FAIR-IMPACT
Expanding FAIR solutions across EOSC

Project Title	Expanding FAIR solutions across EOSC
Project Acronym	FAIR-IMPACT
Grant Agreement No.	101057344
Start Date of Project	2022-06-01
Duration of Project	36 months
Project Website	https://fair-impact.eu/

1 MS4.1 - Semantic artefact governance models: example of community practices

Work Package	**WP4, Metadata and Ontologies**
Lead Author (Org)	**Parham Ramezani (LifeWatch), Nina Grau (INRAE)**
Contributing Author(s) (Org)	**Clement Jonquet (INRAE), Nicola Fiore (LifeWatch)**
Due Date	**30/11/2023**
Date	**30/11/2023**
Version	**V1.0**

Dissemination Level	
X	PU: Public
	PP: Restricted to other programme participants (including the Commission)
	RE: Restricted to a group specified by the consortium (including the Commission)
	CO: Confidential, only for members of the consortium (including the Commission)

Funded by the European Union

2 Versioning and contribution history

Version	Date	Author	Notes
0.1	2023.10.10	Parham Ramezani (LifeWatch)	TOC and V0.1
0.2	2023.10.31	Parham Ramezani (Lifewatch)	Initial write-up summarizing the workshop
0.3	2023.11.08	Nina Grau (INRAE)	Executive summary and introduction
0.4	2023.11.13	Nina Grau (INRAE)	OBO Foundry presentation summary
0.5	2023.11.27	Parham Ramezani (Lifewatch) and Nina Grau(INRAE)	Conclusion
1.0	2023.11.27	Clement Jonquet (INRAE) and Nicola Fiore (LifeWatch)	General proofreading

Disclaimer

FAIR-IMPACT has received funding from the European Commission's Horizon Europe funding programme for research and innovation programme under the Grant Agreement no. 101057344. The content of this document does not represent the opinion of the European Commission, and the European Commission is not responsible for any use that might be made of such content.

3 Table of Contents

Funded by the European Union

TERMINOLOGY

Terminology /Acronym	Description
ABCD	Access to Biological Collection Data
CC	Creative Commons
CGIAR	Integrated Breeding Platform and Consultative Group on International Agricultural Research
CO	Crop Ontology Project
DOIs	Digital Object Identifier
DDI	Data Documentation Initiative
EBI	European Bioinformatics Institute
EOSC	European Open Science Cloud
EFO	Experimental Factor Ontology
ETS	Ecological Trait-data Standard
Exec	Executive Committee
FAQs	Frequently Asked Questions
FAIR	Findable, Accessible, Interoperable and Reusable
GOMO	Governance Operational Model for Ontologies
IRI	Internationalized Resource Identifiers
IVOA	International Virtual Observatory Alliance
IT	Information Technology
KTBL	Kuratorium für Technik und Bauwesen in der Landwirtschaft
LOT	Linked Open Terms
MOD	Ontology Description and Publication Ontology
OBO	Open Biological and Biomedical Ontologies
ODK	Ontology Development Kit
OLS	Ontology Lookup Service
ORCID	Open Researcher and Contributor Identifier
OWL	Web Ontology Language
PURL	Persistent Uniform Resource Locators
RAM	Responsibility Assignment Matrix
RDF	Resource Data Framework
RDFS	RDF Schema
ROR	Research Organization Registry
SA	Semantic Artefacts
SOP	Standard Operating Procedures
SSSOM	Simple Standard for Sharing Ontology Mappings
T4.1	Task 4.1 in Work Package 4: Semantic Artefact Disciplinary Governance
TCG	Technical Coordination Group
TO	Trait Ontology
TS	Terminology Service
UML	Unified Modeling Language
VANN	Vocabulary for Annotating Vocabulary Descriptions
VEP	Vocabulary Enhancement Proposal
WG	Working Groups
W3C	World Wide Web Consortium
WP4	Work Package 4 in the FAIR-IMPACT Project: Metadata and Ontologies

Funded by the European Union

Executive Summary

This report is a synthesis of the FAIR-IMPACT Semantic Artefact Governance Workshop[1] run on 28th September 2023, which presented a review of multiple approaches to community-driven governance of semantic artefacts. This work is developed by Task 4.1 (T4.1) on "Semantic artefact disciplinary governance", which is part of the Work Package 4 about "Metadata and ontologies" in the FAIR-IMPACT project. FAIR-IMPACT supports the implementation of FAIR-enabling (Findable, Accessible, Interoperable, and Reusable) practices, tools and services across scientific communities within the European Open Science Cloud (EOSC) ecosystem.

I2 FAIR principle "(Meta)data use vocabularies that follow FAIR principles", recognize the crucial role of semantic artefacts as a key component for enabling FAIR. Thus, FAIR-IMPACT aims to define the support, governance, and coordination mechanisms needed to ensure their adoption and continued utility in a FAIR EOSC. To contribute to this mission, T4.1's *FAIR-IMPACT Semantic Artefact Governance Workshop* outlined communities-driven governance examples of their semantic artefacts. Indeed, governance for semantic artefacts is crucial as it establishes the frameworks to ensure ethical, clear, transparent, and responsible utilization across diverse applications of their use. The workshop illustrated nine examples of practices for managing semantic artefacts in different communities from agri-food to biomedicine, through astrophysics and industry.

This report includes an introduction of general aspects of quality-verified digital resources and the FAIR principles, the FAIR-IMPACT's T4.1 targeted goals, the workshop methodology applied, the represented use cases and the synthesis of existing semantic artefact governance practices encountered by the communities presented.

[1] https://fair-impact.eu/events/fair-impact-events/fair-impact-semantic-artefact-governance-workshop

1 Introduction

1.1 General information

1.1.1 Research reusability

Reusability is one of the cornerstone of the research process. Indeed, the publication holding the research's final results have to be part of a corpus of documents that defines the state of knowledge validated by the field, called references. However, not whichever document can be used as reference. The document has to be previously validated by a peer's committee to be reused. This process of reuse of approved scientific resources is mandatory for scientific trustworthiness.

In the digital age, the reusability process in the research field has changed and needs to be adapted and reassessed. A set of criteria has been proposed, whose reliability assessment is at the forefront. Firstly, the research output must have to be sufficiently understandable and meaningful to assess their quality. Then, to enrich research reusability, the European Open Science Cloud (EOSC)[2] promotes interdisciplinary research by supporting both machine and human-readable research output. In addition, to further promote research reusability, a set of FAIR principles[3] has been established. These principles not only highlight interoperability and reusability aspects (which are represented by the I and R of principle abbreviation) but also support findability and access (which is represented by the F and A of principle abbreviation), as well as long-term care of research output.

1.1.2 Semantic artefacts

Semantic artefacts (SA) are defined as "machine-actionable and readable formalization of a conceptualisation that enables the sharing and reuse by humans and machines"[4]. This means that they are understandable research outputs and can be accessible with minimal effort using machines, which facilitates the researchers' ability to both reuse and reproduce them. SAs include ontologies, terminologies, taxonomies, thesauri, vocabularies, metadata schemas and other standards.

Thus, SAs are an important resource for research reuse. SAs are meeting many of the criteria previously mentioned for a trusted and efficient reuse of research output, by involving:

- the possibility of quality assessment by the way of their meaningful metadata,
- machine consultation through machine-readable formalization of their metadata, and enhancing the interoperability between scientific disciplines,
- respecting the I2 FAIR principle by being metadata vocabularies.

---

[2] European Commission, Directorate-General for Research and Innovation, Strategic Research and Innovation Agenda (SRIA) of the European Open Science Cloud (EOSC), Publications Office of the European Union, 2022, https://data.europa.eu/doi/10.2777/935288

[3] Wilkinson, M., Dumontier, M., Aalbersberg, I. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. *Sci Data* 3, 160018 (2016). https://doi.org/10.1038/sdata.2016.18

[4] Le Franc, Y., Parland-von Essen, J., Bonino, L., Lehväslaiho, H., Coen, G., & Staiger, C. (2020). D2.2 FAIR Semantics: First recommendations (1.0 DRAFT). FAIRsFAIR. https://doi.org/10.5281/zenodo.3707985

Funded by the European Union

1.1.3 SA reusability barriers

While SA represents the highest level of meaningfulness of knowledge representation integrating interoperability framework, their reusability in the research field is still limited.

Currently, the reutilisation of SAs is hindered by the lack of:

- sufficient and shared metadata which reduces understanding and in turn impact the research output quality-verification,
- harmonization for the standardization of the metadata formats across scientific disciplines which block their interoperability, and
- clear and define lifecycle management that can depreciate the quality and intelligibility of the SAs.

1.2 Goals and scopes

1.2.1 FAIR-IMPACT project

The overall objective of the FAIR-IMPACT project is to realize a FAIR EOSC, that is an EOSC of FAIR research outputs and services. Thus, FAIR-IMPACT supports the implementation of FAIR-enabling practices across scientific communities and research outputs at a European, national, and international level.
While in some areas FAIR-enabling foundations are well established; in others the FAIRification consolidation is still required. Belong to this challenge, the research outputs descriptions with shared and common semantics to make them interoperable and reusable. To do so, the objective of Work Package 4 (WP4) on 'Metadata and ontologies' of FAIR-IMPACT has placed SAs at the forefront of FAIRization promotion and implementation. One of the first subtasks of the WP4 is to outline how SAs are handled and managed into existing communities and consequently generate SAs governance examples which will support in turn the coordination of SA within the EOSC ecosystem.

1.2.2 Objectives

The FAIR-IMPACT Task 4.1's, called “Semantic artefact disciplinary governance”, aims to provide a first a projection of FAIR principles application into multiple domains and for all forms of SA. This document aims to create a clear argument for the need for a governance model around a standard definition of FAIRness, demonstrate how stakeholders will benefit from it, and explore existing governance models from peer internet projects to better understand the decisions that need to be made.
To achieve this goal, we’ve organized our work into three successive actions. To do so, we’ve planned:

1. a desk review in order to determine a definition of governance that can be applied to SAs, as to our knowledge no clear definition of SAs' governance has been found in scientific literature,
2. a review of the governance aspects of existing SAs in various disciplines by interviewing a broad range of communities,
3. to develop generic models of governance of SAs that support and ensure their FAIRification and coordination within the organization.

Funded by
the European Union

In this report, we will showcase 1) the methodology we employed to characterize SA governance, 2) the approach we used to gather governance aspects of communities using SAs, albeit 3) the development of generic models of governance, which is also part of the T4.1, is the next step that will feed another FAIR-IMPACT deliverable.

1.2.1 Next step: models of SA governance

The product of this report will feed into another deliverable called "Report on semantic artefact governance models and disciplinary approaches for inclusion within EOSC" which will document examples of models of SA governance. The goal of the deliverable will then be to propose a set of models that any new community, project or infrastructure coming to EOSC could decide to adopt or embrace for the management of their own SAs.

1.3 Methodology

1.3.1 Desk review

1.3.1.1 Governance vs management

Before starting this section, we believe that it is important here to clarify the differences between management and governance. For this, we used the Weill P. and Ross J.W publication[5] that highlights the contrast between the two terms. The governance refers to what decisions are made and who is involved in the decisions, otherwise the management is focused on the actions to make and implement decisions. Governance can have multiple meanings and frameworks according to the application of what resource is “governed” under his control and the intended objectives. The lato sensus definition of governance is to have the administration of something. Based on this, we decide here to qualify the governance of SAs as the specifications of the decision rights and accountability framework controlling the SAs in a group or community.

1.3.1.2 SA governance

Khatri V. and Brown’s framework design for data governance[6] has been chosen as it combines governance practices used in IT and data (Figure 1), which corroborates with our notion of SA. In the proposed framework for data governance, there are five decision domains that are interconnected: data principles, data quality, metadata, data access, and data lifecycle (Figure 1).

