Modelling Data For A Sustainable Aquaculture

Ahmed Abid, Charlotte Dupont,Franck Le Gall
Easy Global Market
Sophia Antipolis, France
surname.name@eglobalmark.com

Allan Third
The Open University
KMi - Knowledge Media
Milton Keynes, UK
allan.third@open.ac.uk

Frank Kane
Marine Institute
Galway, Ireland
frank.kane@marine.ie

***Abstract*—Combining aquaculture and Internet of Things (IoT) technologies still poses several challenges. IoT technologies are exploited to enhance productivity in aquaculture sites by maintaining precise operating conditions and to avoid undesirable situations. Currently, human intervention is still needed to make good decisions, dependant on the values of parameters detected by sensors, often captured at one point in time in a day. This is not only a time-consuming task but also an inaccurate one since parameters, such as water quality, evolves continuously and affects the whole aquaculture system. This is mainly caused by the lack of information collection, exchange and process automation between different actors in aquaculture farms. To overcome these problems, technologies such as semantic Web technologies and Artificial Intelligence (AI), are bringing new capabilities to organise data in an inter-operable way to then be processed and used for monitoring and decision-making. In this context, we are proposing in this paper a fully semantic based-reference data model for Integrated Multi-Trophic Aquaculture (IMTA) that involves the collection, processing, and sharing of data both between components and between the platform and external aqua-systems by supporting IoT and data information standards. It is based on analysis of the current data models and covers the core concepts required for aquaculture management and has been validated against several business scenarios.**

***Index Terms*—aquaculture, IoT, Semantics, interoperability.**

I. Introduction

Aquaculture is one of the fastest growing foodproducing sector. It represents the farming of aquatic species like fish, crustaceans, molluscs, aquatic plants and micro-algae. Global fisheries production has been relatively stable since the 1980's. However, the fish demand is still increasing, being twice as fast as the population growth since the 1960's [1]. To meet the demand, aquaculture is exponentially growing all over the world, but China is a clear market leader, producing 1/3 of the world annual fish tonnage. In 2016, the percentage of raised fish (51 percent) has outpaced the percentage of caught fish in the world fish production (49 percent). For aquaculture to contribute to food security and safety in a overpopulated and warming world, it needs to be sustainable and with the lowest impact on environment as possible. IMTA is a form of multi-species production, that provides the byproducts, including waste, from one aquatic species as inputs to another. It allows the different species to benefit from each other's waste as fertilizer or food, reducing the impact of waste on the environment. The equilibrium of this kind of aquaculture production is complex to reach, and a lot of parameters, including water quality, needs to be finely managed.

To address all these new developments in Aquaculture, there is a need to collect more data and exploit them using advanced techniques such as big data , AI , etc. which have led to many successful applications in other sectors. For that data needs to be more structured allowing inter-operability and more analysis. Therefore there is a strong need of data modelling for aquaculture.

For instance, monitoring the water quality in the farms is one of the crucial steps in the commercial aquaculture process. It consists mainly of controlling, physical, chemical and biological parameters, such as: oxygen, temperature and pH in water, on an on-going basis in order to maintain precise conditions and avoid undesirable situations that may lead to issues for the aquaculture environment. Currently, aquaculturists are depending on manual testing due to the lack of integration of smart connected devices in aquaculture farms. This limits the sensors that can be used to monitor and controls the water quality. Smarter devices such as camera or movement detection are not yet exploited in such farms. This pushes aquaculturists to make decisions with a significant lack of information on the behaviour of fishes. Such decisions may be sub-optimal and decreases the productivity and thus affects outputs.

Integrating IoT technologies in aquaculture farms will lead to a new era of connected, responsible and efficient aquaculture. This compels the development of new IoT oriented applications in order to collect and control data from heterogeneous sources, and organise it in an inter-operable way for rapid and autonomous decision-making processes. Technologies, such as semantic Web [2] and AI [3] are proposing new capabilities to promote data inter-operability and to develop around it new smart and autonomous applications that promotes the industry. Integrating such technologies through IoT based solutions in aquaculture will improve the decision-making, production control and management for all fish aquaculture systems, including smart feeding and environmental impacts.

