External Scientific Report

APPROVED: 20 NOVEMBER 2024
doi: 10.2903/sp.efsa.2024.EN-9119

Governance Model for Ontologies

(final version – November 2023)

Yannis Tzitzikas, Yannis Marketakis, Pavlos Fafalios, Athina Kritsotaki, Anastasia Axaridou, Michalis Mountantonakis, Chryssoula Bekiari, Maria Theodoridou

Institute of Computer Science, Foundation for Research and Technology – Hellas (FORTH), Greece

Abstract

This report proposes a governance model for managing the activities involving ontologies at EFSA. It contains recommendations about (a) the scoping of the ontology development activities at EFSA (purpose and subject matter), (b) external and internal consistency and interoperability, (c) maintenance and evolution, (d) access policies, (e) IT considerations (aligned with Deliverable D4), (f) outsourcing strategy, (g) training needs related to ontologies. The proposed model is aligned with the existing governance activities of EFSA, such as Rebuild DF. The document is complemented with the adopted governance model of other ontologies.

Key words: ontology governance, ontology interoperability, ontology maintenance and evolution, metadata management

Question number: EFSA-Q-2024-00653
Correspondence: KNOW@efsa.europa.eu

Disclaimer: The present document has been produced and adopted by the bodies identified above as author(s). This task has been carried out exclusively by the author(s) in the context of a contract between the European Food Safety Authority and the author(s), awarded following a tender procedure. The present document is published complying with the transparency principle to which the Authority is subject. It may not be considered as an output adopted by the Authority. The European Food Safety Authority reserves its rights, view, and position as regards the issues addressed and the conclusions reached in the present document, without prejudice to the rights of the authors.

Suggested citation: Tzitzikas Y., Marketakis Y., Fafalios P., Kritsotaki A., Axaridou A., Mountantonakis M., Bekiari C., and Theodoridou M., 2024. Governance Model for Ontologies. EFSA supporting publication 2024:EN-9119. 35 pp. doi:10.2903/sp.efsa.2024.EN-9119

ISSN: 2397-8325

Table of contents

1 Introduction

The scope of this document is addressing the main aspects for a governance model for managing all the activities involving ontologies at EFSA. Such activities include their maintenance, their consistency in terms of interoperability with other ontologies, their evolution, and their access policy. Because of its context, readers of this document are advised to read the corresponding deliverable for ontologies (D2. Ontologies and Case Studies), that provides the background information about ontologies.

The outline of this document is the following:

- **Section 1** introduces the scope and outlines the rest of the document.
- **Section 2** provides background information.
- **Section 3** discusses the methodology followed for producing this deliverable.
- **Section 4** describes the various aspects of the governance model.
- **Section 5** details the governance model of one ontology that has become ISO standard and describes in brief the governance model of other popular ontologies.
- **Section 6** summarizes and concludes the report.
- Detailed recommendations for particular topics related to governance, are given in the appendices of this report, in particular:
  - **Annex A** discusses best practices for documenting an ontology
  - **Annex B** describes various policies for designing the URIs of an ontology
  - **Annex C** provides information for making the URIs dereferenceable.

1.1 Background and terms of reference as provided by the requestor

This contract was awarded by EFSA to:

Contractor: Network Research Belgium (NRB) S.A

Contract title: Ontology Roadmapping and Case Study Implementation

Contract number: SC01 implementing FWC OC/EFSA/DATA/2021/01

2 Background

2.1 Objectives

The adoption of ontologies by EFSA presupposes a model for their management and this includes issues of quality, maintenance, rights, strategy and consistency with the organizational structure and work practices of ESFA. The overall aim is to define the guidelines for producing and maintaining ontologies that are interoperable, are developed using common tools and standards, adopt best practices, and identify the required skills for making them sustainable. The readers of this document are advised to refer to the deliverables D2. Ontologies and Case Studies [1] and D4. Blueprint for Ontology Management [2] since many of the aspects of the current document are described extensively in these documents.

2.2 The notion of "Ontology Governance"

Ontology governance refers to the process of managing and maintaining an ontology. Ontology governance involves defining the policies, standards, and procedures for creating, maintaining, and using the ontology. This includes tasks such as establishing a governance structure, defining roles and responsibilities, setting guidelines for ontology development and maintenance, ensuring data quality, and managing changes to the ontology. Effective ontology governance is essential for ensuring the accuracy, consistency, and usability of the ontology, which in turn can improve data integration, interoperability, and decision-making.

Related publicly available webinars include:

- "GOMO - Governance Operational Model for Ontologies" [3], Alexander Garcia, https://www.youtube.com/watch?v=CYt05vvEsS8
- "From Ontology Engineering to Ontology Governance Models" Dr. Oscar Corcho (WEBIST 2021), https://www.youtube.com/watch?v=xZPZmAZ7UYA

Moreover, there have been proposed generic Ontology Management Architectures [4] that provide some generic steps that include ontology mapping, ontology learning, reasoning with inconsistent ontologies, and query answering. Furthermore, a recent survey [5] focuses on identifying the most relevant elements in the development of an ontology-based solution and how these solutions are being deployed in industry. Finally, an ontology for Data Governance is available [6], for enabling the modelling of the sources, their metadata, but also of the coordination between people and knowledge domains that are included.

3 Methodology for Producing this Deliverable

To produce this deliverable, we used:

- the material and exchanges (emails, zoom meetings) related to deliverables D2 and D4 of this project,
- the established practices for ontology governance,
- our experience in the governance of the ISO 21127 CIDOC CRM ontology,
- the exchanges during virtual meetings organized for the preparation of this particular deliverable, as shown in the following table, and

- the feedback received by EFSA on earlier versions of this deliverable

The following table shows the meetings carried out with various people from EFSA and outside EFSA, while producing this deliverable.

Date	Participants	Topics
Feb 10, 2023	Yannick Spill (EFSA), Allessandra Carinngella (EFSA)	About the expectations from this deliverable and clarifications about the way EFSA functions.
May 15, 2023	Brian Alper (HEvKA-Health Evidence Knowledge Accelerator)	Discussion about ontology governance at HEvKA.
Nov 8, 2023	Davide Gibin (EFSA)	EFSA Catalogue Management policy[1]
Nov 14, 2023	Luca Pasinato (EFSA), Giorgio Centrini (EFSA)	Rebuild DF
Nov 16, 2023	Fabrizio Abbinante (EFSA)	FAIR principles in EFSA[1]

Table 1. Dedicated meetings for this deliverable

4 A governance model for ontologies

The proposed governance model is described subsequently. In the following subsections various aspects are covered, such as access policy, maintenance policy, consistency, compliance, etc. based on established and existing procedures carried out at EFSA. The definition of the acronyms used are given in the "Glossary and Abbreviations" section at the end of this document.

4.1 Scope

Purpose: The wide adoption of ontologies is the result of their competence in describing semantics in a consistent and non-ambiguous manner, enabling the re-use of resources, inferring implicit knowledge and their use for supporting data integration. Taking into consideration EFSA current activities, these competencies are enhanced with further improvements in semantic interoperability, improved data management, facilitation of Artificial Intelligence (AI) tasks, with more and better training data, combination of Machine Learning (ML) with Knowledge Representation (KR), and improvement and speedup of various EFSA services (e.g., automating the process of annotations, classification tasks, etc.).

Subject matter: The Ontology development activities involve several actors and use cases. The main actors are the following:

- **Conceptual Modeller** (creates/extends the conceptual model)
    - This actor is responsible for analysing and designing the entities and their relationships in terms of a conceptual model, that will eventually lead to the creation or extension of an ontology and is important to either be specialized for

[1] Carried out through e-mail exchanges

the domain that will be modelled, or collaborate with domain experts. Essentially the conceptual modeller is responsible for representing the abstract ideas and concepts of a specific domain into concrete entities of a particular ontological model. There are various techniques and tools to use in order to produce the conceptual model, such as UML diagrams, mind maps, and other high-level view-oriented representations. Another activity for this actor is to perform the requirements analysis by collaborating with the stakeholders of the domain in order to understand their needs and create a set of competency queries that will be answerable by the ontology.

- **Ontology Developer** (implements the ontology using tools)
    - This actor is responsible for implementing the defined ontology according to the conceptual model, and needs to have expertise in the existing ontology engineering tools. This actor uses as input the conceptual model of an ontology (created by the conceptual modeller) and produces the formal representation of the ontology using well defined languages, such as RDFS or OWL. The result of this process will be the description of an ontology in a structured and standardized form that can be exchanged and further disseminated. The implementation of the ontology can include additional activities, such as the creation of visualizations, documentation, etc.
- **Ontology User** (uses the ontology)
    - This actor can be any user that uses the ontology, e.g., for a specific need, for performing an analysis, etc. Concerning users that are experts in ontologies and query languages such as SPARQL, they can write such queries for retrieving the information expressed through the ontology. On the contrary, for the non-familiar users, interactive information access services should be created (e.g., query builders and browsing systems).
- **Ontology Custodian** (responsible for the evolution of ontology and its documentation, publishing)
    - This actor is responsible for the ownership and maintenance of the ontology, including its evolution, documentation publishing, organising of activities, communicating with external users, etc. This actor ensures the availability of the ontology when it is requested, the collection of requests for resolving issues and extensions to the ontology, the quality of the ontology, especially as it evolves, as well as its dissemination.

Figure 1 depicts the involved actors and their involvement.

