Ontological modeling of dynamic biodiversity consensus

Robert Hoehndorf[1], Andra Waagmeester[2]

[1]King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
[2]Micelio, Ekeren, Belgium

Abstract
The digitization of biodiversity in under-explored environments, such as the Rub' al Khali (Empty Quarter), relies increasingly on citizen science platforms like iNaturalist. However, the data produced is not static; taxonomic identifications evolve through community consensus, creating a provenance challenge for the Semantic Web. Here, we developed a generalized, configurable workflow and formal OWL 2 DL ontology aligned with the Semanticscience Integrated Ontology (SIO) to model how consensus about taxonomy of observations is reached. We utilized the Rub' al Khali project as a primary case study to demonstrate a system that integrates iNaturalist data with the NCBI Taxonomy to detect epistemic conflicts between agents. Furthermore, we established semantic links to external repositories, utilizing OpenStreetMaps to map taxa to Environment Ontology (ENVO) classes and UniProt to retrieve functional traits, such as heat-shock proteins relevant to desert adaptation. We separated the TBox (consensus logic) from the ABox (observation data), enabling automated reasoning over conflicting evidence and allowing cross-domain queries in the Linked Open Data cloud. Data and source code are available at https://rub-al-khali.bio2vec.net/.

Keywords
OWL-DL, SIO, Biodiversity, Rub' al Khali, Conflict Detection

1. Introduction

The Rub' al Khali, the world's largest sand desert, represents a significant data void in global biodiversity monitoring. To address this gap, we established a digitization project on iNaturalist seeded by research expeditions [1]. While effective for data mobilization, the platform's consensus mechanism, where an observation's identity "flips" based on user voting, presents a semantic challenge. Existing Darwin Core mappings [2] capture only the snapshot of the current state, losing the history of disagreement essential for scientific inquiry. Therefore, we developed a solution utilizing OWL 2 DL aligned with the Semanticscience Integrated Ontology (SIO) [3]. By strictly separating the ontological schema (TBox) from the instance data (ABox), we enable automated reasoning to detect logical inconsistencies in taxonomic assertions. Although developed for the Rub' al Khali study, the architecture is generalized to support any iNaturalist project.

SWAT4HCLS 2026: Semantic Web Applications and Tools for Health Care and Life Sciences, March 23–26, 2026, Amsterdam, The Netherlands
robert.hoehndorf@kaust.edu.sa (R. Hoehndorf); andra@micelio.be (A. Waagmeester)
0000-0001-8149-5890 (R. Hoehndorf); 0000-0001-9773-4008 (A. Waagmeester)

Provenance and evidence are fundamental to scientific rigor, ensuring that data can be traced to its origin and that conclusions are grounded in verifiable facts [4, 5, 6]. They are also important for improving reproducibility. In biodiversity science, these concepts occur, among others, in taxonomic identification, where a specimen or observation serves as the physical evidence supporting an assertion of species identity. However, the rise of citizen science platforms like iNaturalist (https://www.inaturalist.org/) [7] has introduced a new paradigm where identification is not a static expert determination but a dynamic social process [8]. On these platforms, identifications evolve through community consensus; an observation's classified identity may "flip" based on user discussions and changing taxonomies.

This fluidity presents a semantic challenge. Current state-of-the-art provenance models, such as the Provenance Ontology (PROV-O) [9], and biodiversity standards like Darwin Core [2], are largely designed to capture static snapshots of information. They adequately model *that* an identification occurred but do not model the *change* in consensus over time or the conflicting nature of simultaneous assertions. Furthermore, these models do not explicitly distinguish between "ontology" [10, 11] (i.e., the consistent reality of biological entities) and "epistemology" [12, 13] (the often contradictory claims agents make about that reality). While the biological world must be logically consistent (an organism cannot simultaneously belong to two disjoint species) the epistemic layer, i.e., human assertion, often have many contradictions, in particular in a social media context.