In such a process, IoT data are collected, combined with other important data such as feeding patterns and weather conditions, and finally processed by AI algorithms in order to provide recommendations and management decisions, such as feeding management strategies and optimal harvest times. This results in having a system to optimise feeding of the fish depending on external conditions, but also contributes to the safety, resilience (including climate change) and sustainability of the complete aquaculture value chain.

In this process, collecting and combining data, with organising provided from heterogeneous sources in an inter-operable way is a crucial step. Semantic technologies at this level, offer several tools to annotate and describe connected objects according to their specifications in a machine-readable way. Thus, any collected parameter is semantically and automatically annotated, according to an inter-operable provided reference data model. In this context, we exploit in this paper the existing IoT standardized ontology's and semantic Web technologies in order to develop an inter-operable semantic model that covers the core concepts required for aquaculture management. The presented model relies mainly on two levels. (1) A vocabulary level, in which the concepts of aquaculture are defined as a thesaurus, such as Physical, Chemical and Biological Parameter definitions. (2) An inter-operable platform that supports business requirement defined in relation with data model such as sensors and measurement representations. Based on this model, data provided from heterogeneous sources may be organised in an inter-operable way and thus prepared to be processed automatically by AI based applications that will help aquaculturalists in making effective decisions and intervenes automatically in some specified, precise cases.

The rest of this paper is organized as follows. Section II describes the stakes of sustainable aquaculture that can be achieved with IMTA type of production. Section III gives a brief overview of existing related work in both vocabularies and IoT standardised ontologies. Section IV fixes the requirements of designing a data model for the IMTA. Section V details both the main concept hierarchy of the proposed data model and presents the standardised way for interfacing with it. Finally, section VI discusses of the proposed data model, concludes the paper and gives some interesting perspectives.

II. IMTA, THE FUTURE OF AQUACULTURE

The first evidence of aquaculture can be found 3000 years ago in China [4]. Carp was then the first species to be raised for food. It has been introduced also in Ancient Rome, where the Romans began the first oysters farms. It has been slowly developed since. The revolution of aquaculture only began in the 20^{th} century, with the introduction of granulated food and less expensive material. The stagnation of caught fish production since the 1980's and the increased demand for fish by a growing population has largely contributed to the exponential growth of the aquaculture sector in the past decade. However, aquaculture is sometimes criticized for its impact on environment, due to high density of production and uncontrolled release of production waste and byproducts (food and fertilizers) in the environment.

IMTA is a different type of production; more ecologic and sustainable where multiple species are raised in a same area. Although it is not a new concept - rice and fish co-culture is almost 2000 years old in China [5], it has regained a new interest in western world recently. By using the waste of one specie as byproduct for another, it decreases the overall impact of the production on the environment. IMTA has a long history in Asia, however the approach is not the same [6]. In Asia, seaweeds sites have seen the development of smaller invertebrates and fish (the SIF approach) while in the western world, fish aquaculture have added the invertebrates and the seaweeds (the FIS approach). This means that technologies have still to develop in the western IMTA, which is still in its early stages.

To progress the growth of an IMTA that will provide economic stability, water quality, nutrients efficiency and environmental impact must be well managed. In the IMPAQT project, sensors will be deployed to monitor the water quality, the fish behaviour, and the environment in the IMTA pilot sites. A particular effort will be placed on data modelisation of the incoming data. Water quality data will be supplemented by other sensors data analysis (e.g. cameras, satellite remote sensing) and aquaculturists metadata. To give a better analysis of the ecosystem and production and, later, useful and appropriate advice to the end-users, data modelisation is of particular importance. The hierarchy of the data will influence on the decision mechanisms, so a suitable infrastructure must be set even before the sensors are deployed.

