Preprint of article for 12th International Conference on Information and Communication Technologies in Agriculture, Food & Environment (HAICTA 2026)

proceedings

MDPI

Proceeding Paper

A Digital Seafood Sustainability Assistant: Integrating Public Data to Support Consumer Decisions †

Yannis Marketakis [1,2,*], Alessandro Petrocelli [3,4], Maicol Battistini [3,4], Ezmiron Deniku [3,4], Sara Pittonet Gaiarin [3], and Yannis Tzitzikas [1,2]

[1] Institute of Computer Science, Foundation for Research and Technology-Hellas, Heraklion, Greece; marketak@ics.forth.gr (Y.M.); tzitzik@ics.forth.gr (Y.T.)
[2] Computer Science Department, University of Crete, Heraklion, Greece
[3] Trust-IT services, Pisa, Italy; a.petrocelli@trust-itservices.com (A.P.); m.battistini@trust-itservices.com (M.B.); e.deniku@trust-itservices.com (E.D.); s.pittonet@trust-itservices.com (S.P.)
[4] Communication Platforms and Online Solutions - COMMpla, Pisa, Italy
* Correspondence: marketak@ics.forth.gr (Y.M.)
† Presented at the 12th International Conference on Information and Communication Technologies in Agriculture, Food & Environment (HAICTA|2026), Kos Island, Greece, 30 September – 3 October, 2026

Abstract

Seafood sustainability is a complex and multidimensional issue that cannot be assessed through a single criterion or data source. Environmental impacts, stock and fishery status, nutritional value, health implications, provenance, and socio-economic conditions all contribute to a more complete understanding of whether and how seafood products can support sustainable consumption. However, relevant information is typically distributed across multiple public data sources, each covering different aspects of seafood sustainability, using different terminology and levels of detail. As a result, consumers lack the integrated information needed to make holistic seafood choices. This paper presents a digital seafood sustainability assistant that integrates heterogeneous public data into a knowledge base and communicates the resulting information through consumer-friendly factsheets. Through the assistant, consumers can explore seafood species and products across multiple sustainability dimensions, including environmental, nutritional, and socio-economic aspects. By transforming fragmented public data into structured factsheets, our approach supports transparent communication, and helps consumers make more informed and holistic decisions.

Keywords: seafood sustainability; knowledge base; environmental data; nutrition and health data; consumer information

1. Introduction and Motivation

Seafood sustainability is increasingly important for consumers, producers, retailers, and policy makers seeking to support healthier diets and more responsible food systems. However, assessing whether a seafood product is a sustainable choice is not straightforward. Sustainability depends on multiple dimensions, including environmental impacts, stock and fishery status, fishing or farming practices, nutritional value, health implications, provenance, and socio-economic conditions along the supply chain. These dimensions are often interrelated, and focusing on only one of them may lead to incomplete or even misleading conclusions.

Academic Editor: Firstname Lastname

Published: date

Copyright: © 2026 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license.

A consumer, for example, may wish to buy a fish product that is rich in nutrients and suitable for a healthy diet. At the same time, the same consumer may also want to know whether the stock of that species is overexploited, whether the fish was harvested using a low-impact fishing gear with minimum bycatch risk, whether the production process is associated with habitat damage, and whether labour conditions in the producing country raise concerns such as child labour or weak worker protection. Some of these questions can be partially answered using existing publicly available data sources. For instance, one source may provide nutritional composition data, another may report country-level labour or governance indicators. However, answering them together, in a coherent and comparable way, requires an integrated knowledge base that can connect species, stocks, fisheries, products, countries, indicators, and data sources. Although relevant data are available, they are typically fragmented across multiple public repositories. As a result, the information needed to support holistic seafood choices remains difficult to combine, interpret, and reuse.

A further challenge is the absence of a common and user-friendly terminology for communicating seafood sustainability. Technical terms such as stock exploitation status, fishing gear impact, bycatch risk, nutrient composition, or social risk indicators may be meaningful to experts but difficult for consumers to interpret when presented in isolation. Therefore, there is a need not only to integrate heterogeneous data sources, but also to communicate the resulting information in a straightforward, consistent, and user-friendly manner.

