Alliance of Bioversity International and the International Center for Tropical Agriculture (CIAT) of France and co-organization of World fish

The subject of

Development of Small-scale fisheries and Aquaculture Ontology

Jahan Tusnuva

MASTER 2 EURAMA

May 2021

Table of contents

Table of the Figures

List of Abbreviation

CIAT	International Center for Tropical Agriculture
CGIAR	Consultative Group on International Agricultural Research
CoP	Ontologies Community of Practice
FAO	Food and Agriculture Organization
AIMS	Agricultural Information Management Standards
UN	United Nations
URI	Uniform Resource Identifier
EBI	European Bioinformatics Institute
OLS	Ontology Look-up services
CO	Crop Ontology
FAIR	Findable, Accessible, Interoperable, and Reusable
DOI	Digital Object Identifier
ASFA	Aquatic
KOS	Knowledge organization system
UNBIS	United Nations Bibliographic Information System

Acknowledgements

This report concludes a period of 6 months of internship within the “Development of small-scale fisheries and Aquaculture ontology” team of the Alliance Bioversity-CIAT and WorldFish. This internship was my first professional experience in the field of Agri-food and aquaculture sector, and I want to thank all the contributors who provided guidance and inputs into my internship and who helped me during the writing of this report.

I thank the “Development of small-scale fisheries and Aquaculture ontology” team of The Alliance Bioversity CIAT and The WorldFish for allowing me to do this internship. I thank Elizabeth Arnaud, (Alliance Bioversity-CIAT), my internship tutor and I am grateful to get her my supervisor. I also thank Jacqueline Muliro (WorldFish) and Marie-Angelique Laporte (Alliance Bioversity-CIAT), for their well welcoming and sharing experiences.

I thank my tutor Elizabeth Arnaud, Elizabeth for her availability, suggestion, advice, and her patience which allowed me to feed my reflection, as well as Professor Celine Domange of the EURAMA Master Program. I thank all who helped me during this work, I thank Ghazali, Saadiah (WorldFish) and all WorldFish scientist who gave their precious time to provide information to improve these activities.

I especially want to thank Marie Angelique Laporte, Digital Solution Development of the "Knowledge organization" sector for her advice, her ability to describe the linking term and to develop data annotation for ontology. I also special thanks to the FAO AGROVOC team, Kristin Kolshus, (PSUA) and Andrea Turbati for their advice, suggestions, and guidelines. I also thank to Alexander Tilley (WorldFish) for sharing his knowledge and research experiences. Thanks once again my tutor Elizabeth Arnaud for editing my report and correction of my English.

Finally, I thank my parents, my sister, and my partner for their support throughout this experience.

1 Introduction

Ontology plays an important role in Semantic Web for Life Science. The Semantic Web is a technology, with the capability to give well-defined meaning to information and to enable better interaction between human and computers which provides a promising platform for researchers who are interested to link and share data in open repository. The deployment of the Semantic Web includes the ontologies and tagged documents and comparison to the huge number of domains. Semantic Web technologies have unlocked a new approach to manage life science research data.

Currently, the rapid development of the digitalization of research in fishery and aquaculture sector as well the information resources are rapidly increasing. Different type of website provides various types of domain information. Huge numbers of research and survey data are available online but not easily findable and reusable due to a lack of standardization and proper labeling. One solution of is the use of an ontology to provide both human and machine-readable labels.

An ontology is kind of dictionary bearing a formal and logical representation of a disciplinary domain, providing a semantic standard that can be employed to annotate data, and where key concepts are defined, as well as the relationships between concepts (Gruber, et al., 2009) (1).Ontologies provide a common language for different kinds of data to be easily interpretable and interoperable allowing easier aggregation and analysis. Ontologies compile information about the content of a dataset that can be explicitly used by computers that help to sharing and added value in a data. (R., et al., 2012)(2).

Heterogeneous and multidisciplinary data is generated by research on sustainable global agriculture and agri-food systems. This data is analysed and often integrated into predictive models for climate change or decision-making tools for farm production and rural advisory services. Fish is vital source of food, nutrition, and income for millions around the world. Around 800 million people depend on fishing and aquaculture for their livelihoods. Among the fifteen centres of the CGIAR, the WorldFish (CGIAR) research aims to improve the sustainability, productivity and resilience of fish agri-food systems and collects fish-related datasets, which include fish health, diseases, breeding, genetics catch data and amongst others (Arnaud, et al., 9 october 2020)(3)

Harmonizing fish data labelling with controlled vocabularies will enable easier data aggregation, interpretation, and analysis. One available ontology, FISHO (https://bioportal.bioontology.org/ontologies/FISHO)(4), focuses on ichthyology, diversity, and adaptation (Ali, et al., 2017)(5).The Food and Agriculture Organization (FAO) of the United Nations initiated several fisheries ontologies, but the ontologies available remained drafts (Caracciolo, et al., , 2012: 383-405)(6). Therefore, CGIAR and relevant partners formed the Small-scale Fisheries and Aquaculture Ontology Working Group, in May 2019, to compile, update and contribute fishery related terms to existing controlled vocabularies. The working group plans to collaborate with the other

animal science partners towards developing and adopting animal ontologies within CGIAR (Arnaud, et al., 9 october 2020) (3) (https://www.cgiar.org/).

The Ontologies Community of Practice (CoP) is led by the Alliance Bioversity-CIAT and was launched as a part of the CGIAR Platform for Big Data in Agriculture. The CoP aims at harnessing relevant expertise in ontology development and identifying innovative solutions that support quality data annotation. The CoP develops knowledge areas applicable to the Platform with the intent of defining the best practices and instructions for collection of reference ontologies (https://bigdata.cgiar.org/communities-of-practice/ontologies/). The CoP provides the ideal forum for co-learning and knowledge exchange on ontologies and for guiding consistent data annotation, as well as the deployment of quality ontologies in databases and repositories (7). The CoP stimulates exchanges between domain experts and experts in ontology design, knowledge modeling, ontology-driven applications, and semantic web technologies. CGIAR Big Data-Ontologies COP is composed to this day of 144 active members: 35 from universities, 61 from public research institutes, and 48 from the private sector (Arnaud, et al., 9 october 2020)(3).

1.2 Objectives of the project

In 2018, following a webinar given to WorldFish, the data management team proposed to start the ontology with two domains: Small-scale Fisheries and Aquaculture.

The ontology project 's objective is to improve the WorldFish data interoperability into the various projects, databases and repositories by (a) addressing inconsistent use of fish, fisheries, and aquaculture related terms across the datasets, (b) highlighting the missing terms in the main semantic resources, and lastly (c) connect and collaborate with the Community of Practice.

1.3 Actors of the Project

1.3.1 CGIAR (Consultative Group on International Agricultural Research)

The CGIAR (https://www.cgiar.org/) is a global partnership uniting 15 international organizations that take part in research about food-security. It aims to reduce deficiency in poor or developing countries, increase food security, improve human health and nutrition, and protect environment through sustainable management of natural resources. CGIAR collaborates with hundreds of partner organizations including national and regional research organization and private sector with mix of knowledge, skills, and research facilities.
The CGIAR runs a portfolio of 8 agri-food system research programs among the Fish Research Program (https://www.cgiar.org/research/program-platform/fish/).

1.3.2 Alliance Bioversity -CIAT (International Center for Tropical Agriculture)

The Alliance of Bioversity International and CIAT (https://ciat.cgiar.org/alliance/) delivers research-based solutions that harness agricultural biodiversity and sustainably transformation of food systems to improve people's lives in a climate crisis. The Alliance provides food system solutions at the nexus of agriculture, environment, and nutrition. The Alliance leads the CGIAR Big Data Platform in Agriculture and its Ontologies Community of Practice. The Headquarters are in Italy and Colombia with regional offices in Africa, Asia, Central and South America, and Europe.
(https://ciat.cgiar.org/alliance/).

1.3.3 WorldFish

WorldFish (https://www.worldfishcenter.org/tags/cgiar) is an international, nonprofit research organization and member of CGIAR (Consultative Group on International Agricultural Research). The headquarters of WorldFish are in Penang, Malaysia and the center has a global presence in 20 countries through Africa, Asia and the Pacific with 460 staff representing 30 nationalities.

WorldFish research aims to create, innovate, advance, and translate the scientific research on sustainable aquatic food system into an impact on the human well-being and environment. WorldFish aquatic foods research program is key to achieving Zero Hunger and the other UN Sustainable Development Goals by 2030 with nearly across 20 countries in Asia, Africa and the pacific. This organization uses its experience in fisheries and aquaculture to comply with Sustainable Development Goals.

The Research Program on Fish Agri-Food Systems (FISH) led by WorldFish is a multidisciplinary research program which is designed in collaboration with research partners, beneficiaries, and stakeholders. FISH develops and implements research innovations that optimize the individual and joint contributions of aquaculture and small-scale fisheries to reducing poverty, improving food and nutrition security, and sustaining the underlying natural resources and ecosystems services (https://www.worldfishcenter.org/).

1.4 Open access and Open Data Principles in CGIAR

*"*Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and share alike"(8).The CGIAR Open Access and Data Management Policy ("CGIAR OADM Policy") states that Open Access is the "immediate, irrevocable, unlimited and free access to internet to knowledge items for every person worldwide and unrestricted re-use of content (which may be limited to non-commercial use and/or permitted in compliance with the CGIAR Intellectual Asset Principles) subject to proper attribution"(9).

For a decade, CGIAR Research Centers and Research Programs have been working in agricultural development, producing value outputs which are widely known as global public goods. CGIAR adopted an Open Access-Open Data Strategy (OA-OD), and all 15 Centers signed on to the Open Access and Data Management Policy in 2013 (https://cgspace.cgiar.org/bitstream/handle/10947/2876/Open_Access_and_Data_Management_Policy_(the_Policy)_FAQs_October%2010%202013.pdf?sequence=1), acknowledging the need to make all outputs findable, accessible, interoperable, and reusable. According to its Open Access Policy, the Bill and Melinda Gates Foundation that is a major donor for the CGIAR, is committed to information sharing and transparency. The harmonized introduction of OA-OD through CGIAR would not only allow compliance with a broad amount of donor policies (e.g., Bill and Melinda Gates Foundation Open Access Policy, USAID Open Data Policy), it will encourage researchers to improve research efficiency and enhance innovation and impact in an era of complex and large data sets ("big data") *(https://www.cgiar.org/how-we-work/accountability/open-access/)* (Bill and Melinda Gates, 23 July 2014)(10).

1.4.1 Importance for Research and Development

The rationale behind OA-OD is to achieve the maximum impact to advantage the poor, especially smallholder farmers in developing countries. OA-OD improves the visibility, accessibility and impact of Research and Development activities and improves the performance, efficiency, and effectiveness of research activities. It enables interdisciplinary research; supports novel research literature computation; and allows CGIAR research to benefit the global public. It also ensures that the results of research and activities can make the infrastructure necessary for CGIAR to be at the forefront of the OA-OD for agriculture movement more easily and collectively.

In further support to reach the CGIAR's open access objectives, the Platform for Big data in Agriculture was launched in 2017. It aims to mobilize CGIAR data to accelerate research and new data-driven innovations, build data collaboration internally and externally, and leverage CGIAR expertise while claiming a unique leadership voice in digital agriculture. It also supports and promotes open data. The CGIAR Platform for Big Data created new data-driven capabilities, built new digital partnerships and alliances, developed digital innovations which makes data more efficient and accessible (11).

1.5 Findable, Accessible, Interoperable, Reusable (FAIR) data Principles

FAIR is an evolution and concretization of open data concept and aim to facilitate of data holdings of different stakeholders. FAIR data enable knowledge discovery by assisting humans and machines in their discovery of appropriate scientific data and their associated algorithms and workflows. The FAIR Data Principles is a set of guiding principles to make data Findable, Accessible, Interoperable, and Reusable. (Wilkinson , et al., 2016)(12).

Findable – assigning a globally unique and eternally persistent identifier (like a DOI or Handle), describing the data with rich metadata, and making sure it is findable through disciplinary discovery portals.

Accessible – data and metadata should be retrievable in a variety of formats that are sensible to humans and machines using persistent identifiers.

Interoperable – the description of metadata elements should follow community guidelines that use an open, well defined vocabulary.

Reusable – the data should maintain its initial richness. The description of essential, recommended, and optional metadata elements should be machine processable and verifiable, use should be easy, and data should be citable to sustain data sharing and recognize the value of data (13).

1.6 Semantics webs technology

The deployment of the Semantic Web is supported by the ontologies to tag documents to link data from large number of domains and documents that is exist on the Web. Machine learning, technologies are useful to extract the labelled data and publications for analysis and extraction of patterns. Semantic Web technologies have unlocked a new sclabale and flexible approaches to manage and discover life science research data create a relation system by using knowledge graphs that models the information elements of a domain along with the properties and processes existing between the elements of the domain (14).

In agriculture, for data like crop field trials, those technologies are progressively being adopting adopted to not only integrate data and standardize it through ontologies and controlled

vocabularies in general but mainly to get new insights from experimental data though for instance perform better data modelling, enable data consistency, and ensure a scalable data integration (15).

1.7 GARDIAN harvests metadata

Global Agricultural Research Data Innovation Acceleration Network (GARDIAN) have started 2018. GARDIAN is used as a source of current information for the CGIAR website. GARDIAN allows users to discover and access datasets and publications across CGIAR's 13 Centers regardless of where they are archived.

GARDIAN also enables data exploration—the mapping and visual large datasets has made it possible to access data and publications from national, governmental, and non-governmental organizations such as USDA, USAID, and the World Bank. By the end of 2019, GARDIAN had facilitated the discovery of approximately 191570 publications and 38735 datasets from across the agricultural sector, the large majority of which are open (16).

GARDIAN is CGIAR´s flagship data harvester. It enables the discovery of publications and datasets from across the thirty-odd institutional publications and data repositories from CGIAR Centers and beyond. It´s a key component of the Platform´s objective to establish the infrastructures, tools, and approaches to making CGIAR data Findable, Accessible, Interoperable, Reusable (FAIR)(17).The Platform on Big Data, CGIAR's primary objective is to annotate multidisciplinary research data with the appropriate ontologies for publishing on the GARDIAN platform (https://gardian.bigdata.cgiar.org/), CGIAR's metadata repository, and stimulate the ontology content gap filling rather than developing completely new ontologies. (Leonelli, 8 (2013),)(18).
GARDIAN employs text-mining to enrich the associated metadata to enhance discovery, and will soon test data mining techniques with cleaned, well-annotated datasets to enhance interoperability. Plans for GARDIAN include further demonstration of the value of interoperable data via seamless interactivity of discovered data with key analytical or visualization tools, including models and maps (https://bigdata.cgiar.org/resources/gardian/).(19)

1.8 Open access at WorldFish

WorldFish research outputs is regarded as international public goods, and this organization is committed to the widespread dissemination of all information products, and to this end make its products open. This is in line with both the Center's Research Data Management and Open Access Policy and the CGIAR Open Access and Data Management Policy. Opening WorldFish research, including publications, data and tools, ensures that more people can read and apply research findings, thereby increasing the efficiency, reach and impact of the research. Worldfish uses to types of open repositories: all information products are published in the open D-Space repository and data sets in Worldfish Dataverse repositor. (https://digitalarchive.worldfiscenter.org/)(20).

