OLS4: a new Ontology Lookup Service for a growing interdisciplinary knowledge ecosystem

James McLaughlin¹⁰, Josh Lagrimas¹⁰, Haider Iqbal¹, Helen Parkinson¹, Henriette Harmse^{*,10}

¹Samples, Phenotypes and Ontologies Team (SPOT), EMBL-EBI, Wellcome Genome Campus, Hinxton CB10 1SD, United Kingdom *Corresponding author. Samples, Phenotypes and Ontologies Team (SPOT), EMBL-EBI, Wellcome Genome Campus, Hinxton CB10 1SD, United Kingdom. E-mail: henriette007@ebi.ac.uk.

Associate Editor: Peter Robinson

Abstract

Summary: The Ontology Lookup Service (OLS) is an open source search engine for ontologies which is used extensively in the bioinformatics and chemistry communities to annotate biological and biomedical data with ontology terms. Recently, there has been a significant increase in the size and complexity of ontologies due to new scales of biological knowledge, such as spatial transcriptomics, new ontology development methodologies, and curation on an increased scale. Existing Web-based tools for ontology browsing such as BioPortal and OntoBee do not support the full range of definitions used by today's ontologies. In order to support the community going forward, we have developed OLS4, implementing the complete OWL2 specification, internationalization support for multiple languages, and a new user interface with UX enhancements such as links out to external databases. OLS4 has replaced OLS3 to transition.

Availability and implementation: The source code of OLS is available at https://github.com/EBISPOT/ols4 and DOI 10.5281/zenodo.14960290 with Apache 2.0 License. A freely available implementation is accessible at https://www.ebi.ac.uk/ols4.

1 Introduction

The Ontology Lookup Service (OLS) is a search engine for ontologies, first released in 2006 (Côté et al. 2006). It supports users to search for ontology terms during knowledge curation required by the Findable, Accessible, Interoperable, and Reusable principles. Users of OLS include high-throughput phenotyping centers producing and exporting their data, such as members of the International Mouse Phenotyping Consortium (IMPC) (Groza et al. 2023); data integration initiatives such as the OpenTargets Platform for drug target identification and prioritization (Ochoa et al. 2023); and curators of databases including the Genome-Wide Association Study (GWAS) Catalog (Cerezo et al. 2025), Expression Atlas (Moreno et al. 2022), European Genome-Phenome Archive (Freeberg et al. 2022), Polygenic Score Catalog (Lambert et al. 2021), ChEMBL (Gaulton et al. 2012), WormBase (Harris et al. 2010), EuropePMC (Gou et al. 2014), PRIDE (Perez-Riverol et al. 2025), Ensembl (Dyer et al. 2025), IntAct (Hermjakob et al. 2004), CancerModels.org (Perova et al. 2024), BioStudies (Sarkans et al. 2018), and the BioImage Archive (Hartley et al. 2022). Recent applications of OLS include harmonization and standardization of data, for example protocols across phenotyping centers; quality control checks on raw data, e.g. correction of data submission errors and detection of baseline drift due to instrumentation; and tracking the progress of phenotyping efforts by funding bodies.

Successive iterations of OLS have evolved in response to user needs, e.g. as standards for Application Programming Interfaces (APIs) changed from SOAP to REST (Côté *et al.* 2010) and

when new OBO and Web Ontology Language (OWL) version 2 (OWL2) (https://www.w3.org/2007/OWL/draft/ED-owl2-newfeatures-20081202/all.pdf) standards for ontologies were introduced (Jupp et al. 2015). Recently, the scale and complexity of biological and chemical knowledge have increased dramatically. New methodologies such as spatial transcriptomics have changed the resolution of data to single cell; e.g. OLS is used in the Human BioMolecular Atlas Program (Börner et al. 2025); and high performance computing has become more abundant. In turn, ontologies have grown significantly in scale: in December 2016, OLS indexed 158 ontologies with 4 862 923 classes. In December 2024, OLS indexed 266 ontologies with 8 682 322 classes. New authoring tools such as ROBOT templates (Jackson et al. 2019) and Dead Simple OWL Design Patterns (DOSDP) (Osumi-Sutherland et al. 2017) have expedited this process by making the development of ontologies more automated enabling new terms to be added in large quantities. The complexity of ontologies has also increased; e.g. internationalization to support the translation of ontologies into different languages to support a diverse and international user base (Gargano et al. 2024), and features from the OWL2 specification such as disjointness statements and property chains. So far none of the existing opensource solutions [OLS3, BioPortal (Noy et al. 2009), OntoBee (Ong et al. 2017), and AgroPortal (Jonquet et al. 2018)] are able to comprehensively support these use cases.