[5] Weill, P. and Ross, J. W. “IT governance: How top performers manage IT decision rights for superior results”. Harvard Business School Press, Boston, MA, 2004.

[6] Khatri V. and Brown C.V.; Designing data governance. 2010. Commun. ACM. https://doi.org/10.1145/1629175.1629210

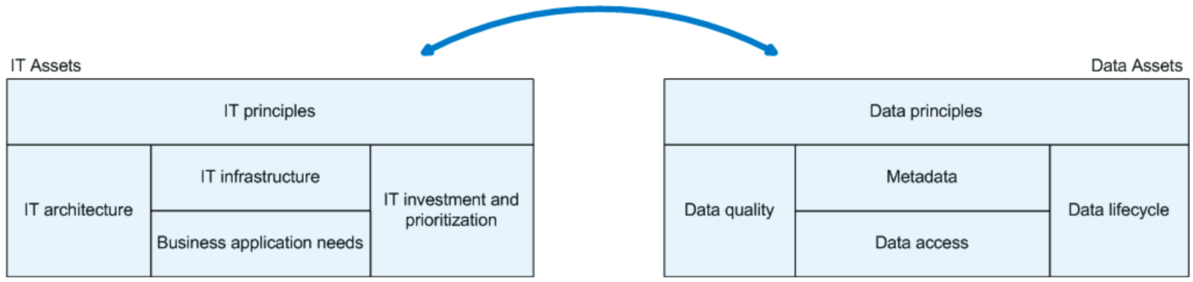

Figure 1. Framework for IT and data decision domains from Khatri V. and Brown C.V.

In addition, these five decision domains also gather research reusability criteria previously described, strengthening to use Khatri V. and Brown's concepts as template to define governance aspects of SA. Indeed,

- the access domain gather FAIR's findability and access criteria,
- the metadata and quality domains gather respectively FAIR's metadata criteria and quality-verified aspect of research reusability,
- the lifecycle domain gather FAIR's long-term care aspects,

will give us instructive information to implement SA governance within FAIR EOSC.

Thus, we've decided to apply these five decision domains within the governance framework of SAs, and to complete the exploration of these aspects we also added a specific domain for the stakeholders involved and the type of decisions they make.

1.3.2 Interview communities

We have decided to gather information about the SA's governance from communities in the form of a workshop.[7] The auditioned communities come from a wide scientific domain; from agronomy to industry and passing through astrophysic and more, guaranteeing the SA application's exhaustivity. The nine communities who agreed to play along are represented below (Figure 2).

[7] Semantic artefact governance workshop website, https://fair-impact.eu/events/fair-impact-events/fair-impact-semantic-artefact-governance-workshop

Funded by
the European Union

1. Crop Ontology
acronym: **CO**
website: https://cropontology.org/
type of SA: ontologies of phenotypic traits and variables

2. Open Biological and Biomedical Ontology
acronym: **OBO Foundry**
website :https://obofoundry.org/
type of SA: biological, biomedecine and domains adjacent to life sciences ontologies and metadata standards

3. BASF Governance operational model for ontologies
acronym: **BASF GOMO**
website: https://basf.com
type of SA: BASF chemical ontologies

4. Vocabularies of the French National Research Institute for Agriculture, Food and the Environment
acronym: **VO INRAE**
website: https://vocabulaires-ouverts.inrae.fr/
type of SA: thesauri and ontologies support for IRNAE researchers

5. International virtual observatory alliance
acronym: **IVOA**
website: https://www.ivoa.net/
SA: astronomical datasets

6. National Research Data Infrastructure in Germany for Biodiversity
acronym: **NFDI4biodiversity**
website: https://www.nfdi4biodiversity.org/
type of SA: taxonomies of living organisms, geographical ontologies and metadata standards

7. European Bioinformatics Institute (EBI)part of the European Molecular Biology Laboratory (EMBL)
acronym: **EMBL-EBI**
website: https://www.ebi.ac.uk/
SA: life science research ontologies

8. The Smart Applications REFerence (SAREF) of the European Telecommunications Standards Institute (ESTI)
acronym: **ETSI SAREF**
website: https://saref.etsi.org/
SA: ETSI smart applications ontologies

9. FAO Multilingual Agricultural Thesaurus
acronym: **AGROVOC**
website: https://www.fao.org/agrovoc/
SA: food and agriculture FAO thesaurus

Figure 2. List of the nine interviewed communities, their acronyms, website and type of SA used.

We presented a set of six guiding questions to these nine communities, recapping the five decision domains of SA governance as well as the stakeholders involved and the decisions made aspects.

The set of questions that were asked and validated by each community's representative is listed below:

- **Q1:** (**Principles**) Goals of the Infrastructure/Project/Research entity
  - What is the nature of the semantic artefacts and where and from whom do they come?
  - Is your group of semantic artefacts hosted by one or several semantic artefact catalogues? Are those catalogues part of your publication processes?
- **Q2: Metadata**
  - Which information do you require to describe your semantic artefacts?
  - Which metadata standards do you use?
- **Q3: Quality**
  - How do you assure the quality of your semantic artefacts?
  - What are the recommended good practices? Are you following guidelines or high level principles?
  - Do you enforce reuses and imports from other semantic artefacts?
  - How do you collect feedback and issues from the users?
- **Q4: Access**
  - Do you have terms and conditions for your semantic artefacts and who is responsible? How are they licensed?
  - Do you have machine accessible endpoints available? Other services to share/support the ontologies?
  - How do you communicate with semantic artefact users and get them notified?

   - How do you ensure the sustainability of your semantic artefacts? (financially speaking but not only)
- **Q5: Lifecycle**
   - How do you deal with the maintenance? Describe the processes to add new terms (method, periodicity and policy)?
   - How do you deal with retirement or obsolescence?
   - How do you manage the versioning?
   - Do you manage different languages (for labels)?
- **Q6: Stakeholders and decisions**
   - Can you list the stakeholders involved in each of these governance aspects? (developper, curator, board, experts, committee…)?
   - How are you taking decisions for each of the governance workflow steps?

2 Summary of the workshop

The workshop was well received, with more than 82 registrations. While the online attendance's peak was around 30 online participants and 10 locals at Lecce in Italy, this number did not decrease during the all 2h workshop, highlighting the participant's alertness for the presentations.

During the introduction, Clement Jonquet presented the FAIR-IMPACT project and EOSC ecosystem, the importance of governance, the governance aspects examined and the surveyed communities. Then, in concise 15-minute sessions, each presenters answered the guiding questions covering the six decision domains related to the governance of SAs within their respective communities. Nine communities agreed to play along.

The workshop's video is now available on FAIR-IMPACT's website[8], with at this date more than 200 views, and the speakers' slides are available on the Zenodo report.

2.1 The Crop Ontology by Elizabeth Arnaud & Marie-Angélique Laporte (time in video 13:08)

2.1.1 Principles

The Crop Ontology Project[9] (CO) was established by the Integrated Breeding Platform and Consultative Group on International Agricultural Research (CGIAR) in 2008. CO aims to be a crop-specific ontology of phenotype traits and variables with semantic relationships, developed and cared for and by an assigned curator from a lead center or a CGIAR institution. Now, it is being developed by a wider community composed of curators and crop experts from CGIAR as well as from partners, universities and consortia. Breeders, geneticists, and food scientists are the primary providers, but they also receive spontaneous submissions. Currently, CO has under his belt 38 crop-specific ontologies which all formed the Crop Ontology project. To facilitate effective governance, three advisory groups have been established, in alignment with the recommendations of the governance and stewardship framework[10]: Curator Committee, Scientific Advisory Committee and Strategy Advisory Committee. The list of committee members can be accessed directly on the website[11] for more information.

2.1.2 Metadata

The CO governance framework recognizes the importance of minimal amount of metadata about the crop ontologies and within the cropontology.org registry, encompassing: ontology name, short descriptions, curator and contributor names and affiliations, version, and links to peer-reviewed papers. While the ontology currently lacks a specific metadata standard and machine-readability, a process was initiated this year to get DOIs along with citation and version information.

2.1.3 Quality

[8] https://fair-impact.eu/events/fair-impact-events/fair-impact-semantic-artefact-governance-workshop

[9] Crop Ontology website, https://cropontology.org/

[10] Crop Ontology Governance and Stewardship document, https://cgspace.cgiar.org/handle/10568/118001

[11] https://cropontology.org/page/MembersAC

Quality management in the CO community is ensured by clear guidelines[12] to assist scientists in utilizing the Trait Dictionary Template[13], which comes with embedded quality control tools to detect errors and discrepancies. The system also includes dedicated helpdesk support and an organized workflow that outlines all the steps for submitting an ontology, ensuring a comprehensive approach to quality management. Additionally, The Planteome GitHub repository[14] serves as the tool for archiving and versioning for many ontologies in CO. The community has established quality criteria for agricultural ontologies with significant input from the Community of Practice. Thus the governance framework controls predominantly 2 aspects: the ontology's accurate curation and its versioning (Figure 3).
Multiple feedback channels, including GitHub issue postings, helpdesk communication, Curator Committee meetings, surveys, and active engagement on the community Forum[15], collectively contribute to the governance of the crop ontologies with a focus on ensuring and enhancing overall quality. CO is clearly looking towards his community with whom they have a close relationship and has put a lot of effort into it through various communication channels and feedback tools. A notable feature of the Crop Ontology's approach is the mapping of crop-specific concepts to the species-neutral Plant Trait Ontology (TO)[16], enhancing trait data accessibility.

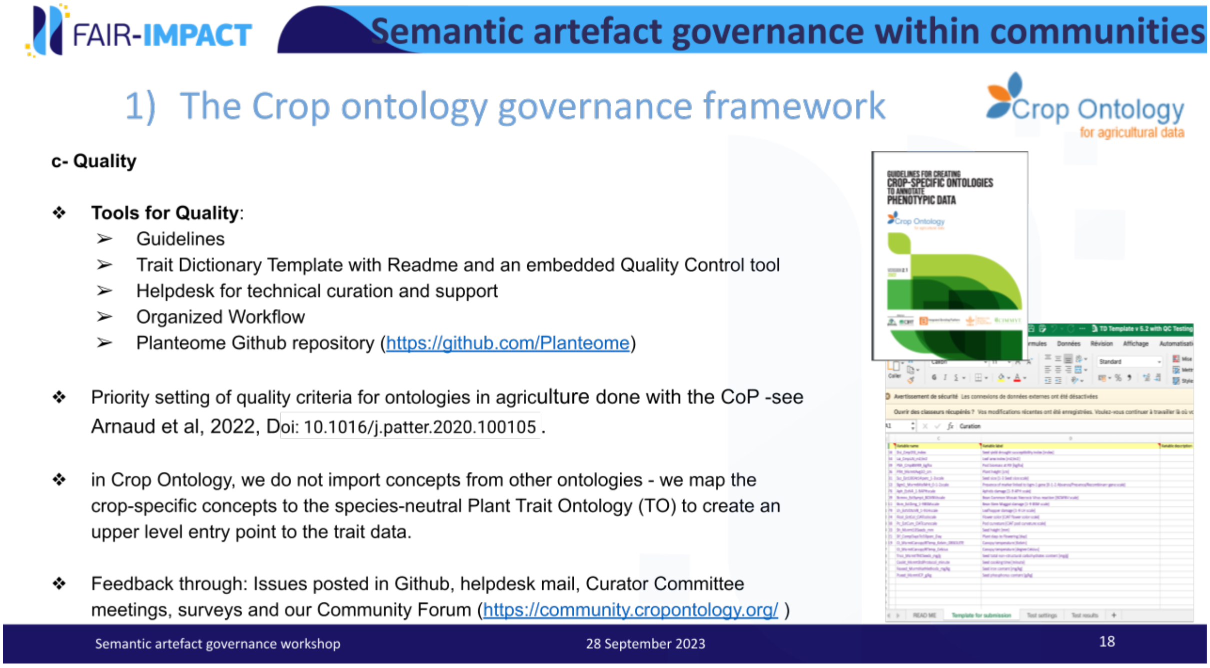

Figure 3.CO's quality verification and versioning process for ontologies, taken from the quality section slide.