1.9 D-Space and Dataverse Repositories

D-Space is an open source repository application that allows to capture, preserve and distribute digital material including text, video, audio and data. D-Space provides a way to manage materials and publications in a professionally maintained repository to give greater visibility and accessibility over time. The main role of the D-space is to Facilitates the capture and ingest of materials, including metadata about the materials, easy access to the materials including

Documents, such as articles, working papers, technical reports, conference papers, books, Theses paper, Data sets and lastly the long-term preservation of the materials. There are over 1000 digital repositories worldwide using the D-Space application for a variety of digital archiving needs. The D-Space has many customizable features and tools for managing digital content, enabling digital preservation and providing accessibility to materials (Valorie, 2011)(21). CGIAR has adopted D-Space as the repository of its research publications.

For example, Figure no (1) shows that WorldFish repository is built on the D-Space platform to facilitate access to WorldFish material. *D-Spaces* built-in organizational structure of Communities and Collections.

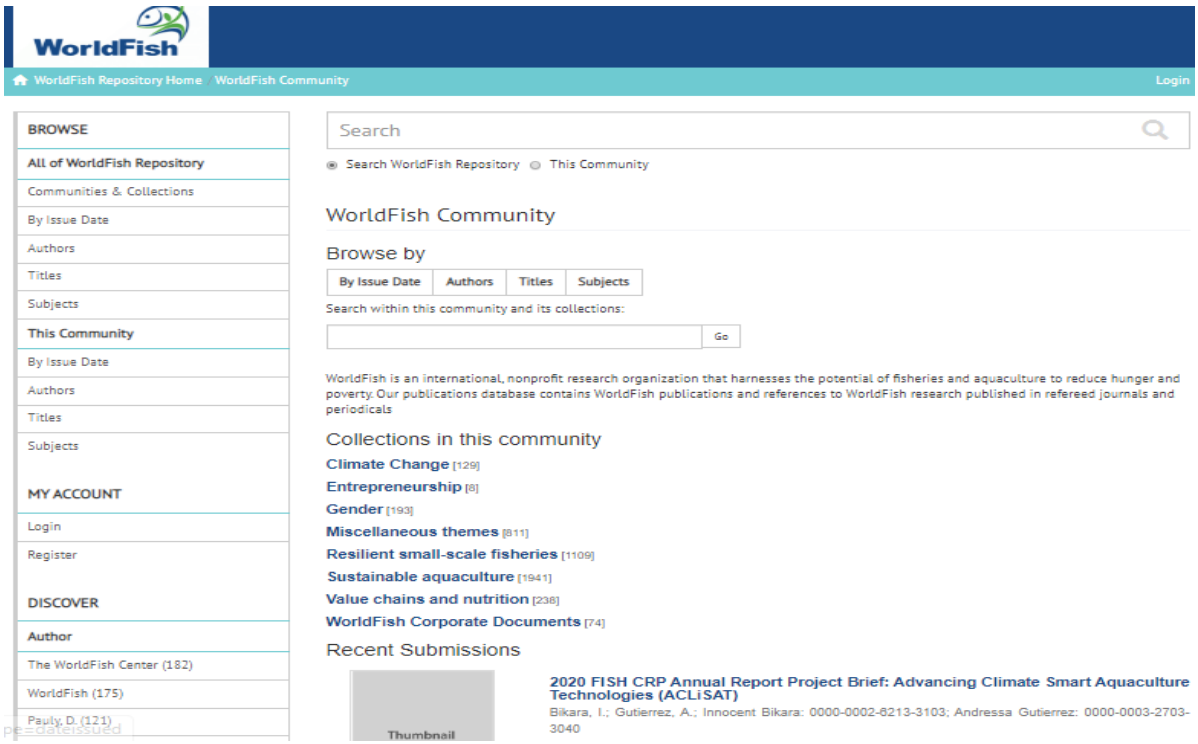


Figure 1: WorldFish open repository or D-spaces, Sources: WorldFish website.

Dataverse is a digital data repository hosted by Harvard University. Dataverse is an Open-source data repository software that is integrated in hundreds leading organizations for the benefit of public by publishing, sharing, extracting, and analyzing data. Harvard Dataverse is the biggest of the existing Dataverse collections with more than 60,000 data sets and is available to researchers from all fields of research. Dataverse enables the allocation of the Digital Object Identifier (DOI) that uniquely identifies the uploaded dataset or publication. Dataverse follows the FAIR data principles enabling access to metadata, data files, dataset terms, and various information. Metadata is defined as the data providing information about one or more aspects of the data; it is used to summarize basic information about data which can make tracking and working with specific data easier (22). Deposits include metadata, data files, and any complementary files such as documentation or codebook or questionary that help to understand the dataset and analysis. A codebook describes the contents, structure, and layout of a data collection. Codebook contains information of variables of data files, methodological details and larger or more complex data

collection instruction. Annexes-1 shown as the codebook are describe the variables details. Metadata is always open for everyone even if data are restricted. However, all datasets have DOI when its Published.

Harvard Dataverse allows to create sub-repositories identified per research institutions. WorldFish has developed its institutional Dataverse (https://dataverse.harvard.edu/ https://dataverse.harvard.edu/dataverse/worldfish) that currently contains 101 Datasets including 431 files of 9 research projects carried out in different territories like Bangladesh, Egypt, Zambia, Nigeria, Myanmar. The WorldFish Dataverse archives has two types of datasets: (a) replication data whereby the results or analysis have been carried out and the results published and (b) project data with restricted access.

Figure 2 shows a record for the consumer study of the Nigeria Scoping study in the WorldFish Dataverse with keywords linked to the AGROVOC thesaurus, a DOI (doi:10.7910/DVN/FJZCJW) and Metadata. (23)

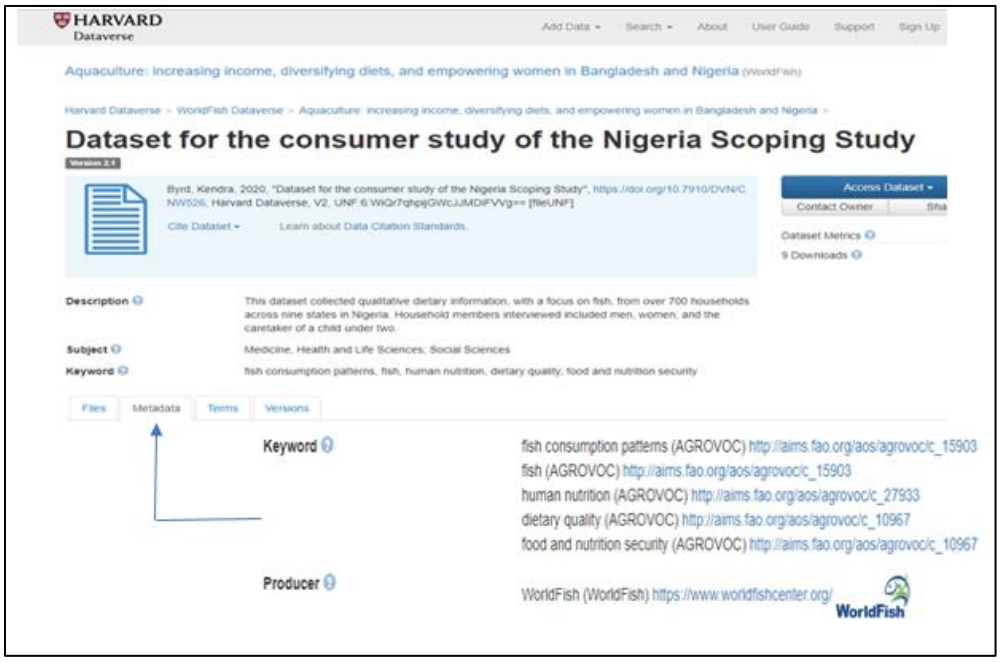


Figure 2:WorldFish Dataverse repositories level of metadata with link of AGROVOC

1.10 What is controlled Vocabularies?

Controlled vocabularies are knowledge organization systems that contain structured set of concepts or terms for organizing and classifying data to ensure its future access and retrieval. The terms are data descriptors related to each other via explicit relationships (hierarchical or

associative). These data descriptors are used to distinguish and define the characteristics of knowledge resources in a specific domain. Using controlled vocabularies to describe the resources makes possible for these resources to be queried, retrieved, analyzed and linked to other relevant information objects. (Harpring, , 2010) (24). Controlled vocabularies are the list of terms (e.g., words, phrases) that is used to tag information in a consistent way. It also provides textual definitions that describe the intended meaning of the classes in vocabularies and standard identifiers for concepts describing a domain. Objective is to facilitate the data publication and data accesses.

For an example figure 3, we can see that subject have always, an ID that provide concept or class which predicate, or property attaches semantics to the subject. Here 3 different properties as follow Pref label = pref term; label = synonym; broader: hierarchy. We can also see that the object can be a concept or class which allow to continue the graph, or it can be called a litteral, which is a string of characters.

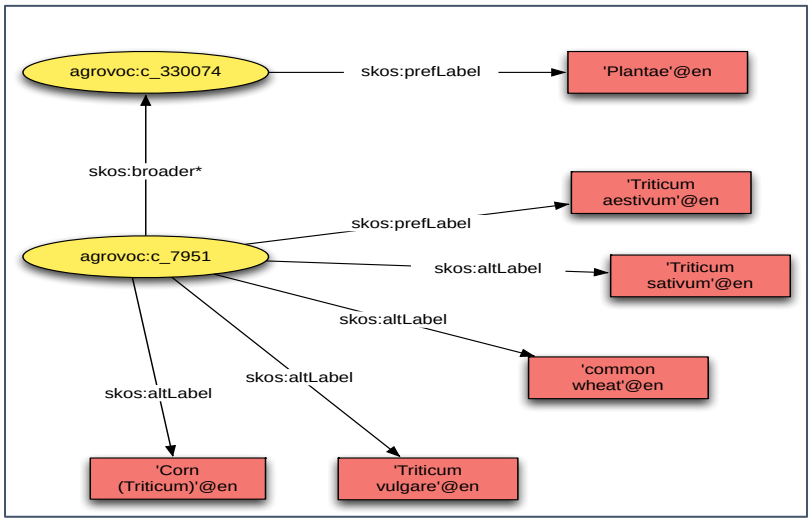


Figure 3Figure 3/Example of Vocabularies; Agrovoc: http://aims.fao.org/aos/agrovoc

1.11 What is AGROVOC?

AGROVOC is a very large-controlled vocabulary thesaurus that consists of around 38,000 concepts with nearly 801,000 terms in up to 40 languages (25). AGROVOC is much used for describing bibliographic references. It is recommended as a source of keywords in the CGIAR core metadata schema that is used by repositories of publications and different data. AGROVOC was established in 1980s as a multilingual structured thesaurus in all fields relating to the agriculture, forestry, fisheries, food, and other related fields such as environment. AGROVOC has almost 22 editors. It is the largest open thesaurus in agriculture sector which is open for everyone and its impact is through providing the access and visibility of data across domains and languages.

AGROVOC has grown considerably to the point that it is now a tool for the organization of explicit knowledge and the development of ontologies, and a multilingual research feature. AGROVOC has been transformed into a concept server at the same time as a terminology-based thesaurus.

It is used to tag content that could be from different Published articles to Datasets or anything that has be published on the web. In this way to simply find to be easy resources information and sematic links.

AGROVOC used the Skosmos Search & browse interface which is simple web-based browser. It is an open-source web-based browser and publishing tool. This interface offers search and browse functionalities, alphabetical and thematic index, structured concept display, visualized concept hierarchy and multilingual user interface. The version of AGROVOC loaded in Skosmos is always the latest release of AGROVOC. Browsing the AGROVOC vocabulary is quite intuitive. In the left panel, it is possible to browse the vocabulary alphabetically or hierarchically. The alphabetical tab shows all terms, including alternative/non-preferred labels. Terms that are not preferred term are not clickable and an arrow links them to the corresponding preferred terms. The hierarchy tab shows only preferred terms. When a user clicks on a concept in the left panel, the right panel shows its properties where have preferred terms in English broader concepts, narrower concepts, alternative labels, labels in other languages, URI, close matches, and exact matches [Fig 4].

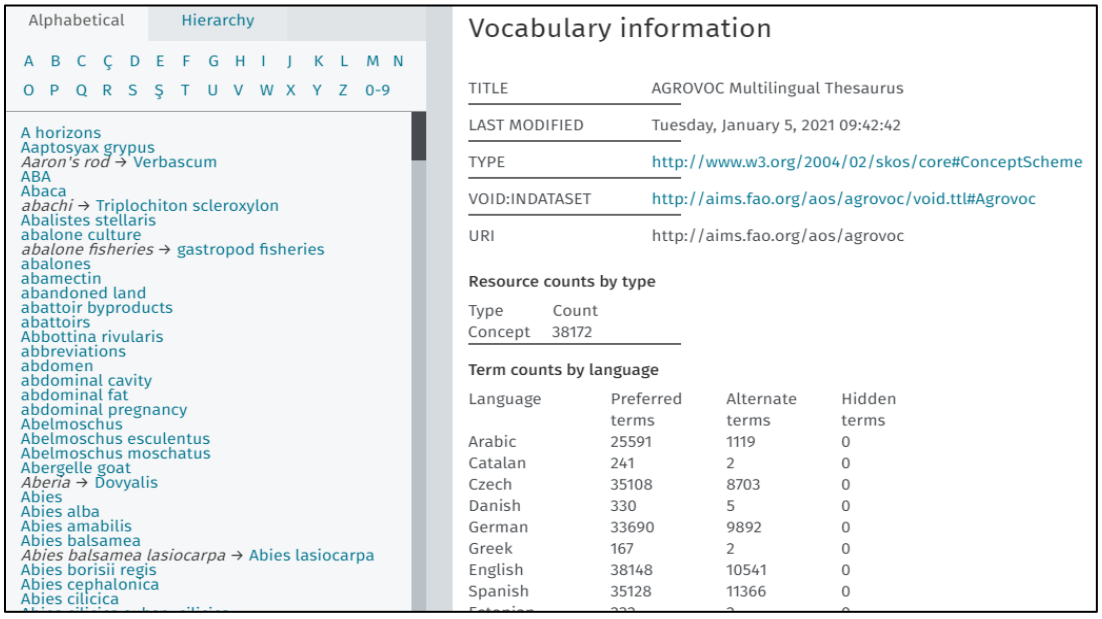


Figure 4:AGROVOC used the Skosmos open search web Search browser with URI, languages

1.12 ASFA (Aquatic Science and Fisheries Abstract)

The ASFA Thesaurus is also a Multilingual controlled vocabulary thesaurus like AGROVOC, and it was established in 1971 with the aim of disseminating aquatic sciences, fisheries, and aquaculture research. The ASFA Thesaurus is an indexing and searching tool. It contains the subject descriptors used to index the content of the Aquatic Sciences and Fisheries Abstracts. It covers the world's literature on the science, technology, management, and conservation of marine, brackish water, and freshwater resources and environments, including their socio-economic and legal aspects. Since 2019, aquatic sciences and fisheries concepts have been added to AGROVOC. ASFA has been integrated in AGROVOC as an independent sub-vocabulary and sharing concepts with AGROVOC and other vocabularies. ASFA uses all the features of the AGROVOC infrastructure, and its

browsing interface is provided by the AGROVOC Skosmos platform. The ASFA Skosmos interface displays only the ASFA scheme, and only shows the ASFA hierarchic relationships. [Fig 5] (26).

