scientific data

Article in Press

Check for updates

OPEN

DATA DESCRIPTOR

A dataset of the European fishing fleet and maritime infrastructure

Enrique Rodriguez-Martin[1]✉, Jorge Álvarez-Fidalgo[2], Manuel Luna[3] & Jose Emilio Labra-Gayo[2]

We present the European Fishing Fleet Knowledge Graph (EFFKG), a provenance-aware semantic dataset integrating heterogeneous maritime data describing the European fishing fleet and the contextual maritime infrastructure required to represent vessels, ports and administrative relationships. Maritime information in Europe is fragmented across international standards, European registries, and national administrative systems, limiting interoperability and cross-source analysis. The dataset integrates vessel registries, fishing gear classifications, geospatial resources, and administrative datasets from authoritative sources, including detailed national-level data from Spain, within a unified Wikibase-based data model. The current release contains approximately 191,000 entities and 3.7 million statements linking vessels, ports, fishing areas, classifications, and administrative entities. Provenance is preserved at the statement level, enabling traceability of integrated values and transparent representation of source-specific discrepancies. Technical validation confirmed high structural consistency, complete referential integrity, and provenance coverage of 99.99% of all statements. The dataset is publicly available through RDF serialisations, Wikibase JSON exports, and a public SPARQL endpoint, supporting fisheries monitoring, maritime governance, and cross-source maritime data analysis.

Background & Summary

Fisheries and maritime activities are a central component of the European Blue Economy, supporting food production, economic development, maritime governance, and the sustainable management of marine resources. These activities are described by a wide range of datasets produced by international organisations, European institutions, and national authorities, including vessel registries, fleet records, fishing gear classifications, port infrastructures, administrative records, regulatory documents, and geospatial reference datasets. Furthermore, poor fleet data and underestimates of fleet size and effort can affect management and stock assessment, contributing to economic losses and to a misunderstanding of impacts on ecosystems, livelihoods, and employment[1].

The need for semantic interoperability in fisheries information systems has been recognised for more than two decades. Early ontology-based approaches in the fishery domain identified the fragmentation of fisheries data across institutions, countries, databases, and scientific domains, and proposed formal conceptual models to mediate between heterogeneous terminologies and schemas [2]. Subsequent marine semantic-web initiatives, such as MarineTLO, demonstrated the value of top-level ontologies, RDF, OWL, and SPARQL for integrating heterogeneous and distributed marine information resources, particularly in biodiversity, species, and fisheries-related datasets [3].

At the infrastructure level, major marine data initiatives such as EMODnet have improved the availability and standardisation of ocean and maritime data through harmonised standards, controlled vocabularies and interoperable services[4]. Complementary work on global ocean observing systems and marine spatial data infrastructures has further demonstrated that technical interoperability alone is insufficient without semantic and organisational interoperability across data providers and governance structures[5–7].

Recent studies have further demonstrated the scientific value of combining fisheries data from multiple institutional sources for large-scale analyses of fishing fleets and maritime governance. At the global level, quantitative analyses integrating vessel registries, operational information and vessel activity data have been used to characterise the industrial fishing fleet and support the implementation of port State control [8]. At the European level, integrated technical, economic and administrative fleet information has been used to assess structural

[1]University of Oviedo, Dept. of Computer Science, Gijón, 33203, Spain. [2]University of Oviedo, Dept. of Computer Science, Oviedo, 33007, Spain. [3]University of Oviedo, Dept. of Business Management, Gijón, 33203, Spain. ✉e-mail: rodriguezmenrique@uniovi.es

Capability	MarineTLO	EMODnet	Digital Twins	RDBFIS	Wikidata	EFFKG
Domain-specific semantic representation	✓		✓			✓
Authoritative fisheries data integration		✓		✓		✓
Operational fisheries analysis and reporting		✓		✓		
Entity reconciliation across heterogeneous registries						✓
Queryable knowledge graph					✓	✓
Statement-level provenance					✓	✓

Table 1. Positioning of EFFKG with respect to representative marine semantic and data initiatives.

characteristics of the fishing fleet and evaluate policy challenges such as the decarbonisation of the trawl sector[9]. These studies illustrate the increasing demand for integrated fisheries information, while also demonstrating that substantial source-specific extraction, cleaning, harmonisation and reconciliation are required before institutionally diverse datasets can be analysed consistently. The resulting integrated datasets are generally developed for specific analytical objectives rather than being published as reusable semantic resources that support interoperability, provenance and long-term reuse. Consequently, these integrated representations are often difficult to reuse, reproduce or extend beyond their original analytical context. Recent marine data integration initiatives have further highlighted the need for interoperable and machine-readable maritime data infrastructures capable of supporting cross-domain analysis and decision support [10,11]. Although these initiatives have significantly advanced marine data integration and interoperability, they address different and largely complementary aspects of the problem. Operational infrastructures such as RDBFIS support the centralised management, validation, analysis and reporting of fisheries data for the Mediterranean and Black Seas[12], while semantic initiatives address complementary requirements such as semantic modelling, cross-registry entity reconciliation, queryable knowledge graphs and provenance management. These complementary roles also provide a basis for potential future interoperability, for example through shared identifiers and standard RDF/SPARQL interfaces where compatible access and reuse conditions permit. Table 1 summarises these complementary capabilities and the corresponding focus of representative initiatives.

Despite these advances, fisheries and maritime fleet data remain highly fragmented. Vessel registries, national fleet records, port datasets, classification standards, and administrative sources are commonly distributed across independent platforms, rely on heterogeneous schemas and identifiers, and are exposed through different access mechanisms. As a result, combining information across sources—for example, linking vessel characteristics with national administrative status, fishing gear classifications, registration ports, or FAO fishing areas— still requires substantial data cleaning, schema alignment and entity reconciliation before meaningful analyses can be performed[1]. This fragmentation limits data reuse, reduces reproducibility, and constrains the development of integrated applications for maritime monitoring and policy-making at the European level. These challenges are closely aligned with the broader FAIR data principles, which emphasise that scientific datasets should be findable, accessible, interoperable, and reusable[13].

Knowledge graphs and Linked Data technologies provide a suitable framework for combining public datasets with differing schemas through shared identifiers, RDF representations and semantic relationships[14,15]. Maritime knowledge graph approaches have demonstrated the applicability of these technologies to vessel, geospatial and event data[16,17], while broader semantic integration frameworks have shown how heterogeneous data sources can be integrated without requiring common internal schemas[18,19].

A further challenge in multi-source integration is provenance. Institutional datasets frequently provide different descriptions of the same entity because of differences in update frequency, administrative scope or source-specific definitions. Data models with explicit provenance preserve these alternative values, enabling traceability, reproducibility and transparent cross-source comparison. Foundational work on data provenance, named graphs, and Semantic Web provenance established these principles[20–22], which are now central to FAIR knowledge graphs and scientific data infrastructures[23].

Wikibase provides an implementation framework well suited to these requirements. Originally developed as the infrastructure underlying Wikidata, it supports entity-centric modelling, multilingual labels, persistent identifiers, statement-level qualifiers and references, RDF export, and SPARQL querying [24,25]. Wikibase and Wikidata have been used as semantic integration infrastructures in multiple scientific domains, including life sciences, cultural heritage, and FAIR knowledge graph development, demonstrating their utility for integrating heterogeneous datasets while preserving provenance and enabling reuse through Linked Data interfaces [26–28].

