Fish Ontology framework for taxonomy-based fish recognition

Najib M. Ali [1] , **Haris A. Khan** [1] , **Amy Y-Hui Then** [1] , **Chong Ving Ching** [1] , **Sarinder Kaur Dhillon** [Corresp. 1]

[1] Institute of Biological Sciences, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author: Sarinder Kaur Dhillon
Email address: sarinder@um.edu.my

Life science ontologies play an important role in semantic web. In the fish and fisheries research field, it is imperative to have an ontology that can automatically provide information for biological objects annotations and links to relevant data pieces. As such, we introduce the Fish Ontology (FO), an automated classification architecture of existing fish taxa which provides taxonomic information of unknown fish based on metadata restrictions. It is designed to support knowledge discovery, providing semantic annotation of fish and fisheries resources, data integration, and information retrieval. The automated classification for unknown specimen is a feature not existing in other known ontologies covering fish species profiling and fisheries data. Examples of automated classification for major groups of fish are demonstrated, showing the inferred information by introducing several restrictions at the species or specimen level. The current version of FO has 1830 classes, includes widely used fisheries terminology, and models major aspects of fish taxonomy, grouping, and character. With more than 30,000 known fish species globally, the FO will be an indispensable tool for fish scientists and other interested users.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3020v1 | CC BY 4.0 Open Access | rec: 13 Jun 2017, publ: 13 Jun 2017

Fish Ontology Framework For Taxonomy-Based Fish Recognition

Najib M. Ali[1], Haris A. Khan[1], Amy Y-Hui Then[2], Chong Ving Ching[2], Sarinder Kaur Dhillon[1]

[1] Data Science and Bioinformatics Laboratory, Bioinformatics Department, Institute Science and Biology, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

[2] Institute Science and Biology, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

Corresponding Author:

Sarinder K. Dhillon[1]

Data Sciences and Bioinformatics Laboratory, Bioinformatics Department, Institute Science and Biology, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

Email address: sarinder@um.edu.my

ABSTRACT

Life science ontologies play an important role in semantic web. In the fish and fisheries research field, it is imperative to have an ontology that can automatically provide information for biological objects annotations and links to relevant data pieces. As such, we introduce the Fish Ontology (FO), an automated classification architecture of existing fish taxa which provides taxonomic information of unknown fish based on metadata restrictions. It is designed to support knowledge discovery, providing semantic annotation of fish and fisheries resources, data integration, and information retrieval. The automated classification for unknown specimen is a feature not existing in other known ontologies covering fish species profiling and fisheries data. Examples of automated classification for major groups of fish are demonstrated, showing the inferred information by introducing several restrictions at the species or specimen level. The current version of FO has 1830 classes, includes widely used fisheries terminology, and models major aspects of fish taxonomy, grouping, and character. With more than 30,000 known fish species globally, the FO will be an indispensable tool for fish scientists and other interested users.

INTRODUCTION

Increasing amount of data produced by a single species has made it harder for fish researchers to manage and provide fish data in a single database. Moreover, the high demand of having related data (metadata) for a single species have encourage researchers to find an alternative for the current database set. Since the arrival of computational automation, it is impractical to generate species data based on human observation (Perez & Benjamins, 1999). The semantic web technology provides a promising platform for biodiversity researchers who are interested to link and share their data in common public repository.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3020v1 | CC BY 4.0 Open Access | rec: 13 Jun 2017, publ: 13 Jun 2017

Nowadays, most of the fish databases are constructed using relational database models, focusing on species related information. Data in these repositories are usually structured based on the researcher's interests and personal needs which in turns restricts the application of a uniform naming standards. Hence a preferable way to provide species data is in the form of an ontology, a structured vocabulary that describes entities of a domain of interest and their relationships (Shadbolt, Hall & Berners-Lee, 2006). This is because both an ontology and a relational database stores same information but in a different structure. However with metadata, ontology can link and map with other related ontologies and use those information to automatically infer and recognize the most suitable results.

There are several important and popular projects in the fish and fisheries domain developed as conventional back-end databases such as FishBase (Froese & Pauly, 2000, 2016), IGFA Fish Database ("Fish Species Database"), The New Zealand Freshwater Fish Database ("NZ Freshwater Fish Database"), The Fish Database of Taiwan (Shao, 2001), Fish Stocking Database("Fish Stocking Database"), FishTraits (Emmanuel & Angermeier, 2013), and Fish Barcode of Life ("FISH-BOL"). While these databases provide extensive and up to date information on fish, they are not based on ontology and hence do not support semantic web deployment unless converted into appropriate formats (Ankolekar et al., 2007). Furthermore, most of them are not created based on semantic web principles (Berners-Lee, 2009) and there are little efforts dedicated to create an automated fish species identification using the semantic web approach. Thus, the work laid out in this paper is created as an effort to address these problems.

To date, no dedicated ontology with automated classification for fish exists, with the exception of the Network of Fisheries Ontology (NFO) (Caracciolo et al., 2012) which focuses on fisheries activity and selected species of commercial interests, and the Marine Top Layer Ontology (MarineTLO) (Tzitzikas et al., 2016) which focuses on marine animal. Both of these ontologies are not primarily focused on fish, and they do not provide automated classification capabilities. Given that the total number of fish species has been estimated at 32,500 globally (Nelson, 2006), an automated and comprehensive fish classification platform would be an indispensable tool for fisheries biologists, marine scientists, and even laypersons with interest in fish.

This paper describes the framework of the Fish Ontology (FO), which we see as an important work for precise and comprehensive semantic annotation of fish resources (e.g. datasets, documents, and models) where it can be used to fill in the gap of distinct terms which are missing in other ontologies. The FO is an effort to develop and maintain a controlled, structured vocabulary of terms which describe fish anatomy, morphology, ecology and various developmental stages. The FO reuses many terms from other ontologies which are related and appropriate for the fish and fisheries domain. Additionally some terms are implemented to add more description to fish, with the intention to provide more diverse search results.

Originally the FO was developed as a data warehouse for several database formats. It has since evolved to host information on captured and observed fish specimen (e.g. data on captured samples, captive specimens, or from observational experiments). After undergoing several modifications, both of these features were merged in the current FO, expanding its functionality to incorporate fish classifications and reasoning capabilities. The current FO framework outlined in this paper (current version v1.0.2, Aug 2016) is designed to facilitate integration with related ontologies in order to achieve bigger aim towards the Big Data Initiative ("IEEE Big Data").

METHODS

In this research, we used Protégé to create, edit and manage the Fish Ontology and all its terms and relationships ("Protégé," 2016). This open access software contains all the tools needed for this research since it contains sufficient plugins to assist in development and visualization of ontology. Furthermore, Protégé provides several reasoner engines such as Hermit, FaCT++, and Pellet, to provide variation in ontology validation and reasoning (Tsarkov & Horrocks, 2006; Sirin et al., 2007; Glimm et al., 2014). There are also variation of visualization tools that are provided by Protégé such as OWLVIZ, Ontograf, and VOWL (Falconer, 2010; Horridge, 2010; Negru, Lohmann & Haag, 2014). Each tools have their own way to visualize the ontology, which lessen the burden and ensure that the FO are created properly.

"The diversity of fishes: biology, evolution, and ecology" was the main reference used in identifying terms and definitions while devising the FO (Helfman et al., 2009). This book is a well-established reference that follows standard fish taxonomy nomenclature proposed by Nelson (Nelson, 2006). Most of the class labels, synonyms and definitions in the FO correspond to those in the reference book. Some of the terms for specimen entries are taken from experimental data such as sampling data provided by Chong et al (Chong, Lee & Lau, 2010), while others are taken from public online entries such as Wikipedia ("Wikipedia") and DBPedia (Heath & Bizer, 2011).

One of the most important aspects in ontology creation is consistency; hence, we sought to follow a standard naming convention while creating the FO. There are no obligatory naming conventions for the creation of Web Ontology Language (OWL) classes and properties, however, we decided to use the Camel Case (also known as Camel Back) notation to ensure that the ontology terms and naming are consistent (Campbell, 2006; Horridge et al., 2011). This naming convention has an advantage of creating more meaningful names by using an expressive sequence of words while respecting the naming constraint. As such, all of the classes in the FO use the Upper Camel Case notation, while all of its properties use the Lower Camel Case notation. Furthermore some properties are appended with the prefixes of 'has', or 'is', as per the convention recommendation (e.g. hasBodyPart, isPartOf). Additionally, this naming convention also helps clarifying the properties to human and to some tools in Protégé (e.g. The "English Prose Tooltip Generator" which uses this naming convention to generate more human readable expressions for class descriptions).