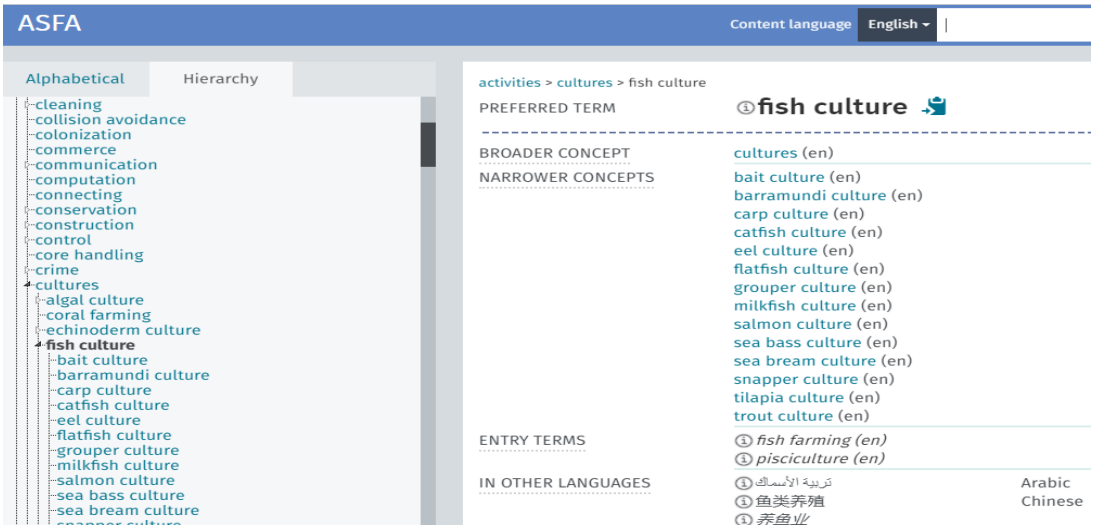


Figure 5: ASFA control vocabulary thesauri

1. 13 The Ontology Lookup Service (OLS)

The Ontology Lookup Service (OLS) (http://www.ebi.ac.uk/ols) provides interactive and programmatic interfaces to query, browse, and navigate an ever-increasing number of biomedical ontologies and controlled vocabularies. Anyone can browse the ontologies through the website. OLS is developed and maintained by the Samples, Phenotypes and Ontologies Team at EMBL-EB. The Ontology Lookup Service (OLS) aims to provide a single point of access to the latest ontology versions. It has included 259 ontologies, 6394,870 terms, 31,301 properties and 497,475 individuals. The core functionality of the OLS provides users with the means to search on controlled vocabulary and ontology terms and synonyms, as well as alternative the relationships between terms and obtain additional metadata (such as definitions, comments, synonyms, or references to other databases and annotations on selected terms. (https://www.ebi.ac.uk/ols/index) (27). [Fig: 6]

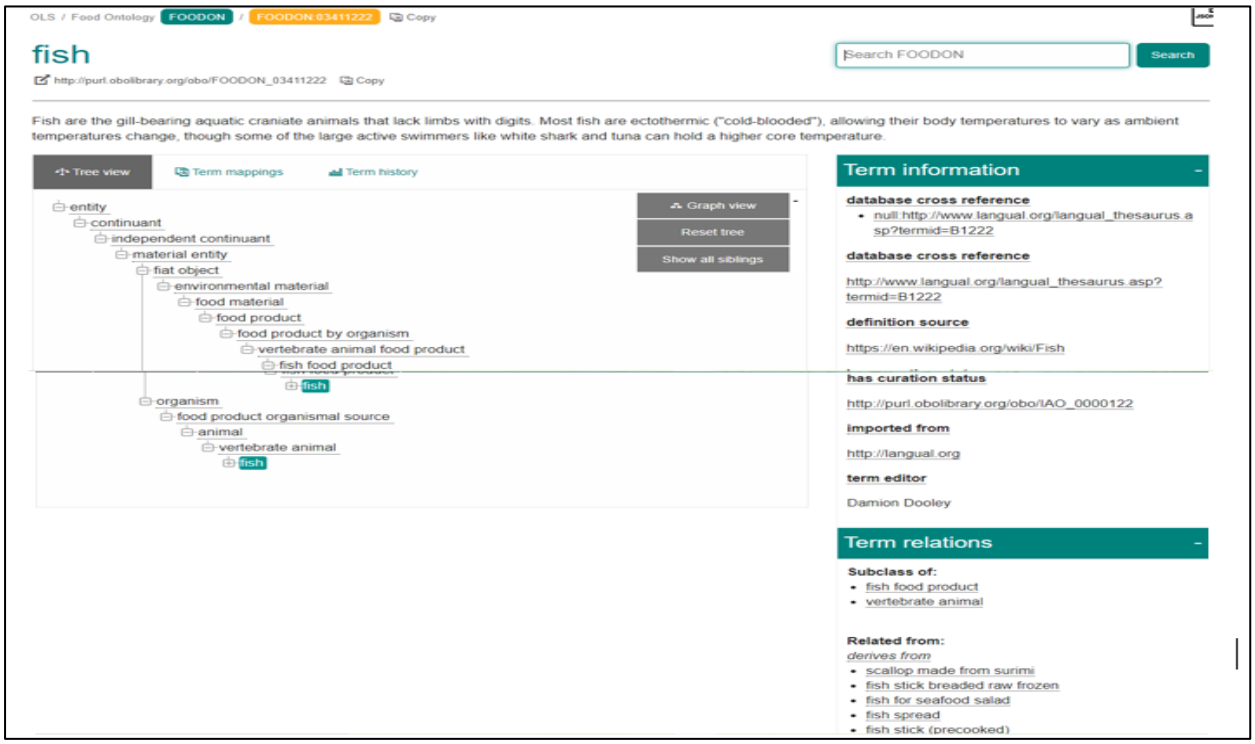


Figure 6: OLS ontological term searching website with URI and term information

2 The challenges

The use of standards for metadata and data annotation plays a key role in addressing the challenge of adequately describing the data sets and annotating the data files in repositories. While the metadata schema, like the CGIAR Core Metadata schema, provides a metadata standard to describe datasets, the Ontology enables the precise labeling of variables in the files using standard naming. While the metadata schema provides a metadata standard to describe datasets, the Ontology enables the precise labeling of variables in the files using standard naming.

Research outputs are properly described using rich metadata to enable better interoperability, easy monitoring, and better transparency.

2.1 FAIR data principles applied to WorldFish-data

WorldFish datasets are not only open, but they are also FAIR. WorldFish like the other CGIAR centers uses the FAIR Data Principles as a guide to make the data Findable, Accessible, Interoperable, and Reusable (FAIR) (Wilkinson , et al., 2016)(28)(29)

The Figure no 7 shows how the WorldFish Datasets scores on the Digital Research Resources (DANS) metrics for FAIR compliance. WorldFish is continuously working towards improving the FAIR score out of 5, Accessible, Interoperable, and Reusable score are 4.04, 4.53, 3.74 and 4.10 accordingly. The development of the fish ontology is one of the aspects of improving the Interoperability score of the datasets. (https://gardian.bigdata.cgiar.org/analytics.center).

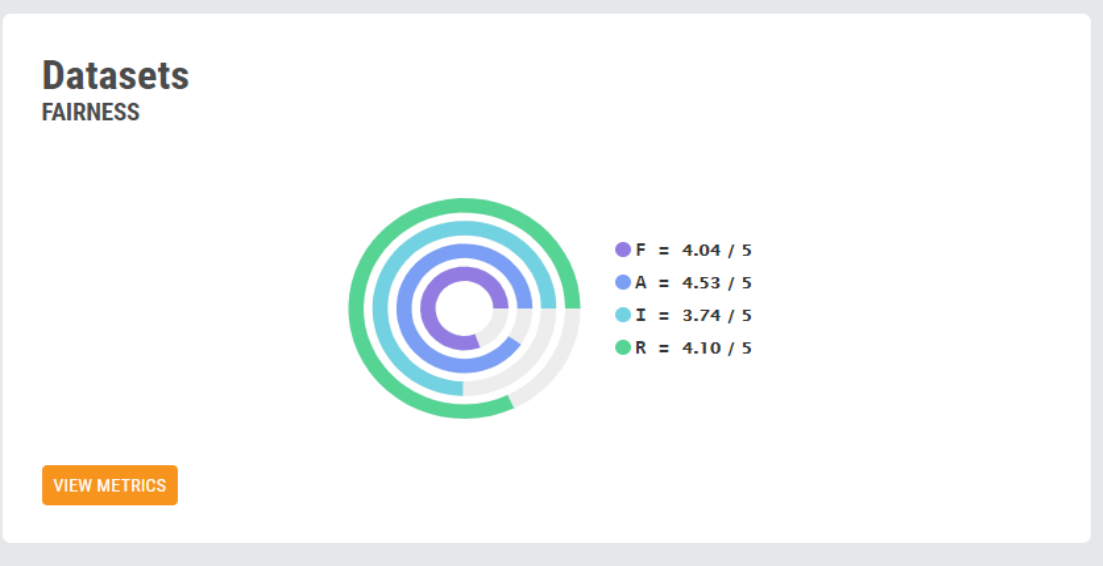


Figure 7: WorldFish Datasets FAIRNESS scoring rate

WorldFish has a total of 101 datasets with 431 files available in the digital repository. To ensure interoperability of FAIR data, datasets must be annotated with an ontology. Currently, only about 5% of the datasets available are annotated due to the lack of ontology.

While WorldFish was working on annotating its datasets with an ontology, one of the key challenges was the limited or lack of ontology in the animal science domain and specifically fish ontology. Some of the concepts had to be borrowed from plant science which in some instances did not make sense in the context of animal science. This led to majority of the data were not a hundred percentage fish related.

3 Methodology

3.1 Selection of the datasets

We selected datasets in the two domains that WorldFish had decided would be the focus of the ontology development. These two domains were Aquaculture and Small-scale fisheries.

Secondly, we looked at their content looking for fish traits, including genetic data, fish species, observations of some parameters of the fish itself. Then, we selected data generated by aquaculture trials that included management practices, environmental factors, schools of fishes and testing of nutritional components. In the small-scale fisheries sub domain, we looked at

monitoring surveys, technology used to fish, quantity of fish catches, type of fish, fisherman, market value of the fish, etc. Initially we started with datasets of The Assessment of Species and Performances in Small-scale Shrimp Farming in Aceh, Indonesia where we had found all the terms.

3.2 Extracting WorldFish concepts

3.2.1 Mapping WorldFish keywords to AGROVOC thesaurus

The keywords in the Dataverse repository are currently linked to AGROVOC. Unfortunately, some of the fish-related concepts are missing which is a major hindrance towards making WorldFish resources more interoperable.

In the initial step was the manual mapping of Dataverse keywords to AGROVOC (FAO) and ASFA (Aquatic sciences and fisheries Abstract). A list was created with terms missing from both AGROVOC and ASFA thesaurus. The Figure [no 8] shows that the WorldFish team had already linked most of the keywords where possible in its data repository to AGROVOC. When the term was exactly the same than in AGROVOC, we qualified the relation as an 'Exact match' and when it was only related, we qualified it as 'Close Match'. The work had to be done manually as they are no algorithm available to do this.

