MCBO: Mammalian Cell Bioprocessing Ontology, A Hub-and-Spoke, IOF-Anchored Application Ontology

Kimberly Robasky[1,*,†], James Morrissey[2,†], Markus Riedl[3,†], Andreas Dräger[4,†], Nicole Borth[3], Michael J. Betenbaugh[2], and Nathan E. Lewis[1,†]

[1] *Center for Molecular Medicine, Complex Carbohydrate Research Center, and Department of Biochemistry and Molecular Biology, University of Georgia, Athens, Georgia, USA*

[2] *Department of Chemical and Biomolecular Engineering, Johns Hopkins University, Baltimore, Maryland, USA*

[3] *Department of Biotechnology, BOKU University, Vienna, Austria*

[4] *Data Analytics and Bioinformatics, Institute of Computer Science, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany*

Abstract

The International Biomanufacturing Network (IBioNe) aims to accelerate discoveries and developments by providing a network of biomanufacturing training and workforce development to educate the next generation of biomanufacturing experts. This effort includes the construction of tools and data for sharing knowledge around bioprocessing with mammalian cells. In this effort, we are developing a central datahub. Currently, knowledge-sharing suffers from several problems, which can be addressed as publicly-available data are curated and assembled. There, tools can be provided to explore and model these data to better understand and optimize mammalian biopharmaceutical manufacturing processes (biomanufacturing of biologics). One of the most impactful challenges is linking protein outputs to known production cell line genetic variation, phenotypes, and bioreactor conditions. Our datahub will serve public data and tools for overcoming these other challenges. In future work, we will publish our semi-automated metadata extraction pipelines (including LLM-assisted approaches) aligned to MCBO, currently in progress. Here, we report on preliminary work and design plans for a new application ontology, the *Mammalian Cell Bioprocessing Ontology* (MCBO), for organizing and accelerating efforts in AI-ready data to enable our understanding of protein production, facilitating more efficient biopharmaceutical manufacturing of biologics. Availability: https://github.com/lewiscelllabs/mcbo

Keywords

Bioprocessing, application ontology

* Corresponding author.
† These authors contributed equally.
kimberly.robasky@uga.edu (K. Robasky); rmorri68@jh.edu (J. Morrissey); markus.riedl@boku.ac.at (M. Riedl); andreas.draeger@informatik.uni-halle.de (A. Dräger); nicole.borth@boku.ac.at (N. Borth); beten@jhu.edu (M. Betenbaugh); natelewis@uga.edu (N. E. Lewis)
0000-0002-0090-8698 (K. Robasky); 0000-0003-0528-2257 (J. Morrissey); 0000-0001-6599-4702 (M. Riedl); 0000-0002-1240-5553 (A. Dräger); 0000-0001-6324-9338 (N. Borth); 0000-0002-6336-4659 (M. Betenbaugh); 0000-0001-7700-3654 (N. E. Lewis)

1. Introduction

Mammalian cells, such as Chinese hamster ovary (CHO) or Human Embryonic Kidney (HEK) cells, are vital in biopharmaceutical manufacturing because they can produce therapeutic proteins, such as monoclonal antibodies, with human-like post-translational modifications, which reduces their immunogenicity in patients[1]. CHO cells are of particular importance. Their ability to be cultured at large industrial scales with high production yields, along with their strong, established history of regulatory approval for their derived products, makes them the "gold standard" for producing biologics[2]. Not only are CHO cells central to biologics manufacturing, they come with decades of accumulated public data[3] and extensive proprietary datasets maintained by industrial partners. These include diverse cell lines, each with proprietary knock-ins or expression cassettes, and rich phenotypic profiles generated under a wide range of cultivation conditions. Yet, despite this data abundance, it remains frustratingly difficult to answer even basic queries (e.g., "*what culture conditions best support a given protein of interest?*", or "*which mutations improve yield for a particular CHO lineage?*").