This paper presents the digital seafood sustainability assistant developed in the context of the VeriFish project[1], which integrates heterogeneous public data into a knowledge base and communicates the resulting information through interactive, consumer-friendly factsheets. The contribution is an end-to-end approach for transforming fragmented public data into structured and accessible information that supports more holistic seafood choices. The scope of this paper is limited to the design and feasibility demonstration of the data integration pipeline and assistant; systematic user evaluation is left for future work.

2. System Overview

The proposed system follows an end-to-end pipeline that transforms fragmented public data into consumer-facing sustainability information. As illustrated in Figure 1, the pipeline starts from heterogeneous public data sources covering environmental, fisheries, aquaculture, nutritional, health-related, and socio-economic aspects. These data are collected, harmonized, and semantically integrated into a common knowledge base, where entities such as species, stocks, fisheries, food products, countries, indicators, and data sources are interconnected.

The knowledge base, described in the next section, aligns identifiers, concepts, and terminologies used by the original sources. For example, a fish species can be connected to stocks, fisheries, fishing practices, nutrient values, harvesting or farming countries, and socio-economic indicators.

On top of the knowledge base, the VeriFish Indicator Framework [1] orchestrates the available information into meaningful sustainability dimensions. The framework translates raw and heterogeneous data into quantifiable and comparable measures, such as scores, that make complex information easier to interpret across environmental, nutrition and health, and socio-economic aspects. These measures facilitate the rapid identification

[1] VeriFish: The sustainability indicator framework to communicate responsible aquafood production and consumption patterns (HORIZON-MISS-2023-OCEAN-01, GA number: 101156426) https://verifish.info/

https://doi.org/10.3390/xxxxx

and understanding of key sustainability information, helping consumers compare seafood species and products more easily and make more informed choices.

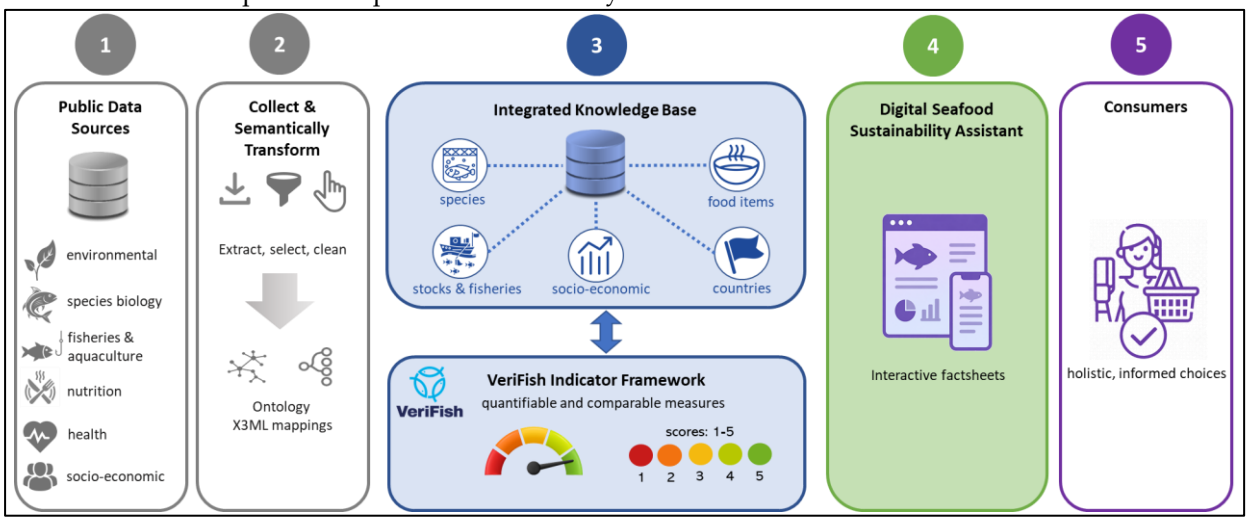


Figure 1. Overview of the pipeline, from heterogeneous public data, to a semantically integrated knowledge base and indicator framework, and the digital seafood assistant.