2 Materials and methods

OLS4 is the new version of the Ontology Lookup Service. The OLS4 data-load and backend are implemented in Java

Received: 23 January 2025; Revised: 24 March 2025; Editorial Decision: 23 April 2025; Accepted: 1 May 2025

© The Author(s) 2025. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

11 and Spring Boot. OLS4 uses Neo4j as a graph database and Solr for full text search. We use Neo4j rather than an RDF triplestore as Neo4j has strong support for recursive queries, used in the OLS tree view and API to retrieve all ancestors and descendants of an ontology node. The labeled property graph (LPG) structure used by Neo4j also enables provenance and reference information attached to OWL axioms to be represented as properties of graph edges. In future, this LPG representation could enable OLS data to be represented and queried using the emerging KGX standard for knowledge graph interoperability (Caufield *et al.* 2023) and incorporated into wider biomedical knowledge graphs such as BioCypher (Lobentanzer *et al.* 2023).

The architecture has been simplified from OLS3; Neo4j is now used as a standalone server rather than an embedded database, and ontology metadata is also stored in Solr removing the need for a MongoDB instance. The motivation for this simplification was to reduce the complexity of deploying OLS outside of its primary instance at EMBL-EBI, for use cases such as the MONARCH Initiative OLS and the NFDI4Chem Terminology Service (Steinbeck et al. 2023). The ETL pipeline has also been simplified by eliminating redundant processing to improve scaling of the dataload. OLS4 dataloads for the complete set of ontologies take on average 6% of the time used by OLS3. These faster dataloads allow users to see updates to ontologies more quickly and reflect changes in knowledge, particularly important when biological knowledge develops rapidly which is an important issue for pandemic preparedness; OLS is currently being used as part of the European Viral Outbreak Response Alliance project.

In OLS4, the Neo4j and Solr schemas are dynamic and depend on the annotation properties used in the OWL entities in the source ontologies. OWL entities are translated from RDF to a lossless JSON representation, which is then stored complete and unmodified in both Neo4j and Solr alongside the extracted queryable properties. Queries to Neo4j and Solr include this JSON representation which is used to generate API responses and the frontend pages. In order to maintain backward compatibility, the API is implemented using view classes which match the previous OLS3 data model, but abstract from the underlying OLS4 data model. The use of abstract views enables multiple API versions to be built over the same underlying data model, and changes to the API without reloading data each time to deliver updates delivering new API use cases more quickly to users.

The OLS4 frontend has been rewritten to communicate with the backend exclusively using HTTP APIs rather than by directly accessing the internal data model, offering an option to build new front ends for an OLS4 backend instance, such as third party interfaces tailored to specific ontologies/ use cases. Custom instances of OLS3 such as the NFDI4Chem terminology service, which previously had to run divergent OLS instances with locally modified backend code, will in future be able to use the latest OLS backend code coupled to a customized frontend reducing the overhead of keeping the backend code synchronized.

3 Results

OLS4 has many new features including full implementation of the OWL2 specification; annotations on annotations; internationalization support; cross-references between ontology terms; and BioRegistry (Hoyt *et al.* 2022) integration. The OWL2 specification has been implemented comprehensively and tested using a suite of test cases based on both the OWL2 Primer and example test-cases extracted from biological and biomedical ontologies, including the Experimental Factor Ontology (EFO) (Malone et al. 2009) and the MONDO Disease Ontology (Vasilevsky et al. 2022). OLS4 is therefore able to support ontologies using OWL2 features; e.g. in the MONDO disease ontology where OWL2 disjointness is asserted between extrapulmonary tuberculosis (MONDO:0000368) versus pulmonary tuberculosis (MONDO:0006052); and in the Relation Ontology (RO) where the OWL2 property chain regulates=directly regulates ->directly regulates is defined to describe transitive regulation relations. These OWL2 definitions are now visible in the OLS browser and API. In addition to OWL2 ontologies, OLS4 loads schemas defined using rdfs: Class hierarchies, providing users with a standard API to access both OWL ontologies and commonly used schemas such as Dublin Core (https://www.rfc-editor.org/rfc/rfc2413) and Schema.org (Guha et al. 2016), which are in turn used by OWL2 ontologies.