Ontologies provide a structured, hierarchical, and machine-readable representation of a domain, making them a natural solution to these challenges. By formally capturing relationships among cultivation conditions, genetic modifications, and phenotypic outcomes, they can transform scattered datasets into interoperable knowledge that is both human-interpretable and machine-actionable. Beyond this core role, ontologies also complement Large Language Models (LLMs) by adding explicit background knowledge and worldview models to reconcile conflicting perspectives in LLM training data, while enabling tasks such as semantic search and LLM prompt expansion through synonym mappings and domain-specific terminology. Because ontology reasoning is computationally lightweight[4] compared to LLM inference[5] and avoids the overhead of traversals in relational or document-based databases, it provides a practical, cost-efficient foundation for scalable bioprocess data integration[6, 7].

While the broader biomedical community has developed mature ontologies (Gene Ontology[8], ChEBI[9], Cell Line Ontology[10]), and the systems biology community has created sophisticated metabolic models (CHO-GEMs like *i*CHO3K[11]), no comprehensive formal ontology exists yet, specifically for mammalian cell cultivation experimental workflows, for applications in bioprocess development. Existing efforts focus on either broad biological concepts or mathematical modeling frameworks, leaving a gap in practical experimental metadata representation. Recent initiatives by NIST's Industrial Ontology Foundry (IOF [12, 13]) for biopharmaceutical manufacturing show promise, using Basic Formal Ontology (BFO[14]) as a foundation for manufacturing process ontologies. However, these efforts have not yet addressed the specific needs of mammalian cell cultivation experimental data capture and integration.

To address this gap, we present a lightweight but extensible mammalian cell cultivation application ontology[15] (AO). Our AO (spoke) is anchored to the IOF (hub), and is authored in *Web Ontology Language* (OWL 2), is *Descriptive Logic* (DL)-compatible, so it already supports standard DL reasoners (e.g., HermiT, ELK). While we have not yet added a *Simple Knowledge Organization System* (SKOS) layer, the model is fully compatible with SKOS for

synonym and vocabulary management, providing a clear path for standards-compliant expansion. Our goal is not to over-engineer from the start, but to offer a practical first step: a modular, open-source framework that can be deployed internally (e.g., behind firewalls), supports data validation, *Electronic Lab Notebook* (ELN) rendering, and enables progressive enhancement toward *Findable, Accessible, Interoperable, Reusable* (FAIR) data goals and emerging repositories.

2. Scope and Competency Questions

We use these competency questions to set the *minimum necessary scope* for MCBO and to guide which upper-ontology commitments and design patterns must be supported. Subsequent sections describe the resulting design choices and how they were implemented and evaluated.

The primary motivation for this ontology is to lay the groundwork for the datahub we are building to support mining of public data. To that end, the scope is limited to the bare essentials needed to cover the metadata currently captured, and in that way, we avoid over-engineering. More terms will be added as needed, with high ontological commitment, leaving less ambiguity, as is appropriate for an AO. The first release of the datahub will focus on transcriptomic data and key bioprocess parameters, with an emphasis on understanding their interconnections. As we are just now building the datahub, we define rudimentary *Competency Questions* (CQs) for testing MCBO coverage as follows:

CQ1: Under what culture conditions (pH, dissolved oxygen, temperature) do the cells (e.g., CHO-K1) reach peak recombinant protein productivity?

CQ2: Which CHO cell lines have been engineered to overexpress gene Y?

CQ3: Which nutrient concentrations are most associated with viable cell density above Z at day 6 of culture?

CQ4: How does the expression of gene X vary between clone A and clone B?

CQ5: What pathways are differentially expressed under Fed-batch vs Perfusion in CHO-K1?

We further simplify these CQs in the evaluation section in order to align with the limited metadata currently curated. We will continue to version and share the ontology on GitHub as more data, terms, and CQs are added to support the burgeoning datahub.

3. Ontology Design

The *Mammalian Cell Bioprocessing Ontology* (MCBO) targets mammalian cell cultivation experiments, and is engineered for practical utility without sacrificing semantic rigor. It is formally represented using constructs from the OWL 2 ontology language with standard RDFS constructs and aligns with best practices in biomedical ontology engineering.