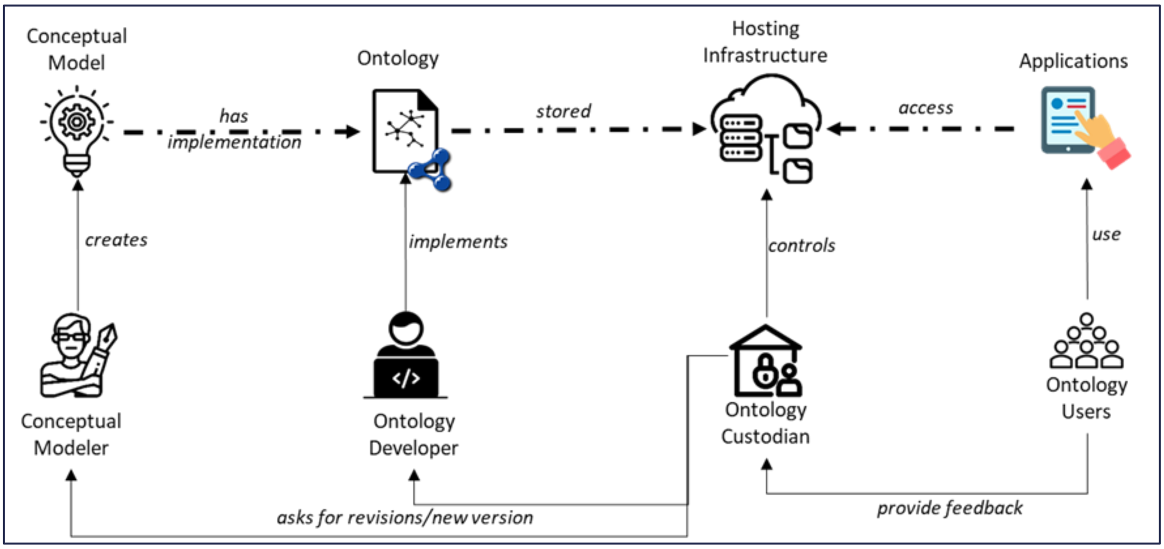

Figure 1. The main actors involved in ontology development activities

4.2 External Consistency

As regards **external consistency**, the ontologies developed by EFSA should be interoperable with existing ontologies. We can identify the following aspects of external consistency:

A) Format Consistency. Compliance with World Wide Web Consortium (W3C) standards (RDF, OWL, etc.), as described in Deliverable D4.

B) Conceptual Consistency. With the top-level classes and relations of other (related) ontologies and domains to facilitate mappings, transformations and query interoperability. The relevant ontologies are described in Deliverable D2.

C) Publishing Consistency. All best practices for publishing an ontology should be followed, i.e., Linked Data [14] and FAIR principles **Error! Reference source not found.**. They can nclude documentation (e.g. scope notes for all concepts and relations), design of simple and consistent URIs, dereferenceable URIs, publish of the ontologies in related catalogues (e.g. at Linked Open Vocabularies), versioning, preparation and publishing of metadata about the ontology (e.g. using the vocabulary VoID[2]). Section 2 of D4 [2] contains a detailed list of tools for supporting this requirement. More detailed descriptions of the documentation of ontologies can be found in the appendices of this document. More specifically:

- **Annex A** discusses best practices for documenting an ontology
- **Annex B** describes various policies for designing the URIs of an ontology
- **Annex C** provides information for making the URIs dereferenceable.

4.3 Internal Consistency

Internal consistency refers to the harmonization and reliability of data across the organization, ensuring uniformity, accuracy, and coherence in data definitions, formats, and quality. This reliability empowers better decision-making, supports streamlined operations, and strengthens

[2] http://www.w3.org/TR/void

trust in data, serving as a cornerstone for effective data-driven strategies within the organization. It is important therefore that, that the ontological resources that are used within EFSA are compliant with existing management policies. For this reason, at first we describe the related background that is either adopted by EFSA (i.e. SSD2, DAMA2, Rebuild DF), or is of interest to EFSA (i.e. OBO), and then make suggestions for internal consistency.

4.3.1 Standard Sample Description ver. 2.0 (SSD2)

The Standard Sample Description ver. 2.0 (SSD2) [16] provides a specification for describing food and feed samples, including detailed characteristics such as its origin, production method, processing, packaging and any relevant analytical parameter. It specifies the data elements and the data structure to describe several types of samples and results coming from laboratory analytical measurements. The logical model of SSD2 is composed of: (a) data elements definitions and structure, (b) controlled terminologies, and (c) business rules to ensure the validity of the information supplied. Its purpose is to ensure consistency and comparability of data gathered from various sources. The main scope of SSD2 is to support the data collection and transmission of sample data and the results of analytical measurement for data collection in the chemical and biological domain. Although it has been primarily designed for the collection of analytical results submitted to EFSA data collections, it can be applied also in different domains, by configuring and customizing it accordingly.

4.3.2 Rebuild DF

In the context of re-engineering and sharing data collection, data storage, data management and data analysis solutions in collaboration with other EU Agencies, EFSA focuses on connectivity, interoperability and co-creation of data and data analysis and implementing interfaces to allow automatic transfer of EFSA metadata to the European Union Open Data Portal[3] and IPCHEM Portal[4], and publication in the Knowledge Junction of public datasets collected by EFSA and contained in the EFSA Scientific Data Warehouse. The Data Management and Data Analysis (DAMA) 2.0 activity aims to improve data interoperability and facilitate data exchange. DAMA2 is EFSA's cluster of projects, started in 2021, aiming to continue the implementation of a strategic data management infrastructure and functionality, enabling digitalisation of processes used to deliver risk assessment and risk communication within EFSA's remit and in collaboration with EU agencies within the ENVI group. The DAMA2 is based on an Event sourced approach for high scale production requirements and real time data (automatically flowing data). Since DAMA2 was a temporary cluster of projects, by the time of writing this deliverable, DAMA2 cluster is considered obsolete. Following DAMA2 cluster, EFSA has adopted projects under specific programmes (e.g. Data & Evidende Programme – DEEP). One of the main related projects under DEEP is is Rebuilt DF.

The Rebuild DF project aims to address data management challenges within EFSA by building a coherent and scalable integrated architecture. It aims to create a better system for collecting and analysing data, managing terminology, ensuring data security, and maintaining continuity of services. The project involves collaborating with member states and EU partners to rebuild

[3] https://data.europa.eu/en
[4] https://ipchem.jrc.ec.europa.eu/

the Data Collection Framework and develop new tools. The project will also follow FAIR data principles when developing APIs.

4.3.3 OBO Foundry

OBO Foundry[5] develops a family of interoperable ontologies that are both logically well-formed and scientifically accurate (based on a set of established principles for ontology development), for the biomedical and biological domain. It uses a sourceforge repository to maintain ontologies and ontology versions. It aims at creating a collection of well-documented and well-defined biomedical ontologies that are designed to work with one another. The Foundry initiative serves to align ontology development on the basis of an upper-level ontology, the BFO ontology (Basic Formal Ontology)[6]. BFO narrowly focuses on the task of providing a genuine upper ontology which can be used in support of domain ontologies developed for scientific research, as for example in biomedicine within the framework of the OBO Foundry. Thus, BFO does not contain physical, chemical, biological or other terms, which would properly fall within the special sciences domains.

The OBO Foundry effort aims to integrate biomedical results and carry out analysis in bioinformatics. Currently, there are more than a hundred ontologies that follow the OBO Foundry principles.

4.3.4 Suggestions

The core objective of existing projects within EFSA (e.g. Rebuild DF) is to provide a scalable integrated data architecture in order to improve the collection and analysis of data. The adoption of ontologies could further enhance the activities towards this. To ensure the consistency of data within EFSA, the adopted ontologies should be clear, non-ambiguous, logically compatible and non-conflicting with other ontologies. The same applies for the relations and the properties of the ontologies. Overall, the adopted ontologies should accurately represent the food safety domain and should reflect the consensus and understanding of experts in the field, avoiding contradictions or inaccuracies with established knowledge. Finally, they should not interfere with policies, and best practices that have been clearly identified by other data-relevant-related projects (e.g. Rebuild DF) that address different aspects or views, such as ensuring data security, service availability and scalability, etc.

4.4 Maintenance policy

Ontology Evolution. As described already, ontologies are used to formally represent the knowledge of a particular domain. Since our knowledge is not static but instead changes over time, these changes should be reflected in the ontology itself. Such occasions may be corrections in the conceptualisation, adapting the ontology to changing facts in the real world in order to catch up with the current reality, etc. It is therefore mandatory to support a versioning scheme for ontologies (see "Versioning" paragraph below), as well as an evolution lifecycle.

Similarly, to document versions, ontologies adopt a versioning scheme in order to facilitate collaborative development, better monitoring of the evolution of the ontology, explicit connection

[5] https://obofoundry.org/
[6] https://basic-formal-ontology.org/

with the documentation, and eventually mutual agreement on the ontology coverage. Ontology versions as a reference when collecting new requirements for expanding or updating an ontology, as well as for the identification of issues and their resolution.

The evolution of ontologies should be carried out using evolution phases, in a similar manner as with the software evolution lifecycle. During each phase, the collected requirements are prioritized and addressed accordingly. The requirements concern updates in the ontology either for its expansion or for the resolution of concrete issues. The duration of each phase should ideally be predefined. At the end of the phase, and once the updates are validated, the corresponding scope notes (i.e. the detailed documentation of the ontology) are produced and all the relevant material is made publicly available. The latest released version becomes the default version of the ontology.

Maintenance Policy. The maintenance policy should be clearly defined and agreed. This includes many different aspects about the lifecycle of an ontology, such as: how often a new version will be released, how requirements and feedback is collected, how extraordinary versions or patches are communicated, etc. The reader is advised to see the section related to the CIDOC CRM Governance Model.

Custodians. A group of different stakeholders with complementary expertise and knowledge should be in charge of the ontology maintenance. The roles of the group members are clearly defined. They are in charge of receiving and assessing the feedback and new requirements, taking the required decisions and defining the strategy for the evolution of the ontology.

Relation with Catalogue Management. Being digital resources themselves, ontologies should be based on terminologies content and be aligned with the update process and governance of terminologies. Therefore, they should be aligned with the catalogue management policies. This means that they will be described with respect to the catalogue semantics, their access will be managed by the catalogue, and their relations and dependencies with other catalogue resources will be properly and accurately described.