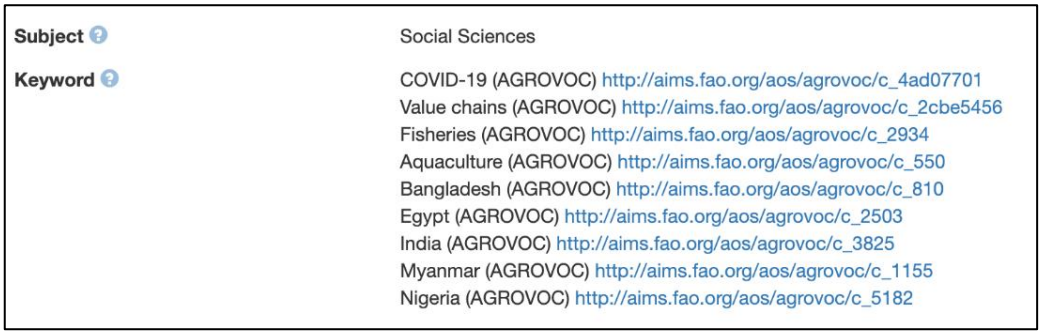
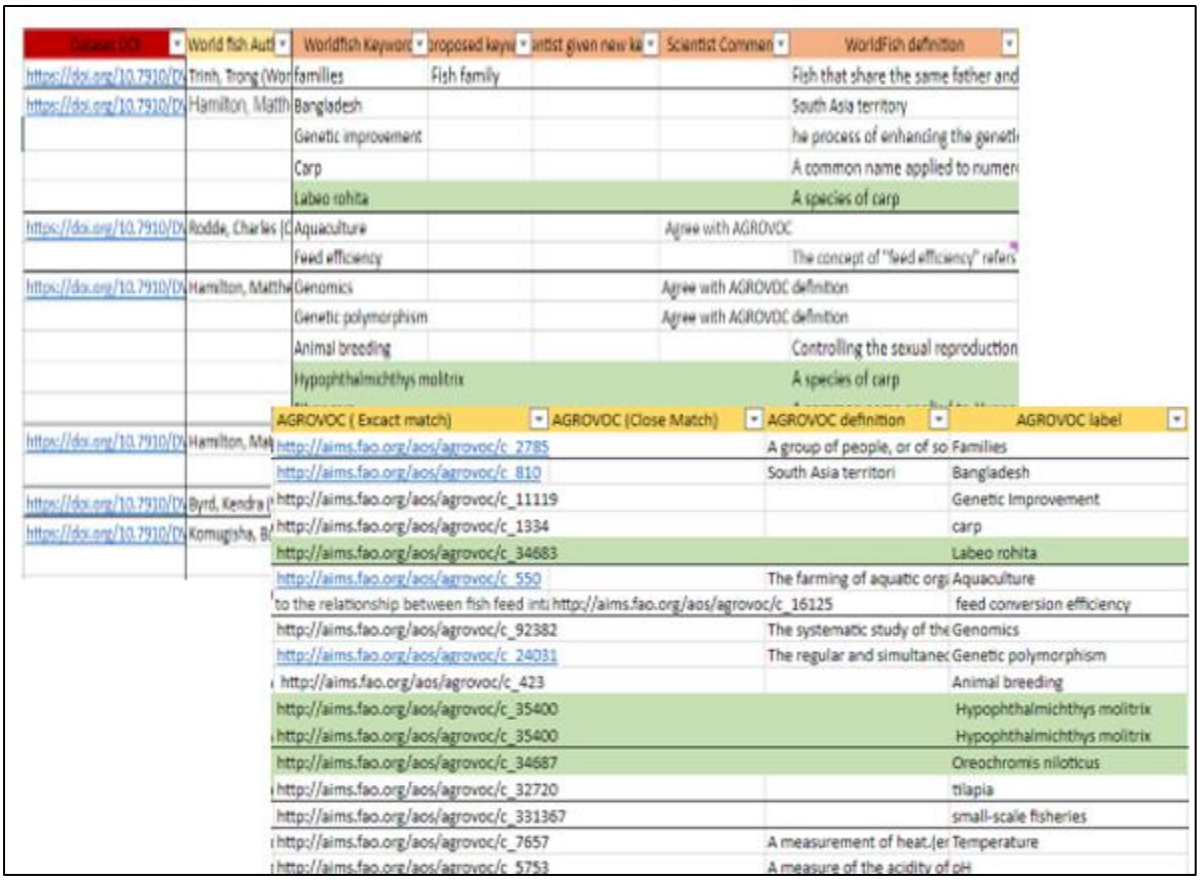


Figure 8:WorldFish Keywords linked with the AGROVOC

Keywords were extracted from the repository to an excel sheet, separating the keywords that had already been linked from the ones that could not be matched to AGROVOC because the terms were missing. A total Keywords of 210 extracted from WorldFish Dataverse [Figure no 9]



Dataset DOI	World fish Aut	Worldfish Keywor	proposed keyw	antist given new ke	Scientist Commen	WorldFish definition
https://doi.org/10.7910/DV	Trinh, Trong (Wor	families	Fish family			Fish that share the same father and
https://doi.org/10.7910/DV	Hamilton, Matth	Bangladesh				South Asia territory
		Genetic improvement				he process of enhancing the geneti
		Carp				A common name applied to numer
		Labeo rohita				A species of carp
https://doi.org/10.7910/DV	Rodde, Charles (C	Aquaculture			Agree with AGROVOC	
		Feed efficiency				The concept of "feed efficiency" refers
https://doi.org/10.7910/DV	Hamilton, Matthe	Genomics			Agree with AGROVOC definition	
		Genetic polymorphism			Agree with AGROVOC definition	
		Animal breeding				Controlling the sexual reproduction
		Hypophthalmichthys molitrix				A species of carp

AGROVOC (Excact match)	AGROVOC (Close Match)	AGROVOC definition	AGROVOC label
http://aims.fao.org/aos/agrovoc/c_2785		A group of people, or of so	Families
http://aims.fao.org/aos/agrovoc/c_810		South Asia territori	Bangladesh
http://aims.fao.org/aos/agrovoc/c_11119			Genetic Improvement
http://aims.fao.org/aos/agrovoc/c_1334			carp
http://aims.fao.org/aos/agrovoc/c_34683			Labeo rohita
http://aims.fao.org/aos/agrovoc/c_550		The farming of aquatic org	Aquaculture
to the relationship between fish feed int	http://aims.fao.org/aos/agrovoc/c_16125		feed conversion efficiency
http://aims.fao.org/aos/agrovoc/c_92382		The systematic study of the	Genomics
http://aims.fao.org/aos/agrovoc/c_24031		The regular and simultanec	Genetic polymorphism
http://aims.fao.org/aos/agrovoc/c_423			Animal breeding
http://aims.fao.org/aos/agrovoc/c_35400			Hypophthalmichthys molitrix
http://aims.fao.org/aos/agrovoc/c_35400			Hypophthalmichthys molitrix
http://aims.fao.org/aos/agrovoc/c_34687			Oreochromis niloticus
http://aims.fao.org/aos/agrovoc/c_32720			tilapia
http://aims.fao.org/aos/agrovoc/c_331367			small-scale fisheries
http://aims.fao.org/aos/agrovoc/c_7657		A measurement of heat.(er	Temperature
http://aims.fao.org/aos/agrovoc/c_5753		A measure of the acidity of pH	

Figure 9:The List of Extracted keywords from the WorldFish Dataverse

Out of all the keywords, 113 are mapped to AGROVOC, with definition and without definition, 94 keywords were not found in AGROVOC, 7 keywords were mapped in ASFA and almost 89 keywords are not found neither in AGROVOC nor in ASFA thesaurus. The following table 1, summarizes the keywords mapping.

	Mapped Keywords	Keywords not found	Missing keywords in both thesauri
AGROVOC	113	94	89
ASFA	7		

Table 1. summarizes the keywords mapping

To support an accurate mapping, each concept must be defined to avoid confusion due to synonymy or different contexts of application. Therefore, we reached out to the WorldFish scientists to get definitions of the terms. The process took about three weeks of communication between the team and scientists and we were able to get definitions of the keywords along with new improved keywords. We selected almost 210 keywords for the mapping to AGROVOC. We checked that all terms and that a maximum of keywords had definition. We then compiled this list for validation by Mrs Kristin Kholsus, the FAO AGROVOC curator. She provided feedback to improve the list. She indicated that in general the proposed mapping was of quality and she agreed with most of it as follows Annex-2. 18 WorldFish concepts were immediately added into the AGROVOC March release. For example, the term 'Fish genetics' was added to AGROVOC with the unique identifier 'c_07a98e20'.

Hereunder are some examples of the reasons for rejection of terms. Terms that are composed by too many concepts must be broken down to the mapping. In our list, there were some compound concepts that could not be mapped to a single concept in AGROVOC, like "Human nutrition and health" that should be mapped to both AGROVOC concepts "Nutrition" and "Health". The rule is to breakdown such a concept into two: "Human nutrition" and" Human Health".

In some cases, the meaning of the proposed concept was not clear, and definition was not added like "Fry trader" while some terms were too specific for AGROVOC like "g18", "g17" that both indicate populations of genetically Improved Farm Tilapia which is developed in 18 generation and 17 generations for experimental purpose.

Other concepts were rejected when it was in fact a commercial product like "ONT" Acronym for Oxford Nanopore Technologies.

Sometimes, we submitted new terms that in fact already exists in AGROVOC like "Water Temperature '

All the notes and comments were important to push up and will help the ontology development. In the next steps, we clarified the keywords and added missing definitions with the scientists.

3.2.2 Extracting and mapping the variables from the dataset/codebook/Questionnaire form/ Research Papers

To make the data interoperable, variables used to collect data must be mapped to an ontology. This enables computers to find and discover data when querying the web or a database. Variable names contained in selected WorldFish data files were extracted for the mapping with ontological concepts. A search in the Ontology Lookup Service enabled to identify the best matching ontological concepts. A large range of concepts could not be found due to the lack of a relevant ontology for fisheries and aquaculture.

WorldFish Dataverse has 101 datasets including 431 files. Every dataset has a raw and verified data, questionnaire, codebook and sometimes have published papers. A dataset should be connected to the codebook that provides the protocol, the survey form with definitions of terms and objectives of the questions or variables, and the data in the questionnaires. We have taken

the variable name from the raw and verified data files. In the data file sometimes variable names are difficult to understand or in short form. To better understand we have used the codebook and questionnaire or sometimes we can get information through published paper.

For example, we extracted variables of the data set entitled " Replication Data : for Influence of seasons, habitat sanctuaries, gears and environmental variables on the catches of hilsa shad (Tenualosa ilisha) in Bangladesh waters " available in the WorldFish Dataverse (harvard.edu). Figure 10 present the codebook, questionaries, Method Documentation which is store in the data files.

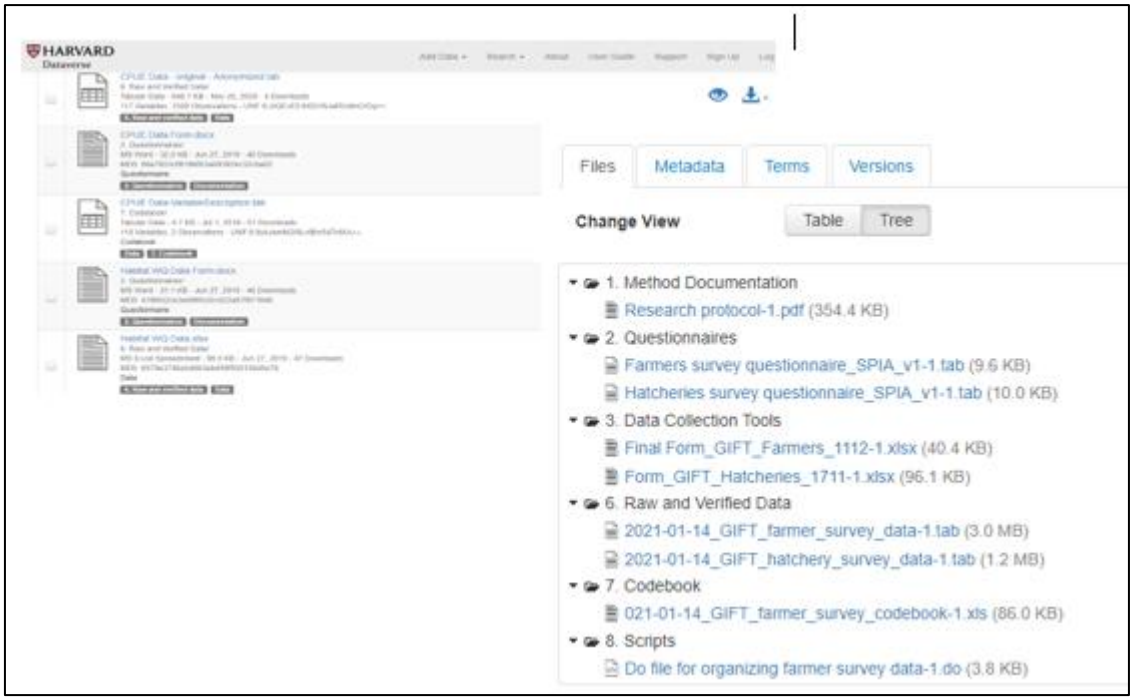


Figure 10:WorldFish dataverse Datasets, codebook, and questionnaires

We also used a data set entitled "Quantitative Data - The Assessment of Species and Performances in Small-scale Shrimp Farming in Aceh, Indonesia "(16)

3.3 Extract concepts -using the Crop Ontology Template as a model

The Crop Ontology is an ontology for Crop Traits developed by CGIAR. CO proposes an Excel Dictionary Template to guide users to create their own trait ontology. We used this template to get a simple standardized framework for development of ontologies. This template is a two-dimension table. The columns represent the properties of the variable, trait, method and scale concepts. The rows represent the instances of variables and their corresponding trait, method and scale. Within such framework, each variable is unique and can only appear once. However, a given trait, method or scale might appear in more than one variable. [Fig no 11](30)

Variable	Trait	Method	Scale
Variable ID	**Trait ID**	**Method ID**	**Scale ID**
Variable name	Trait	Method	Scale name
Variable synonyms	Trait class	Method class	Scale class
Context of use	Trait description	Method description	Decimal places
Growth stage	Trait synonyms	Formula	Lower limit
Variable status	Main trait abbreviation	Method reference	Upper limit
Variable Xref	Alternative trait abbreviations		Scale Xref
Institution	Entity		Category 1
Scientist	Attribute		Category 2
Date	Trait status		...
Language	Trait Xref		Category n

Figure 11: CO template for development of crop breeding platform

To support our work, we have used this template to format the variables of small-scale fisheries and aquaculture ontology.

3.4 Search terms in existing ontologies using OLS

We can find data everywhere and it has multi-format like databases, excel files, tabulated files, publications, reports and word files. For the development of our ontology, we have taken the variables' names directly from the columns' headers of data files and compiled them in the Excel Temple. We selected variables' names that could be for annotating the data files. Then, extracted terms were searched in the Ontology lookup services of used service called EMB-EBI OLS, AgroPortal, Ontobee. They all will provide access to a large range of ontologies.

We searched the species name 'Penaues *monodon* 'in the OLS and found a terms identifier in the NCBI Taxonomy with the identifier 'NCBTaxon-6687 '. We included the URI in our Excel file. If the term is not directly found, we can use the closest term: for example, 'Pond size' could not be found in OLS. In this situation we searched for the most related terms to find its URI. 'Pond' from the environmental Ontology is a close related term which indicates the water body, with the definition 'A body of water, usually of smaller size than a lake'. When we have copied the URI (URI http://purl.obolibrary.org/obo/ENVO_00000033) and placed it in the excel file. In below figure no 12, we can show the list of variables' names, Column of OLS exact match, OLS close match and OLs definition.

