TraitBank: Practical semantics for organism attribute data

Cynthia S. Parr[§a], Nathan Wilson[b], Katja Schulz[a], Patrick Leary[b], Jennifer Hammock[a], Jeremy Rice[b], Robert J. Corrigan Jr.[a]

[a]*National Museum of Natural History, Smithsonian Institution, Washington, DC*
[b]*Marine Biological Laboratory, Woods Hole, United States of America*
[§]*Corresponding author, parrc@si.edu*

Abstract. Encyclopedia of Life (EOL) has developed a new web-accessible repository, TraitBank (http://eol.org/traitbank), to better serve scientific discovery. EOL's TraitBank aggregates and manages attribute (trait) data across the tree of life including morphological descriptors, life history characteristics, habitat preferences, and interactions with other organisms. This paper describes how TraitBank uses Darwin Core and other standards to ingest and subsequently manage trait data in a Virtuoso triple store in a way that leverages EOL's extensive existing infrastructure. We add to and improve the semantics of both data and metadata in order to improve interoperability across the domains of morphology, ecology, and genomics. The system takes a semantic approach and also emphasizes practicality and ease of use for experts and non-experts. In addition to aggregating trait data in existing literature or databases, TraitBank contributes to community-based ontologies and sets the stage for a rapid rise in annotations about attributes on specimens and citizen science observations.

Keywords: biodiversity, ontologies, Semantic Web, traits, ecology, evolution, taxonomy, aggregation

1. Introduction

The word "trait" can broadly include any measurable characteristic, phenotype, property, or attribute of individuals or groups of the same taxon (type of organism). Over the history of biological description, biologists have used very different approaches to capturing and managing this kind of information for their own analyses of morphology, behavior, life history, and ecological interactions. Most descriptions exist primarily in free text or data tables in published papers, though some communities are starting to annotate those papers [1], extract information from text [24], and build special-purpose databases, e.g., TRY for plants [13], SeaLifeBase[1] for marine organisms. Increasingly, researchers are archiving datasets associated with published studies in open data repositories such as DRYAD [26]. While this is a critical development, relatively unstructured and idiosyncratic metadata about attribute definitions means that individual data items in these datasets are not easily discovered or repurposed. Adding to the substantial legacy of descriptive data, with its various degrees of "computability" and discoverability, recent efforts to automate the process of description and measurement are accelerating the pace of data generation, e.g. [3, 12] .

Many people including the phylogenetic, ecological, and conservation research communities will need effective ways to discover and consume available data in the coming era of data-intensive science. For example, researchers and educators in marine environmental modeling need high-quality inputs about large numbers of species in order to understand the current and historical distributions of species; how these distributions are impacted by environmental changes such as climate change, overharvesting, or invasive species; how biological communities function to provide ecosystem services; and what could happen to these services under future scenarios that change the composition of these communities.

Other projects (e.g., Ocean Biogeographic Information Systems [2], Global Biodiversity Information Facility[3]) compile occurrence-level or abiotic information necessary to address these kinds of questions. However, no project has provided trait

---

[1] http://sealifebase.org

[2] http://iobis.org/
[3] http://gbif.org

data on a global scale and across the major organismal groups. These characteristics, summarized at the species and clade level, are needed to enable large scale modeling of biological communities [10]. They have also been identified by DIVERSITAS and the Global Earth Observation Biodiversity Observation Network (GEO BON) as likely to be required by the Intergovernmental science-policy Platform on Biodiversity and Ecosystem Services (IPBES) [9, 21]. Aggregating and standardizing these data, making them freely re-usable, and providing discovery mechanisms for them facilitates rapid analyses for these urgent problems.

This paper describes TraitBank™, a system designed by the Encyclopedia of Life (EOL) team to address these problems and to integrate seamlessly with the existing EOL platform [18], which provides text and multimedia about taxa across the entire tree of life – nearly two million species including bacteria, archaebacteria, plants, fungi, and animals.[4] It first describes our approaches to semantics and usability, then details TraitBank's implementation, and finally evaluates the system with respect to

system performance, implications for interoperability, and impact on community and provider processes.

2. Approach

Our overall approach to TraitBank is consistent with EOL's philosophy of being a Content Creation Community [23]. Thus, the goal was to leverage EOL's existing, traditional relational database framework as much as possible, aggregating and adding value to data from partners through semantic integration and curation. When adding value, semantics from existing standards and ontologies are to be used as much as possible. At the same time, the goal is to avoid heavy reasoning in order to achieve lightweight integration and leave more principled inference to end users who may have highly specific goals. Emphasis is to be on usability for both scientific experts and non-expert users.

