OntoPortal Workshop 2023 Report

Clement Jonquet, Syphax Bouazzouni, John Graybeal, Jennifer L. Vendetti

OntoPortal Workshop 2023 Report

Hosted by Lifewatch,
Lecce, Italy, September 26-29th, 2023

Clement Jonquet, Syphax Bouazzouni, John Graybeal, Jennifer Vendetti and the members of the OntoPortal Alliance

Participants: Guillaume Alviset, Clement Jonquet, Ilaria Rosati, John Graybeal, Naouel Karam, Mark Musen, Martina Pulieri, Giuseppe Turrisi, Enrica Nestola, Marko Harasic, Nicola Fiore, Arkopaul Sarkar, Michael Dorf, Alex Skrenchuk, Syphax Bouazzouni, Sirine Mhedhbi, Nasreddine Bouchemel
Online participants: Christelle Pierkot, Jennifer Vendetti, Xiaolin Yang, Zhe Wang, Hedi Karray
Other guests: Cosimo Vallo, Baptiste Cecconi, Alexandra Kokkinaki, Philip Strömert, Roman Baum, Robert Rovetto

Summary of the event

The OntoPortal Alliance is a consortium of several research and infrastructure teams and a company dedicated to promoting the development of ontology repositories —in science and other disciplines— based on the open, collaboratively developed OntoPortal open source software. Teams in the Alliance develop and maintain several openly accessible semantic resource repositories in multiple disciplines. The list of repositories is available on the Alliance web page and the teams are listed on GitHub.

The OntoPortal Alliance members and other partners gathered from Sept. 26 to 29th, 2023 in Lecce for the 2nd OntoPortal Workshop. The meeting gathered around 30 persons including management, research and technical profiles (25 on site and 5 remotely). The 2023 workshop's main goals were to consolidate the OntoPortal Alliance organization and shared agenda, and to inform and engage the larger scientific community interested in the OntoPortal technology.

The program included several sessions (technical, content, management) and a set of public sessions including a general presentation of OntoPortal, a user panel and Q&A session with representatives of different communities and a tutorial. The event was also associated with the FAIR-IMPACT Semantic Artefact governance workshop. Topics selected at the agenda included: multilingualism, federated services and metadata, ontology versioning, documentation and training, mappings, code management, and general governance of the Alliance.

In preparation of the event, a survey gathered the partners' perceptions with respect to each topic selected for the agenda of the meeting. In addition, each team presented short updates of their project (available here (private folder)).

The event demonstrated the clear motivation and commitment of the Alliance members to collaborate on the OntoPortal technology and the implementation of ontology-based services in general. This document is a summary of the discussions and the decisions taken.

OntoPortal philosophy and statement

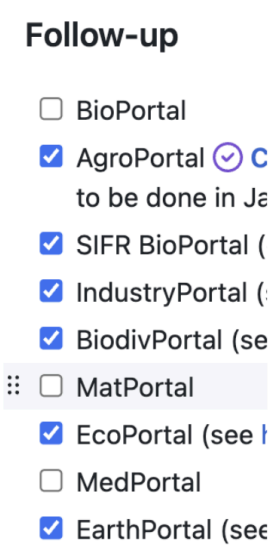
- OntoPortal's motivation is to **ease the use of ontologies,** especially for users not familiar with semantic technologies. OntoPortal becomes useful in the life cycle of semantic resources at the point when releasing a specific version to archive or share become necessary.
- OntoPortal installations **do not own ontologies**, they just serve them. Most portals have adopted an open approach with a flexible editorial policy, where users can upload resources themselves, and the evaluation is left to the community. The OntoPortal project teams do not decide if an ontology is good or not.

- It is accepted that semantic resources **can be hosted in several OntoPortal installations**, but better coordination of the content is needed. An approach to federating our content and services is necessary typically involving a canonical portal for ontologies.
- The OntoPortal Alliance does not yet have a **legal entity but will engage in the process of creating one** to ease the governance and decision making as well.
- OntoPortal default technology is as modular as we can make it and contains the basic components that everyone wants. The default OntoPortal code is mainly contributed and consolidated by Stanford with contributions (as pull requests) from other partners.
- We adopt an **open source philosophy** in which we believe that by investing efforts in developing code for the shared project, we can progress more and more efficiently for our own project. From 2022 and now on we must rely on a **centralized repository on GitHub** https://github.com/ontoportal/.
- Any new functional feature developed by OntoPortal developers should be d**esigned as a new module that can be parameterized** in the config files and activated (or not).
- We rely on **standardized semantic web formats** and technologies to develop generic, domain agnostic, and compliant software that can be applied to a wide range of uses that may not be anticipated.