As for the terms and structures involving taxonomic rank and hierarchy, we referred to the Vertebrate Taxonomy Ontology (VTO) (Midford et al., 2013) and imported several of its major classes (with subclasses and all the annotations) in order to demonstrate the functionality of the FO. We also considered biodiversity standard such as the Darwin Core (Wieczorek et al., 2012), and other related ontologies such as the Zebrafish Information Network (ZFIN) (Sprague, 2003), as the references for the FO vocabulary creation. As an example, we imported the class "Chordata" and all of the subclasses for the genus Rastrelliger and Chiloscyllium from the VTO, and reused the terms "Location" and "Taxon" from Darwin Core in our FO. Some generic terms like "Species" were adopted due to their usage in many popular ontologies. The summary of imported classes is shown in Table 1.

Ontology or Standard	Number of classes
Zebrafish Anatomy and Stage Ontology (ZFA, ZFS)	2
Darwin Core	2
Vertebrate Taxonomy Ontology (VTO)	1345
NCBI organismal classification (NCBITaxon)	13
Total	1362

Table 1. Statistic of imported or integrated class and properties.

The FO is created with the aim of integration and standardization; thus it is imperative to ensure that the created ontology is unique and the terms that are used in the ontology have not been used elsewhere. There are many ways to create a unique identifier (ID); however following an example of globally accepted guideline will ease future integration with the FO. Furthermore by having a unique ID, no other ontology is allowed to use the same ontology terms, ensuring its originality. As such, we adopted the guidelines issued by the Open Biological and Biomedical Ontologies (OBO Foundry) (Smith, Lewis & Ashburner, 2006; Smith et al., 2007) and created each term in the FO using an ID which starts with the prefix 'FO' followed by unique digit numbers (e.g. "FO_XXXXXXX" where X is a digit).

There are many tools created for ontology validation such as the inference and rule engine. However, it is apparent that human validation is still mandatory in the current state of practice for ontology learning (Zhou, 2007). Furthermore, most ontology learning results have mainly been evaluated by domain experts manually. As such, a logical evaluation was done by fish experts to verify the naming of concepts and to validate the hierarchy of the terms which the FO presented. Criteria such as accuracy, complexity, semantic consistency, terms redundancy, naturalness, precision, completeness, and verifiability were checked using questions such as "what if we do not know the name of the specimen?", "what if we only know its common name?", "what if the specimen is similar to certain kind of known specimen?", or "what if we were to find a completely unknown specimen?". Figure 1 shows the full workflow of the fish ontology creation.

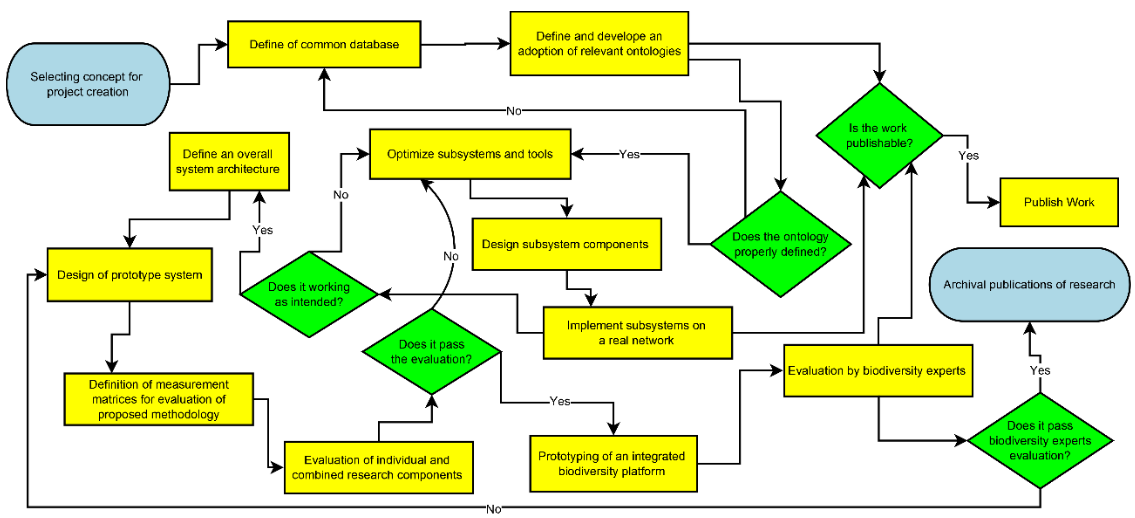

Figure 1. Workflow for Fish Ontology creation.

In this work, we show the applicability of the FO on several areas such as determining if a specimen is a fish, determining the type of fish based on characteristic(s), morphology, name, or taxonomic rank, determining its conservation status (extant or extinct), and determining whether or not it is an ancient species.

RESULTS

Fish Ontology Framework and Content

The Fish Ontology proposed in this paper contains 1830 classes, and 27 object properties. It is the first of its kind to provide automated fish classification based on taxonomic rank, group, name and characteristics. As of the current release, it contains around 500 species names, with 1223 synonyms, 8 fish group, and 9 fish characteristics. A graphical illustration of several main classes in the FO and its integration with other ontologies such as the VTO is demonstrated in Figure 2. The online version of the ontology can be accessed at https://mohdnajib1985.github.io/FOWebPage/. The OWL file for the FO with all of its imported classes is available as a supplementary material (Refer to Additional File FishOntology.owl).

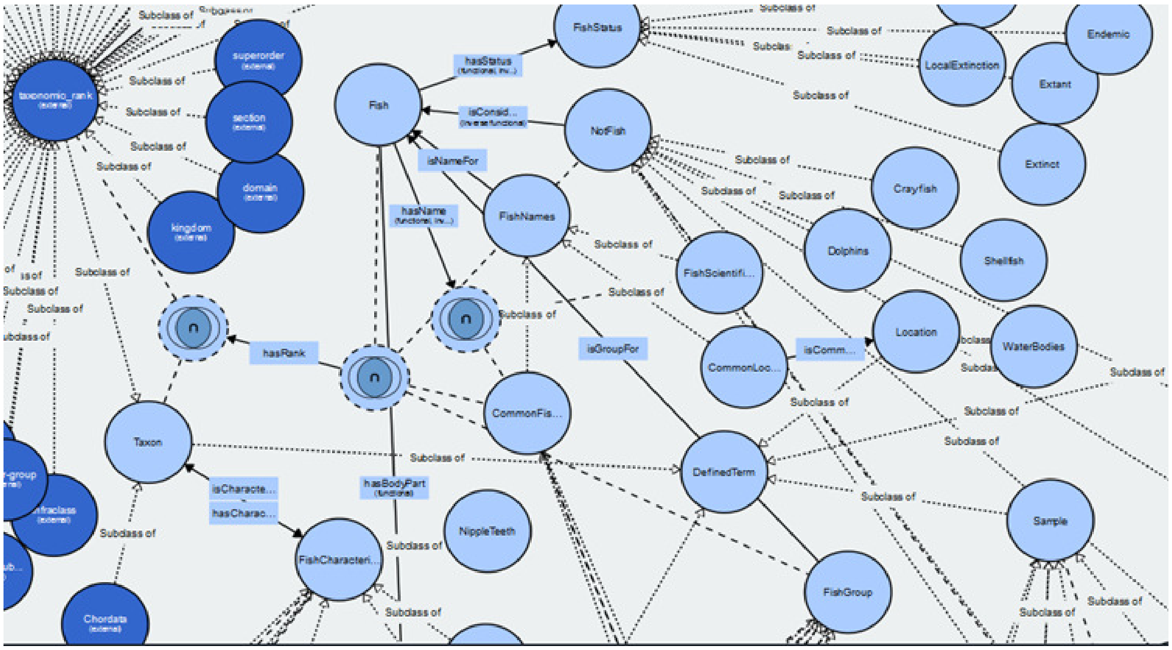

Figure 2. The Fish Ontology (FO) architecture. A portion of the FO is shown here on how the classes are related to each other and to other ontology classes. The dark blue circles represent terms from other ontologies while light blue circles represent terms from the FO.