2.1.4 Access

The CO maintains a commitment to openness and accessibility. It maintains a transparent approach with no formal terms and conditions, primarily governed by the governance and

[12] https://cgspace.cgiar.org/handle/10568/110906
[13] https://cgspace.cgiar.org/handle/10568/114769?show=full
[14] https://github.com/orgs/Planteome/repositories
[15] https://community.cropontology.org/
[16] https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=1cc36a8c1366976f0195f6de73e396f8d1081071

stewardship framework[17] and the CC-BY 4.0 License. CO offers readily accessible ontologies through APIs, integrating with the Breeding API and various in-house functionalities. Community engagement is fostered through diverse communication channels, including a forum, GitHub alerts, emails, CGIAR Ontology Working Group meetings, and webinars, alongside active participation in relevant conferences. CO's collaboration with the Alliance Bioversity-CIAT team underscores its dedication to support and sustain these ontologies, ensuring their ongoing availability and usability within the community. Furthermore, most of the crop ontologies of CO are indexed and replicated by two ontology repositories: the Ontology Lookup Service (OLS)[18] of the EMBL-EBI, and AgroPortal[19]. It is also indexed (metadata only) in FAIRsharing[20].

2.1.5 Lifecycle

The CO employs a robust lifecycle management approach, which has been coordinated by the Alliance Team since 2009. This approach includes ongoing maintenance for crop-specific ontologies, tool maintenance, versioned repositories on GitHub, and proposal writing. New terms require curator approval, and outdated terms are marked as 'OBSOLETE'. Currently, the ontologies are available in English only. Financial sustainability is ensured through support from research and data projects, including backing from entities such as the CGIAR Generation Challenge project, Climate Change CRP, NFS Planteome, etc, along with small grants for expert consultations and training from partners, securing the project's continued growth and sustainability.

2.1.6 Stakeholders and decisions

The decision-making process within the Crop Ontology project is a multi-leveled approach involving various stakeholders. Decisions to publish proposed ontologies are made by the Project Coordinator and Data Steward, while Curators and the Crop Expert Group are responsible for the development and updates of crop-specific ontologies. The extension of domains and the removal of obsolete terms in ontologies are determined by the Scientific Advisory Committee and the Curator Committee, and the Strategy Advisory Committee is consulted for strategic technical development decisions. Community consultation is ensured through the Community Forum, surveys, and webinars. Website improvements, which are the role of the alliance team, are guided by curator feedback. Through their community's tight and active communication, CO notices the updates, promotes their use, and receives feedback to improve adaptability to their users. The influence of project leaders offering financial support, guided by Advisory Committees, is significant in shaping the project's content direction and focus, especially in areas like on-farm trials and food technology.

2.2 OBO Foundry by Deepak Unni (time in video 26:38)

2.2.1 Principles

[17] https://cgspace.cgiar.org/handle/10568/118001
[18] https://www.ebi.ac.uk/ols/
[19] AgroPortal document , https://www.sciencedirect.com/science/article/pii/S0168169916309541
[20] https://fairsharing.org/

Funded by
the European Union

The OBO Foundry[21] started in 2003 with a goal of developing ontologies that are interoperable and scientifically accurate. To do so, the OBO Foundry organization focuses on four main aspects: developing standards for a unified representation of ontologies, developing infrastructure for effective and scalable ontology management, creating OBO principles[22] which are the baseline for the development of open and FAIR ontologies, and building an active community that facilitates the growth and development of shared best practices. OBO Foundry is also a registry of ontologies, referred to as the OBO Registry, and mainly focused on biological, biomedical and other domains that are adjacent to life sciences, such as environmental, with 184 active ontologies as of 2023. To ensure interoperability within the distributed ontologies, the OBO Foundry requires ontologies to have at least one OWL representation in RDF-XML syntax to ensure that ontologies share the same syntax. Even if additional formats are allowed, this "common format" (OBO Foundry principle n°2) is a mandatory principle for all registered ontologies. Regarding ontology exploration, OBO Foundry does not host its own ontology browser but provides link out to services such as EBI Ontology Lookup Service, BioPortal and Ontobee, where OBO Foundry ontologies are indexed and available for exploration.

2.2.2 Metadata

The OBO Foundry has a clear metadata management policy. They consider two types of metadata: 1) metadata about the ontology itself represented using standard vocabularies and 2) a YAML file that contains metadata about the ontology, as provided by the maintainers, and gives an overview of ontology's maintenance, versioning, usage, and activity (Figure 4).

Metadata elements:

- **id:** Unique name (typically the ontology prefix)
- **title:** The full name
- **description:** A short description of the ontology
- **domain:** The domain of the ontology
- **browsers:** Default browser for this ontology
- **contact:** Contact person
- **dependencies:** Other ontologies that are dependencies
- **license:** The license for the ontology
- **preferredPrefix:** The preferred prefix for term CURIEs
- **products:** Products that are created for this ontology
- **publications:** Relevant publications
- **repository:** The repository where the ontology is maintained
- **tracker:** The issue tracker for community engagement
- **usages:** Documented usage of this ontology
- **activity status:** Whether this ontology is still active

Figure 4. Examples of OBO Foundry metadata elements, extracted from the workshop presentation

The YAML file with metadata elements is used by the OBO Registry, where several metadata standards are used such as RDF Schema, SKOS, Dublin Core, PROV Ontology and facilitate ontology's discoverability and accessibility through machine-actionable metadata. Moreover, the OBO Foundry provides another level of metadata which came from their own ontology: OBO Metadata Ontology (OMO). OMO represents metadata that are used to annotate ontology terms and ontology metadata for all OBO ontologies.

2.2.3 Quality

The OBO Foundry principles are aimed at improving and sustaining an ontology's interoperability over time. To check compliance to the OBO Principles, the OBO Foundry has

[21] OBO Foundry website, http://obofoundry.org/

[22] http://obofoundry.org/principles/fp-000-summary.html

Funded by the European Union

implemented an OBO dashboard[23] which operationalized many of the OBO principles by running computational checks (Figure 5). A report for each principle and checks per ontology is provided, with recommendations on how to fix any violations of the principles. From 2024 onwards, passing all the checks of the OBO dashboard will be mandatory for all ontologies registered in the OBO Foundry. One key aspect of ontology' quality which OBO Foundry enforces is to import terms from other ontologies, especially when terms are outside the scope of an ontology. Gathering feedback and information from their community is well structured, with specific channels for each topic of discussion, such as the OBO Discuss mailing list and Slack for general discussion, or the GitHub issue trackers for specific discussions. If there are discussions that are specific to an ontology then these discussions are directed to the individual ontology issue tracker.

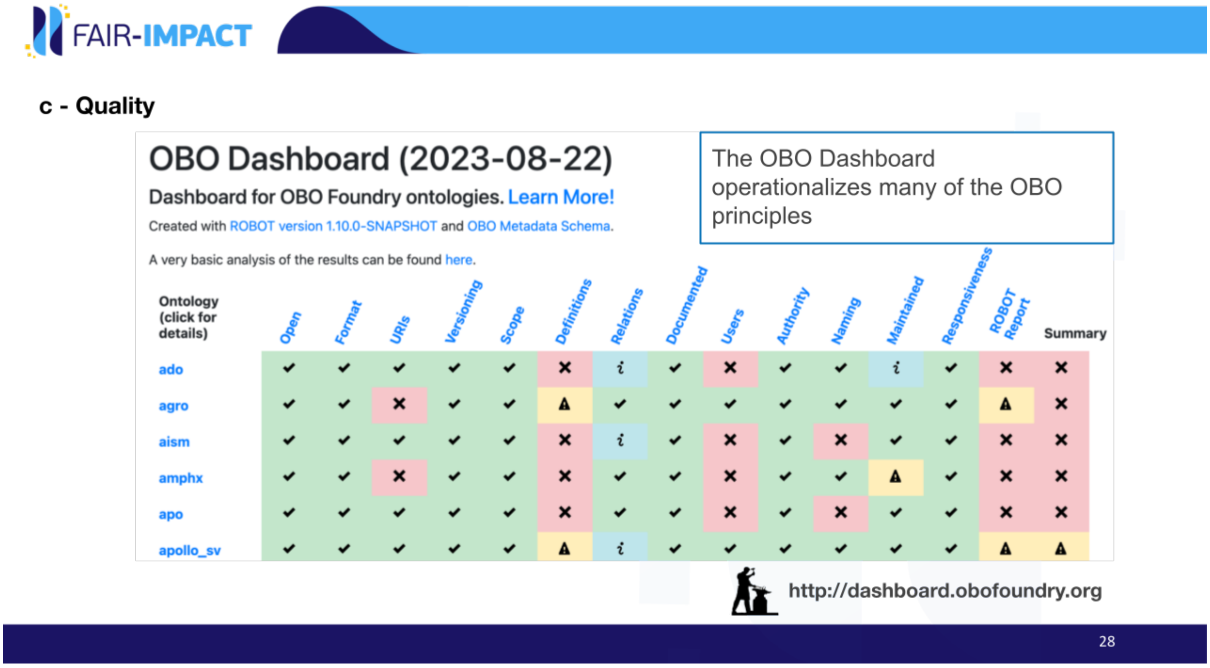

Figure 5. The OBO Foundry automated OBO Dashboard tool checks the ontologies' compliance with OBO principles, extracted from the quality slide.

2.2.4 Access

Ontology's access is conveyed by the first principle of OBO Foundry which is to be openly available without any constraint. To do so, ontologies are distributed via CC BY or CC0 license. OBO Foundry does not have its own service for ontology exploration but does keep track of individual ontology-specific endpoints and exploration services like OLS, BioPortal and Ontobee.

2.2.5 Lifecycle

The Ontology Development Kit (ODK)[24] is provided by the OBO Foundry, along with recommendations, on how to create and maintain an ontology by leveraging ontology tools like ROBOT, owltools, dosdp-tools. ODK has a set of executable workflows which can be used for various applications such as managing the ontology's continuous integration, development, quality control, managing imports from other ontologies and preparing new

[23] http://dashboard.obofoundry.org/

[24] https://github.com/INCATools/ontology-development-kit

Funded by the European Union

releases.

2.2.6 Stakeholders and decisions

OBO Foundry is governed by a volunteer team, - the OBO Operations Committee - mainly composed of ontology maintainers, engineers, and stakeholders.[25] This committee is divided into 3 working groups (WG): The Technical WG is mainly focused on the metadata aspect, within OBO and across OBO ontologies, with respectively curation and harmonization actions, and the website management. The Editorial WG is responsible for defining the wording of both current and new OBO principles, as well as the wording of Standard Operating Procedures (SOP) and Frequently Asked Questions (FAQs) on the OBO Foundry website. The Outreach WG is in charge of communication, which involves supporting mailing lists, preparing educational materials, and spreading news through the quarterly OBO newsletter. As governed by a volunteer team that can be fairly distributed, OBO Foundry also set up a Governance task team that advises the Operations Committee. This task team consists of members who are also part of the OBO Operations Committee but also outside the committee and makes recommendations for improving the governance with better processes and documents, such as the codification of a code of conduct. Recently, the OBO Foundry has once again improved the coordination of its governance but this time within the three WGs through a new function: the "role based system". This system clarifies responsibilities and scope according to well defined roles and empowers members of the OBO Operation Committee.