Here we present the European Fishing Fleet Knowledge Graph (EFFKG), a semantic dataset with explicit statement-level provenance that addresses the fragmentation of European fisheries information by integrating authoritative records from distinct registries into a unified Wikibase-based knowledge graph through entity reconciliation. The dataset combines international classification standards, European registries, global geospatial resources, and national administrative datasets, providing an entity-centric representation that links vessels, ports, administrative units, fishing areas, classification systems, equipment, and source-specific provenance metadata.

The implemented graph comprises approximately 191,000 entities and 3.7 million statements, integrating data from authoritative sources. It includes broad coverage of the European fishing fleet through the European Fleet Register and is enriched with detailed national-level data from Spain. Spain was selected as the first Member State integrated as a high-resolution national layer because comprehensive administrative data are publicly available and provide substantial coverage of the European fishing fleet. This first version therefore

Dataset / resource	Provider	Source / access point	Main contribution
International Standard Statistical Classification of Fishery Vessels by Vessel Types (ISSCFV)	Food and Agriculture Organization	FAO Open Knowledge Repository	Standardised vessel classification hierarchy and vessel-type categories.
International Standard Statistical Classification of Fishing Gear (ISSCFG)	Food and Agriculture Organization	FAO Open Knowledge Repository	Fishing gear classification taxonomy and gear-category definitions.
FAO Major Fishing Areas	Food and Agriculture Organization	FAO Fisheries Portal	Maritime spatial divisions and fishing-area hierarchies.
World Port Index (WPI)	National Geospatial-Intelligence Agency	NGA World Port Index portal	Port identifiers, geospatial coordinates, and port metadata.
European Fleet Register	Directorate-General for Maritime Affairs and Fisheries (European Commission)	European Fleet Register	Primary vessel registry, administrative identifiers, and technical vessel metadata.
Official Journal of the European Union	Publications Office of the European Union	EUR-Lex portal	Normative definitions used to interpret and validate vessel properties reported in the European Fleet Register.
Liste des ports du système ERS avec données de position	Centre National de Surveillance des Pêches (France)	data.gouv.fr	European ERS port identifiers, port metadata, and georeferenced port-position information.
Spanish General Register of the Fishing Fleet	Ministerio de Agricultura, Pesca y Alimentación (Spain)	Official web consultation service	Additional vessel identifiers, operational metadata, and technical characteristics.
Boletín Oficial del Estado (BOE)	Government of Spain	BOE	Information related to administrative boundaries and territorial metadata.
SIGNA	Instituto Geográfico Nacional (Spain)	IGN Download Centre	Geospatial contextual information.
Situación de la flota pesquera española	Ministerio de Agricultura, Pesca y Alimentación (Spain)	MAPA Portal	Information about number of vessels in Spanish ports per year.

Table 2. Input datasets and supporting source resources used in the construction of EFFKG v1.1.0, including their institutional access points.

represents the first step of an extensible integration strategy intended to incorporate additional Member State data as compatible public resources become available.

The dataset is constructed using an authority-based integration approach, in which supranational sources define shared reference structures while national datasets contribute complementary or alternative values. This design enables multiple source-specific values for the same attribute to coexist as parallel statements, each associated with explicit provenance. Rather than enforcing a single global truth, EFFKG preserves source-level variability and supports transparent comparison of discrepancies across institutional datasets.

By enabling unified querying across previously disconnected resources, EFFKG provides a semantic framework for analyses that would otherwise require substantial manual reconciliation across independently maintained sources. Representative examples include cross-country comparisons of vessel registration, source-aware identification of conflicting vessel characteristics between European and national registries, analyses of fishing gear distributions within FAO fishing areas, and integrated spatial analyses linking vessels, ports and administrative regions. These analytical scenarios illustrate how the combination of harmonised identifiers, explicit semantic relationships and statement-level provenance enables reproducible cross-domain analyses that cannot be readily performed using the original data sources in isolation.

Beyond these representative use cases, EFFKG supports applications in fisheries monitoring, maritime governance, semantic data integration, geospatial analysis, data validation and source-aware auditing. Its extensible design supports the incorporation of additional Member State datasets and maritime domains, positioning the resource as a reusable semantic infrastructure for the European Blue Economy.

Methods

Data Sources. The dataset was constructed by integrating heterogeneous maritime data originating from international standards, European registries, and national institutional sources. All integrated resources were obtained from publicly accessible repositories and correspond to the most recent versions available at the time of access.

EFFKG v1.1.0 was constructed using international classification standards and maritime reference resources, including ISSCFV[29], ISSCFG[30], FAO Major Fishing Areas[31], and the World Port Index[32]. It further integrated European and national administrative resources, including the European Fleet Register[33], the Spanish General Register of the Fishing Fleet[34], the annual *Situación de la Flota Pesquera Española* report[35], geospatial information provided through the Spanish National Geographic Institute's SIGNA platform[36], and the List of ERS System Ports maintained by the Centre National de Surveillance des Pêches[37]. The Official Journal of the European Union[38] was additionally used as a normative source to interpret and validate vessel-property definitions reported in the European Fleet Register.

The principal datasets and supporting source resources used in the construction of EFFKG v1.1.0 are summarised in Table 2, including their providers, access points, and main contribution to the dataset. Complementary reproducibility information, including whether each resource is dynamically updated, the version or snapshot used, and its original licence or reuse conditions, is provided in Table 3. For resources contributing data represented in the knowledge graph, statement-level provenance references identify the originating source through their canonical access URLs. The complete description of each referenced resource, including provider, version or snapshot, retrieval information, licensing conditions, and record-identification mechanisms, is distributed as

Source resource	Dynamic	Version / snapshot	Licence / reuse conditions
International Standard Statistical Classification of Fishery Vessels by Vessel Types (ISSCFV)		Rev. 1 (2019)	FAO copyright; reuse subject to the FAO Terms and Conditions. Non-commercial reuse with attribution is permitted; commercial reuse requires permission.
International Standard Statistical Classification of Fishing Gear (ISSCFG)		2016.1	CC 3.0 IGO
FAO Major Fishing Areas	✓	March 2026	FAO copyright; reuse subject to the FAO Terms and Conditions. Non-commercial reuse with attribution is permitted; commercial reuse requires permission.
World Port Index	✓	March 2026	Public domain (Title 17 U.S.C.)
European Fleet Register	✓	December 2025	CC BY 4.0
Official Journal of the European Union (EUR-Lex)	✓	March 2026	CC BY 4.0
Liste des ports du système ERS avec données de position	✓	March 2026	Etalab Open Licence 2.0
Spanish General Register of the Fishing Fleet	✓	March 2026	Law 37/2007, as amended by Royal Decree-Law 24/2021
Boletín Oficial del Estado (BOE)		BOE-A-2007-10951	Law 37/2007, as amended by Royal Decree-Law 24/2021
SIGNA	✓	February 2026	CC BY 4.0
Situación de la flota pesquera española		2025 edition	Law 37/2007, as amended by Royal Decree-Law 24/2021

Table 3. Access, reproducibility and reuse information for the source resources listed in Table 2. Resources marked with ✓ correspond to continuously updated online services. For these resources, the reported snapshots identify the versions integrated into the current EFFKG release.

part of the machine-readable documentation package accompanying the released dataset. The resulting EFFKG dataset is released under the Creative Commons Attribution 4.0 International (CC BY 4.0) licence.