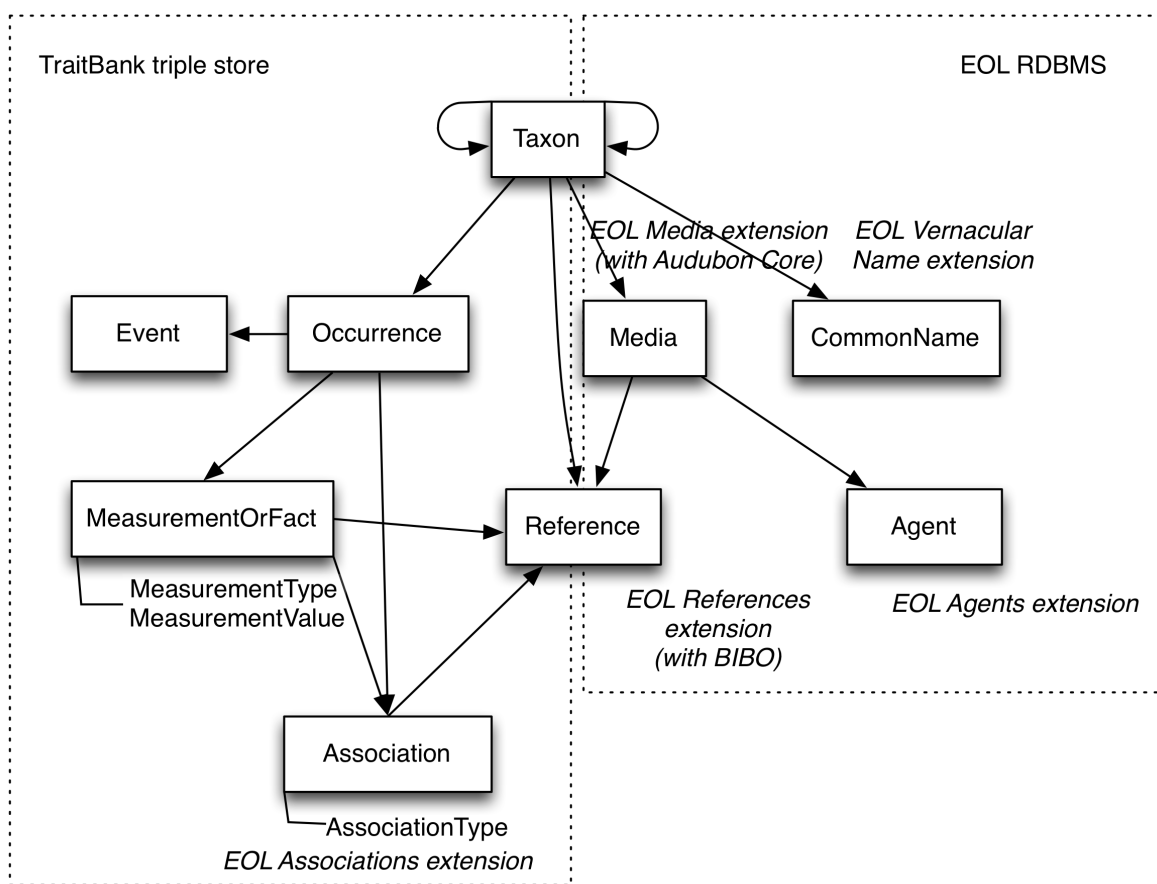### 2.1. General data model

Fig. 1 Data model and architecture for TraitBank and EOL. All elements are from Darwin Core unless an EOL-specific extension is specified. Only important properties are indicated. TraitBank elements may hold only pointers to elements managed in the EOL relational database management system, like names and references.

To represent trait data, we chose to both use and extend Darwin Core [27] (Figure 1). At the heart of each trait record is an `Occurrence`, where the identity of the taxon and context in which the trait was observed or measured may be recorded (e.g. geospatial information, dates, life stages, individual counts). The Darwin Core field `MeasurementOrFact` holds the basics of the trait measured and some other metadata. In particular, `MeasurementType` describes what was measured (ideally, a Uniform Resource Identifier, URI, for an ontology term) and `MeasurementValue` holds a number or a term from a controlled vocabulary or ontology. Measurement metadata might include, for example, `Unit` (from the Units of Measurement Ontology, UO [5]), `Accuracy`, and `MeasurementMethod`.

