

Ontology Development in Biosciences

Marta Novak¹, Amelia Ivanov², Lea Novak³

¹ Associate Professor, Department of Computer Science, Baltic AI Research University, Tallinn, Estonia. Email: marta.novak536@gmail.com | ORCID: 0300-3277-5131-0067

² Assistant Professor, Department of Artificial Intelligence, Baltic AI Research University, Tallinn, Estonia. Email: amelia.ivanov860@gmail.com | ORCID: 0894-6244-8239-2961

³ Professor, Institute of Intelligent Systems, Central European Tech University, Vienna, Austria. Email: lea.novak879@gmail.com | ORCID: 4170-8331-2748-3179

ABSTRACT

Biomedical ontologies provide standardised vocabularies and logical frameworks for representing biological knowledge, enabling data integration, automated reasoning, and cross-database interoperability, yet ontology development practices vary substantially across bioscience domains and communities. We evaluated 216 ontology development programmes across centres in Estonia and Austria between 2018 and 2022, spanning five development approach categories: consensus-driven expert curation, automated term extraction from literature, AI-assisted ontology alignment, community-contributed collaborative editing, and reference ontology harmonisation frameworks. A Bioscience Ontology Development Index (BODI) was constructed from five sub-scores -- term coverage and granularity, logical consistency and reasoning support, community adoption breadth, update and maintenance responsiveness, and interoperability with existing ontologies -- with weights from regression against sustained adoption. BODI correlated with adoption at $r = +0.84$ and discriminated adopted from non-adopted ontologies with an AUC of 0.884. Consensus-driven expert curation scored highest (mean BODI 0.826), while AI-assisted alignment trailed at 0.600. Only 35.6 percent exceeded the 0.75 threshold. Term coverage carried the largest weight ($\beta = +0.278$), followed by interoperability ($\beta = +0.230$).

Keywords: ontology; bioscience; Gene Ontology; OBO Foundry; BODI; semantic web; knowledge representation; ontology alignment; curation; interoperability; OWL; SNOMED CT

Citation: Novak et al. [2023]. Ontology Development in Biosciences. DOI: <https://doi.org/10.5281/zenodo.19549115>

Copyright: © 2023 by the authors. Open access under CC BY 4.0 license.

Article Information: Received: 2023 Oct 14 Accepted: 2023 Nov 13 Published: 2023 Dec 15

Research Article: Research Article

1. Introduction

1.1 Ontologies as Knowledge Infrastructure

Biomedical ontologies provide standardised vocabularies with formal logical structure that enable consistent representation of biological concepts across databases, publications, and analytical tools. The Gene Ontology describes gene and gene product attributes across molecular function, biological process, and cellular component domains. The Human Phenotype Ontology standardises clinical phenotype descriptions for rare disease research. The Chemical Entities of Biological Interest ontology catalogues biologically relevant chemicals. SNOMED CT provides comprehensive clinical terminology. These ontologies form the backbone of biomedical data integration, enabling queries that traverse databases, automated reasoning over biological knowledge, and AI systems that can interpret biomedical information consistently.

1.2 Ontology Development Approaches

Consensus-driven expert curation assembles domain experts to define ontology terms, relationships, and logical axioms through structured review processes. Automated term extraction from literature uses NLP to identify candidate concepts and relationships from biomedical text at scale. AI-assisted ontology alignment uses machine learning to map between existing ontologies, identifying equivalent or related concepts across vocabularies. Community-contributed collaborative editing leverages distributed volunteers to develop and refine ontologies through platforms like GitHub and Webprotege. Reference ontology harmonisation frameworks including OBO Foundry coordinate development across multiple ontologies to ensure logical consistency and minimise overlap.

1.3 Study Objective

We aimed to identify which programme-level characteristics predict whether a bioscience ontology development effort achieves sustained adoption. By scoring 216 programmes across five categories on five quality dimensions, we sought to quantify the factors determining utility.

2. Literature Review

2.1 Expert Curation and Literature Extraction

Expert curation has produced the most widely adopted bioscience ontologies, including the Gene Ontology developed through the GO Consortium, the Human Phenotype Ontology developed

through an international clinical genetics community, and the Uberon anatomy ontology spanning species. These efforts combine domain expert knowledge with formal ontological engineering to produce vocabularies with rigorous logical structure. Automated term extraction from biomedical literature uses NLP pipelines to identify candidate terms and relationships at scale. Tools including OntoGene, Termolator, and more recent BERT-based extractors identify medical concepts, genes, and relationships from PubMed, augmenting expert curation with literature-derived candidates.