III. SEMANTIC TECHNOLOGIES AND IOT STANDARDS OVERVIEW

The Semantic Web [7] is a set of technologies designed to support the publication of machine-readable data in a similar way to the Webs support for the publication of human-readable pages; in particular, the concepts of decentralisation and of linking are central to both the Semantic and ordinary Webs. The combinations of semantic Web and IoT technologies offer several tools to annotate and describe connected objects according to their specification in a machine-readable way. Below, we present an overview of existing standardised ontologies for IoT that may be referred in the proposed reference model next.

The oneM2M [8] was proposed as a global open standard for Machine to Machine Communications and the IoT, providing a common M2M service layer with common service functions to connect and interwork devices. The oneM2M ontology gives more attention on machine-to-machine than Web-based applications. The SAREF (Smart Appliances REFerence) ontology [9] was proposed by the European Commission in close cooperation with ETSI (European Telecommunications Standards Institute) to provide a modular and domain-independent semantic layer for smart equipment designed to perform a specific task. It is defined as device-centric ontology including sensors and actuators, which focuses on functions and measurements given by devices. The Web Of Thing ontology[1] (WoT) is defined as a thing-centric formal model and a common representation for things on the Web. A Thing is an abstraction of a physical or virtual entity which interacts and participates in the WoT, and a Thing Description describes the metadata and interfaces of Things. The SSN (Semantic

[1] https://www.w3.org/TR/wot-thing-description/

Sensor Network) and SOSA (Sensor, Observation, Sample, and Actuator)[2] are a set of ontologies that describe sensors, actuators, samplers as well as their observations, actuation, and sampling activities. The ISG CIM (Industry Specification Group for cross- cutting Context Information Management), proposed a cross-domain based data-centric ontology that supports the data management by the CIM RESTful APIs named NGSI-LD[3]. The NGSI-LD information model covers several domain application such as Smart City, Smart Agriculture, and Smart Industry. It makes easier the creation of real-world entities, relationships and properties; moreover,the information model is expressive enough to connect and federate other existing information models, using JSON-LD[4].

In the the next section, all requirements for designing a reference data model for IMTA will be fixed according to the description in previous sections.

IV. Vocabularies and Standards Requirements for designing an IMTA Data Model

The provided data model platform needs to represent and share data relating to multiple different kinds of entity, and so needs to use vocabularies relevant to each. In this section, we briefly summarise required topics and relevant existing vocabularies for each, indicating which ones we have chosen to align with. Requirements are categorised in four main parts detailed below.

A. Aquaculture, Fisheries and Marine Environments

The Food and Agriculture Organization (FAO) of the United Nations[5] maintains the AGROVOC vocabulary [10], which contains terms for over 36k concepts (at the time of writing) covering all areas of the FAOs interest, in food, nutrition, agriculture, fisheries, forestry and the environment, and associated concepts such as location, time, activities, measures, and so on. In the same context, the Network of Fisheries Ontologies[6], associated with the FAO, and created during the EU-funded NeOn project, appears no longer to be active, with multiple broken links from its description page. For these purposes, the proposed Data Model may be aligned with AGROVOC wherever possible.

B. Physical, Chemical and Biological Parameters

Physical, chemical and biological parameters are widely covered in terms of data standards, but, being a highly diverse set of concepts, not necessarily in single vocabularies. Individual standards for parameters relevant to aquaculture include WaterML[7]; the Unified Medical Language System[8], which, though focused on human medicine, contains concepts which apply equally well to other species, such as heart rate, stress, prescription, diagnosis, and so on; and the range of vocabularies available on BioPortal[9].

C. Sensors and Measurements

Sensors and measurements are useful in an extremely wide range of fields and there have been a number of approaches to the sharing of sensor readings. In Section III we presented an overview of standardised IoT ontologies. Measurements in the proposed data model will be modelled in the form of the Semantic Sensor Network Ontology (SSNO), which provides the ability to describe communicating sensors and their data in a Semantic Web friendly way. The proposed data model has to ensure also the compatibility with NGSI-LD in order to maintain the flexibility of discovering and querying relevant information in a standardised way.