Interactions among species, for example predator-prey relationships, are handled using a new Darwin Core extension named Associations. This extension is similar to `MeasurementOrFact,` but with `AssociationType` indicating the type of relationship among taxa, (eg: X feeds on Y) and values are references back to other rows in the Occurrence extension.

If the `MeasurementType` involves a statistical operation, e.g. mean or maximum, this is indicated as metadata with a statistical modifier field usually referencing the Semanticscience Integrated Ontology (SIO)[6]. Similarly, if the measurement is specific to a life stage or sex these are also indicated using the Phenotypic Quality Ontology (PATO) [7] and the Uber Anatomy Ontology (UBERON) ontologies [16]. Thus, even though there are a variety of ways that body mass is reported by sources (means, extreme highs, extreme lows, for adults or juveniles, for males or females) we decided that all should share the same `MeasurementType` in TraitBank.

In some cases, Occurrences are part of a multi-occurrence `Event` with its own metadata. In general, this approach should accommodate not only specimen- or individual-observation-level measurements but also measurements summarized from literature or large datasets.

As with other content on EOL, provenance in TraitBank is handled using rich attribution metadata via fields from Dublin Core[7] and Darwin Core, with structured references supported using an internal EOL ontology based on the Bibliographic Ontology (BIBO)[8].

2.2. Taxonomic semantics

EOL manages multiple dynamic biological taxonomic hierarchies. Rather than attempt to fully capture this complex interaction semantically [6], the TraitBank system reflects data structures already developed to represent multiple classifications within the Encyclopedia of Life relational database [18]. Within TraitBank, scientific names are designated with the Darwin Core property `scientificName` and are typically associated with Taxon URIs that have the rdf:type of `Taxon`. These in turn are associated using the `taxonConceptID` predicate to an EOL taxon page URL, e.g., http://eol.org/pages/328615. These Taxon URIs serve both to associate a data point with a particular page and to describe the parent/child relationships between the taxa. The parent/child relationships use the `parentNameUsageID` predicate.

2.3. Semantics of attributes

At least initially, most data aggregated by TraitBank are not expected to be "born semantic – they will come from literature sources or conventional databases. Here, semantics will need to be added by applying formally-defined terms *post hoc*. As EOL's scope includes all organismal diversity and aims to cover whatever attributes are needed by its users, the team did not want to limit the system to a pre-determined set of attribute URIs in a formal ontology we would maintain. Rather, we decided on a strategy of seeking URIs as much as possible from existing ontologies for the type of the attribute (e.g. habitat descriptors from the Environments Ontology[9]) and where possible, for value of attributes (e.g., a particular type of habitat) as well as for most metadata describing the context of the measurement (e.g., life stage). As large datasets are ingested, new terms may be proposed by EOL staff to managers of existing ontologies. If these changes are not made immediately, EOL will create and define provisional terms. The intention is to ensure that new terms discovered in source datasets become part of the most relevant ontologies where they can be managed in the long term by domain experts and readily discovered by the largest number

of users and semantic web developers. EOL staff will also mint terms for concepts that are similar to but not exact matches to existing terms. Datasets that are already available with URIs will be preserved when harvested by EOL for TraitBank.

2.4. Reasoning

Because of the complexity of semantic reasoning and the challenges of reasoning across highly heterogeneous or web-scale datasets [20, 25] we decided that usage of semantic reasoning would be limited at first, with additional reasoning to be added later as the system matures and as demand requires. For example, conversion relationships of units (e.g. from g to kg), logarithmic transformations, and inverse relationships (for example, `preysUpon` and `hasPredator`) are straightforward. Eventually, reasoning could be expanded to infer values based on phylogeny, or to leverage semantic similarity for searches. We plan to use iterative processes to relate new and provisional terms to each other and to existing ontologies.

2.5. Usability

TraitBank's initial use case was integration into Encyclopedia of Life pages, so interface design had to suit the general users who are currently EOL's primary audience. However, expert biologists are a major target audience and they will need more detailed metadata to determine fitness for research use. Therefore, the structured data interfaces on EOL were designed to provide obvious overviews of "quick facts" and term definitions for non-scientists, while still giving scientists access to highly technical metadata, including rich provenance.