The European Fleet Register was used as the primary reference dataset for vessel entities, ensuring broad coverage of the European fishing fleet. National-level sources, particularly from Spain in this version, were incorporated to enrich this representation with additional identifiers, operational metadata, and contextual information not available at the supranational level.

The original data sources are distributed in heterogeneous formats, including CSV, XLSX, XML, geospatial web services, and controlled classification standards.

The source inputs used to construct EFFKG v1.1.0 were retained in the public source-code repository to support reproduction of the released dataset. These include source files obtained directly from the institutional providers and tabular representations prepared from resources originally distributed through other formats or consultation services. Together with the released processing code, these archived inputs provide the fixed input set required to reproduce the EFFKG v1.1.0 integration pipeline. The institutional access points reported in Table 2 provide access to the corresponding source resources, either through direct file downloads or through their public consultation and geospatial services. These access points allow the same acquisition and processing workflow to be applied to subsequently available source data when generating updated versions of the dataset. For dynamically updated resources, Table 3 identifies the snapshot used for EFFKG v1.1.0.

The released dataset does not contain personal information relating to vessel owners, operators or crew. EFFKG integrates only publicly available institutional data describing vessels, maritime infrastructure, administrative entities and associated classifications. Personal identifiers and other personal data are not collected, modelled or redistributed as part of the released knowledge graph.

Scope and inclusion criteria. The dataset covers the European fishing fleet and selected components of its associated maritime infrastructure. Vessel entities were derived from the European Fleet Register, which was used as the primary source to ensure broad coverage of active fishing vessels at the European level.

Additional data were incorporated from national administrative sources to enrich vessel descriptions with attributes not available in supranational datasets. In this version, Spain constituted the first high-resolution national layer, providing extended information on vessel identifiers, administrative status, and operational context. These data were integrated as complementary semantic layers aligned with the European reference dataset.

Maritime infrastructure was included to provide the contextual entities required to represent vessel registration and related administrative relationships. EFFKG does not reproduce the underlying port gazetteers exhaustively. Port entities were included when they were required to represent a registration place reported for an integrated vessel or when they belonged to the pre-existing domain-specific administrative layer. External port resources, principally the World Port Index and the CNSP ERS port list, were used as reference sources for resolving and enriching these entities.

Entities were included based on their presence in at least one authoritative source and their ability to be mapped to the unified semantic model. When multiple sources described the same real-world entity, records were reconciled and integrated as described in the harmonisation process.

Data Model. The dataset was represented as a knowledge graph using an entity-property model that captured vessels, infrastructure, classifications, and administrative structures within a common data model. This version comprises approximately 191,000 entities and 3.7 million statements.

The model was centred on vessel entities, which constituted the primary unit of integration. Vessels were linked to technical characteristics (e.g. length, engine power), administrative attributes (e.g. registration status, identifiers), classification categories (e.g. vessel types and fishing gear), and operational context (e.g. associated ports and fishing areas).

Classification systems, including ISSCFV and ISSCFG, were represented as hierarchical structures and linked to vessel entities through standardised relationships. This enabled consistent categorisation across heterogeneous data sources.

Maritime infrastructure was represented through port entities, which were associated with geospatial identifiers and administrative jurisdictions. Ports were linked to vessels through registration-place relationships derived from the integrated source registries and supported multiple labels and name statements to account for linguistic and regional variations.

Administrative units were modelled hierarchically, covering both terrestrial divisions (e.g. regions) and maritime jurisdictions (e.g. FAO fishing areas). These entities provided the spatial and regulatory context for vessels and infrastructure.

The implemented Wikibase schema currently defines 60 properties and multiple domain-specific entity categories derived from the source standards.

A key feature of the dataset was the use of statement-based representations with explicit provenance. Each attribute value was associated with the authoritative source resource from which it was obtained, allowing multiple values for the same property to coexist when discrepancies arise. Statement-level provenance was represented using Wikibase references, where the source property records the canonical access URL of the corresponding source resource. These references intentionally provided a compact identifier of the originating resource rather than repeating release-level metadata. The complete provenance description for every referenced resource,including provider, version or snapshot, retrieval information, licensing conditions, and record-identification mechanisms,is distributed separately as part of the accompanying machine-readable documentation package. Together with the persistent identifiers preserved in EFFKG entities (e.g. Community Fleet Register identifiers, registration numbers, port codes and controlled classification codes), this information enables integrated values to be traced back to their corresponding source records whenever the external resource continues to provide record-level access. This design preserves lightweight statement-level provenance while ensuring reproducibility of the released dataset.

Figure 1 illustrates the structure of the data model and its main entity types and relationships.

Harmonisation and entity reconciliation. The integration process followed a hierarchical harmonisation strategy based on an authority model. Datasets were organised according to their institutional level, with supranational sources (e.g. European registries and international standards) defining canonical properties and structures, and national datasets providing complementary or higher-resolution values.

Attributes describing the same conceptual aspect were aligned across datasets through schema mapping. Equivalent attributes across sources were identified by comparing field definitions, units, identifiers, and controlled vocabularies. When equivalent attributes were identified, the values were retained with explicit provenance.

Entity reconciliation was prioritised for records that constituted the principal units of integration. Vessel records were reconciled deterministically through exact matching of valid Community Fleet Register (CFR) identifiers. Records sharing the same CFR were represented as a single Wikibase entity, while source-specific values were preserved as separately referenced statements. National vessel records lacking a valid CFR were retained as source-specific vessel entities identified by their national registration code.

Ports were used primarily as contextual entities for vessel registration places. During vessel ingestion, each registration-place value was first resolved against existing port entities. When no existing entity was found, the value was compared with normalised port names indexed from the WPI and CNSP reference resources. A new port entity was created only when a corresponding reference record was available and the port was required for a vessel–port relationship. This process was a selective reference-resolution step rather than an exhaustive reconciliation or import of the underlying port gazetteers.

The vessel-port relationship represented in EFFKG corresponds specifically to the registration place reported by the source vessel registry; it does not represent a home port, operating port, fishing port, landing port, or observed area of activity. The relationship reflects the value available in the corresponding source snapshot and should not be interpreted as a complete temporal history of vessel registration. Where different source registries provided differing registration-place values for the same reconciled vessel, these values were retained as separate statements with explicit source references rather than being collapsed into a single preferred value.

Data inconsistencies were addressed through mapping and harmonisation procedures, including:

- Mapping of coded values to controlled entities and classification systems.
- Representation of multilingual textual fields using multiple labels and statements.
- Preservation of alternative or conflicting values as parallel statements with source-specific provenance.

This approach prioritised identifier-based reconciliation and authoritative data sources while preserving source-level variability and traceability across integrated datasets.

Data processing pipeline. The dataset was constructed through a multi-stage processing workflow combining source-specific preparation, tabular consolidation, Wikibase import, and post-import validation.

