

OPEN
ARTICLE

Assessing semantic interoperability in environmental sciences: variety of approaches and semantic artefacts

Cristina Di Muri^{1,6}✉, Martina Pulieri^{2,3,6}✉, Davide Raho¹, Alexandra N. Muresan¹, Andrea Tarallo¹, Jessica Titocci¹, Enrica Nestola¹, Alberto Basset^{1,2,4}, Sabrina Mazzoni^{2,5} & Ilaria Rosati^{1,3}

The integration and reuse of digital research products can be only ensured through the adoption of machine-actionable (meta)data standards enriched with semantic artefacts. This study compiles 540 semantic artefacts in environmental sciences to: *i.* examine their coverage in scientific domains and topics; *ii.* assess key aspects of their FAIRness; and *iii.* evaluate management and governance concerns. The analyses showed that the majority of semantic artefacts concern the terrestrial biosphere domain, and that a small portion of the total failed to meet the FAIR principles. For example, 5.5% of semantic artefacts were not available in semantic catalogues, 8% were not built with standard model languages and formats, 24.6% were published without usage licences and 22.4% without version information or with divergent versions across catalogues in which they were available. This investigation discusses common semantic practices, outlines existing gaps and suggests potential solutions to address semantic interoperability challenges in some of the resources originally designed to guarantee it.

Introduction

In the age of technology-driven and machine-assisted research, data are no longer only those gathered by humans through observations, laboratory/field experiments and hypothesis testing but also those arising in digital form (*e.g.* lines of code, software, or AI-generated data). In addition, with the ever-increasing automation and instrumentation advances, machines have become generators of data as well as analytical assistants, such as in the case of autonomous monitoring systems, remote sensing devices, sensors and so forth. All such types of digital data require to be properly interpreted and reused by different researchers and computers to be fully exploited¹.

Semantic interoperability allows the exact transmission of the format and the meaning of research products as well as the precise exchange of information between parties², hence, it represents one of the fundamental pillars of the FAIR principles³, a set of guidelines to improve the Findability, Accessibility, Interoperability, and Reusability of scientific data and other research products including workflows, software and semantic artefacts^{4–6}.

The adoption of semantic artefacts is required to achieve semantic interoperability, to describe the meaning of data and relations among them. Semantic artefacts are machine-readable, -interpretable and -actionable formalisations of concepts that can be used and exchanged to encode and predictably decode information, thus enabling the discovery, integration and reuse of information by both humans and machines⁷. Semantic artefacts provide a framework for conflict resolution between terms developed independently across different disciplines or groups of interest.

Thus far, different terms may have been used to describe the same concept or, alternatively, the same term may have been used to express several concepts across different disciplines/communities, thereby impeding

¹National Research Council of Italy (CNR), Research Institute on Terrestrial Ecosystems (IRET), Lecce, 73100, Italy.

²University of Salento, Department of Biological and Environmental Sciences and Technologies, Lecce, 73100, Italy. ³LifeWatch Italy, Lecce, 73100, Italy. ⁴LifeWatch ERIC, Lecce, 73100, Italy. ⁵National Research Council of Italy (CNR), Institute for Sustainable Plant Protection (IPSP), Sesto Fiorentino, 50019 FI, Italy. ⁶These authors contributed

equally: Cristina Di Muri, Martina Pulieri. ✉e-mail: cristina.dimuri@cnr.it; martina.pulieri@unisalento.it

interoperability. For such a reason, despite semantic artefacts are created and used to align research products with the FAIR principles, their variety and heterogeneity are considered a source of lack of semantic interoperability².

Semantic artefacts may have a broad range of formalisations, from loose sets of terms such as lists, glossaries, and categorisation schemes to higher-order logic constructs such as thesauri and ontologies^{8,9}. As a consequence, they are built using different standard models (e.g. RDFS, OWL, SKOS) and serialisation schemes (e.g. XML, XML Schema, JSON, RDF/XML, OWL/XML, JSON-LD, Turtle, N-Triples, etc.)⁷.

Semantic artefacts are often stored and shared by means of semantic catalogues that facilitate their discoverability and access, support their management and use, and in some cases provide additional editing tools and services¹⁰. These catalogues can be registries, namely, metadata catalogues of semantic artefacts, and repositories that store and offer access to semantic artefacts metadata as well as to their content^{2,7,8}. Even though these catalogues facilitate the discovery of and access to semantic artefacts in accordance to the FAIR principles, the distribution of such resources over multiple catalogues could, in some cases, lead to duplications and inconsistencies that compromise their long term sustainability.

Accounting for the value of semantic interoperability in interdisciplinary research, such as that of environmental sciences, an extensive collection and analysis of existing semantic artefacts in such a context can be critical to outline the state of the art of semantic interoperability and to highlight existing gaps.

This study was initiated in the framework of the project “Italian Integrated Environmental Research Infrastructures System” (ITINERIS), a multidisciplinary project coordinating a network of national nodes from 22 environmental Research Infrastructures (RIs). One of the main aims of the project is to build the Italian Hub of RIs providing centralised access to existing and novel FAIR data and related services (e.g. codes, software and any other digital research outputs; Digital Objects; DOs)¹¹ used to study environmental processes in the atmosphere, marine, terrestrial biosphere, and geosphere land surface domains. The achievement of such an objective requires the utilisation of FAIR semantic artefacts that can facilitate the search and retrieval of interdisciplinary DOs available through the Hub and enable their reuse. This endeavour can accelerate the synthesis of existing scientific evidence and the generation of novel knowledge supporting decision-making to tackle current and expected environmental challenges¹².

With this in mind, an extensive search of existing semantic artefacts was carried out starting from the semantic repositories used by the ITINERIS RIs. The search was then widened to include further semantic artefacts and catalogues focusing on the environmental domain.

The collection of semantic artefacts and their metadata was performed to *i.* examine their coverage in topics and environmental domains; *ii.* assess key properties that influence their findability, accessibility, interoperability and reusability; *iii.* evaluate the management, governance and long term sustainability of semantic artefacts available across multiple catalogues. Specifically, the analysis described in *ii.* consisted in the evaluation of 13 metadata properties associated with seven FAIR sub-principles considered relevant for publishing FAIR semantic artefacts¹³. The properties and associated FAIR sub-principles (in brackets) included in the analysis are: identifiers (F1), inclusion in semantic catalogues (F4), status (A1), formality level, language and format (I1), description (R1), usage licence (R1.1) and version (R1.2).

This study can provide useful insights to the ITINERIS RIs that still do not fully embrace semantic interoperability practices and support their future implementation. In addition, and beyond ITINERIS, this analysis is pivotal to emphasise the strengths and weaknesses of semantic artefacts in the field of environmental sciences and also to showcase existing gaps and challenges still limiting cross-domain semantic interoperability.

Results

Coverage: environmental domains and topics. A total number of 540 semantic artefacts were compiled¹⁴ and classified, according to the ITINERIS project, in the four environmental domains. Specifically, 225 out of 540 resources provided terms exclusively for the terrestrial biosphere domain (41.7%; Fig. 1), 60 for the geosphere land surface domain (11.1%; Fig. 1), 48 for the marine domain (8.9%; Fig. 1), and four for the atmosphere domain (0.6%; Fig. 1). In addition, 143 semantic artefacts covered all environmental domains (26.5%; Fig. 1) and 60 concerned multiple domains (Fig. 1).