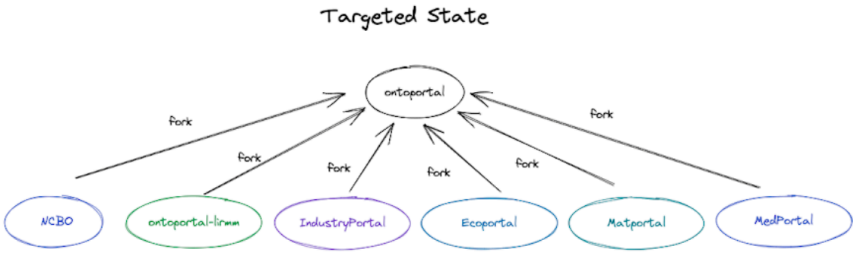
Presentations

OntoPortal year in review

Throughout this past year, significant progress has been made across various fronts within OntoPortal. A notable achievement was the successful collaboration on GitHub, a platform pivotal to our advancements with the new “ontoportal” repository adopted by the Alliance. It stands as an open invitation for contributions and a pillar to our open source project model. The centralized GitHub repository is a valuable resource, particularly in maintaining a robust connection with the master code branch, albeit uncertainties exist regarding MatPortal and MedPortal. BioPortal's unique situation as the older repository and distributor of virtual appliances requires specific attention, but work has begun to reverse the code flow. Six repositories (out of the 9 public) are also relying on the ontoportal-lirmm branch which is now very much advanced in terms of functionalities and remodeling.



3. From now on we must work to move to a centralized repository on GitHub https://github.com/ontoportal/ and work as an “open source” project on this repo.
   a. Beside the code repository and readme files, the issues, teams, discussions, and actions mechanisms of GitHub will be used.
   b. Alliance partners shall “fork” the code repositories to enable traceability and collaborative contributions via pull requests.
   c. BioPortal will have to make a list of the tasks required to make BioPortal a fork of the OntoPortal repo, rather than a master.
   d. A system must be developed by which pull requests to the OntoPortal repo can be efficiently evaluated and accepted (or not) by the OntoPortal community.



Connected 6 out of 9 OntoPortals to master code branch
https://github.com/ontoportal/ontoportal-project/issues/6

The introduction of the new OntoPortal website was also a big achievement. While the developer and admin guides have been established online, the absence of a user guide remains a challenge, owing to the complexity of encompassing all OntoPortal versions. Initiatives in this direction commenced previously but faced interruptions. A DemoPortal is also now operational on the default master branch. Significantly, strides have been made in Docker packaging, contributing to substantial progress. Decisions established previously concerning feature design as switchable modules may have found some application in AgroPortal, but more work is required in this area.

A resource paper published at ISWC 2023 marked a significant effort to communicate OntoPortal's relevance externally. This paper introduces OntoPortal as a tool for constructing ontology repositories and semantic artefact catalogues. We detailed OntoPortal's features and showcased existing and upcoming public repositories developed using this technology, all overseen by the Alliance. We have 53 unique appliance IDs that had more than 2 check-ins in 2023. This shows the technology is used beyond our 9 public repositories.

Over the last year, while various meetings have occurred, and work in sub-groups (Bio/Agro or Eco/Agro or Earth/Agro) has demonstrated efficiency; the efficacy of topic orientation for sessions and leadership remains in question, suggesting the need to reassess this approach. Collaborative exchanges have shown promise, with fruitful GitHub decision-making and multiple merged pull requests, yet the ideal collaboration/commitment model is still work in progress. The pursuit of joint funding to sustain initiatives has been highlighted as pivotal. The annual workshop, however, stands as a testament to our collective efforts and commitment to advancing OntoPortal.

Output from the team presentations and demonstrations

BioPortal

BioPortal Ruby on Rails: we developed new functionality to support ontology change requests. This feature seeks to simplify the contribution process by enabling easy submission of change requests, automatically converting them into GitHub issues, and utilizing KGCL for descriptions while simultaneously generating pull requests. This streamlined functionality has already been activated for prominent ontologies such as MONDO, GO, UBERON, and ENVO. We made several infrastructure upgrades including advancement to Ruby 3.0.6, Ruby on Rails 6.1.7.3, and migrated from Universal Analytics to Google Analytics 4. We modified our fork structure to transition our codebase from an upstream to a downstream repository in the ontportal GitHub organization, and began submitting pull requests of our continued work to ontoportal_web_ui.

BioPortal REST API: we improved our compatibility with AllegroGraph versions 7.3 and 7.4, including resolution of a serious issue that prevented paginated SPARQL queries from completing properly. This work is in preparation for our transition from 4store to Allegrograph, scheduled for November. We improved the goo library via code refactoring and making the unit test suite configurable. Progress was also made in reviewing and merging several pull requests from the AgroPortal team. We modified

the Annotator to make dictionary generation configurable as one way to address sluggish ontology processing. Our analytics services were upgraded from Universal Analytics to Google Analytics 4.

BioPortal DevOps: we released a new version of the OntoPortal Appliance (v3.2.0) and began the process of transitioning from the CentOS Linux 7 operating system to Ubuntu 22.04. (The CentOS Linux 7 operating system will reach EOL on June 30, 2024). We published new docker-compose scripts and DIP (Docker Interaction Program) commands that allow developers to easily spin-up the whole REST API or just dependent software in a Docker container. We implemented new GitHub Actions for auto-launch of unit tests when pull requests are submitted, one-click application deployment, and publishing images to the ontoportal Docker Hub community organization.