To accurately model such a dynamic set of assertions, we believe that a system should satisfy three high-level requirements. First, the system should track the history of statements agents make about taxon identifications (i.e., asserting that observation o belongs to taxon t), identifying which statement is currently "active" or "actual". Second, the system must separate the "statement" (the claim) from the "reality" (the biological truth), allowing for the representation of conflicting claims without corrupting the logical consistency of the underlying ontology. And third, the model should have a mechanism to relate statements to the states of affairs or facts that would have to hold true in the world if the statements were accurate, thereby exposing contradictions when they occur.

Here, we present an application note describing a deployed semantic system rather than a theoretical contribution. Our system addresses the three requirements listed and primarily applies it to the Rub' al Khali (Empty Quarter) desert as example. We developed a configurable workflow and OWL 2 DL ontology aligned with the Semanticscience Integrated Ontology (SIO) [3]. We implement requirements (a) and (b) by modeling identifications as reified processes within SIO, separating the TBox (consensus logic) from the ABox (observation data). We satisfy requirement (c) utilizing Semantic Web Rule Language (SWRL) [14] rules to test for contradictory statements, enabling the automated detection of epistemic conflicts between agents. Although we use the Rub' al Khali desert as example, we provide a general structure for integrating shifting taxon classifications and conflicting evidence into the Linked Open Data [15] cloud.

2. Ontology Design

To capture the changing nature of biodiversity consensus in a social network like iNaturalist, we formalized the identification process not as a static attribute of an observation, but as a

distinct event in time. We grounded this model in the Semanticscience Integrated Ontology (SIO) [3], treating the act of identification as a process that generates an epistemic assertion about the world.

2.1. Formalizing Assertions as Processes

In our model, an identification is a reified entity classified as an `IdentificationAct`, a subclass of `sio:process`. This distinguishes the *assertion* (what an agent claims, the output of the process) from the *reality* (the classification of the biological organism, independent of any identification acts or assertions about it). We defined the structure of these acts using the following SIO relations:

- **Agency:** The act is performed by an `sio:Agent` (mapped to iNaturalist users), and linked to the identification act using `has agent`.
- **Targeting:** The process operates on a specific physical entity, the `sio:Observation`, linked via `has target`.
- **Output:** The outcome of the process is a `sio:Taxon` class, representing the agent's classification claim, linked via `has output`.

This formalization allows multiple agents to make distinct, potentially contradictory assertions about the same observation without creating logical inconsistencies in the ABox. The identification act process acts as the reification of the statement made by the agent (that a certain observation is of a certain taxonomic class). While SIO provides classes for "statement" or "proposition" as well, we found it not necessary to use them; the process serves as the "source" of the statement made by the agent, and it clearly locatable in time.

2.2. Causal Chains and Active Status

To reconstruct the history of consensus, we modeled the temporal sequence of identifications as a causal chain. We introduced the property `is successor of` to link each new identification act to the one immediately preceding it for the same observation. This creates a linked list of assertions that preserves the complete provenance of the debate.

Within this chain, we assign that class `ActiveIdentification` only to the final node in the sequence. An identification is considered active if and only if it has not been superseded by a subsequent act by the same agent. This distinction is critical for reasoning. While the ontology stores all historical identifications, conflict detection operates only on assertions that hold the `ActiveIdentification` status.

2.3. Taxonomic consistency and conflicting assertions

We used a hybrid taxonomy to ground taxon assertions. We combined the iNaturalist hierarchy with the NCBI Taxonomy to form a backbone of class subsumptions. To detect contradicting assertions, we generated `isIncompatibleWith` assertions for all diverging branches of the taxonomic tree.