Variable ID	Variable name	Definition	No	ith t	AG	OLS controlled vocabullary(EXact mato	OLS colse match	OLS definition
	Farmer Code	Species						
	A1	*Penaues monodon*	*A1 and*	*Yes*		http://purl.obolibrary.org/obo/NCBITaxon_6687		
	A2	*Litopenaeus vannamei*	*A2,A3*	*Yes*		http://purl.obolibrary.org/obo/NCBITaxon_6689		
	PS	Pond Size			http://aims.fao.org/aos/agrovoc/c_6105		http://purl.obolibrary.org/obo/ENVO_0000	A body of water, usually of sm
	SR	Survival Rate				http://purl.obolibrary.org/obo/NCIT_C28192		The percentage of people in a
	TimCyc	Time cycle					http://purl.obolibrary.org/obo/PATO_00001	A quality in which events occu
	TotHarv	Total harvest						
	ShrSize	Shrimp size						
	Revenue	Revenue obtained					http://purl.obolibrary.org/obo/NCIT_C1544	The income that a governmen
	RevShr	Revenue shrimp						
	RevMilk	Revenue milkfish			http://aims.fao.org/aos/agrovoc/c_4834			
	RevTil	Revenue tilapia			http://aims.fao.org/aos/agrovoc/c_32720			
	TotRev	Total revenue						
	PondBottom Prepa	Pond bottom preparation					http://purl.obolibrary.org/obo/ENVO_0000	Pond Bed is a the ground su
	FerUrea	Fertilizer urea				http://purl.obolibrary.org/obo/CHEBI_16199		
	FerSP36	Fertilizer SP36						

Figure 12: The list of Variable names, URI link of OLS and OLS definition

4 Results

4.1 List of missing keywords by using AGROVOC

We are searched our keywords in AGROVOC to get the URI. We have enlisted all the missing keywords in a separate excel file. Our total keywords were 210 and among them 95 keywords or 45% keywords we did not found in AGROVOC. [Fig no 13]

Dataset DOI	world fish author	Worldfish Keyword- Not found	Scientist given new Keywo	Scientist comment	Word fish DEFINITON
https://doi.org/10.7910/DVN/DBZ7IV	Delamare-Deboutteville, Jerome (WorldFish	DNA quantification	Nucleic acid quantitation		In molecular biology, quantitation of nucleic acids is commonly performed to determine the average concentrations of DNA or RNA
		Bacterial genomes	Aquatic Animal's pathogens		Bacterial genomes are generally smaller and less variant in size amo
		Oxford nanopore technologies			Oxford Nanopore Technologies (ONT) Limited is a UK-based company which is developing and selling nanopore sequencing
		Sample metadata	Sample records		Sample metadata or sample records are data about a particular bio
		ONT			Acronym for Oxford Nanopore Technologies
		Sequencing info			
https://doi.org/10.7910/DVN/TZS58I	Trinh, Trong (WorldFish)	GIFT			Genetically Improved Farmed Tilapia, an genetically improved
		g18			GIFT generation 18 developed at Jitra for experimental purpo
		Jitra			The core breeding nucleus of GIFT (terminated on 31 Decemb
https://doi.org/10.7910/DVN/RZCGED	Hamilton, Matthew (WorldFish)	Rohu			A common name applied to *Labeo rohita*
https://doi.org/10.7910/DVN/1HTKPJ	Rodde, Charles (CIRAD, UMR ISEM; ISEM, U	Individual rearing			A rearing design where fish are isolated (one per aquarium) that e
		Feeding rate			The amount of feed given by day to a fish relatively to its body weig
		Fasting tolerance			The ability of fish to cope with feed deprivation, i.e. the ability to e
https://doi.org/10.7910/DVN/YODXEY	Trinh, Trong (WorldFish)	g17 Penang	GIFT G17 designated		GIFT generation 17 developed at Penang, which is the core br
https://doi.org/10.7910/DVN/ID2SXK	Byrd, Kendra (WorldFish)	human nutrition and health	Human nutrition	In Agrovoc as is	
		food and nutrition security	Food security		Physical and economic access, by all people at all times, to the
https://doi.org/10.7910/DVN/CNW526	Byrd, Kendra (WorldFish)	fish consumption patterns	Dietary patterns		A pattern of food consumption, and an overview of the diet a

Figure 13:The list of Missing term in the AGROVOC

4.2 Added new Terms to AGROVOC

Whenever a concept is missing in the AGROVOC thesaurus, a new descriptor must be added, following the rules and Policy term should be included into AGROVOC. The AGROVOC rules are following below:

- Check in dictionaries and other thesauri on the Web if the term exists, and check for the most used term (spelling, etc.). A word may have more than one spelling example for labour - labor; centre – center, etc.
- Check if a s definition might be useful to ensure the correct use of the new descriptor.
- For names of animals and plants, look for taxonomic names and common/local names. Examples: Labeo rohita – Labeo rohita, rohu, rui, or roho, (decide which one will be descriptor and which ones will be non-descriptors). There is no rule for making a scientific name or a common name descriptor or non-descriptor. Both types of terms can be descriptor and non-descriptor based on the frequency one is used. More recently FAO is adopting a policy to make them both descriptors and link them.
- For names of countries, use the official names as defined in the FAO Terminology http://www.fao.org/faoterm/nocs.asp?lang=EN, and/or in the UNBIS (United Nations Bibliographic Information System) thesaurus. Generally, all countries recognized by the UN are already in AGROVOC. As AGROVOC geographical entities follow a specific UN standard, just limit to add the translations for existing geographical entities.
- For entering a new term into AGROVOC, follow the ISO standard Documentation-Guidelines for the establishment and development of multilingual thesauries, if possible, terms should be nouns or noun phrases, but in some multiword terms, phrases may be used. Avoid commas, periods, and hyphens. (31)

Following the rules, we submitted our terms and the AGROVOC curation team who approved 18 keywords among 210 keywords. The terms were included in the March release of AGROVOC. Once, definitions will be provided for remaining 61 keywords, then it would be evaluated and integrated. Terms newly added to AGROVOC are then available into two different languages. The below figure 14 present the list of all approved keywords and next added keywords.

Not found in ASFA and AGROVOC: Worldfish Ke	notes KK Jan 2021 - in general, would need definitions to evaluate fully	action taken Feb 2021	new AGROVOC URI
fish genetics	might be added to AGROVOC	added fish genetics	c_07a98e20
Policy coherence	might be added to AGROVOC, is in FAOTERM with definition	added policy coherence c_095d612d	c_095d612d
Coastal Fishing Communities	fishing communities exists /c_331405, would I be narrower term NT) of that?	added coastal fishing communities c_1478f1b7 but definition source provided in file does not actually mention coastal fishing communities, it is for artisanal fishing. Not all coastal fishing communities are artisinal. Definitions must be precise and must match	c_1478f1b7
Inland fisheries governance	might be added to AGROVOC	added fisheries governance c_1864f588 and inland fisheries governance c_150480be	c_150480be
Bacterial genomes	could look at this	added bacterial genomes c_17c11d03 but good definition would be helpful	c_17c11d03
Human-crocodile conflicts	definition? AGROVOC has human-wildlife relations altterm human-wildlife conflicts, could add this as new NT	added human-crocodile conflicts c_324c7713 with definition :any interaction which results in negative effects on human social, economic or cultural life, or on the conservation of crocodilian species and/or their habitats" from IUCN-CSG	c_324c7713
Health economics	might be added to AGROVOC, is in FAOTERM	added health economics c_4be91522	c_4be91522
Individual rearing	RT individual feeding, could look at this but not only for fish context, under rearing systems	added individual rearing c_4dbc0544 under rearing techniques, mass rearing existed. Definition amended as not only used for fish	c_4dbc0544
Fish value chains	could look at this	fisheries value chains, altterm fish value	c_812ebeeb
Value chain governance	might be added to AGROVOC	added value chain governance c_8a5e09c9 with FAOTERM definition	c_8a5e09c9
Social Resilience	might be added to AGROVOC, is in FAOTERM	added social resilience	c_9d50046e
Seafood safety	might be added to AGROVOC	added seafood safety c_c0ac2de9	c_c0ac2de9
Fasting tolerance	might be added to AGROVOC. could look at this, but not only relevant to fish	added fasting tolerance c_c0fb841c but with amended definition for organisms not only	c_c0fb841c
fish trade	might be added to AGROVOC, look at commercial activities	added fish trade c_d6ead2ca, was in FAOTERM	c_d6ead2ca

Figure 14:The list of newly added keywords and would be added later keywords

4.3 Ontological terms

The concepts that were extracted and compiled for the ontology helped structuring the key elements of each knowledge domain: Aquaculture and Small-scale Fisheries.

Developing an ontology is not only about extracting the concepts but also giving definitions, context of use and semantically link the concepts. The source of each definition should be recorded alongside the definition itself.

Most of the terms we needed to describe fish production or fisheries were missing on the available ontologies. However, we could find some useful concepts in the NCBI taxon ontology for the fish species, in the Environmental Ontology (ENVO) for the environmental conditions, and in the Chemicals Ontology (CHEBI) for chemical components included in intrants. [Figure 15;]

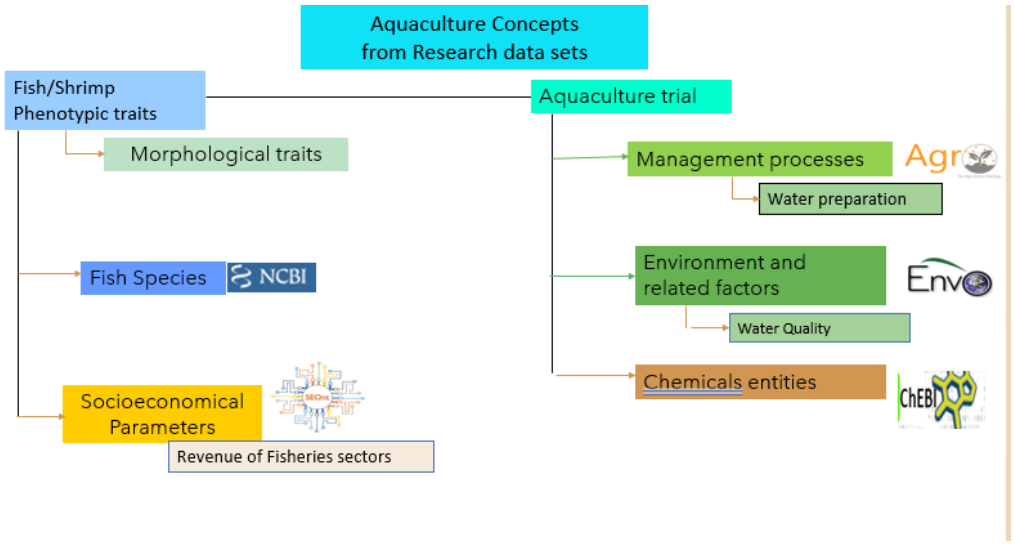


Figure 15:Ontological terms from research dataset

4.4 Knowledge organization model

The Ontology team developed the knowledge organization model is an important step for the development of ontology and must be developed by scientists who are the domain experts. It helps understanding the connections between the different information elements and capture the full coverage of the domain. Knowledge organization models are created to organize information and promote knowledge management (Zeng, 2008). It is accommodating different types of terms like classifications, taxonomies, subject headings, components, elements, environmental measurements etc.

The design of a Knowledge organization model for Aquaculture started with the “aquaculture trial” as the central concept. Aquaculture trials are often studying about how shrimps or fishes are growing (“fish species”, “weight”, “size”) according to environmental measurement process like fish nutrition process and fish harvesting process [Figure 16].

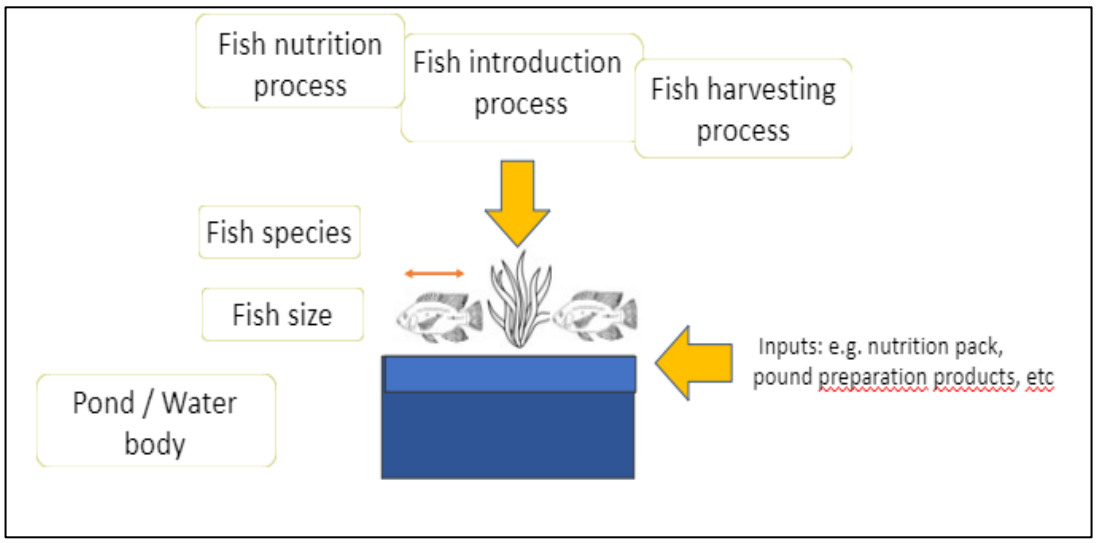


Figure 16:Initial concept of Knowledge organization system

Figure 17 presents the resulting knowledge model in February 2020. In an aquaculture experiment, all processes "occurr_in" the "pond" and are performed by the farmer and the scientist. In such a model, fish "is participant" to all described processes (harvesting, nutrition). Fish "has_ qualities" like fish size, fish lipid content and fish egg number and these qualities will be a monitored all along the experiment's duration. Moreover, there have been occurring fertilization process and could be used different type of fish fertilizer component like Limestone which is coming from Chemical ontology. The water in the pond "has qualities" like "freshwater", "saline water", that can be found in the Environmental Ontology, while "Ph", "Temperature" are concepts available in the PATO. Fish nutrition is developed by providing fish meal which is a AGRO ontology. All elements are measured with precise units that can be found is the Unite ontology. The next step was to extend the Knowledge model by adding the variables extracted from the Worldfish datasets. [Annex 3].