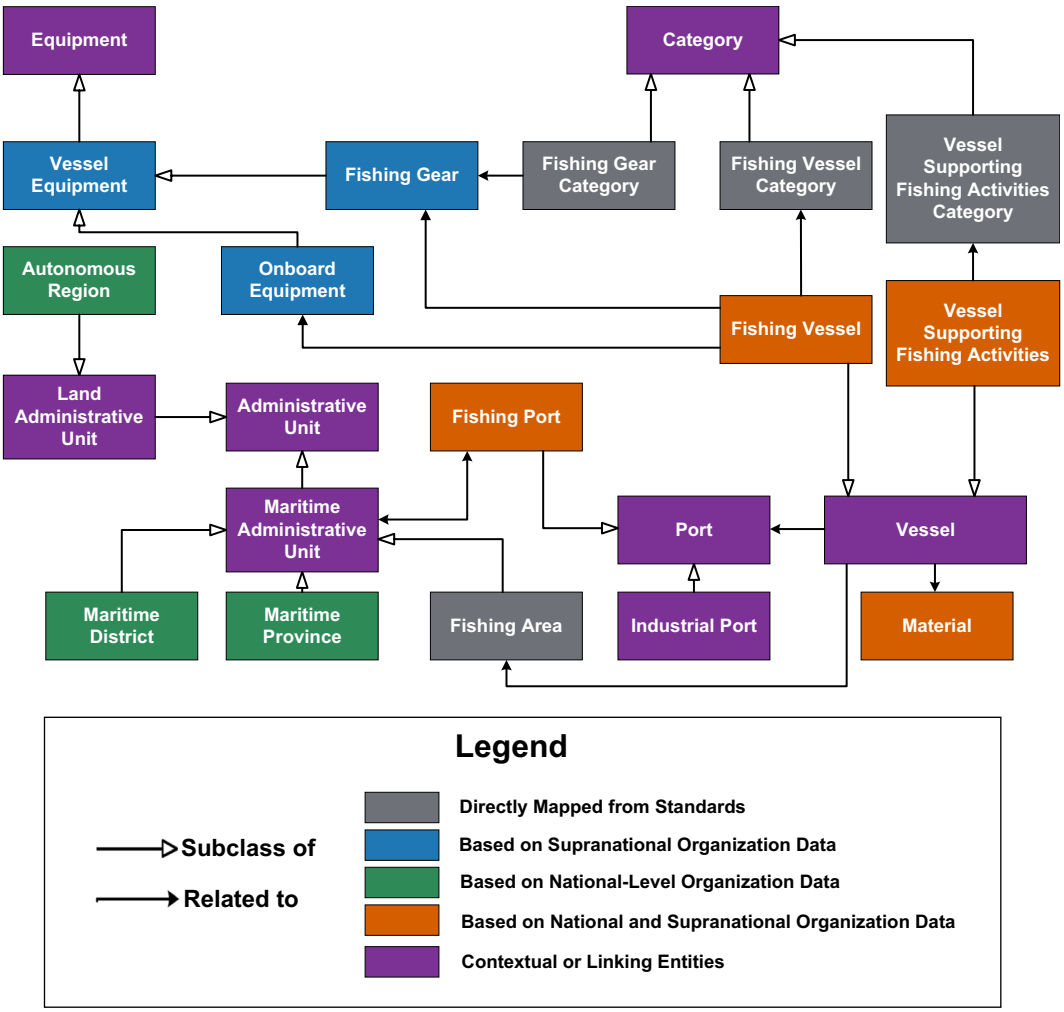


Fig. 1 Overview of the EFFKG data model showing the main entity types and their relationships.

1. **Source acquisition and preparation**. Source datasets were retrieved from official repositories and services. The European Fleet Register was split by country, national vessel records were prepared for integration, and port datasets were filtered to the target geographic scope.
2. **Preprocessing and normalisation**. Input records were cleaned and standardised, including identifier normalisation, text cleaning, coordinate formatting, date handling, and mapping of coded values to controlled categories.
3. **Tabular consolidation**. Intermediate CSV files were generated for vessels, port reference resources, port statistics, and supporting entities. European and national vessel records sharing a valid CFR identifier were consolidated through exact CFR matching. Port-source records were cleaned, geographically filtered, and indexed to support registration-place resolution during vessel ingestion.
4. **Schema mapping and entity reconciliation**. Fields in the prepared records were mapped to Wikibase properties and controlled entities. Contextual entities required by these records, including ports, were resolved against existing graph entities and external reference indexes before new entities were created when necessary.
5. **Knowledge graph import**. Prepared and consolidated records were converted into Wikibase entities, statements, qualifiers, and references. Source attribution was attached at statement level during import.
6. **Post-import validation and release**. The resulting graph was validated through automated SPARQL queries and schema-inference procedures. Validation outputs, inferred schema artefacts, RDF serialisations, and Wikibase JSON exports were generated for the released dataset.

The complete processing workflow was implemented through reproducible Python pipelines distributed both through the public source-code repository and as a versioned code archive accompanying the Zenodo release. The public source-code repository additionally contains the fixed source inputs used for EFFKG v1.1.0, allowing the processing workflow to be executed against the same input data used to construct the published dataset. Figure 2 summarizes the workflow.

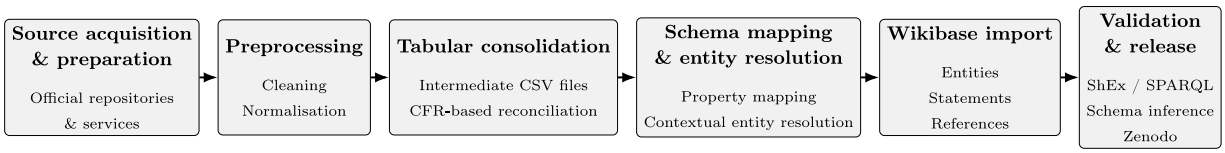


Fig. 2 Overview of the EFFKG data processing workflow. Data were retrieved from authoritative repositories and services, preprocessed and consolidated into intermediate tabular representations, with vessel records reconciled through persistent identifiers. Source fields and contextual entities were subsequently mapped to the Wikibase data model, imported as entities, statements and references, and finally validated and published as versioned dataset releases.

Implementation. The knowledge graph was implemented using Wikibase, which provides an entity–property data model with native support for multi-valued statements, qualifiers, and references. This framework enables the representation of maritime information drawn from multiple institutional sources while preserving explicit provenance information at the statement level.

Each entity was represented through persistent identifiers associated with semantic properties describing administrative, technical, geospatial, and classification-related attributes. Source-specific values originating from different institutional sources were preserved as parallel statements linked to their corresponding provenance references, allowing transparent comparison of alternative values without enforcing a single canonical representation.

Data Records

The dataset is publicly released as a versioned resource and provides a unified representation of the European fishing fleet and its associated maritime infrastructure. The released records describe vessels, infrastructure, classifications, administrative entities, and provenance relationships derived from the integration workflow described in the Methods section.

All data records are distributed in interoperable formats to support reuse, querying, and downstream analysis.