The classes in the FO are created as a base for integration with other ontology. We incorporated classes from other ontologies for modelling FO classes such as the VTO into our ontology to further enhance its automatic recognition capabilities. For the "Taxon" class, it is organized in single inheritance, up to species level whenever possible to increase the reasoning capabilities and expand its scope by further including relationship and annotations to the terms. This includes imported classes, which are linked to their respective class types. Each FO branch is organized hierarchically by the means of "is_a" (or subclass of) relationship, by appropriately placing it under a single root term. Each classes have annotations to enrich its meaning and purpose. Examples of the relationships are shown in Table 2.

Properties	Explanation	Example
is_a	A subclass in OWL	Overharvesting is_a CausesOfThreat
hasRank (FO:0000097)	Describe a term which has a taxonomic rank	Carpet Shark hasRank of Orectolobiformes
isNameFor (FO:0000235)	Describe a name for some other class	FishNames isNameFor Fish
isGroupFor (FO:0000171)	Describe a group of some class	FishGroup isGroupFor Fish
isPartOf (FO:0000280)	Describe a situation where the class is part of something	PreflexionLarva isPartOf Larva

Table 2. Example of relationships in the Fish Ontology.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3020v1 | CC BY 4.0 Open Access | rec: 13 Jun 2017, publ: 13 Jun 2017

The FO is capable to classify jawless fish, early jawed fish and living fossil fish in the current version. Furthermore the development for classifications of several highly diverse groups, such as bony fishes, advanced jawed fish, sharks, skates, and rays, are still ongoing. The FO contains 253 classes dedicated to fish studies and 38 classes related to fish sampling processes. These classes are well suited for describing sample and specimen related terms. In combination with suitable classes, relations, and annotations, we believe that the FO can be utilized for automated fish species recognition through sample and specimen data. Some of the classes such as “FishSampling” and “FishName” are structured in a multiple inheritance structure, with some classes being subclasses of the same class; an example is the class “Trap” which is the subclass of “FishingGear” and “FishSamplingMethod”. As aforementioned, most of the new terms were created based on the reference book (Helfman et al., 2009) because to the best of our knowledge, there are no suitable ontologies from which we could import the classes in these areas, while some of the terms that we found were poorly defined and structured. However, we have included cross-references of several classes for potential mapping to relevant external resources, including the FishBase, Teleost Taxonomy Ontology (TTO), and National Centre of Biotechnology Information Taxonomy Database (NCBITaxon) (Froese & Pauly, 2000, 2016; Midford et al., 2010; Federhen, 2012). Table 3 shows the statistics of cross referencing of the FO classes to other resources.

Resources	Number of Cross References
NCBITaxon	264
Teleost Taxonomy Ontology (TTO)	317
PaleoDB	1091
Marine Top Layer Ontology (MarineTLO)	14
Gene Ontology (GO)	2
Total	1688

Table 3. Statistics for the Fish Ontology cross references.

Inference Capabilities

We have created relationships which allow a specimen (and sample) to be inferred and automatically analyzed in the area of fish grouping, taxonomic rank, and common fish names. We focused most of our modelling activities on these aspects. The specimen (and sample) which is not inferred would only be shown as a subclasses of “Sample” or “Specimen” classes; however after being inferred using the reasoner provided by Protégé, each one of them will be properly classified according to their respective parameters. Furthermore the inferred results can show which individual shares the same trait(s) as the sample and suggest what kind of group it fits into based on its characteristics. The FO also provides a structure to determine whether a specimen or a species is actually a fish or otherwise by using the reasoning capability. Figure 3 shows the results of the inferring process which shows whether the specimen is a fish or not, and what group it ranks in the taxonomic hierarchy.

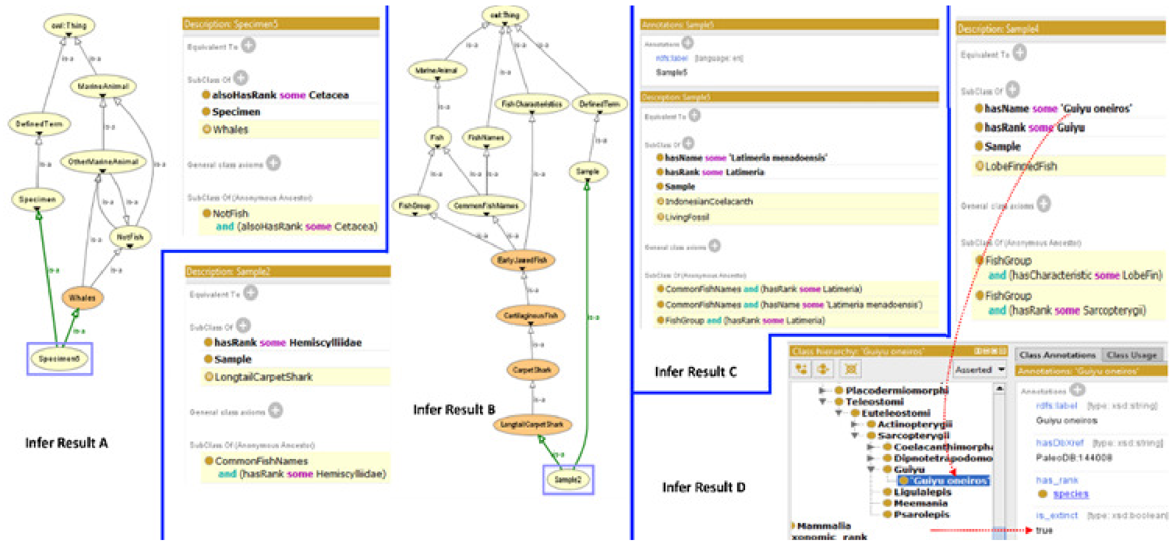

Figure 3. Results from the Fish Ontology inferring process. Infer Result A shows how the specimen (Specimen5) is recognized by the reasoner as a "Whale" and leads the reasoner to recognize it as "OtherMarineAnimal" and "NotFish". Infer Result B shows how the sample (Sample2) is recognized as a "LongtailCarpetShark", which leads the reasoner to recognize that it is a fish. Infer Result C shows that Sample5 recognized as a "LivingFossil" while Infer Result D shows how Sample4 is actually an extinct species.

The FO is created using Web Ontology Language (OWL); thus it is possible to query its data as you would query RDF/XML files using triple based query language such as SPARQL (Prud'hommeaux & Seaborne, 2008). Compared to a SPARQL query, an improved results were obtained using the SPARQL-DL query, which could query inferred data in the ontology (Sirin & Parsia, 2007). An example of query results obtained from SPARQL and SPARQL-DL is shown in Figure 4. The results show that several new classes are found in the query results, which are obtained from the imported class and integrated terms from other ontologies. The results shown in Figure 5 show how more data can be retrieved in the FO through the Description Logic (DL) query, a feature not existed in any relational databases.

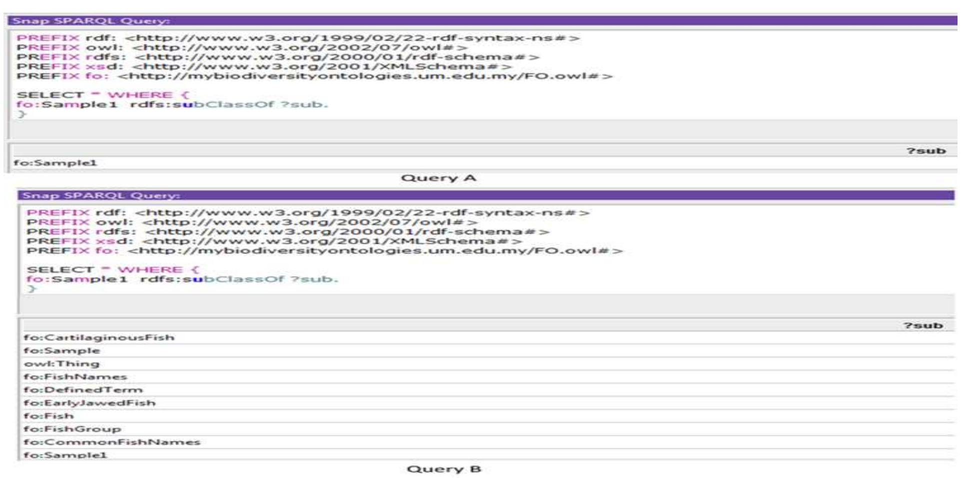

Figure 4. A sample query to check the inferred results. Results from Query A are retrieved before the inferring process, while results from Query B are retrieved after the inferring process.