A text mining analysis on semantic artefacts’ titles was carried out to identify the topics covered (Fig. 2). The results showed that the most frequent terms were parameter (*freq* = 43), anatomy (*freq* = 24), environment (*freq* = 23), development (*freq* = 19), phenotype (*freq* = 15), biological (*freq* = 14), coastal (*freq* = 14) and marine (*freq* = 14), method (*freq* = 13), trait (*freq* = 12), observation (*freq* = 11), plant (*freq* = 11), taxonomy (*freq* = 11), data (*freq* = 10), sensor (*freq* = 10), and unit (*freq* = 10).

FAIRness of semantic artefacts. *Findability: identifiers and semantic catalogues.* The findability of semantic artefacts was evaluated by performing an analysis of identifiers (F1; mod:URI; dct:identifier) and of semantic catalogues in which the semantic artefacts were available (F4; schema:includedInDataCatalog).

All semantic artefacts were entities of the digital realm that can be identified with Uniform Resource Identifiers (URIs), Persistent Uniform Resource Locators (PURLs) or through Uniform Resource Locators (URLs) of the web pages in which they were available. In addition, 100 semantic artefacts were also minted with Digital Object Identifiers (DOIs).

As about their online distribution, 510 semantic artefacts were included in the semantic catalogues considered in this study (Table 1) and 30 were not (Fig. 3). Specifically, 145 semantic artefacts were available in the NERC Vocabulary Server, 121 in Bioregistry, 119 in BioPortal, 116 in Research Vocabularies Australia (ARDC), 115 in EMBL-EBI Ontology Lookup Service (OLS), 112 in FAIRsharing, 100 in AberOwl, 85 in Ontobee, 83 in Open Biological and Biomedical Ontology (OBO) Foundry, 78 in the I-ADOPT Catalogue of Terminologies, 71 in Basic Register of Thesauri, Ontologies and Classifications (BARTOC), 69 in AgroPortal, 18 in EcoPortal, 17 in Linked Open Vocabularies (LOV), 14 in the German Federation for Biological data

Catalogue name (acronym)	Type	Semantic artefacts
AberOWL	Repository	1,422
AgroPortal	Repository	150
Basic Register of Thesauri, Ontologies and Classifications (BARTOC)	Registry	3,429
BioPortal	Repository	1,044
Bioregistry	Registry	1,650
NERC Vocabulary Server (NVS)	Repository	292
EcoPortal	Repository	25
FAIRsharing	Registry	833
German Federation for Biological data (GFBio)	Repository	29
I-ADOPT Catalogue of Terminologies	Registry	84
Linked Open TERminology REsources (LOTERRÉ)	Repository	65
Linked Open Vocabularies (LOV)	Repository	812
Marine Metadata Interoperability Ontology Registry and Repository (MMI-ORR)	Repository	330
Ontobee	Repository	262
Open Biological and Biomedical Ontology (OBO) Foundry	Registry	185
EMBL-EBI Ontology Lookup Service (OLS)	Repository	242
Research Vocabularies Australia (ARDC)	Repository	443

Table 1. List of semantic catalogues, in alphabetical order, used to gather semantic artefacts. The table includes the catalogues' full names and acronyms with URLs, the type of catalogue (repository or registry), and the total number of semantic artefacts available.

and systems to understand the context in which these terms are defined, and to retrieve the correct information. Overall, seven formality levels were found including ontologies ($N = 199$), thesauri ($N = 153$), glossaries ($N = 145$), gazetteers ($N = 18$), categorisation schemes ($N = 13$), subject headings ($N = 7$), and authority files ($N = 5$).

These semantic artefacts were expressed in different languages and formats and, in most cases, they were available and downloadable in multiple formats (Fig. 4). A total of 497 semantic artefacts were built using standard languages and 43 were not. Specifically, 295 were built in SKOS, 201 in OWL, and one in RDFS¹⁴. The 497 semantic artefacts with a standard language were expressed as RDF/XML ($N = 439$; Fig. 4), Turtle ($N = 271$; Fig. 4), JSON-LD ($N = 253$; Fig. 4), OWL/XML ($N = 168$; Fig. 4), N-Triples ($N = 131$; Fig. 4), Notation3 ($N = 55$; Fig. 4), OBO ($N = 51$; Fig. 4), RDF/JSON, N-Quads, and TriG ($N = 47$; Fig. 4), Sesam Binary RDF and TriX ($N = 40$; Fig. 4), SKOS-Core ($N = 20$; Fig. 4) and other less common formats ($freq \leq 8$). In addition, many of these artefacts were also available as CSV ($N = 172$) and PDF ($N = 10$). As for those semantic artefacts with no standard language ($N = 43$), 25 were available as XML, 12 as HTML, four as PDF, and three as CSV (Fig. 4).

Reusability: descriptions, usage licences and versions. The reusability was evaluated by analysing descriptions (R1; schema:description), usage licences (R.1.1; dct:licence) and versions available (R.1.2; pav:version).

Overall, 492 out of 540 semantic artefacts were published with a description or a brief abstract. Results of text mining analysis performed on descriptions showed that the most frequent terms were data ($freq = 139$), parameter ($freq = 45$), environment ($freq = 44$), plant ($freq = 42$), geoscience ($freq = 37$), trait and unit ($freq = 36$), development ($freq = 35$), anatomy, biological and marine ($freq = 34$), species ($freq = 32$), and biodiversity ($freq = 31$).

Usage licences were available for 407 semantic artefacts and, the majority of them ($N = 347$), were licensed under the Creative Common (CC) BY licence (64.3%; Fig. 5) and 35 were CC0 (6.5%; Fig. 5). In addition, 14 semantic artefacts were licensed under other types of CC BY licences (2.6%; Fig. 5) including CC BY-SA ($N = 5$), CC BY-NC-ND ($N = 3$), CC BY-NC-SA ($N = 3$), CC BY-ND ($N = 2$), and CC BY-NC ($N = 1$). Other types of licences were also found (2.0%; Fig. 5) including MIT ($N = 3$), BSD-3-Clause ($N = 3$), ODC-By ($N = 2$), and public licences ($N = 2$). Licences were not specified in 133 instances (24.6%; Fig. 5).

Version information was available for 419 semantic artefacts and different formats were used to present this information. The semantic versioning (*i.e.* MAJOR version.MINOR version.PATCH version¹⁵) was used in 273 instances, whereas dates were provided in 139 cases and, in seven cases, mixed formats (dates and semantic versioning) or alphanumeric strings (*i.e.* 3ed) were used.

Management and governance. To evaluate the management of semantic artefacts accessible from multiple catalogues, a versioning analysis was carried out on those resources available in at least two of the 16 catalogues considered for this analysis (the I-ADOPT catalogue does not provide version information or the download option; Table 2). The version information was extracted for each semantic artefact and catalogue in which the resource was available. This collection was followed by the calculation of the versions' alignment. In short, the number of times the resource had a unique value was counted and, such value, was then divided by the number of catalogues in which the resource and its version were available. When a semantic artefact had the same format and version across different catalogues, the resource was considered aligned. Conversely, if the version differed or it was expressed in a different format, the resource was considered misaligned.

The versioning analysis was performed on 177 semantic artefacts (Fig. 6). In fact, of the 540 semantic artefacts collected, 207 resources were shared over at least two semantic catalogues and, only 177 of them, were available with version information and, hence, version comparisons could be performed.