All artefacts distributed in the Zenodo record correspond to the same EFFKG v1.1.0 release. The deposited RDF and Wikibase JSON files were exported from the EFFKG Wikibase state corresponding to v1.1.0, and the archived validation reports were generated against that same release state. The accompanying source-code archive contains the processing, population, export, and validation implementation associated with the same dataset version. The public Wikibase instance serves as the continuously maintained representation of EFFKG and may evolve as new data and corrections are incorporated; each subsequent EFFKG release freezes this evolving state into a new versioned snapshot. Accordingly, the Zenodo deposit constitutes the immutable reference snapshot for EFFKG v1.1.0. SHA-256 checksums for the deposited files are provided in the accompanying `SHA256SUMS.txt` file to support independent file-integrity verification.

Data repository and access. The dataset is publicly released through the Zenodo repository and is available as a versioned and citable resource[39]. The deposited dataset corresponds to version 1.1.0.

The released resources include:

- Zenodo dataset https://doi.org/10.5281/zenodo.21885296
- Wikibase landing page: https://effkg.wikibase.cloud/wiki/EFFKG
- Competency Questions (executable queries): https://effkg.wikibase.cloud/wiki/Competency_Questions
- Public SPARQL endpoint: https://effkg.wikibase.cloud/query/
- Public source-code repository: https://github.com/EnriqueJRodriguez/EFFKG

All distributed files correspond to frozen snapshots of the integrated knowledge graph and are preserved to support reproducibility and long-term reuse. Data archives, schema artefacts, and validation resources are distributed through the Zenodo repository associated with the dataset DOI.

File structure and formats. The released dataset corresponds to a frozen snapshot of the EFFKG knowledge graph distributed through the Zenodo repository (version 1.1.0).

The deposited package includes:

- RDF serialisations and native Wikibase JSON exports of the complete knowledge graph.
- Schema definitions and Shape Expressions (ShEx) artefacts describing the implemented data model.
- Validation outputs, integrity reports, and SPARQL quality-assessment queries.
- Archived source code for the data-processing, Wikibase population, export, and validation workflows, together with fixed software dependencies and execution documentation.
- Machine-readable documentation, including entity-class and property dictionaries, property provenance mappings, and the source-resource catalogue.
- SHA-256 checksums for independent integrity verification of the deposited release files.

Archive	Format	Contents
effkg-v1.1.0-data	TAR.GZ	Compressed archive containing RDF serialisations and Wikibase JSON exports of the integrated knowledge graph.
effkg-v1.1.0-data	ZIP	Compressed archive containing RDF serialisations and Wikibase JSON exports of the integrated knowledge graph.
schema	ZIP	Schema definitions and Shape Expressions (ShEx) artefacts describing the implemented data model.
validation	ZIP	Validation outputs, integrity reports, and SPARQL quality-assessment queries.
code	ZIP	Archived source code corresponding to the EFFKG v1.1.0 release, including the data-processing, Wikibase population, export, and validation workflows, together with fixed software dependencies, documented input requirements, and execution instructions.
documentation	ZIP	Machine-readable documentation package comprising entity-class and property dictionaries, property provenance mappings, and a source-resource catalogue describing every provenance resource referenced within the knowledge graph.
SHA256SUMS	TXT	SHA-256 checksums for the deposited release files, enabling independent verification of file integrity.

Table 4. Distributed artefacts included in the EFFKG v1.1.0 release.

The dataset is distributed as compressed archives intended for offline reuse, graph-based analysis, and reproducible research workflows. The principal artefacts distributed with this version are summarised in Table 4.

The development version of the source code is maintained through the public GitHub repository referenced in the Code Availability section, while the code corresponding to each released dataset version is archived with the associated Zenodo release.

Data structure and representation. The dataset follows an entity–property model in which each entity is described through a set of statements. Each statement consists of:

- a property (attribute).
- a value (entity or literal).
- optional qualifiers (e.g. temporal context or units).
- provenance references linking the statement to its originating source.

This structure allows multiple values for the same property to coexist, supporting the representation of discrepancies across sources while preserving traceability and source attribution.

To facilitate reuse of the semantic model, the released dataset includes a machine-readable documentation package comprising four complementary resources. The property dictionary documents all Wikibase properties, including persistent identifiers, definitions, datatypes, domains, ranges, cardinalities, qualifier and reference roles, and controlled values where applicable. A complementary entity-class dictionary describes the principal entity classes, inheritance relationships, and direct and inherited properties. A property provenance mapping records the source datasets contributing values to each implemented property. Finally, a source-resource catalogue documents every canonical resource referenced through statement-level provenance, including provider information, canonical URLs, integrated versions or snapshots, retrieval metadata, licensing conditions, representation within EFFKG, and the mechanisms required to retrieve the corresponding source records. Together, these resources provide a complete machine-readable description of the semantic model and provenance architecture implemented in this dataset version.

Provenance and source attribution. Data records include explicit provenance information at the statement level. Statement-level references use the Wikibase source property to identify the authoritative resource from which each value was obtained. Within EFFKG, this property records the canonical access URL of the corresponding source resource, providing a compact identifier that uniquely associates every statement with its origin. The complete provenance associated with each referenced resource—including provider, version or snapshot, retrieval information, licensing conditions, and record-identification mechanisms—is documented separately in the accompanying source-resource catalogue distributed with every dataset release. Together with the persistent identifiers preserved in the corresponding EFFKG entities, this information enables users to retrieve the original source records whenever the external resources continue to provide record-level access. This provenance model supports transparent comparison of values originating from different authoritative resources while maintaining lightweight statement-level references and reproducible dataset releases.

Versioning and updates. The dataset follows a semantic versioning scheme (`MAJOR.MINOR.PATCH`). Each released version is associated with a persistent Digital Object Identifier (DOI) and represents a stable, citable snapshot of the dataset.

New versions may include updates to source datasets, corrections, schema extensions, and the integration of additional maritime resources. All previous versions remain publicly accessible to support reproducibility and long-term reuse.

Dataset component	Count
Vessel entities (active)	71,787
Vessel entities (historical)	116,383
Fishing port entities	1,997
Fishing areas	114
Administrative units	155
Fishing gear classes	71
Vessel categories	66
Total entities	190,629
Statements	3,707,656
References	4,033,845
Qualifiers	101,639
Multilingual entity coverage	99.97%

Table 5. Summary statistics of the EFFKG v1.1.0 release. Counts correspond to the source snapshots reported in Table 3. Entity-class counts are reported for the major categories. Vessel counts distinguish active and historical vessels according to the status recorded in the integrated source registries.

Entity category	Description	Example properties
Vessels	Fishing vessels and vessel-related operational entities	CFR, registration number, LOA, GT tonnage, main engine power, fishing gear category.
Ports	Maritime infrastructure and harbour entities	Name, code, coordinates, belongs to administrative unit.
Administrative units	Maritime and terrestrial administrative jurisdictions	Name, code, contains administrative unit, has ports.
Categories	Controlled classification systems and taxonomies	Code, abbreviation, has parent category, subclass of.
Fishing areas	Spatial maritime management and regulatory areas	Name, code, subarea of.
Equipment	On-board systems and fishing equipment	Code, category of, vessel–equipment associations.