AgroPortal

A general update about the AgroPortal project was presented and many new features in development or already deployed were demoed. AgroPortal is currently supported by the French ANR D2KAB project (ANR-18-CE23-0017), the OntoPortal Alliance *International Research Team* of the NUMEV Labex (ANR-10-LABX-0020), as well as T4.2 of Horizon Europe FAIR-IMPACT project (No 101057344). By the end of D2KAB (august 2024), AgroPortal shall be adopted and officially transferred to INRAE. FAIR-IMPACT in particular is seen as a chamber in which we can push for OntoPortal at the level of multiple communities and for the European Open Science Cloud (EOSC). Some work done within these projects was mentioned including: (i) Work on the next version of MOD revisited as a DCAT2 extension/profile, aiming for its widespread adoption and implementation across platforms; (ii) Enabling the availability of O'FAIREe in diverse portals such as EcoPortal, EarthPortal, BioDivPortal; (iii) the work on FAIR-by-design ontology methodology.
A general update of the AgroPortal organization and team (3/4 persons now) was presented. The portal now hosts 156 ontologies but an increase in candidates awaiting addition (150), signals substantial growth. The progression through the releases done since the last OntoPortal workshop in 2022 includes: mappings, SKOS support, administration tools, and merging back with EcoPortal. The recent developments (almost released) were presented on the staging server too: (i) Multilingual content support and internationalized user interfaces; (ii) Brand new and modernized user interfaces; (iii) A consolidated metadata model with a revised submission workflow.; and (iv) A new agent model to support person and organizations metadata values.

EcoPortal

New implementations have been presented, with a particular focus on metadata extension to enable the FAIRness assessment with the O'FAIRe tool and to facilitate DataCite compatible exports to enable DOI requests. Discussions have revolved around content negotiation using ShowVoc and shared interests in negotiating URIs, with certain URIs currently resolving to ShowVoc instead of EcoPortal. Resolving URI issues remains a priority, aiming to circumvent the need for creating ShowVoc installations, and acknowledging a general gap in URI resolution within the semantic web community. Discussions further delve into the necessity for content negotiation services within OntoPortals, emphasizing the need for flexible ontology identifiers that can resolve to various repositories, rather than producing different products for the same ontology based on content negotiation directives.
Efforts focused on streamlining user logins across platforms, adopting a single sign-on (SSO) system via KeyCloak to merge accounts from different providers. The Eco/AgroPortal SSO prototype showcases various provider options (login with Google, GitHub, ORCID) plus KeyCloak. The EcoPortal demonstration emphasized monitoring FAIR scores and metadata export for DataCite, while discussions delved into addressing challenges in resolving IRIs and refining editing roles in tools like VocBench.

MatPortal

No presentation at the workshop.

MedPortal

The Data Platform supporting the National Population Health Data Center incorporates the PHDA website, serving as the platform for submitted data. A pivotal focus lies in creating a robust environment for biomedical standards applications, enabling machine-understandable content standards through layered metadata and ontologies. Notably, the portal houses 60 ontologies enriched with Chinese content, following a recent GitHub installation that facilitated improvements. These enhancements include refined bilingual handling, ensuring each ontology contains bilingual metadata—an accomplishment achieved through extensive manual data curation. Two primary ontology types exist: either a Chinese version with a new name or an ontology supporting multiple languages, although the consistent display of preferred names remains a challenge. Two primary use cases—CDE Portal and CDE Tools—demonstrate the ontology's utility in describing data elements, value domains, and CaseReport Forms (CRFs), offering integration capabilities with Excel and CDE lookup tools.

IndustryPortal

Initiated in 2016, the Industrial Ontologies Foundry (IOF) gained momentum in January 2022 with the deployment of IndustryPortal, a significant outcome within H2020 OntoCommons. IndustryPortal hosts 106 ontologies, 51 users, and spanning diverse domains. To enhance FAIRness, they registered in FAIRsharing, crafted a metadata curation guide, and established a dedicated page for hosting. Their Meta-FAIR initiative amalgamates FAIRness assessment tools.
Focusing on mapping enhancements and the utilization of SSSOM as a standard model for metadata, intending to create a separate database for mapping metadata extraction from ontology source files. Discussions also touch upon the integration of Shibboleth email verification for security and the careful evaluation of ontology content and structural quality metrics, signaling a comprehensive and forward-thinking approach toward OntoPortal's evolution.

EarthPortal

The current earth sciences landscape revealed a bunch of separate SKOSMOS systems (e.g., in France, the solid earth, continental surfaces and atmospheric vocabularies), and other systems based on VocPrez (e.g. the NVS for the marine domain) or UKGovLD (e.g., the terravocabulary linked data registry), highlighting the necessity for harmonization in the earth sciences. Not only there are lots of different systems, but there is a strong need for centralizing semantic artifacts as they rarely are easily findable but also to harmonize practices around semantics. There is a need to establish upload policies and conduct communication and training sessions, understanding user needs, and listen to their feature requests.
Future work focuses on:

- Integrating other semantic artefacts that are not yet present in repositories (e.g., GCMD, Climate and Forecast Standard names, …)
- Engaging other communities of the Earth System, such as the climate community or the community working on water.
- Writing documentation and training for users
- Collaborating on semantic analyzers and FAIRness assessments with EU project FAIR-EASE
- Connecting the EarthPortal to a data repository (the national data repository EaSy Data) and demonstrate the added value to have these two applications connected.
- Exploring vocabulary/ontology editing through a potential VocBench pipeline
- Add new features according to the needs defined by the communities:
    - A vocabulary edition pipeline (VocBench)

- Native SSSOM mappings support
- A better integration of SKOS vocabularies

BiodivPortal

Historically rooted in the GFBio Terminology Service (GFBio TS), which encompassed 22 in-house terminologies and 7 external ones (3.4 million classes and 13,000 individuals). While gaining numerous new capabilities, the shift to BiodivPortal resulted in the loss of Linked Data Deployment, multilingualism, and access to individuals. Version management emerges as a vital consideration. For diff calculation between ontology versions, efforts are underway to upgrade the Bubastis tool and extend OAK's capabilities to align with KGCL standards. Redesigning widgets previously connected to the GFBio TS for connection to BiodivPortal is in progress. Additionally, plans include adopting I-ADOPT and exploring the use of the Annotator to structure observation variables using a subset of ontologies for each I-ADOPT component. A new project, TS4NFDI, aims to harmonize multiple Terminology Services in the NFDI based on OntoPortal, OLS, and SKOSMOS, with a central access point and repository managed sustainably by NFDI. A central SSSOM based mapping repository will serve as a point of truth and bridge between ontologies from the NFDI network.

Newcomers and interested parties

Insights into astronomical initiatives were offered by Baptiste Cecconi and Robert Rovetto, outlining their efforts within IVOA, IPDA, and IHDEA, spanning distinct services. Rovetto aims to materialize 'AstroPortal,' extending its scope to encompass diverse space activities, while Cecconi's work involves an OntoPortal prototype housing various ontologies, still in the exploration phase, emphasizing cross-community interoperability. Alexandra Kokkinaki presented insights into UKRI's Environmental Data Service (EDS), contemplating the creation of a comprehensive portal consolidating vocabulary services and environmental resources in the United Kingdom. The proposed EDSPortal entails a wishlist encompassing Dockerizing, SSSOM, Sparql, triple management, and O'FAIRe. Discussions delved into potential convergence with EarthPortal.

Summary of the public sessions

Public session on ontology development lifecycle

The workshop included a virtual public session on September 27th, 16h-18h CEST.

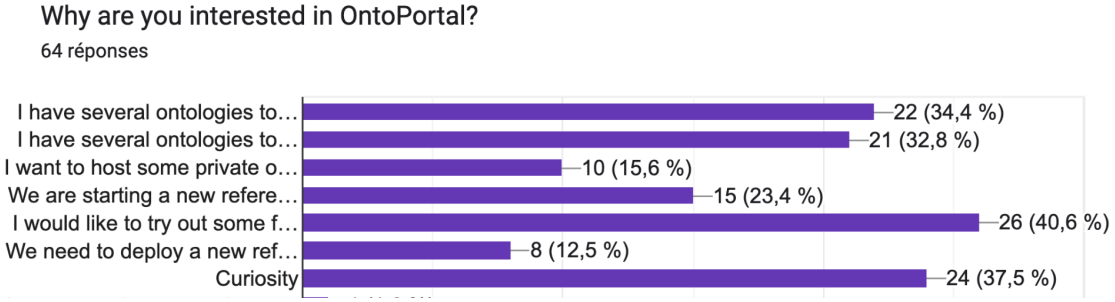
- **16h-16h20:** General introduction to OntoPortal for Ontology Development Lifecycle (C. Jonquet)

- **16h20-17h20:** User panel discussion with «representative user from portals »
  - Anthony Juehne (BioPortal)
  - Sophie Aubin (AgroPortal)
  - Martina Pulieri (EcoPortal)
  - Yan Zhu (MedPortal)
  - Thierry Carval (EarthPortal)
  - Felicitas Löffler (BiodivPortal)
  - Patricia Casla (IndustryPortal)

- **17h20-18h:**Tutorial on “Fast Vocabulary Management with OntoPortal and SKOSPlay” (J. Graybeal)
- Illustration with OntoPortal and VocBench (N. Fiore)

The session was pre-announced with a short registration with 64 respondents (mostly ontology developers, ontology-based application developers, data managers, and semantic web experts). The form included a question about the interest in OntoPortal with the following possible non exclusive responses:

- I have several ontologies to deal with and I like to facilitate this.
- I have several ontologies to manage for a project or group
- I want to host some private ontologies and/or process private data
- We are starting a new reference community ontology repository
- I would like to try out some features included in OntoPortal on some ontologies (Search, Annotator, Mappings, etc.)
- We need to deploy a new reference repository in another language than English
- Curiosity



After the general introduction to OntoPortal, Clement Jonquet led the user representative panel with some questions and responses with the panelists summarized here:

Can you share your best (B) and worst (W) experiences working with your favorite OntoPortal?

