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] , **Manas Gaur** [2] , **Sarinder Kaur Dhillon**
Corresp. [1]

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

[2] Wright State University, Kno.e.sis Center, Dayton, Ohio, United States of America

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

Life science ontologies play an important role in semantic web. Given the diversity in fish species and associated wealth of information, it is imperative to develop an ontology capable of linking and integrating this information in an automated fashion. As such, we introduce the Fish Ontology (FO), an automated classification architecture of existing fish taxa which provides taxonomic information on unknown fish based on metadata restrictions. It is designed to support knowledge discovery, provide semantic annotation of fish and fisheries resources, data integration, and information retrieval. Automated classification for unknown specimens is a unique feature that currently does not appear to exist in other known ontologies. 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.

1 **Fish Ontology Framework For Taxonomy-Based Fish**
2 **Recognition**

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

4 [1]Institute of Biological Sciences, Faculty of Science, University of Malaya, Kuala Lumpur, Malaysia

5 [2] Kno.e.sis Center, Wright State University, Dayton, Ohio, United States of America

7 Corresponding Author:

8 Sarinder K. Dhillon[1]

9 Email address: sarinder@um.edu.my

11 **ABSTRACT**

12 Life science ontologies play an important role in semantic web. Given the diversity in fish species and
13 associated wealth of information, it is imperative to develop an ontology capable of linking and integrating
14 this information in an automated fashion. As such, we introduce the Fish Ontology (FO), an automated
15 classification architecture of existing fish taxa which provides taxonomic information on unknown fish
16 based on metadata restrictions. It is designed to support knowledge discovery, provide semantic
17 annotation of fish and fisheries resources, data integration, and information retrieval. Automated
18 classification for unknown specimens is a unique feature that currently does not appear to exist in other
19 known ontologies. Examples of automated classification for major groups of fish are demonstrated,
20 showing the inferred information by introducing several restrictions at the species or specimen level. The
21 current version of FO has 1830 classes, includes widely used fisheries terminology, and models major
22 aspects of fish taxonomy, grouping, and character. With more than 30,000 known fish species globally,
23 the FO will be an indispensable tool for fish scientists and other interested users.

24

25 **INTRODUCTION**

26 Increasing amount of data produced on a single species has made it harder for fish researchers to manage
27 and provide fish data in a single database. Moreover, the high demand for related data (metadata) for a
28 single species is driving researchers to find a better alternative for the current database structure
29 (discussed further below). The semantic web technology, with the capability to give well-defined meaning
30 to information and to enable better cooperation between human and computer, provides a promising
31 platform for biodiversity researchers who are interested to link and share their data in common public
32 repository.
33   Presently, most of the fish databases are constructed using relational database models, focusing on
34 species related information. Data in these repositories are usually structured based on the researcher's
35 interests and personal needs, which in turn restrict the application of a uniform naming standard. Hence
36 a preferred way to provide species data is in the form of an ontology, a structured vocabulary that

37 describes entities of a domain of interest and their relationships (Shadbolt, Hall & Berners-Lee, 2006). A
38 relational database focuses on the data whereas, an ontology provide meaning to the data with the help
39 of metadata. Using metadata, an ontology can be linked and mapped to other related ontologies and this
40 information can be used to automatically infer and recognize the most suitable result for a given search.
41   There are several important and popular projects in the fish and fisheries domain developed as
42 conventional back-end databases such as the Catalog of Fishes ("Catalog of Fishes Classification"),
43 FishBase (Froese & Pauly, 2000), IGFA Fish Database ("Fish Species Database"), The New Zealand
44 Freshwater Fish Database ("NZ Freshwater Fish Database"), The Fish Database of Taiwan (Shao, 2001),
45 Fish Stocking Database("Fish Stocking Database"), FishTraits (Emmanuel & Angermeier, 2013), and Fish
46 Barcode of Life ("FISH-BOL"). While these databases provide extensive and up to date information on fish,
47 they are not based on ontology and hence do not support semantic web deployment unless converted
48 into appropriate formats (Ankolekar et al., 2007). Furthermore, most of them are not created based on
49 semantic web principles (Berners-Lee) and there is little effort dedicated to create an automated fish
50 species identification using the semantic web approach. An ontology can enhance information providing
51 capability for a database by the use of metadata to discover and gather new information from other
52 databases, or by linking them to create a better information network. Thus, the work laid out in this paper
53 is created as an effort to address these problems.
54   To date, no dedicated ontology with automated classification for fish exists, with the exception of
55 the Network of Fisheries Ontology (NFO) (Caracciolo et al., 2012) which focuses on fisheries activity and
56 selected species of commercial interests, and the Marine Top Layer Ontology (MarineTLO) (Tzitzikas et
57 al., 2016) which focuses on marine animal. Both of these ontologies are not primarily focused on fish, and
58 they do not possess automated classification capability. Given that the total number of fish species has
59 been estimated at 32,000 to 40,000 globally (Nelson, 2006; Chapman, 2009; Eschmeyer et al., 2010), an
60 automated and comprehensive fish classification platform would be an indispensable tool for fisheries
61 biologists, marine scientists, and even laypersons with interest in fish. Thus, in this paper, we propose a
62 fish-based ontology that is able to automate group classification, and to link terms used by research on
63 the fish domain with related terms from other research domain.
64   This paper describes the framework of the Fish Ontology (FO) for precise and comprehensive
65 semantic annotation of fish resources (e.g. datasets, documents, and models) where it can be used to fill
66 in the gap of distinct terms which are missing in other ontologies. The FO is an effort to develop and
67 maintain a controlled, structured vocabulary of terms which describe fish anatomy, morphology, ecology
68 and various developmental stages. The FO reuses many terms from other ontologies which are related to
69 and appropriate for the fish and fisheries domain. Additionally some terms such as "Location", "Shape",
70 and "Threat" are implemented to add more description to fish, with the intention to provide more diverse
71 search results.
72   Originally the FO was developed as a data warehouse for several database formats. It has since
73 evolved to host information on captured and observed fish specimen (e.g. data on captured samples,
74 captive specimens, or from observational experiments). After undergoing several modifications based on
75 reviews by fish experts, both of these features were merged in the current FO, expanding its functionality
76 to incorporate fish classification and reasoning capability. The FO framework outlined in this paper
77 (current version v1.0.2, Aug 2016) is designed to facilitate integration with related ontologies which is in
78 line with the Big Data Initiative ("IEEE Big Data") that aims for diverse analytic options.
79