3. A Knowledge Base for Integrated Data and Indicators

The knowledge base acts as the semantic integration layer of the system, connecting information that is typically stored separately, such as species characteristics, stock and fishery status, food products, nutrient values, harvesting or production countries, and socio-economic conditions. By linking these entities, it enables complex cross-domain questions that would be difficult to address using the original sources independently.

To handle the integration of heterogeneous and thematically diverse data sources, the system relies on Semantic Web [2] and semantic data integration workflows. More specifically, it uses X3ML framework [3], a semantic data integration framework that supports schema mappings, URI [4] generation, and data transformation processes. It uses declarative mappings to describe how different source schemas are transformed into a unified ontology. In our approach X3ML mappings are defined between each data source and the common semantic model, which is an extended version of a top-level ontology for the marine domain, called MarineTLO [5]. These mappings are then executed to transform the source records into knowledge base entities and relations. Mapping correctness was checked through manual inspection of representative transformed records and iterative refinement of the mappings; when overlapping or conflicting values occur, source provenance is preserved so that values can be interpreted in relation to their original dataset.

This mapping-based approach improves the transparency and reproducibility of the integration pipeline. Since the transformation logic is explicitly captured in reusable mappings, the knowledge base can be reconstructed when updated versions of the public data sources become available. This allows the system to incorporate more recent data while preserving a consistent integration process and maintaining traceability between facts, indicators, and their original sources.

In its current version, the knowledge base integrates data from 22 public data sources, covering environmental, ecological, fisheries, aquaculture, nutrition and health, and socio-economic dimensions. These sources include, among others, datasets on species ecology and biology (e.g., FishBase [6], GBIF [7]), stocks and fisheries (e.g., GRSF [8]), nutrient composition (e.g., EuroFIR FoodEXplorer [9]), labour rights and working conditions (e.g., ILO Conventions [10], ITUC Global Rights Index [11]), and other aspects. For each integrated source, the pipeline records provenance metadata, including the source name,

access point, and dataset version (when available), enabling future reconstruction and comparison across knowledge base releases.

Table 1 summarizes the scale and thematic coverage of the knowledge base, while Figure 2 illustrates the semantic harmonization of entities across sources. In the example shown, 'Mullus barbatus' is linked to stock and fishery records, conservation status, biological and taxonomic descriptions, food products and nutrient values, and country-level socio-economic indicators. The transformation enables fragmented source-specific records to support coherent cross-domain querying and interpretation from multiple sustainability perspectives. The integrated entities also provide the basis for indicators, which translate data into comparable scores or categorical values for communication through the assistant.

Table 1. Occurrence statistics for selected entity types in the integrated knowledge base

Entity	Count	Entity	Count
Species	13,582	**Food items**	179
Stocks	5,445	**Nutritional attributes**	77
Fisheries	17,509	**Nutrient values**	9,331
Water areas	7,732	**Aquaculture indicators**	746
Countries	204	**Socio-economic values**	6,285