2.3 BASF by Paola Espinoza Arias (time in video 43:05)

2.3.1 Principles

BASF[26] is a chemical company engaged in the production, marketing, and sale of various chemical products. Its diverse product line includes solvents, adhesives, surfactants, fuel additives, electronic chemicals, etc. However, within the chemical sector, the challenge of standardization often arises due to the fragmented and heterogeneous nature of data, alongside ad-hoc practices and inadequate stakeholder coordination. To tackle these issues, the BASF Governance Operational Model for Ontologies (GOMO) framework[27] has been introduced. GOMO establishes consistent and standardized methodologies and techniques for ontology development, thus reducing ad-hoc practices and facilitating the reusability and interoperability of ontologies. These ontologies, which originate from collective stakeholder needs, operational divisions, and community groups, are centralized in a core catalogue. This catalogue currently houses 73 ontologies. This centralized approach streamlines access and utilization of ontologies.

2.3.2 Metadata

To comprehensively describe SAs, a set of mandatory and optional metadata fields have been established. These include essential information such as the title, creator, contributors, creation date, version details, preferred namespace URI, licensing terms, and ownership details. Additionally, optional elements encompass aspects like b

[25] https://obofoundry.org/docs/OperationsCommittee.html

[26] https://www.basf.com/fr/fr.html

[27](GOMO) framework publication, https://zenodo.org/records/7007495

Funded by the European Union

ackward compatibility, organizational units, and diagrams. To ensure standardization and compatibility, recognized metadata standards[28] including RDFS, SKOS, oboInOwl, DCTerms, Schema.org, and the GOMO metadata vocabulary –which is not currently available outside BASF– are employed. **2.3.3 Quality**

GOMO has established a set of ontology assurance methods to validate the necessary standards and ensure the quality of SAs (Figure 6). These methods is employed across 10 standards organized in six categories, 1) including Persistent Uniform Resource Locators (PURLs) to maintain resource stability, 2) deprecation procedures for obsoleting the OWL entities, 3) conventions for character sequences in identifiers, 4) mandatory metadata requirements, 5) version control repositories for tracking changes, and well-defined 6) documentation processes. Once the ontology is deemed ready for deployment in a production environment, the related pipeline is executed. If errors are detected during this process, a detailed report is generated, serving as a valuable guideline for resolving any issues. Additionally, GOMO provides multiple channels for collecting feedback from users and the community, including GitLab issue templates, email, and internal team communications, ensuring that user input is considered and issues are addressed effectively.

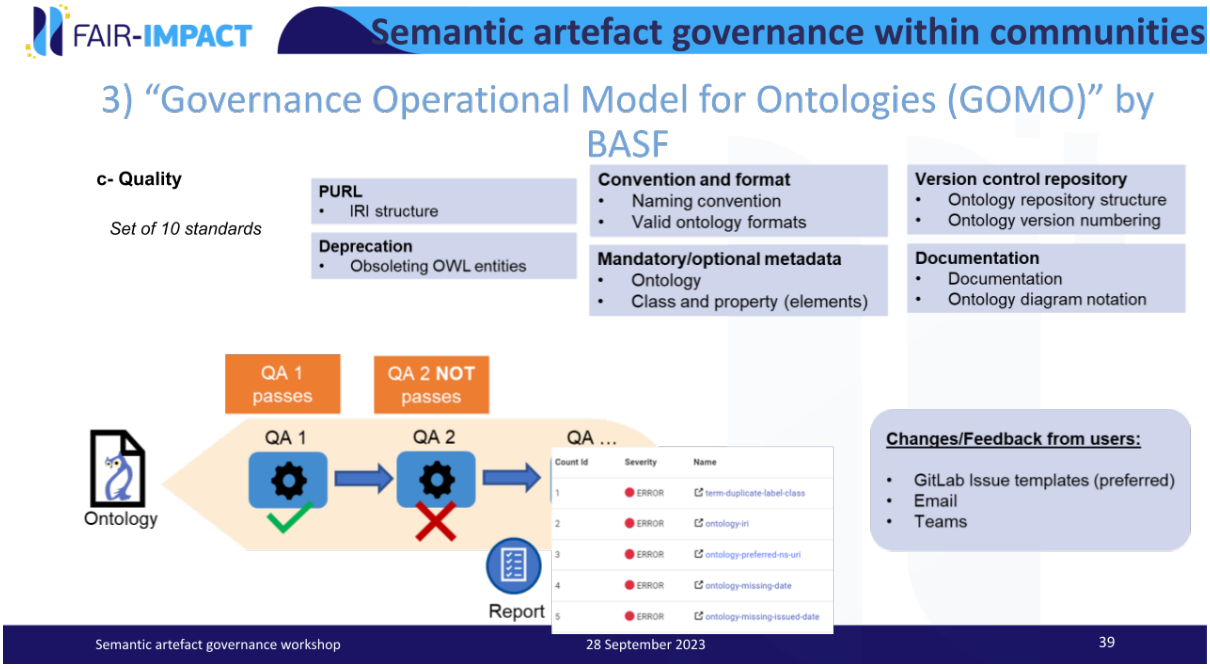

Figure 6. The six categories of GOMO guiding the development and maintenance of BASF ontologies, slide taken from the quality section.

2.3.4 Access

Ontologies are categorized according to confidentiality levels, and specific licenses are assigned to regulate accessibility based on these classifications. This ensures a structured approach to determining the level of access to the ontological data. Third parties are involved in creating these licenses. Machine-accessible endpoints, vital for ontology accessibility and usability, are supported by tools like OpenLink Virtuoso. This includes a specific instance of

[28]https://www.researchgate.net/publication/346881646_Best_Practices_for_Implementing_FAIR_Vocabularies_and_Ontologies_on_the_Web

Funded by the European Union

Virtuoso and a well-described REST API using OpenAPI specifications. BASF, in particular, has a legal department that provides clear terms and conditions.

2.3.5 Lifecycle

GOMO comprises four interconnected components—Governance Principles, Standards, Best Practices, Trainings, and Outreach—designed to support ontology development across its lifecycle stages. Aligned with The Linked Open Terms (LOT)[29] Methodology's workflow of Requirement, Implementation, Publication, and Maintenance, each stage involves various tasks governed by GOMO components (Figure 7). These components may apply universally or to specific steps; for example, the Documentation Principle is exclusive to the Publication stage, while the FAIRness Principle spans all stages. Furthermore, GOMO defines community roles and proposes through Responsibility Assignment Matrix (RAM), also known as RACI matrix, outlining the roles' responsibilities throughout the ontology development lifecycle. Importantly, GOMO adopts a comprehensive approach to ontology retirement or obsolescence, opting to deprecate specific terms rather than deleting entire ontologies, because the ontologies must persist. Deprecation is accompanied by mandatory metadata, including "deprecated" status and comments, as well as optional information like "replaced by" references.

Versioning within GOMO is managed in a Git-based environment, implementing standard software development practices like semantic versioning[30]. Regarding language, while English is the preferred language, GOMO allows the inclusion of other languages in labels and as synonyms, ensuring accessibility and usability across linguistic preferences. Ontology sustainability practices differ for core and domain ontologies; core ontologies receive funding from a common budget and are managed by a permanent team, while domain ontologies are financed by owners and maintained by domain-specific communities to ensure relevance and up-to-date content.

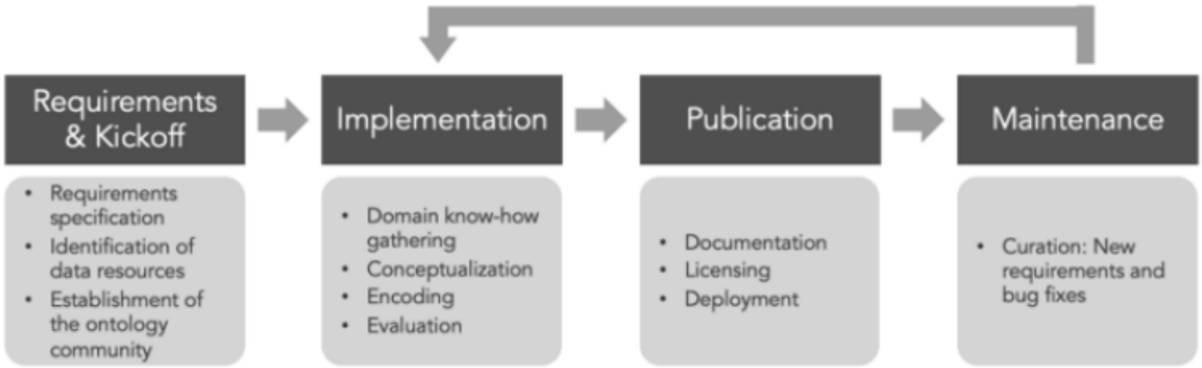

Figure 7. Ontology lifecycle developed in GOMO, figure extracted from Ana Iglesias-Molina and al.[31]

2.3.6 Stakeholders and decisions

For BASF, underpinning their governance model is the fundamental principle that all ontologies are supported by an associated community. To efficiently allocate responsibilities

---

[29] https://cdn.semantic-web.com/wp-content/uploads/2017/05/LOD-the-Essentials_0.pdf

[30] https://semver.org/

[31] Ana Iglesias-Molina and al. Ontology Management in an Industrial Environment: The BASF Governance Operational Model for Ontologies (GOMO). Zenodo. 2022. doi: 10.5281/zenodo.700749

Funded by the European Union

among different actors, a RACI Matrix is employed (Figure 8). The RACI model, derived from the four key responsibilities (Responsible, Accountable, Consulted, and Informed), is a valuable tool for clarifying and defining roles and responsibilities within cross-functional or departmental projects and processes. Responsible individuals are hands-on team members who actively contribute to the development and completion of project deliverables. Accountable parties ensure adherence to project deadlines and overall project completion, often falling into the informed category as well. Consulted individuals provide guidance and feedback, playing a role in decision-making, especially in areas like legal advice or organizational-wide projects. Informed persons, typically business owners or stakeholders, stay informed about project updates without direct involvement in decision-making, contributing to a project perspective. RACI ensures that everyone involved in the ontology development and management process understands their roles and contributions, leading to more effective collaboration and successful outcomes. Finally, the governance models are maintained by the Data Semantics Team at BASF.

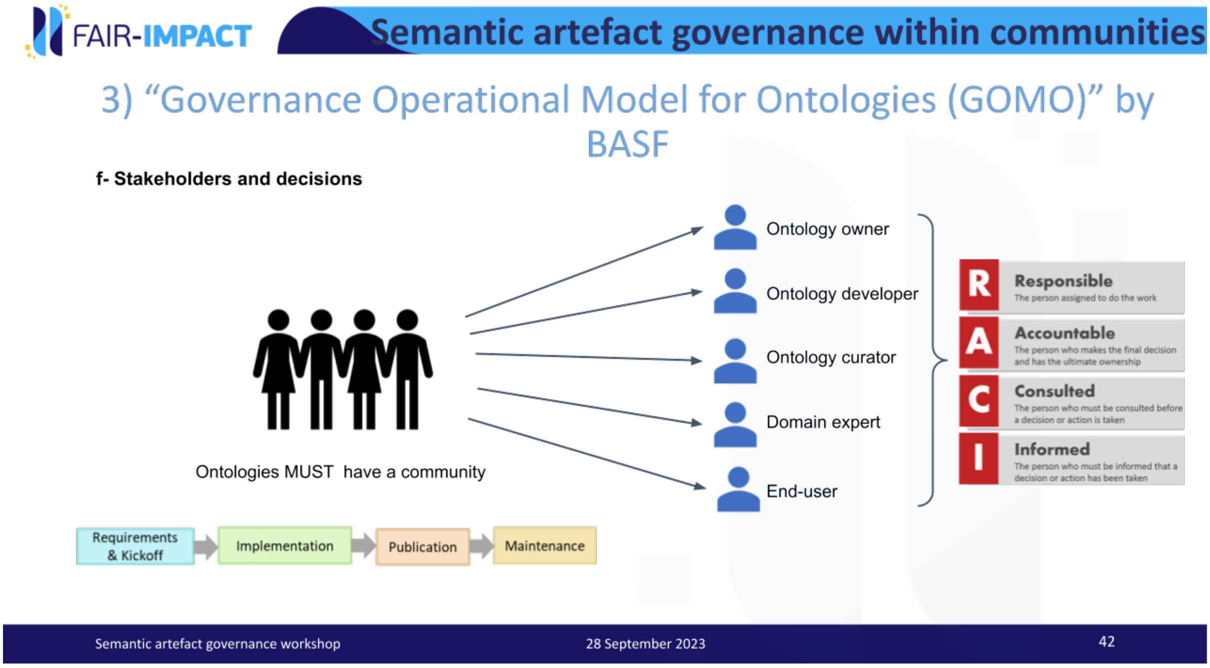

Figure 8.Roles and responsibilities defined in GOMO guiding who is involved and to what extent to each activity of the ontology lifecycle, slide taken from the stakeholders and decisions section.