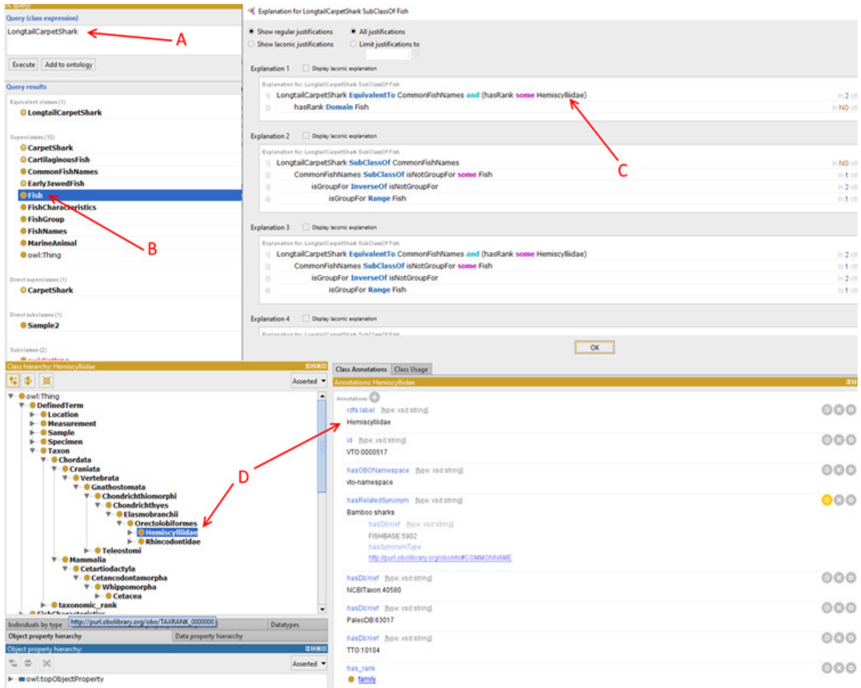

Figure 5. Types of information obtained from the Description Logic (DL) query. The DL query shows how a long tail carpet shark is inferred in the DL query (A). In B, the shark is inferred as Fish. In C, the DL query shows what kind of fish it is while in D, the shark fish rank is subsequently inferred.

Evaluation

To evaluate the quality of the FO, we follow the Gruber method for ontology construction (Gruber, 1995). There are 5 criteria that are highlighted in his research which is clarity, coherence, extendibility, minimal encoding bias, and minimal ontological commitment. To measure the clarity level of the FO, the ontology definitions should be objective and independent of the social and computational context. In FO, all the definitions are stated in such a way that the number of possible interpretations of a concept would be restricted. The clarity test results for the FO are divided into 6 parts which are:

1. No Cardinality Restriction on Transitive Properties
2. No Meta-Class
3. No Subclasses of RDF Classes
4. No Super or Sub-Properties of Annotation Properties
5. Transitive Properties cannot be Functional

Results for tests 1 and 5 are shown in Figure 6 below. Since fish data are large in volume, there is a need to add more data over the time. As such, there is no cardinality restriction assigned to any transitive properties in the FO. Figure 6 also shows that the transitive properties in the FO are not functional because it relates to more than one instance via the property. As for tests 2, 3 and 4, Figures 7 and 8 show that there are no meta-classes, no properties with a class as a range, and no sub-classes of RDF classes in the FO. Furthermore, since we used the Protégé as the development tool, all the 5 tests are automatically filtered, because these criteria are automatically flagged in the latest Protégé version.

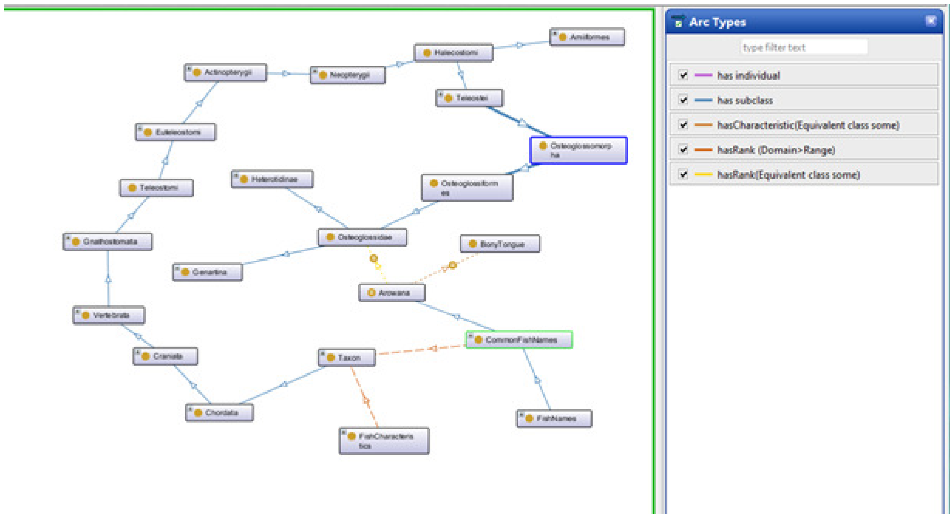

Figure 6. Results for clarity tests (1 and 5).

To ensure the coherence quality of the FO, the definitions of concepts given in the ontology should be consistent. While building the FO, the inferences drawn from the ontology must be consistent with its definitions and axioms. Only inferences consistent with existing definitions should be allowed. Most of the inferred terms from the ontology are consistent with its definition and axioms. As an example, in Figure 3 when the FO inferred that specimen5 is a whale, it also inferred that it is not a fish, and it also shown the correct taxon rank. The formal part of the FO is checked by following these 6 consistency criteria listed below and ensuring that all return true:

1. Domain of a Property should not be empty
2. Domain of a Property should not contain redundant Classes
3. Range of a Property should not contain redundant Classes
4. Inverse of Symmetric Property must be Symmetric Property
5. Inverse Property must have matching Range and Domain

The usage of software (Protégé) forces the user to always be wary about an empty domain, redundant classes, and properties. As such, tests 1 to 3 are achieved and can be further viewed through the ontology itself. For test 4, we provide an example of the property "isSimilarTo". The class "CosmoidScales" is related to the class "PlacoidScales" via the "isSimilarTo" property. Then we can infer that "PlacoidScales" must also be related to "CosmoidScales" via the "isSimilarTo" property. Figure 7 shows the results of coherence test using the Ontology Debugger Tool from Protégé. The coherence test from this tool checks for possible faulty axioms. The ontology passed the coherence test provided by this tool. Figure 8 shows the results for test 5 displaying that the properties hasCharacteristic and isCharacteristicFor have matching range and domain.

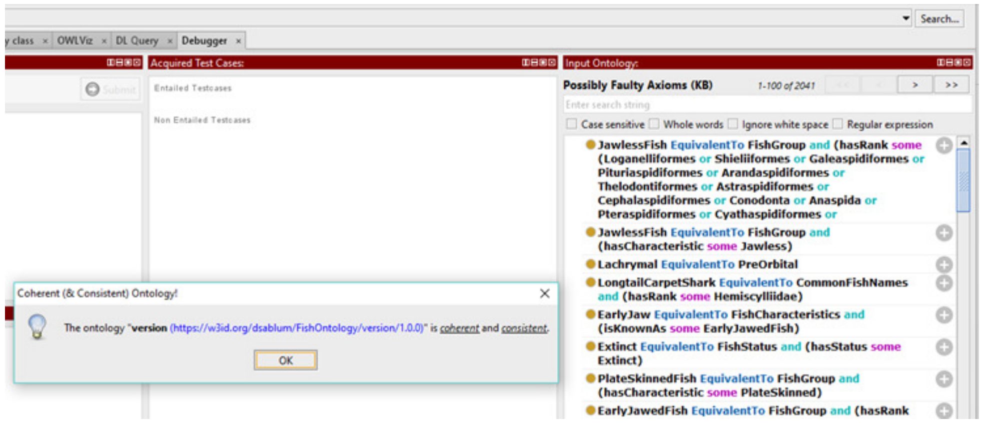

Figure 7. Results of the coherence test using Protégé Ontology Debugger tool.