Units of measurement, where not included in any of the existing mentioned vocabularies, exhibit similar variability to physical, chemical, and biological parameters, in that they appear in many different vocabularies and use cases. Some widely-used standard vocabularies of units include the LOINC[10] ontology for healthcare measurements, and the Units of Measurement Ontology[11].

D. General Purpose Concepts

The core distinctive data features of the IMPAQT project are generally covered by the above vocabularies concepts relating to aquaculture and the marine environment, sensors, observable parameters, and units. There are a number of other concepts which are useful across many domains and for which well-established standards exist, which we mention only very briefly here. Data relating to people is very commonly represented using the Friend of a Friend ontology[12], dates, times and latitudes/longitudes by the Time Ontology and the Basic Geo[13] vocabulary, and data provenance by Prov-O[14], many of which are W3C initiatives or standards. Data provenance is particularly important in highly-regulated contexts, or commercially-sensitive ones, where the source and trustworthiness of data, and how it was measured or calculated, can have significant effects.

V. A proposed Reference Data Model for IMTA

A. Vocabulary of the proposed Data Model

According to requirements fixed in the previous section, Figure 1 presents its main concepts hierarchy.

The presented Reference Data Model has is developed using the Simple Knowledge Organization System (SKOS)[15] ontology. SKOS is designed as framework for modelling relationships between vocabularies and facilitates the relating data models. SKOS is expressed in RDF[16] every valid SKOS

[2]https://www.w3.org/TR/vocab-ssn/
[3]https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp31_NGSI_API.pdf
[4]https://json-ld.org/
[5]http://www.fao.org
[6]http://aims.fao.org/network-fisheries-ontologies
[7]http://www.opengeospatial.org/standards/waterml
[8]https://uts.nlm.nih.gov/uts.html
[9]https://bioportal.bioontology.org
[10]https://loinc.org
[11]https://bioportal.bioontology.org/ontologies/UO
[12]http://xmlns.com/foaf/spec/
[13]https://www.w3.org/2003/01/geo/
[14]https://www.w3.org/TR/prov-o/
[15]https://www.w3.org/TR/2008/WD-skos-reference-20080829/skos.html
[16]https://www.w3.org/RDF/