- B: Guiding users to leverage ontologies where possible.
- W: Hard to get new communities used to ontologies.
- W: Filling up the metadata is sometimes too difficult for producers.
- W: Uploading a modular ontology had some problems, and had to be merged in a single file.
- B: Looking for classes and ontologies.
- B: DOI to link resources to papers.
- B: O'FAIRe tool integration.
- B: Management of SKOS resources is a real plus.

The OntoPortal Alliance is discussing the "federation" of all/some portals (metadata, search, content, …) Do you have any suggestions on this?

- Useful to have a federated system among the portals to deal with similar content.
- The concept and class search would be nice to have with federated portals.
- Search, referencing other resources, and dealing with duplicates/overlaps.
- Keep domain-specific ontology services, so users can decide whether to stay in a specific domain or generalize.
- For specific repos domain-specific services make sense, so only think of this as an add-on.
- Must find the right ontologies, which means the ability to find the right portals.
- Really in favor of federation because duplicate ontologies are a nightmare.
- Be careful with keeping each portal's identity.
- Each portal has its own editorial process and constraints.
- Metadata authoring facilitated by the federation.

A recurrent topic is how to connect OntoPortals to related databases and possibly enable semantic search, and semantic indexing of their content. Do you have any suggestions on this?

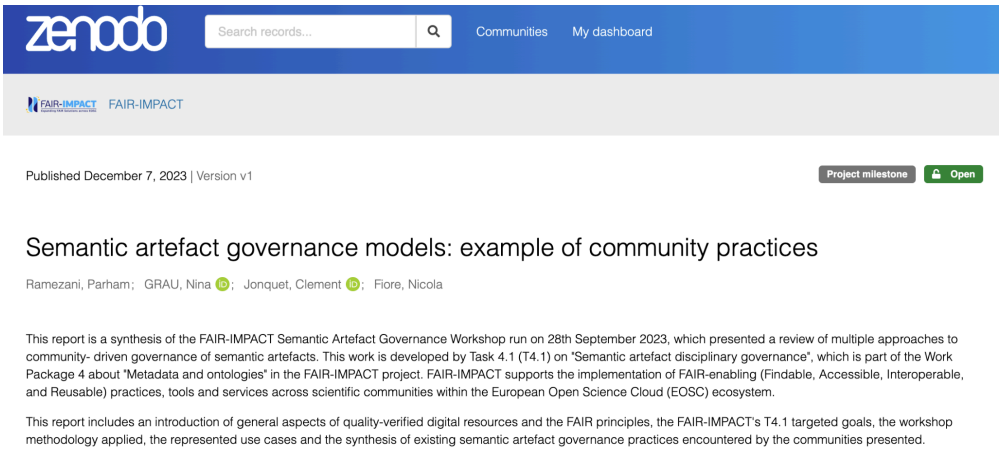
- Need ability to select ontologies for autocompletion (exists in the widget but needs revamping).
- Why not come to an ontology repository to explore the resources in a database.
- OntoPortals provide ways to get to data itself, that's desirable.
- We convince communities to create ontologies, but then we don't show them the added value of using them for datasets or databases.
- Understanding why the related topics are brought up, and why they are related, is very important.
- Getting a list of databases or datasets that use terms hosted in the portal is interesting to see what an ontology is used for. Diving into the datasets from the ontology terms maybe a different tool than the ontology portal.
- Getting annotated text back is also important: don't drop the Annotator.

What would be the ideal feature OntoPortal should develop to complete/enhance its portfolio?

- Management of multilingualism.
- Text annotation feature.
- SPARQL endpoint
- Federated semantic queries on different systems
- Editing would be great. Being able to work with versions is key.
- Display ontologies with properties, classes, and instances linked together to give a broad view of ontology itself.

Semantic artefact governance workshop

Report of the workshop has been produced by FAIR-IMPACT in MS4.1: 10.5281/zenodo.10287010



Summary of the topic sessions

Multilingual OntoPortal

The primary goals are centered around three key objectives: displaying ontologies content in multiple languages, enabling cross-language search functionality, and internationalizing the user interface (UI). In the new proposition presented by AgroPortal, the presentation language for ontology labels is determined by the default language of the portal. Users have the option to switch between a

predefined list of languages, set by the ontology admin through a metadata property (omv:hasOntologyLanguage). Additionally, users can choose to view all values in all available languages.
Previous limitations in indexing all values, particularly preferred labels, have had an impact on search results. Efforts were made to address this issue by default fetching and indexing values in all languages defined in the original parsed resource.
Translation initiatives for the UI are in progress, with a focus on soliciting contributions from non-English speakers, to internationalize the UI in the following languages: French, Italian, German, Spanish, and Chinese.
Its important to note the multilingual implementation done in AgroPortal is very much incorporated in the new user interface and rely on some metadata properties not available in the OntoPortal master branch.