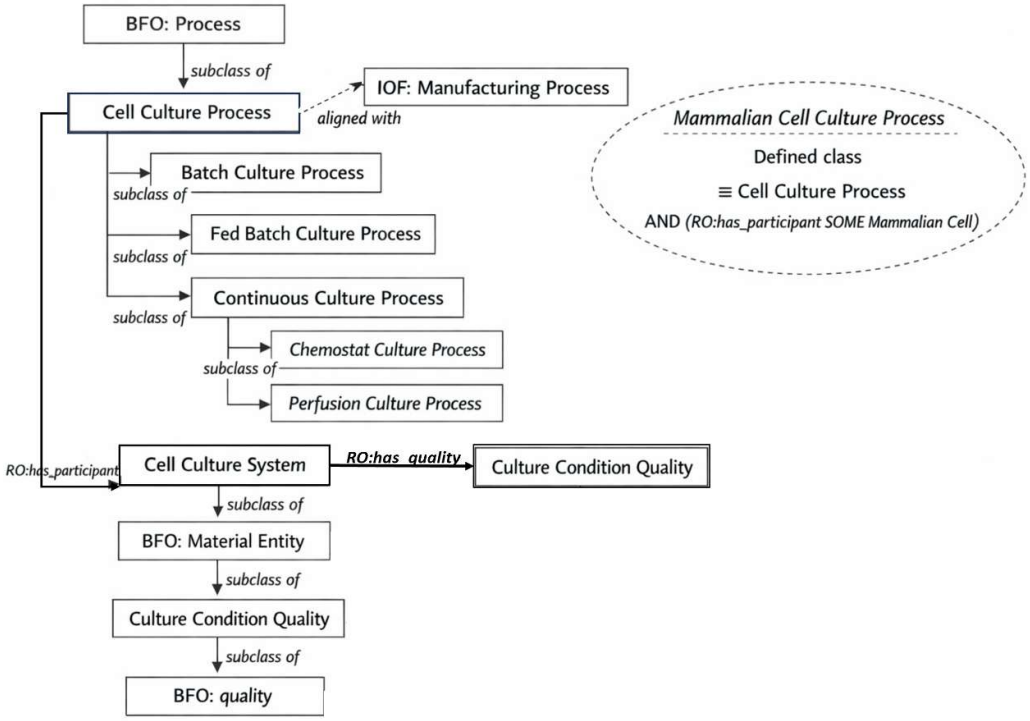

Figure 1 Core process and material entity modeling pattern in MCBO. Generic cell culture processes are organized independently of organism type, with batch, fed-batch, and continuous subclasses. Mammalian cell culture processes are represented as OWL defined classes using participant restrictions rather than subclassing. Culture conditions are modeled as qualities of the material cell culture system, consistent with BFO principles. Alignment with the IOF Manufacturing Process is shown conceptually without asserting subclass relations.

MCBO is an application ontology that reuses BFO and IOF classes where appropriate but does not yet claim full ontological normalization under the IOF Core. Process-level constructs are currently aligned to IOF production-related classes as pragmatic anchors for downstream data integration, rather than as finalized commitments to business-level process semantics.

At the core is the abstract concept *CellCultureProcess*, from which specific operational modalities inherit: *BatchCultureProcess* (closed-system culture), *FedBatchCultureProcess* (batch with discrete nutrient additions), *PerfusionCultureProcess* (continuous nutrient feed with waste removal and cell retention), and *ChemostatCultureProcess* (continuous nutrient feed with waste removal). Each subclass inherits properties from its parent while introducing specialized attributes.

We distinguish generic cell culture from mammalian cell culture. **MammalianCellCultureProcess** is modeled as a *defined class*: *CellCultureProcess* and (*RO:has_participant* some *MammalianCell*). This supports automatic classification and prevents inadvertently treating yeast/bacterial culture processes as mammalian.