2.4 INRAE Vocabularies by Sophie Aubin (time in video 56:41)

2.4.1 Principles

INRAE, France's National Research Institute for Agriculture, Food, and Environment, established in January 2020, is at the forefront of driving transformative changes in agriculture, food and environment scientific research. INRAE researcher teams produce[32] and use various SAs from Wheat Trait and Phenotype Ontology to the Terminology of French Bread Descriptors, each tailored to specific domains while support services work with more

[32] The INRAE group in AgroPortal contains 30 semantic artefacts to date.

Funded by the European Union

transversal ones like the INRAE Thesaurus[33]. The Vocabulaires Ouverts support service[34] assists INRAE collaborators in the creation and utilization of SAs.

2.4.2 Metadata

The minimal metadata requirements for semantic resources, as recommended, encompass essential details such as the semantic resource's name, information about the authors and contributors, including unique identifiers such as Open Researcher and Contributor Identifier (ORCID) or Research Organization Registry (ROR), if available, contact information, a descriptive text in both French and English, version specifics, and licensing information. To standardize this metadata, various standards are employed, including Metadata for the Ontology Description and Publication Ontology (MOD), the Data Catalogue Vocabulary (DCAT), Datacite, the Data Documentation Initiative (DDI), and Dublin Core. These standards are either already implemented in SA repositories or recommended by FAIRness assessment tools like O'FAIRe[35] and FAIR checker[36], ensuring consistency and adherence to best practices in metadata presentation and accessibility.

2.4.3 Quality

The SAs creator teams and experts are actively encouraged to adhere to good practices, guidelines, and high-level principles, rather than rigid policies, to enhance the quality of their creations. These principles include the FAIR principles, the Linked Open Terms methodology, and sometimes the OBO Foundry principles for open, sustainable, and interoperable ontologies. While the reuse of existing SA is recommended, it often remains challenging due to issues related to sustainability and the absence of well-defined practices for maintenance and user support. In cases where direct reuse is unfeasible, mapping to existing SAs is promoted, relying on the Simple Standard for Sharing Ontology Mappings (SSSOM) framework. To maintain high technical and syntactic quality, tools and methods are employed, including SKOS Play! Tester and VocBench ICV for thesaurus and SA catalogues such as AgroPortal validation for ontologies. Scientific quality is ensured through the involvement of editorial committees and the engagement of experts in the respective domains.

2.4.4 Access

INRAE's vocabulary adheres to the national policy "Plan pour la science ouverte" and follows the FAIR principles (Figure 9). The use of License Ouverte Etalab, equivalent to CC-BY 4.0, ensures an open and transparent framework for these SAs. To enhance accessibility, machine-accessible endpoints are available, including APIs accessible through AgroPortal, Skosmos, and Loterre, allowing users to programmatically interact with the data. For more advanced querying, Sparql endpoints are provided through AgroPortal, Loterre, and an institutional triple store. Effective communication with users is achieved through dedicated website and email channels. However, there is a recognized concern regarding the sustainability of these artefacts within research unit productions, highlighting the need for long-term support and relevance in the research community.

---

[33] https://consultation.vocabulaires-ouverts.inrae.fr/thesaurus-inrae

[34] https://vocabulaires-ouverts.inrae.fr/

[35] O'FAIRe document, https://hal-lirmm.ccsd.cnrs.fr/lirmm-03630543v3/document

[36] FAIR checker website, https://fair-checker.france-bioinformatique.fr/

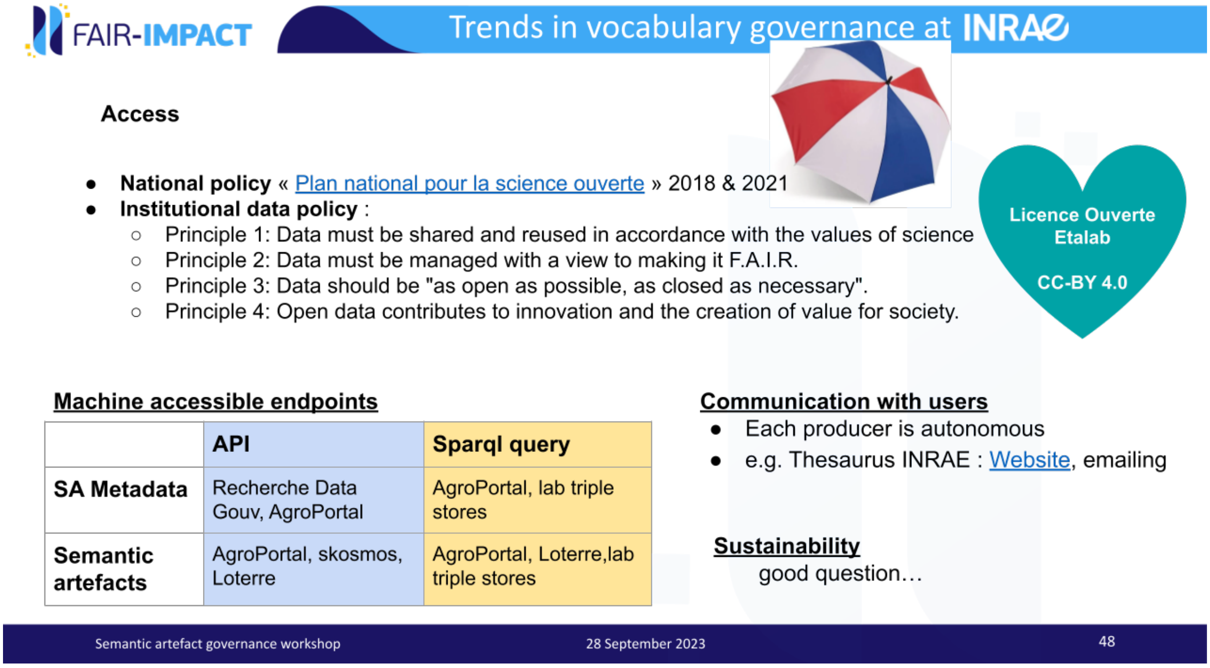

Figure 9. The SA access recommendations adhered by INRAE, taken from the access section

2.4.5 Lifecycle

In the life cycle of scientific artefacts as illustrated by the workflow for INRAE Thesaurus, users have the opportunity to initiate modifications to specific SA. The requests are processed by the dedicated Editorial Committee, which collaborates closely with domain experts throughout this task. This committee operates with the support of various tools and resources including editorial guidelines, regular meetings, task groups, shared spreadsheets, and VocBench. The management of concept obsolescence is achieved through the application of the OWL deprecated feature. This lifecycle typically involves the issuance of approximately three releases per year, ensuring that the artefacts remain up-to-date and relevant to the scientific community.

2.4.6 Stakeholders and decisions

In the process of developing and maintaining SAs, most decisions are made by the authors. On the other hand, INRAE science support services are more focused on providing guidance through guidelines and standards, relying on projects like FAIR-IMPACT. The subsequent table provides a concise overview of decision-making processes across various aspects and their respective decision-makers. AgroPortal is also becoming a more prominent platform at INRAE on which the institute will rely more and more in the future. Ultimately, AgroPortal, despite being an open and public platform; will be managed and driven by INRAE in the long term.

	How to take decisions	**Who is responsible**
Publication repository	Institutional recommendations	SA authors
Minimal metadata	Community recommendations	SA authors
Ontology quality	Community recommendations	SA authors/curators

License / access policy	European/National policy	INRAE
Maintenance	Community recommendations	SA authors/curators

Table 1. Stakeholders and decisions for semantic artefacts at INRAE, table extracted from INRAE's presentation.

2.5 IVOA by Baptiste Cecconi (time in video 1:10:34)

2.5.1 Principles

The International Virtual Observatory Alliance (IVOA), founded in June 2002, aims to foster global coordination and collaboration to facilitate the development and deployment of tools, systems, and structures crucial for the integrated and interoperable use of astronomical archives as a virtual observatory. The IVOA's objective is to create and maintain an interoperability framework for astronomy data, covering service registries, schemas, protocols, and vocabularies. IVOA working groups primarily produce two types of SAs: data models (schemas) and controlled lists of terms (vocabularies), occasionally drawing inspiration from external sources. These resources can be accessed via the following web pages: Schemas[37], Vocabularies[38], Specifications[39].

2.5.2 Metadata

The IVOA places great importance on their own metadata standards; an overarching data model, based on Unified Modeling Language (UML), is used to document IVOA schemas, allowing for domain-specific customization. Vocabularies are utilized to create controlled lists for standards for effective findability, accessibility, and reusability of astronomical data. These vocabularies are now being written using SKOS or OWL information and are made available in RDF/XML format, a significant step towards improved integration and interoperability within ontology systems and engines, albeit still in a prototype phase.

2.5.3 Quality

In the IVOA standards process, a structured sequence of steps is employed to ensure the review and consensus-building for progressing documents from Working Drafts to Proposed Recommendations and, ultimately, IVOA Recommendations. This process involves document preparation, review, response to comments, final voting, and approval by the Executive Committee, culminating in the status of IVOA Recommendation (Figure 10). Notably, IVOA utilizes a diverse feedback collection mechanism encompassing the Vocabulary Enhancement Proposal (VEP) process for vocabulary-related discussions. Furthermore, it actively solicits input through mailing lists and GitHub issues in the document development repository.

[37] https://www.ivoa.net/xml/

[38] https://www.ivoa.net/rdf/

[39] https://www.ivoa.net/documents/

Funded by the European Union

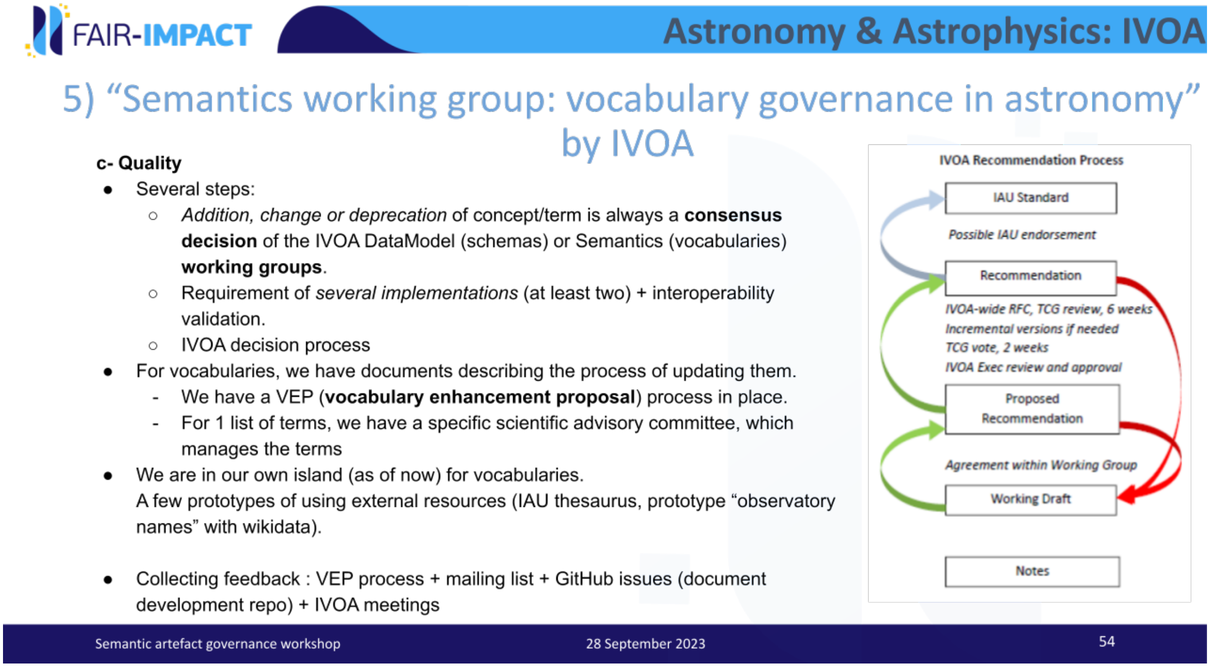

Figure 10. Slide depicting the IVOA's decision process for the SA quality assessment, taken from the quality section.