80 **METHODS**

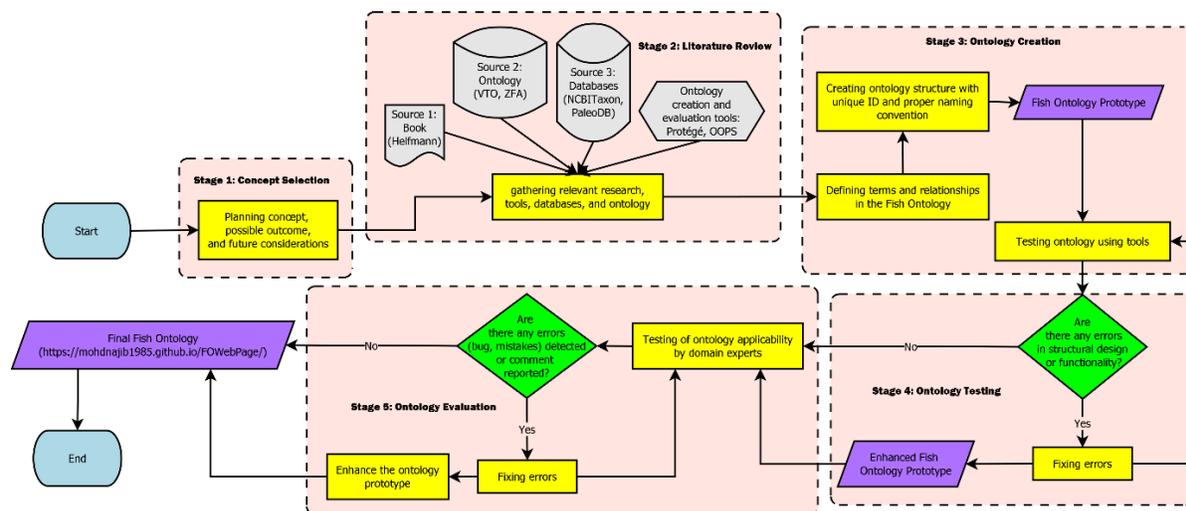
81 We used Protégé to create, edit and manage the Fish Ontology and all its terms and relationships
82 ("Protégé," 2016). This open access software contains all the tools needed for this research since it
83 contains sufficient plugins to assist in development and visualization of ontology. Furthermore, Protégé

84   provides several reasoner engines such as Hermit, FaCT++, and Pellet, to provide variation in ontology
85   validation and reasoning (Tsarkov & Horrocks, 2006; Sirin et al., 2007; Glimm et al., 2014). There are also
86   various visualization tools that are provided by Protégé such as OWLVIZ, Ontograf, and VOWL (Falconer,
87   2010; Horridge, 2010; Negru, Lohmann & Haag, 2014). Each tool visualizes the ontology differently.
88   The FO is created using Web Ontology Language (OWL) which allows us to query using triple based
89   query languages such as SPARQL (Prud'hommeaux & Seaborne, 2008), SPARQL-DL (Sirin & Parsia, 2007),
90   and Description Logic (DL) (Baader et al., 2003). A triple query (composed of subject-predicate-object) can
91   perform more complex query compared to a relational database query (composed of columns and rows),
92   and is able to retrieve more information due to the semantic web architecture which enables them to pull
93   data from URIs or URLs with related metadata (Alexander, 2013).
94   "The Diversity of Fishes: Biology, Evolution, and Ecology" was the main reference used in identifying
95   terms and definitions while devising the FO (Helfman et al., 2009). This book is a well-established
96   reference that follows standard fish taxonomy nomenclature proposed by Nelson (Nelson, 2006). Most of
97   the class labels, synonyms and definitions in the FO correspond to those in the reference book. Some of
98   the terms for specimen entries are taken from experimental data such as sampling data provided by Chong
99   et al (Chong, Lee & Lau, 2010), while others are taken from public online entries such as Wikipedia
100 ("Wikipedia") and DBpedia (Heath & Bizer, 2011). We also incorporated classes from other ontologies
101 such as the VTO into our ontology to model the FO classes and enhance its automatic recognition
102 capabilities.
103   One of the most important aspects in ontology creation is consistency; hence, we sought to follow a
104 standard naming convention while creating the FO. There are no obligatory naming conventions for the
105 creation of OWL classes and properties; however, we decided to use the Camel Case (also known as Camel
106 Back) notation to ensure that the ontology terms and naming are consistent (Campbell, 2006; Horridge et
107 al., 2011). This naming convention has the advantage of creating more meaningful names by using an
108 expressive sequence of words while respecting the naming constraint (Horridge et al., 2011). As such, all
109 of the classes in the FO use the Upper Camel Case notation, while all of its properties use the Lower Camel
110 Case notation. Furthermore some properties are appended with the prefixes of 'has', or 'is', as per the
111 convention recommendation (e.g. "hasBodyPart", "isPartOf"). This naming convention helps clarify the
112 properties to human and to some tools in Protégé (e.g. The "English Prose Tooltip Generator" which uses
113 this naming convention to generate more human readable expressions for class description).
114   As for the terms and structures involving taxonomic rank and hierarchy, we referred to the
115 Vertebrate Taxonomy Ontology (VTO) (Midford et al., 2013) and imported several of its major classes
116 (with subclasses and all the annotations) in order to demonstrate the functionality of the FO. We also
117 considered the biodiversity standard outlined in the Darwin Core (Wieczorek et al., 2012), and other
118 related ontologies such as the Zebrafish Information Network (ZFIN) (Sprague, 2003), as the references
119 for the FO vocabulary creation. These ontologies are related to fish, popular in their domain, frequently
120 used, and regularly updated by the research community. Hence, they are the most relevant choices as the
121 main vocabulary provider for fish rank and terms for the FO. As an example, we imported the class
122 "Chordata" and all of the subclasses for the genus *Rastrelliger* and *Chiloscyllium* from the VTO, and reused
123 the terms "Location" and "Taxon" from the Darwin Core in the FO. Some generic terms like "Species" were
124 adopted due to their usage in many popular ontologies. The summary of imported classes is shown in
125 Table 1.
126

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

127
128   **Table 1.** Statistic of imported or integrated class and properties.
129