Relationships among concepts are captured with *owl:ObjectProperty* (e.g., *RO:has_participant, RO:has_quality*, *inCulturePhase*), literal attributes with *owl:DatatypeProperty* (e.g., *hasCollectionDay*, *hasViableCellDensity*), hierarchical structure with *rdfs:subClassOf*, and typing with *rdf:type*; when needed, *owl:Restriction* axioms provide tighter domain constraints. Variables and measurements follow a consistent pattern in which measurement classes (e.g., *CellViabilityMeasurement*, *ProductivityMeasurement*, *QualityMeasurement*, *GeneExpressionMeasurement*) specialize the OBO class *IAO:0000109* (*measurement datum*) and carry standardized values and units (e.g., *hasConcentrationValue*, *hasConcentrationUnit*, *haspH*, *hasTemperature*, *hasDissolvedOxygen*, *hasTiterValue*). These formalized relationships enable standardized annotation of experimental observations. Datasets, runs, and measurements are modeled as **information content entities** (IAO) that are about processes/material entities, while the underlying experimental processes and culture systems are BFO-aligned continuants/occurrents.

In MCBO, **culture conditions are modeled as qualities of a material entity (the *cell culture system*, e.g., bioreactor + medium + cells)**, not as qualities of the process. Accordingly, a **CellCultureProcess** *RO:has_participant* some **CellCultureSystem**, and the **CellCultureSystem** *RO:has_quality* some **CultureConditionQuality** (e.g., temperature, pH, dissolved oxygen setpoint). This preserves the BFO distinction that **processes do not bear qualities**; rather, qualities inhere in independent continuants that participate in processes. Figure 1 illustrates this pattern.

MCBO adheres to modular reuse, aligning with the Systems Biology Ontology (SBO)[16], Ontology for Biomedical Investigations (OBI)[17], and Units of Measurement Ontology (UO)[18] to maximize interoperability. Together, this semantic backbone enables expressive metadata annotation, reproducible modeling, ontology-driven form generation, automated and computationally-efficient reasoning, consistent data integration, and downstream schema validation.

MCBO is intentionally designed as a **lightweight application ontology** that prioritizes semantic correctness, reuse, and extensibility over exhaustive domain coverage. All core modeling decisions follow Basic Formal Ontology (BFO) constraints, with explicit reuse of Industrial Ontology Foundry (IOF), Relations Ontology (RO), Ontology for Biomedical Investigations (OBI), Information Artifact Ontology (IAO), and Units of Measurement Ontology (UO) where applicable.

To maintain BFO consistency, MCBO adopts the following design commitments:

- **Processes vs. continuants** are strictly separated. Experimental activities such as cultivation, feeding, and sampling are modeled as BFO:processes. Physical systems (e.g., bioreactors, media, cells) are modeled as material entities.
- **Qualities inhere only in continuants**, not processes. Culture conditions (e.g., temperature, pH, dissolved oxygen) are modeled as qualities of a *CellCultureSystem*, which participates in a *CellCultureProcess*. This avoids the common modeling error of assigning qualities directly to processes.

- **Organism specificity is handled through defined classes**, not hard-coded subclassing. *MammalianCellCultureProcess* is defined via participant restrictions rather than asserted as a primitive subclass, enabling automatic classification and preventing unintended inclusion of non-mammalian systems (e.g., yeast or bacterial cultures).
- **Information artifacts are explicitly distinguished from biological reality**. Datasets, runs, and measurements are modeled as IAO information content entities that are *about* processes and material entities, while the experimental systems themselves remain BFO-aligned continuants and occurrents.

MCBO reuses existing object properties from RO wherever possible. New relations are introduced only when no suitable standard property exists, and are documented with domain, range, and textual definitions. Measurement modeling follows a consistent pattern in which specialized measurement classes subclass *IAO:0000109 (measurement datum)* and carry standardized values and units.