Federated services (search and more) and metadata, governance of the content

In summary:

- The federation does not necessarily involve all the portals. Technically speaking it's possible to federate with all, but it can be customized portal per portal.
- We shall adopt a canonical portal for a "resource" approach
- A resource is here understood as any object an OntoPortal instance is dealing with (ontologies, notes, projects, users, agents…) with a primary focus on ontologies.
- The federation of metadata will only be possible with a portal following the same metadata model. Currently we have two such models.
- Federation is not a response to handle multiple versioning. A canonical portal for a given ontology shall be in charge of hosting and serving the latest version of the ontology. If a portal (like what BioPortal did after some user request) hosts multiple versions of an ontology (as primary object i.e., with different acronym IDs), those shall be seen as "separated" ontologies in the federation.
- The canonical model also brings up the question of harmonizing IDs so duplicates are detected. It means we shall start to work on a file to identify the duplicates.
- Federated queries can create technical challenges (delays, connections, etc.) that could produce a bad user experience.
- MegaPortal (a portal unique to rule everything all) has come up a number of times in the federation discussion but this is not the choice the group is taking.

Elements on this topic moved to GitHub: https://github.com/ontoportal/ontoportal-project/issues/38

Ontology versioning, diffs, and historical views

The group discussed the need to support differences in ontologies, emphasizing a preference for a visual representation. Changes should also encompass ontology metadata. Standardized descriptions through KGCL[1] were deemed essential, with a consensus on the necessity for changes, though opinions varied on the desired complexity. A debate arose on prioritizing ontology-level versus term-level history views, with 75% favoring the former and 45% the latter. Addressing the broader topic of versioning processes throughout an ontology's life cycle, there was agreement on the need for improvement. An agreement to follow the recommendation published in https://hal.science/hal-04094847 was strong. Discussions covered the importance of version information in annotations, and advocating for semantic versioning to distinguish major, minor, and trivial changes. BiodivPortal exemplified diverse versioning scenarios, though challenges in computability were acknowledged. The plan involved extending KGCL to cover additional changes, in collaboration with Chris Mungall's group. The choice between Bubastis and KGCL formats favored the

[1] https://incatools.github.io/kgcl/

latter. BiodivPortal developed some ideas with respect to versioning and diff and will make propositions.
On the side, the terminology shift from “release” dates to “creation” or “modification” dates was proposed, aligning with the previous paper approach. CJ noted the variation in tools' understanding of CreationDate versus ModificationDate and suggested accommodating ontologies not using dc: properties.

Documentation and training

Identified issues include partially outdated documentation due to the new UI and the absence of user documentation, with plans to address these gaps through contributions.
During the workshop Cosimo organized the beginning of coordinated training and documentation development across OntoPortal. The goal is not to do all our documentation, but coordinate the documentation each team develops.
It was discussed to organize a small working group on these topics, not necessarily engaging all the portals. It was also suggested that someone (one team) “jump in the pool and do it” and work for 3 months on all the historical content we all have and will produce the documentation for all. Then a small working group can amend that and customize it a bit.
There is a on-going open GitHub discussion on this matter.

Mappings repository and SSSOM compliance

This section was about presenting IndustryPortal results on mappings. No reporting.

OntoPortal code management

The workshop introduced the need for better and more harmonized code management in OntoPortal. After connecting codes to OntoPortal as decided in the 2022 workshop, the focus shall now be on updating and enhancing procedures for code updates and shared contribution. The key aspects include:

- Updating upstream OntoPortal code: the current state involves an outdated BioPortal version (ncbo branch) in the ontoportal upstream code, with the goal for 2024 being synchronization with the latest ontoportal-lirmm branch.
- Notifying and assisting OntoPortal forks: they should be informed of updates through pull requests, enabling them to choose whether to merge or not.
- Updating appliance infrastructure: currently, manual updates are required, but future exploration of Docker adoption aims to simplify updating, adding, or removing infrastructure components.

The action items include synchronizing ontoportal upstream code, assisting forks, establishing a notification mechanism for code releases, and progressing with Dockerization.
Additionally, for BioPortal a parallel deployment is considered to inspect or implement new instrumentation for scalability and debugging, focusing on metrics such as page speeds, API calls, and generated SPARQL queries.
Furthermore, discussions were held on the integration of different editors, such as VocBench, Protege and SousLeSens, into OntoPortal. Considerations were made regarding the establishment of a protocol to facilitate a seamless connection between these editors and the OntoPortal system.

Governance of the Alliance, project proposals and funding

The session revolved around governance, partnerships, and the need for formalizing agreements within the group. Key topics included aligning with industry standards, enhancing collaboration within the group, and discussing the formation of a formal agreement. We agreed on formalizing the partnership through the signing of an agreement or Memorandum of Understanding (MoU) first. We all believe these engagements are necessary for the Alliance to continue its mission. If we assume the

Alliance will grow (because semantic needs are getting much stronger), then we should invest time to be ready for that. The MoU shall be a first step toward the creation of a legal entity (association or non-profit) that will give the Alliance more stability and offer an official entity that can be contacted.
Concerns were raised regarding MatPortal's engagement status and LifeWatch's investment, emphasizing the need for reliability.
Discussions centered on the group's growth assumptions, responsibilities of members, and the balance between technological focus and skill development (maybe we could have the Alliance focus "not only" on the technology but also on developing skills around semantics).
Everyone agreed that it's important to lay out what you get by becoming a member, and what your responsibilities and contributions are too. Contribution to the group seems essential. Maybe it could be in-kind, or even financial. Perhaps efforts just need to be better directed.