Versioning. The version of the ontology can adopt any desired scheme. The most commonly used versioning scheme, which is also adopted for software components, is the major.minor.patch semantic versioning scheme[7] [7]. It relies on three digits concatenated with dots indicating the version number for each part. More specifically their semantics are:

- major: this number changes (e.g. incremented) when changes to the delivered artefact are made that create incompatibilities with previous versions.
- minor: this number changes when changes to the delivered artefact are made that are backwards compatible.
- patch: this number changes when changes to the delivered artefact are made in order to resolve issues and bugs.

The version tag of an ontology it is recommended to be applied in both, the digital container of the ontology (e.g. as a label in the filename of the related file), as well as in the ontology itself

[7] https://semver.org/

(with the use of proper semantic annotation properties), so that it will always follow the ontology when ingested in a semantic repository with other data.

The catalogue maintenance policy that currently exists at EFSA is shown in Figure 2[8]. Everything starts from a new request for an addition or change in the catalogue. The responsible requestor adds their request by filling the appropriate form which is available in the ticketing system. The generation of the tickets informs the Duty Officer to validate and prioritize the ticket, which is then assigned to the Data Steward for validation and assessment. Finally, it is being dispatched by the Catalogue Manager, who is responsible for updating the catalogue after possibly iterating with the Data Steward to ask for clarifications about the new request.

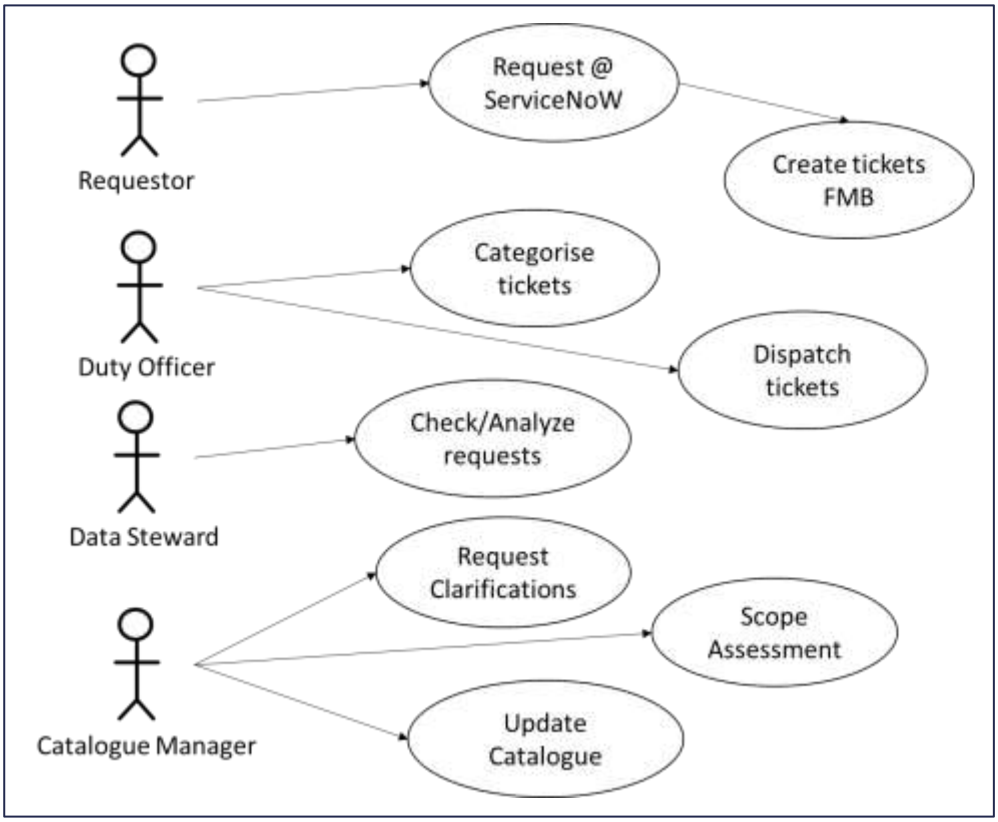

Figure 2. Use case diagram for the maintenance of EFSA catalogues and relevant actors

As regards ontologies, the key users might differ. Based on the users defined in Figure 1, requests are derived from a community of users working with the ontology and requesting for an addition or update (i.e. so they are the requestors). Then the request should be handled by the conceptual modeller, who is responsible for updating the ontology model accordingly, and then by the ontology developer who is responsible for implementing the updates. Finally, the ontology custodians will integrate the update and disseminate it to the ontology users. The following use case diagram demonstrates the maintenance policy for ontologies.

[8] Based on the internal Work instruction document (WIN_SOP008/10) Date: 2021.07.01

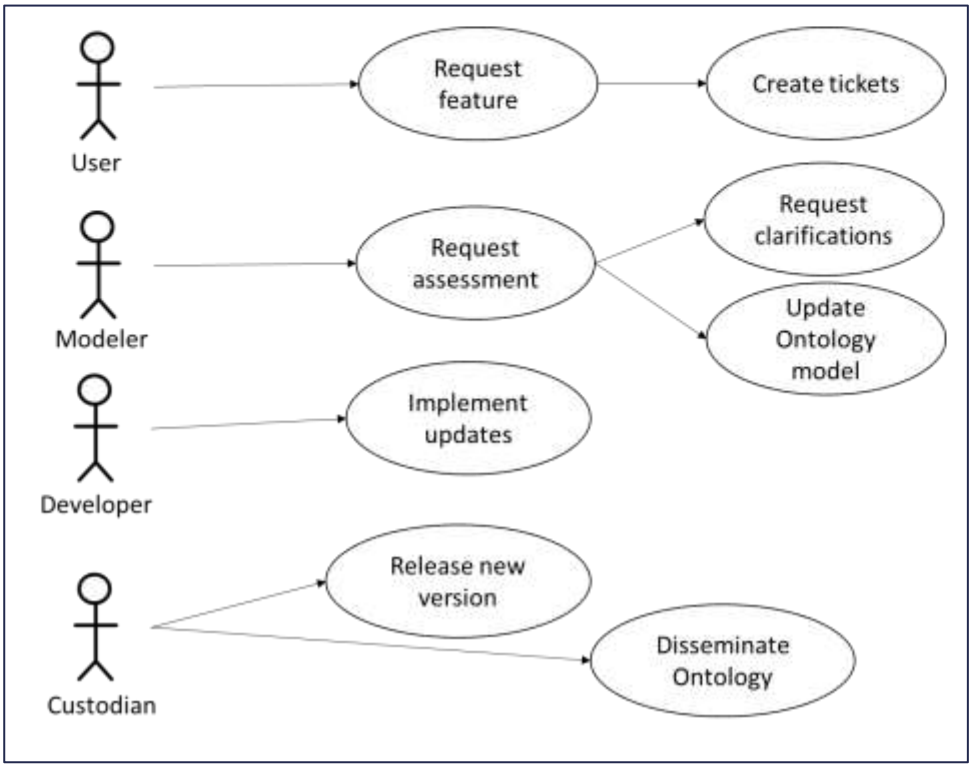

Figure 3. Use case diagram for the maintenance of Ontologies and relevant users

We should note here that all the tools and services for ontology creation, maintenance, migration and querying, including links to software systems and APIs are described in deliverable D4 [2]. EFSA uses Microsoft tools as well as open source tools. Commercial tools (e.g. SAP, Distiller) are decided by the IT council of EFSA.

4.5 Access Policy

In general, licenses are used to establish the legal permission for using a product. They provide the legal agreement that outlines the terms and conditions under which a product can be used, distributed, or modified. In the same manner, an ontology should be licensed as well. Choosing the best license is not a straightforward decision, and it should be decided based on the particular case or domain. A general comment is that since the main purpose of ontologies is to be publicly shared and distributed, in order to foster community involvement, as well as be interoperable, they should have a license that avoids conflicts in case they have to be extended or combined with others. The best way to achieve this, is to release an ontology without any restrictions (i.e. using the "CC0 - No Rights Reserved" license[9]). However, there might also be reasons for retaining some rights of attribution, since other rights can be used according to the needs. In general, the choice of the proper license for distributing an ontology should be decided by the creators and maintainers of the ontology, having in mind the intended use and overall purpose of the ontology. This should also be aligned with the licensing policy that EFSA has for distributing

[9] https://creativecommons.org/share-your-work/public-domain/cc0/

their products. In general, EFSA uses open licenses (e.g. CC BY 4.0[10] for publishing datasets on Zenodo).

The most common licenses that are used for distributing ontologies (and beyond) are:

- **Creative Commons License[11]** that is often used for ontologies that are designed to be openly shared and reused by others. The Creative Commons License allows others to use, distribute, and modify the ontology as long as they give appropriate credit to the creator and use the ontology for non-commercial purposes.
- **Open Data Commons Open Database License (ODbL)[12]** that is specifically designed for open data and allows others to use, distribute, and modify the ontology as long as they share any changes they make to the ontology and attribute the original creator.