This version of MCBO does **not yet claim full normalization under IOF Core or IOF Product Production Process semantics**. Alignment to IOF Manufacturing Process is treated as a pragmatic anchor for interoperability rather than a final commitment to business-level process aggregation. These alignments will be refined as IOF biopharmaceutical manufacturing extensions mature.

Finally, MCBO enforces basic ontology quality controls. All public-facing classes are expected to carry textual definitions, labels, and provenance annotations. Automated quality checks (e.g., missing definitions, orphan classes, duplicate labels) are run as part of development and reported alongside releases. This ensures MCBO remains evaluable as an ontology artifact and maintainable as it evolves.

4. Ontology Development

Our ontology development followed an agile, stakeholder-driven methodology designed to address immediate data integration challenges in the CHO biopharmaceutical manufacturing community. Rather than pursuing comprehensive domain coverage from the outset, we adopted an iterative approach, prioritizing practical utility and extensibility. Initial development was motivated by extensive consultation with industry partners. The academic community articulated sophisticated use cases that industry partners deemed overly complex for immediate implementation. Conversely, industry partners struggled to define specific functional requirements, requesting instead a general "knowledge base" without clear specifications. This gap led to our decision to develop a no-code data exploration solution as a bridge between communities. Through stakeholder consultation, one core use case emerged: *"What datasets exist that up-regulate my proteins of interest, and how are they characterized regarding genetic mutations, phenotypes, metabolic states, inputs, and cultivation conditions?"* This question drove our initial ontological scope and semantic modeling decisions as we parse and curate metadata across the vast landscape of publicly available data.

MCBO design was constrained by several technical imperatives: supporting agile schema evolution as new use cases emerge, enabling auto-generation of data collection interfaces

from semantic specifications, and facilitating LLM-assisted extraction of structured metadata from unstructured publications and ELNs. The AO was developed using Python-based tooling with *Terse RDF Triple Language* (TTL) representation, prioritizing integration with web technologies and schema validation frameworks. Development leveraged LLM-assisted analysis of existing publications and semi-structured data to identify common terminology and metadata patterns.

5. Evaluation

We assessed the ontology's ability to represent real-world experimental metadata by selecting three of our manually annotated, recent cultivation studies from the literature[19–21]. For each study, we identified key experimental metadata (cultivation strategy, cell line, feeding protocols, measurement types, etc.) and determined what percentage could be formally represented using our ontology terms. Domain experts (JM, NL) reviewed the ontology structure, term definitions, and hierarchical relationships for accuracy and completeness relative to current cultivation practices. Feedback focused on whether the ontology captures the essential concepts needed for experimental annotation and data integration. Comprehensive competency question validation will be performed after sufficient gene expression datasets have been integrated into the datahub repository we are developing. This incremental approach aligns with our core design principle of simple, progressive development. Once adequate data coverage is achieved, we will test the ontology's ability to support queries such as "*What cultivation conditions correlate with increased production of protein X*?" and measure query precision and recall.

As of this writing, we have successfully integrated 724 curated bioprocessing sample runs from published studies, where a run refers to RNA-Seq dataset collected on a sample. All evaluation artifacts are available in the github project repository, including (i) the RDF graph used for evaluation (or a publicly shareable subset), (ii) the SPARQL queries for each competency question, and (iii) query outputs sufficient to reproduce counts reported here. Runs are distributed across Batch (518), Fed-batch (135), Perfusion (49), and Unknown (22) processes. Results follow:

CQ1 (Culture optimization): 161 results correlating culture qualities (pH, dissolved oxygen, temperature) with medium-to-high productivity measurements

CQ2 (Cell engineering): The ontology supported retrieval of three engineered CHO line entries (CHO-DXB11, CHO-K1, CHO-S) where overexpression of target genes, using reported product type as a proxy for overexpressed gene(s), was reported. Productivity measurements across multiple runs were aggregated at the cell line level to support comparison. CQ2 intentionally operates at the cell line level, reflecting the biological interpretation of genetic engineering as a property of the lineage rather than individual culture runs.