Table 6. Main entity categories represented in the dataset.

Data Overview

The released dataset contains entities representing vessels, maritime infrastructure, classifications, administrative structures, and geospatial resources derived from the integrated maritime datasets. Table 5 provides a statistical overview of EFFKG v1.1.0, including the major entity classes and the overall numbers of statements, references, qualifiers and multilingual labels. The main entity categories represented in the dataset are summarised in Table 6.

Each entity is identified by a persistent identifier and described through statements linking properties, values, qualifiers, and provenance references.

Reconciliation outcomes. The identifier-based reconciliation described in the Methods section resulted in 27,379 of 27,773 (98.58%) distinct valid CFR identifiers from the Spanish vessel sources being matched with records in the European Fleet Register. These records were consolidated through exact CFR matching, and no vessel entities required textual reconciliation. National vessel records lacking a valid CFR were retained as independent entities identified by their source-specific national registration codes.

Technical Validation

The dataset was validated through a combination of structural schema conformance checks, provenance verification, referential integrity validation, and domain-specific plausibility assessments. Validation procedures were implemented through SPARQL-based quality-assessment queries executed over the complete integrated knowledge graph.

Overall, the validation process confirmed high structural consistency, extensive provenance coverage, complete referential integrity for evaluated entity relationships, and high plausibility across quantitative, temporal, and geospatial attributes.

Schema conformance and structural validation. Structural validation was performed using Shape Expressions (ShEx)[40] constraints defined for the main entity categories represented in the knowledge graph, including vessels, ports, administrative units, fishing areas, and classification entities.

Validation assessed conformance between the integrated graph and the expected data model, including:

- Presence of mandatory and optional properties.
- Datatype consistency.
- Cardinality constraints.
- Qualifier usage.
- Structural consistency of entity relationships.

To assess structural consistency, an inferred schema was automatically generated from the complete graph and compared against the expected structural definitions implemented for the dataset.

Validation confirmed that the graph structure closely follows the expected data model across all major entity categories. Core identifiers, classification relations, administrative hierarchies, geospatial attributes, and entity-valued relationships were consistently represented according to the defined schema constraints.

Coverage analysis further indicated that attributes expected to exhibit broad availability in the original institutional datasets retained similarly high availability after integration. More heterogeneous coverage was observed for attributes inherently dependent on specific institutional or national-level sources, reflecting differences in source-level availability rather than information loss introduced during integration.

The inferred schema generated during the structural validation process is included as part of the released validation artefacts described in the Data Records section.

Provenance validation. Provenance validation evaluated the presence and consistency of statement-level references across 3,707,656 statements, of which 3,707,230 contained at least one provenance reference, resulting in an overall provenance coverage of 99.99%.

Complete or near-complete provenance coverage was observed across the principal domain-relevant properties represented in the dataset, including vessel identifiers, technical vessel characteristics, registration metadata, classification relations, and fishing-area associations.

The limited number of statements lacking provenance references was concentrated in structural or ontology-related properties, including `subclass of`, `category of`, and a small number of auxiliary administrative or geospatial relations. These cases primarily correspond to schema-level or classification metadata rather than operational vessel data.

The provenance validation outputs are included in the released validation artefacts.

Referential integrity validation. Referential integrity validation was performed to verify that entity-valued relationships resolve to existing entities within the knowledge graph and do not produce orphan or unresolved references.

Validation evaluated entity-valued relations defined in the Wikibase schema, including vessel category assignments, fishing gear categories, administrative hierarchies, fishing-area relations, port associations, equipment relations, and ontology-level entity links.

A total of 1,190,190 entity-valued links distributed across 18 relationship properties were evaluated during the validation process. No broken or orphan references were detected, resulting in an overall broken-link rate of 0%.

All evaluated relationships successfully resolved to existing entities within the graph, indicating that entity linking and relationship construction were consistently preserved throughout the integration process.

The referential-integrity validation outputs and SPARQL quality-assessment queries are included in the released validation artefacts.

Domain plausibility validation. Domain-specific plausibility checks were applied to detect semantically invalid or inconsistent values across quantitative, temporal, and geospatial attributes.

Validation assessed quantitative vessel measurements, lifecycle-related temporal relations, and geospatial coordinate consistency. Evaluated checks included verification of positive LOA, GT tonnage, and engine-power values, temporal consistency between construction, entry-into-service, retirement, and destruction dates, and validation of geographic coordinate ranges for port entities.

A total of 809,299 values were evaluated through domain plausibility checks, of which 809,096 satisfied the expected domain constraints, resulting in an overall plausibility rate of 99.98%.

Quantitative vessel attributes exhibited complete plausibility compliance, including LOA values, GT tonnage measurements, engine-power values, and geospatial coordinate validation. Temporal consistency validation identified a limited number of anomalous date relationships involving construction, entry-into-service, and retirement dates.

The detected anomalies primarily reflect inconsistencies inherited from the original institutional source datasets rather than errors introduced during knowledge graph construction. In accordance with the provenance-preserving integration strategy adopted in this work, anomalous source values were retained together with their associated provenance metadata.

The domain-validation outputs are included in the released validation artefacts.

Limitations of validation. Despite the applied validation procedures, several limitations remain.

Vessel reconciliation is based predominantly on persistent identifiers and therefore depends on the completeness and correctness of the identifiers provided by the original registries. Port resolution from vessel registration-place values may be affected by spelling variation, multilingual names, homonymous locations, and incomplete reference metadata. Because ports are contextual entities rather than the principal units of cross-source integration, exhaustive manual verification of port-name resolutions for all vessels was outside the scope of this dataset version.

Differences in update frequency, publication schedules, and maintenance practices across institutional data sources may also introduce temporal inconsistencies between integrated datasets.

Attribute completeness remains heterogeneous across entity categories because certain variables are only available from specific institutional or national-level sources. Consequently, the absence of a given attribute does not necessarily indicate missing integration, but may instead reflect limited source-level availability.

Competency question	Demonstrated capability
Which ports in Galicia contain more than 50 registered fishing vessels?	Administrative hierarchy traversal, spatial aggregation and integrated vessel–port relationships.
Which vessels have inconsistent main-engine power values between the European Fleet Register and the Spanish Fleet Register?	Provenance-aware comparison of multiple authoritative sources and identification of conflicting attribute values.
How many Spanish fishing vessels use bottom trawl gears within FAO Area 37?[a]	Integration of vessel characteristics, fishing gear classifications and fishing-area information across heterogeneous datasets.
What is the most common fishing gear in each European national fleet?	Cross-country comparative analysis enabled by harmonised vessel registration and fishing gear classifications.

Table 7. Representative competency questions supported by EFFKG. [a]FAO fishing-area information is currently available only for the Spanish national dataset integrated into EFFKG.

Validation primarily focused on structural consistency, referential integrity, provenance coverage, and domain-level plausibility rather than exhaustive manual verification of all source-provided factual statements.

Consequently, some inconsistencies inherited from the original datasets may remain represented in the knowledge graph rather than being silently corrected during integration.

Finally, although automated validation procedures were applied over the complete graph, domain plausibility constraints necessarily represent only a partial approximation of all semantic inconsistencies that may exist in real-world fisheries and maritime operational data.