To integrate well with the EOL user experience, TraitBank attributes and their definitions need to be organized and displayed in a meaningful way on EOL webpages, whether they are from an external ontology or are provisional. Thus, EOL site administrators will need a to 1) import ontology term labels and definitions into the EOL system [4], 2) add to these as necessary, 3) specify EOL-specific groupings of terms and the order of display. EOL curators (thousands of volunteers with additional control over site content) will also need well-designed tools to hide erroneous values and select values to highlight on the site (e.g., in a box on the taxon page Overview tab).

3. Implementation

To ensure that TraitBank functionality would meet the needs of the scientific community, and to build a stakeholder base of users ready to use the new functionality, we convened workshops and an advisory panel. Scientists who attended workshops sponsored by EOL's Biodiversity Synthesis Center at the Field Museum over a period of four years have generated high-level community requirements. A workshop in Washington, DC September 2012 brought together more than twenty experts from biology and computer science, including semantics, to focus specifically on computability opportunities and requirements for future research and development. Finally, quarterly teleconferences with an 11-person panel of scientists and technologists drawn from the above workshops informed iterative design and development during implementation.

3.1. Architecture and technology

TraitBank is built on the RDF triple store integrated into the open source edition of the OpenLink Virtuoso Universal Server [10]. This datastore is accessed by EOL's application servers and backend data harvesting engine [18]. We considered neo4j [11] however, based on internal testing, using an RDF triple store made it easier to import and blend standard URI-based ontologies, URIs provided by content partners, and when necessary newly minted EOL URIs. The SPARQL[12] query language works well to efficiently query complex chains of relationships including recursive queries needed for traversing taxonomic hierarchies. All code is available under an MIT open source license.[13]

3.2. Data import and sources

Initial datasets (summarized in Table 1) were chosen to represent a large number of taxa and attribute types, with special attention to attributes that would be beneficial for marine biodiversity science. Most data are imported from other databases via PHP connectors or uploaded to the repository via Darwin Core Archive files with extensions for measurements

---

[10] http://virtuoso.openlinksw.com/
[11] http://www.neo4j.org/

[12] http://www.w3.org/TR/rdf-sparql-query/
[13] https://github.com/eol

Table 1. TraitBank contents as of 14 March 2014 as retrieved from http://eol.org/statistics. Trait types include both MeasurementTypes and AssociationTypes.

Datasets	30
Trait types	277
Individual data records	3,827,779
Taxa with at least one data record	383,297
Total triples	72,135,629

and facts, associations, occurrences, events, and references. Manual bulk uploads are supported by a custom spreadsheet template that can easily be converted to a Darwin Core Archive.

Most datasets were not previously annotated with ontology terms so EOL staff sought appropriate terms using tools such as the Bioontologies Bioportal,[14] Ontobee [28], and Gramene[15]. EOL staff created some new terms by using the post-composition tool Term Genie[16] to define attributes by combining classes from PATO and GO or UBERON. For example, `WoodDensity` and `CellShape` are two new terms now available in the Ontology of Biological Attributes (OBA). Table 2 lists some of the ontologies from which we have drawn terms.

Of 277 attribute types, 184 were provisional terms created by EOL staff, more than originally anticipated. In part this is because most of our data sets focus on life history and ecology where there is currently little ontology coverage. In part this was because existing terms were often not defined exactly as needed. For example, `Head-Body Length`[17] is

a concept with a more specific definition than `Body Length`.[18] Rather than lose semantics, or assume an rdf:sameAs relationship, we opted to create a new term that could be related to an existing term in the future. One dataset, Polytraits, was already annotated with URIs [5].

Most locations provided as text strings in data records were ingested only as text strings, though we plan to attempt to convert these to URIs from GAZ[19] supplemented with Geonames[20] in the future.

To facilitate filtered searches by taxon, we populated TraitBank with parent/child relationships from the National Center for Biotechnology Information (NCBI)[21] and Catalogue of Life [22] classifications currently in EOL. Names at all levels of these classifications were included.

Direct addition of data records through the web interface has been implemented for staff administrators; further improvement is required before this feature can be made public.

3.3. Data access

TraitBank data is displayed on the Overview (e.g., http://eol.org/pages/328615/overview) and Data tabs (e.g., http://eol.org/pages/328615/data) for each EOL Taxon Page, and is made available through the EOL data search facility[22] and through data downloads initiated from the search facility.