4.6 Dissemination and Training

As regards training, it depends on the role of each involved actor. In Section 4.1 we identified the different actors that are involved. Below, we can distinguish the following training categories and identify the involved actors:

- conceptual modelling training (for conceptual modellers)
- ontology representation/implementation using W3C standards and tools (for ontology developers)
- ontology publishing and maintenance (for ontology custodians), accessing and querying the ontology for enabling the creation of ontology-based services (for ontology users)

The persons involved have to be trained according to their role. The tasks and tools mentioned in deliverable D4 can be used to identify the skills that will be required. Since EFSA has a Learning Management System (LMS), that platform could be exploited as well for sharing the related tutorials and guides. Moreover, the EFSA YouTube channel[13] can be used as well. For aiding the task of training, and also of publishing new datasets based on a specific ontology, one option is to create an ontology-specific data portal [13]. The rationale for building a portal for datasets expressed with respect to a particular ontology, is that the community which is interested in that ontology has various incentives to publish its datasets in such a portal. Indeed, for the data owners it is important for browsing successful use cases, for enabling the discoverability and reusability of their datasets, and for finding other datasets that could be easily integrated. On the other hand, for the ontology creators, it is important for inspecting the use of the ontology, e.g., for spotting errors and/or for guiding the use and evolution of the ontology, for checking for power-law distributions concerning the usage of the ontology and others. As an example, recently such a portal for the ISO Standard CIDOC-CRM was created [13], which provides the following functionality: i) browsing of all the available CIDOC-CRM datasets by supporting ontology-based statistics and visualisations about each dataset, ii) searching for specific classes

10 https://creativecommons.org/licenses/by/4.0/
11 https://creativecommons.org/
12 https://opendatacommons.org/licenses/odbl/summary/
13 https://www.youtube.com/@EFSA_EU

and properties, iii) discovery of the most frequent (CIDOC-CRM) properties and classes, and iv) measurements regarding the commonalities between pairs of datasets.

Depending on the competence of the group in charge of maintaining and evolving an ontology, it could be beneficial to delegate some activities to external stakeholders that have the required competence and capacity. For example, the custodians of an ontology have perfectly designed and documented the conceptual part of the ontology, but they are missing the technical background in order to publish it, which in turn consists of many different activities/steps (e.g, implementation based on a standard for better exploitation and dissemination in the community, designing namespace and permanent URIs, etc.). In this case, external assistance is required. In general, we could identify the following strategy options:

- Buy/use the required tools and services that were identified in deliverable D4. Blueprint for Ontology Management. In this deliverable, we report the existing tools that are used for carrying out particular activities (e.g. implement an ontology, visualize an ontology), and we provide suggestions about the recommended tools to use for each case (most of them are open source tools that can be used for free).
- Build your own tools and services, for carrying out a particular task. Of course, this means that the team has the required competence to build them.
- Outsource the building of tools to an external party in order to implement them.
- Outsource the services directly to an external party that has the required experience to implement them. An indicative scenario is the definition of a formal top-level ontology for the food domain (that is also one of the proposed case studies in D2); its design, conceptual modelling, implementation and documentation could be completely outsourced to an external stakeholder (i.e. such as FORTH) for delivering this.

4.7 Standards Operating Procedures (SOP) and Work Instructions (WIN)

Governance is a complex activity where several procedures and actors are involved to preserve a consistent and harmonic function of an ecosystem, accomplishing particular objectives. Specific tasks have to be assigned to the actors and finally follow a simple or hierarchical, controlled execution order. Thus, "who" is doing "what", "when", and "where" can be a challenge for project managers and organizations. A popular solution for organising cases like Ontology Governance, is the use of SOP and WIN.

A Standard Operating Procedure (SOP) is a document that formalizes the tasks in an organization, linking them to the people responsible for them and to the resources used in each stage. The document contains the following essential sections:

a. Purpose: The section with the goals, objectives, and issues that the SOP aims to solve.
b. Procedures: The section that describes the operating procedures in detail to properly guide the actors at the implementation level according to standards, regulations, restrictions going through specific instructions. It provides either a step-by-step guidance, with plain or hierarchical structure, or a flowchart, containing conditions for decision making towards the correct path of an execution plan.
c. Scope: A description for the extent and coverage of the SOP, and the involvement of the actors' roles.

d. Responsibilities: The section that defines the involved actors and responsibilities.

Work Instructions extend the described procedures of a SOP in terms of details about how to proceed with the related tasks. They can be separate documents gathering all the steps regarding a specific procedure. Hence, it is very important to persist such information for each task explained in a document. Standardizing the steps of a procedure, a. enables the involved actors to be consistently aware and have a reference guidance of how to work to achieve a result, and b. helps highlight the errors or missing information and keep a base for instruction improvements.

Some procedures that could be described SOP documents and define the steps to accomplish the particular tasks they require, are:

- New ontology development
- Maintenance of an existing ontology
- Issues management (discussions, conflicts, decisions, feedback and updates from domain experts)

Next, we present some distinct tasks related to the ontology governance procedures, that can be described with WINs:

- Ontology review (regarding new/outdated/obsolete concepts and relationships)
- Ontology alignment to other ontologies of the domain (external consistency)
- Ontology alignment to standards such as RDF, OWL, and SPARQL (external consistency)
- Version control procedure to track changes, maintain a history of the ontology, document changes and the reasons behind them
- Ontology publishing to inform the involved users and the community for a new version
- Quality evaluation to check for inconsistencies, redundancies, errors in the ontology (or errors of how the ontology is used by data owners)
- Ontology documentation, including its purpose, scope, and usage guidelines.
- User feedback from ontology users and domain experts, managing suggestions and incorporating relevant improvements.
- Test cases and evaluation of the performance and correctness of the ontology.
- Data integration to ensure that the ontology integrates well with relevant data and existing applications
- Security to protect the ontology from unauthorized access or modification.
- Training and education for users and developers to understand the ontology and its updates.
- Backup and recovery procedures to prevent data loss and facilitate recovery in case of unexpected issues.

In the concluding section of this document, we proposed an overall workflow demonstrating the maintenance of ontologies. This workflow shows the different user roles, the main activities, and their exchange flow. This workflow could be used as a basis for a concrete SOP/WIN tailored for the needs EFSA (i.e. using EFSA communication portals, infrastructure, etc.).

5 Related Examples of Governance Models

Here we describe one successful governance model and a few other examples.

5.1 The Governance Model of the ISO CIDOC CRM Ontology

Here, we provide details about the governance model of the ISO CIDOC-CRM ontology (ISO 21127:2014) [12].

5.1.1 Brief Description

In brief, a special committee has been established (in which more than 40 organizations participate) and the committee members meet 3-4 times per year, they discuss requests for changes and extensions and they produce new releases. There is a website[14] with all information about this ontology and the related activities for using it: ontology versions, mappings, compatible models, use cases, best practices, activities (meetings), issues, community, news etc. A screenshot of the website is shown in Figure 4.

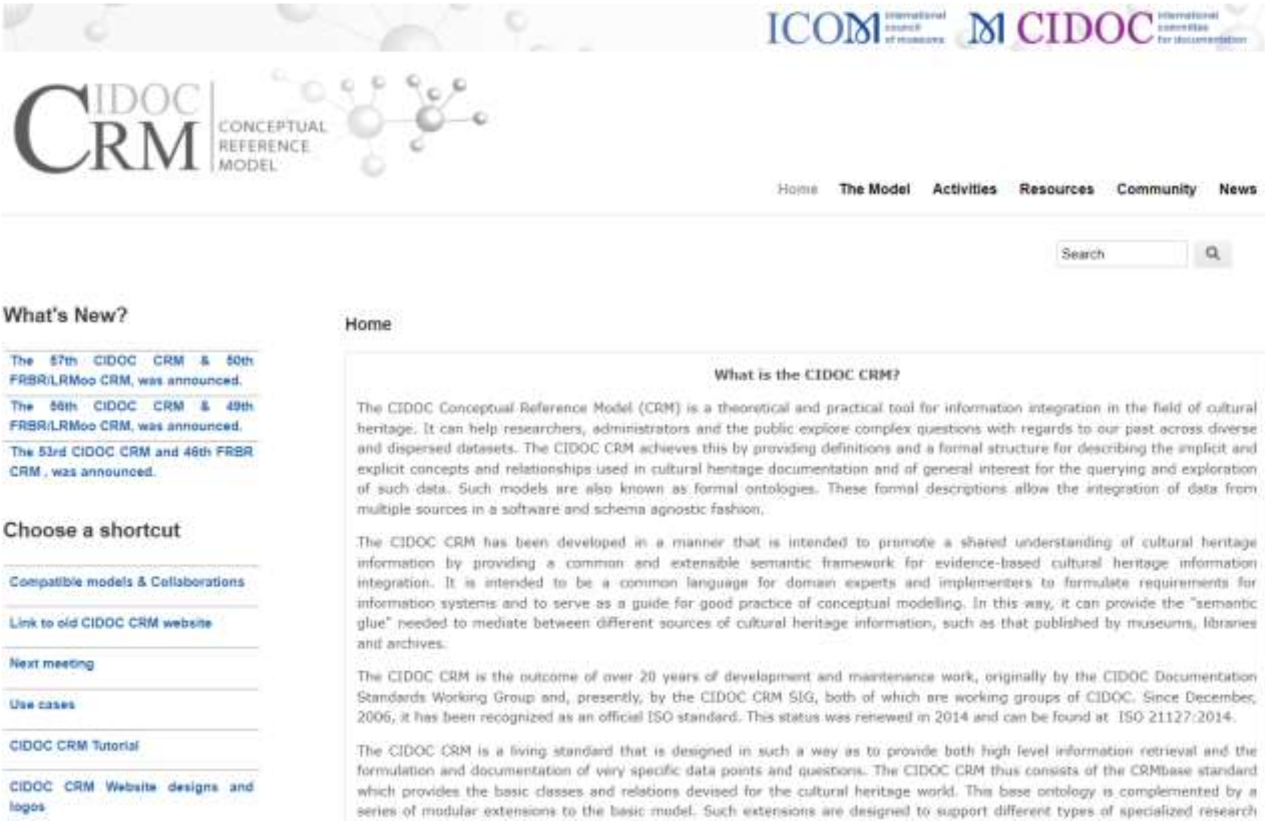

Figure 4. The web site of the ISO CIDOC CRM ontology CIDOC CRM

5.1.2 Detailed Description

Below we describe in detail the governance model of CIDOC CRM. The description covers

- the history of CIDOC CRM,

[14] https://www.cidoc-crm.org/

- the issue management adopted (categories of issues, formulation and templates, issue management workflow, decision process), and
- memberships, voting, meetings and minutes.