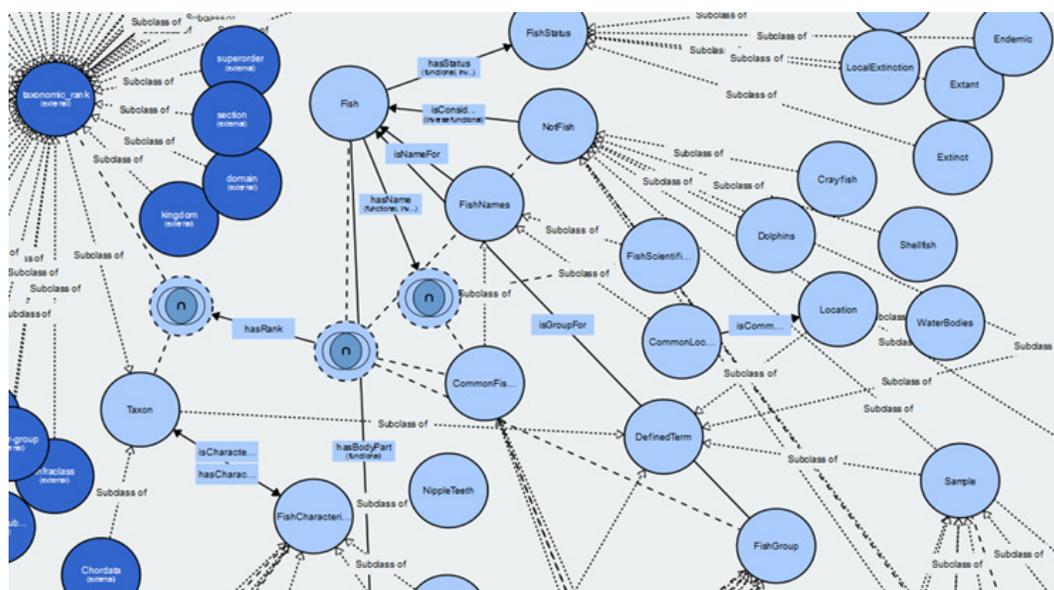
130   The FO is created with the aim of integration and standardization; thus it is imperative to ensure that
131 the terms in the ontology have a unique identifier (ID) that has not been used in other ontologies. A unique
132 ID for a term allows cross-referencing between related databases and ontologies, without the confusion
133 of same existing terms with different functions. There are many ways to create a unique ID; however
134 following an example of a globally accepted guideline will ease future integration with the FO. As such,
135 we adopted the guidelines issued by the Open Biological and Biomedical Ontologies (OBO Foundry) ("The
136 OBO Foundry"; Smith et al., 2007) and created each term in the FO using an ID which starts with the prefix
137 'FO' followed by unique digit numbers (e.g. "FO_XXXXXX" where X is a digit).
138   There are many tools created for ontology validation such as the inference and rule engine. However,
139 it is apparent that human validation is still mandatory in the current state of practice for ontology
140 learning (Zhou, 2007). Furthermore, most ontology learning results have mainly been evaluated by
141 domain experts manually. As such, a logical evaluation was conducted by fish experts to verify the naming
142 of concepts and to validate the hierarchy of the terms, which the FO presented. Criteria such as accuracy,
143 complexity, semantic consistency, terms redundancy, naturalness, precision, completeness, and
144 verifiability were checked using questions such as "what if we do not know the name of the specimen?",
145 "what if we only know its common name?", "what if the specimen is similar to certain kind of known
146 specimen?", or "what if we were to find a completely unknown specimen?".
147

148
149
150   **Figure 1.** Workflow for the Fish Ontology creation.
151

152   The workflow for creating the FO in this paper is separated into five stages which are: (1) Concept
153 Selection, (2) Literature Review, (3) Ontology Creation, (4) Ontology Testing, and (5) Ontology Evaluation.
154 Figure 1 shows the workflow of the FO creation. In the concept selection stage, we first decided on the
155 ontology concept, its possible structure, and future considerations needed to create the ontology. During
156 the literature review, relevant research, such as papers, books, and tools were gathered. Databases and
157 ontologies that are relevant for adoption into the FO were also researched at this stage. Next, in the
158 ontology creation stage, terms and relationship to be used in the ontology were defined and a proper

159 structure of the ontology was created. Additionally, at this stage, the naming convention used for the
160 terms were selected, and any relevant databases and ontologies were imported to the ontology.
161 Subsequently, in the ontology testing stage, the functionality of the FO was tested to detect and fix
162 problems with the terms, structure, or relationships. Finally, during the evaluation, the ontology was
163 evaluated by the fish and biodiversity researchers for its applicability. At this stage, mistakes, bugs, or
164 comments were collected and resolved to improve the ontology.
165   In this work, we show the applicability of the FO on several areas such as determining if a specimen
166 is a fish, determining the type of fish based on characteristic(s), morphology, name, or taxonomic rank,
167 determining its conservation status (extant or extinct), and determining whether or not it is an ancient
168 species. Examples of its applicability are presented in the Results section.
169

170 **RESULTS**

171 Fish Ontology Framework and Content

172 The Fish Ontology proposed in this paper contains 1830 classes, 27 object properties, 500 species names,
173 with 1223 synonyms, 8 fish groups, and 9 fish characteristics. It is the first of its kind to provide automated
174 fish classification based on taxonomic rank, group, name and characteristics. Given that the FO is intended
175 for broad classification of fish, including common extinct ones, the FO is able to classify jawless fish, early
176 jawed fish and living fossil fish. A graphical illustration of several main classes in the FO and its integration
177 with other ontologies such as the VTO is demonstrated in Figure 2. The online version of the ontology can
178 be accessed at https://mohdnajib1985.github.io/FOWebPage/. The OWL file for the FO with all of its
179 imported classes is available as a supplementary material (Refer to Additional File FishOntology.owl).
180

181
182
183 **Figure 2.** The Fish Ontology (FO) architecture. A portion of the FO is shown here on how the classes are
184   related to each other and to other ontology classes. The dark blue circles represent terms from other
185   ontologies while light blue circles represent terms from the FO.
186   The classes in the FO are created as a base for integration with other ontologies mentioned in Table
187 1, and with any related ontologies that might be useful for fish recognition such as the popular Gene
188 Ontology (GO). For the "Taxon" class, it is organized in single inheritance, up to the species level whenever
189 possible to increase the reasoning capabilities and expand its scope by including relationship(s) and

190 annotation to the terms. This includes imported classes, which are linked to their respective class types.
191 Each FO branch is organized hierarchically by the means of "is_a" (or subclass of) relationship, by
192 appropriately placing it under a single root term. Each class has annotations to enrich its meaning and
193 purpose. Examples of the relationships are shown in Table 2.
194

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

195
196 **Table 2.** Example of relationships in the Fish Ontology.
197
198 The FO contains 253 classes dedicated to fish studies and 38 classes related to fish sampling
199 processes. These classes are well suited for describing sample and specimen related terms. In combination
200 with suitable classes, relations, and annotations, the utility of the FO for automated fish species
201 recognition through sample and specimen data is likely to be improved. Some of the classes such as
202 "FishSampling" and "FishName" are structured in a multiple inheritance structure, with some classes being
203 subclasses of the same class; an example is the class "Trap" which is the subclass of "FishingGear" and
204 "FishSamplingMethod". As aforementioned, most of the new terms were created based on the reference
205 book (Helfman et al., 2009) because to the best of our knowledge, there are no suitable ontologies from
206 which we could import these classes, while some of the terms that we found were poorly defined and
207 structured. However, we have included cross-references of several classes for potential mapping to
208 relevant external resources, including the FishBase, Teleost Taxonomy Ontology (TTO), and National
209 Centre of Biotechnology Information Taxonomy Database (NCBITaxon) (Froese & Pauly, 2000; Midford et
210 al., 2010; Federhen, 2011). Table 3 shows the statistics of cross referencing of the FO classes to other
211 resources.
212

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

213
214 **Table 3.** Statistics for the Fish Ontology cross references.
215 Inference Capabilities

216 We have created relationships which allow a specimen (and sample) to be inferred and automatically
217 analyzed in the areas of fish grouping, taxonomic rank, and common fish names. We focused most of our