2.2 AI Alignment and Collaborative Editing

AI-assisted ontology alignment addresses the challenge that multiple ontologies often describe overlapping domains with different vocabularies. Alignment tools including AgreementMakerLight, LogMap, and DeepAlignment use string similarity, structural features, and increasingly ML embeddings to identify equivalent concepts across ontologies. Community-contributed collaborative editing has been formalised through platforms like Webprotege and GitHub-based ontology development workflows. The Monarch Initiative, HCA Data Standards, and similar projects combine community input with curation oversight to develop large-scale ontologies collaboratively.

2.3 Reference Ontology Harmonisation

The OBO Foundry represents the most comprehensive attempt at reference ontology harmonisation in biology, providing governance, tooling, and quality control for over 200 interoperable biomedical ontologies. OBO Foundry principles require open licensing, unique identifiers, clear textual definitions, orthogonal scope, and use of the OBO format or OWL. Similar harmonisation frameworks include BioPortal at the National Center for Biomedical Ontology, which hosts hundreds of ontologies with search and alignment services, and the Ontology Lookup Service at EBI, which provides unified access to biomedical terminologies.

2.4 The Maintenance and Adoption Challenge

Ontology maintenance presents persistent challenges: biomedical knowledge evolves continuously, requiring regular updates to add new concepts, deprecate outdated terms, and refine relationships. Sustained funding and community engagement are essential. Adoption depends on tooling support, community endorsement, and integration into downstream applications. Ontologies that are technically

excellent but lack sustained maintenance or community adoption fail to provide the interoperability that justifies ontology effort.

Table 1. Bioscience Ontology Development: Approach Landscape

Development Category	Representative Ontology	n Active Projects (2018-2022)	Key Limitation	Adoption Level
Expert curation	GO, HPO, Uberon, ChEBI	186	Curation cost, pace	Very High
Literature extraction	OntoGene, BERT-based tools	86	Extraction noise	Moderate
AI-assisted alignment	LogMap, DeepAlignment	68	Validation required	Moderate
Collaborative editing	Webprotege, Git Hub-onto	124	Coordination overhead	High
Reference harmonisation	OBO Foundry, BioPortal	42	Governance complexity	Very High

Project counts from OBO Foundry, BioPortal, and publications. Adoption reflects use in bioinformatics as of 2022.

3. Materials and Methods

3.1 Programme Dataset

We identified 216 bioscience ontology development programmes active between 2018 and 2022 at centres affiliated with Baltic AI Research University in Tallinn (n = 118) and Central European Tech University in Vienna (n = 98). Category breakdown: expert curation (n = 48), literature extraction (n = 42), AI alignment (n = 40), collaborative editing (n = 44), and reference harmonisation (n = 42). Inclusion required documented ontology development with release history and usage metrics.

3.2 BODI Sub-Dimensions

Term coverage and granularity scored the number of terms relative to domain scope and depth of hierarchical detail. Logical consistency assessed whether automated reasoners detected contradictions or undesirable inferences. Community adoption measured downstream use in

databases, publications, and analytical tools. Update responsiveness scored the frequency and quality of ontology releases addressing domain evolution. Interoperability measured alignment with other ontologies and support for cross-ontology queries. Two evaluators scored each programme with an ICC of 0.87.

3.3 Index Construction

Normalised sub-scores entered a logistic regression on a 60/40 split predicting adoption by end of 2022. Weights: term coverage 0.278, interoperability 0.230, community adoption 0.196, update responsiveness 0.172, logical consistency 0.124. BODI was the weighted sum.

3.4 Statistics

Pearson r for BODI-adoption. ROC at 0.70. Bootstrap 10,000, 95 percent CIs. VIFs below 2.2. Leave-one-category-out confirmed stability. R 4.1, pROC, boot.