The conversation explored options like a 501(c)3, AISBL, or a commercial entity. John originally wrote a draft of status/document for a 501(c)3 but we never looked at it as a group. Different perspectives emerged on creating a commercial company and the balance between individual commitments and organizational strategies. Who will put the time in to make it happen? Others believed our energy should still focus on getting grants (possibly big ones) together, for instance making OntoPortal a real EOSC component.
Creating a commercial entity to take advantage of tenders was a little more concerning for some, but not out of the question. Several felt the attractiveness of the group is around the commitments of its individuals toward the common good, not any organizational commitment or corporate strategy.
We need clear Terms of Reference about how people are expected to use/re-use the fruits of our labors.

Feature requests identified by users and partners

Legend

- Reported in 2022 and achieved in 2023 in either one specific portal or at the OntoPortal level.
- Reported in 2022 not yet achieved but moved to GitHub for further discussion.
- Reported in 2022 and not yet addressed or not yet ready to be moved to GitHub for discussion and realization.
- New, reported in 2023.
- Added the following tags in blue:
    - **(done {in PortalNames})**
    - **(ongoing)**
    - **(planned)**

General

- Refactor the ontology submission process so that a user is invited to come back after parsing to validate/curate the outputs of the process and eventually fix a few issues and edit metadata BEFORE full availability in the OntoPortal. **(ongoing)**
- Refactor the ontoportal.org website (https://ontoportal.org) and use it as a landing page for the technology, the alliance, and the documentation. **(done)**
- Multilingual support for ontology content **(done in AgroPortal)**
- Multilingual support for the UI **(ongoing)**
- Better support of SKOS (something more aligned to what SKOSMOS does) **(ongoing)**
- Make the UI easily and more customizable **(ongoing)**
- Creation of different user roles (curators, admin of a group of ontologies).
- SHACL shapes-based validation of semantic resources (what we need in OntoPortal, no lessons on how to develop an ontology). Semantic validation functionality that will

automatically check a semantic resource with respect to identified guidelines (plug-in architecture).
- Refactor the notes to be more modern and connected to the semantic resource developer practices. Support a unique, solid, and relevant term proposal process (currently OntoPortal has three: “Add proposal”, “Provisional classes” and OntoloBridge) **(planned)**
- Consolidate all the documentation available (deployer guide, developer guide, user guide) and set up a system that allows factorizing documentation between OntoPortal installations. Better documentation is needed (at a very level) maybe a series of webinars or training materials. **(planned)**
- Develop a link to an ontology editor specific to a semantic resource similar to what EcoPortal has done for VocBench (can be based on metadata attribute) **(planned)**
- Better support ontology diffs and improve the perception of what items have changed between versions and overall ontology evolution. **(planned)**
- Select a shared classification to define and harmonize our categories across portals **(planned)**
- Link OntoPortal semantic resources to external datasets using or annotated by the semantic resources **(planned)**
- Implement a read-only version of OntoPortal i.e., an installation with no user contributions (ontologies, projects, mappings, notes, login).
- Video tutorial for the installation of OntoPortal.
- Tutorial on how to setup a development environment for a new OntoPortal developer
- Re-develop an OntoPortal plug-in for Protégé.
- Better support of ontology imports and re-uses of objects (classes, properties, instances, concepts) from other semantic resources.
- Better handling of ontologies that are knowledge bases (i.e., OWL model with many instance data)
- Better support of the Annotator as many many data are still in text form.
- The portal needs to facilitate ontology competency questions e.g., by showing the type of questions/queries an ontology allows to answer.

Metadata

- Re-align EcoPortal and AgroPortal metadata model **(done in AgroPortal and EcoPortal)**
- Merge the DOI feature of EcoPortal for other portals with a rich metadata model. **(done in EcoPortal)**
- Refactor the Summary page for portals with a rich metadata model. **(done in AgroPortal)**
- Whatever the metadata model, consolidate the automatic metadata extraction and generation to ease the management/curation of metadata. **(ongoing)**
- Better capture semantic resource status (manage obsolete resources) **(done in AgroPortal)**
- Implement a federated metadata search service in each portal and cross portal. **(planned)**
- Refresh cache after ontology submission for every other user to see a new submission **(planned)**
- Enable a parameterizable metadata model. **(planned)**
- Implement an incrementally augmentable minimum metadata model allowing any portal to incorporate new metadata properties ‘one-by-one’.
- Whatever the metadata model, find a way to import metadata from external sources e.g., a VOAF file or OBO Foundry YAML description.
- Better description of the OntoPortal pages (not only Summary) with Schema.org for better indexing by Google.
- Evaluate the best way to adopt a more advanced metadata model in BioPortal (ideally compatible with the rich model of AgroPortal, and with requirements for FAIR ontologies).
- History view for the evolution of a semantic resource (historical influences between ontologies view as a graph)

- Connect to tools to Vidoco to produce HTML doc for (small) semantic resources

Mappings

- Full support of SSSOM for the mapping repository. **(ongoing)**
- Implement better views for mappings including an evaluation view and an overview of mappings between any ontologies.
- Support bulk upload and download of mappings **(ongoing)**
- Connect to third-party tools for ontology alignment. The appropriate way needs to be clarified as many tools are not generic or pluggable. Not necessarily process the cartesian product of mappings all the ontologies with all the other ones.
- Enhance the use of mappings inside the portal itself.
- Support evolution of mappings in the repository when mapped semantic resources evolve
- Make existing mappings selectable/viewable by their origin (e.g., show only API mappings, or only automated mappings).