OLS4 improves annotation support by implementing annotations on annotations (sometimes termed reification), making references and provenance associated with ontology axioms visible in the web interface. For example, UBERON (Mungall et al. 2012) is a widely used multispecies anatomy ontology. The UBERON term for "lung" contains homology notes derived from The evolution of organ systems (Schmidt-Rhaesa 2007), a link which is now visible in the corresponding OLS page for attribution and cross-referencing. Full internationalization of annotations has also been added in OLS4; ontologies are browsable in multiple languages and OLS4 displays a language picker listing all languages present in the ontology. When a language is selected, annotations are displayed in the language selected by the user where possible. This functionality has been demonstrated in the Human Phenotype Ontology (HPO) (Gargano et al. 2024) which is now accessible in multiple languages in the main OLS instance, enabling curators to map to consistent phenotype terms across language barriers.

Another significant development in OLS4 is the handling of cross-references between ontology terms. The majority of the ontologies indexed by OLS do not exist in isolation, but reuse terms from other ontologies. For example, the EFO imports chemical terms from the Chemical Entities of Biological Interest ontology. In OLS4, such imported terms are labeled in the tree with a tag linking to the defining ontology (Fig. 1). This functionality is critical to support the Unified Phenotype Ontology (uPheno) (Matentzoglu et al. 2024), which aggregates terms from multiple phenotype ontologies often with the same name; without the defining ontology tags it would be unclear the difference between, e.g. HPO and MP (Smith et al. 2005) terms, in the tree view. In addition to cross-references between ontology terms, OLS4 also automatically creates external links using the Bioregistry (Hoyt et al. 2022), enabling users to easily navigate from ontologies to external databases, e.g. from a gene to a genome sequence in GenBank (Sayers et al. 2023).

Altogether these features allow OLS4 to deliver a range of new use cases for the ontology community, and to make ontologies more interoperable with other biological and chemical resources. OLS4 is now in production at EMBL-EBI and served approximately 50 million requests from



Figure 1. An ontology term from the Experimental Factor Ontology (EFO) viewed in OLS4. References to terms in different ontologies are tagged with links to the corresponding defining ontologies for easy navigation.

approximately 200 000 unique hosts between November 2023 and February 2024.

4 Discussion

Future work will include support for the Simple Standard for Sharing Ontological Mappings (SSSOM) (Matentzoglu *et al.* 2022). Mappings between ontology terms are used, e.g. to map phenotype terms between human phenotypes and model organisms such as mouse and zebrafish. While some mappings are present in ontologies, often represented as hasDbXref properties, SSSOM allows multiple different mapping sets to be defined with associated mapping metadata, which is important as mappings are often subjective and project-dependent. We plan to add support to load and display alternative sets of mappings depending on user preference, e.g. to allow users to choose between different HPO to MP mappings provided by MGI, IMPC, and Pistoia Alliance.

In future, we also plan to implement more sophisticated search capabilities, such as searching for a specific annotation with a specific value and searching in a specific branch of an ontology. For example, EFO terms used to annotate studies in the GWAS Catalog are annotated with a property gwas trait=true. Searching for terms with this annotation would allow the impact of deprecating or moving a term on annotated datasets to be assessed. Limiting a search to a specific branch of an ontology would allow users interested in, e. g. cardiology to limit their searches to terms underneath "heart disease" in MONDO or "heart" in UBERON.

OLS serves curators, annotators, data resource producers, and ontology developers. It has been designed to meet the needs of these user groups as well as to scale for larger and more complex ontologies. For curators and annotators, OLS4 provides a richer view of terms including complete OWL2 axiomatization to help users select appropriate terms and navigate between terms, a unique feature of OLS4 among open source ontology browsers. OLS4 also adds multiple language support which enables bio-curators to search for terms using labels in their native languages, a feature so far only supported by AgroPortal but useful for human disease communities who also need common names. For data resource producers, the faster dataload will enable resources to load and link to the latest versions of ontology terms in days without waiting several weeks for OLS to update. For ontology developers, the new display features of OLS will enable new use cases to be delivered in ontologies with visibility to users, as has been demonstrated for HPO and MP internationalized editions. OLS4 will also help to prevent the proliferation of terms across multiple ontologies re-defining the same concepts in different contexts, by adding ontology tags to make

the presentation of the relationship between ontologies more prominent and easier to navigate.