Table 2. Ontologies used in TraitBank

Subject Areas	Ontology	Example terms
Statistics	Semanticscience Integrated Ontology (SIO)	mean, minimal value, standard deviation
Units of measure	Units of Measurement Ontology (UO) [8]	meter, years, degree Celsius
Habitat information	Environments Ontology (EnvO)	wetland, desert, snow field
Attributes of organisms	Phenotype Quality Ontology (PATO) [15]	aerobic, conical, evergreen
Plant attributes	Plant Trait Ontology (TO) [11]	flower color, life cycle habit, salt tolerance
Animal attributes	Vertebrate Trait Ontology (VT) [17]	body mass, total life span, onset of fertility
Animal natural history	Animal Natural History and Life History Ontology (ETHAN) [19]	nocturnal, oviparous, scavenger

---

[14] http://bioportal.bioontologies.org/
[15] http://archive.gramene.org/plant_ontology/#to
[16] http://www.berkeleybop.org/software/termgenie
[17] http://eol.org/schema/terms/HeadBodyLength

[18] http://purl.obolibrary.org/obo/VT_0001256
[19] http://bioportal.bioontology.org/ontologies/GAZ
[20] http://www.geonames.org/
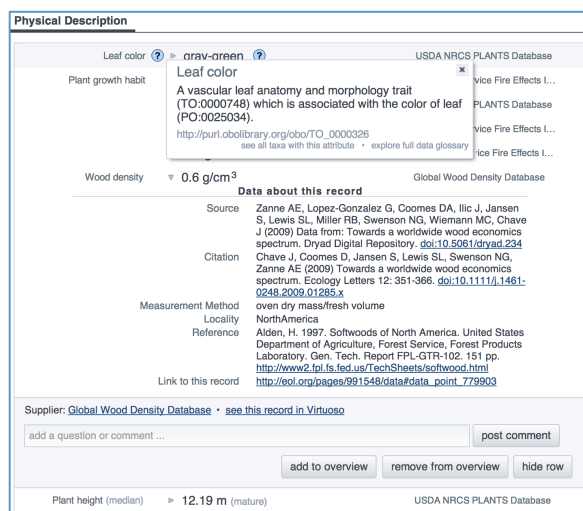[21] http://www.ncbi.nlm.nih.gov/taxonomy/
[22] http://eol.org/data_search

Fig. 2. Part of a data tab in EOL. Wood density is expanded to show rich metadata. User selects the ? to see a definition of Leaf Color

Fig. 3. Data search interface for TraitBank.

The Overview tab is the information center of each EOL Taxon Page. In addition to a brief description, images, a map, and a taxonomic hierarchy, this tab features a selection of trait data, with a link to the more comprehensive data presentation in the Data tab. The default view of the Data tab shows a simple list of trait labels, values, and data providers, ordered by subject (Distribution, Physical Description, Ecology, etc.). A dynamic user interface (Figure 2) gives access to the metadata for each record (provenance, methods, life stage, sex, sample size, etc.), curation and commenting tools (see below), and URIs and definitions for traits, metadata types, and categorical data values. For higher taxa (genera, families, etc.) a Data Summaries subtab provides the range of values represented in TraitBank for key physical, ecological, and life history traits of taxonomic children.

TraitBank provides a basic search interface (Figure 3) prompting users to select an attribute type for searching. The user may refine the search by choosing a taxonomic group, and/or by specifying a value or range of values for the attribute. Filtering by taxonomic group is inclusive, so if either NCBI or the Catalog of Life classifications indicates that a taxon is a child in the selected group, then any associated data records are included in the search results. While scientific names are stored in the TraitBank triple store, the resolution of synonyms and homonyms continues to be handled by previously developed EOL systems which primarily rely on a MySQL database. A button in the search interface allows users to initiate a job compiling
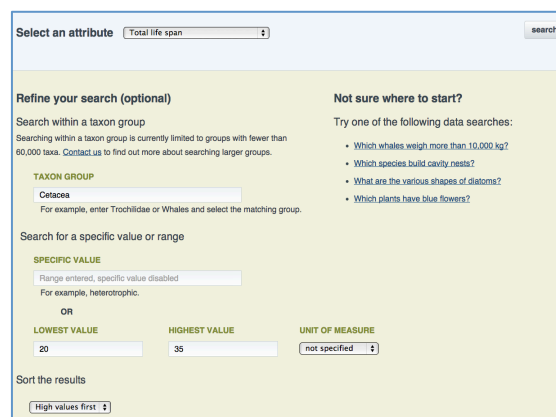
search results into a download file for future retrieval.