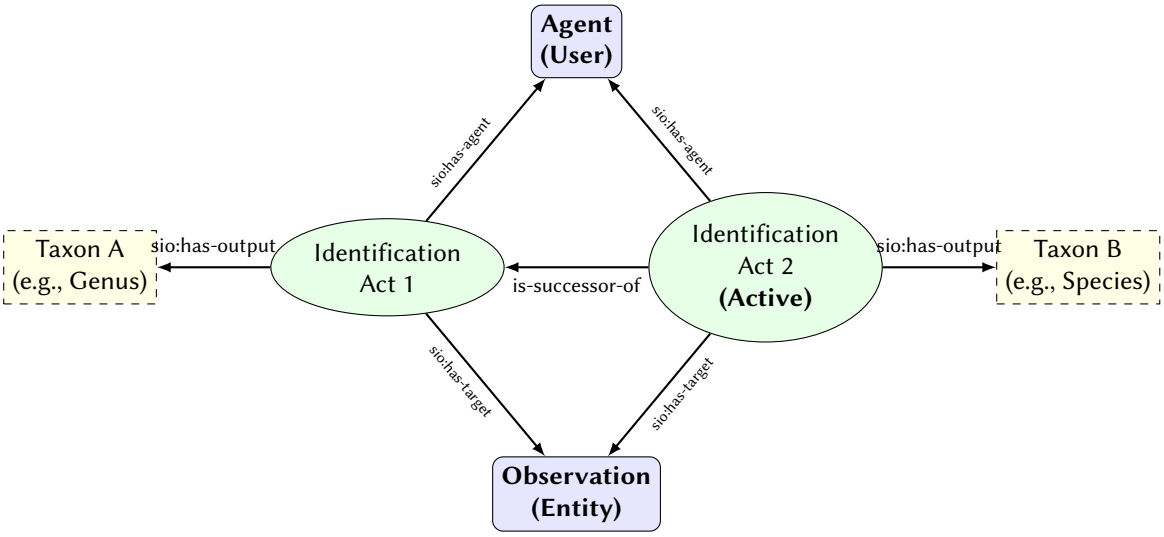

Figure 1: The causal model of identification. Agents perform identification acts that target an observation and output a taxon. The `is-successor-of` property creates a provenance chain, allowing the system to distinguish superseded assertions from the currently **Active** identification.

After we materialized the incompatibility statements, we implemented conflict detection using SWRL rules that target the intersection of the epistemic (active identifications) and ontological layers. A conflict is flagged if and only if:

1. Two distinct `IdentificationAct` individuals target the same observation.
2. Both acts possess the `ActiveIdentification` status.
3. The acts output Taxon classes that are explicitly disjoint via `isIncompatibleWith`.

The corresponding SWRL rule is implemented like this:

$$\begin{aligned} \texttt{ActiveID}(?a1) \land \texttt{ActiveID}(?a2) \land \texttt{target}(?a1, ?o) \land \texttt{target}(?a2, ?o) \\ \land \texttt{output}(?a1, ?t1) \land \texttt{output}(?a2, ?t2) \\ \land \texttt{isIncompatibleWith}(?t1, ?t2) \\ \rightarrow \texttt{ConflictingObservation}(?o) \end{aligned} \tag{1}$$

3. Semantic integration

To extend the utility of the dataset beyond simple occurrence records, and to demonstrate the ability to integrate multiple different datasets, we established semantic links to external ontologies. These links bridge the gap between taxonomic names, environmental context, and functional traits. We developed a multi-stage workflow to resolve entities against the NCBI Taxonomy [16] and the Environment Ontology (ENVO) [17], and used the resulting identifiers to federate queries with UniProt [18].

We aligned iNaturalist taxa with the NCBI Taxonomy to facilitate cross-domain data interoperability. To achieve this mapping, we first parsed the `ncbitaxon.obo` ontology file to

generate a local index of valid NCBI scientific names and identifiers. Subsequently, we iterated through all unique taxa within the dataset, matching iNaturalist names against this index. For successful matches, we materialized an `owl:Class` corresponding to the NCBI identifier (e.g., `obo:NCBITaxon_1234`), asserting it as a superclass of the local iNaturalist taxon. To preserve the hierarchical integrity of the external ontology, we also retrieved and materialized the direct parent of each mapped taxon from the OBO structure. This approach allows the reasoner to infer disjointness based on the NCBI taxonomy while retaining the granularity of the iNaturalist community taxonomy.