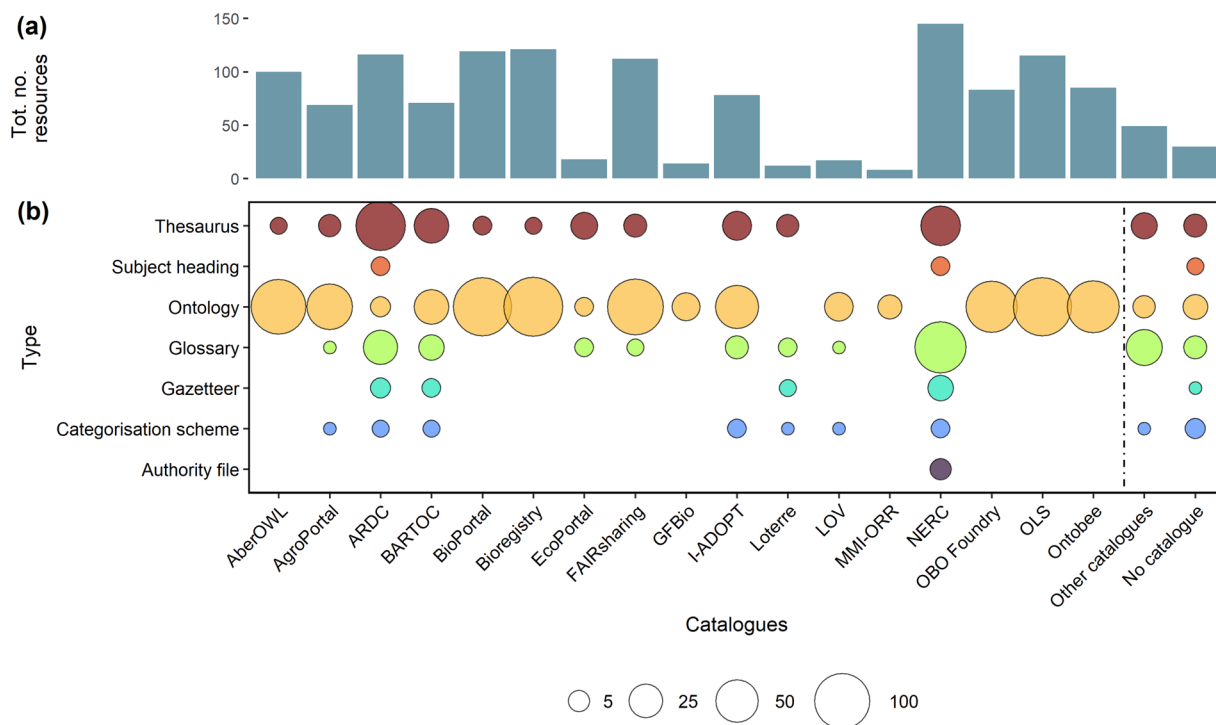


Fig. 3 Distribution, number and type of semantic artefacts. The barplot (a) shows the total number of semantic artefacts per catalogue. The bubble chart (b) describes the distribution and the type of semantic artefacts across the 17 semantic catalogues considered in this study. In “Other catalogues” and “No catalogue” are shown respectively the number and the type of semantic artefacts found in catalogues not considered within this study and those not available in any catalogue.

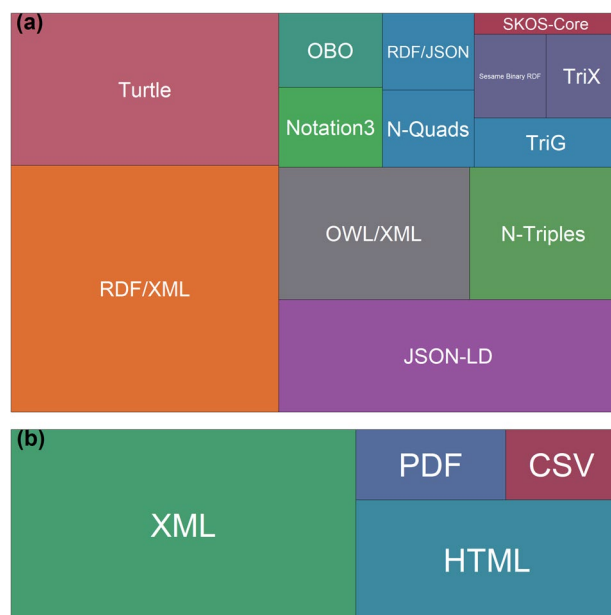


Fig. 4 Treemaps showing the digital representation formats available for semantic artefacts built with standard models (a) and with no standard model (b). The box sizes are proportional to the number of formats.

Overall, 68 semantic artefacts had aligned versions across different catalogues (value = 0; Fig. 6) and, such resources, were shared between a minimum of two and a maximum of eight catalogues. A total of 109 semantic artefacts (61.8%; Fig. 6) had misaligned versions across the different catalogues in which such resources were available. Of the 109 misaligned semantic artefacts, 32 had the maximum misalignment (value = 1; Fig. 6) and were shared between two and three catalogues. The remaining resources had a misalignment proportion ranging between 0.25 and 0.75 (Fig. 6) and were available in between three and nine catalogues.

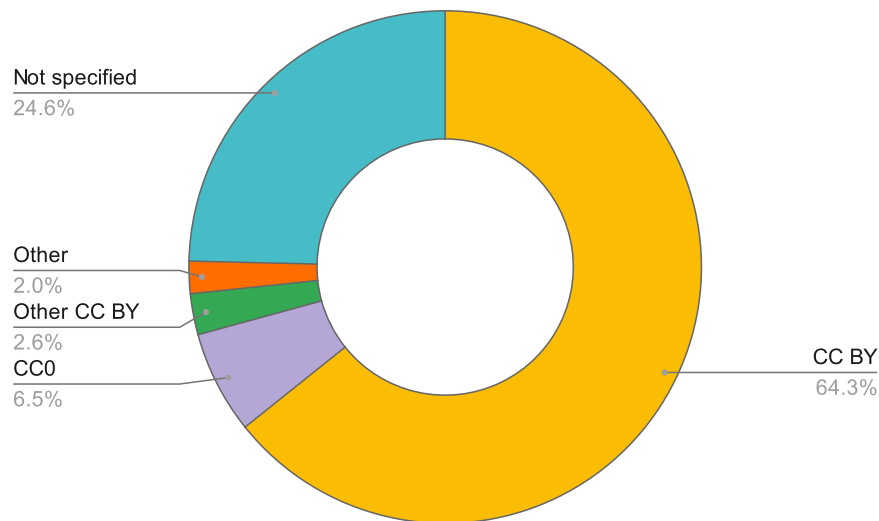


Fig. 5 Donut chart with percentages of the different licence types of the semantic artefacts gathered within this study.

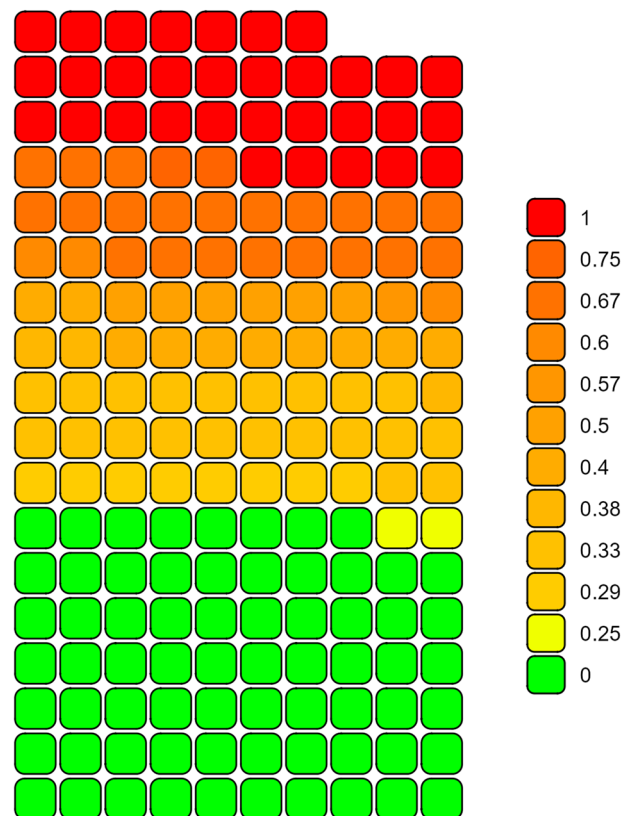


Fig. 6 Waffle chart showing the versions' misalignment of semantic artefacts available with version information on multiple (at least two) semantic catalogues. The misalignment is expressed as the proportion between the number of times in which each resource had a unique value and the number of catalogues in which the resource was available with a specified version. In the figure, each square represents a single semantic artefact ($N = 177$) and the proportion of misalignment varies from 0 (no existing misalignment among catalogues) to 1 (maximum misalignment among catalogues).