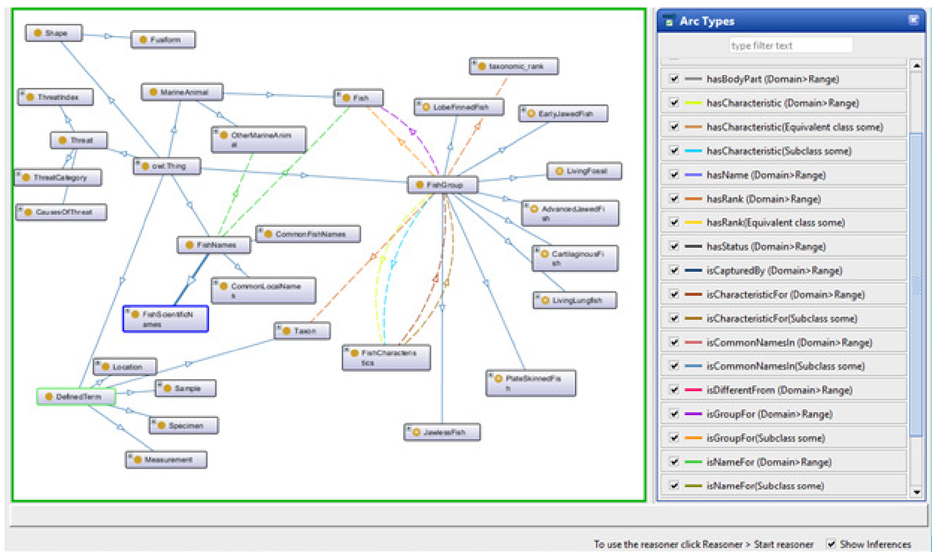

Figure 8. Results for clarity test (2, 3, and 4), coherence test (5).

It should be possible to extend the ontology without altering their existing definitions. As such, the need for easy ontology extension is prioritized while creating the FO. New knowledge emerges everyday so there may be a need to add new concepts and relationships to the existing ontology. This paper explained how the FO are extended provided that it has integrated terms from other ontology. The design consists of concepts, classification hierarchy represented as classes, from general to specific is important to make the FO extendable. Applying reasoning to the FO helps to define new concepts (generated from an ontology) from defined generic concepts (books and other databases). By placing any related concepts derived from other generic concepts in its class hierarchy, the FO represent information that defines a specimen. Creation of classes and annotations that may be useful for future integration such as genetic content will further enhance FO's extendibility. Table 1 and 3 show the extendibility of the FO. Since the first design, we have considered integrating terms from other ontologies into the FO. By placing any related concepts derived from other generic concepts in its class hierarchy, the FO represents information that defines a fish specimen, linking it with terms from other ontologies. Creation of classes and annotations that may be useful for future integration such as "genetic content" will further enhance FO's extendibility.

The ontological commitment can give a meaning of "a mapping between a language and something which can be called an ontology". Ontology modelers sometimes have a vague idea of the role each concept will play such as their semantic interconnections, within the ontology. If necessary, they can annotate new development ideas during the next update (De Nicola, Missikoff & Navigli, 2005). As such, an ontology should make as few claims as possible about the domain while still supporting the intended knowledge sharing. Since the FO reuses existing concepts (from books, databases and other ontology) and proposes only a few new concepts, it has low ontological commitment. The low ontology commitment makes the FO more extensible and reusable. Also, since most of the new concepts are from notable books and published journal articles, the concepts will be more widely accepted among the user community (Helfman et al., 2009; Last et al., 2010; Chong, Lee & Lau, 2010).

Encoding bias occurs when a representation choice is made for the convenience of notation or implementation. By minimizing encoding bias, knowledge-sharing agents may be implemented in different representation systems and styles of representation. An ontology that is independent of the issues of implementing language is considered to have minimal encoding bias. Also, the conceptualization of the ontology should be specified at the knowledge level and must be independent of symbol-level encoding. While developing the FO, the choices of using OWL as the representation language and to stick with terms from books, database, and related ontology (shown in Table 1, Table 2, and Table 3), are intended to reduce the encoding bias. Furthermore, Figure 7 also shows that there are no errors regarding encoding bias.

To strengthen the results of the FO evaluation, we use an online ontology evaluation tool named OOPS! Ontology Pitfall Scanner (OOPS) (Poveda-Villalón, Gómez-Pérez & Suárez-Figueroa, 2014). OOPS uses a checklist to ensure that best practices of ontology creation are followed and that the bad practices are avoided. The inventor created a catalog of bad practices and automated the detection of as many of them as possible (41 currently). The evaluation of the FO using the OOPS tools is shown in Figure 9. There are 1794 cases listed in the minor pitfall categories, 19 cases in 4 important pitfall categories, and 11 cases in 4 critical pitfall categories. Compared to the ontology debugger tools in the Protégé, there are many error flags that can be found in the FO by using OOPS. However, most of them are minor, and the important and critical pitfalls problems are mostly caused by the same features in the FO, and is further elaborated in discussion.

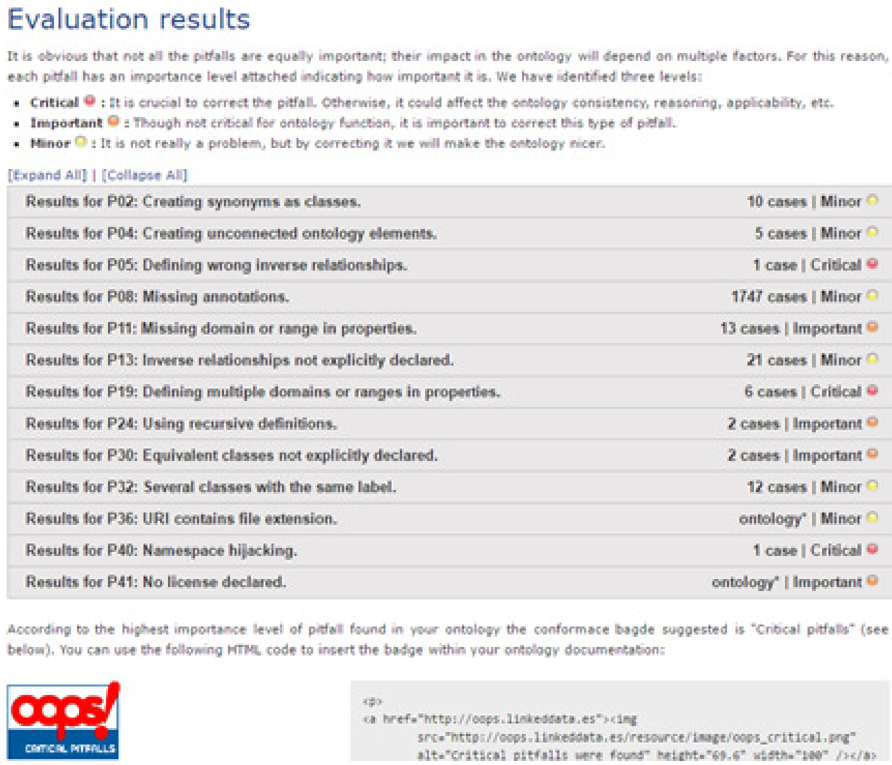

Figure 9. Results of evaluation using the Ontology Pitfall Scanner tool (Poveda-Villalón, Gómez-Pérez & Suárez-Figueroa, 2014).

DISCUSSIONS

In this paper we developed a Fish Ontology framework which is a general-purpose ontology that allows integration of domain-specific biodiversity ontologies containing standard terms and relationships. The design of the FO is flexible enough to accommodate any biodiversity ontology containing data or knowledge about fish. Even in cases where integration can be difficult, the FO can be tweaked in order to incorporate new biodiversity related ontology. One example is linking the FO to the MarineTLO which is an upper level ontology for marine species (Tzitzikas et al., 2016). The MarineTLO does not have a class named "Fish" that can map to data from the FO; however since the MarineTLO provides classes of taxonomic rank such as "Species" and "Genus", and related classes such as "MarineAnimal" and "Specimen", the FO provides classes and annotations to link these classes. The same steps can be done with the ZFIN, which contains "zebrafish anatomical entity" and "Stages" as main classes. The FO provides complementing classes to match the classes provided by the ZFIN such as "FishAnatomicalEntity" and "OtherStagesTerminology".

The FO is able to prepare captured and observed fish specimen data, mapped and structured in a way that the meaning is expressed in a machine understandable format. Since data representation in the form of an ontology allows the information to be linked by using semantic web applications, we envision several practical cases of real life applications using this ontology. As shown in the results, the FO can infer conservation and evolutionary statuses of a fish as well as show related characteristics, e.g. early jawed fish, which are useful information for interested museum visitors. The FO's ability to infer location and habitat of the fish can be useful for students or researchers. They can use the FO to identify species using local names, since all fish names in the FO are linked to other database repositories. Linkage of the FO to

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3020v1 | CC BY 4.0 Open Access | rec: 13 Jun 2017, publ: 13 Jun 2017

other ontologies via reusing of terms allows the search for relevant information such as genetic data of a specific fish species. In this way, the FO is able to produce new knowledge which is useful to biologists.