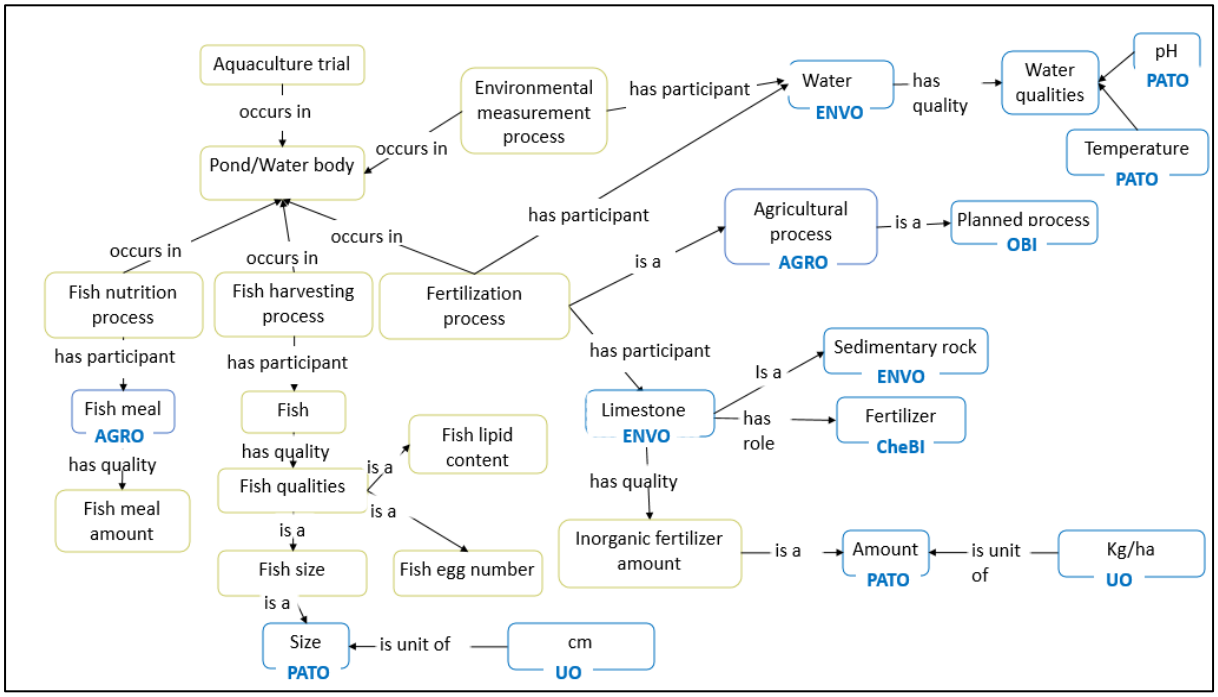


Figure 17:Develop aquacultural trail and linked all domain elements

4.5 First Draft of Small-scale fisheries and Aquaculture Ontology Framework Process

Based on this Knowledge model, Marie-Angélique Laporte, Ontology expert, used Protégé to create, edit and manage the Aquaculture and Small Fisheries Ontology and all its terms and relationship). Protégé is a most used software for creating and maintaining ontology (Musen, 2015) (32). This open access software contains all the tools needed for research since it contains sufficient plugins to assist in development and visualization of ontology. (Ali, et al., 2017).

For development of ontology, it is important to get a comprehensive list of terms without overlap between concepts. The small-scale fisheries and aquaculture ontology is currently structured into three parts. First part is the development of a hierarchy of classes -Small-scale fisheries ontology, second step is the classification of the concepts extracted from data and third is the definition of the concepts' properties using a statement method. Terms describing the types of small-scale fisheries and aquaculture are grouped under 'small-scale fisheries' which is defined as 'a type of information' that is used to support "statement method". Here we have created definitions of the concepts or terms in the hierarchy part and continue by describing properties of these concepts or terms. The development started with the most general concept of the domain followed by subsequent specialization concepts. For example, figure 18 shows the hierarchy of classes starting with the general concept which is "small-scale fisheries". After, the general concept was separated into several subclasses like "agricultural household survey", "gear", "gear qualities", "agricultural process", "catch", "water body", "Fish species". Then, each subclass was categorized adding sub- subclasses like "type of gear", "gear length", "weight and so on.

In the “annotations” section, the RFDS label is indicated along with the ORCID for the creator of the concept, and in the description section, details to those concepts or terms were added, like individual link and terms’ descriptions which is internally link together.

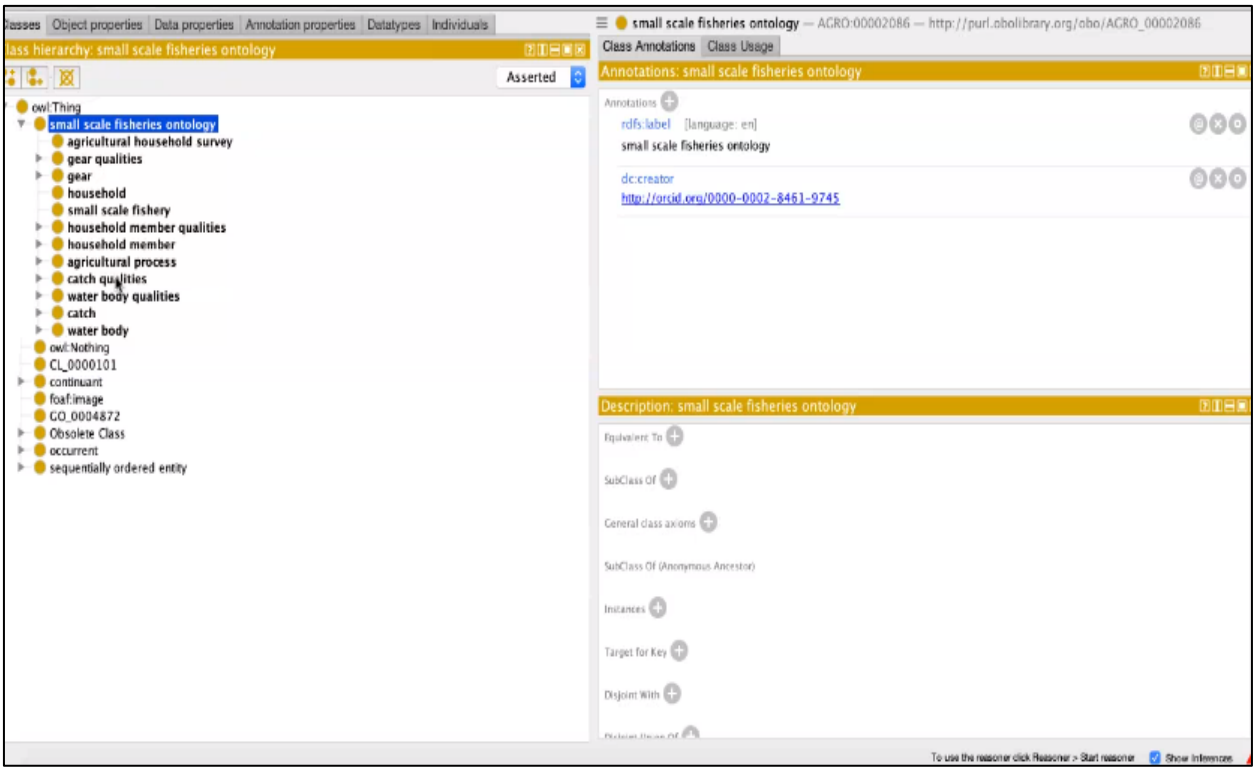


Figure 18: Draft of small-scale fisheries ontology framework process

5.Discussion:

5.1 What were the problems encountered and how the solutions applied helped resolved it?

5.1.1 Long distance Communication and different time Zone

In every work communication is very important. This project is collaborating with WorldFish and Alliance Bioversity-CIAT both organizations are situated into two different continents. So, there have a difference of timing and between two country nearly +8 hour time distance. Moreover because of the COVID-19 pandemic, we have worked remotely. To alleviate this situation, the team used the institutional online collaborative Microsoft SharePoint platform available for the small-scale fisheries and aquaculture ontology working group. We used this platform to upload and share files versions. In this folder we are share our activities, different type of datafiles, documents. My supervisor fixed a weekly meeting day and time to regularly assess the progress made and discussion options and next steps with the team. [Annex-4]

5.1.2 Experiences extracted and mapped Keywords from WorldFish Dataverse

The WorldFish Dataverse has published different types of fish and aquaculture research datasets, files, questions, and technological innovation which is the best practices of open repositories of data. It was difficult to collect and extracted all keywords from the metadata. After extracted all

keywords, I started to manually map all keywords by using different type of controlled vocabulary thesauri like AGROVOC, ASFA and Bioportal. The Alliance Bioversity-CIAT ontology teams and WorldFish have providing guidance and interacting with teams that can facilitate the annotation process for developing solution. Developing or completing ontologies, as well as recommending annotation support tools, all tasks are well-defined Ontologies CoP of the CGIAR Platform. For supporting of the Alliance bioversity-CIAT ontology team and WorldFish we are overcome all difficulties.

5.1.3 Manual search for ontological terms

In general, when annotating datasets, scientists and data managers first need to manually check if relevant ontology terms exist. They also need to be familiar with the terms used in the original files because, for example, fish traits names are often decided by different groups of scientists without any coordination.

The terms are manually searched using the ontology look-up services of the main registries (e.g., European Bioinformatics Institute [EBI] Ontology Lookup Service [OLS] or Bioportal (Ontobee). The set of manually extracted terms will be used to validate the results of a Machine Learning algorithm developed by University of Sheffield, UK, for extracting ontology concepts from survey questionnaires.

5.2 What could not be resolved?

We have tried to resolve all the obstacles which had been create in the project. During Mapped keywords definition are important. There are some keywords does not have the definition, so we try to get the definition from the scientist what they are used in their dataset. Once we will get the definition to added this in our separated excel files. We have submitted our mapped keywords to the FAO AGROVOC and nearly 18 keywords they approved among 210 keywords and its had published March 2021. To add all the keywords AGROVOC need proper definition. We will collaborate with the scientist to get rest of the keyword's definition. In a short period, we get some keywords definition and submitted the files to the AGROVOC core teams. This time we have divided ours files into three tabs. First tab addresses the FAO AGROVOC previous comment where we can view some concepts definitions have need to be updated, some keywords provide new AGROVOC URI and some terms needed to be clear explanation. We were getting some definition but still in our file there are some keywords definitions was pending [annex 5]. In the second tab that was important where we have gathered the updated definition from collaborate with the scientists and last tab no definition represent figure 19. So we will use the FAO AGROVOC network to find the missing concepts or could be able to update the concepts. It will be a continuous process and will be contribute more concepts with the FAO AGROVOC. Besides, we try to build

our first draft of ontology framework for this we need more domain, vocabularies from the datasets.

Dataset DOI	Keyword	Definition	notes KK
https://doi.org/10.791	Fish value chains	A network of stakeholders involved in the growing, processing, and selling food fish that consumers eat—fron	could look at this
	Fish trade	All transactions involving buying and selling of fish, both domestically and internationally, including barter	could look at thiss if we have nothing that matches
https://doi.org/10.791	Performance assessment	A rigorous analysis of aquaculture production at the farm level, particularly focusing on growth, yield and pr	not same as performance testing? c_24061
	Abbassa Strain	In 2002, the Abbassa Strain (AS) of Nile tilapia (Oreochromis niloticus) was initiated by the WorldFish Center	definition?
https://doi.org/10.791	Seafood safety	An integral part of food and nutritional security that involves protecting the fish/seafood supply from microb	might be added to AGROVOC
https://doi.org/10.791	Aquaculture performance as	A rigorous analysis of aquaculture production at the farm level, particularly focusing on growth, yield and pr	not related to performance testing? c_24061
https://doi.org/10.791	Integrated performance asse	A rigorous analysis of aquaculture production at the farm level, extending the analysis beyond growth, yield	not related to performance testing? c_24061
	Genetically improved farmed tilapia	An improved strain of Nile tilapia (Oreochromis Niloticus) created by selective breeding methods.	
https://doi.org/10.791	Fish Traits & Preferences	Qualitative ranking of consumers' preferences for heritable morphometric traits and meristic characteristics	AGROVOC generally does not include compound concepts. Could look at fish traits Do you men feeding preferences?
	Poor Markets	Markets where the capacity to provide care for access to and control over resources and assets, income, heal	definition? As opposed to wealthy?
	Gender-responsive Fish Bree	Tilapia selective breeding programme that incorporates sex-disaggregated preferences and needs of value ch	definition?
https://doi.org/10.791	Human-crocodile conflicts	a situation in which an animal eats, kills, hurts, scares or inflicts economic, physical or psychological losses	AGROVOC has human-wildlife relations altterm human-wildlife conflicts, could add this as new NT
	Melanesia	a region known as the black islands	look at this, Melanesia is a subregion of Oceania
https://doi.org/10.791	Fish Consumption	Fish consumption refers to the consumption of fish and fishery products in a meal.	fish consumption exists. WorldFish definition refers not only to fish but possibly to dietary patterns?
	Dietary Diversity	Dietary diversity is a qualitative measure of food consumption that reflects household access to a variety of f	could look at this if we do not have it under another name
https://doi.org/10.791	Social and gender analysis	An examination to understand the complex nature of social and gender dynamics in a particular context.	
		As defined by the CGIAR Research Program on Aquatic Agricultural Systems (AAS), a gender-transformative approach to development goes beyond the "symptoms" of gender inequality to address "the social norms, attitudes, behaviors, and social systems that underlie them" (AAS 2012, 3). This approach entails engaging groups in critically examining, challenging and questioning gender norms and power relations	

From Kristin | No definition yet | Updated Definitions

Figure 19:The list of updated definition and submit AGROVOC core team

5.3 What are the next steps?

The development of ontology is a long-term process. Initially we have started to extract and mapped the vocabularies. The project team will continue extracting and mapping more keywords, adding more variables' names from the WorldFish datafiles and search more ontological term. The Ontology development and compilation of variables for the mall-scale fisheries will continue with the support of lead scientist and database developers of the project entitled "Automated Analytics System for Small-Scale Fisheries in Timor-Leste (PeskAAS) ". This system using a digital data pipeline, directly collecting data from the fishing boats, and aims to put important data in the hands of fisheries officers, researchers and local stakeholders. This will enable them to better understand the contribution of fish and fisheries to local livelihoods and food security.

The ontology team will engage more with other institutions working on fisheries and aquaculture research and Work on interoperating with ENVO and SDGIO and set up an interest group.

5.4 Ontology will help the open data publishing at Worldfish and will beneficiate other projects on Fisheries and Aquaculture

The Development and practices of ontologies play an important role in life science of fish and aquaculture sector by the development of semantic web**.** Given the diversity in fisheries and aquaculture sector and the associated wealth of information. It is imperative to develop an ontology which can link and integrating all related information. The development of small-scale Fisheries and Aquaculture Ontology will be structured by various variables from the existing datasets of WorldFish which provides fisheries and aquaculture information on metadata. It is

designed by the support of knowledge organization system, semantic annotation of fisheries and aquaculture resources, and different type of survey data integration.