Acknowledgments

The authors would like to thank Nicolas Matentzoglu (Semanticly Ltd), David Osumi-Sutherland (Wellcome Trust Sanger Institute), and the OBO Foundry Community for testing and feedback.

Conflict of interest: None declared.

Funding

J.M., J.L., H.I., H.P., and H.H. are supported in part by EMBL-EBI Core Funds. J.M., H.I., H.P., and H.H. are supported by European Union's HORIZON program grant number 101131959. J.M. was supported by the Chan-Zuckerberg Initiative award for the Human Cell Atlas Data Coordination Platform from 2020 to 2023; Office of the Director, National Institutes of Health (R24-OD011883, OT2OD033756); and NIH National Human Genome Research Institute Phenomics First Resource, NIH-NHGRI # 5RM1 HG010860, a Center of Excellence in Genomic Science. H.H. was supported by the European Union's Horizon 2020 research, and innovation program grant numbers 824087 (European Open Science Cloud Life from June 2020 to August 2023) and 825575 (European Joint Programme on Rare Diseases from June 2020 to December 2023). J.M., H.I. and H.H are supported in part by EVORA. The EVORA project has received funding from the European Union's HORIZON programme under grant agreement No 101131959.

References

- Börner K, Blood PD, Silverstein JC et al. Human biomolecular atlas program (hubmap): 3D human reference atlas construction and usage. Nat Methods 2025;22:845–60. https://doi.org/10.1038/s41592-024-02563-5
- Caufield JH, Putman T, Schaper K *et al.* KG-Hub-building and exchanging biological knowledge graphs. *Bioinformatics* 2023;**39**: btad418. https://doi.org/10.1093/bioinformatics/btad418
- Cerezo M, Sollis E, Ji Y et al. The NHGRI-EBI GWAS catalog: standards for reusability, sustainability and diversity. Nucleic Acids Res 2025;53:D998–1005.
- Côté R, Reisinger F, Martens L et al. The ontology lookup service: bigger and better. Nucleic Acids Res, 2010;38:W155–60.
- Côté RG, Jones P, Apweiler R *et al.* The Ontology Lookup Service, a lightweight cross-platform tool for controlled vocabulary queries. *BMC Bioinformatics* 2006;7:W155–60.
- Dyer SC, Austine-Orimoloye O, Azov AG et al. Ensembl 2025. Nucleic Acids Res 2025;53:D948–57. https://doi.org/10.1093/nar/gkae1071
- Freeberg MA, Fromont LA, D'Altri T et al. The European Genome– Phenome Archive in 2021. Nucleic Acids Res 2022;50:D980–7.
- Gargano MA, Matentzoglu N, Coleman B *et al.* The human phenotype ontology in 2024: phenotypes around the world. *Nucleic Acids Res* 2024;**52**:D1333–46.
- Gaulton A, Bellis LJ, Bento AP et al. Chembl: a large-scale bioactivity database for drug discovery. Nucleic Acids Res 2012;40:D1100–7.
- Gou Y, Graff F, Rossiter P, *et al.* Europe PMC: a full-text literature database for the life sciences and platform for innovation. *Nucleic Acids Res* 2014;43:D1042–8.
- Groza T, Gomez FL, Mashhadi HH et al. The international mouse phenotyping consortium: comprehensive knockout phenotyping underpinning the study of human disease. Nucleic Acids Res 2023;51: D1038–45. https://doi.org/10.1093/nar/gkac972