The current version of the FO can utilize specimen grouping and characteristics to determine whether a specimen is a fish or otherwise, provide taxonomic information and heredity of a characteristic rank, and determine conservation status, evolutionary status (ancient or modern) and type (ancient species is a jawless fish). This version uses simple character classification where the user provides the necessary character for the specimen. As an example, the user can specify that “Sample 1 has the characteristic of Plate Skinned”, and manually add the characteristic of “Plate Skinned” into the FO. We believe the ideal version should contain anatomical and phenotype data from several classes in the ontology such as “AnatomicalCharacters”, “MeristicCharacters”, “MolecularCharacters”, and “Morphometric Characters” and these features will be included in the near future. These classes can be useful for pattern recognition, and species taxon recognition studies. The power of the FO lies in its ability to automate group classification, and ability to link the terms used by fish domain researchers, and other researchers outside the domain.

There are other resources that model animal taxonomy which can be used to build the FO, such as the NCBITaxon which is an automatic translation of the NCBI taxonomy database into OBO or OWL format (Federhen, 2012). However, the NCBITaxon differs from the FO where it models only the taxonomic ranks without fish characters and nomenclature. The NCBITaxon also has different hierarchical organization and definitions compared to the VTO which is used as the main reference for taxonomic characters and rank in the FO. The VTO is directly imported to the FO because it is built by following several taxonomic resources including the NCBI Taxonomy, the Paleobiology Database (PaleoDB), and the Teleost Taxonomy Ontology (TTO), which suits the need of the FO for a comprehensive fish taxonomy information (Foote, 2006; Dahdul et al., 2010). One of the most distinctive values of the VTO compared to others is its broad taxonomic coverage of the vertebrates. The NCBITaxon however excludes many extant and nearly all extinct taxa while largely includes only species associated with archived genetic data, complemented by data from the PaleoDB and the TTO to provide an authoritative hierarchy and a richer set of names for specific taxonomic groups (Midford et al., 2013). Having said that, we incorporate any taxon ranks covered by the NCBITaxon but not by the VTO, such as “Protanguilla palau” and “Oxudercinae”. More examples on the differences between the main reference book, the VTO, and the NCBITaxon, as well as what the FO uses are shown in Table 4.

In general, we follow the information such as synonym, name, fish grouping, group rank, fisheries, and fish studies related terms provided by Helfman et al. as the main structure of the FO and adopt the usage of the VTO for taxonomic hierarchy, taxonomic related information, and terms related to taxonomic rank. In most cases, the taxonomy structure of the VTO is followed as it is a regularly updated ontology. There are exception in adopting the classes in the FO, such as the class “Mammalia” which the VTO classified as under “Sarcopterygii” (meaning that it is derived from fish). There are differing views on this classification and we opted not to follow the structure provided by the VTO, since some classification stated that fish is not a mammal. The use of adopted terms and concepts from our main references is further clarified with domain experts (Amy Y. Then, Chong V. Ching) in order to represent and map the appropriate contents to reflect the diverse aspects of fish (Helfman et al., 2009). The new terms are checked for its suitability to be adopted as a standard vocabulary for fish scientists. Proposing new vocabulary in biodiversity is not uncommon, since ontologies in this domain are presently insufficient and many are under development. Available standard vocabulary is not comprehensive enough to cover all the terms needed to make an ontology in the fish domain. In most cases, new terms must be proposed based on the rationale utilized in the ontology. One such example is that of Hymenoptera Anatomy Ontology, where new terms had to be proposed to expand the ontology (Seltmann et al., 2012, 2013).

Regarding ontology evaluation, there are reasons a number of errors were flagged by the Ontology Pitfall Scanner tool (OOPS) but none can be detected by using the tools from Protégé. The most apparent reason is because the scope of evaluation for both methods are different. In Protégé, only the classes and its relationship structures created in the ontology are being evaluated, while in OOPS, the classes, relationships, mapping and future integration problems are being evaluated, giving different results. One of the most important features in the FO is reusing of terms from other ontologies to reduce term redundancy in global usage. As such, many terms and structures related to fish and fisheries are taken from other ontology such as the VTO, with proper indications and reference that they are taken from its source. The idea is to reduce terms redundancy in global usage. However, since most of the terms are directly used in the FO, the OOPS tool flag these occurrences as critical errors such as “P24: Using recursive definitions”, “P32: Several classes with same labels”, and “P40: Namespace hijacking”.

Other pitfalls such as P02, P04, P08, P11, P13, P30, P36, and P41 (refer to Figure 7) are considered acceptable since there are constantly new items to be added to the ontology along with the necessary annotations, relations and property constraints. As for the pitfall “P19: Defining multiple domains or ranges in properties”, this is usually due to how the ontology is modelled. Unlike a typical ontology that use inferring capabilities to discover new relationships, we also use the inferring capabilities for automated fish species recognition. Therefore instead of using 1 to 1 relationships for the domain and range to restrict the use of the property, the usage of the property is enlarged so that it is more reliable for automated species discovery.

Term Example	Helfman (2009)	VTO	NCBITaxon	FO
Furcacaudiformes (order)	Classified as Subclass of Thelodonti (superclass).	Classified as subclass of Agnatha (class).	Not classified.	Follows and reuses the VTO terms.
JawlessFish	Contains species and information for jawless fish species.	No classes and annotations found, but related species are classified.	No classes and annotations found, but related species are classified.	Follows Helfman (2009) for labeling.
LobeFinnedFish	Classify it as Actinopterygii (page 4).	No classes and annotations found, but related species are classified.	Classified as Coelacanthiformes	Follow Helfman (2009) for classification and labeling.
Gobiidae (family)	Listed and classified as family.	Listed and classified as family.	Listed and classified as family.	Follows and reuses the VTO terms.
Oxudercinae (subfamily)	Not listed or classified.	Not listed or classified.	Classified as a subclass of Gobiidae (family).	Follows and reuses the VTO classification up to the lowest existing taxonomic terms covered (Family Gobiidae). Adopts NCBITaxon terms for Subfamily Oxudercinae onwards.

Table 4. Term adoption example in the Fish Ontology.

The current FO version covers the terms for fish domain which are not well described by other ontologies, particularly those related with automatic classifications, annotations and relationships. There are however some terms in the FO created using parameters rarely used outside of this domain, such as "FishDatabases" which are for any known databases for fish, or "GasBladder" which is a specific organ for "Actinopterygii". The differences between the FO and other fish related ontologies and databases is its ability to provide automated classification of unknown specimen. Table 5 further elaborate the differences of the approach done by the Fish Ontology, compared to other related ontology and popular databases. Please take note that this evaluation are based on publicly accessed information for all the databases and ontologies involved.

	Fishbase	MarineTLO	NFO	FO
Domain Coverage	Fish and Fisheries	Marine Life	Fisheries	Fish
Ontology Based	No	Yes	Yes	Yes
Underlying Sources	33500 Species, 319000 Common names, 58100 Pictures, 53800 References information from the FishBase Consortium and 2270 Collaborators	FLOD (Fisheries Linked Open Data), ECOSCOPE (A Knowledge Base About Marine Ecosystems), WORMS (World Register of Marine Species), DBpedia, and FishBase	ISSCAAP (International Standard Statistical Classification of Aquatic Animals and Plants), AGROVOC (a portmanteau of agriculture and vocabulary) thesaurus, ASFA (Aquatic Sciences and Fisheries Abstracts) thesaurus, and FIGIS (Fisheries Global Information System) data	TTO, NCBITaxon, and VTO (with linked information from FishBase and PaleoDB)
Fish Information Provided	Common Name, Scientific Name (both species and genus, and species id), Information by Family, by country/island, by ecosystem, or by specific topic	Species, Scientific Names, Common Names, Predators, Authorships, Ecosystems, Countries, Water Areas, Vessels, Gears, EEZ, Bibliography, Statistical Indicators	Imported data sources in the owl file cover the topic of water areas, species taxonomic classification, ISSCAAP commercial classification, Aquatic resources, Land areas, Fisheries commodities, Vessel types and size, Gear types, AGROVOC data and ASFA data.	Species, Taxon Information, Fish Name, classes related to fish studies and fisheries.
Difference in fish searching concept	When searching for a fish species in FishBase, details	Searching a fish species through the MarineTLO owl file	Searching a fish species through the NFO owl file is also	When FO search for a fish, it provide its taxon information,