Discussion

This study has brought forth a plethora of semantic artefacts within the environmental domain¹⁴. Despite their abundance, it is widely acknowledged that their adoption remains deficient or suboptimal, particularly within certain scientific communities¹⁶.

Catalogue	URI	Identifier	Status	Deprecated	In catalogue	Formality	Language	Syntax	Licence	Version
AberOWL	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
AgroPortal	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
BARTOC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
BioPortal	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Bioregistry	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
NERC	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
EcoPortal	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
FAIRsharing	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
GFBio	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
I-ADOPT	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
LOTERRRE	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
LOV	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>
MMI-ORR	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Ontobee	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
OBO Foundry	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>
OLS	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
ARDC	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

Table 2. List of semantic catalogues and their exposed metadata. Black squares indicate the available metadata, white squares the metadata not available. Note: the “Title”, “Acronym”, and “Description” properties were not included in the table as they were available in all catalogues with I-ADOPT not showing the “Acronym” and “Description” properties. Refer to methods section for the full names of each metadata property.

The limited use of semantic artefacts could be attributed to the lack of appreciation and understanding of their impact but also, to some extent, to their partial or absent compliance with established good practices and principles (*i.e.* FAIR principles), ultimately hampering the (re)usability of these resources¹⁷. The poor metadata descriptions (*e.g.* missing licences, versions, descriptions) and the lack of standards (*e.g.* languages and formats) coupled with inconsistencies of information provided by different semantic catalogues, exacerbates this challenge determining a failure to meet the semantic interoperability.

In this context, the main findings of this study are discussed to highlight the key issues limiting the FAIRness and the flawless management of semantic artefacts in the environmental domain, and to provide a glimpse of existing gaps and challenges that prevent the advancements toward the envisioned web of linked data¹⁸.

Semantic artefacts in environmental sciences. This study performed an extensive analysis of 540 semantic artefacts in the environmental domain (Fig. 1, 3). These semantic artefacts have emerged independently, resulting in distinct models tailored and optimised for specific operations and scopes². This evolution has led to the creation of semantic silos, potentially holding redundancies and overlaps.

The classification of semantic artefacts in environmental domains and the text mining analysis on titles and descriptions led to comparable results in terms of coverage and representation of environmental entities. These results showed that the majority of semantic artefacts described biological entities including anatomies, developmental stages, phenotypes, traits and taxonomies (Fig. 2) of specific taxonomic groups or species, many of which represented model organisms largely used in the experimental research. Many other semantic artefacts described cross-domain entities including parameters, environments, observations and methods (Fig. 2). These cross-domain semantic artefacts can be considered valuable resources for multidisciplinary projects and initiatives involving heterogeneous research communities, such as the ITINERIS project.

The frequency analysis resulting from text mining showed that only a limited number of terms had the highest frequencies. This result could indicate existing overlaps and topics overrepresentation and a more in-depth analysis would be required to support such findings. Furthermore, an investigation of the degree of alignments among terms could be used to measure the interoperability among these semantic artefacts. The overrepresentation in the biosphere domain might also have been generated by the adoption of the ITINERIS classification. According to ITINERIS, all organisms on Earth, independently of their habitats and ecosystems, are considered part of the biosphere. Nevertheless, the consistency of coverage analysis performed on domains and topics demonstrated that these resources represent a large portion of semantic artefacts in environmental sciences.

FAIRness of semantic artefacts in environmental sciences. FAIR principles do not provide specific indications or implementation solutions as they are rather a set of guidelines to produce machine-actionable DOs¹⁷. While this flexibility has contributed to the rapid and widespread adoption of FAIR principles across different stakeholder communities, on the other side, it has generated a number of specific community-tailored solutions with the inherent risk of being incompatible among them¹⁹.

In this study, technical, technological and management choices related to the publication of semantic artefacts were analysed by examining core properties used to evaluate their FAIRness. The selected properties have been originally listed to design the framework of the O’FAIRe tool²⁰, an online, metadata-based automatic

FAIR assessment of semantic artefacts, already integrated into a number of semantic repositories^{21,22}. Although the O'FAIRe tool allows a quantitative FAIR assessment of semantic artefacts, this can only be integrated in catalogues adopting the Metadata Ontology Description and publication (MOD)²³ as metadata schema. The O'FAIRe score is computed based on the presence/absence of statements provided by the semantic artefacts' creators rather than on the assessment of metadata content. The challenge of automated FAIR assessments lies in the inability of such services to generate unified evaluations based on the interpretation of different standards, formats and types of descriptors used within the metadata²⁴. To overcome such limitations, this study focused on the evaluation of content and publication choices of semantic artefacts available across different or multiple semantic catalogues. Specifically, the analysis was performed on 13 core properties considered pivotal for the publication of FAIR semantic artefacts (Table 2). The analysis presented herein facilitated a broader and detailed evaluation of semantic artefacts distributed over multiple catalogues, each characterised by distinct metadata schemas and web interface visualisations^{25,26}.

The analysis of persistent identifiers and the inclusion in semantic catalogues was used to evaluate the findability of semantic artefacts.

Thus far, multiple guidelines have been published on the construction of Persistent Identifiers (PIDs)^{27,28}. By following the first FAIR principle (F1), both data and metadata of any DO should be identified with PIDs²⁹. For semantic artefacts, this entails the assignment of PIDs to semantic artefacts and concepts/classes included therein as well as to their metadata³⁰. This study focused on the collection of PIDs used to identify semantic artefacts (mod:URI) and DOIs minted to their metadata (dct:identifier). The results showed that only 18.5% of semantic artefacts metadata were assigned with DOIs, however such result depends on the management practices of semantic artefacts' catalogues in which the resources are available. For instance, semantic artefacts available in FAIRsharing are automatically assigned with DOIs³¹, and EcoPortal provides DOIs upon request²², whereas those available in OBO Foundry are identified with PURLs³². DOIs and PURLs are persistent identifiers, however, while PURLs can identify semantic artefacts, concepts/classes and their metadata, hence enabling machine interoperability, DOIs can only be minted to the resource landing page. Landing pages represent dead-ends for machines because it is not always clear which of the potentially many links available in the web page corresponds to the objects identified by the DOI^{29,30}.