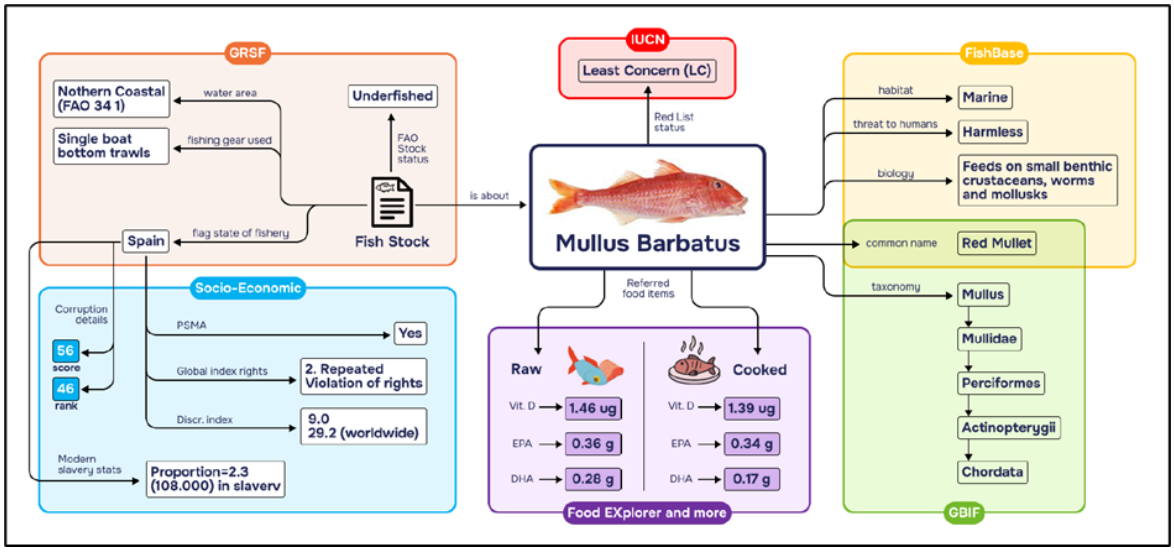


Figure 2. Harmonized and semantically interconnected entities in the knowledge base, showing how information from heterogeneous public data sources is integrated.

Indicator scores are calculated through rule-based normalization. Source-specific values are mapped to a common 1-5 scale, where higher values indicate lower risk or more favourable conditions. Binary variables, ordinal indices, and continuous values are transformed using predefined mappings or threshold bands; when several indicators contribute to a dimension, their normalized scores are averaged with equal weighting, unless otherwise specified by the indicator definition. Missing values are not imputed, but are excluded from aggregate calculations and remain visible to users as unavailable evidence. Country-level socio-economic indicators are used as proxies when fishery-specific labour evidence is unavailable and should be interpreted as contextual risk signals rather than direct assessments of the labour conditions of a specific fishery. More details about the VeriFish Indicator Framework and the scoring formulas are given in [1,12].

4. Digital Seafood Sustainability Assistant

The digital seafood sustainability assistant is implemented as a web application that exposes the contents of the knowledge base to end users through interactive and consumer-friendly factsheets. A web-based implementation was selected to support interoperability across different systems, operating systems, and devices, allowing users to access the assistant without specialized software. In addition, the responsive design enables straightforward access from smartphones and other mobile devices, making the assistant suitable for everyday use and on-the-go consultation. This design also facilitates future integration with external platforms and services.

The application does not query the knowledge base directly. Instead, data are exposed through an access layer comprising dedicated services and a programmatic API[2], which translates application requests into SPARQL [13] queries executed over the knowledge base. This layer hides the complexity of the underlying semantic representation and enables the web application to retrieve structured information about species, ecology, stocks, fisheries, food products, nutritional values and socio-economic resources in a consistent way.

The assistant[3] uses these services to generate factsheets for seafood species and products. Each factsheet brings together information from multiple sustainability dimensions. Rather than presenting raw records from separate databases, the application organizes the information into homogeneous sections, explanatory fields, and comparable indicator values. This allows consumers to quickly identify relevant sustainability information and explore the evidence behind each aspect when needed. For example, Figure 3 presents indicative screenshots from the digital seafood sustainability assistant. The interface provides an overview of species biology and ecology, a summary of nutritional facts, detailed nutrient values for specific seafood products, and socio-economic indicators for countries associated with harvesting or production. These views illustrate how information originating from different domains and data sources is presented through a homogeneous interface, allowing users to move from general factsheets to more detailed evidence when needed.

The assistant follows a modern web application architecture, with a responsive frontend implemented using Kotlin and Compose Multiplatform[4], and a Supabase[5]/PostgreSQL[6] backend for application-specific data. External AI services are used only for auxiliary functions, such as multilingual species-name support and image-related processing. Overall, the architecture follows a Model-View-Controller (MVC) [14] pattern; the knowledge base and services provide the data layer, the web interface presents the factsheets, and the application logic coordinates API interactions. This separation supports maintainability and future extensions.