These limitations are partially mitigated through the dataset's explicit provenance model, which preserves source-specific representations and supports transparent cross-source comparison, reproducibility, and downstream quality assessment.

Usage Notes

This section provides practical notes for accessing, querying, and interpreting the released dataset.

Access and querying. The dataset can be accessed through RDF dumps, Wikibase JSON exports, the Wikibase web interface, and a public SPARQL endpoint. The released artefacts support both offline processing workflows and interactive graph querying.

The SPARQL endpoint enables retrieval of information across vessels, ports, administrative entities, fishing areas, classifications, and provenance metadata. Example query collections and additional query documentation are provided through the dataset repository and the Wikibase query interface.

Representative competency questions. The analytical scenarios outlined in the *Background & Summary* section are illustrated through a set of representative competency questions. Rather than constituting an exhaustive catalogue of supported analyses, these examples demonstrate how the semantic integration of fisheries datasets from multiple providers enables realistic cross-domain analytical workflows that would otherwise require substantial manual reconciliation across independent data sources.

Table 7 presents four representative competency questions covering the principal analytical opportunities enabled by EFFKG, including spatial analyses, source-resolved comparison of authoritative registries, integration of fishing gear classifications with FAO fishing areas, and cross-country analyses of European fishing fleets. For each competency question, a corresponding executable SPARQL query is provided through the dedicated *Competency Questions* page of the EFFKG Wikibase documentation (Data Records), allowing users to execute, inspect and adapt the queries directly through the public Wikibase Query Service.

Together, these competency questions illustrate how EFFKG transforms independently maintained institutional datasets into a unified analytical resource. By harmonising identifiers, preserving statement-level provenance and explicitly representing semantic relationships across vessels, ports, administrative units, fishing areas and classification systems, the knowledge graph enables reproducible analyses that are difficult to perform using the original datasets in isolation. Although representative rather than exhaustive, these examples provide a practical demonstration of the scientific opportunities enabled by the semantic integration approach adopted in EFFKG.

Interpretation of multi-source data. The dataset adopts an integration model with explicit source attribution in which multiple values for the same attribute may coexist when provided by different institutional sources.

Users should therefore consider that:

- Different values for the same property may originate from different datasets.
- Provenance references identify the originating source of each value.
- No single source is enforced as globally authoritative across all entity types and attributes.

For example, vessel characteristics or administrative attributes may differ between supranational and national registries. These differences are preserved explicitly in order to support transparent cross-source comparison and reproducible downstream analyses.

Best practices and limitations. Certain attributes are only available for subsets of entities because their availability depends on the coverage provided by the original institutional datasets. Consequently, missing values do not necessarily indicate incomplete integration.

Differences in update frequency and publication schedules across integrated sources may also introduce temporal inconsistencies between records originating from different sources.

Users performing source-sensitive analyses are encouraged to include provenance constraints in their SPARQL queries and to consider the temporal and institutional context associated with integrated values.

EFFKG does not currently incorporate dynamic vessel-tracking data and should therefore not be used to infer vessel trajectories or fishing activity patterns. Future extensions could consider integration with AIS, VMS, or other dynamic maritime data sources where appropriate access, licensing, privacy, and reuse conditions permit.

Data availability

The EFFKG dataset is publicly available through the Zenodo repository: https://doi.org/10.5281/zenodo.21885296. Repository landing page: https://zenodo.org/records/21885296. The released version (v1.1.0) includes the RDF serialisations and Wikibase-compatible JSON exports of the complete knowledge graph, together with the accompanying resources described in the Data Records section. A public SPARQL endpoint for interactive querying is available at: https://effkg.wikibase.cloud/query/. The Wikibase web interface is available at: https://effkg.wikibase.cloud/wiki/EFFKG

Code availability

The source code used for data preprocessing, entity reconciliation, validation, and knowledge graph generation is publicly available through the EFFKG source-code repository. The code version corresponding to the released EFFKG v1.1.0 dataset is additionally archived together with the dataset in the Zenodo record (DOI: https://doi.org/10.5281/zenodo.21885296), providing a persistent snapshot of the implementation associated with this release. The repository contains the scripts and processing workflows required to generate the released dataset and reproduce the integration pipeline described in the Methods section, including data harmonisation, validation procedures, RDF export, and Wikibase data transformation. Fixed software dependencies, documented input requirements, and execution instructions are provided with the source code. The source code is distributed under the MIT license.

Received: 28 May 2026; Accepted: 10 October 2026;
Published: xx xx xxxx

References

1. Rousseau, Y., Watson, R. A., Blanchard, J. L. & Fulton, E. A. Evolution of global marine fishing fleets and the response of fished resources. *Proc. Natl. Acad. Sci.* **116**, 12238–12243 https://doi.org/10.1073/pnas.1820344116 (2019).
2. Gangemi, A. *et al.* A formal ontological framework for semantic interoperability in the fishery domain (2002).
3. Tzitzikas, Y. *et al.* Unifying heterogeneous and distributed information about marine species through the top level ontology MarineTLO. *Program* **50**, 16–40 https://doi.org/10.1108/PROG-10-2014-0072 (2016).
4. Martín Míguez, B. *et al.* The European Marine Observation and Data Network (EMODnet): Visions and Roles of the Gateway to Marine Data in Europe. *Front. Mar. Sci.* **6**, 313 https://doi.org/10.3389/fmars.2019.00313 (2019).
5. Snowden, D. *et al.* Data Interoperability Between Elements of the Global Ocean Observing System. *Front. Mar. Sci.* **6**, 442 https://doi.org/10.3389/fmars.2019.00442 (2019).
6. Buck, J. J. H. *et al.* Ocean Data Product Integration Through Innovation—The Next Level of Data Interoperability. *Front. Mar. Sci.* **6**, 32 https://doi.org/10.3389/fmars.2019.00032 (2019).
7. Development of Spatial Data Infrastructures for Marine Data Management. Tech. Rep., Open Geospatial Consortium, Inc., https://doi.org/10.62973/19-025r1 (2019).
8. Núñez-Sánchez, M. J., Dalén, D. & Pérez-Rojas, L. A quantitative data analysis in relation to size, scope and behavioural patterns for the implementation of a port state control regime of the global industrial fishing vessel fleet. *WMU J. Marit. Aff.* **20**, 457–481 https://doi.org/10.1007/s13437-021-00251-1 (2021).
9. Nunez-Sanchez, M. J., Pérez Pérez, J. A. & Rojas, L. P. Are there sufficient incentives to decarbonize the trawl fleet in the EU? *Mar. Policy* **188**, 107062 https://doi.org/10.1016/j.marpol.2026.107062 (2026).
10. Palma, R. *et al.* Iliad Approach for Semantic Interoperability in Digital Twins of the Ocean. In *OCEANS 2025 Brest*, 1–10, https://doi.org/10.1109/OCEANS58557.2025.11104546 (IEEE, 2025).
11. Berre, A. J. *et al.* Digital Twins of the Ocean – Interoperability Pipeline Architecture. In *OCEANS 2025 Brest*, 1–5, https://doi.org/10.1109/OCEANS58557.2025.11104322 (IEEE, 2025).
12. Kavadas, S., Chamodrakas, I., Bitetto, I., *et al.* Final Report of the RDBFIS: Hosting, Maintenance and Further Development of the Regional Database for the Mediterranean and the Black Sea. Tech. Rep., Publications Office of the European Union, https://doi.org/10.2926/5619525 (2025).
13. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 https://doi.org/10.1038/sdata.2016.18 (2016).
14. Berners-Lee, T. Linked Data – Design Issues. W3C (2006).
15. Bizer, C., Heath, T. & Berners-Lee, T. Linked Data – The Story So Far. *Int. J. Semantic Web Inf. Syst.* **5**, 1–22 https://doi.org/10.4018/jswis.2009081901 (2009).
16. Camossi, E., *et al.* Maritime Linked Data for Situational Awareness Heterogeneous Sensor Networks. In *OCEANS 2021: San Diego – Porto*, 1–10, https://doi.org/10.23919/OCEANS44145.2021.9705759 (IEEE, 2021).
17. Zhang, Y. *et al.* Multi-Modal Spatio-Temporal Knowledge Graph of Ship Management. *Appl. Sci.* **13**, 9393 https://doi.org/10.3390/app13169393 (2023).
18. Calvanese, D. *et al.* Ontology-based data integration in EPNet: Production and distribution of food during the Roman Empire. *Eng. Appl. Artif. Intell.* **51**, 212–229 https://doi.org/10.1016/j.engappai.2016.01.005 (2016).
19. Zeginis, D., Kalampokis, E., Palma, R., Atkinson, R. & Tarabanis, K. A semantic meta-model for data integration and exploitation in precision agriculture and livestock farming. *Semantic Web* **15**, 1165–1193 https://doi.org/10.3233/SW-233156 (2024).
20. Buneman, P., Khanna, S., & Wang-Chiew, T. Why and Where: A Characterization of Data Provenance. In *Database Theory—ICDT 2001*, 316–330, https://doi.org/10.1007/3-540-44503-X_20 (Springer, 2001).