Semantic catalogues that do not have agreement with external services providing PIDs, use different types of URIs nevertheless ensuring the long term findability and accessibility of semantic artefacts and their metadata through their infrastructures. This study found that 5.5% of semantic artefacts were not included in any catalogue (Fig. 3) and, in a few cases, semantic artefacts were published in generalist data catalogues that do not offer the services to manage them (Fig. 3). In such cases, semantic artefacts are accessible through URLs of projects/initiatives-specific web pages. These resources face inherent challenges in terms of discoverability and sustainability, both in the short and long term⁷. This vulnerability is exacerbated by the risk of becoming lost when the hosting website/web page becomes no longer accessible or maintained.

Status and maintenance are considered key properties to assess the accessibility of semantic artefacts, however, this information was not provided for the majority of them. In addition, only a few catalogues formally report this information using different terminologies (Table 2). For example, some catalogues use software development status (e.g. AgroPortal, EcoPortal and BioPortal), whereas other catalogues adopt custom terminologies to describe the level of development of semantic artefacts (e.g. FAIRsharing, OBO Foundry), thus generating comparability issues and confusion for end-users. Overall, only 2.4% of semantic artefacts were no longer maintained, although this value could vary in time as it is strictly related to the responsiveness of semantic artefacts' creators as well as to the promptness of catalogues' updates³². Very often semantic artefacts are created *ad-hoc* for specific projects and, the lifespan of these resources is likely to be tight to the project duration. In this context, the role of semantic catalogues is fundamental to ensure the long term accessibility to these resources, even if they are no longer maintained, and to enable a persistent reference of those DOs that were annotated with them⁸.

The interoperability principle was analysed by considering formality levels, representation languages and formats. Predominantly, semantic artefacts comprised ontologies and thesauri, while the remaining resources consisted of sets of terms characterised by simpler structures and functions, such as glossaries, gazetteers, and categorisation schemes (Fig. 3). The observed variety of formality levels used to represent environmental concepts mirrors the heterogeneity of themes, levels of complexity, and scopes (Figs. 1, 2) for which these semantic artefacts were developed³³. Ontologies were the most numerous component of this collection, primarily due to their capacity to precisely represent complex natural phenomena and entities such as environments, organisms and their interrelations³⁴. Moreover, their ability to generate novel knowledge offers a significant advantage in the multidisciplinary field of environmental sciences as it facilitates and/or automates new linkages and relationships in the networks of entities^{33,35}. The majority of semantic artefacts were expressed as machine-actionable constructs, however, about the 8% of them lacked standard representation languages and formats (Fig. 4) and did not adhere to semantic web standards³⁶. These semantic artefacts could be actively used by large communities that have invested significant time and effort in establishing agreed-upon terms. Transforming these resources into machine-actionable semantic artefacts holds the potential to amplify their reusability, concurrently limiting the duplication of efforts and minimising semantic overlaps if meaningful linkages are built^{37,38}.

The reusability of semantic artefacts was examined by evaluating available descriptions, usage licences and versions. Descriptions, summaries or abstracts of semantic artefacts can help end-users in understanding the general scope, content and communities of reference for which the resource was conceived and developed²⁴. Descriptions should be exhaustive in providing detailed information that can enhance their reusability, however, 8.9% of semantic artefacts were published without a description. The results of the text mining analysis performed on this property did not show significant differences with the one performed on titles, thus indicating that descriptions did not provide additional details. A further essential prerequisite to enable the reuse

of existing semantic artefacts is the availability of clear usage licences^{39,40}. Despite “usage licence” being listed among the minimum mandatory metadata for FAIR semantic artefacts⁴¹, 24.6% of the collected resources did not specify licences in their metadata (Fig. 5). The version is also a key aspect of semantic artefacts’ reusability whereby the semantic versioning is the recommended format to provide version information about semantic artefacts⁴². Semantic versioning enables the meaningful communication of the significance of a version change in the release of any DO¹⁵. The adoption of semantic versioning for the publication of semantic artefacts can facilitate, and in some instances, automate operations such as the diffing (*i.e.* the computation of changes between versions) and the retrieval of the most recent version by machines^{42,43}. The study findings showed that 22.4% of semantic artefacts were published without version information while semantic versioning was used in 50.5% of cases and dates or other formats were used in the remaining instances. Another study, which focused on the analysis of ontologies indexed in OLS, also revealed a lack of version information or differences in version formats²⁴. The study reported that only 2.5% of ontologies used the semantic versioning whereas the majority of them used date formats and 31.7% did not provide machine-readable versions²⁴.

Management and governance: current challenges and future perspectives. The storage and publication of semantic artefacts through semantic catalogues is highly recommended, however, those available in multiple catalogues can also generate management and governance challenges. An illustrative example is the existence of divergent versions of the same semantic artefact across different catalogues¹⁰. The versioning analysis found that 61.8% of the semantic artefacts available in at least two catalogues had distinct versions of the same resource (Fig. 6). This misalignment not only impacts the versions’ consistency but also affects other metadata properties evolving with it (*i.e.* identifiers, status, download links, relations to earlier versions, date and so forth)⁴².

Other inconsistencies, although not analysed here, can exist among catalogues. Such inconsistencies can be associated with different metadata properties and can generate confusion in end-user experiences as well as management issues when creating services for information exchange among systems. For example, when data collection occurred, discrepancies between catalogues were found in titles (*e.g.* the “Ontology of units of Measure” in OLS is the “Units of Measurement Ontology” in BioPortal; the “Core terms of Audubon Core” in LOV is the “Audiovisual Core Term List” in the GBIF Repository of Schemas), acronyms (*e.g.* the “Plant Trait Ontology” is PTO in BioPortal and TO in FAIRsharing and in OLS; “PATO” is the “Phenotypic Quality Ontology” in BioPortal and the “Phenotype And Trait Ontology” in OBO Foundry) and status (the “Anatomical Entity Ontology” is “Inactive” in OBO Foundry, in “Production” in BioPortal and “Ready” in FAIRsharing).

The community of practitioners, including developers and maintainers of semantic artefacts’ catalogues, are currently working on possible solutions to mitigate these challenges, such as in the case of the OntoPortal Alliance¹⁰. The consortium involves several research institutions and infrastructure teams dedicated to the development of semantic repositories in various disciplines based on the open and collaboratively developed OntoPortal software¹⁰. Within this consortium, developers and managers of the nine existing catalogues constantly strive to advance FAIR solutions for both semantic artefacts and their catalogues. This commitment involves the adoption of shared management and governance practices⁴⁴, including efforts to minimise the replications of semantic artefacts across different repositories or to ensure the alignment of resources in instances where replications exist¹⁰.

The FAIR-IMPACT project is a further notable example of such efforts and it includes different tasks and work packages focusing on metadata, ontologies and semantic interoperability⁴⁵. The project members are actively working on mapping existing FAIR assessment methods and developing a unique and integrated methodology for a consistent quantitative evaluation of semantic artefacts⁴⁶. This methodology will enable a pre-publication assessment of semantic artefacts and the comparison of FAIR assessments performed with different tools, overcoming some of the challenges that have emerged within this study. Furthermore, to overcome existing semantic silos, FAIR-IMPACT members are designing solutions for FAIR mappings. These mappings will be enriched with provenance information and other associated metadata allowing the seamless transfer of information through the assignment of specific URIs⁴⁷.

Recommendations and concluding remarks. This study presents a thorough analysis of semantic artefacts within the environmental domain available online and through different catalogues. This analysis cannot be considered comprehensive due to the continuous release of novel semantic artefacts and catalogues. A noteworthy instance is the official launch of [BiodivPortal](#) and [EarthPortal](#), featuring environmental semantic artefacts that hold relevance for both this study and the ITINERIS project. Such repositories were launched after the conclusion of the collection and analysis, and as such, they were not incorporated into the current study. However, the list of semantic artefacts¹⁴ should not be seen as a static document, but rather as a dynamic product capable of ongoing updates and integration in response to the publication of novel and pertinent resources.