Download files are formatted as CSV (comma separated values) tables making it easy to load into common spreadsheet applications such as Microsoft Excel or to parse in any programming language. Each data row includes a unique EOL identifier for the associated taxon along with its scientific name and one common name if available. Parent/child information is not provided. Each data row specifies the term label (e.g., "body mass"), the value (38.5), and units (kg) when appropriate. Many unit types are automatically normalized into comparable values. However, the raw value and units are also provided. All corresponding URIs are provided. Finally, the metadata includes the provenance information along with information about how the measurement was made and appropriate context information such as life stage or geographical location.

Most EOL Science Advisors expressed strong interest in data downloads in a simple comprehensible format such as CSV so this was the first machine-readable format developed. There has also been significant interest in a web-based application programming interface (API) to support simple data-driven web-applications, for example using JSON-LD [14]. Other methods of exposing TraitBank data as linked open data [2] or as a SPARQL endpoint are under consideration.

3.4. Data curation

Any registered EOL member can review TraitBank content and report problems by adding comments to individual data records. EOL also has over 1,000 volunteer curators with "full curator"

status based on their professional/scientific credentials. These curators have the power to remove incorrect or suspect TraitBank records from public view. Flagged records remain visible to other curators and can be restored if flagged in error. EOL comments and curator actions are generally reported to data providers, but this feature is still under development for TraitBank records. Once it has been activated, TraitBank data providers will benefit from the quality control activities of the EOL community. EOL curators also participate in the selection of trait data for the Overview tabs of individual taxon pages. This activity is particularly important to ensure that the most interesting and informative records are highlighted for taxa of interest to a wide audience.

3.5. Term management and data glossary

TraitBank terms, their definitions, URIs, and other relevant metadata are maintained through an admin interface (the Known URIs tool) that supports the categorization of terms as measurement types, values, associations, metadata and the mapping of measurement types to subjects (e.g., Ecology, Life History and Behavior). Administrators can also rank terms to prioritize the default display sequence on Overview and Data tabs and can designate key terms to be featured in Data summaries for higher taxa. Administrators can create a limited set of semantic relationships between terms (e.g., equivalent, inverse). This term management system feeds a full data glossary[23] with definitions, URIs, and quick links to simple queries for all known TraitBank traits, trait values, association types, and metadata. In addition to this full glossary, the Data tab of each taxon page features a filtered, taxon-specific glossary of relevant terms.

Using the Known URIs tool we also added local definitions to those imported URIs that were not defined in their source ontologies. We recognize this is not optimal but wanted users to find definitions for all TraitBank terms in the data glossary.

4. Evaluation and Conclusions

TraitBank was released after a limited private beta test followed by a public beta period, with each test followed by a survey. In addition, there were informal demonstrations to communities at several conferences. In general, utility and performance met expectations. A formal evaluation has not been conducted but we present here preliminary findings.

4.1. System performance

Early in the development of TraitBank large, query-based dataset downloads proved impossible to handle interactively. We therefore implemented a background queuing system that allows users to request download files to be generated asynchronously. These requests are processed behind the scenes and when finished, the requesting user is emailed a link to the resulting file. The number of results was not the only limiting factor; queries which filter by large taxonomic groups (> 100,000 members, including sub-species and higher taxa) were unacceptably slow for interactive display often taking more than a minute and resulting in a timeout. As a result queries were optimized and filtered searches were limited to groups with less than 60,000 members. This allowed searches for some large groups of particular interest such as the birds (Aves, currently 42,263 members) and mammals (Mammalia, currently 32,863 members). For all attributes, it is still possible to display example results across the entire tree of life. However, in some extreme cases (e.g., 'habitat') downloads of these data can take hours to prepare.

4.2. Implications for interoperability

TraitBank fosters semantic interoperability both within a domain and across domains by preferring to apply to its ingested datasets URIs from existing ontologies which are used in other systems. Use of semantic technologies is already more prevalent in genomics, morphology, ecology, and developmental biology, so it makes sense to link newly exposed or annotated trait information to these efforts. On the other hand, where these systems do not yet capture knowledge adequately (e.g. missing terms, missing relations, missing definitions, complex taxonomic and nomenclatural semantics), our approach still allows progress in knowledge management and sharing in the most practical sense, even if not all elements of the system are interoperable.

Of particular note is our use of statistical modifiers. There are myriad ways and scales at which organismal measurements are made and reported as their source studies require. Summary statistics or data ranges may or may not be accompanied by the

---

[23] http://eol.org/data_glossary

individual measurements upon which they are based. Our approach offers the advantage of normalizing to common terms, which are then available for discovery and reasoning, without losing the semantics of the statistical operations involved in each study.