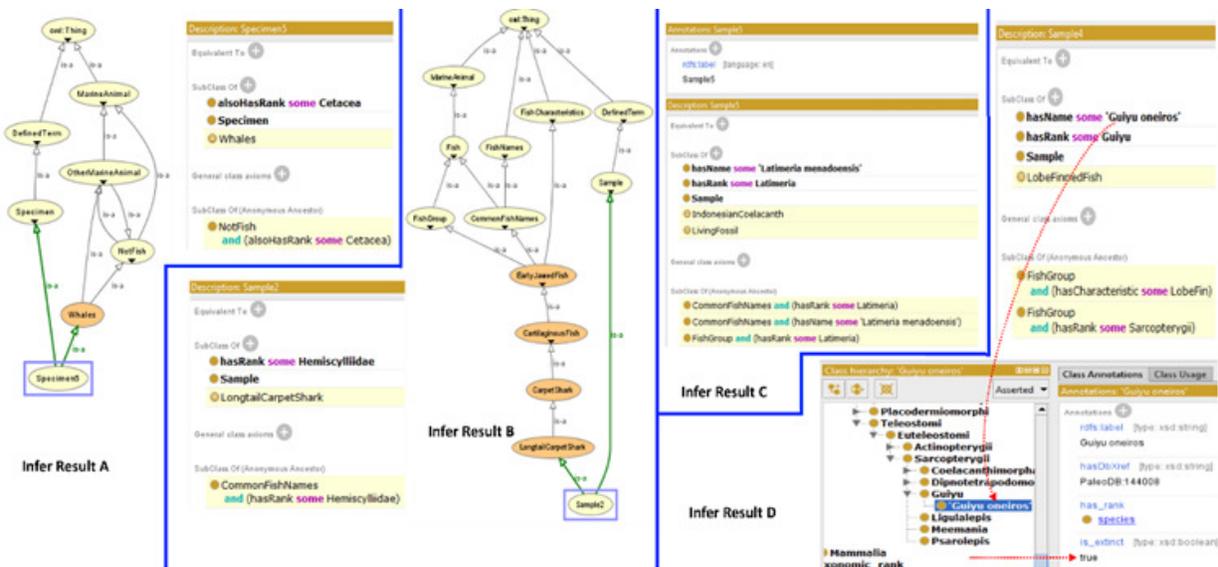
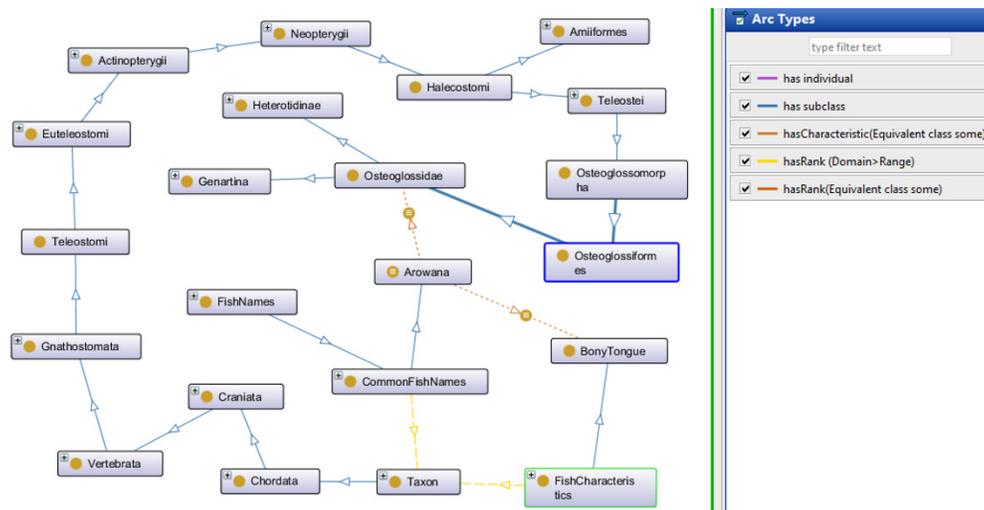
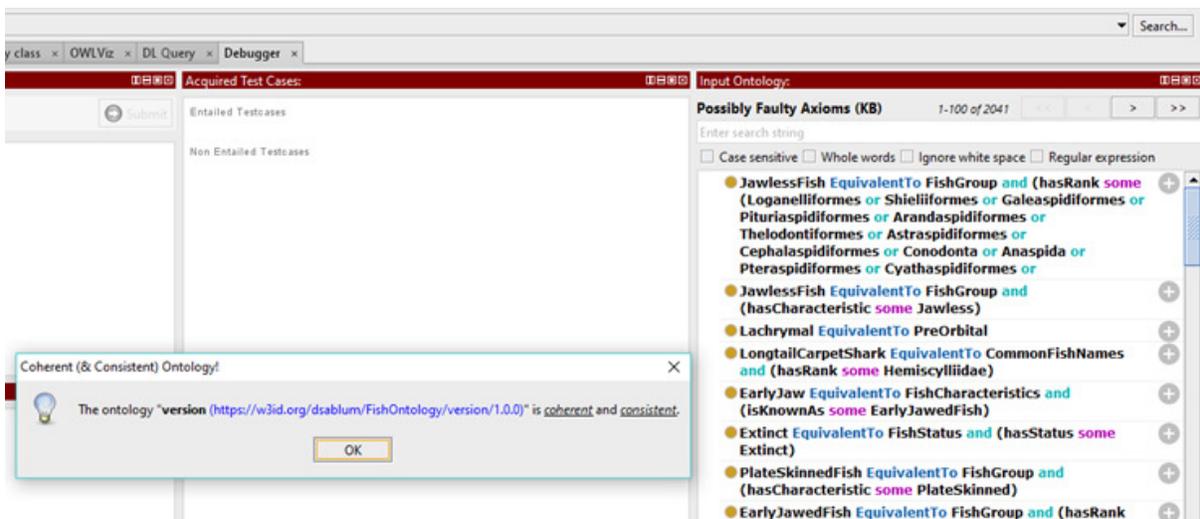
218 modelling activities on these aspects. The specimen (and sample) which is not inferred would only be
219 shown as subclasses of "Sample" or "Specimen" classes; however after being inferred using the reasoner
220 provided by Protégé, each one of them will be properly classified according to their respective parameters.
221 Furthermore the inferred results can show which individual shares the same trait(s) as the sample and
222 suggest what kind of group it fits into based on its characteristics.
223   The FO also provides a structure to determine whether a specimen or a species is actually a fish or
224 otherwise by using the reasoning capability. Figure 3 shows the results of the inferring process which
225 demonstrates the classification of specimen as a fish or otherwise, and what group it belongs to in the
226 taxonomic hierarchy. Infer Result A shows how a specimen (Specimen5) is recognized by the reasoner as
227 a "Whale" and leads the reasoner to recognize it as "OtherMarineAnimal" and "NotFish". Infer Result B
228 shows how a sample (Sample2) is recognized as a "LongtailCarpetShark", which leads the reasoner to
229 recognize that it is a fish. Infer Result C shows that Sample5 is recognized as a "LivingFossil" while Infer
230 Result D shows that Sample4 is actually a species called *Guiyu oneiros* and subsequently recognized as an
231 extinct species.
232

233
234
235   **Figure 3.** Results from the Fish Ontology inferring process.
236
237   Sample queries are presented in Figure 4 using SPARQL (Query A) and SPARQL-DL (Query B).
238 Improved results were obtained using the SPARQL-DL query, which could query inferred data in the
239 ontology compared to a SPARQL query. As shown in Figure 4, new classes were found in Query B results,
240 which are obtained from the imported class and integrated terms from other ontologies. The results in
241 Figure 5 show how additional data can be retrieved from the FO using the DL query.
242

Figure 4. A sample of query to check the inferred results. Results from Query A (using SPARQL) were retrieved before the inferring process, while results from Query B (SPARQL-DL) were 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 rank in the fish taxonomic structure is subsequently inferred.