5.1.2.1 THE HISTORY OF CIDOC CRM

Over many years, CIDOC and the CIDOC Documentation Standards Working Group (DSWG) have engaged in the creation of a general data model for museums, with a particular focus on information interchange. Until 1994, the product of these activities had been the CIDOC Relational Data Model. In the interim meeting in March 1996 in Crete, the DSWG decided to engage in an object-oriented approach in order to benefit from its expressive power and extensibility for dealing with the necessary diversity and complexity of data structures in the domain. This effort resulted in 1999 in the first complete edition of the "CIDOC Conceptual Reference Model" (CRM), a product of the intensive voluntary work of a variety of contributors. In order to fully exploit the potential of the CRM as a means of enabling information interchange and integration in the museum community and beyond, CIDOC decided in London 1999 to submit the CRM to ISO for standardization. ISO, in contrast to CIDOC, has the procedures and authority to create and declare well-defined, valid editions of international recommendations. The CIDOC CRM was accepted as a working draft by ISO/TC46/SC4 in September 2000. Since 9/12/2006 it has been the official standard ISO 21127:2006. In December 2014 a new version of the standard became available ISO21127:2014 (based on version 5.0.4 of the CIDOC CRM).

In practice, this means that CIDOC makes use of the services of ISO and collaborates with the respective ISO committees to bring the CRM to the suitable final form and to ensure the widest possible agreement with the broader international community. It is thereby understood that CIDOC offers the lead in the collaboration with the ISO committees by inviting and bringing together interested stakeholders of all kinds and from different areas, into an effective working group. This group engages in all appropriate actions for bringing the objectives of the CRM forward - foster application, dissemination and ensure that the final form of the CRM represents the needs of all interest groups.

The work is not thought to end with the completion of the standardization process. The CRM is an extensible model and there is a continuous need to elaborate new areas, as well as keep the model in line with the inevitable changes and progress of conceptualization for information integration. In addition, there are a lot of application issues, including the development of tools, guides for best practice etc., which are benefited from a central forum for harmonization.

5.1.2.2 THE CRM-SIG GROUP

In consideration of this, CIDOC decided in Ottawa, August 2000 to initiate the Conceptual Reference Model Special Interest Group (CRM-SIG) for the maintenance of the model. Martin Doerr from FORTH has been nominated as first chair of this group. It is a working group under the patronage of CIDOC that is not restricted to CIDOC members. The rule is that at least by two thirds, the members of the CMR-SIG also be CIDOC members, otherwise, it would not be able to retain its status as a CIDOC Working Group. Its members represent organizations. The CRM-SIG is empowered to seek funding to carry out and promote its objectives.

More than 40 organizations participate in the CRM-SIG today[15]. Each organization participates through its representative. CRM-SIG meets 3-4 times per year. At the last meeting of each year the members plan the next year's meetings. The members vote by raising their hands for the place and time of the forthcoming meetings. The organizations that host the meetings are posted on the official website of the CIDOC-CRM. During the period of COVID19 the CRM-SIG meetings were virtual only, however, since autumn 2022 they became hybrid with both physical and virtual participation.

The website[11] of CIDOC CRM contains information about this ontology and the related activities for using it: ontology versions, mappings, compatible models, use cases, issues, best practices, activities (meetings), community, news etc. More detailed information about the operation of the CRM-SIG is given in the sequel.

5.1.2.3 CRM Issues Discussion Process for reaching an agreement

The CIDOC CRM grows and evolves through community engagement in the standard and the constant revision of the existing specification as well as the proposal for additions, modifications or deletions to it.

Changes to the CIDOC CRM necessarily happen through a formalized, documented process in order to ensure community engagement with the topics and an awareness of the reasons behind the evolution and changes of the model. The CIDOC CRM changes through proposing, discussing and eventually resolving issues. An issue is a problem posed by a community member for debate and decision. Issues are discussed and resolved either through the CIDOC CRM-SIG mailing list (crm-sig@ics.forth.gr) or during the regular meetings of the CRM-SIG. Highly involved issues tend to be debated during face-to-face meetings. For the sake of enabling all to consider and prepare for such conversations, issues should be raised at least two weeks prior to a meeting.

Issues are categorized into the following types:

- **Model Change**: issues of this type have to do with a modification to the CRM model itself, adding, modifying or deleting classes or properties or changing their scope notes in a meaningful way.
- **Editorial**: issues of this type do not change the meaning of the model but have to do with providing additional material within the specification which clarifies and illuminates the use and functioning of the model (such as redrafting the examples of classes and properties to best reflect their semantics) or with improving the text and editing of typos.
- **Community**: issues of this type are related to the growth and development of the CIDOC CRM community, things like adding members or changing members' representatives, organizing conferences, advertising the groups' work etc.
- **Additional Documents**: issues of this type have to do with the generation of documents that are extra to the CIDOC CRM specification but which may help illustrate the standard or how to use it.

[15] November 2023

Anyone, regardless of whether they are a representative of a CIDOC CRM member institution may raise an issue at any time. An issue is raised either on the CIDOC CRM-SIG mailing list or during the course of a CIDOC CRM-SIG meeting.

5.1.2.4 Issue Formulation and Workflow

In order for issues to be addressed in a managed and reasonable way, they need to be formulated and addressed in such a manner that the problem posed is clearly understood and that the proposed solution, if available, is formally outlined in such a way that CRM-SIG members can vote in a Yes or No manner (in person or virtually) to accept it.

Initial Issue and Proposal Formulation

An issue should be formulated in the following manner:

- **Title**: propose a title of the issue which is somehow an index for the subject to be discussed
- **Background**: write a resume of what prompts the raising of the issue and other contextual background so that other CRM-SIG members understand the origin of the question. The background information should indicate clearly what problem has arisen and provide example material and evidence where possible. Examples of the case are of particular importance.
- **Proposal**: formulate a proposal to solve the issue which can be voted on by other CRM-SIG members. The proposal should be answerable in a yes or no format.
- **Status**: open (by default)
- **Date**: start of the issue
- **Issue Raiser**: person who raise the issue
- **Forum Where Raised**: CRM-SIG mailing list (referring to the corresponding CRM-SIG event)

Whether formulated on the CRM-SIG mailing list or during the course of a CIDOC CRM-SIG meeting, these issues must then be registered in the issue list of the CIDOC CRM-SIG site.

Issue Discussion

Before an issue can be decided, it is discussed by the group. Group members can pose questions, raise objections, offer changes and advice. If the discussion finds no significant problem with the proposal then it can be voted on immediately. Otherwise, a new proposal should be formulated considering the group's discussion.

Proposal Reformulation

It is often the case that proposals for solving an issue will have to go through several permutations before arriving at a state in which they can ultimately be accepted. When an issue has been discussed and a proposal not accepted, a new proposal taking into account the group discussion should be put together, as a continuation of the discussion. This reformation should include:

- **Original Issue/Proposal**: description of the issue, preserving the exact same words that were used for describing it
- **Discussion Context**: include any discussions (e.g. emails) or meeting minutes
- **New Proposal**: rearticulation of a new proposal which can be voted yes/no

Issue Concluding

Issues are discussed on this basis (*Issue Discussion* and *Proposal Reformulation* steps) until they are solved. An issue is considered solved when either

- a proposal is accepted which closes the issue
- the issue is declared no longer of interest / need
- there is not enough evidence to proceed

When an issue is closed, the following data should be added to the issue:

- **Outcome**: a summary of the final decision leading to the closing of the issue and references to any follow up issues
- **Status**: can be changed to 'closed' or 'paused'
- **Closing Date**: the date issue was closed

Manner of Decision of Issues

The CIDOC CRM-SIG operates on a democratic basis. Decisions are put to the member community. CIDOC CRM-SIG members' votes (i.e. votes cast by the representatives of the institutions making up the group) can either be taken through the mailing list or at an in-person meeting. Votes can only be taken on issues formulated in the procedures outlined above.

The possibilities for voting are: Yes and No.

If a vote is called and all votes are in favour of the proposed solution, then the decision passes. This is as true for a vote within a CRM-SIG session as it is for a vote via email.

Where there are negative votes, it is considered good practice that CRM-SIG members opposing a proposed solution back their objection with supporting arguments. The points raised can then be addressed and resolved before a decision is reached. This means that either a new discussion will take place, which will lead to the formulation of a new proposal and vote; or, alternatively, the decision can be reached by majority - provided that any objections raised have been thoroughly discussed among CRM-SIG members. In general, the CRM-SIG strives for consensus.

In the event that they see no merit to a proposed solution, CRM-SIG members can exercise their right to veto an e-vote and block the decision making. Vetoed decisions have to be discussed in a face-to-face (or virtual/hybrid) CRM-SIG meeting.

In the course of a CIDOC CRM-SIG meeting if a decision has been taken, it cannot be undone within the course of that same meeting.

There is no quorum rule for voting. However, accepting the outcome of a vote where only a few SIG members have voted to resolve a particular issue in a given manner, is not considered good practice. In principle, at least the CRM-SIG members who have contributed to the discussion that lead up to a concrete proposal to resolve the issue need to participate in the voting procedure. An issue once decided and closed cannot be undone. If there is a challenge to it, a new issue must start.

5.1.2.5 Long-term Issue Management - Documenting the Provenance

Although the CRM grows and develops through the opening and resolving of issues, the overall mass of issues and the queue that develops needs to be properly managed over time in order to avoid becoming bogged down in minor, poorly documented and poorly understood issues. In order to help keep track of issues over time, it is important they are curated over time and accounted for with regards to their origin, their meaning and their relevance.