This study, the collection of semantic artefacts, and all its new versions that may be published in the future, serve as a point of reference within the ongoing ITINERIS project, as well as for related initiatives and the broader environmental research community. In response to the gaps identified here, ITINERIS has established a task group on semantic interoperability and allocated funding to provide technical support and solutions to foster semantic interoperability among environmental RIs and beyond. With this study, the task group has laid the groundwork for the upcoming project developments which will consist in the publication of novel FAIR semantic artefacts, and in the extension and/or FAIRification of the existing ones. In addition, a terminology service is going to be designed and developed by taking advantage of the collection and assessment of semantic artefacts and catalogues performed in this study. The ITINERIS terminology service is conceived as a mediator between users, either humans or machines, and providers of semantic artefacts. Such mediation will facilitate the search and discovery of FAIR semantic artefacts and their meaningful links. These developments will foster

the reuse of semantic artefacts, hence, avoiding overlaps and replications, and will improve their interoperability by bridging alignments among them.

To preemptively address semantic inconsistencies, adopting this approach is not just advisable but imperative, as already widely acknowledged at national and international levels²⁶. This avenue represents the most meaningful way to enable FAIR science in practice and to achieve the envisioned full machine interoperation in the semantic web.

Methods

The collection of semantic artefacts (Fig. 7) was mainly performed through a search within 17 semantic repositories and registries listed in Table 1. These catalogues were selected as they offer access to a wide array of semantic artefacts relevant to different environmental domains including the atmosphere, marine, terrestrial biosphere and geosphere land surface domains of the ITINERIS project. The catalogues with less than 500 semantic artefacts were thoroughly examined either manually through their web interfaces or by using APIs or SPARQL interfaces. The catalogues with a larger number of semantic artefacts (> 500) were browsed by performing manual queries and filtering by specific keywords or relevant subjects/domains (check Table 1 for the number of semantic artefacts included in each catalogue). In addition to this preliminary investigation, the [Google search engine](#) was used to retrieve further semantic artefacts not included in the catalogues listed in Table 1. The Google search was performed using keywords such as “semantic resource” or “semantic artefact” together with environmental domain-specific keywords such as “marine” or “atmosphere” or “biosphere” or “geosphere” or “land surface”. During the collection, the semantic artefacts representing subsets of a main resource were not included in the list (e.g. the SWEET Ontology subsets *Material Element*, *Property Space Direction*, *Material Rock* contain classes belonging to the main SWEET ontology), whereas extensions were added (e.g. the SAREF4ENVI is a declared extension of the SAREF Ontology and includes different classes). The selected semantic artefacts were finally listed in a spreadsheet¹⁴. The collection of semantic artefacts included the harmonisation of the compiled metadata properties (Fig. 7). In fact, the catalogues adopt different metadata schemas and different labels to describe properties of semantic artefacts in their web interfaces. Considering such heterogeneity, the compiled properties were harmonised using, whenever possible, the labels from the Metadata for Ontology Description and publication (MOD v.2.0), and the following properties were collected:

- **Title:** A name given to the resource.
- **Acronym:** Often used as an identifier within some ontology platforms such as BioPortal or OBO Foundry.
- **Description:** A description of the item.
- **URI:** The URI of the ontology which is described by this metadata. When available, resource identifiers (i.e. URI or PURL) were used. If no resource identifier was available, the web page URL was added.
- **Identifier:** An unambiguous reference to the resource within a given context. This property was filled with the DOIs of the semantic artefacts when available.
- **Status:** The tracking information for the contents of the ontology.
- **Deprecated:** Specifies if the semantic artefact IRI is deprecated.
- **Included in data catalogue:** A data catalogue which contains this dataset. This property includes a list of all catalogues considered within this study (Table 1) and in which the semantic artefacts can be found.
- **Has formality level:** The level of formality of the semantic artefact (e.g. ontology, thesaurus, etc.). If not specified in the exposed metadata (Table 2), the formality level was assigned according to the classification described in previous studies^{9,35,48}. For those semantic artefacts stating the formality level either in the title or in the description, the data models and formats were checked and, in case of inconsistencies, the formality level was adjusted according to the chosen classifications^{9,35,48}. In addition, the semantic artefacts classified in the original description under the general umbrella term of “vocabulary” were checked to identify the most suitable formality level following the above-mentioned classifications.
- **Has representation language:** A language that is used to create an ontology.
- **Has syntax:** The syntax followed in the creation of the semantic artefact (e.g. XML, XML Schema, JSON, RDF/XML, OWL/XML, JSON-LD, Turtle, N-Triples, etc.). For each semantic artefact, all formats available across the different catalogues and web pages were extracted.
- **Theme:** A main category of the resource. The ITINERIS classification of the four environmental domains (i.e. atmosphere, marine, terrestrial biosphere, geosphere land surface) was used to categorise the shortlisted resources.
- **Licence:** A legal document giving official permission to do something with the resource. When not provided in the exposed metadata, licences were retrieved from the downloaded file of each resource. In case the semantic artefact was not provided in any downloadable format, the web page was searched and, if the licence could not be found, “Not Specified” was used as a value.
- **Version:** The version number of the resource. The most recent version of each semantic artefact was included. In addition, a further search was performed to gather and compare the resource versions across all catalogues in which the same resource was available to assess eventual discrepancies.
- **Source accessed on:** The resource is related to a source which was originally accessed or consulted on the given date as part of creating or authoring the resource.

Table 2 provides an overview of the exposed properties considered within this analysis for the 17 semantic catalogues used to compile the list of semantic artefacts.

Data analysis and data visualisation were performed in the R-Studio interface, R engine version 4.3.0⁴⁹ (Fig. 7).

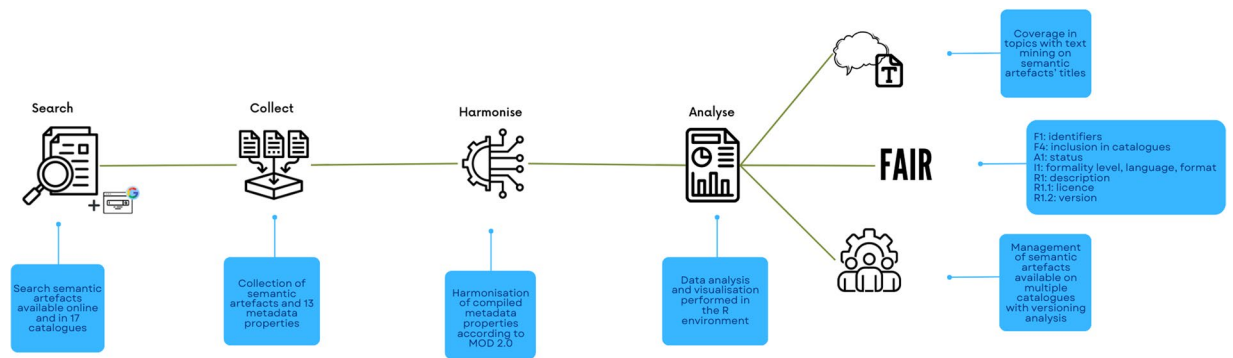


Fig. 7 Collection and analysis of semantic artefacts. The diagram shows the collection of semantic artefacts and their metadata using 17 catalogues and a Google search. The “MOD 2.0” was used to harmonise the metadata properties analysed in the study. The analyses included an evaluation of topics’ coverage, FAIRness and management.