An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them and build a fixable framework. So that researchers and others can get automated fish classification, information about fish, fisheries and aquaculture. They can be able to link all the terms which is used by fisheries and aquaculture domain. The development of ontology practices where researcher will be able to share common understanding of the structure of information between human or machine. The researcher is enabled to reuse, separate and analyze domain knowledge. In this project covers domains knowledge such as fish traits, fish species, management processes, environmental factors, socio-economic parameters, and survey data concepts. By this dataset will be easily share and understand, integrated with the existing resources .so develop of ontology will help the open data publishing at WorldFish and will be benefits other projects on Fisheries and aquaculture.

5.5 What I have achieved from this activity

- I have got experiences to mapped existing controlled or common vocabularies or keywords which is existed in WorldFish dataset. Extracted and compilation all the keywords from the dataset of WorldFish and publication repositories and create a new list of keywords.

- Select the key datasets from the WorldFish databases and tag them using standard vocabularies or Ontologies. Knowing about different type of controlled vocabulary thesaurus like AGROVOC, ASFA and ontological side like OLS, Ontobee and Bioportal etc. Create a list of extracted terms to AGROVOC (FAO) or relevant vocabularies like ASFA vocabulary thesaurus and search the ontological term by using OLS, Ontobee and Bioportal thesaurus.

- Moreover, get experiences about knowledge organization system and the data annotation for the development of first draft of small-scale Fisheries and aquaculture Ontology framework with the support of the Alliance of Bioversity-CIAT and WorldFish.

- In January:

    - I presented during a webinar the project progress to the CGIAR Ontology Working Group composed by 35 scientists and data managers.

    - I attended the meeting of the Livestock Ontology Working Group coordinated by the University of Edinburgh, UK

6 Conclusion

The Development and practices of ontologies play an important role in Fish-related science as it supports the use of new type of databases called graph databases that logically organize the knowledge and with the development of semantic web technologies or Machine Learning**.** Given the diversity in fisheries and aquaculture sector and the associated wealth of information. It is imperative to develop an ontology which can link and integrate all related information. The development of small-scale Fisheries and Aquaculture Ontology will be structured by various variables from the existing datasets of WorldFish which provides metadata on fisheries and aquaculture information. It is designed by the support of knowledge organization system, semantic annotation of fisheries and aquaculture resources, and different type of survey data integration.

An ontology defines a common vocabulary for researchers who need to share information in a domain. It includes machine-interpretable definitions of basic concepts in the domain and relations among them and build a flexible framework. So that researchers and others can get automate classification of information about fish, fisheries and aquaculture. The researcher is enabled to reuse, separate and analyze domain knowledge. thanks to this work, dataset will be easily shared, interpreted, and integrated easily with the existing resources. Therefore, the ontology will help the open data publishing at WorldFish and will be benefits other projects on Fisheries and aquaculture.

The small-scale fisheries and aquaculture ontology is a resource that attempts to summarize the knowledge of domains constantly evolving and then, will always be a work in progress to respond to scientific progress.

The potential usage of the ontology is huge so once an advanced version will be ready and validated by scientists, it will be published in the public domain, with a creative Common Licebnse by CC 4.0 as per the Ontology best practices. The ontology could be used as a framework to build Semantic Web systems for wider data integration in the fish and fishery sector. Such a semantic resource will secure the quality, usability, and sustainability of a comprehensive set of semantic web resources for agri-food science and will support the production of innovations for agricultural development.

Annexes:

The list of Annexes

1 Additional Codebook-G0_silver_carp in Bangladesh from Worldfish

2 Additional list of WorldFish Dataverse with WorldFish definition, AGROVOC definition and note of AGROVOC core team

3 Additional figure of small-scale fisheries trail

4 *Communication website in SharePoint to contract with the team member*

5 Additional list of updated definition to the AGROVOC

Annex-1: Additional Codebook-G0_silver_carp in Bangladesh from Worldfish

Variable Name	Description
EBV	Estimated Breeding Value (phenotypic scale after log10 transformation of raw data)
LINE	Line identifier ('Control', 'Negative' or 'Selection' [i.e. positive selection]. Also identifies fish to be euthanised for the study of 'new traits')
Temperature	The temperature of the water in the pond. It was measured in degrees celcius (◦C). The morning reading was taken at 8:00 am and all readings for the afternoon were taken at 2:00pm .
DO (mg/l)	Dissolved oxygen (DO) in the water available for uptake by the fish measured in mg/l. The morning reading was taken at 8:00 am and the afternoon reading at 2:00 pm.
pH	The pH or level of acidity or alkalinity of the water in the pond. The morning reading was taken at 8:00 am and the afternoon reading at 2:00 pm

Figure 20:Codebook-G0_silver_carp in Bangladesh from Worldfish

Annex-2: Additional list of WorldFish Dataverse with WorldFish definition, AGROVOC definition and note of AGROVOC core team

Dataset DOI	World fish Author	Worldfish Keyword	proposed keyword	to be check with the scientis	Scientist given new keyword	Scientist Comments	WorldFish definition	AGROVOC Controlled Vocabulary(Exact match)	AGROVOC Controlled Vocabulary(Close Match)	AGROVOC definition	AGROVOC label	AGROVOC	note	Notes KK
https://doi.org/10.7	Trinh, Trong (WorldF	families	Fish family	yes			Fish that share the same father and the same mother, i.e., full-	http://aims.fao.org/aos/agrovoc/c_27		A group of people, or of some other animals,	Families	AGROVOC	Fish family is not found, its a	I would use "families" or "siblings". AGROVOC would not include families for
https://doi.org/10.7	Hamilton, Matthew (	Bangladesh					South Asia territory	http://aims.fao.org		South Asia territori	Bangladesh	AGROVOC		Mapping OK
		Genetic		Yes			he process of enhancing the	http://aims.fao.org			Genetic Improvement	AGROVOC		Mapping OK
		Carp					A common name applied to	http://aims.fao.org			carp	AGROVOC	Carp are various	Mapping OK
		Labeo rohita					A species of carp	http://aims.fao.org			Labeo rohita	AGROVOC	Its a scientific name	Mapping OK
https://doi.org/10.7	Rodde, Charles (CIF	Aquaculture				Agree with		http://aims.fao.org		The farming of aquatic	Aquaculture	AGROVOC	its a farming of	Mapping OK
		Feed efficiency		yes			The concept of "feed efficiency"		http://aims.fao.		feed conversion	AGROVOC	In my sense Feed	use feed intake c_16126 ? Plus feed
https://doi.org/10.7	Hamilton, Matthew (	Genomics				Agree with		http://aims.fao.org		The systematic study	Genomics	AGROVOC		Mapping OK
		Genetic		yes		Agree with		http://aims.fao.org		The regular and	Genetic	AGROVOC		Mapping OK
		Animal breeding		yes			Controlling the sexual	http://aims.fao.or			Animal breeding	AGROVOC		Mapping OK
		Hypophthalmichth					A species of carp	http://aims.fao.org			Hypophthalmichthys	AGROVOC	Its a scientific name	Mapping OK
		Silver carp					A common name applied	http://aims.fao.org			Hypophthalmichthys	AGROVOC	its a general name	Mapping OK
https://doi.org/10.7	Hamilton, Matthew (	Oreochromis					A species of tilapia	http://aims.fao.org			Oreochromis niloticus	AGROVOC	Its a scientific name	Mapping OK
		Tilapia					A common name applied to	http://aims.fao.org			tilapia	AGROVOC		Mapping OK
https://doi.org/10.7	Byrd, Kendra (World	small-scale		Yes		In		http://aims.fao.org			small-scale fisheries	AGROVOC		Mapping OK
https://doi.org/10.7	Komugisha, Basilita	Temperature	water				The temperature of the water in the	http://aims.fao.org		A measurement of	Temperature	AGROVOC	Water temperature is	water temperature exists c_16063
		pH	water PH				The pH or level of acidity or alkalinity	http://aims.fao.org		A measure of the	pH	AGROVOC	Water PH is not	Mapping OK. Water pH is not on list of
		Dissolved oxygen	water				Dissolved oxygen (DO) in the water	http://aims.fao.org			Dissolved oxygen	AGROVOC	Water Dissolved	Are you sure these are different?
https://doi.org/10.7	Tilley, Alex (WorldFi	Fish					Fishery resources	http://aims.fao.or		fishery products	Fish	AGROVOC	I think fish is not a	Use fishes. See AGROVOC scope note
		Fisheries					Capture of fishery resources for	http://aims.fao.org			Fisheries	AGROVOC		Mapping OK
		Monitoring	monitoring	Yes			A system to collect and manage	http://aims.fao.or			Monitoring	AGROVOC	related to the fish	Mapping OK. Use also monitoring sytems
		Management	fish	Yes	Fisheries manage		The establishment of rules governing	http://aims.fao.or			Management	AGROVOC		Mapping OK. Use also fishery
		Digital	Digital	yes			Computerized devices, methods and		http://aims.fao.		digital technology	AGROVOC		Mapping OK
		Pipeline	Data analytical	yes			A sequence of actions performed with						agrovoc pipeline	suggest data analysis, data management
https://doi.org/10.7	Baten, MD. Abdul (V	Gender equality				Agree with		http://aims.fao.org		The state in which	Gender equality	AGROVOC		Mapping OK
		Partnerships for		yes		Agree with			http://aims.fao.	Goal 17. Strengthen the	Goal 17 Partnerships	AGROVOC		Mapping OK
		Good health and		yes		Agree with			http://aims.fao.	Goal 3. Ensure healthy	Goal 3 Good health	AGROVOC		Mapping OK
		Responsible		yes		Agree with			http://ajms.f	Goal 12. Ensure	Goal 12 Responsible	AGROVOC		Mapping OK
		No poverty		yes		Agree with			http://aims.fao.	oal 1. End poverty in all	Goal 1 No poverty (en)	AGROVOC		Mapping OK
		Zero hunger		yes		Agree with			http://aims.fao.	Goal 2. End hunger,	Goal 2 Zero	AGROVOC		Mapping OK
		Life on land		yes		Agree with			http://aims.fao.	Goal 15. Protect,	Goal 15 Life on	AGROVOC		Mapping OK
		Reduce inequality		yes		Agree with			http://aims.fao.o	Goal 10. Reduce	Goal 10 Reduced	AGROVOC		Mapping OK
		Life below water		yes		Can			http://aims.fao.	Goal 14. Conserve and	Goal 14 Life below	AGROVOC		Mapping OK
https://doi.org/10.7	Ali, Shimaa (WorldF	Boric acid					Boric Acid is a weakly acidic hydrate	http://aims.fao.org			Boric acid	AGROVOC	its one kind of	Mapping OK

Figure 21: Additional list of WorldFish Dataverse with WorldFish definition, AGROVOC definition and note of AGROVOC core team

Annex -3: Additional figure of small-scale fisheries trail

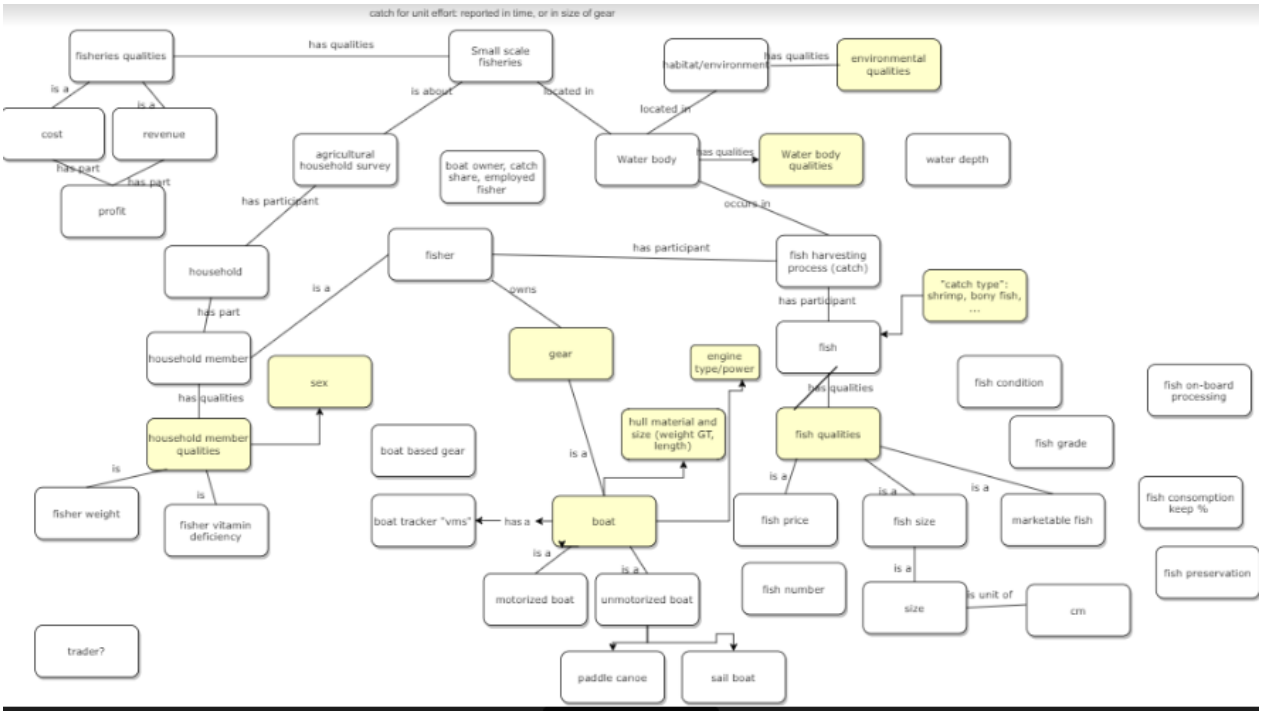


Figure 22: Additional figure of small-scale fisheries trail

Annex-4: **Communication website in SharePoint to contract with the team member**