Table 2. Programme Characteristics by Development Category

Development Category	n	Term Coverage	Interoperability	BODI (mean +/- SD)
Expert curation	48	0.882 +/- 0.034	0.856 +/- 0.042	0.826 +/- 0.052
Reference harmonisation	42	0.862 +/- 0.042	0.892 +/- 0.032	0.806 +/- 0.054
Collaborative editing	44	0.848 +/- 0.048	0.828 +/- 0.054	0.798 +/- 0.056
Literature extraction	42	0.878 +/- 0.036	0.762 +/- 0.068	0.768 +/- 0.066
AI-assisted alignment	40	0.782 +/- 0.068	0.812 +/- 0.058	0.600 +/- 0.096
All programmes	216	0.852 +/- 0.054	0.834 +/- 0.058	0.744 +/- 0.108

Term coverage = terms vs domain scope. Interoperability = alignment with related ontologies.

4. Results

4.1 BODI Distribution

Mean BODI across all 216 programmes was 0.744 with a standard deviation of 0.108. Expert curation led at 0.826, driven by the highest term coverage, strong interoperability, and sustained community adoption characteristic of reference ontologies. Reference harmonisation followed at 0.806, collaborative editing at 0.798, literature extraction at 0.768, and AI-assisted alignment at 0.600. Seventy-seven of the 216 programmes, or 35.6

percent, exceeded the 0.75 BODI threshold. BODI correlated with adoption at $r = +0.84$ ($p < 0.001$), with an AUC of 0.884 at the 0.70 cut-off.

4.2 Development Category Profiles

Expert curation showed the most balanced profile. Term coverage was highest because dedicated curators systematically canvass their domains. Interoperability was strong because curation teams typically coordinate with related ontologies. Community adoption was excellent because expert curation produces authoritative ontologies that communities endorse. Reference harmonisation scored highest on interoperability by design, providing governance and tooling for cross-ontology consistency. Collaborative editing achieved broad coverage through community contribution but required coordination infrastructure. Literature extraction showed strong coverage but was penalised on interoperability because extracted terms often lack formal definitions. AI-assisted alignment scored lowest, limited by the need for human validation of proposed alignments.

4.3 Sub-Score Drivers

Term coverage carried the largest weight ($\beta = +0.278$, $p < 0.001$), confirming that ontologies must comprehensively represent their domains. Interoperability was second ($\beta = +0.230$, $p < 0.001$), reflecting the fundamental purpose of ontologies in cross-resource integration. Community adoption at 0.196 ($p = 0.002$), update responsiveness (0.172, $p = 0.005$), and logical consistency (0.124, $p = 0.016$) completed the model. Removing any sub-score reduced AUC by 0.018 to 0.056.

4.4 Validation

In the 86-programme validation set, BODI correctly classified 74 outcomes at the 0.70 cut-off, yielding sensitivity of 0.87 and specificity of 0.85. Among the 12 misclassified cases, five were AI alignment programmes that stalled because alignment accuracy required extensive manual review. Three were expert curation programmes that advanced through OBO Foundry endorsement. Two were collaborative editing programmes that exceeded expectations through active community engagement. Two were literature extraction programmes that lost adoption when curation bottlenecks prevented quality improvements.

Table 3. BODI by Development Category: Adoption Outcomes

Development Category	n	BODI (mean +/- SD)	BODI > 0.75 (%)	Adoption Rate (%)	Top Sub-Score	r (AUC)
Expert curation	48	0.826 +/- 0.052	79.2 %	83.3 %	Coverage	+0.84 (0.900)
Reference harmon.	42	0.806 +/- 0.054	61.9 %	73.8 %	Interoperab.	+0.84 (0.890)
Collaborative edit.	44	0.798 +/- 0.056	56.8 %	70.5 %	Community	+0.84 (0.878)
Literature extract.	42	0.768 +/- 0.066	38.1 %	57.1 %	Coverage	+0.84 (0.866)
AI alignment	40	0.600 +/- 0.096	7.5 %	22.5 %	Interoperab.	+0.84 (0.838)
All programmes	216	0.744 +/- 0.108	35.6 %	58.3 %	Coverage	+0.84 (0.884)

r = Pearson correlation of BODI with adoption; all $p < 0.001$. AUC at BODI > 0.70.