Data manipulation was carried out with the packages *dplyr* v.1.1.2⁵⁰, *tidyverse* v.2.0.0⁵¹, *tidyr* v.1.2.0⁵², and *reshape* v.2.1⁵³. To find the most frequent topics covered by the semantic artefacts gathered herein, a text mining analysis was carried out on the full title of the semantic artefacts and their descriptions by using *tm* package v.0.7-11⁵⁴. The package *wordcloud* v.2.6⁵⁵ was used to display results of text mining analysis performed on titles. To clean the text, the functions *removePunctuation*, *removeNumbers*, *stripWhitespace*, and *content_transformer* within *tm* were used, respectively, to eliminate punctuation marks, numbers, and extra white spaces and to convert all letters to lowercase. The *tm* function *removeWords* was applied to remove stopwords and common words such as semantic, ontology, thesaurus, vocabulary, glossary, terms, and terminology as well as project-specific words such as NERC, Argo, SeaDataNet etc.

As for the versioning analysis, only the semantic artefacts shared with a specified version among at least two catalogues were considered. To calculate the versions’ alignment across semantic catalogues, the number of times in which the resource had a unique version value was counted and, such value, was then divided by the number of catalogues in which the resource was available. The resulting values ranged between 0 and 1, whereby 0 represented a perfect alignment between versions of the same semantic artefact across different semantic catalogues and 1 represented the maximum misalignment of versions, *i.e.* all catalogues in which the same resource was available had different versions of that semantic artefact. Notably, when version formats of the same semantic artefact were not directly comparable, versions were also considered misaligned (*e.g.* instances in which version number and date were used in different catalogues as versions of the same resource).

Lastly, images were plotted using *ggplot2* v.3.4.2⁵⁶ and the packages *treemapify* v.2.5⁵⁷, and *ggvenn* v.0.1.10⁵⁸ were used respectively for the treemaps and the Venn diagram. The waffle chart was generated using *waffle* v.1.0.1⁵⁹ and *hrbrthemes* v.0.8.0⁶⁰.

Data availability

The list of semantic artefacts and associated metadata is available on OSF¹⁴ at: <https://doi.org/10.17605/osf.io/axy3s>.

Code availability

The R code used for data analysis and visualisation is available within the project “Collection of semantic artefacts in the environmental domains” of the OSF repository¹⁴ at: <https://doi.org/10.17605/osf.io/axy3s>.

Received: 11 January 2024; Accepted: 22 July 2024;

Published online: 27 September 2024

References

1. Michener, W. K. & Jones, M. B. Ecoinformatics: supporting ecology as a data-intensive science. *Trends in Ecology & Evolution* **27**, 85–93 (2012).
2. David, R. *et al.* Converging on a semantic interoperability framework for the european data space for science, research and innovation (EOSC). *2nd Workshop on Ontologies for FAIR and FAIR Ontologies (Onto4FAIR), 9th Joint Ontology Workshops (JOWO 2023)*. Zenodo <https://doi.org/10.5281/zenodo.8042997> (2023).
3. Wilkinson, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* **3**, 1–9 (2016).
4. de Visser, C. *et al.* Ten quick tips for building FAIR workflows. *PLoS Computational Biology* **19**, e1011369 (2023).
5. Barker, M. *et al.* Introducing the FAIR Principles for research software. *Scientific Data* **9**, 622 (2022).
6. Cox, S. J., Gonzalez-Beltran, A. N., Magagna, B. & Marinescu, M.-C. Ten simple rules for making a vocabulary FAIR. *PLoS Computational Biology* **17**, e1009041 (2021).
7. Le Franc, Y. *et al.* D2.2 FAIR Semantics: First recommendations. Zenodo <https://doi.org/10.5281/zenodo.5361930> (2021).
8. Corcho, O. *et al.* A maturity model for catalogues of semantic artefacts. *Scientific Data* **11**, 479 (2024).
9. Zeng, M. L. Knowledge Organization Systems (KOS). *Knowledge Organization* **35**, 160–182 (2008).
10. Jonquet, C. *et al.* Ontology repositories and semantic artefact catalogues with the OntoPortal technology. *ISWC 2023 - 22nd International Semantic Web Conference. HAL Open Science* <https://hal.science/hal-04088537> (2023).
11. Schwardmann, U. Digital Objects - FAIR Digital Objects: Which services are required? *Data Science Journal* **19** (2020).
12. Balbi, S. *et al.* The global environmental agenda urgently needs a semantic web of knowledge. *Environmental Evidence* **11** (2022).