21. Carroll, J. J., Bizer, C., Hayes, P., & Stickler, P. Named graphs, provenance and trust. In *Proceedings of the 14th International Conference on World Wide Web*, 613, https://doi.org/10.1145/1060745.1060835 (ACM Press, 2005).
22. Hartig, O. Provenance information in the Web of Data. *LDOW***538** (2009).
23. Simmhan, Y. L., Plale, B. & Gannon, D. A survey of data provenance in e-science. *ACM SIGMOD Rec.* **34**, 31–36 https://doi.org/10.1145/1084805.1084812 (2005).
24. Vrandečić, D. & Krötzsch, M. Wikidata: a free collaborative knowledgebase. *Commun. ACM* **57**, 78–85 https://doi.org/10.1145/2629489 (2014).
25. Erxleben, F., Günther, M., Krötzsch, M., Mendez, J., & Vrandečić, D. Introducing Wikidata to the Linked Data Web. In *The Semantic Web – ISWC 2014*, 50–65 https://doi.org/10.1007/978-3-319-11964-9_4 (Springer, 2014).
26. Waagmeester, A. *et al.* Wikidata as a knowledge graph for the life sciences. *eLife* **9**, e52614 https://doi.org/10.7554/eLife.52614 (2020).
27. Diefenbach, D., Wilde, M. D. & Alipio, S. Wikibase as an Infrastructure for Knowledge Graphs: The EU Knowledge Graph. In *The Semantic Web – ISWC 2021*, 631–647, (Springer, 2021).
28. Thiery, F., Schenk, F., Stefan, D., Distel, A.-K. & Thiery, P. A. From Stones and Wells to Tephra: Leveraging Wikidata and Wikibase for Interdisciplinary Humanities Linked Data. *J. Open Humanit. Data* **12**, 10 https://doi.org/10.5334/johd.425 (2026).
29. Food and Agriculture Organization of the United Nations. International Standard Statistical Classification of Fishery Vessels by Vessel Types (ISSCFV – Vessel Type). Rev. 1 https://openknowledge.fao.org/handle/20.500.14283/cb5201en (2019).
30. Food and Agriculture Organization of the United Nations. International Standard Statistical Classification of Fishing Gear (ISSCFG, 2016). 2016.1 https://openknowledge.fao.org/handle/20.500.14283/bt988e (2016).
31. Food and Agriculture Organization of the United Nations. FAO Major Fishing Areas Accessed March 2026. https://www.fao.org/fishery/area/search/en (2026).
32. National Geospatial-Intelligence Agency. World Port Index Snapshot integrated in EFFKG: March 2026. https://msi.nga.mil/Publications/WPI (2026).
33. Directorate-General for Maritime Affairs and Fisheries, European Commission. European Fleet Register Snapshot integrated in EFFKG: December 2025. https://vessel-register.oceans-and-fisheries.ec.europa.eu/ (2025).
34. Ministerio de Agricultura, Pesca y Alimentación. Registro General de la Flota Pesquera Accessed March 2026. https://servicio.pesca.mapama.es/censo/ConsultaBuqueRegistro/Buques/Search (2026).
35. Ministerio de Agricultura, Pesca y Alimentación. *Situación de la Flota Pesquera Española* 2025 edition. https://www.mapa.gob.es/es/pesca/temas/registro-flota/informacion-sobre-flota-pesquera (2025).
36. Instituto Geográfico Nacional. SIGNA Snapshot integrated in EFFKG: February 2026. https://signa.ign.es/signa/ (2026).
37. Centre National de Surveillance des Pêches. *Liste des ports du système ERS avec données de position* Accessed March 2026. https://www.data.gouv.fr/datasets/liste-des-ports-du-systeme-ers-avec-donnees-de-position (2026).
38. Publications Office of the European Union. Commission Implementing Regulation (EU) 2017/1130 of 19 June 2017 establishing a fisheries data exchange system. EUR-Lex, accessed March 2026. http://data.europa.eu/eli/reg_impl/2017/1130/oj (2017).
39. Rodriguez-Martin, E., Álvarez-Fidalgo, J., Luna, M., & Labra-Gayo, J. E. EFFKG: European Fishing Fleet Knowledge Graph. Zenodo https://doi.org/10.5281/zenodo.21885296 (2026).
40. Prud'hommeaux, E., Labra Gayo, J. E., & Solbrig, H. Shape Expressions: an RDF validation and transformation language. In *Proceedings of the 10th International Conference on Semantic Systems*, 32–40 https://doi.org/10.1145/2660517.2660523 (ACM, 2014).

Author contributions

E.R. contributed to conceptualization, software development, data curation, validation, methodology, and manuscript writing. J.A. contributed to software development, data curation, validation, and manuscript review and editing. M.L. contributed to supervision and manuscript review. J.E.L. contributed to supervision, project administration, funding acquisition, and manuscript review. All authors reviewed and approved the final manuscript.

Funding

This work was partially funded by the Spanish Research Agency project "Dando forma a los grafos de conocimiento e interoperables" (PID2024-157010OB-I00), and by the regional project SEK-25-GRUGIC-24-089.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to E.R.-M.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2026

This Article in Press is shared early to give you faster access to new research. It is citable and carries a permanent DOI. The final edited version will replace it automatically.