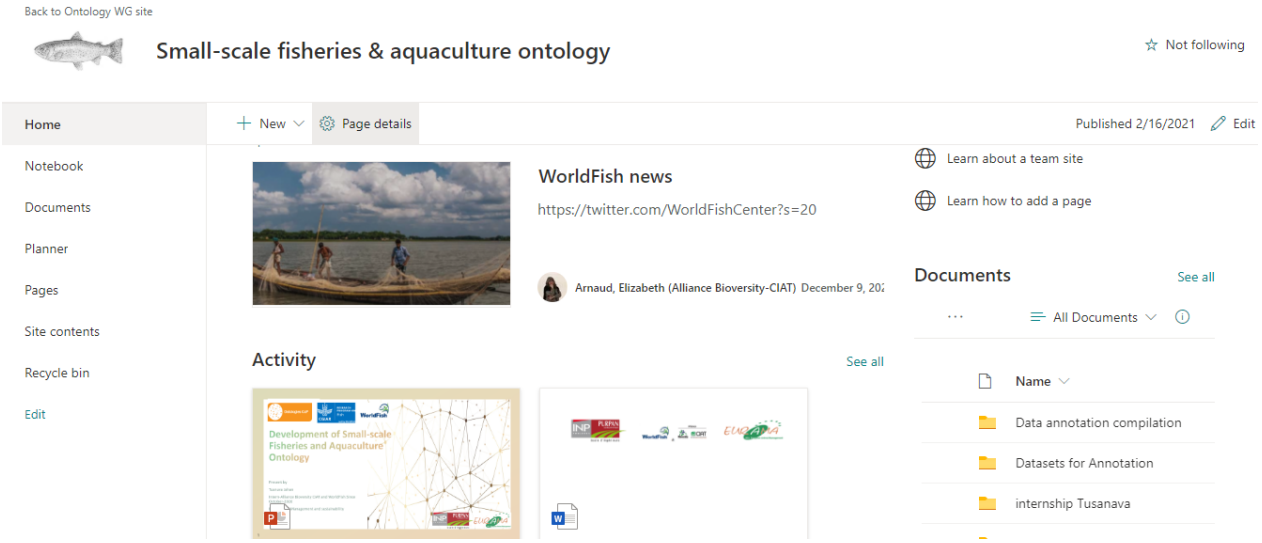
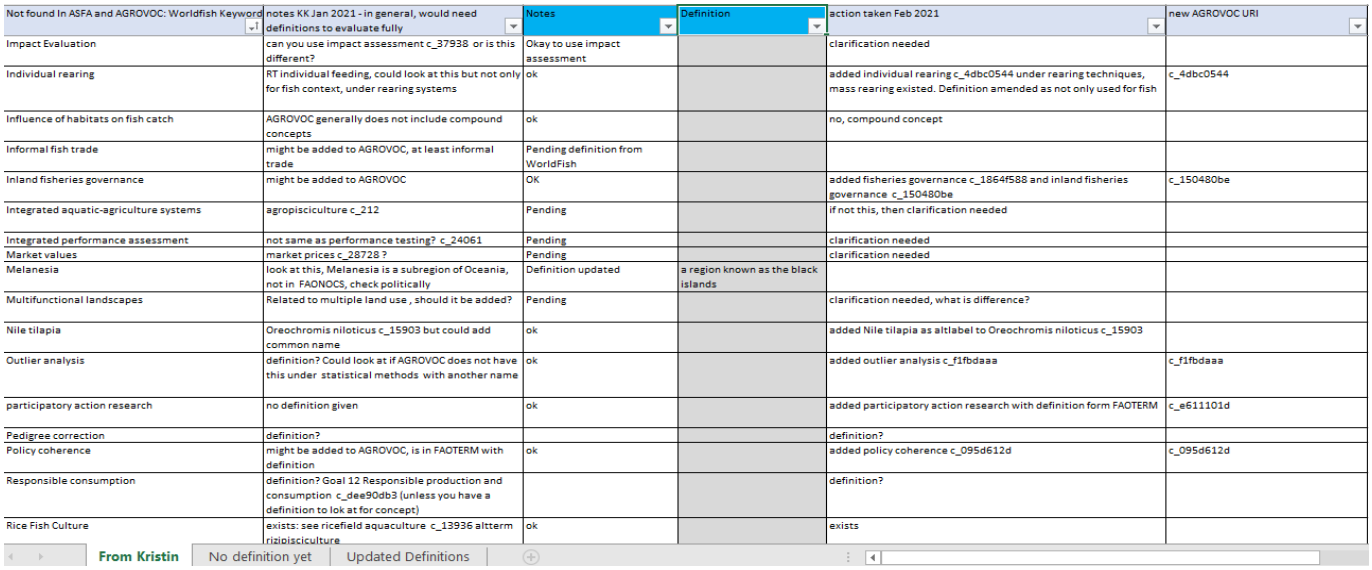


Figure 23: Communication website in SharePoint to contract with the team member

Annex-5: Additional list of updated definition to the AGROVOC



Not found In ASFA and AGROVOC: Worldfish Keyword	notes KK Jan 2021 - in general, would need definitions to evaluate fully	Notes	Definition	action taken Feb 2021	new AGROVOC URI
Impact Evaluation	can you use impact assessment c_37938 or is this different?	Okay to use impact assessment		clarification needed	
Individual rearing	RT individual feeding, could look at this but not only for fish context, under rearing systems	ok		added individual rearing c_4dbc0544 under rearing techniques, mass rearing existed. Definition amended as not only used for fish	c_4dbc0544
Influence of habitats on fish catch	AGROVOC generally does not include compound concepts	ok		no, compound concept	
Informal fish trade	might be added to AGROVOC, at least informal trade	Pending definition from WorldFish			
Inland fisheries governance	might be added to AGROVOC	OK		added fisheries governance c_1864f588 and inland fisheries governance c_150480be	c_150480be
Integrated aquatic-agriculture systems	agropisciculture c_212	Pending		if not this, then clarification needed	
Integrated performance assessment	not same as performance testing? c_24061	Pending		clarification needed	
Market values	market prices c_28728 ?	Pending		clarification needed	
Melanesia	look at this, Melanesia is a subregion of Oceania, not in FAONOCS, check politically	Definition updated	a region known as the black islands		
Multifunctional landscapes	Related to multiple land use , should it be added?	Pending		clarification needed, what is difference?	
Nile tilapia	Oreochromis niloticus c_15903 but could add common name	ok		added Nile tilapia as altlabel to Oreochromis niloticus c_15903	
Outlier analysis	definition? Could look at if AGROVOC does not have this under statistical methods with another name	ok		added outlier analysis c_f1fbdaaa	c_f1fbdaaa
participatory action research	no definition given	ok		added participatory action research with definition form FAOTERM	c_e611101d
Pedigree correction	definition?			definition?	
Policy coherence	might be added to AGROVOC, is in FAOTERM with definition	ok		added policy coherence c_095d612d	c_095d612d
Responsible consumption	definition? Goal 12 Responsible production and consumption c_dee90db3 (unless you have a definition to lok at for concept)			definition?	
Rice Fish Culture	exists: see ricefield aquaculture c_13936 altterm rizipisciculture	ok		exists	

From Kristin | No definition yet | Updated Definitions

Figure 24:Additional list of updated definition to the AGROVOC comment of AGROVOC core team.

Data set's location

1. Nuraini, Retno, 2018, "The Assessment of Species and Performances in Small-scale Shrimp Farming in Aceh, Indonesia", https://doi.org/10.7910/DVN/1732LM, Harvard Dataverse, V1, UNF:6: nhxclsNiaHhYy6ydtr21Ww== [fileUNF]

2. Mohammad Jalilur Rahman; S. M. Nurul Amin; Md. Nahiduzzaman; Md. Abdul Wahab; Nicholas Romano, 2019, "Replication Data for: Influence of seasons, habitat sanctuaries, gears and environmental variables on the catches of hilsa shad (Tenualosa ilisha) in Bangladesh waters", https://doi.org/10.7910/DVN/COVGMO, Harvard Dataverse, V4, UNF:6:D/bBx3mzwQwifccAQqpNNg== [fileUNF]

3. Manyungwa-Pasani, Chikondi L.; Hara, Mafaniso; Chimatiro, Sloans, 2018, "Replication Data for: An investigation of women participation in fish value chains and value chain governance in Malawi: A case of Msaka on Lake Malawi and Kachulu on Lake Chilwa", https://doi.org/10.7910/DVN/3QZUZI, Harvard Dataverse, V1, UNF:6:sNA3FhUPKgiA8iUol+/JJw== [fileUNF]

Bibliography:

1. Gruber T. Liu L. Tamer Özsu M. Ontology.in: Encyclopedia of Database Systems. Springer-Verlag, 2009.
2. Walls R.L, Athreya B, Cooper L, Elser J. Gandolfo M.A, Jaiswal P., Mungall C.J., Preece J, Rensing S, Smith B, et al. Ontologies as integrative tools for plant science, Am. J. Bot., 99 (2012), pp. 1263-1275
3. Arnaud Elizabeth; Laporte Marie Angélique; Kim Soonho; Aubert Céline; Sabina Leonelli;"The Ontologies Community of Practice: A CGIAR Initiative for Big Data in Agrifood Systems" Volume 1, Issue 7, 9 October 2020.
4. Bio-portal **Fish Ontology;** August 28, 2017; Weblink -Available at - https://bioportal.bioontology.org/ontologies/FISHO,
5. Najib M. Ali, Haris A. Khan, Amy Y-Hui Then, Chong Ving Ching, Manas Gaur, Sarinder Kaur Dhillon, et al (2017) "Fish Ontology framework for taxonomy-based fish recognition" [online article] Available at *PeerJ* 5:e3811 https://doi.org/10.7717/peerj.3811
6. Caracciolo C; Heguiabehere J.Gangemi A.Baldassarre C.Keizer J.Taconet M.Knowledge management at FAO: a case study on network of ontologies in fisheries.in: Suárez-Figueroa M. Gómez-Pérez A. Motta E. Gangemi A. Ontology Engineering in a Networked World. Springer, 2012: 383-405
7. The International Food Policy Research Institute "Ontology Communities of practice" GCIAR platform for Big Data in Agriculture, https://bigdata.cgiar.org/communities-of-practice/ontologies/
8. Open definition 2.1, "DEFINING OPEN IN OPEN DATA,OPEN CONTENT AND OPEN KNOWLEDGE" Open Knowledge foundation, Retrieved 22 January 2017 [Online]. Available at https://opendefinition.org/od/2.1/en/
9. Open Access and Data Management Policy (the Policy) FAQs; CGSpace, Available at https://cgspace.cgiar.org/bitstream/handle/10947/2876/Open_Access_and_Data_Management_Policy_(the_Policy)_FAQs_October%2010%202013.pdf?sequence=1)
10. Bill and Melinda Gates, These Open Access and Data Management Implementation Guidelines are effective as of 23 July, 2014; Available at- https://www.cgiar.org/open-data-and-open-access/ and https://www.cgiar.org/how-we-work/accountability/open-access/
11. CGIAR annual Report; Performance Report 2018; Food systems Innovations for Impact in Agriculture.[Online] Available at- https://www.cgiar.org/annual-report/performance-report-2018/ and https://www.cgiar.org/open-data-and-open-access/
12. Wilkinson M.D.;Dumontier M.;Aalbersberg Ij.J.;Appleton G.;Axton M.;Baak A.;Blomberg N.; Boiten J.-W.;da Silva Santos L.B. ; Bourne P.E. The FAIR guiding principles for scientific data management and stewardship. *Scientific Data.* 2016; 3: 160018 ; https://www.go-fair.org/fair-principles/

13. CGIAR Climate change, Agriculture and food security: "OPEN ACCESS AND FAIR PRINCIPLES"[Online] Available at https://ccafs.cgiar.org/open-access-and-fair-principles.

14. Aldo Gangemi, Frehiwot Fisseha, Johannes Keizer, Jos Lehmann, Anita Liang "A Core Ontology of Fishery and its use in the Fishery Ontology Service Project".
15. A. Maedche; S. Staab "Ontology learning for the Semantic Web"IEEE publisher IEEE Intelligent Systems ,Volume: 16, Issue: 2, Mar-Apr 2001.

16. CGIAR Platform for Big Data in Agriculture Annual Report 2018, https://gardian.bigdata.cgiar.org

17. Mariannne McDade; "FAIR is the new Open when it comes to big data" Communications Coordinator CGIAR Platform for Big Data in Agriculture; Cali Colombia, March 7, 2020.

18. S. Leonelli Global data for local science: assessing the scale of data infrastructures in biological and biomedical research BioSocieties, 8 (2013), pp. 449-465
19. GARDIAN; CGIAR for Big Data Platform-Global Agricultural Research Data Innovation Acceleration Network. https://bigdata.cgiar.org/resources/gardian/.

20. WorldFish digitalarchive https://digitalarchive.worldfishcenter.org/

21. Valorie Hollister, Introduction on D-spaces; Knowledgebase ; Dec 02, 2011 [Online] Available at-https://wiki.lyrasis.org/pages/viewpage.action?pageId=25467341#space-menu-link-content
22. "A Guardian Guide to your Metadata". theguardian.com. Guardian News and Media Limited. 12 June 2013. Archived from the original on 6 March 2016.

23. Harverd Dataverse [Online] Weblink https://dataverse.harvard.edu/dataverse/worldfish
24. Harpring, Patricia. Introduction to controlled vocabularies: terminology for art, architecture, and other cultural works. Getty Publications, 2010. https://www.getty.edu/research/publications/electronic_publications/intro_controlled_voca/
25. FAO organization AGROVOC 2020 [Online] Available at http://www.fao.org/agrovoc/about
26. AGROVOC Editorial Community. ASFA; 23.06.2020, [online] Available link at http://www.fao.org/agrovoc/news/agrovoc-editorial-community-asfa

27. OBO Technical Working Group. 2016. The OBO Foundry. [Online] https://www.ebi.ac.uk/ols/index .(accessed 24 May 2016)
28. M.D. Wilkinson, M. Dumontier, Ij.J. Aalbersberg, G. Appleton, M. Axton, A. Baak, N. Blomberg, J.-W. Boiten, L.B. da Silva Santos, P.E. Bourne, et al. The FAIR guiding principles for scientific data management and stewardship Sci. Data, 3 (2016), p. 160018
29. Weblink https://www.force11.org/group/fairgroup/fairprinciples
30. Julian Pietragalla, Integrated Breeding Platform

31. Léo Valette, Rosemary Shrestha, Marie-Angélique Laporte, Tom Hazekamp, Elizabeth Arnaud, Template beta version 5, July 2015, Integrated Breeding Platform and Crop Ontology. Available at crop ontology website

https://sites.google.com/a/cgxchange.org/cropontologycommunity/home/guidelines-for-the-trait-dictionary-template-v-5

32. S. Dagobert, Gauri Salokhe, Gudrun Johannsen, Margherita Sin **"**AGROVOC - The FAO Agricultural Multilingual Thesaurus" Basic Guidelines for Managing AGROVOC. Research gate. January 2008

33. Mark A. Musen "The protégé project: a look back and a look forward" Research Article, Volume 1, Issue 4, AI Matters publication June 2015. Available at https://doi.org/10.1145/2757001.2757003.