The CRM-SIG should therefore have an account at the start of each group meeting of how many issues are open of what kind for each family model. At the end of each group meeting the number of issues closed and new ones raised should also be indicated. Documenting the provenance (i.e. the issue raiser) can help to trace the raison d'etre of an issue and lead to potential closing of issues if they pass in relevance over time.

Issues that are over two years old should be considered for staleness and potentially put on pause or closed for lack of interest. The issues that fall in this category should be considered within a session of the group meeting in order to ensure the continuous curation and pruning of issues and help avoid overloading the meeting with outdated and irrelevant issues with no use case or supporter.

5.1.2.6 Membership

Members of CIDOC CRM-SIG are its member institutions. Member institutions receive one vote. This vote is cast by their representative or a representative of their representative.

5.1.2.7 CIDOC CRM SIG Meetings

As stated earlier, the CIDOC CRM-SIG meets three or four times a year, the meetings being hosted by member institutions of the CRM-SIG. These meetings are either in physical form or virtual/hybrid meetings. To ensure that the discussions leading up to any decision-making are well-documented and recoverable after the meetings are over, the group of the CIDOC CRM editors have been recording the sessions of the meetings. The recordings are locally stored by the secretary of the group, to be used only for the purpose of keeping minutes. After the minutes of a meeting have been published, the records are destroyed, as they are not meant to be shared with anyone.

5.2 Examples of More Governance Models

Here we describe in brief more examples of governance models related to successful ontologies (schema.org, Gene Ontology, and/or W3C ontologies).

5.2.1 The Governance of Schema.org

Schema.org is a collaborative, community activity with a mission to create, maintain, and promote schemas for structured data on the Internet, on web pages, in email messages, and beyond. It was founded by Google, Microsoft, Yahoo and Yandex. Schema.org vocabularies are developed by an open community process, using the *public-schemaorg@w3.org* mailing list and through GitHub. As regards governance, there is an 'about' page describing the people involved and the role of each one[16].

The same page describes the *Community Group* and the *Steering Group*. In brief, Schema.org is organized via two groups: a small Steering Group responsible for high level oversight of the project (including approval of new releases), and a larger Community Group which handles the day to day activity of schema evolution, discussion and integration.

The forum[17] of the community groups allows discussing all changes, additions and extensions to schema.org. In addition to providing a public setting for the day-to-day operation of the project, it serves as the mechanism for reviewing extensions and as a liaison point for all parties developing independent extensions to the schema.org core. The releases are described in a dedicated webpage[18].

5.2.2 The Gene Ontology

The Gene Ontology (GO) resource[19] is the world's largest source of information on the functions of genes. This knowledge is both human-readable and machine-readable, and is a foundation for computational analysis of large-scale molecular biology and genetics experiments in biomedical research. The GO portal is used for providing all necessary information, including details about the process for suggesting new terms[20].

As regards evolution, they use a ticketing system supported by github[21] as well as a helpdesk[22].

5.2.3 The FoodEx2 Maintenance Policy

FoodEx2 is a comprehensive food and feed classification and description system developed and maintained by EFSA with the aim to keep it up to date and to cover the needs of data collections in different domains. To keep this terminology and sample coding framework fit-for-purpose, meeting evolving scientific and legislative requirements, a regular maintenance of this system is of utmost importance.

An annual maintenance of the system has taken place since April 2015 when the major revision (FoodEx2 revision 2) was performed. A maintenance technical group composed of EFSA staff with expertise in different domains (food consumption, chemical contaminants, pesticides residues, etc.) was set-up to evaluate proposals and requests provided by users and

[16] https://schema.org/docs/about.html
[17] https://www.w3.org/community/schemaorg/
[18] https://schema.org/docs/releases.html
[19] http://geneontology.org/
[20] http://geneontology.org/docs/contributing-to-go-terms/
[21] https://github.com/geneontology/go-ontology/issues
[22] http://help.geneontology.org/

stakeholders and take decisions regarding their implementation. The first maintenance was carried out in 2015 followed by five major releases performed between 2016 and 2022.

The suggestions provided during the period between January 2022 and January 2023 by users and stakeholders of the FoodEx2 system were collected through written feedback and used to update the system. Each request for addition or changes in the structure was considered and based on expert judgement, the technical group decided whether a change was needed, or the request was already covered by the available terms and facet descriptors. All changes were made following the structure and the logic of the system as described in the report on the development of FoodEx2 revision2. The tools used for implementing the maintenance were Microsoft Excel and the Catalogue Browser, a user-friendly tool developed in Java by EFSA staff, in particular the Integrated Data Unit (IDATA), to help create FoodEx2 codes and to facilitate the management of the MTX catalogue [8].

A technical report [8] describes the outcome of the maintenance processes, which contains addition of new terms, inclusion of a novel hierarchy, reportability changes of some terms and amendments to existing terms including enrichment of implicit facets and changes of the hierarchical relationships.

5.2.4 Health Evidence Knowledge Accelerator (HEvKA) governance

The Health Evidence Knowledge Accelerator (HEvKA)[23] is an open virtual group working to accelerate the identification, processing, and dissemination of knowledge about health by developing standards and support tools for computable expression of evidence and guidance. It is composed of 14 active working groups, that meet on a weekly basis (15 times per week), aiming to accelerate the development and implementation of standards for healthcare-related evidence exchange.

HEvKA activities are coordinated based on the Fast Evidence Interoperability Resources (FEvIR) Platform[24], a website and cloud-based system for creating, editing, viewing, and sharing scientific knowledge in an electronic form designed to standardize data exchange using the Fast Healthcare Interoperability Resources (FHIR) standard[25]. The FEvIR Platform is in development to support living systematic reviews, living guidelines, sharing across citation repositories, knowledge portals, and other projects related to scientific communication. It includes free tools to report, view and share structured summaries of recommendations, the justification of recommendations (such as the evidence-to-decision framework ratings), the supporting evidence (summary of findings and details for individual studies and outcomes), and citations for each of these summaries and components. Structured summaries enable re-use for adaptation and implementation.

6 Conclusion

This report proposed a governance model for managing the activities involving ontologies at EFSA. It contained recommendations about (a) the scoping of the ontology development

---

[23] https://fevir.net/resources/Project/29272
[24] https://fevir.net
[25] https://ecqi.healthit.gov/fhir

activities at EFSA (actors and use cases), (b) how to achieve external and internal consistency and interoperability, (c) how to maintain and support the evolution of ontologies, (d) access policies to adopt, (e) IT considerations (aligned with Deliverable D4), (f) outsourcing strategy, (g) training needs related to ontologies. The proposed model will be aligned with the existing governance activities of EFSA, such as Rebuild DF. Moreover the report describes in detail the governance model of an ontology that is now an ISO standard, as well as the governance of a few other popular ontologies. Finally, the report contains detailed guidelines for Ontology Documentation (Annex A), Definition of URIs (Annex B), and Dereferenceable URIs (Annex C).

The following figure summarizes the different activities and the corresponding actors that are needed to support the life cycle of an ontology. Overall, the workflow demonstrates the maintenance procedures of an ontology. Apart from the different user roles, a catalogue portal is used for collecting requests for issues and updates from a community of users, and for disseminating any updates of the ontology. This workflow is a combination of policies and procedures that are carried out in the governance models of other ontologies (e.g. CIDOC CRM) and, catalogue management policies at EFSA. It can be used as a basis for formulating concrete WIN documents, based on the infrastructure and technology adopted.

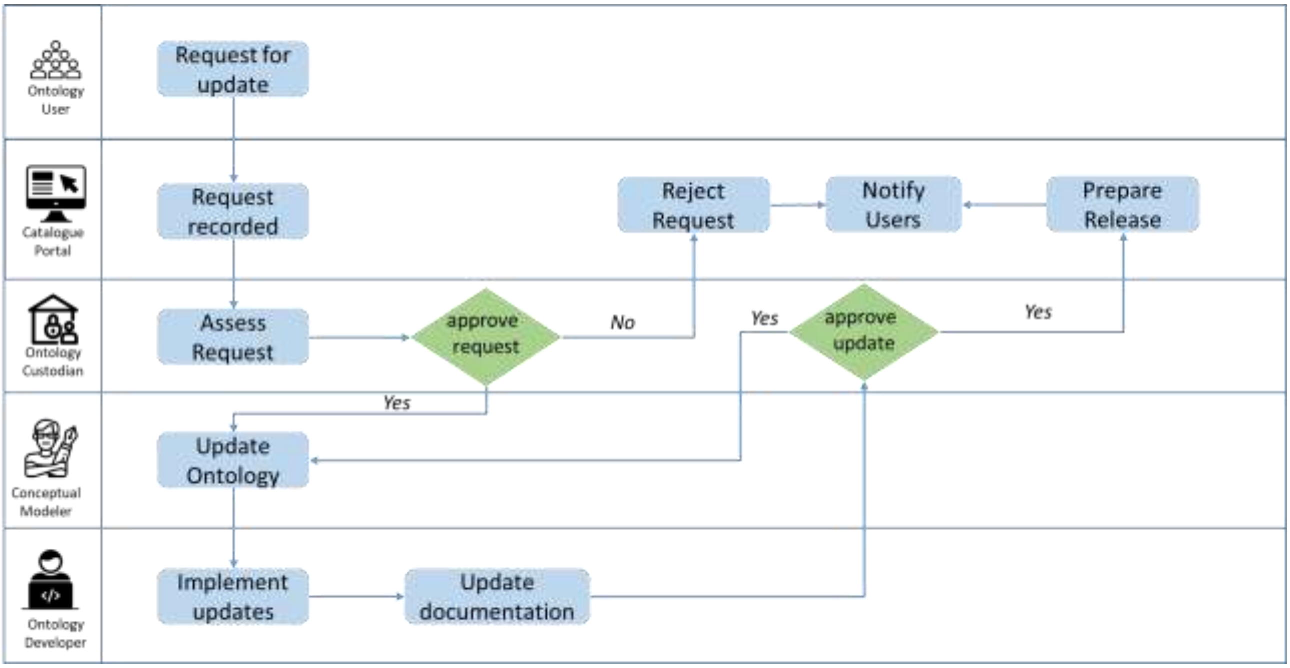

Figure 5. Ontology maintenance workflow

References

[1]. Tzitzikas Y., Fafalios P., Kritsotaki A., Axaridou A., Marketakis Y., Mountantonakis M., Bekiari C., and Theodoridou M., 2023. Ontologies and Case Studies, EFSA supporting publication 2023, May 2023.