4.3. Impact on community and provider processes

Our data model is compatible with existing Darwin Core-based data sharing processes. Thus, as museum collections (for example) annotate their specimens with measurements, this information can be included as `MeasurementsOrFacts` or `Associations` in the Darwin Core Archives they make available for indexing by the Global Biodiversity Information Facility.

Similarly, citizen science observation projects can (and some iNaturalist[24] Projects already do) ask observers to provide more than just the basic time-and-place information about the organisms they are seeing. Each project may choose to control the `MeasurementTypes` used for annotation; existing usage on EOL will allow projects to identify and choose commonly used `MeasurementTypes`.

Many of the datasets aggregated by TraitBank contained traits and trait values that could not be mapped to current ontologies. This suggests that TraitBank has the potential to substantially contribute to meaningful growth in community-maintained domain ontologies. We anticipate the use of ontologies in TraitBank will result in increased or expanded usage of ontologies in research applications.

For specific providers such as Global Biotic Interactions[25], PolyTraits[26], and EOL-Environments[27] TraitBank provides a live example of re-use that exposes their data to broader audiences and promotes significant community curation.

The TraitBank system represents initial steps towards scalable methods for mobilizing and integrating organism attribute data so that it can be discovered and re-used for a wide range of use cases, from simple fact-finding to "big data" modeling studies. Future work will focus on exposing TraitBank data in more advanced machine-readable formats, replacing provisional EOL terms with community-managed terms, and exploring the best

use of reasoning within the EOL-TraitBank framework.

5. Acknowledgements

6. References

[1] Balhoff, J. P., Dahdul, W. M., Kothari, C. R., Lapp, H., Lundberg, J. G., Mabee, P., … Vision, T. J. (2010). Phenex: Ontological Annotation of Phenotypic Diversity. *PLoS ONE*, *5*(5), 10.

[2] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked Data - The Story So Far. *International Journal on Semantic Web and Information Systems*, *5*(3), 1–22. doi:10.4018/jswis.2009081901

[3] Burleigh, J. G., Alphonse, K., Alverson, A. J., Bik, H. M., Blank, C., Cirranello, A. L., … Yu, M. (2013). Next-generation phenomics for the Tree of Life. *PLoS Currents*, (JUNE).

[4] Courtot, M., Gibson, F., Lister, A. L., Malone, J., Schober, D., Ruttenberg, … Ruttenberg, A. (2011). MIREOT: The minimum information to reference an external ontology term. *Applied Ontology*, *6*, 23–33. doi:10.3233/AO-2011-0087

[5] Faulwetter, S., Markantonatou, V., Pavloudi, C., Papageorgiou, N., Keklikoglou, K., Chatzinikolaou, E., … Arvanitidis, C. (2014). Polytraits: A database on biological traits of marine polychaetes. *Biodiversity Data Journal*, *2*, e1024. doi:0.3897/BDJ.2.e1024

[6] Franz, N. M., & Thau, D. (2010). Biological taxonomy and ontology development: Scope

---

[24] http://inaturalist.org
[25] http://globalbioticinteractions.wordpress.org
[26] http://polytraits.lifewatchgreece.eu/
[27] http://environments-eol.blogspot.com/

and limitations. *Biodiversity Informatics*, *7*, 45–66.

[7] Gkoutos, G. V, Mungall, C., Dolken, S., Ashburner, M., Lewis, S., Hancock, J., … Robinson, P. N. (2009). Entity/quality-based logical definitions for the human skeletal phenome using PATO. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference*, *2009*, 7069–7072. doi:10.1109/IEMBS.2009.5333362

[8] Gkoutos, G. V, Schofield, P. N., & Hoehndorf, R. (2012). The Units Ontology: a tool for integrating units of measurement in science. *Database : The Journal of Biological Databases and Curation*, *2012*, bas033. doi:10.1093/database/bas033

[9] Guisan, A. (2014). Biodiversity: Predictive traits to the rescue. *Nature Climate Change*, *4*(3), 175–176. doi:10.1038/nclimate2157

[10] Harfoot, M., & Roberts, D. (2014). Taxonomy: Call for ecosystem modelling data. *Nature*, *505*(7482), 160. doi:10.1038/505160a

[11] Jaiswal, P., Ware, D., Ni, J., Chang, K., Zhao, W., Schmidt, S., … McCouch, S. (2002). Gramene: development and integration of trait and gene ontologies for rice. *Comparative and Functional Genomics*, *3*(2), 132–136.