	such as names (common, scientific, other language), taxon classifications, environment, climate, range, distribution, size, weight, age, short description, biology, life cycle, mating behavior, main references, IUCN redlist status, threat to human, and human uses will be provided (if available). Furthermore, other information such as the species countries, FAO areas, occurrences, ecology, genetics, internet sources, special reports, tools, and xml data sources are available as additional information sources.	is not possible. However its competency query v4 suggested that it cover wide range of search topics such as finding a species, its scientific name, its WORMS classification, prey and predator information, references, images, general terms, identifiers, competitors, biotic type of predator, assignment data, its biological environment, common name with complementary information, and water areas with their FAO codes.	not possible. However it's imported data sources suggested the you can get information on fish species' ISSCAAP classification, ASFIS list (covers names and extensive details of species taxonomic rank), Aquatic Sciences and Fisheries Abstracts (ASFA) bibliographic database (links to FAO Fish Finder Fact Sheets which cover synonyms, FAO names, scientific names with original description, diagnostic features, Geographical distribution, habitat and biology, size, interest to fisheries, local names, source of information and Bibliography)	scientific name, common name, synonym, and links to TTO, FishBase and PaleoDB (if available). When unknown species is inferred in the FO, it can find whether a specimen or a sample is a fish or not fish, providing its taxon rank, full name, its characteristic, grouping, and its extinction status. Future concepts will allows it to provide data on fish morphology, genetic content and other fish species related information such as country maturity and other related information (like fishbase). FO infers the type of fish based on parameters provided.

Table 5. Differences between FO with other related ontology and database.

FishBase has a wide range of fish information. You can search most of the important topic from the portal, and each information has informative related links. However, it does not provide semantic web search. The search function is provided in the Portal as a relational based search. The MarineTLO covers all information on marine species including fish. It is used as the underlying backbone for many application in iMarine Gateway. However, since there is no definitive proof that the MarineTLO is directly used in their application, we evaluate its search based on its competency query V4. The NFO provided by the FAO is still a prototype and it is mentioned on their website that the ontology currently provided is still a draft version. Since this is a prototype, we can't make a proper evaluation to the ontology, however, we are interested in discussing their concept as the comparison to the FO. We compare NFO based on the owl file provided in the FAO web portal. Compared to other ontologies and databases, the FO is the only one focuses primarily on fish with the intention to bring automated fish recognition through the use of ontology.

We envision the FO to expand by incorporating additional components such as fish models, fisheries parameters, gene annotations and other relevant information as aforementioned. We will focus on parameters that influence the grouping process such as shape and characteristic recognition, and anatomical metric distinctions. Other than including more terms and defined relationships, we are considering to increase granularity by linking to more relevant and established ontologies, such as the Gene Ontology (GO), Zebrafish Information Network (ZFIN), Vertebrate Skeletal Ontology (VSAO, Protein Ontology (PO), and the Monogenean Haptoral Bar Ontology (MHBI) (Ashburner et al., 2000; Sprague, 2003; Dahdul et al., 2010; Natale et al., 2011; Abu et al., 2013). In the near future we aim to integrate the FO with other ongoing efforts in our research group such as the Otolith Ontology, Monogenean Ontology, and the MHBI Ontology (Abu et al., 2013). The annotation of fish and fisheries resources in the FO and other related ontologies is a response to the emerging need for data sharing and integration especially for fish data resources (Ashburner et al., 2000; Gangemi et al., 2004; Bizer et al., 2009; Dahdul et al., 2010, 2012, Midford et al., 2010, 2013; Natale et al., 2011; Schriml et al., 2012; Federhen, 2012; Tzitzikas et al., 2013; Van Slyke et al., 2014; Pesquita et al., 2014). There is also a possibility to link related ontologies to the existing fish databases using the FO as a mediator(“Fish Stocking Database”, “FISH-BOL”, “Fish Species Database”, “NZ Freshwater Fish Database”; Shao, 2001; Froese & Pauly, 2016).

CONCLUSIONS

An ontology for fish which covers all the terms useful for fish and fisheries research are introduced. The FO are created by importing on the existing ontologies such as VTO, ZFA, and TTO. As conclusion, we are confident that the FO presented in this article could be used as a framework to build semantic web systems for data integration to be applied in biodiversity research in the fish and fishery domain.

ACKNOWLEDGEMENT

This research was supported by the University of Malaya Research Grants (FRGS 2014-2, RG053/09SUS, RP008/2012A and PG212-2014B).

ADDITIONAL INFORMATION AND DECLARATIONS

COMPETING INTERESTS

The authors declare there are no competing interests.

AUTHOR CONTRIBUTIONS

Mohd Najib Mohd Ali analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper, created and developed the ontology.

Haris Ali Khan, Sarinder Kaur Dhillon, and Amy Yee Then reviewed drafts of the paper, contributed to the development of the ontology by making active changes to it.

Amy Yee Then reviewed drafts of the paper, contributed to the development of the ontology by providing insight of the ontology structure with her expertise in fish and fisheries knowledge.

Chong Ving Ching reviewed drafts of the paper, contributed to the development of the ontology by providing his data on fish and fisheries research.

Sarinder Kaur Dhillon reviewed drafts of the paper, contributed to the development of the ontology with support regarding the ontology development.

REFERENCES

Abu A., Lim S., Sidhu A., Dhillon S. 2013a. Biodiversity image retrieval framework for monogeneans. *Systematics and Biodiversity* 11:1–15.

Abu A., Susan L., Sidhu A., Dhillon S. 2013b. Semantic representation of monogenean haptoral Bar image annotation. *BMC bioinformatics* 14:48.

Ankolekar A., Krotzsch M., Tran T., Vrandecic D. 2007. The Two Cultures. In: *WWW '07: Proceedings of the 16th international conference on World Wide Web*. New York, New York, USA: ACM Press, 825–834.

Ashburner M., Ball CA., Blake JA., Botstein D., Butler H., Cherry JM., Davis AP., Dolinski K., Dwight SS., Eppig JT., Harris MA., Hill DP., Issel-Tarver L., Kasarskis A., Lewis S., Matese JC., Richardson JE., Ringwald M., Rubin GM., Sherlock G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nature genetics* 25:25–9.

Berners-Lee T. 2009.Linked Data - Design Issues. *Available at http://www.w3.org/DesignIssues/LinkedData.html* (accessed September 10, 2016).

Bizer C., Lehmann J., Kobilarov G., Auer S., Becker C., Cyganiak R., Hellmann S. 2009. DBpedia - A crystallization point for the Web of Data. *Web Semantics: Science, Services and Agents on the World Wide Web* 7:154–165.

Campbell GA. 2006. Ontology for Call Control.

Caracciolo C. 2007. Revised and enhanced fisheries ontologies.

Caracciolo C., Heguiabehere J., Gangemi A., Baldassarre C., Johannes K., Taconet M. 2012. Knowledge management at FAO: A case study on network of ontologies in fisheries. *Ontology Engineering in a Networked World*:383–405. DOI: 10.1007/978-3-642-24794-1_18.

Chong VC., Lee PKY., Lau CM. 2010. Diversity, extinction risk and conservation of Malaysian fishes. *Journal of fish biology* 76:2009–66.

Dahdul WM., Balhoff JP., Blackburn DC., Diehl AD., Haendel MA., Hall BK., Lapp H., Lundberg JG., Mungall CJ., Ringwald M., Segerdell E., Van Slyke CE., Vickaryous MK., Westerfield M., Mabee PM. 2012. A unified anatomy ontology of the vertebrate skeletal system. *PloS one* 7:e51070.

Dahdul WM., Lundberg JG., Midford PE., Balhoff JP., Lapp H., Vision TJ., Haendel MA., Westerfield M., Mabee PM. 2010. The teleost anatomy ontology: anatomical representation for the genomics age. *Systematic biology* 59:369–83.

Emmanuel FA., Angermeier PL. 2013.FishTraits. *Available at http://www.fishtraits.info/* (accessed May 22, 2016).

Falconer S. 2010.OntoGraf - Protege Wiki. *Available at http://protegewiki.stanford.edu/wiki/OntoGraf* (accessed May 25, 2016).

Federhen S. 2012. The NCBI Taxonomy Database. *Nucleic Acids Research* 40:136–143.

FISH-BOL. *Available at http://www.fishbol.org/* (accessed May 22, 2016).