13. Amdouni, E. & Jonquet, C. FAIR or FAIRer? An integrated quantitative FAIRness assessment grid for semantic resources and ontologies. *15th International Conference on Metadata and Semantics Research (MTSR 2021)*. HAL Open Science <https://hal.umontpellier.fr/lirmm-03208544v3> (2021).
14. Pulieri, M. *et al.* Collection of semantic artefacts in the environmental domain. OSF <https://doi.org/10.17605/osf.io/axy3s> (2024).
15. Preston-Werner, T. Semantic Versioning 2.0.0. *Semantic versioning* <https://semver.org/spec/v2.0.0.html> (2013).
16. Drury, B., Fernandes, R., Moura, M.-F. & de Andrade Lopes, A. A survey of semantic web technology for agriculture. *Information Processing in Agriculture* **6**, 487–501 (2019).
17. Wilkinson, S. R. *et al.* F*** workflows: when parts of FAIR are missing. In *Proceedings of the 2022 IEEE 18th International Conference on e-Science (e-Science)* (2022).
18. Berners-Lee, T. & Hendler, J. Publishing on the semantic web. *Nature* **410**, 1023–1024 (2001).
19. Jacobsen, A. *et al.* FAIR principles: Interpretations and implementation considerations. *Data Intelligence* **2**, 10–29 (2020).
20. Amdouni, E., Bouazzouni, S. & Jonquet, C. O'FAIRE makes you an offer: metadata-based automatic FAIRness assessment for ontologies and semantic resources. *International Journal of Metadata, Semantics and Ontologies* **16**, 16–46 (2022).
21. Amdouni, E., Bouazzouni, S. & Jonquet, C. O'FAIRE: Ontology FAIRness Evaluator in the AgroPortal semantic resource repository. In *Proceedings of the European Semantic Web Conference (ESWC 2022)* (2022).
22. Tarallo, A., Pulieri, M., Ramezani, P. & Rosati, I. Advancements in EcoPortal: Enhancing functionalities for the ecological domain semantic artefacts repository. *FAIR Connect* **2**, 1–7 (2024).
23. Dutta, B., Nandini, D. & Shahi, G. K. MOD: metadata for ontology description and publication. *Dublin Core Papers* (2015).
24. Xu, F. *et al.* Features of a FAIR vocabulary. *Journal of Biomedical Semantics* **14**, 6 (2023).
25. Nyberg Åkerström, W. *et al.* Developing and implementing the semantic interoperability recommendations of the EOSC Interoperability Framework. *Zenodo* <https://doi.org/10.5281/zenodo.10843882> (2024).
26. Corcho, O. *et al.* EOSC interoperability framework: report from the EOSC Executive Board Working Groups FAIR and Architecture. *European Commission Publication Office* <https://doi.org/10.2777/620649> (2021).
27. Splendiani, A., Gray, A. J. G., Evelo, C., Willighagen, E. & Rocca-Serra, P. The FAIR Cookbook: FCB006. *FAIR Cookbook* <https://faircookbook.elixir-europe.org/content/recipes/findability/identifiers.html#fcb-find-identifiers> (2022).
28. Gleim, L. C. & Decker, S. Timestamped URLs as persistent identifiers. In *Managing the Evolution and Preservation of the Data Web Workshop (MEPDaW'20)*, *International Semantic Web Conference (ISWC)* (2020).
29. Bonino da Silva Santos, L. O., Prince Sales, T., Fonseca, C. & Guizzardi, G. Towards a conceptual model for the FAIR Digital Object Framework (IOS Press, 2023).
30. Le Franc, Y. *et al.* D2.2 FAIR Semantics: First recommendations. *Zenodo* <https://doi.org/10.5281/zenodo.3707985> (2020).
31. Sansone, S.-A. *et al.* FAIRsharing as a community approach to standards, repositories and policies. *Nature biotechnology* **37**, 358–367 (2019).
32. Jackson, R. *et al.* OBO Foundry in 2021: operationalizing open data principles to evaluate ontologies. *Database* **2021**, 2021, baab069 (2021).
33. Mazzocchi, F. Knowledge Organization System (KOS): an introductory critical account. *Knowledge Organization* **45**, 54–78 (2018).
34. Hitzler, P., Krötzsch, M., Parsia, B., Patel-Schneider, P. F. & Rudolph, S. OWL 2 web ontology language primer (second edition). *W3C recommendation* <http://www.w3.org/TR/owl-primer> (2012).
35. Hodge, G. M. Systems Of Knowledge Organization For Digital Libraries: Beyond Traditional Authority Files (Digital Library Federation, Council on Library and Information Resources, 2000).
36. Soiland-Reyes, S., Goble, C. & Groth, P. Evaluating FAIR Digital Object and Linked Data as a distributed object system. *Peer J Computer Science* **10**, e1781 (2024).
37. Zeng, M. L. & Mayr, P. Knowledge Organization Systems (KOS) in the Semantic Web: a multi-dimensional review. *International Journal on Digital Libraries* **20**, 209–230 (2019).
38. Rodríguez-Revello, J., Barba-González, C., Rybinski, M. & Navas-Delgado, I. KNIT: Ontology reusability through knowledge graph exploration. *Expert Systems with Applications* **228**, 120239 (2023).
39. David, R. *et al.* "Be sustainable": EOSC-Life recommendations for implementation of FAIR principles in life science data handling. *The EMBO Journal* **42**, e115008 (2023).
40. Fernández-López, M., Poveda-Villalón, M., Suárez-Figueroa, M. C. & Gómez-Pérez, A. Why are ontologies not reused across the same domain? *Journal of Web Semantics* **57**, 100492 (2019).
41. Jonquet, C., Dutta, B., da Silva Santos, L. O. B., Pergl, R. & Le Franc, Y. Common minimum metadata for FAIR semantic artefacts. *Onto4FAIR 2023 - 2nd Workshop on Ontologies for FAIR and FAIR Ontologies*. HAL Open Science <https://hal.science/hal-04106533> (2023).
42. Jonquet, C. & Poveda-Villalón, M. About versioning ontologies or any digital objects with clear semantics. *DaMaLOS 2023 - 3rd Workshop on Metadata and Research (objects) Management for Linked Open Science*. HAL Open Science <https://hal.science/hal-04094847> (2023).
43. Garijo, D. & Poveda-Villalón, M. Applications and practices in ontology design, extraction, and reasoning, vol. 49, chap. Best practices for implementing FAIR vocabularies and ontologies on the web (IOS Press Amsterdam, 2020).
44. Ramezani, P., Grau, N., Jonquet, C. & Fiore, N. M4.1 Semantic artefact governance models: example of community practices. *Zenodo* <https://doi.org/10.5281/zenodo.10287011> (2023).
45. Parland-von Essen, J. & Dillo, I. FAIR-IMPACT. *First International Conference on FAIR Digital Objects* <https://doi.org/10.3897/rio.8.e96144> (2022).
46. Garijo, D. *et al.* M5.3 Semantic artefact assessment methodology. *Zenodo* <https://doi.org/10.5281/zenodo.8305173> (2023).
47. Matentzoglou, N. *et al.* A Simple Standard for Sharing Ontological Mappings (SSOM). *Database* **2022**, baac035 (2022).
48. Zeng, M. L. KOS types vocabulary. *Networked Knowledge Organization Systems Dublin Core Application Profile (NKOS AP)* <https://nkos.dublincore.org/nkos-type.html> (2019).
49. R Core Team. *R: A Language and Environment for Statistical Computing*. <https://www.R-project.org/> (2021).
50. Wickham, H., François, R., Henry, L. & Müller, K. *dplyr: A Grammar of Data Manipulation*. <https://dplyr.tidyverse.org> (2021).
51. Wickham, H. *et al.* Welcome to the Tidyverse. *Journal of Open Source Software* **4**, 1686 (2019).
52. Wickham, H. & Girlich, M. *tidyr: Tidy Messy Data*. <https://CRAN.R-project.org/package=tidyr> (2022).
53. Wickham, H. Reshaping data with the reshape package. *Journal of Statistical Software* **21**, 1–20 (2007).
54. Feinerer, I. & Hornik, K. *tm: Text Mining Package*. <https://CRAN.R-project.org/package=tm> (2023).
55. Fellows, I. *Package 'wordcloud'*. <https://cran.r-project.org/package=wordcloud> (2018).
56. Wickham, H. *ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://cran.r-project.org/package=ggplot2> (2024).
57. Wilkins, D. *treemapify: Draw treemaps in 'ggplot2'*. <https://CRAN.R-project.org/package=treemapify> (2021).
58. Yan, L. *ggvenn: Draw Venn Diagram by 'ggplot2'*. <https://cran.r-project.org/package=ggvenn> (2021).
59. Rudis, B., Gandy, D., Breza, A., Jütte, M. & Campbel, P. *waffle: Create Waffle Chart Visualizations*. <https://cran.r-project.org/package=waffle> (2019).
60. Rudis, B. *et al.* *hrbrthemes: Additional Themes, Theme Components and Utilities for 'ggplot2'*. <https://cran.r-project.org/web/packages/hrbrthemes/index.html> (2020).

Acknowledgements

The study has been funded by EU - Next Generation EU Mission 4 “Education and Research” - Component 2: “From research to business” - Investment 3.1: “Fund for the realisation of an integrated system of research and innovation infrastructures” - Project IR0000032 - ITINERIS - Italian Integrated Environmental Research Infrastructures System - CUP B53C22002150006. The authors acknowledge the Research Infrastructures participating in the ITINERIS project with their Italian nodes: ACTRIS, ANAEE, ATLaS, CeTRA, DANUBIUS, DISSCO, e-LTER, ECORD, EMPHASIS, EMSO, EUFAR, Euro-Argo, EuroFleets, Geoscience, IBISBA, ICOS, JERICO, LIFEWATCH, LNS, N/R Laura Bassi, SIOS, SMINO. Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Commission. Neither the European Union nor European Commission can be held responsible for them. M.P., S.M. were supported by LifeWatch Italy through the project “LifeWatchPLUS”-CIR01_00028. The publication was funded by the CNR project USEit - Using operational synergies for the study and integrated management of invasive alien species in Italy. The Authors are extremely grateful to the Editor and the two Reviewers for the comments received and which highly improved the readability and the quality of the presented study.

Author contributions

I.R. conceived the study, C.D.M., M.P., A.N.M., E.N., A.T., J.T. and S.M. performed data collection, M.P. and D.R. analysed the results, C.D.M. wrote the first full draft of the manuscript and all authors critically reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to C.D. or M.P.

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) 2024