW (2021) Preserve a Voucher Specimen! The Critical Need for Integrating Natural History Collections in Infectious Disease Studies. mBio 12 (1). https://doi.org/10.1128/mbio.02698-20
- Vassallo V, Felicetti A (2020) Towards an ontological cross-disciplinary solution for multidisciplinary data: VI-SEEM data management and the FAIR principles.

International Journal on Digital Libraries1-11. https://doi.org/10.1007/s00799-020-00285-5

- Wheeler QD, Raven PH, Wilson EO (2004) Taxonomy: impediment or expedient? Science 303 (5656): 285-285. https://doi.org/10.1126/science.303.5656.285
- Wilkinson M, Dumontier M, Aalbersberg I, et al. (2016) The FAIR Guiding Principles for scientific data management and stewardship. Scientific data 3 (1): 1-9. https://doi.org/10.1038/sdata.2016.18

Endnotes

*1 ENVRI is a community of Environmental Research Infrastructures, projects, networks and other diverse stakeholders interested in environmental Research Infrastructure matters. The community also includes e-infrastructures supporting the Research Infrastructures in data solutions.

*2 ENVRI FAIR (2019-2023) project aimed to assist Research Infrastructures in developing a set of FAIR data services, enhancing efficiency, supporting innovation, enabling data- and knowledge-based decisions and connecting the ENVRI Cluster to the European Open Science Cloud (EOSC). The ENVRI cluster, a Science Cluster (SC) dedicated to Environmental Sciences, comprises European RIs developed under the ESFRI (European Strategy Forum on Research Infrastructures) for coordinating long-term initiatives in environmental monitoring and promoting data and resource accessibility at the European scale. In the ecology and biodiversity fields within the ENVRI SC, the following RIs participated in this use-case: AnaEE, eLTER, ICOS, Life Watch, DiSSCo and DANUBIUS.

*3 https://lifewatch.be/en/2022-news-WoRMS-15th-anniversary-story-7: "*Taxonomy is described sometimes as a science and sometimes as an art, but really it's a battleground*".