[12] Kao, R. H., Gibson, C. M., Gallery, R. E., Meier, C. L., Barnett, D. T., Docherty, K. M., … Schimel, D. (2012). NEON terrestrial field observations: designing continental-scale, standardized sampling. *Ecosphere*, *3*(12), 1–17. doi:10.1890/ES12-00196.1

[13] Kattge, J., Díaz, S., Lavorel, S., Prentice, I. C., Leadley, P., Bönisch, G., … Wirth, C. (2011). TRY - a global database of plant traits. *Global Change Biology*, *17*(9), 2905–2935. doi:10.1111/j.1365-2486.2011.02451.x

[14] Lanthaler, M., & Gütl, C. (2012). On using JSON-LD to create evolvable RESTful

services. In *Proceedings of the 3rd International Workshop on RESTful Design WSREST 2012 at WWW2012* (pp. 25–32). ACM Press. doi:10.1145/2307819.2307827

[15] Mabee, P. M., Ashburner, M., Cronk, Q., Gkoutos, G. V, Haendel, M., Segerdell, E., … Westerfield, M. (2007). Phenotype ontologies: the bridge between genomics and evolution. *Trends in Ecology & Evolution*, *22*(7), 345–50. doi:10.1016/j.tree.2007.03.013

[16] Mungall, C. J., Torniai, C., Gkoutos, G. V, Lewis, S. E., & Haendel, M. A. (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology*. doi:10.1186/gb-2012-13-1-r5

[17] Park, C. A., Bello, S. M., Smith, C. L., Hu, Z.-L., Munzenmaier, D. H., Nigam, R., … Reecy, J. M. (2013). The Vertebrate Trait Ontology: a controlled vocabulary for the annotation of trait data across species. *Journal of Biomedical Semantics*, *4*(1), 13. doi:10.1186/2041-1480-4-13

[18] Parr, C. S., Wilson, N., Leary, P., Schulz, K. S., Lans, K., Walley, L., … Corrigan Jr., R. J. (in review). The Encyclopedia of Life v2: Providing Global Access to Knowledge About Life on Earth. *Biodiversity Data Journal*.

[19] Parr, C., Sachs, J., Parafiynyk, A., Wang, T., Espinosa, R., & Finin, T. (2006). ETHAN: the Evolutionary Trees and Natural History Ontology. University of Maryland, Baltimore County.

[20] Payne, P. R. O. (2012). Chapter 1: Biomedical knowledge integration. *PLoS Computational Biology*, *8*(12), e1002826. doi:10.1371/journal.pcbi.1002826

[21] Pereira, H. M., Ferrier, S., Walters, M., Geller, G. N., Jongman, R. H. G., Scholes, R. J., … Wegmann, M. (2013). Essential Biodiversity Variables. *Science* , *339* (6117), 277–278. doi:10.1126/science.1229931

[22] Roskov, Y., Kunze, T., Paglinawan, L., Abucay, L., Orrell, T., Nicolson, D., …

Didžiulis, V. (2013). Species 2000 & ITIS Catalogue of Life monthly checklist. Retrieved from catalogueoflife.org

[23] Rotman, D., Procita, K., Hansen, D., Parr, C. S., & Preece, J. (2012). Supporting Content Curation Communities : The Case of the Encyclopedia of Life. *Journal of the American Society for Information Science and Technology*, *63*(6), 1–29. doi:10.1002/asi.22633

[24] Thessen, A. E., & Parr, C. S. (2014). Knowledge Extraction and Semantic Annotation of Text from the Encyclopedia of Life. *PLoS ONE*, *9*(3), e89550. doi:10.1371/journal.pone.0089550

[25] Urbani, J. (2013). Three Laws Learned from Web-scale Reasoning. In *2013 AAAI Fall Symposium Series*.

[26] Vision, T. (2010). The Dryad Digital Repository: Published evolutionary data as a part of the greater data ecosystem. *Nature Precedings*, (713), 1–1. doi:10.1038/npre.2010.4595.1

[27] Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., … Vieglais, D. (2012). Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE*, *7*(1), e29715. doi:10.1371/journal.pone.0029715

[28] Xiang, Z., Mungall, C. J., Ruttenberg, A., & He, Y. (2011). Ontobee: A Linked Data Server and Browser for Ontology Terms. *Proceedings of the 2nd International Conference on Biomedical Ontologies (ICBO), July 28-30, 2011, Buffalo, NY, USA*, 279–281.