2.5.4 Access

IVOA licenses its vocabularies under CC0 (Creative Commons Zero) and associated documents under CC-BY (Creative Commons Attribution). Schemas related to specific documents follow the document's license. Vocabularies are accessible through Internationalized Resource Identifiers (IRIs), allowing users to resolve to specific controlled lists. While there's an OntoPortal prototype –experimented in the context of FAIR-IMPACT T4.2–, it's not yet integrated into a workflow, and formal notification plans are lacking. Community interaction mainly occurs via mailing lists. IVOA's sustainability for vocabulary development relies on project-based funding, with no dedicated funding at the organizational level.

2.5.5 Lifecycle

In the context of the life cycle, the maintenance process for IVOA vocabularies involves the submission of VEP, where community discussions ensue until a consensus is reached on new terms, definition updates, or deprecation. Following consensus, the vocabulary is updated, and the IVOA Technical Coordination Group (TCG) endorses this update, potentially prompting additional discussions at the IVOA level. Schemas adhere to a stringent versioned release process, utilizing GitHub issues and mailing lists for proposing, discussing, and drafting new specifications.

Vocabulary versioning is primarily based on the release date, with the IRI format incorporating this date. The retirement of terms in vocabularies is infrequent. Conversely, there is no formal requirement for the retirement or obsolescence of schemas. Major schema versions are typically updated to accommodate new features, and while some older services may not receive active maintenance, they continue to function, leading to client implementations often retaining older versions of protocols. Lastly, it's noteworthy that all labels and terms in IVOA vocabularies are in English.

2.5.6 Stakeholders and decisions

Decision-making within the IVOA is characterized by a bottom-up approach, where consensus is cultivated within working groups involving diverse stakeholders, including software developers, researchers, data curators, and more. This consensus is typically reached within working groups; if needed, a vote occurs within the TCG and the Executive Committee (Exec). The TCG comprises chairs and vice-chairs of working groups, with renewal every three years, while the Exec includes representatives from national and regional virtual observatories. Additionally, the Committee on Science Priority plays a crucial role in gathering science requirements from communities, shaping the development of standards, and ensuring alignment with real-world needs. The decision process can be found on a published document[40].

2.6 NFDI4Biodiversity by Naouel Karam (time in video 01:26:29)

2.6.1 Principles

NFDI4Biodiversity, part of Germany's National Research Data Infrastructure (NFDI), involves around 50 partner institutions, including research institutions, museums, IT services, etc. A key component of the NFDI4Biodiversity project is the Research Data Commons cloud-based platform, known as NFDI-RDC, conceived as an expandable infrastructure empowering data sharing and data-centered projects. This facilitates users to store, analyze, and share diverse data types. A self-contained collection of data sets within the RDC is referred to as a *data product* serving for a specific analytical purpose.

The NFDI4Biodiversity initiative covers various SAs, such as structured taxonomies of living organisms and geographical ontologies, essential for biodiversity research. These resources are hosted within their OntoPortal instance named BiodivPortal[41].

2.6.2 Metadata

NFDI4Biodiversity employs metadata and data standards like Access to Biological Collection Data (ABCD) and Ecological Trait-data Standard (ETS) to organize and document diverse data types related to biodiversity. By leveraging DC, SKOS, and OWL, NFDI4Biodiversity can provide structured, standardized, and semantically rich metadata and knowledge representation for biodiversity data, making it accessible and interoperable for researchers and users in the biodiversity community.

2.6.3 Quality

NFDI4Biodiversity maintains a robust quality assurance approach, particularly for taxonomies and OWL ontologies (Figure 11). It involves a standardized pipeline to transform taxonomies into OWL, ensuring the format and the content accuracy of OWL ontologies. The quality of the content of original taxonomies are upheld through collaboration with project partners and domain experts. OWL ontologies undergo validation with reasoning tools to ensure logical consistency. Release notes are enforced, guaranteeing transparency. Best practices for reusing reference ontologies are promoted. The FAIR principles are assessed with the O'FAIRe tool. User feedback is actively collected through GitHub issues. Centralized workflows through BiodivPortal are planned to ensure consistent quality standards across the project.

---

[40] https://www.ivoa.net/documents/DocStd/20170517/

[41] BiodivPortal, https://biodivportal.gfbio.org/

Funded by the European Union

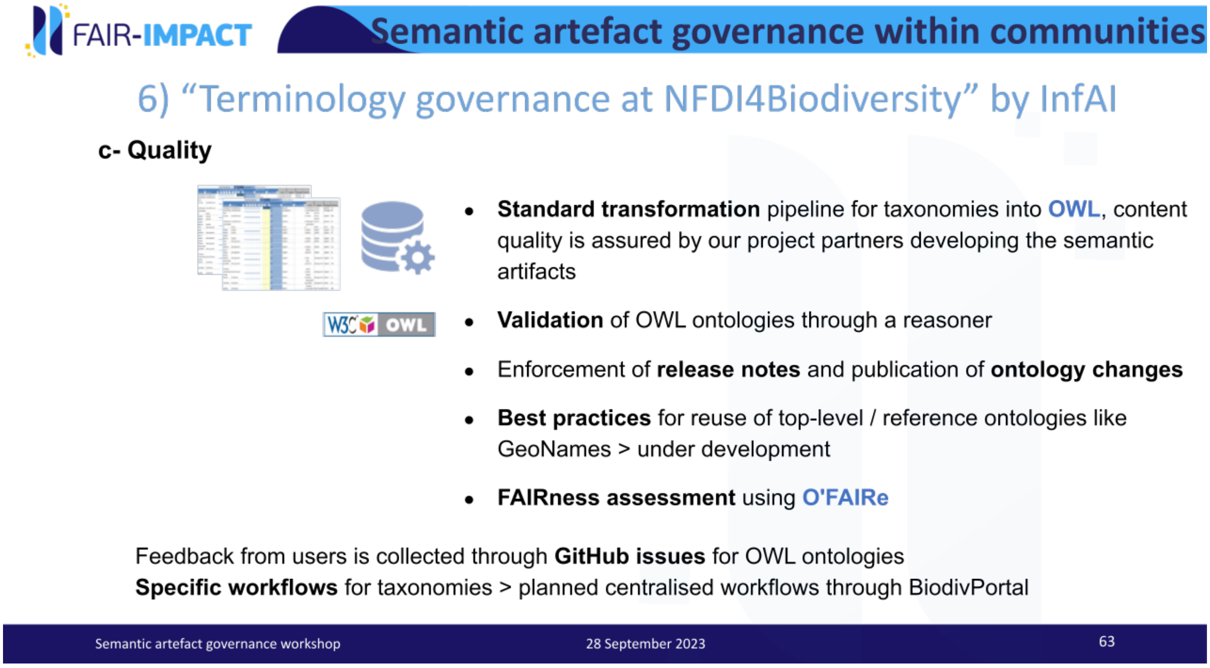

Figure 11. Slide of NFDI4Biodiversity ontology transformation process, from the quality section .

2.6.4 Access

NFDI4Biodiversity prioritizes widespread access to its SAs, employing a CC-BY license for sharing and adaptation. Access is facilitated through a SPARQL endpoint, the legacy GFBio Terminology Service (TS), and BiodivPortal. Communication and updates are managed via mailing lists, and financial support from institutions and national agencies ensures sustainability, enabling ongoing development and accessibility of these resources.

2.6.5 Lifecycle

NFDI4Biodiversity makes it a point of honor to control their SA lifecycle. Maintenance is a collaborative effort, with each institution maintaining its data and plans for term suggestion workflows. Periodical releases on GitHub keep ontologies up to date. Retirement or obsolescence of terms is handled by linking them to accepted alternatives. Versioning includes release dates for taxonomies and Semantic Versioning for ontologies. Multiple languages, including English, German, and numerous other languages for some of geographical ontologies, ensure broad accessibility and usability.

2.6.6 Stakeholders and decisions

Taxonomies Editorial Committees oversee taxonomy-related decisions and quality control. Task Groups like the ABCD Task Group work collaboratively to establish consensus on data and metadata standards. Ontology Managers are responsible for ontology structure, content, and versioning decisions. Experts in biodiversity domains contribute domain-specific expertise and validate artefacts. Taxonomies Editorial Boards ensure the quality and accuracy of taxonomic information by making informed decisions about taxonomy updates and changes. Task Groups drive decisions on standards and best practices for biodiversity data through consensus-building processes

2.7 EMBL-EBI ontology by Zoe Pendlington & Henriette Harmse (time in video 01:36:15)

2.7.1 Principles

As part of the European Molecular Biology Laboratory, the European Bioinformatics Institute (EBI) is an international organisation dedicated to research and services in bioinformatics. EMBL-EBI's primary mission is to translate data into knowledge and to accomplish this, EMBL-EBI relies on the Experimental Factor Ontology (EFO). However, EMBL-EBI complements EFO with over 30 other ontologies to meet diverse use cases across various data types and domains.

EMBL-EBI's EFO, an open-access application ontology, is crucial for describing experimental variables in molecular biology, covering disease, anatomy, cell types, and more. The Ontology Lookup Service (OLS) streamlines access to diverse ontologies, and OxO facilitates efficient mapping between terms for improved data integration.

2.7.2 Metadata

In the OLS, all ontologies are imported from OWL 2 RDF ontologies, adhering to the standards established by The World Wide Web Consortium (W3C). Each ontology available in OLS is required to provide the following essential metadata:

- A Persistent Uniform Resource Locator (PURL) that serves as the download link for the ontology;
- An exclusive and easily recognizable abbreviation or prefix assigned to the ontology. This unique identifier is used for efficient searching and integration into data processing pipelines.

While the above metadata is mandatory, OLS encourages ontologies to include the following additional information:

- Details about the licensing terms and conditions associated with the ontology
- A concise title for the ontology, providing a brief description of its purpose.
- A more comprehensive narrative that describes the ontology's content, scope, and relevance.

Regarding mapping between ontologies, OxO is transitioning to use the Standard for the Scientific and Scholarly Open Mapping (SSSOM). This standard defines a structured and consistent way to specify mappings between ontologies.

2.7.3 Quality

To maintain the quality of the EFO ontology, a multi-faceted approach is employed. This includes continuous integration on GitHub, where processes regularly test the ontology to detect and resolve issues. Changes to the ontology are made through pull requests, allowing for discussion, review, and validation before merging. Local testing via the Ontology Development Kit[42] is also conducted at each release to ensure high quality.