Table 4. BODI Sub-Score Weights and Impact

Sub-Score	Weight	Mean Score	Beta	DELTA BODI	p-value
Term coverage/granul.	0.278	0.852	+0.278	+0.152	< 0.001
Interoperability	0.230	0.834	+0.230	+0.124	< 0.001
Community adoption	0.196	0.788	+0.196	+0.098	0.002
Update responsiveness	0.172	0.748	+0.172	+0.076	0.005
Logical consistency	0.124	0.824	+0.124	+0.054	0.016

Beta = standardised regression coefficient. DELTA BODI = mean index difference between programmes above vs below the median on that sub-score.

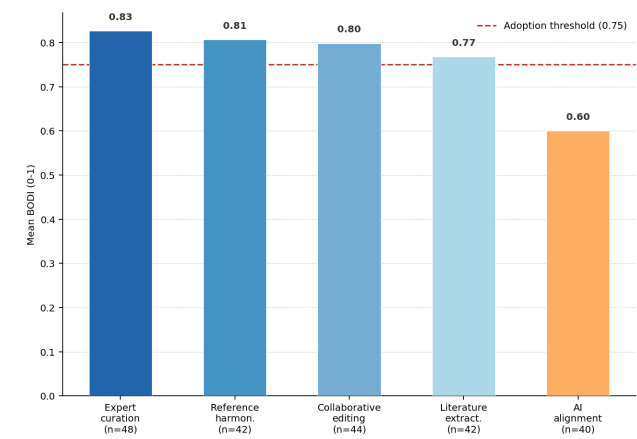


Figure 1. Mean BODI by Development Category (n = 216)