To capture the ecological context of observations without manual annotation, we implemented an automated geospatial lookup utilizing the OpenStreetMap (OSM) Overpass API (https://wiki.openstreetmap.org/wiki/Overpass_API). For each observation possessing geospatial coordinates, we queried the API for surrounding features tagged with `natural` or `landuse` keys within a 100 meter radius. We use a custom mapping logic to translate these OSM tags into formal Environment Ontology (ENVO) classes. For instance, we mapped the tag `natural=sand` to `ENVO:00000115` (sand desert) and `natural=scrub` to `ENVO:00000302` (shrubland). In cases where the API returned no specific features for observations within the project boundaries, we applied a fallback strategy, classifying the location based on a manual lookup table. In particular, we classified all entities in the Rub' al Khali project as `ENVO:00000115` (sand desert).

This process linked specific observations to semantic environment types, and enables queries that filter biodiversity data based on ecological characteristics. The alignment with NCBI Taxonomy allows us to federate queries directly with other knowledge bases that rely on NCBI taxonomy, such as the UniProt SPARQL endpoint. By utilizing the NCBI taxonomic identifier as a shared key, we retrieve functional protein data, such as stress response proteins, associated with the observed taxa. This integration effectively connects the macro-scale observation (organism in a desert) with micro-scale genomic evidence (proteins related to heat resistance), bypassing the need to locally warehouse proteomic data.

```
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>
PREFIX sio: <http://semanticscience.org/resource/>
PREFIX up: <http://purl.uniprot.org/core/>
PREFIX project: <https://rub-al-khali.bio2vec.net/consensus/>

SELECT DISTINCT ?upTaxon ?taxonName ?speciesName ?protein ?functionLabel
WHERE {
  {
    # Step 1: Select Local Taxa (e.g., Genera/Families)
    SELECT DISTINCT ?upTaxon
    WHERE {
      GRAPH <https://rub-al-khali.bio2vec.net/consensus/rub-al-khali> {
        ?idAct a project:ActiveIdentification ;
               sio:has-output ?localTaxon .
        ?localTaxon rdfs:subClassOf ?ncbiClass .
        FILTER(STRSTARTS(STR(?ncbiClass), "http://purl.obolibrary.org/obo/
            NCBITaxon_"))
```

```
        BIND(IRI(CONCAT("http://purl.uniprot.org/taxonomy/", STRAFTER(STR(?
            ncbiClass), "NCBITaxon_"))) AS ?upTaxon)
      }
    }
    # Larger limits will run in timeouts
    LIMIT 5
  }

  # Step 2: Federate to UniProt
  SERVICE <https://sparql.uniprot.org/sparql> {

    # Get the taxon name
    ?upTaxon up:scientificName ?taxonName .

    # Get the sub-taxa (including species) and their names
    ?species rdfs:subClassOf* ?upTaxon .
    ?species up:scientificName ?speciesName .

    ?protein up:organism ?species .

    # We look for proteins classified with "Response to Heat" (GO:0009408) or any
        of its children
    ?protein up:classifiedWith ?goTerm .
    ?goTerm rdfs:subClassOf* <http://purl.obolibrary.org/obo/GO_0009408> .

    # Get the name of the specific function found (e.g., "cellular response to
        heat")
    ?goTerm rdfs:label ?functionLabel .
  }
}
ORDER BY ?upTaxon
```

Listing 1: Federated SPARQL query retrieving taxa observed in desert environments and their associated heat-shock proteins from UniProt.

4. Results: Instantiation (ABox)

We instantiated the system with data from the Rub' al Khali project [19] (https://www.inaturalist.org/projects/rub-al-khali), a research project to catalog microbial biodiversity in the largest continuous sand desert of the world. As of February 2026, the dataset comprises 227 observations and 546 identification acts. The hybrid taxonomy contains 269 distinct taxa, of which 80 were successfully mapped to NCBI IDs, while 14 remained specific to iNaturalist. The remaining taxa serve as intermediate nodes in the hierarchy.