Fish Species Database. *Available at https://www.igfa.org/fish/fish-database.aspx* (accessed May 22, 2016).

Fish Stocking Database. *Available at http://www.glfc.org/fishstocking/* (accessed May 22, 2016).

Foote M. 2006.The Paleobiology Database. *Available at http://fossilworks.org/?page=paleodb*

Froese R., Pauly D. 2000. *FishBase 2000: Concepts, Designs and Data Source*. WorldFish.

Froese R., Pauly D. 2016.FishBase. *Available at http://www.fishbase.org/* (accessed May 22, 2016).

Gangemi A., Fisseha F., Keizer J., Lehmann J., Liang A., Pettman I., Sini M., Taconet M. 2004. A core ontology of fishery and its use in the fishery ontology service project. *EKAW04 Workshop on Core Ontologies in Ontology Engineering*.

Glimm B., Horrocks I., Motik B., Stoilos G., Wang Z. 2014. HermiT: An OWL 2 Reasoner. *Journal of Automated Reasoning* 53:245–269.

Gruber TR. 1995. Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies* 43:907–928.

Heath T., Bizer C. 2011. *Linked Data: Evolving the Web Into a Global Data Space*. Morgan & Claypool Publishers.

Helfman GS., Collette BB., Facey DE., Bowen BW. 2009. *The diversity of fishes: biology, evolution, and ecology*.

Horridge M. 2010.Owlviz. *Available at http://protegewiki.stanford.edu/wiki/OWLViz* (accessed May 25, 2016).

Horridge M., Knublauch H., Rector A., Stevens R., Wroe C., Jupp S., Moulton G., Drummond N., Brandt S. 2011. A Practical Guide To Building OWL Ontologies Using Protégè 4 and CO-ODE Tools Edition 1.3. *Matrix*:0–107.

IEEE Big Data. *Available at http://www.bigdata.ieee.org* (accessed May 28, 2016).

Last PR., White WT., Caira JN., Dharmadi., Fahmi., Jensen K., Lim a PK., Manjaji-Matsumoto BM., Naylor GJP., Pogonoski JJ., Stevens JD., Yearsley GK. 2010. *Sharks and Rays of Borneo*.

Midford P., Balhoff J., Dahdul W., Kothari C., Lapp H., Lundberg J., Mabee P., Vision T., Westerfield M. 2010. The Teleost Taxonomy Ontology. *Nature Precedings*.

Midford P., Dececchi T., Balhoff J., Dahdul W., Ibrahim N., Lapp H., Lundberg J., Mabee P., Sereno P., Westerfield M., Vision T., Blackburn D. 2013. The vertebrate taxonomy ontology: a framework for reasoning across model organism and species phenotypes. *Journal of biomedical semantics* 4:34.

Natale DA., Arighi CN., Barker WC., Blake JA., Bult CJ., Caudy M., Drabkin HJ., D'Eustachio P., Evsikov A V., Huang H., Nchoutmboube J., Roberts N V., Smith B., Zhang J., Wu CH. 2011. The Protein Ontology: a structured representation of protein forms and complexes. *Nucleic acids research* 39:D539-45.

Negru S., Lohmann S., Haag F. 2014.Visual Notation for OWL Ontologies. *Available at http://vowl.visualdataweb.org/v2/*

Nelson JS. 2006. *Fishes of the World*. John Wiley & Sons.

De Nicola A., Missikoff M., Navigli R. 2005. A Proposal for a Unified Process for Ontology Building: UPON. In: *Proceedings of the 16th International Conference on Database and Expert Systems Applications*. 655–664.

NZ Freshwater Fish Database. *Available at https://www.niwa.co.nz/our-services/online-*

services/freshwater-fish-database (accessed May 22, 2016).

Perez AG., Benjamins VR. 1999. Overview of Knowledge Sharing and Reuse Components : Ontologies and Problem-Solving Methods. *IJCAI-99 workshop on Ontologies and Problem-Solving Method (KRR5)*:1–15.

Pesquita C., Ferreira JD., Couto FM., Silva MJ. 2014. The epidemiology ontology: an ontology for the semantic annotation of epidemiological resources. *Journal of biomedical semantics* 5:4.

Poveda-Villalón M., Gómez-Pérez A., Suárez-Figueroa MC. 2014. OOPS! (OntOlogy Pitfall Scanner!): supporting ontology evaluation on-line. *International Journal on Semantic Web & Information Systems* 10:7–34.

Protégé 2016. *Available at http://protege.stanford.edu/* (accessed May 23, 2016).

Prud'hommeaux E., Seaborne A. 2008.SPARQL Query Language for RDF. *Available at http://www.w3.org/TR/rdf-sparql-query/* (accessed May 23, 2016).

Schriml LM., Arze C., Nadendla S., Chang YWW., Mazaitis M., Felix V., Feng G., Kibbe WA. 2012. Disease ontology: A backbone for disease semantic integration. *Nucleic Acids Research* 40:940–946.

Seltmann K., Pénzes Z., Yoder M., Bertone M., Deans A. 2013. Utilizing Descriptive Statements from the Biodiversity Heritage Library to Expand the Hymenoptera Anatomy Ontology. *PLoS ONE* 8:e55674.

Seltmann K., Yoder M., Miko I., Forshage M., Bertone M., Agosti D., Austin A., Balhoff J., Borowiec M., Brady S., Broad G., Brothers D., Burks R., Buffington M., Campbell H., Dew K., Ernst A., Fernandez-Triana J., Gates M., Gibson G., Jennings J., Johnson N., Karlsson D., Kawada R., Krogmann L., Kula R., Mullins P., Ohl M., Rasmussen C., Ronquist F., Schulmeister S., Sharkey M., Talamas E., Tucker E., Vilhelmsen L., Ward PS., Wharton R., Deans A. 2012. A hymenopterists' guide to the Hymenoptera Anatomy Ontology: utility, clarification, and future directions. *Journal of Hymenoptera Research* 27:67–88.

Shadbolt N., Hall W., Berners-Lee T. 2006. The semantic web revisited. *IEEE Intelligent Systems* 21:96–101.

Shao K-T. 2001.Fish Database of Taiwan. *Available at http://fishdb.sinica.edu.tw/eng/home.php* (accessed May 22, 2016).

Sirin E., Parsia B. 2007. SPARQL-DL: SPARQL query for OWL-DL. *CEUR Workshop Proceedings* 258.

Sirin E., Parsia B., Grau BC., Kalyanpur A., Katz Y. 2007. Pellet: A practical OWL-DL reasoner. *Web Semantics: Science, Services and Agents on the World Wide Web* 5:51–53.

Van Slyke C., Bradford Y., Westerfield M., Haendel M. 2014. The zebrafish anatomy and stage ontologies: representing the anatomy and development of Danio rerio. *Journal of biomedical semantics* 5:12.

Smith B., Ashburner M., Rosse C., Bard J., Bug W., Ceusters W., Goldberg LJ., Eilbeck K., Ireland A., Mungall CJ., Leontis N., Rocca-Serra P., Ruttenberg A., Sansone S-A., Scheuermann RH., Shah N., Whetzel PL., Lewis S. 2007. The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature biotechnology* 25:1251–1255.

Smith B., Lewis S., Ashburner M. 2006.The OBO Foundry. *Available at http://obofoundry.org/* (accessed May 24, 2016).

Sprague J. 2003. The Zebrafish Information Network (ZFIN): the zebrafish model organism database. *Nucleic Acids Research* 31:241–243.

Tsarkov D., Horrocks I. 2006. FaCT++ Description Logic Reasoner: System Description. *Proceedings of the Third International Joint Conference (IJCAR)* 4130:292–297.

Tzitzikas Y., Allocca C., Bekiari C., Marketakis Y., Fafalios P., Doerr M., Minadakis N., Patkos T., Candela L. 2013. Integrating Heterogeneous and Distributed Information about Marine Species through a Top Level Ontology. In: *7th Metadata and Semantic Research Conference*.

Tzitzikas Y., Allocca C., Bekiari C., Marketakis Y., Fafalios P., Doerr M., Minadakis N., Patkos T., Candela L. 2016. Unifying heterogeneous and distributed information about marine species through the top level ontology MarineTLO. *Program* 50:16–40.

Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T., Vieglais D. 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. *PLoS ONE* 7:e29715.

Wikipedia. *Available at http://www.wikipedia.org*

Zhou L. 2007. Ontology learning: state of the art and open issues. *Information Technology and Management* 8:241–252.