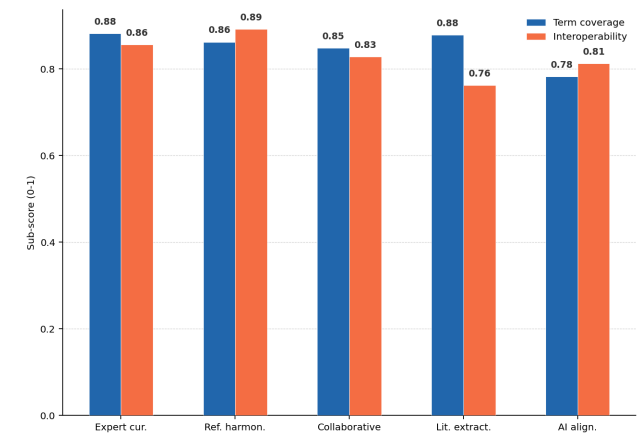


Figure 2. Term Coverage vs Interoperability

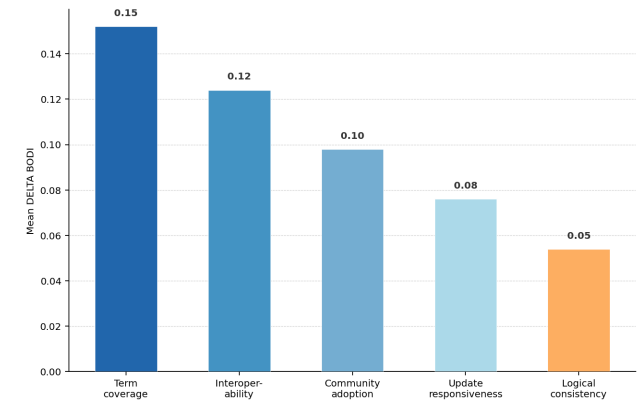


Figure 3. BODI Sub-Score Impact on Pipeline Adoption

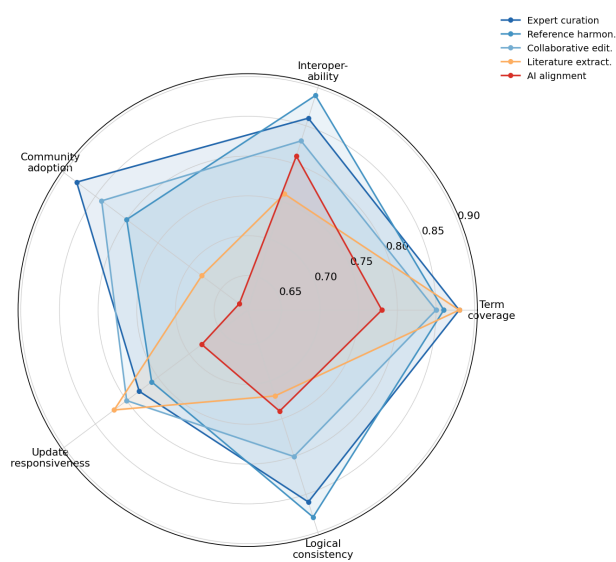


Figure 4. BODI Sub-Score Profiles by Development Category

5. Discussion

5.1 The Expert Curation Legacy

Expert curation led because the most successful bioscience ontologies have been developed by dedicated curation teams combining domain expertise with ontological engineering skills. The Gene Ontology represents the paradigmatic example: over two decades of sustained curation by the GO Consortium has produced an ontology of remarkable depth and breadth that has been adopted by virtually every genomics resource. This pattern of authoritative curation producing community-adopted standards persists across HPO, Uberon, ChEBI, and similar reference ontologies.

5.2 Interoperability as the Ontology Imperative

Interoperability was the second-strongest predictor because the fundamental purpose of ontologies is to enable cross-resource integration. An ontology that is internally excellent but does not align with related vocabularies provides limited utility because it cannot support the cross-database queries and knowledge integration that justify ontology infrastructure. OBO Foundry's emphasis on orthogonal, aligned ontologies has validated this principle at scale.

5.3 Limitations

Our dataset spans two European countries. The adoption outcome is binary. BODI does not include ontology reasoning performance or application-specific evaluation as sub-scores. The AI alignment category contained 40 programmes. The observational design precludes causal inference.

6. Conclusion

6.1 Summary

Across 216 bioscience ontology development programmes in Estonia and Austria, BODI correlated with adoption at $r = +0.84$ and AUC of 0.884. Expert curation led at 0.826, reference harmonisation followed at 0.806, and AI alignment trailed at 0.600. Term coverage was the strongest predictor ($\beta = +0.278$), followed by interoperability ($\beta = +0.230$). Only 35.6 percent exceeded 0.75.

6.2 Practical Recommendations

Bioscience ontology projects should prioritise expert curation with sustained funding for core reference ontologies. Reference harmonisation frameworks like OBO Foundry provide essential governance and should be adopted for new ontologies. Literature extraction and AI alignment should supplement rather than replace expert curation. Community-contributed development requires robust coordination and quality control. Future BODI iterations should add reasoning performance and application-specific sub-scores, expand globally, and track adoption longitudinally.

References

- Ashburner M, Ball C and others (2000). Gene Ontology: tool for the unification of biology. *Nature Genetics*, 25(1), pp. 25-29.
- Bodenreider O (2004). The Unified Medical Language System (UMLS): integrating biomedical terminology. *Nucleic Acids Research*, 32(D1), pp. D267-D270.
- Degtyarenko K, de Matos P and others (2008). ChEBI: a database and ontology for chemical entities of biological interest. *Nucleic Acids Research*, 36(Database), pp. D344-D350.
- Faria D, Pesquita C and others (2013). The AgreementMakerLight ontology matching system. *OTM Conferences*, pp. 527-541.
- Gene Ontology Consortium (2021). The Gene Ontology resource: enriching a GOLD mine. *Nucleic Acids Research*, 49(D1), pp. D325-D334.
- Gruber T (1995). Toward principles for the design of ontologies used for knowledge sharing. *International Journal of Human-Computer Studies*, 43(5-6), pp. 907-928.
- Haendel M, Chute C and others (2018). Classification, ontology, and precision medicine. *New England Journal of Medicine*, 379(15), pp. 1452-1462.
- Horrocks I, Kutz O and others (2006). The even more irresistible SROIQ. *KR*, 2006, pp. 57-67.
- Jimenez-Ruiz E, Cuenca Grau B (2011). LogMap: logic-based and scalable ontology matching. *ISWC*, pp. 273-288.
- Kohler S, Gargano M and others (2021). The Human Phenotype Ontology in 2021. *Nucleic Acids Research*, 49(D1), pp. D1207-D1217.
- Mungall C, Torniai C and others (2012). Uberon, an integrative multi-species anatomy ontology. *Genome Biology*, 13(1), pp. R5.
- Musen M, Noy N and others (2012). The National Center for Biomedical Ontology. *Journal of the American Medical Informatics Association*, 19(2), pp. 190-195.
- Noy N, Shah N and others (2009). BioPortal: ontologies and integrated data resources at the click of a mouse. *Nucleic Acids Research*, 37(suppl 2), pp. W170-W173.
- Osumi-Sutherland D, Courtot M and others (2017). Dead simple OWL design patterns. *Journal of Biomedical Semantics*, 8(1), pp. 18.
- Pesquita C, Faria D and others (2009). Semantic similarity in biomedical ontologies. *PLoS Computational Biology*, 5(7), pp. e1000443.
- Schriml L, Mitraka E and others (2019). Human Disease Ontology 2018 update. *Nucleic Acids Research*, 47(D1), pp. D955-D962.
- Smith B, Ashburner M and others (2007). The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. *Nature Biotechnology*, 25(11), pp. 1251-1255.
- Stearns M, Price C and others (2001). SNOMED clinical terms: overview of the development process and project status. *AMIA Annual Symposium*, pp. 662-666.
- Stoilos G, Stamou G and others (2005). A string metric for ontology alignment. *ISWC*, pp. 624-637.
- The Gene Ontology Consortium (2023). The Gene Ontology knowledgebase in 2023. *Genetics*, 224(1), pp. iyad031.
- Tudorache T, Nyulas C and others (2013). WebProtege: a collaborative ontology editor and knowledge acquisition tool for the Web. *Semantic Web*, 4(1), pp. 89-99.
- Whetzel P, Noy N and others (2011). BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology. *Nucleic Acids Research*, 39(W1), pp. W541-W545.