The HermiT reasoner successfully classified the ontology (and the ontology was consistent), detecting **10 conflicting observations**. These conflicts represent scientific disagreement or uncertainty that persists despite community discussion. For example, Observation

obs_321457657 was flagged due to active disagreement between two experts on the specific species classification. Two such conflicts are illustrated in Figure 2.

To automatically detect these conflicts, the algorithm has created and materialized over 33,000 incompatibility assertions. While this increased the size of the dataset substantially, it also ensured that the reasoner could detect conflicts at any level of the taxonomic tree, not limited to direct siblings, and not limited to species-level conflicts.

5. Discussion

5.1. Implementation and LLM Assistance

The implementation and deployment of the software infrastructure described in this study relied on Large Language Model (LLM) assistance. Specifically, we used the Gemini-CLI tool with the `gemini-3-pro-preview` for all coding tasks. We manually reviewed the output, but no source code was directly authored by a human. The complete implementation process, including the website, RDF store and SPARQL endpoint, and containerization, resulted in a functional system within approximately two hours.

We also used Gemini-CLI for the actual deployment of the knowledge graph and website. This process involved connecting to a remote server, obtaining root privileges, and configuring an Nginx web server. Specific configuration tasks included setting up SSL certificates via Let's Encrypt and implementing Linked Open Data (LOD) resolution through custom redirect statements pointing to SPARQL DESCRIBE statements for all coined IRIs. Additionally, the LLM model did the systemd service creation, firewall configuration, and port redirection to prevent conflicts with other services running on the same system. We only performed minor manual interventions during deployment (such as configuring the public key used by the LLM to connect through ssh), and the total deployment time was approximately 10 minutes.

We also attempted to use the LLM for generating SPARQL queries to query the resulting graph. While the model usually produced syntactically correct SPARQL, the queries often contained semantic errors; one main confusion was mixing up queries for super-class and sub-class. Furthermore, the generated queries often exhibited high complexity, resulting in execution timeouts. In particular, the LLM created queries that attempted to federate simultaneously to multiple endpoints, specifically to expand the ENVO hierarchy via Ontobee (to find all specific desert environments) and then retrieving protein functions from UniProt of all the taxa found in the expanded environments. Mostly, we found the LLM-generated queries required some correction before they could be run successfully.

Finally, we also tried to use the LLM for drafting or modifying sections of this manuscript. For example, we used prompts such as "Modify main.tex to describe the changes you just made..."). These attempts all failed to produce scientific text that was acceptable to the authors. Consequently, this manuscript contains no LLM-generated text.

However, what did work well was to *first* write the theoretical concept as precisely as possible (such as the use of the causal chain) and then ask the LLM to implement this based on the written specification. This observation suggests a specific workflow for integrating LLMs into scientific software development, where a specification is written first before any implementation.

5.2. Limitations and future work

One main limitation of our implementation is that the mappings between community-specific terminologies (iNaturalist) and formal ontologies or taxonomies (NCBI, ENVO) remains imprecise. The string-matching approach we implemented has many limitations that have been overcome in more specialized mapping workflows or machine-learning based ontology alignment approaches [20].

Our current method of materializing `isIncompatibleWith` statements for all disjoint taxa scales quadratically. As the taxonomic identifications grow, this approach will become computationally too expensive for storage and querying. One possible improvement could be to create a temporary knowledge base only for consistency checking, using a reasoner to find whether it is inconsistent, and then generate explanations to identify the specific statements causing the inconsistency. This could be further expaned with paraconsistent reasoning or minimal repair strategies to handle contradictions without halting the system.

5.3. Conclusions

We developed a semantic model based on the Semanticscience Integrated Ontology to model biodiversity identification in social media as a dynamic process rather than a static attribute. By separating epistemic assertions from ontological reality, we enabled the automated detection of taxonomic conflicts within the Rub' al Khali dataset and established interoperability with genomic and environmental repositories. Consequently, our semantic model can contribute to changing occurrence records into a living knowledge graph that preserves the history of how scientific consensus is created and prioritizes conflicting identifications for expert review.