EFO adheres to OBO Foundry principles, emphasizing openness and interoperability. It actively imports from domain ontologies to enhance coverage and accuracy, promotes the addition of synonyms and cross-references for improved usability, and contributes to domain ontologies, enriching the broader ontological landscape. In terms of reuse and import, dynamic imports efficiently incorporate external ontologies, ensuring accurate and up-to-date data integration. EFO maintains an effective feedback system through a user mailing list

[42] https://www.researchgate.net/publication/364273465_Ontology_Development_Kit_a_toolkit_for_building_maintaining_and_standardizing_biomedical_ontologies

for discussions and feedback, along with GitHub's issue tracking system for problem reporting and managing discussions.

2.7.4 Access

EFO, OLS, and OxO are licensed under Apache 2.0, promoting open access with certain permissions and limitations. These resources find extensive use across nonprofit and for-profit organizations, demonstrating their broad applicability.

Communication mainly happens through mailing lists, fostering discussions among users and stakeholders. Periodic user events further facilitate face-to-face interactions. These resources primarily receive external funding, underscoring their significance in scientific and biomedical applications. Internal funds are also allocated due to their critical role in supporting research and data integration at EBI.

2.7.5 Lifecycle

The EFO operates as a user-driven system, primarily fueled by term requests from curators aimed at enhancing its hierarchy to suit data requirements (Figure 12). This life cycle begins with term request initiation, where users, often curators, propose new terms for EFO. An assessment determines whether the term is available in other ontologies through services like OLS; if so, it is imported or referenced. When not found in external ontologies, EFO generates a new term with a unique identifier, mapping it to the source term. Monthly updates keep EFO current, crucial for EBI databases, while an obsoletion process ensures that retired terms are appropriately handled, marked as obsolete with replacement or comments to prevent data disruption.

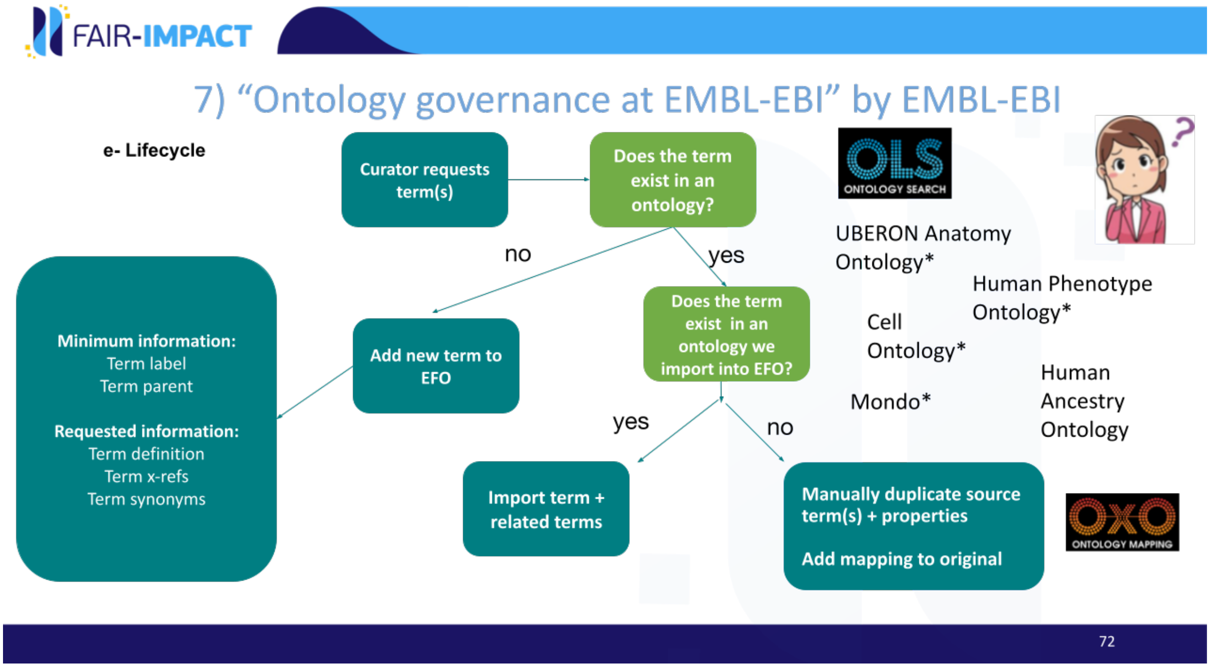

Figure 12. Slide of terms curation process of the EMBL-EBI using EFO, OLS and OXO tools, extracted from the lifecycle section.

2.7.6 Stakeholders and decisions

Stakeholders and decisions for EFO, OLS, and OxO are distinctly influenced by the expertise of various contributors. EFO relies on curators and domain experts for precision, while OLS predominantly involves developers focusing on functionality and performance, with additional input from curators and experts. OxO's changes are largely driven by developers, along with insights from curators and domain experts regarding ontology mapping. Prioritization leans on funders' guidance, with community feedback playing a crucial role. Occasionally, the Principal Investigator (PI) provides strategic direction and development decisions for these resources.

2.8 ETSI SAREF ontologies by Maxime Lefrançois (time in video 01:52:08)

2.8.1 Principles

SAREF, the Smart Applications REFerence ontology, is a shared model of consensus that facilitates the matching of existing assets in the smart applications domain. SAREF is modular by nature, and each module is versioned. SAREF includes a core ontology, domain-independent reference ontology patterns, and extensions for verticals (Figure 13). The documentation portal accessible at https://saref.etsi.org/ enhances its value for stakeholders.

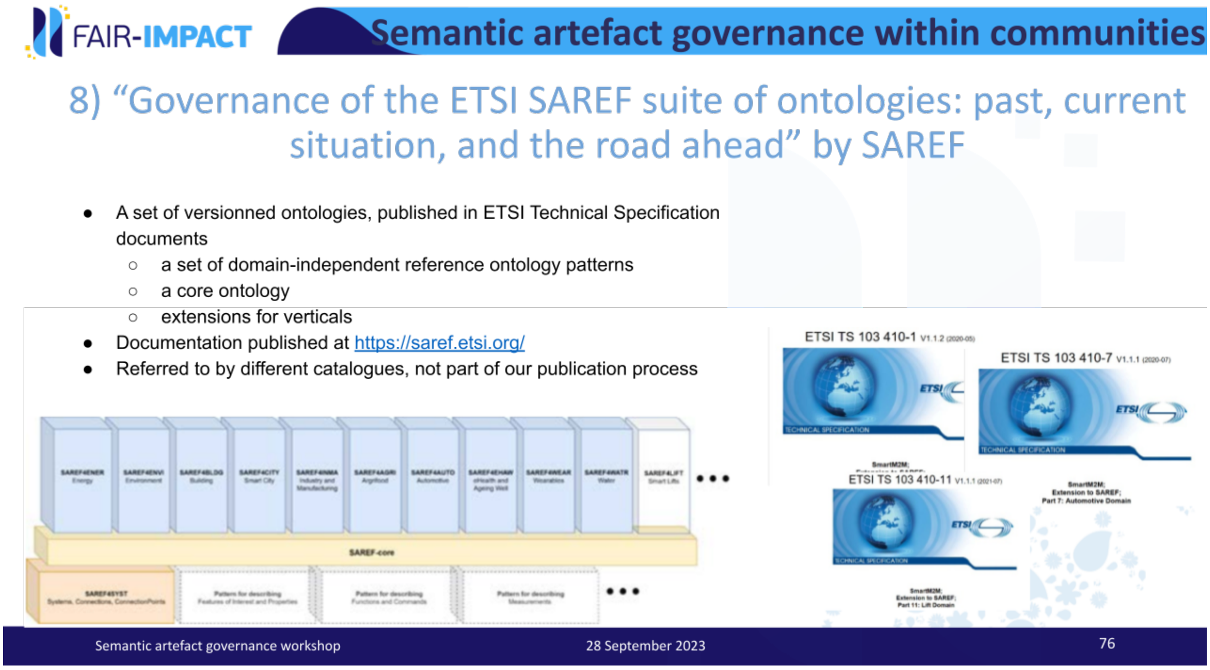

Figure 13. Work items of ETSI governance, extracted from the ETSI SAREF presentation.

2.8.2 Metadata

Metadata standards, such as OWL, Dublin Core Terms, and vocabulary for annotating vocabulary descriptions (VANN), provide comprehensive information about the ontology, while individual term metadata is described using RDF Schema (RDFS), a widely accepted language for characterizing SAREF's terms.

2.8.3 Quality

The SAREF-Pipeline software offers automated testing of SAREF and its extensions, ensuring correctness and compliance with the SAREF development framework specified in ETSI Technical Specification 103 673. Users can use it through a graphical user Interface or a command-line interface. It is configured to run in continuous integration and continuous delivery pipelines on the SAREF portal. SAREF emphasizes reusing terms and established standards like SOSA/SSN, OWL-Time, and GeoSPARQL. Feedback and issue tracking occur via

Funded by the European Union

the ETSI Labs platform https://saref.etsi.org/sources/, enhancing community engagement and issue resolution.

2.8.4 Access

SAREF is released under a BSD-3 License. Currently, contributions to the ETSI Labs platform are limited to ETSI members, but opening non-ETSI members is on its way. There is no SPARQL endpoint for SAREF, but a term lookup service is provided on the SAREF documentation portal. There is no direct user communication, but it is possible for ETSI Labs users to rely on GitLab's notification capabilities.

2.8.5 Lifecycle

ETSI TS 103 673, "SAREF Development Framework and Workflow" specifies guidelines for SAREF ontology development and management, covering crucial components. It defines the roles and responsibilities of actors involved, from developers to maintainers. The document outlines three distinct workflows: proposing new extensions or versions, development, and release. Additionally, it specifies how the git repository should be structured, how the sources of the ontology and the examples should look like, which metadata should be used on the ontology and each term, among other things.

The work on SAREF is governed by the ETSI SmartM2M technical committee through a program that consists of work items, each leading to an ETSI technical document, or a new version of an existing ETSI technical document. Work occurs on the ETSI Labs platform, emphasizing consensus among participating institutions. The workflow adheres to ETSI TS 103 673, with no explicit mention of retirement, but standard deprecation practices. Semantic versioning guides updates, and version management is facilitated through version branches. The documentation portal features redirections for version accessibility, and language tags, at least "en" (English), are consistently incorporated into the ontology.

2.8.6 Stakeholders and decisions

SAREF's governance is overseen by the European Commission and the ETSI SmartM2M Technical Committee. It employs a modular ontology structure with individual versioning for flexibility. Governance is inherited from ETSI, following their established processes. Development adheres to ETSI Technical Specification 103 673. Contributions originate from diverse sources, including EU projects and industrial associations, enhancing the ontology to meet evolving smart applications requirements. Different task forces were funded by the European Commission, and ETSI supports ongoing maintenance. Ongoing task forces STF 641 and STF 653 focus on (1) drafting a European Norm on SAREF (EN 303 760 SAREF Guidelines for IoT Semantic Interoperability), (2) investigating potential extensions for Digital Twins, (3) homogenizing and facilitating the use of SAREF and existing extensions using common ontology patterns.

2.9 FAO Agrovoc by Imma Subirats-Coll (the recorded presentation link is included on the Zenodo of this report)

2.9.1 Principles

The FAO Multilingual Agricultural Thesaurus (AGROVOC) is a global, multilingual resource covering topics related to FAO's areas of interest. Developed collaboratively by 34 organizations from 24 countries, it's monthly updated for relevance. Accessible in multiple languages and as a Linked Open Data (LOD) set, hosted by the University of Tor Vergata, it ensures standardized and interoperable access.

2.9.2 Metadata

AGROVOC's metadata includes crucial elements: dataset name, licensing terms, a concise but informative description, publisher details, creation and modification dates, a download link, and comprehensive content statistics, ensuring consistency through standards like DublinCore, FOAF, VoID, and LIME, crucial for a multilingual thesaurus.