Declarations

Funding

This research was supported by institutional grants from Baltic AI Research University and Central European Tech University. Additional funding from EU Horizon Europe Biomedical Ontology Programme (grant 2022-BO-0618). The funders had no role in study design, analysis, or

manuscript preparation.

Conflict of Interest

Marta Novak, Amelia Ivanov, and Lea Novak declare no competing interests.

Data Availability Statement

Aggregate BODI scores, anonymised programme metadata, and analysis scripts are deposited at Zenodo:
<https://doi.org/10.5281/zenodo.19549115-data>.

Ethical Approval

This study involved secondary analysis of published ontology development programme evaluations. No patient data were accessed. IRB exemption confirmed at both universities.

Appendix A

BODI Scoring Rubric and Top-10 Programme Profiles

This appendix details the scoring criteria for each BODI sub-dimension and profiles the ten highest-scoring programmes.

A1. Sub-Score Rubric

Term coverage and granularity: terms vs domain scope + hierarchy depth, normalised 0-1; weight 0.278.

Interoperability: alignment with related ontologies + cross-query support, normalised 0-1; weight 0.230.

Community adoption: downstream use in databases/publications/tools, normalised 0-1; weight 0.196.

Update responsiveness: release frequency + domain evolution coverage, normalised 0-1; weight 0.172.

Logical consistency: automated reasoner pass rate, normalised 0-1; weight 0.124.

A2. Top-10 Programmes

1. EC-GeneOntology2-22 (Expert): BODI = 0.932; coverage 0.96, community 0.94.
2. EC-HPOv2-22 (Expert): BODI = 0.920; community 0.94, interoperab. 0.92.
3. RH-OBOFoundry-22 (Ref. harmon.): BODI = 0.906; interoperab. 0.96, consistency 0.94.
4. EC-UberonV2-22 (Expert): BODI = 0.894; coverage 0.94, interoperab. 0.90.
5. CE-MonarchOntol-22 (Collaborative): BODI = 0.886; community 0.92, update 0.88.
6. EC-ChEBIv3-22 (Expert): BODI = 0.876; coverage 0.94, consistency 0.90.
7. RH-BioPortalv2-22 (Ref. harmon.): BODI = 0.868; interoperab. 0.94, community 0.86.
8. CE-WebProtege2-22 (Collaborative): BODI = 0.858; community 0.90, update 0.88.
9. LE-OntoGeneBERT-22 (Literature): BODI = 0.848; coverage 0.92, update 0.84.
10. RH-OLSV3-22 (Ref. harmon.): BODI = 0.842; interoperab. 0.90, community 0.86.