Data and software availability

The source code for the consensus workflow, the website, and the ontology files is available at https://github.com/bio-ontology-research-group/inat-consensus under the BSD 2-clause license. The dataset and knowledge graph are hosted at https://consensus.rubalkhali.science/. The iNaturalist project data is available at https://www.inaturalist.org/projects/rub-al-khali.

Declaration on Generative AI

During the preparation of this work, the authors used Gemini 3 Pro Preview through the Gemini-CLI tool to implement and deploy the website. The authors also used the Gemini 3 Pro Preview model to create text content for this manuscript based on the results of the generated code; however, this text content is not included in the manuscript as the authors did not find the created content valuable or reflecting their own views.

Acknowledgements

References

[1] A. Waagmeester, Y. Yamamoto, R. Hoehndorf, D. Steinberg, N. Queralt-Rosinach, J. Koblitz, T. Nakazato, S. Ikeda, DBCLS BioHackathon 2025 report on the WikiBlitz, BioHackrXiv Preprint (2025). URL: https://doi.org/10.37044/osf.io/7s6da_v1. doi:10.37044/osf.io/7s6da_v1.

[2] J. Wieczorek, D. Bloom, R. Guralnick, S. Blum, M. Döring, R. Giovanni, T. Robertson, D. Vieglais, Darwin core: an evolving community-developed biodiversity data standard, PLoS One 7 (2012) e29715.

[3] M. Dumontier, C. J. Baker, J. Baran, A. Callahan, L. Chepelev, J. Cruz-Toledo, N. R. Del Rio, G. Duck, L. I. Furlong, N. Keath, D. Klassen, J. P. McCusker, N. Queralt-Rosinach, M. Samwald, N. Villanueva-Rosales, M. D. Wilkinson, R. Hoehndorf, The semantic-science integrated ontology (sio) for biomedical research and knowledge discovery, Journal of Biomedical Semantics 5 (2014). URL: http://dx.doi.org/10.1186/2041-1480-5-14. doi:10.1186/2041-1480-5-14.

[4] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L. B. da Silva Santos, P. E. Bourne, et al., The FAIR guiding principles for scientific data management and stewardship, Scientific Data 3 (2016) 160018. doi:10.1038/sdata.2016.18.

[5] J. Freire, D. Koop, E. Santos, C. T. Silva, Provenance for computational tasks: A survey, Computing in Science & Engineering 10 (2008) 11–21. doi:10.1109/MCSE.2008.79.

[6] M. C. Chibucos, C. J. Mungall, R. Balakrishnan, K. R. Christie, R. P. Huntley, O. White, J. A. Blake, S. E. Lewis, M. Giglio, Standardized description of scientific evidence using the evidence ontology (eco), Database 2014 (2014) bau075–bau075. URL: http://dx.doi.org/10.1093/database/bau075. doi:10.1093/database/bau075.

[7] iNaturalist, inaturalist: A community for naturalists, https://www.inaturalist.org, 2026. Accessed: 2026-02-16.

[8] T. Lubiana, R. Littauer, S. Leachman, J. Ainali, M. Karingamadathil, A. Waagmeester, H. M. Meudt, D. Taraborelli, Wiki loves inaturalist: How wikimedians integrate inaturalist content on wikipedia, wikidata, and wikimedia commons, Biodiversity Information Science and Standards 9 (2025) e181155. URL: https://doi.org/10.3897/biss.9.181155. doi:10.3897/biss.9.181155. arXiv:https://doi.org/10.3897/biss.9.181155.

[9] T. Lebo, S. Sanguino, D. McGuinness, et al., PROV-O: The PROV Ontology, Technical Report, W3C Recommendation, 2013. URL: https://www.w3.org/TR/prov-o/.