5. Conclusion

This paper presented a digital seafood sustainability assistant that transforms fragmented and thematically diverse public data into interactive, consumer-friendly factsheets. The work focuses on architectural and data-integration feasibility, demonstrating how heterogeneous sources can be semantically integrated into a knowledge base, organized through indicators, and exposed through a web application. This provides the

[2] https://demos.isl.ics.forth.gr/verifish/verifish-api/

[3] The digital assistant is accessible online: https://verifish.info/verifish-web-app/ and is also available as a smartphone application for iOS and Android devices

[4] https://kotlinlang.org/multiplatform/

[5] https://supabase.com/

[6] https://www.postgresql.org/

technical basis for cross-domain exploration of seafood information. The proposed architecture supports transparency, traceability, and reproducibility, since data can be linked back to their original sources and the integration pipeline can be repeated when updated datasets become available.

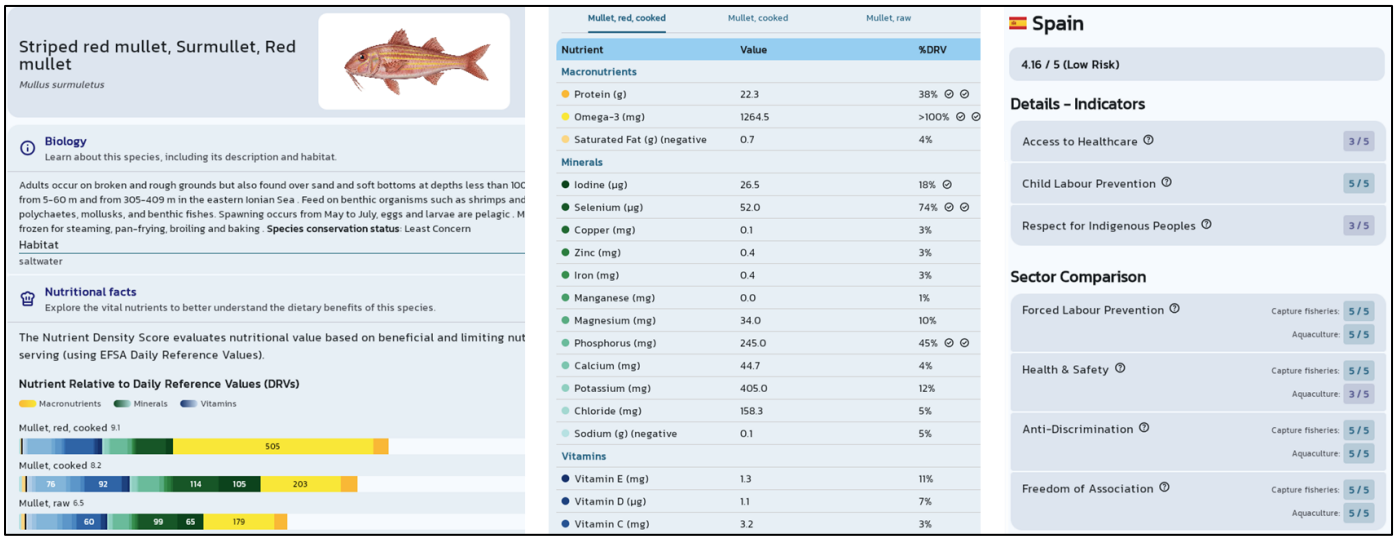


Figure 3. Screenshots from the digital seafood sustainability assistant showing species-level information (left), detailed nutrient values (middle), and country-level socio-economic indicators (right).

Future work will focus on expanding the coverage of data sources and indicators, improving the explanatory content of the factsheets, and conducting a systematic evaluation of the assistant. The evaluation will include usability testing, task-based assessment of whether users can interpret indicators and compare seafood options, and expert validation of the factsheet content, indicator presentation, and scoring logic.