2.9.3 Quality

AGROVOC prioritizes SA quality through technical adherence to standards, scope refinement, and multilingual localization. Quality improvement work is continuous to improve and maintain coherence, while also needing to retain legacy data. Tools such as SPARQL are used to identify outliers or data that does not follow standards (strange characters, duplicate labels) or incorrect input, and consultation with experts, emphasizing precise terminology localization. Standards and guidelines followed include ANSI/NISO Z39.19-2005 and ISO 25964. AGROVOC commits to consulting subject matter experts and primary authorities like the International Committee on Taxonomy of Viruses.

AGROVOC adopts a collaborative approach, encouraging editors to consult other thesauri for equivalence and linguistic accuracy. Alignments to selected thesauri enhance interoperability within a broader semantic landscape. For continuous improvement, AGROVOC employs a dynamic feedback mechanism, gathering user inputs through a collaborative editing platform, email, an editorial community, annual meetings, and other online interactions. Technical questions are resolved through exchanges with editors, and statistical analysis and regional task forces, like in Latin America, provide insights for enhancements and expansion options

2.9.4 Access

AGROVOC employs a straightforward CC-BY 4.0 license for legal clarity and accessibility. It offers machine-accessible endpoints through Skosmos, including Web Services and a SPARQL endpoint, along with a user-friendly Search & Browse interface. Communication happens through a mailing list, website, and social media. FAO oversees maintenance, including technical management and editorial coordination. Collaborative efforts involve multiple institutions, and volunteer contributions enhance content utility.

2.9.5 Lifecycle

AGROVOC maintains continuous updates through a collaborative team of editors using VocBench (Figure 14). Proposed changes undergo thorough review and validation, with batch imports considered exceptionally. User suggestions are welcomed via email, and monthly releases provide downloadable files and lists of new labels by language. The infrastructure is consistently updated. For retirement, concepts are marked as deprecated, signaled by label removal, dct:isReplacedBy references, and a historical note. While lacking explicit versioning, the latest release is available monthly, encouraging users to utilize the most current data.

Funded by
the European Union

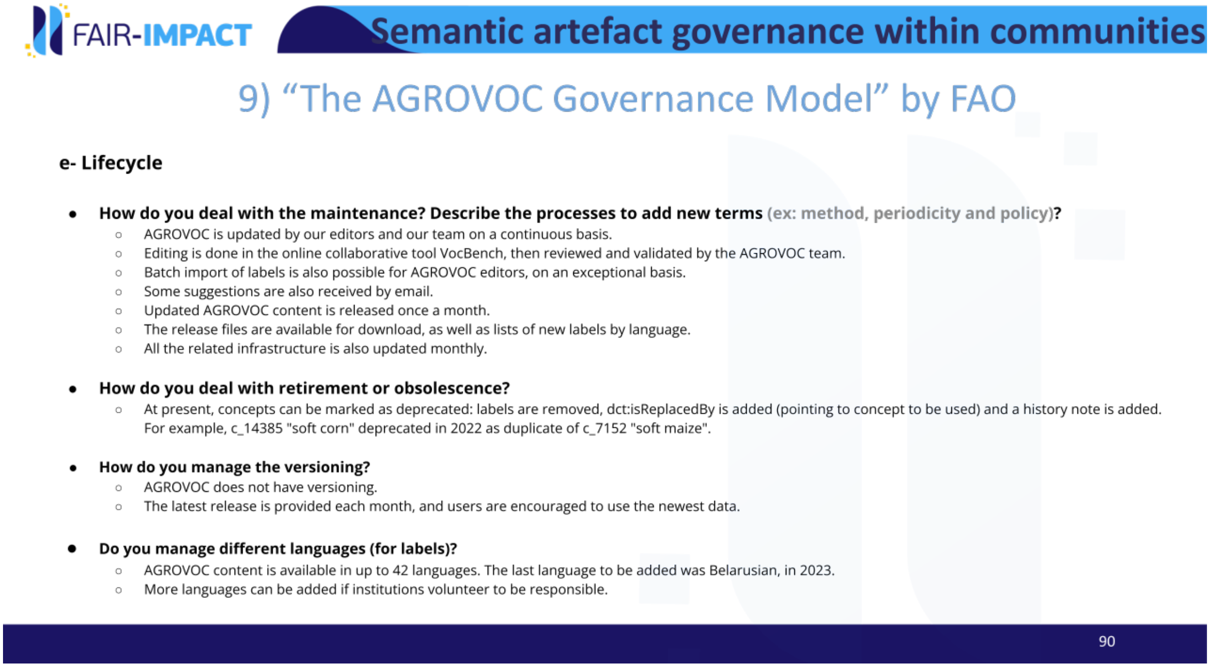

Figure 14. Ontologies's features managed by AGROVOC, from the lifecycle section.

2.9.6 Stakeholders and decisions

In AGROVOC's governance, key stakeholders include FAO, The Kuratorium für Technik und Bauwesen in der Landwirtschaft (KTBL), and Tor Vergata University for development, while curatorial responsibilities involve FAO and a team of 40 experts from 34 organizations across 24 countries. The core board comprises the FAO and KTBL team, and decision-making follows established guidelines, seeking consensus through annual meetings of the AGROVOC Editorial community. External experts are consulted for technical terminology, and individual editors are engaged as needed. For single concepts, the AGROVOC/KTBL team maintains consistency and coherence in line with AGROVOC Editorial Guidelines, continually improving curation workflows.

3 Conclusion and next steps

Our decision domains framework has adequately covered the governance aspects of all nine interviewed communities, as each community has succeeded in answering our designed SA governance questions. Thus, we've been able to explore the diversity of SA governance initiatives revealing a rich landscape of principles, quality assurance practices, metadata strategies, access mechanisms, lifecycle management, and stakeholder dynamics. To summarise,

1. The Crop Ontology demonstrated a robust governance framework, emphasizing quality, accessibility, and stakeholder engagement. Through a well-defined structure involving advisory committees and clear quality management practices, CO ensures the accurate curation and versioning of ontologies (Figure 3. Extracted from the quality section).
2. OBO Foundry stands out for its dedication to interoperability and scientific accuracy. The organization manages a library of ontologies with a focus on life sciences,

employing documented principles to ensure uniformity (Figure 5. Extracted from the quality section).
3. BASF's GOMO framework establishes a set of principles, standards and best practices for ontology development, all supported by a community-driven governance approach (Figure 6 & 8. Respectively, extracted from quality and stakeholders sections).
4. INRAE Vocabularies showcased a decentralized approach to governance. The institute encourages open science , including FAIR principles (Figure 9. Extracted from the access section), adherence to best practices, and employs tools to ensure technical and scientific quality).
5. IVOA stands out for its commitment to fostering global collaboration and interoperability in the field of astronomy, employing structured standards within the consensus decision process (Figure 10. Extracted from the quality section).
6. NFDI4Biodiversity prioritizes sharing data, services and tools through extensive collaboration among stakeholders in the biodiversity research domain-standard transformation pipeline (Figure 11. Extracted from the quality section).
7. The EMBL-EBI aims to generate knowledge through data, and to do so they implement multiple tools to reach this objective, such as by developing ontology's experimental factor, search and mapping (Figure 12. Extracted from the lifecycle section).
8. ETSI SmartM2M develops SAREF, a modular and versioned ontology for the IoT. They established a development framework and workflow which is supported by the SAREF Pipeline software and the SAREF Portal (Figure 13. Extracted from principles section).
9. AGROVOC's governance is a collaborative process allowing flexibility in versioning and maintenance of their ontologies due to their board community and lifespan data, which some dates back to 1981 (Figure 14. Extracted from lifecycle section).

Common themes across these initiatives include a commitment to open access, adherence to standards, and a strong emphasis on community engagement and feedback mechanisms (Table 2). These diverse case studies underscore the importance of governance in ensuring the effectiveness, quality, and sustainability of SA.

Thus, our next step in T4.1 is to compare all these SA initiatives and derive a set of main models and guidelines on 'SA Governance'. Upon examining the diversity of SA practices in the surveyed communities, we will be able to identify key and recurring aspects of governance. By extracting the foundations of SAs governance, we would be able to generate initial requirements and models. Thus, the modeling governance of SAs based on community-based stewardship practices will facilitate the synchronicity of ensuring connected SAs across scientific communities. The goal will be to propose a set of best practices that any new community, project or infrastructure coming to EOSC could decide to adopt or embrace for the management of their own SAs.

Funded by
the European Union

Table 2.Comparative analysis of SA governance across interviewed communities

Community	Metadata Standards	Repositories	Quality assessment	Feedback	lifecycle management approach
CO	-No specific standard	-Agroportal -OLS -ELIXIR	-Guidelines -Control tool -Helpdesk -Workflows	-Github -helpdesk -meetings -Forum	- Team decisions - Maintenance decisions: ontology, versioning via github and website - Terms decisions: new submission or removal - Financially supported by research and data
OBO Foundry	-RDFS -PROV-O -XSD -SKOS -DC -FOAF -DOAP -OMO	-OLS -BioPortal -Ontobee	-The OBO Dashboard Tool	-Discuss mailing list -Issue Tracker -ontology tracker -GitHub	- Ontology development kit (ODK) - Automated tool, leverages from other ontology tools like ROBOT, owltools, dospd-tools - Functionalities : continuous integration, development, quality control, managing imports from other ontologies and preparing new releases
BASF	-rdfs -skos -oboInOwl -dcterms -schema -GOMO	-BASF core catalogue (OLS-based)	-Workflows	-GitLab Issue -Discuss mailing list	- GOMO workflow of Requirement, Implementation, Publication, and Maintenance - Outlining the roles' responsibilities throughout Responsibility Assignment Matrix - obsolescence method - Semantic versioning
Voc INRAE	-MOD -DCAT -Datacite -DDI -DC	-AgroPortal -BioPortal -LOTERRE	-FAIR principles -Linked Open Terms -OBO Foundry principles -SKOS Play! Tester -VocBench ICV -Protégé -Ontoportal validation	-editorial committees -contribution of experts	- editorial guidelines - regular meetings - task groups - shared spreadsheets - VocBench - obsolescence method - Three releases per year

Community	Metadata Standards	Repositories	Quality assessment	Feedback	lifecycle management approach
IVOA	-their own standards -SKOS -OWL	-Prototype OntoPortal	-Workflows -vocabulary enhancement -proposal process	-mailing list -GitHub issues	- Consensus decision through IVOA standards process - Maintenance new terms, definition updates and deprecation - Vocabulary versioning and versioned release process
NFDI 4 Biodiversity		-BiodivPortal -gfbio	-FAIRness assessment -Standard transformation pipeline -reasoner	-GitHub issues -GitHub issues -Specific workflows	- Term suggestion and obsolescence through BiodivPortal workflows - Distributed Maintenance by institutions editorial committee - Periodical versioning and releases
EMBL-EBI	-RDF -OWL 2	-OLS	-integration on GitHub -tests run via ODK -Good Practices -Reuse and import	-Efo-users mailing list -GitHub issues	- Term suggestion and obsolescence through term requests process - Monthly updates
ETSI SAREF ontology	-rdfs -owl -dcterms -vann	-their own Term Lookup Service	- ETSI TS 103 673 - SAREF-Pipeline software - reuse of some identified standards	-ETSI Labs	- Three distinct workflows based on the SAREF development framework: proposing new extensions or versions, development, and release. - Consensus decision in ETSI Labs - Semantic versioning
AGROVOC (FAO)	-DC -FOAF -VoID -LIME	-the Skosmos Search & browse interface		-meetings -editorial community -mailing lists	- Collaborative work using VocBench - User term suggestions via email - Monthly updates

Funded by the European Union