256 Evaluation

257 To evaluate the quality of the Fish Ontology, we follow the Gruber method for ontology construction
258 (Gruber, 1995). There are five criteria highlighted which are clarity, coherence, extendibility, minimal
259 encoding bias, and minimal ontological commitment. To measure the clarity level of the FO, the ontology
260 definitions should be objective and independent of the social and computational context. In FO, all the
261 definitions are stated in such a way that the number of possible interpretations of a concept would be
262 restricted. The clarity test results for the FO are divided into five parts, which are:
263
264 1.   No Cardinality Restriction on Transitive Properties
265 2.   No Meta-Class
266 3.   No Subclasses of RDF Classes
267 4.   No Super or Sub-Properties of Annotation Properties
268 5.   Transitive Properties cannot be Functional
269
270   Results for clarity test 1 and clarity test 5 are shown in Figure 6 below. Since fish data are large in
271 volume, there is a need to add more data to the FO over time. As such, there is no cardinality restriction
272 assigned to any transitive properties in the FO. Figure 6 also shows that the transitive properties in the
273 FO are not functional because it relates to more than one instance via the property. As for clarity test 2,
274 clarity test 3 and clarity test 4, Figure 7 and Figure 8 show that there are no meta-classes, no properties
275 with a class as a range, and no sub-classes of RDF classes in the FO. All the five clarity tests are
276 automatically performed in the latest Protégé version.
277

278
279
280 **Figure 6.** Results for clarity test 1 and clarity test 5.
281
282   To ensure the coherence quality of the FO, the definitions of concepts given in the ontology as well
283 as the inferences drawn from the ontology must be consistent with its definitions and axioms. Based on
284 our evaluation, most of the inferred terms from the FO appeared to be consistent with its definition and
285 axioms. As an example, in Figure 3 when the FO inferred that specimen5 is a whale, it also inferred that
286 it is not a fish, and it also showed the correct taxon rank. The formal part of the FO is checked by
287 following these 5 consistency criteria listed below and ensuring that all return true:

288   1.   Domain of a Property should not be empty
289   2.   Domain of a Property should not contain redundant Classes
290   3.   Range of a Property should not contain redundant Classes
291   4.   Inverse of Symmetric Property must be Symmetric Property
292   5.   Inverse Property must have matching Range and Domain
293

296   **Figure 7.** Results of the coherence test using Protégé Ontology Debugger tool.
297

298   The usage of Protégé forces the user to always be wary about an empty domain, redundant classes,
299   and properties. As such, coherence tests 1 to 3 are achieved and can be further viewed via the ontology
300   itself. For coherence test 4, we provide an example of the property "isSimilarTo". The class
301   "CosmoidScales" is related to the class "PlacoidScales" via the "isSimilarTo" property. Then we can infer
302   that "PlacoidScales" must also be related to "CosmoidScales" via the "isSimilarTo" property. Figure 7
303   shows the results of coherence test using the Ontology Debugger Tool from Protégé. The coherence test
304   from this tool checks for possible faulty axioms. The ontology passed the coherence test provided by this
305   tool. Figure 8 shows the results for test 5 displaying that the properties "hasCharacteristic" and
306   "isCharacteristicFor" have matching range and domain.
307   An ontology should be extensible which means allowing addition of new concepts, according to the
308   current development in the field (Barbosa & da Silva, 2001). In this paper, the FO is made extensible via
309   the design consisting of concepts, classification hierarchy represented as classes, from general to specific.
310   Applying reasoning to the FO helps to define new concepts (generated from an ontology) from defined
311   generic concepts (books and other databases). As such, the FO place any related concepts derived from
312   other generic concepts in its class hierarchy to represent information that defines a specimen. Classes and
313   annotations that may be useful for future integration such as genetic are added, since such content will
314   further enhance FO's extendibility. Table 1 and Table 3 show all the classes that are linked or cross-
315   referenced by the FO to demonstrate the extendibility of the FO. Since the first FO design, we have
316   considered integrating terms from other ontologies into the FO. By placing any related concepts derived
317   from other generic concepts in its class hierarchy, the FO represents information that defines a fish
318   specimen, linking it with terms from other ontologies.
319

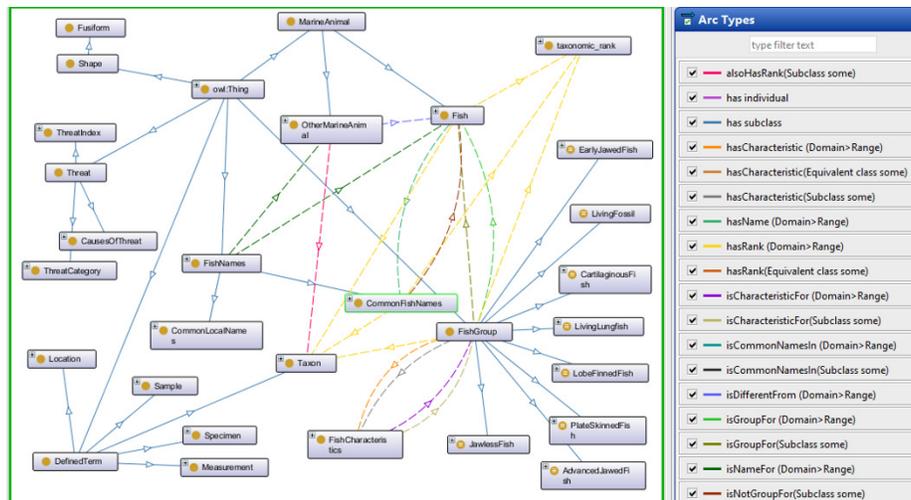



Figure 8. Results for clarity tests (2, 3, and 4) and coherence test (5).

It is a preferred practice to make an ontology which would require minimal ontological commitment so that it is sufficient to support the intended knowledge sharing (Man, 2013). 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 subsequent ontology updates (Nicola D., Missikoff & Navigli, 2005). As such, an ontology should make as few claims as possible about the domain while still supporting the intended knowledge sharing. By reusing existing concepts from books, databases and other ontologies on fish, the FO has low ontological commitment, making it more extendible and reusable (Freitas, Stuckenschmidt & Noy, 2005). Also, since most of the new concepts introduced in the FO are from books and published journal articles, they are likely to be more understandable and preferred 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. As shown in Table 1, Table 2, and Table 3, encoding bias in the FO is reduced by the choices of using OWL as the representation language, and of adopting terms from books, database, and related ontology. As demonstrated in Figure 7, the Protégé Ontology Debugger tools have fully examined all possible encoding biases in the ontology and have cleared the FO as coherent and consistent.