[2]. Tzitzikas Y., Marketakis Y., Mountantonakis M., Fafalios P., Theodoridou M., Axaridou A., Kritsotaki A., and Bekiari C., 2023. Blueprint for Ontology Management, EFSA supporting publication 2023, May 2023.

[3]. Ana Iglesias-Molina, José Antonio Bernabé-Díaz, Prashant Deshmukh, Paola Espinoza-Arias, Aaron Ayllón-Benítez, Alba Fernández-Izquierdo, José María Ponce-Bernabé, Sara Pérez, Edna Ruckhaus, Oscar Corcho, & José Luis Sánchez-Fernández. (2022, August 18). Ontology Management in an Industrial Environment: The BASF Governance Operational Model for Ontologies (GOMO). https://doi.org/10.5281/zenodo.7007495
[4]. Bloehdorn, S., Haase, P., Huang, Z., Sure, Y., Völker, J., van Harmelen, F. and Studer, R., 2009. Ontology management (pp. 3-20). Springer Berlin Heidelberg.
[5]. Fraga, Alvaro Luis, Marcela Vegetti, and Horacio Pascual Leone. "Ontology-based solutions for interoperability among product lifecycle management systems: A systematic literature review." Journal of Industrial Information Integration 20 (2020): 100176.
[6]. Muñoz, A., Martí, L. and Sanchez-Pi, N., 2021. Data Governance, a Knowledge Model Through Ontologies. In Technologies and Innovation: 7th International Conference, CITI 2021, Guayaquil, Ecuador, November 22–25, 2021, Proceedings 7 (pp. 18-32). Springer International Publishing.
[7]. Raemaekers, S., Van Deursen, A. and Visser, J., 2014, September. Semantic versioning versus breaking changes: A study of the maven repository. In 2014 IEEE 14th International Working Conference on Source Code Analysis and Manipulation (pp. 215-224). IEEE.
[8]. EFSA (European Food Safety Authority), Salfinger A, Gibin D, Niforou K and Ioannidou S, 2023. FoodEx2 maintenance 2022. EFSA supporting publication 2023:EN-7900. 21 pp. doi:10.2903/sp.efsa.2023.EN-7900
[9]. Garijo, Daniel, and María Poveda-Villalón. "Best Practices for Implementing FAIR Vocabularies and Ontologies on the Web." In Applications and Practices in Ontology Design, Extraction, and Reasoning, pp. 39-54. IOS Press, 2020.
[10]. Tzitzikas, Y., Allocca, C., Bekiari, C., Marketakis, Y., Fafalios, P., Doerr, M., Minadakis, N., Patkos, T. and Candela, L., 2016. Unifying heterogeneous and distributed information about marine species through the top level ontology MarineTLO. Program, 50(1), pp.16-40.
[11]. Berners-Lee, T., Fielding, R. and Masinter, L., 2005. Uniform resource identifier (URI): Generic syntax (No. rfc3986).
[12]. Doerr, M., 2003. The CIDOC conceptual reference module: an ontological approach to semantic interoperability of metadata. AI magazine, 24(3), pp.75-75.
[13]. Mountantonakis, Michalis, Ioannis Theocharakis, and Yannis Tzitzikas. "Why We Need Ontology-Specific Data Portals: A Case Study for CIDOC-CRM." (2023).
[14]. Bizer, C., Heath, T., & Berners-Lee, T. (2011). Linked data: The story so far. In Semantic services, interoperability and web applications: emerging concepts (pp. 205-227). IGI global.
[15]. Jacobsen, A., de Miranda Azevedo, R., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C.T. and Goble, C., 2020. FAIR principles: interpretations and implementation considerations. Data intelligence, 2(1-2), pp.10-29.
[16]. European Food Safety Authority, 2013. Standard Sample Description ver. 2.0. EFSA Journal, 11(10), p.3424.

Glossary and Abbreviations

AI	Artificial Intelligence
API	Application Programming Interface
CRM	Conceptual Reference Model
DAMA2	Data Management and Data Analysis 2
EFSA	The European Food Safety Authority
FEvIR	Fast Evidence Interoperability Resources
GO	Gene Ontology
HEvKA	Health Evidence Knowledge Accelerator
HTML	HyperText Markup Language
HTTP	HyperText Transfer Protocol
IDATA	Integrated Data Unit
ISO	International Organization for Standardization
IT	Information Technology
FAIR	Findable Accessible Interoperable Reusable
KR	Knowledge Representation
LMS	Learning Management System
LOD	Linked Open Data
ML	Machine Learning
OBO	Open Biological and Biomedical Ontologies.
OWL	Web Ontology Language
SIG	Special Interest Group

SOP	Standard Operating Procedure
SPARQL	SPARQL Protocol and RDF Query Language
SSD2	Standard Sample Description ver. 2.0
RDF	Resource Description Framework
RDFS	Resource Description Framework Schema
UML	Unified Modeling Language
URI	Uniform Resource Identifier
VoID	Vocabulary of Interlinked Datasets
W3C	World Wide Web Consortium
WIN	Work Instructions
XML	eXtensible Markup Language

Annex A

Ontology Documentation

There are related best practices for ontology documentation [9].

The documentation of an ontology is critical, as it provides the context, accurate definitions and examples of the different concepts that are included in an ontology. The documentation can be either incorporated with the ontology itself, or as an external collection of documents that accompany the ontology. The former includes metadata and comments about the ontology and its resources and are described in the same file with the ontology (i.e. using RDF/S or OWL). The latter are external documents prepared either manually or using dedicated tools and services. A detailed description about the state of the art in ontology documentation tools can be found in the Deliverable D4 [2].

Ontology metadata can be categorized into metadata about the ontology per se, and metadata about the ontology entities (i.e. classes, properties and individuals). The former provide an overview of the ontology, describe the community and users that developed and maintains it, and report on the policies for accessing and using it. The below snippet shows the metadata of the MarineTLO ontology [10].

```
<owl:Ontology rdf:about="http://www.ics.forth.gr/isl/ontology/MarineTLO/">

   <dcterms:modified rdf:datatype="&xsd;date">2017-01-05</dcterms:modified>

   <dcterms:issued rdf:datatype="&xsd;date">2014-04-24</dcterms:issued>

   <owl:versionInfo rdf:datatype="&xsd;decimal">5.0</owl:versionInfo>

   <dcterms:title>MarineTLO Ontology</dcterms:title>

   <dcterms:description>

          MarineTLO is a top-level ontology for the marine domain (also applicable to the terrestrial domain).

   </dcterms:description>

   <dcterms:creator> Chrysoula Bekiari </dcterms:creator>

   <dcterms:creator> Martin Doerr </dcterms:creator>

   <dcterms:creator rdf:resource="https://www.ics.forth.gr/~tzitzik/"> Yannis Tzitzikas </dcterms:creator>

   <dcterms:creator> Carlo Allocca </dcterms:creator>

   <dcterms:creator> Julien Barde</dcterms:creator>

   <dcterms:creator> Nikos Minadakis </dcterms:creator>

   <dcterms:creator rdf:resource="https://www.ics.forth.gr/~marketak/">Yannis Marketakis </dcterms:creator>

   <dcterms:creator rdf:resource="https://www.ics.forth.gr/~fafalios/"> Pavlos Fafalios</dcterms:creator>

   <dcterms:license rdf:resource="http://creativecommons.org/licenses/by/3.0/"/>
```

```xml
    <dcterms:rights>Copyright © FORTH ICS</dcterms:rights>

    <dcterms:publisher rdf:resource="http://www.ics.forth.gr/isl"/>

    <owl:versionIRI rdf:resource="http://www.ics.forth.gr/isl/ontology/MarineTLO/"/>

</owl:Ontology>
```

Garijo and Poveda-Villalón [9] provide their recommended and optional annotation properties for describing ontologies and their entities, along with the candidate properties that can be reused from existing ontologies, vocabularies and standards. Their recommendations are shown in Table 2 and Table 3[26].

Property	Annotation Property	Rationale	Usage
License	dcterms:license	Usage conditions	Recommended
Creator	dcterms:creator	Provenance and attribution	Recommended
Contributor	dcterms:contributor	Provenance and attribution	Recommended
Creation Date	dcterms:created	Provenance	Recommended
Previous Version	owl:priorVersion	Provenance and comparison	Recommended
Namespace URI	vann:preferredNamespaceUri	Identifying the ontology	Recommended
Version IRI	owl:versionIRI	Versioning	Recommended
Namespace prefix	vann:preferredNamespacePrefix	Identifying the ontology	Recommended
Title	dcterms:title	Understanding	Recommended
Description	dcterms:description	Understanding	Recommended
Citation	dcterms:bibliographicCitation	Credit	Recommended
Abstract	dcterms:abstract	Additional information	Optional
See also	rdfs:seeAlso	Additional information	Optional
Status	sw:status	Maturity information	Optional

[26] Namespace prefixes: rdf: http://www.w3.org/2000/01/rdf-schema#, owl: http://www.w3.org/2002/07/owl#, bibo: http://purl.org/ontology/bibo/, foaf: http://xmlns.com/foaf/0.1/, dcterms: http://purl.org/dc/terms/, vaem: http://www.linkedmodel.org/schema/vaem, vann: http://purl.org/vocab/vann/, sw: http://www.w3.org/2003/06/sw-vocab-status/ns#

Backwards compatibility	owl:backwardCompatibleWith	Version compatibility	Optional
Incompatibility	owl:incompatibleWith	Version compatibility	Optional
Modification date	dcterms:modified	Provenance and timeliness	Optional
Issued date	dcterms:issued	Provenance and timeliness	Optional
Source	dcterms:source	Provenance	Optional
Publisher	dcterms:published	Provenance	Optional
DOI	bibo:doi	Bibliographic information	Optional
Logo	foaf:logo	Identifying the ontology	Optional
Diagram	foaf:depiction	Visual documentation	Optional

Table 2. Recommended and optional metadata for describing ontologies (source [9]).