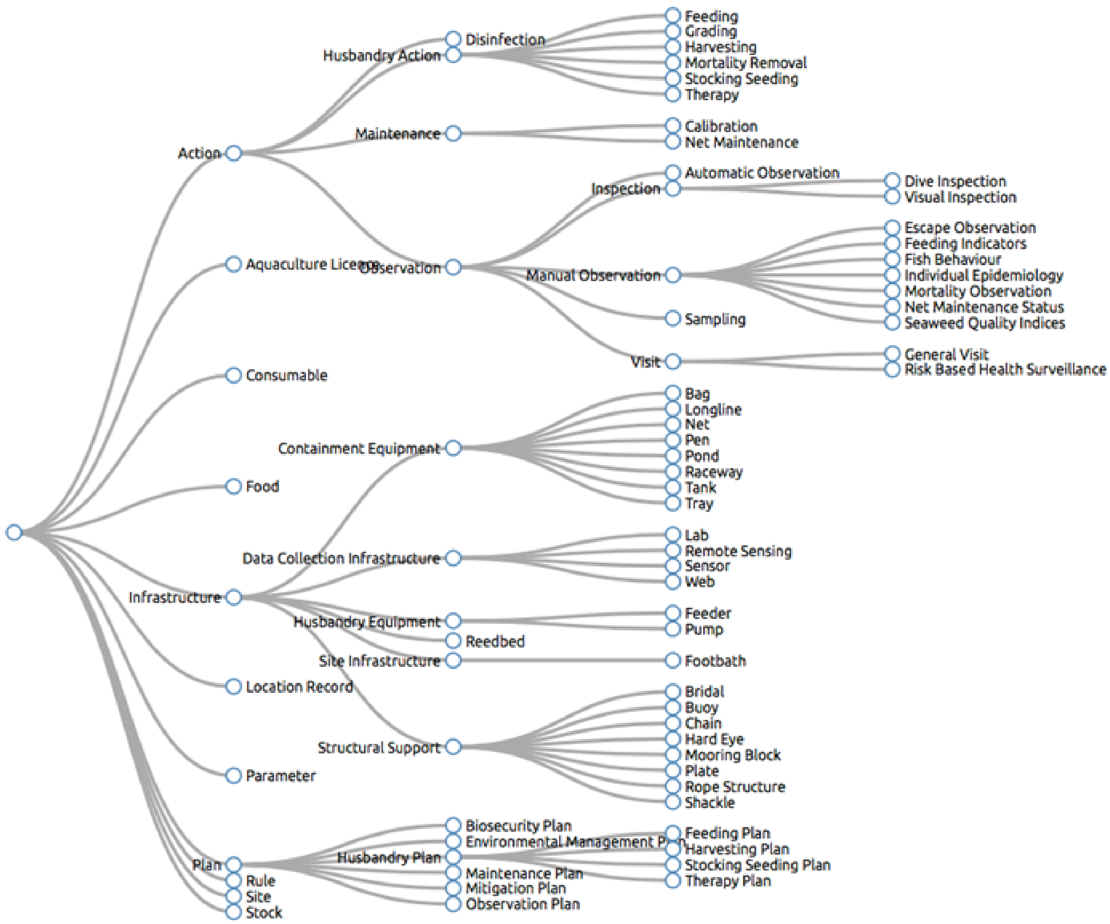

Fig. 1. Concept hierarchy of the proposed Reference Data Model

document is a valid RDF document, and is interoperable with the full Semantic Web technology stack.

The main top-level concepts in this data model are *Site*, *Infrastructure*, *Stock*, *Action*, and *Plan*, representing records of assets and activities in an aquaculture context. Of these, the most complex are *Infrastructure* and *Action*. Other top-level concepts, such as *Location History*, are required in order to model aspects of these main concepts, but are not at the core of the model. All classes in the model support objects being annotated with a free-text comment for human readability. A *Site* entity records the geographical location of a site, and its regulatory status, i.e., the licence covering its activities.

Every piece of Infrastructure has an *identifier* and *number*, a *type* or *brand*, a *size*, *purchase* and *end of use dates*, a *location*, and *records* of its history with regard to *locations*, *inspections*, and *maintenance*. It contains *subclasses* for *Husbandry Equipment* (e.g., a feeder or a pump), Containment Equipment for holding stock (e.g., pen, longline, tank, and so on), *Structural Supports* (rope structure, mooring-block, buoy, and so on), other *non-specific equipment* or *infrastructure*, which may have measurable parameters or maintenance schedules, (e.g. reedbed, sumps, etc.), and *Data Collection Infrastructure* (representing sources of data). Different physical pieces of infrastructure have different relevant sizes, e.g., length, volume, circumference, depth, as well as, e.g., mooring depth.

The *Data Collection Infrastructure* includes the physical infrastructure and serves as as a logical representation of sources of data which may not directly relate to any particular piece physical hardware on site. For example, data sources such as the Web, remote sensing, or external laboratories. Data from each piece of *Data Collection Infrastructure* will be annotated in the system with provenance metadata.

The *Stock* concept is the core concept of the model, and everything in some way is related to the management of a sites stock, which is identified by *species*, *year-class* and a site-specified sub-identifier. *Stock* is recorded with its history of introduction to the site, and previous transport and therapy (e.g., vaccination) records. Particularly important are the interactions, intended generally to capture factors relating to the interaction between different units of stock, in terms of physical proximity and influence on the surrounding environment. Examples might include release or absorption of particular kinds of waste or nutrient, or parasite reduction or prevention. The modelling of interactions is essential for the flexible modelling of multi-trophic systems.