To strengthen the results of the FO evaluation, we use an online ontology evaluation tool named 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 paragraphs 6 and 7 in the discussion section.

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

DISCUSSION

In this paper, we developed a Fish Ontology framework which is a general-purpose ontology that allows integration of fish related ontologies containing standard terms and relationships. The design of the FO is flexible enough to accommodate any 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 to 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

379   other ontologies via reusing of terms allows the search for relevant information such as genetic data of a
380   specific fish species. In this way, the FO is able to produce new knowledge which is useful to biologists.
381   The current version of the FO can utilize specimen grouping and characteristics to determine
382   whether a specimen is a fish or otherwise, provide taxonomic information and heredity of a characteristic
383   rank, and determine conservation status, evolutionary status (ancient or modern) and type (jawless fish
384   is an ancient species). The power of the FO lies in its ability to automate group classification, and ability
385   to link the terms used by fish domain researchers, and other researchers outside the domain. This version
386   uses simple character classification where the user provides the necessary character for the specimen. As
387   an example, the user can specify that "Sample 1 has the characteristic of Plate Skinned", and manually
388   add the characteristic of "Plate Skinned" into the FO. We believe the ideal FO version should also contain
389   anatomical and phenotype data from several classes in the ontology such as "AnatomicalCharacters",
390   "MeristicCharacters", "MolecularCharacters", and "Morphometric Characters" and these features will be
391   included in the near future. These classes can be useful for pattern recognition, and species taxon
392   recognition studies.
393   In general, we extract information such as synonym, name, fish grouping, group rank, fisheries, and
394   other fish-related terms from Helfman et al. to form the general structure of the FO. We adopt the usage
395   of the VTO for taxonomic hierarchy, taxonomic related information, and terms related to taxonomic rank.
396   In most cases, the taxonomic structure of the VTO is followed as it is a regularly updated ontology.
397   However, there are exceptions to this, such as the class "Mammalia" which is placed as a subclass of
398   "Sarcopterygii" in the VTO. Although this classification is consistent with the cladistics standpoint that
399   mammals are derived from fish, explicit classification of mammals under Sarcopterygii in the FO would
400   result in the erroneous recognition of a fish-like mammal, e.g. whale, as both a fish and a mammal. We
401   have therefore placed Mammalia under the higher taxonomic rank of Chordata and made annotation
402   within the FO to highlight this choice.
403   The NCBITaxon, an automatic translation of the NCBI taxonomy database into OBO or OWL format
404   (Federhen, 2011), is also used by the FO as a secondary source for terms regarding taxonomic rank. Both
405   the VTO and NCBITaxon differ in hierarchical organization and definitions. One of the most distinctive
406   feature of the VTO compared to the NCBITaxon is its broad taxonomic coverage of vertebrates. The
407   NCBITaxon excludes many extant and nearly all extinct taxa while largely includes only species associated
408   with archived genetic data, complemented by data from the PaleoDB and the TTO to provide an
409   authoritative hierarchy and a richer set of names for specific taxonomic groups (Midford et al., 2013).
410   Therefore the VTO is more relevant to the FO's purpose for a comprehensive fish taxonomy information
411   (Foote; Dahdul et al., 2010), since the VTO is built based on several taxonomic resources including the
412   NCBITaxon the Paleobiology Database (PaleoDB), and the Teleost Taxonomy Ontology (TTO). Having said
413   that, any taxonomic ranks covered by the NCBITaxon but are not covered by the VTO, such as the species
414   *Protanguilla palau* and the subfamily Oxudercinae are incorporated in the FO to improve coverage of fish
415   data. More examples on the differences between the main reference book, the VTO, and the NCBITaxon,
416   as well as what the FO uses are shown in Table 4.
417   New knowledge emerges everyday so there may be a need to add new concepts and relationships to
418   the existing ontology. Proposing new vocabulary in biodiversity is not uncommon, since ontologies in this
419   domain are presently insufficient and many are under development. Available standard vocabulary is not
420   comprehensive enough to cover all the terms needed to make an ontology in the fish domain. In most
421   cases, new terms must be proposed based on the rationale utilized in the ontology. One such example is
422   that of Hymenoptera Anatomy Ontology, where new terms had to be proposed to expand the ontology
423   (Seltmann et al., 2012, 2013). It should be possible to extend an ontology without altering the existing
424   definitions. As such, the need for easy ontology extension is prioritized while creating the FO. The new
425   terms are checked for its suitability to be adopted as a standard vocabulary for fish scientists. The use of
426   adopted terms and concepts from our main references is further clarified with domain experts (Amy Y.

427 Then, Chong V. Ching) in order to represent and map the appropriate contents to reflect the diverse
428 aspects of fish (Helfman et al., 2009).
429

Term Example	Helfman et al. (2009)	VTO (Midford et al., 2013)	NCBITaxon (Federhen, 2011)	Fish Ontology (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 et al. (2009) for labeling.
LobeFinnedFish	Classify as Sarcopterygii (page 4).	No classes and annotations found, but related species are classified.	Classified as Coelacanthiformes	Follows Helfman et al. (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.	Not listed.	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.

430
431   **Table 4.** Term adoption example in the Fish Ontology.
432

433 Regarding ontology evaluation, there are reasons a number of errors were flagged by the OntOlogy
434 Pitfall Scanner (OOPS!) but none can be detected by using the tools from Protégé. The most apparent
435 reason is because the scope of evaluation for both methods are different. In Protégé, only the classes and
436 its relationship structures created in the ontology are being evaluated, while in OOPS!, the classes,
437 relationships, mapping and future integration problems are being evaluated, giving different results. A
438 number of errors detected by OOPS! can be attributed to the important FO feature of reusing terms from
439 other ontologies in order to reduce redundancy in global usage. As mentioned earlier, terms and
440 structures taken from other ontologies have their own unique ID and metadata to indicate associated
441 function. However, since most of these terms are directly used in the FO, the OOPS! tool flagged these
442 occurrences as critical errors such as "P24: Using recursive definitions", "P32: Several classes with same
443 labels", and "P40: Namespace hijacking".
444 Other pitfalls such as P02, P04, P08, P11, P13, P30, P36, and P41 (refer to Figure 7) are considered
445 acceptable since there are constantly new items to be added to the ontology along with the necessary
446 annotations, relations and property constraints. As an example, errors flagged under pitfall "P19: Defining

447 multiple domains or ranges in properties", is due to the modeling of FO for increased inference capacity
448 (Poveda-Villalón, Gómez-Pérez & Suárez-Figueroa, 2014). In a typical ontology, inferring capabilities is
449 used to discover new relationships. In our work, we used inferring capabilities for automated fish species
450 recognition. As such, we had to avoid using 1 to 1 relationships for the domain and range, instead we
451 expanded the domain and range of each property for a more reliable automated species discovery.
452 Although OOPS! flagged these property expansions as pitfall errors, we deem these as minor for the
453 purpose of the FO development.
454

	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 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 red list 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.	Searching a fish species through the MarineTLO owl file is not possible. However its competency query v4 suggested that it covers a wide range of search topics such as species and 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.	Searching a fish species through the NFO owl file is also 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)	When FO search for a fish, it provide its taxon information, 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.