Federation

- Implement a view to list (with linked logos) the public OntoPortal installation in the OntoPortal Alliance. A commitment to the Alliance means a presence on the page on all the portals. **(planned)**
- Better coordination when a semantic resource is hosted in multiple repositories in the alliance. **(planned)**
    - Avoid repeating the authoring of metadata from one portal to another.
    - Should be easy for final users to find content on different OntoPortal installations
- Implement a federated search of content. Grab either statistics (number of hits) or even results (to produce snippets) in the querying interface. **(planned)**
- Implement a cross-portal metadata search service that will facilitate the identification of the semantic resources. **(planned)**
- Federated Recommender. When a query is passed to the recommender, select the option to query also the other Recommenders.

DevOps and environment

- Docker container-based setup/installation of OntoPortal **(ongoing in BioPortal and AgroPortal)**
- Implement and describe an easily deployable environment test **(ongoing)**
- Regression testing automatically on GitHub code updates **(ongoing)**
- A configurable OntoPortal appliance
- Implement data (ontology) migration scripts
- Better support versioning of code and identification for which code is currently running on which appliance
- Move to a more generic (and GDPR compliant) tool for analytics beyond Google Analytics.
- Support multiple open-source triplestore backends. **(ongoing)**
- Finalize support and transition to AllegroGraph **(done in BioPortal)**
- Clean up a lot of code bits that are workarounds for 4store performance issues **(ongoing)**
- Implement full SPARQL 1.1 compliance to support any triplestore (if not complete already) **(ongoing)**
- Implement a triple data migration script to move data from a 4store to another triplestore **(ongoing)**
- Add monitoring tools to OntoPortals **(ongoing)**
- Create a SPARQL query editor/viewer. Minimally plug in the viewer of the underlying used triple store (4store and AllegroGraph have one). **(ongoing)**

- Address the end of support of CentOS **(ongoing)**
- Create an option to enable the use of a separate public SPARQL triplestore. **(ongoing)**
- Design other views to interact with the SPARQL triplestore e.g. RelFinder, AdvancedSearch, etc.
- Implement an entry point to a unique knowledge graph with the content (not duplicated) of all OntoPortal installations.
- Ultimately OntoPortal could be installed by "pushing one button" within the appropriate research infrastructure

Other technical aspects

- Update Bootstrap version **(done in BioPortal)**
- Upgrading to Rails 7 **(done in AgroPortal)**
- Migrate and define all our reusable UI components as view components and use a lookbook framework for previewing/testing the UI components **(done in AgroPortal)**
- Optimizing process for large ontologies **(done in AgroPortal)**
- Eliminate the old JS libraries.
    - Adopt Stimulus and Hotwire to avoid JavaScript **(done in AgroPortal)**
    - Adopt ViewComponents as reusable view components **(done in AgroPortal)**
    - Update our JS so we aren't forced to use jquery **(ongoing)**
- Refactor notifications (changes, notes, processing status, errors, new version, etc.). e.g. so ontologies/users will know when an ontology changes **(planned)**
- Low-level caching in Rails may be the solution for a better caching in the UI (standard Rails cache will not work for our case) **(planned)**
- Refresh the widgets **(planned)**
- Refactor multiple AJAX calls **(planned)**
- Break up the ontology parsing process into smaller steps, more modular and traceable and executable in parallel **(planned)**
- Use a better queue management tool like Sidekiq **(planned)**
- Avoid looping over classes multiple times **(planned)**
- Revise the mechanisms to deal with obsolete classes (branch and custom property) and see what's going on with the obsolete classes graphs
- Move the generation of mgrep dictionary process to once per day
- Audit on graphs cleaning in the triplestore when a submission is deleted

Future meetings

1. From now on the Alliance meetings will follow this structure:
    a. OntoPortal governance monthly meeting (2nd Thursday of each month at 1500 GMT). Organized by Jenifer Vendetti (deputy Clement Jonquet). Invites sent out to the OntoPortal Alliance management mailing list ontoportal-alliance@lists.stanford.edu
    b. OntoPortal technical meetings (every month, in alternance of 2 weeks with the manager meeting). Organized by Syphax Bouazzouni. Invites sent out to developer mailing list ontoportal-dev@lists.stanford.edu
2. Topic meetings are not the selected approach anymore (as it was not successful in 2022-2023). Still, they are possible in complement of regular meetings. No need to have everyone in the Alliance represented to move these topics forward.
3. We shall reproduce this annual meeting in 2024.

Other observations

- We covered as many topics as possible. List of uncovered items for future meetings. See also the topic survey done in preparation for the 2023 workshop.
- We did not reach out to a specific category of stakeholder: the persons who have installed the OntoPortal Appliance but have not used it to make a public ontology repository. Stanford has a list (pre-license time and post-license time).