Actions represent records of activities occurring as part of site management. These group loosely into activities that relate to stock, to infrastructure, and to data collection. Every action can have a planned and actual date and time, and a priority, and takes place on a site. *Husbandry Actions* relate to stock, and are carried out in relation to a particular unit of stock. These include stocking or seeding, feeding, administering therapies, harvesting, grading and mortality removal. *Infrastructure actions* relate primarily to *Maintenance*, with records for, e.g., net maintenance and sensor calibration. It is particularly important to IMTA for the system to know the calibration status of data collection infrastructure, in order to be able to interpret readings appropriately at particular times.

The *actions of Observation* are related to parameters which can be measured, and are carried out on a site, unit of stock, or piece of infrastructure. They include *Inspection* (visual or dives) and *Visit* (for, e.g., regulatory or health checks), *Automatic Observation* (from sensors or calculations for example, FCR is a calculated metric), *Manual Observation* (seaweed quality indices, mortalities, escapes, feeding indicators, or fish behaviour), and *Sampling*. *Simply recording actions* as they are carried out does not make for a management platform. *Actions* can be grouped into *Plans*, which describe series of actions, with frequencies, priorities, and rules for situational triggers and contextual applicability.

B. Interfacing with the Proposed Reference data model

In this section we present the main standardised interfacing way supported by the proposed data model. SPARQL[17] query language is the main interfacing method used for querying RDF data. Dealing with a SPARQL End-point, requires knowledge not widely available within software developers and offering a unique such interface would limit the adoption of the IMPAQT system.

The presented data model will provide RESTful interfaces, common in the web development community. Compatibility with NGSI-LD is considered as an important option to benefit from the FIWARE software ecosystem . While most of the FIWARE generic enablers are only NGSI 9/10[18] compliant (not implementing the full semantic logic), announcement of NGSI-LD compatible enablers has been made by major FIWARE players and will exist within the first half of 2019. In the meantime, a wrapper to the FIWARE orion context broker, recognised as a CEF building block[19], is already proposed. In the initial version of NGSI, context is based on an entity-attribute model[20]. Supporting such standards will offer to the IMPAQT system to allow getting information in different ways. For example, a publish-subscribe interface is supported to subscribe to events such as publication of a new data. More complex queries can be run, such as geoqueries. In that case a user can ask to get notified if an event occur within a certain geometry.

[17]https://www.w3.org/TR/rdf-sparql-query/

[18]https://forge.fiware.org/plugins/mediawiki/wiki/fiware/index.php/NGSI-9/NGSI-10_information_model

[19]https://ec.europa.eu/cefdigital/wiki/pages/viewpage.action?pageId=71763006

[20]https://fr.slideshare.net/FI-WARE/fiware-ngsi-managing-context-information-at-large-scale

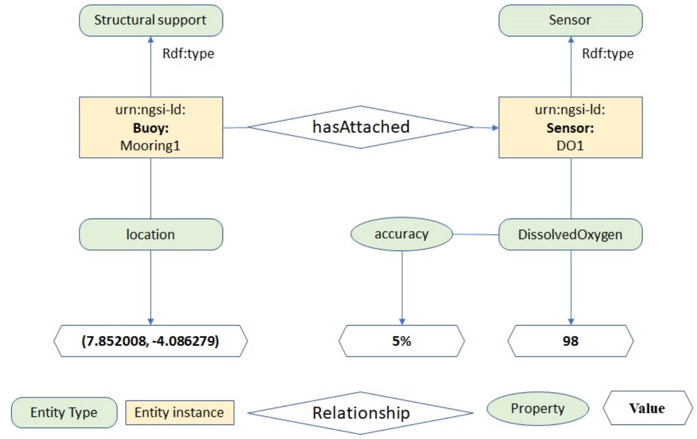

Fig. 2. example of floating sensor modelling

Figure 2, presents an example of modelling floating sensor supporting the NGSI-LD standards. The APIs offers then various way to query data. In this example, it is possible to:

- Query by Id: *GET /entities/urn:ngsi-ld:Buoy:Mooring1*,
- Query by type: *GET /entities?type=Buoy*,
- provide Geoqueries Alerte: */entities?georel=near;maxDistance=2000geometry= pointcoordinates=* $[-2.35, 40.78]$.