455
456   **Table 5.** Differences between Fish Ontology with other related ontology and database.
457

458   The FO covers the terms for fish domain which are not well described by other ontologies, such as
459 terms used for fish studies ("maturity", "age") or sampling experiment ("weight", "length"), particularly
460 those related to automatic classifications, annotations, and relationships. There are however some terms
461 in the FO created using parameters rarely used outside of this domain, such as "FishDatabases" which are
462 for any known databases for fish, or "GasBladder" which is a specific organ for "Actinopterygii". The
463 differences between the FO and other fish related ontologies and databases is its ability to provide
464 automated classification of unknown specimen.

465   There has been efforts to create ontologies for recognition purposes such as Hymenoptera (Balhoff
466 et al., 2013), however in the fish domain, ontologies were created to focus more on classification rather
467 than recognition, such as the MarineTLO, NFO, TTO and the ZFA. In this paper, the FO was created to focus
468 on automated fish recognition. The comparison of FO with other related ontologies in the fish domain is
469 presented in Table 5. For the purpose of this paper, we considered ontologies which are most related to
470 Fish Ontology while not taking into consideration systems that use these ontologies as their underlying
471 framework. FishBase was included as it is the most referred portal in this domain (600000 Visits/Month).
472 Development of the FO for classification of several highly diverse groups, such as bony fishes, advanced
473 jawed fish, sharks, skates, and rays, is an ongoing effort.

474   We envision the FO to expand by incorporating additional components such as fish models, fisheries
475 parameters, gene annotations and other relevant information as aforementioned. These parameters will
476 further enhance fish recognition capabilities to recognize fish based on physical features or gene
477 annotations. We will focus on parameters that influence the grouping process such as shape and
478 characteristic recognition, and anatomical metric distinctions. Other than including more terms and
479 defined relationships, we are considering to increase granularity by linking to other relevant and
480 established ontologies, such as the Gene Ontology (GO), the Zebrafish Information Network (ZFIN), the
481 Vertebrate Skeletal Ontology (VSAO), and the Protein Ontology (PO) (Ashburner et al., 2000; Sprague,
482 2003; Dahdul et al., 2010; Natale et al., 2011). In the near future we aim to integrate the FO with other
483 ongoing efforts in our research group such as the Otolith Ontology, Monogenean Ontology, and the
484 Monogenean Haptoral Bar Ontology (MHBI) (Abu et al., 2013). There is also consideration to link related
485 ontologies to existing fish databases using the FO as a mediator ("Fish Stocking Database", "FISH-BOL",
486 "Fish Species Database", "NZ Freshwater Fish Database"; Froese & Pauly; Shao, 2001). Furthermore we
487 also hope to evolve the FO so that in the future, our other ongoing works on different type of fish related
488 recognition tools or technique can be applied to enhance its inferencing capabilities (Abu et al., 2013a;
489 Leow et al., 2015; See et al., 2016; Salimi et al., 2016; Wong et al., 2016; Kalafi et al., 2016).

490   The annotation of fish and fisheries resources in the FO and other related ontologies is a response
491 to the emerging need for data sharing and integration especially for fish data resources (Ashburner et al.,
492 2000; Gangemi et al., 2004; Bizer et al., 2009; Dahdul et al., 2010, 2012, Midford et al., 2010, 2013;
493 Federhen, 2011; Natale et al., 2011; Schriml et al., 2012; Tzitzikas et al., 2013; Van Slyke et al., 2014;
494 Pesquita et al., 2014) and will be highly relevant for the future of fish and fisheries related research.

495

CONCLUSION

497

498 An ontology for the fish and fisheries domain with automated fish recognition is introduced and discussed
499 in this paper. The Fish Ontology (FO) is a new ontology with the feature of taxonomic-based recognition
500 of fish by importing existing ontologies related to fish such as the VTO, ZFA, and TTO. The ontology infers
501 information based on criteria such as names, rank, or characteristics, thus allowing recognition from
502 specimen characteristics. The base terms are taken or imported from related ontologies or naming
503 standards which enhanced the FO's fish recognition and cross-referencing capabilities. The potential
504 usage of the ontology is huge especially as a comprehensive information provider for interested users

505 such as fishermen, museums, restaurants, or for research purposes. More importantly, the FO could be
506 used as a framework to build semantic web systems for data integration to be applied in biodiversity
507 research in the fish and fishery domain.
508
509
510

511 **REFERENCES**

512

513 Abu A., Lim S., Sidhu A., Dhillon S. 2013a. Biodiversity image retrieval framework for monogeneans.
514   *Systematics and Biodiversity* 11:19–33.

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

517 Alexander K. 2013.The Difference Between a Triplestore and a Relational Database. *Available at*
518   *http://www.krisalexander.com/uncategorized/2013/07/16/the-difference-between-a-triplestore-*
519   *and-a-relational-database/* (accessed July 27, 2017).

520 Ankolekar A., Krotzsch M., Tran T., Vrandecic D. 2007. The two cultures: Mashing up Web 2.0 and the
521   Semantic Web. In: *Proceedings of the 16th international conference on World Wide Web*. ACM
522   Press, 825–834.

523 Ashburner M., Ball CA., Blake JA., Botstein D., Butler H., Cherry JM., Davis AP., Dolinski K., Dwight SS.,
524   Eppig JT., Harris MA., Hill DP., Issel-Tarver L., Kasarskis A., Lewis S., Matese JC., Richardson JE.,
525   Ringwald M., Rubin GM., Sherlock G. 2000. Gene Ontology: Tool for The Unification of Biology.
526   *Nature Genetics* 25:25–29.

527 Baader F., Calvanese D., McGuinness DL., Patel-Schneider P., Nardi D. 2003. *The description logic*
528   *handbook: theory, implementation, and applications*. Cambridge, England: Cambridge university
529   press.

530 Balhoff JP., Mikó I., Yoder MJ., Mullins PL., Deans AR. 2013. A Semantic Model for Species Description
531   Applied to the Ensign Wasps (Hymenoptera: Evaniidae) of New Caledonia. *Systematic Biology*
532   62:639–659.

533 Barbosa GP., da Silva FQB. 2001. An Electronic Marketplace Architecture Based on Technology of
534   Intelligent Agents and Knowledge. In: *E-Commerce Agents, Marketplace Solutions, Security Issues,*
535   *and Supply and Demand*. Springer-Verlag, 39–60.