- Guha R, Brickley D, Macbeth S. Schema.org: evolution of structured data on the web. *Commun ACM* 2016;59:44–51.
- Harris TW, Antoshechkin I, Bieri T et al. Wormbase: a comprehensive resource for nematode research. Nucleic Acids Res 2010;38:D463–67.
- Hartley M, Kleywegt G, Patwardhan A et al. The bioimage archive—building a home for life-sciences microscopy data. J Mol Biol 2022;434:167505.
- Hermjakob H, Montecchi-Palazzi L, Lewington C et al. Intact: an open source molecular interaction database. Nucleic Acids Res 2004; 32:D452–55.
- Hoyt CT, Balk M, Callahan TJ et al. Unifying the identification of biomedical entities with the bioregistry. Scientific Data 2022;9:714.
- Jackson R, Balhoff J, Douglass E *et al*. Robot: a tool for automating ontology workflows. *BMC Bioinformatics* 2019;20:407.
- Jonquet C, Toulet A, Arnaud E et al. Agroportal: a vocabulary and ontology repository for agronomy. Comput Electron Agric 2018; 144:126–43.
- Jupp S, Burdett T, Leroy C et al. A new ontology lookup service at EMBL-EBI. In: Proceedings of the 8th International Conference on Semantic Web Applications and Tools for Life Sciences, Cambridge, 2015. p. 118–9. Cambridgeshire, UK: CEUR Workshop Proceedings, 2015.
- Lambert SA, Gil L, Jupp S *et al*. The polygenic score catalog as an open database for reproducibility and systematic evaluation. *Nat Genet* 2021;53:420–5.
- Lobentanzer S, Aloy P, Baumbach J et al. Democratizing knowledge representation with biocypher. Nat Biotechnol 2023;41:1056–9.
- Malone J, Adamusiak T, Holloway E et al. Developing an application ontology for annotation of experimental variables—experimental factor ontology. Nat Prec 2009. https://doi.org/10.1038/npre.2009.3806.1
- Matentzoglu N, Balhoff JP, Bello SM et al. A simple standard for sharing ontological mappings (SSSOM). Database 2022;2022:baac035. https://doi.org/10.1093/database/baac035
- Matentzoglu N, Bello S, Stefancsik R et al. The unified phenotype ontology (uPheno): a framework for cross-species integrative phenomics. Genetics 2025;229:yaf027. https://doi.org/10.1093/genetics/iyaf027
- Moreno P, Fexova S, George N et al. Expression atlas update: gene and protein expression in multiple species. Nucleic Acids Res 2022;50:D129–40.
- Mungall C, Torniai C, Gkoutos G et al. Uberon, an integrative multispecies anatomy ontology. Genome Biol 2012;13:R5.
- Noy NF, Shah NH, Whetzel PL *et al.* Bioportal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Res* 2009;37:W170–3.
- Ochoa D, Hercules A, Carmona M et al. The next-generation open targets platform: reimagined, redesigned, rebuilt. *Nucleic Acids Res* 2023;51:D1353–59.
- Ong E, Xiang Z, Zhao B et al. Ontobee: a linked ontology data server to support ontology term dereferencing, linkage, query and integration. Nucleic Acids Res 2017;45:D347–52.
- Osumi-Sutherland D, Courtot M, Balhoff J et al. Dead simple owl design patterns. J Biomed Semantics 2017;8:18.
- Perez-Riverol Y, Bandla C, Kundu DJ et al. The pride database at 20 years: 2025 update. Nucleic Acids Res 2025;53:D543-53.
- Perova Z, Martinez M, Mandloi T et al. Abstract 6910: cancermodels. org: an open global cancer research platform for patient-derived cancer models. Cancer Res 2024;84:6910. https://doi.org/10.1158/ 1538-7445.am2024-6910
- Sarkans U, Gostev M, Athar A *et al.* The biostudies database-one stop shop for all data supporting a life sciences study. *Nucleic Acids Res* 2018;46:D1266–70.
- Sayers EW, Cavanaugh M, Clark K et al. Genbank 2024 update. Nucleic Acids Res 2023;52:D134–7. https://doi.org/10.1093/nar/gkad903
- Schmidt-Rhaesa A. The Evolution of Organ Systems. London, England: Oxford University Press, 2007.
- Smith C, Goldsmith C-A, Eppig J. The mammalian phenotype ontology as a tool for annotating, analyzing and comparing phenotypic information. *Genome Biol* 2005;6:R7.
- Steinbeck C, Koepler O, Herres-Pawlis S et al. Nfdi4chem—a research data network for international chemistry. Chem Int 2023;45:8–13.
- Vasilevsky N, Matentzoglu N, Toro S *et al.* Mondo: unifying diseases for the world, by the world. bioRxiv, https://doi.org/10.1101/2022. 04.13.22273750, 2022, preprint: not peer reviewed.

© The Author(s) 2025. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited. Bioinformatics, 2025, 41, 1–4 https://doi.org/10.1093/bioinformatics/btaf279

Applications Note