Author Contributions: Conceptualization, Y.M., A.P., S.P., and Y.T.; methodology, Y.M., A.P., and S.P.; software, Y.M., A.P., M.B., and E.D.; validation, S.P. and Y.T.; formal analysis, Y.M.; investigation, A.P., M.B., and E.D.; resources, Y.M. and A.P.; data curation, Y.M., A.P., M.B., and E.D.; writing—original draft preparation, Y.M., A.P., M.B., E.D., S.P., and Y.T.; writing—review and editing, Y.M., A.P., M.B., E.D., S.P., and Y.T.; visualization, A.P., M.B., and E.D.; supervision, S.P. and Y.T.; project administration, S.P.; funding acquisition, S.P. All authors have read and agreed to the published version of the manuscript.

Funding: This work was carried out in the context of the VeriFish project, funded by the European Union under the Horizon Europe programme, Grant Agreement No. 101156426.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data are publicly available through the VeriFish web application.

Acknowledgments: The authors would like to acknowledge the members of the VeriFish consortium for their contributions to the definition of the knowledge base and user-facing communication requirements.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

AI	Artificial Intelligence
API	Application Programming Interface
GBIF	Global Biodiversity Information Facility
GRSF	Global Record of Stocks and Fisheries
ILO	International Labour Organization
ITUC	International Trade Union Confederation
URI	Uniform Resource Identifier

References

1. Astley, S. (2024) VeriFish Indicator Framework defined- D2.1. Zenodo. doi:10.5281/zenodo.14384281.
2. Antoniou, G. and Van Harmelen, F., 2004. A semantic web primer. MIT press.
3. Marketakis, Y., Minadakis, N., Kondylakis, H., Konsolaki, K., Samaritakis, G., Theodoridou, M., Flouris, G. and Doerr, M., 2017. X3ML mapping framework for information integration in cultural heritage and beyond. International Journal on Digital Libraries, 18(4), pp.301-319.
4. Masinter, L., Berners-Lee, T. and Fielding, R.T., 2005. Uniform resource identifier (URI): Generic syntax. Network Working Group
5. Tzitzikas, Y., Allocca, C., Bekiari, C., Marketakis, Y., Fafalios, P., Doerr, M., Minadakis, N., Patkos, T. and Candela, L., 2016. Unifying heterogeneous and distributed information about marine species through the top level ontology MarineTLO. Program, 50(1), pp.16-40.
6. Froese, R. and Pauly, D., 2010. FishBase [online: http://www.fishbase.org/]
7. Yesson, C., Brewer, P.W., Sutton, T., Caithness, N., Pahwa, J.S., Burgess, M., Gray, W.A., White, R.J., Jones, A.C., Bisby, F.A. and Culham, A., 2007. How global is the global biodiversity information facility?. PloS one, 2(11), p.e1124.
8. D4Science. 2020. i-Marine - Global Record of Stocks and Fisheries Catalogue. In: i-Marine - SDG 14.4.1 VRE. Pisa. [Cited 3 November 2021]. https://i-marine.d4science.org/web/grsf/data-catalogue
9. Durazzo, A., Astley, S., Kapsokefalou, M., Costa, H.S., Mantur-Vierendeel, A., Pijls, L., Bucchini, L., Glibetić, M., Presser, K. and Finglas, P., 2022. Food composition data and tools online and their use in research and policy: EuroFIR AISBL contribution in 2022. Nutrients, 14(22), p.4788.
10. Standing, G., 2010. The international labour organization. New Political Economy, 15(2), pp.307-318.
11. International Trade Union Confederation, 2025. 2025 ITUC Global Rights Index: The world's worst countries for workers. Brussels: ITUC. Available at: https://www.ituc-csi.org/global-rights-index [Accessed 14 May 2026].
12. Marketakis, Y. et al. (2026) D2.5 VeriFish Knowledge Base. Zenodo. doi:10.5281/zenodo.19495262.
13. Hogan, A., 2020. SPARQL query language. In The Web of Data (pp. 323-448). Cham: Springer International Publishing.
14. Deacon, J., 2009. Model-view-controller (mvc) architecture. [Online][Cited: 10 March 2006.] http://www.jdl.co.uk/briefings/MVC.pdf, 28, p.61.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.