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

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

541 Campbell GA. 2006. *Ontology for Call Control*. Department of Computing Science and Mathematics,
542   University of Stirling.

543 Caracciolo C., Heguiabehere J., Gangemi A., Baldassarre C., Johannes K., Taconet M. 2012. Knowledge
544   management at FAO: A case study on network of ontologies in fisheries. *Ontology Engineering in a*
545   *Networked World*:383–405.

546 Catalog of Fishes Classification. *Available at http://www.calacademy.org/scientists/catalog-of-fishes-*
547   *classification/* (accessed July 27, 2017).

548 Chapman AD. 2009. *Numbers of Living Species in Australia and the World*.

549 Chong VC., Lee PKY., Lau CM. 2010. Diversity, extinction risk and conservation of Malaysian fishes.
550   *Journal of fish biology* 76:2009–2066.

551 Dahdul WM., Balhoff JP., Blackburn DC., Diehl AD., Haendel MA., Hall BK., Lapp H., Lundberg JG.,
552   Mungall CJ., Ringwald M., Segerdell E., Van Slyke CE., Vickaryous MK., Westerfield M., Mabee PM.
553   2012. A unified anatomy ontology of the vertebrate skeletal system. *PloS one* 7:e51070.

554 Dahdul WM., Lundberg JG., Midford PE., Balhoff JP., Lapp H., Vision TJ., Haendel MA., Westerfield M.,
555   Mabee PM. 2010. The teleost anatomy ontology: anatomical representation for the genomics age.
556   *Systematic biology* 59:369–383.

557 Emmanuel FA., Angermeier PL. 2013.FishTraits. *Available at http://www.fishtraits.info/* (accessed May
558   22, 2016).

559 Eschmeyer WN., Fricke R., Fong JD., Polack DA. 2010. Marine fish diversity: history of knowledge and
560   discovery (Pisces). *Zootaxa* 2525:19–50.

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

563 Federhen S. 2011. The NCBI Taxonomy Database. *Nucleic Acids Research* 40:136–143.

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

565 Fish Species Database. *Available at https://www.igfa.org/fish/fish-database.aspx* (accessed May 22,
566   2016).

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

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

569 Freitas F., Stuckenschmidt H., Noy NF. 2005. Guest editor's introduction: Ontology issues and
570   applications. *Journal of the Brazilian Computer Society* 11:5–16.

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

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

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

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

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

580 Heath T., Bizer C. 2011. Linked Data: Evolving the Web Into a Global Data Space. *Synthesis lectures on*
581   *the semantic web: theory and technology* 1:136.

582 Helfman GS., Collette BB., Facey DE., Bowen BW. 2009. *The diversity of fishes: biology, evolution, and*
583   *ecology*. John Wiley & Sons.

584   Horridge M. 2010.Owlviz. *Available at http://protegewiki.stanford.edu/wiki/OWLViz* (accessed May 25,
585   2016).

586   Horridge M., Knublauch H., Rector A., Stevens R., Wroe C., Jupp S., Moulton G., Drummond N., Brandt S.
587   2011. A Practical Guide To Building OWL Ontologies Using Protégè 4 and CO-ODE Tools Edition 1.3.
588   *The University Of Manchester*:107.

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

590   Kalafi EY., Tan WB., Town C., Dhillon SK. 2016. Automated identification of Monogeneans using digital
591   image processing and K-nearest neighbour approaches. *BMC Bioinformatics* 17:755–778.

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

595   Leow LK., Chew L-L., Chong VC., Dhillon SK. 2015. Automated identification of copepods using digital
596   image processing and artificial neural network. *BMC Bioinformatics* 16:S4.

597   Man D. 2013. Ontologies in Computer Science. *Didactica Mathematica* 31:43–46.

598   Midford P., Balhoff J., Dahdul W., Kothari C., Lapp H., Lundberg J., Mabee P., Vision T., Westerfield M.
599   2010. The Teleost Taxonomy Ontology. *Nature Precedings*.

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

603   Natale DA., Arighi CN., Barker WC., Blake JA., Bult CJ., Caudy M., Drabkin HJ., D'Eustachio P., Evsikov A
604   V., Huang H., Nchoutmboube J., Roberts N V., Smith B., Zhang J., Wu CH. 2011. The Protein
605   Ontology: a structured representation of protein forms and complexes. *Nucleic acids research*
606   39:D539–D545.

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

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

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

613   NZ Freshwater Fish Database. *Available at https://www.niwa.co.nz/our-services/online-*
614   *services/freshwater-fish-database* (accessed May 22, 2016).

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

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

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

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

623   Salimi N., Loh KH., Kaur Dhillon S., Chong VC. 2016. Fully-automated identification of fish species based

624   on otolith contour: using short-time Fourier transform and discriminant analysis (STFT-DA). *PeerJ*
625   4:e1664.

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

628 See M., Marsham S., Chang CW., Chong VC., Sasekumar A., Dhillon SK., Loh KH. 2016. The use of otolith
629   morphometrics in determining the size and species identification of eight mullets (Mugiliformes:
630   Mugilidae) from Malaysia. *Sains Malaysiana* 45:735–743.

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

633 Seltmann K., Yoder M., Miko I., Forshage M., Bertone M., Agosti D., Austin A., Balhoff J., Borowiec M.,
634   Brady S., Broad G., Brothers D., Burks R., Buffington M., Campbell H., Dew K., Ernst A., Fernandez-
635   Triana J., Gates M., Gibson G., Jennings J., Johnson N., Karlsson D., Kawada R., Krogmann L., Kula R.,
636   Mullins P., Ohl M., Rasmussen C., Ronquist F., Schulmeister S., Sharkey M., Talamas E., Tucker E.,
637   Vilhelmsen L., Ward PS., Wharton R., Deans A. 2012. A hymenopterists' guide to the Hymenoptera
638   Anatomy Ontology: utility, clarification, and future directions. *Journal of Hymenoptera Research*
639   27:67–88.

640 Shadbolt N., Hall W., Berners-Lee T. 2006. The semantic web revisited. *IEEE Intelligent Systems* 21:96–
641   101.

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

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

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

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

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

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

656 The OBO Foundry. *Available at http://obofoundry.org/* (accessed May 24, 2016).

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

659 Tzitzikas Y., Allocca C., Bekiari C., Marketakis Y., Fafalios P., Doerr M., Minadakis N., Patkos T., Candela L.
660   2013. Integrating Heterogeneous and Distributed Information about Marine Species through a Top
661   Level Ontology. In: *Research Conference on Metadata and Semantic Research*. 289–301.

662 Tzitzikas Y., Allocca C., Bekiari C., Marketakis Y., Fafalios P., Doerr M., Minadakis N., Patkos T., Candela L.
663   2016. Unifying heterogeneous and distributed information about marine species through the top
664   level ontology MarineTLO. *Program* 50:16–40.

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

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

668   Wong JY., Chu C., Chong VC., Dhillon SK., Loh KH. 2016. Automated otolith image classification with
669   multiple views: an evaluation on Sciaenidae. *Journal of Fish Biology* 89:1324–1344.

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

672