Property	Annotation Property	Rationale	Usage
Label	rdfs:label	Readability	Recommended
Definition	rdfs:comment	Understanding	Recommended
Example	vann:example	Understanding	Optional
Status	sw:term status	Understanding	Optional
Rationale	vaem:rationale	Understanding	Optional
Source	dcterms:source	Provenance	Optional
Defined By	rdfs:isDefinedBy	Provenance	Optional

Table 3. Recommended and optional properties for describing ontology terms (source [9]).

Another form of documenting an ontology is the creation of external documents describing several aspects of it. Such documents usually contain more extensive descriptions and are accompanied by diagrams, figures, examples, etc., and they can be either traditional static documents (e.g. a PDF document) or interactive HTML pages. The creation of such a document can be a very time-consuming process, so nowadays there are several tools and systems that support the automatic construction of the documentation of an ontology using their original serialization (i.e. the user provides the RDF/XML serialization of an ontology and its documentation in the form of a collection of HTML pages is created). It is generally recommended to create the documentation of an ontology in the form of an HTML page to allow users understand and navigate on the Web. This enables as well the URI dereferencing procedure (more details about this are given in Annex C. Figure 6 shows an indicative screenshot of the documentation (created as a collection of HTML pages) of MarineTLO Ontology.

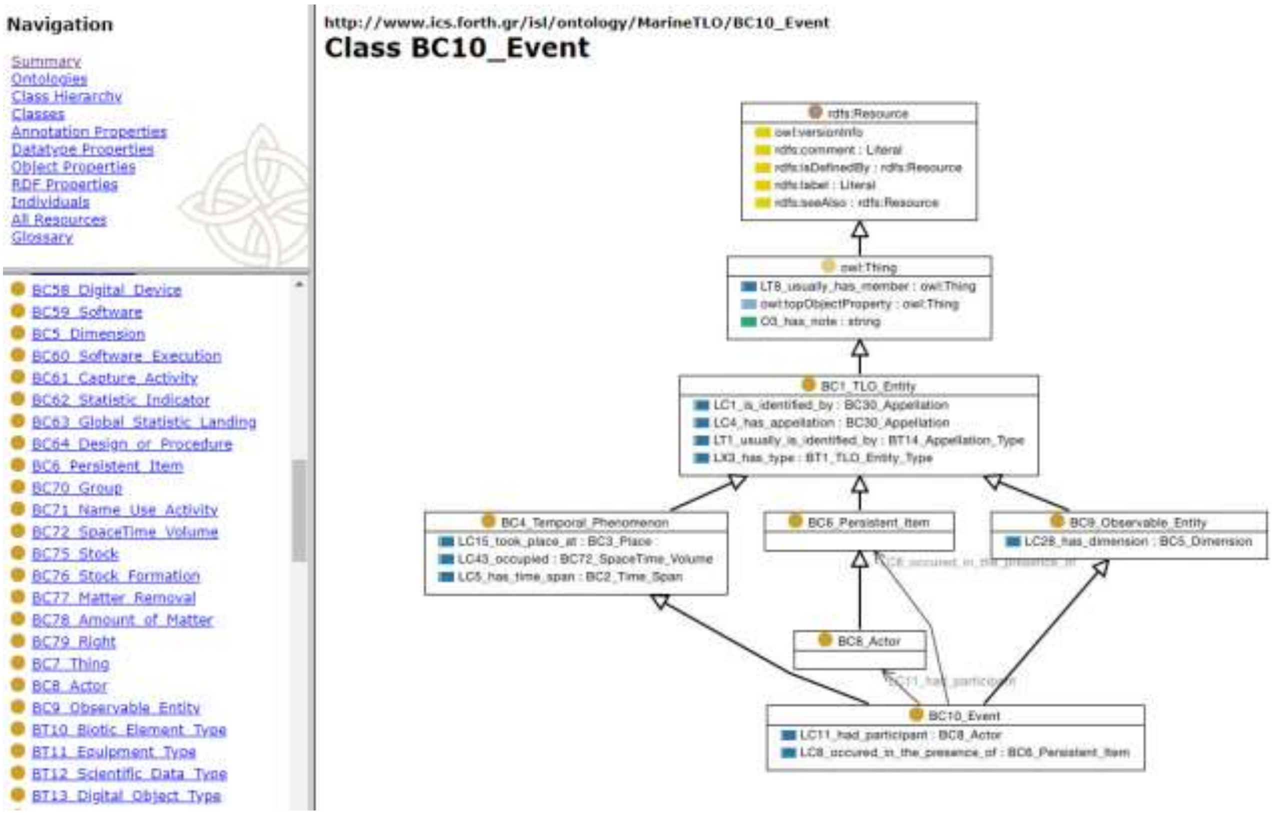

Figure 6. Documentation of MarineTLO ontology as HTML pages.

Annex B

Definition of URIs

Uniform Resource Identifiers (URI) [11] are generic naming constructs that are used for identifying resources. In the ontologies world, they are used for identifying all the resources of an ontology (i.e. classes, properties, etc.). In general, a URI consists mainly of two parts: the namespace section that describes the general context (i.e. the domain) and the trailing part that is the local name of a resource (i.e. a class name). For example, the URI http://www.cidoc-crm.org/cidoc-crm/E5_Event that refers to the class of the ISO 21127:2014 CIDOC-CRM, has namespace http://www.cidoc-crm.org/cidoc-crm/ and the name of the class is E5_Event.

As described above, the namespace prefix of the ontology is related to the domain of the ontology. It can be a combination of the ontology name and targeted domain. In any case the selection of a simple and short name to be used in the namespace prefix, will help others remember and efficiently reuse the ontology. If the ontology name consists of many words, then the definition of an acronym is a very good practice (i.e. foaf is used for the Friend of a Friend Ontology[27]). The same principle applies also for the local names of the resources (i.e. classes and properties). Simple, short and non-ambiguous names are the best candidates for naming classes and properties (i.e. foaf:Person).

Another aspect is the selection of the character that will be used for distinguishing the namespace of the ontology and the local names of its resources. Usually it is either a hash ('#') or a slash character ('/'). For example FOAF uses the hash character approach (e.g. http://xmlns.com/foaf/0.1/#term_Person), while CIDOC-CRM uses the slash character approach (e.g. http://www.cidoc-crm.org/cidoc-crm/E5_Event). The selection of the approach has an impact also on the way the ontology will be documented online, because technically the hash character approach will require all the documentation details (e.g. text, images, diagrams, examples, etc.) to be added in a single web page with internal HTML links for the various resources, while with the slash character approach the documentation can be split across different html pages. As a general comment, it is suggested to use the slash character approach for ontologies with a considerable amount of resources.

Of course, ontologies are evolving and several versions are developed (as described in more detail in Section 4.4). The usual convention is that the namespace prefix of an ontology does not include any information about the version of the ontology. It is discouraged to include version numbers as part of the ontology URI, as it would deeply affect interoperating with its instances (e.g. older instances of an ontology could be used with newer versions of the ontology itself). Therefore, the suggested practice is to use a namespace prefix for the ontology that will not include information about the version as part of its URI, and will always point to the latest version of the ontology.

[27] http://xmlns.com/foaf/0.1/

Annex C

Dereferenceable URIs

As already discussed, URIs are used for identifying several parts of an ontology. It is evident that URIs can be effectively used by machines, however ontologies should be made available in human readable manner as well. A technical solution to this is to enable different views for a URI depending on the agent requesting it. In other words, the plan is to deliver the HTML documentation of an ontology when a URI is requested from a web browser, which means that it is a human requesting that information, and to resolve to a standard RDF serialization of the ontology when a URI is requested from a different place (e.g. from an API call, an ontology editor, etc.). This is implemented using the ontology dereferencing policy[28].

The HTTP URI dereferencing mechanism relies on HTTP accept headers and the hosting server content negotiation policy to be able to redirect an incoming request properly. More specifically, the configuration of the hosting server should be updated to handle in a different manner the incoming requests with HTTP ACCEPT header "application/rdf+xml" compared to requests with HTTP ACCEPT header set to "text/html". The first request should be redirected to return the RDF serialization of the ontology, while the second one should point to the HTML documentation. The actual configuration that is required to support content negotiation depends on the technical details of the hosting server. A common configuration on top of Apache2 web server[29] relies on updating the htaccess configuration file. The following snippet shows an indicative configuration for accessing the CIDOC CRM ontology.

```
# Rewrite rule for accessing the HTML version.

RewriteCond %{HTTP_ACCEPT} !application/rdf\+xml.*(text/html|application/xhtml\+xml)

RewriteCond %{HTTP_ACCEPT} text/html [OR]

RewriteCond %{HTTP_USER_AGENT} ^Mozilla/.*

RewriteRule ^$ https://cidoc-crm.org/index.html [R=303,L]



# Rewrite rule to serve the RDF/XML serialization

RewriteCond %{HTTP_ACCEPT} application/rdf\+xml

RewriteRule ^$ https://cidoc-crm.org/rdfs/7.1.2/CIDOC_CRM_v7.1.2.rdf [R=303,L]
```

28 https://www.w3.org/2001/tag/doc/httpRange-14/2007-05-31/HttpRange-14
29 https://httpd.apache.org/