VI. Conclusion

In this paper we presented our first results of designing a new reference data model for the IMTA. The presented model was developed from pilot practices and the IMAPQT[21] project requirements which aims to build an appropriate management system for IMTA with advanced monitoring, IoT technology, modelling and data analytics. It follows an IoT based approach as follow. First, the IoT layer collects data from heterogeneous sensors and is responsible for providing a common service layer to interact with these devices. Then, these data are agreed and processed in a smart collection system called Data Acquisition System (DAS). The IoT layer also provides device management capabilities and abstracts the resources.

Among this process, the proposed data model in this paper is defined as a SKOS model that supports RDF based semantic annotation of data streams derived from the IoT layer. Including semantic annotation for collected data will enhance the DAS system for reasoning, analysing and monitoring data in an interoberable way and thus providing adequate decision automatically by following decision strategies related to the data model. Finally, the IMPAQT data model is expected to be interfaced with standardised data information models, which facilitates exposing functionalities (APIs) to be then queried by external actors.

[21]https://impaqtproject.eu/

VII. Acknowledgment

The research leading to these results has been supported by the IMPAQT project, which receives funding from the European Union's Horizon 2020 Research and Innovation Program under grant agreement No. 774109. The content of this article does not reflect the official opinion of the European Union. Responsibility for the information and views expressed in this article lies entirely with the authors.

References

[1] A. O. of the United Nations. Fisheries Department, *The State of World Fisheries and Aquaculture, 2018*. Food & Agriculture Org., 2018, vol. 12.

[2] P. Barnaghi, W. Wang, C. Henson, and K. Taylor, "Semantics for the internet of things: early progress and back to the future," *International Journal on Semantic Web and Information Systems (IJSWIS)*, vol. 8, no. 1, pp. 1–21, 2012.

[3] R. Khan, S. U. Khan, R. Zaheer, and S. Khan, "Future internet: the internet of things architecture, possible applications and key challenges," in *2012 10th international conference on frontiers of information technology*. IEEE, 2012, pp. 257–260.

[4] M. C. Beveridge and D. C. Little, "The history of aquaculture in traditional societies," *Ecological aquaculture. The evolution of the Blue Revolution*, pp. 3–29, 2002.

[5] K. T. MacKay, *Rice-fish culture in China*. IDRC, 1995.

[6] T. Chopin, "Why is integrated multi-trophic aquaculture (imta) developing more easily, and at a larger scale, in china than in the western world?" vol. 21, pp. 18–19, 12 2018.

[7] T. Berners-Lee, J. Hendler, O. Lassila *et al.*, "The semantic web," *Scientific american*, vol. 284, no. 5, pp. 28–37, 2001.

[8] J. Swetina, G. Lu, P. Jacobs, F. Ennesser, and J. Song, "Toward a standardized common m2m service layer platform: Introduction to onem2m," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 20–26, 2014.

[9] L. Daniele, F. den Hartog, and J. Roes, "Created in close interaction with the industry: the smart appliances reference (saref) ontology," in *International Workshop Formal Ontologies Meet Industries*. Springer, 2015, pp. 100–112.

[10] C. Caracciolo, A. Stellato, A. Morshed, G. Johannsen, S. Rajbhandari, Y. Jaques, and J. Keizer, "The agrovoc linked dataset," *Semantic Web*, vol. 4, no. 3, pp. 341–348, 2013.