[10] Aristotle, Metaphysics, Clarendon Press, Oxford, 1924. See specifically Book Gamma, 1003a20-32, for the definition of the study of being qua being.

[11] W. V. Quine, On what there is, The Review of Metaphysics 2 (1948) 21–38.

[12] Plato, Theaetetus, Routledge & Kegan Paul, London, 1935. Classically dated c. 369 BCE. See 201c-d for the account of knowledge as true belief with an account (logos).

[13] A. I. Goldman, A causal theory of knowing, The Journal of Philosophy 64 (1967) 357–372.

[14] I. Horrocks, P. F. Patel-Schneider, H. Boley, S. Tabet, B. Grosof, M. Dean, SWRL: A Semantic Web Rule Language Combining OWL and RuleML, W3C Member Submission, World Wide

Web Consortium (W3C), 2004. URL: https://www.w3.org/Submission/SWRL/.
[15] C. Bizer, T. Heath, T. Berners-Lee, Linked data — the story so far, International Journal on Semantic Web and Information Systems (IJSWIS) 5 (2009) 1–22. doi:10.4018/jswis.2009081901.
[16] S. Federhen, The ncbi taxonomy database, Nucleic acids research 40 (2012) D136–D143.
[17] P. L. Buttigieg, N. Morrison, B. Smith, C. J. Mungall, S. E. Lewis, E. Consortium, The environment ontology: contextualising biological and biomedical entities, Journal of biomedical semantics 4 (2013) 43.
[18] S. S. I. of Bioinformatics RDF Group Members, The sib swiss institute of bioinformatics semantic web of data, Nucleic Acids Research 52 (2023) D44–D51. URL: https://doi.org/10.1093/nar/gkad902. doi:10.1093/nar/gkad902. arXiv:https://academic.oup.com/nar/article-pdf/52/D1/D44/55040312/gkad902.pdf.
[19] iNaturalist contributors, Rub' al khali project, https://www.inaturalist.org/projects/rub-al-khali, 2025. URL: https://www.inaturalist.org/projects/rub-al-khali, iNaturalist project page, accessed 21 December 2025.
[20] M. A. N. Pour, A. Algergawy, E. Blomqvist, P. Buche, J. Chen, P. G. Cotovio, A. Coulet, J. Cufi, H. Dong, D. Faria, L. Ferraz, S. Hertling, Y. He, I. Horrocks, L. Ibanescu, S. Jain, E. Jiménez-Ruiz, N. Karam, F. Kraus, P. Lambrix, H. Li, Y. Li, P. Monnin, H. Paulheim, C. Pesquita, A. Sharma, P. Shvaiko, M. Silva, G. Sousa, C. Trojahn, J. Vataščinová, B. Yaman, O. Zamazal, L. Zhou, Results of the Ontology Alignment Evaluation Initiative 2024, in: E. Jiménez-Ruiz, O. Hassanzadeh, C. Trojahn, S. Hertling, H. Li, P. Shvaiko, J. Euzenat (Eds.), Proceedings of the 19th International Workshop on Ontology Matching (OM 2024) co-located with the 23rd International Semantic Web Conference (ISWC 2024), volume 3897 of *CEUR Workshop Proceedings*, CEUR-WS.org, Baltimore, MD, USA, 2024, pp. 64–97. URL: https://ceur-ws.org/Vol-3897/oaei2024_paper0.pdf.

(a) Disagreement: *Heliotropium arbainense* vs. *H. digynum*.

(b) Disagreement: *Haloxylon persicum* (Amaranthaceae) vs. Genus *Calligonum* (Polygonaceae).

Figure 2: Examples of taxonomic identification conflicts in citizen science data of the Rub' al Khali project. (a) A species-level conflict within *Heliotropium*. (b) A high-level disagreement spanning distinct families (Amaranthaceae vs. Polygonaceae). From the photographs it is obvious that the images were taken by non-experts and lack the specific diagnostic features that trained botanists or ecologists would focus on. This reduced information contributes to the ambiguity and the magnitude of the observed classification discrepancies.