CQ5 (Process comparison): 4 process types identified for comparative analysis (listed above).

All competency questions executed within sub-second response times on the 724-sample dataset. Multi-hop graph traversals successfully processed culture condition-productivity correlations. Reuse of OBO Foundry ontologies (OBI, BFO, ChEBI, GO) ensures interoperability with existing life sciences psemantic infrastructure. CQ1 results enable systematic culture optimization analysis across multiple studies; insights not easily

achievable with traditional tabular approaches. The ontology successfully harmonized heterogeneous experimental data into query-able knowledge graphs. Gene expression integration (CQ4) requires additional RNA-seq processing workflows, and the current dataset also lacks detailed nutrient concentration linkages. Overall, MCBO demonstrates effective bioprocessing data integration with 724 real observations and some limited practical analytical capabilities. The framework provides a foundation for systematic bioprocess optimization and cross-study comparative analysis.

Due to licensing/privacy constraints on curated metadata, we provide a public, runnable subset graph (eval/graph.sample.ttl) and corresponding query outputs; the full evaluation graph can be regenerated locally from restricted metadata using the provided pipeline.

6. Discussion

The ontology presented here addresses a critical gap in the mammalian bioproduction community: the lack of an appropriately specific, standardized, structured, and interoperable metadata framework for the development and integration of modern data and analysis workflows for the datahub we are developing. By defining a data ontology ahead of defining a structured data schema, we lay the foundation for unforeseen use cases and future-proof data modeling that would not otherwise be achievable. Rather than replacing existing ontologies such as the Systems Biology Ontology (SBO), our AO introduces specialized, non-overlapping (orthogonal) terms, while also integrating existing ontologies that are themselves orthogonal. Our work builds on and extends existing ontologies with practical, cultivation-specific terms and schema-anchored constraints. This layered approach adds actionable structure to existing vocabularies and fills gaps relevant to bioprocessing, where data curation remains ad hoc or proprietary. By making this system fully open-source and ontology-driven, it is deployable behind institutional or corporate firewalls, offering a pathway for industrial partners to integrate public and private data without exposing proprietary information. This separation of structure from content enables federated queries, interoperable ELNs, and foundation model refinement to specific organizational needs, on harmonized data, while maintaining IP sensitivity.

With these foundations, our datahub will support agentic and human-in-the-loop workflows where AI agents can execute semantic and traditional queries, subset, and visualize mammalian bioprocessing-related data without requiring users to write code. Bench scientists can explore large public datasets empowered by semantic search capabilities (e.g., "Create a dataset with Fed-batch cultures of CHO-K1 with glucose limitation and high IgG yield"), while AI agents render heatmaps, volcano plots, or growth curves based on structured metadata and gene expression counts. Users can invoke workflows to train predictive models for key outcomes (titer, viability) and trace feature importances to suggest underlying mechanisms (nutrient depletion, metabolic bottlenecks). To bootstrap this system, we are developing LLM-powered pipelines that parse unstructured metadata from unstructured legacy data (e.g., ELNs, publications) using our ontology and schema (in development) as target representations. This technology stack automates generation of structured, provenance-aware metadata records and pre-populated dataset submission forms. Authors can review AI-generated summaries of

experimental variables and metadata fields rather than manual data entry, reducing friction and increasing consistency across submissions to our datahub.

Our ontology-schema framework serves as a scalable substrate for building intelligent, user-friendly, and future-proof mammalian bioprocessing data infrastructures, paving the way toward value-rich, harmonized data curation, interoperable tooling, and AI-native research and manufacturing environments while grounding innovation in principled semantic foundations.

7. Future work

This initial release of the *Mammalian Cell Bioprocessing Ontology* (MCBO) represents the first formal application ontology specifically addressing mammalian cell cultivation experimental workflows in bioprocessing research, filling a significant gap in the biopharmaceutical manufacturing semantic infrastructure. While existing genome-scale metabolic models (CHO-GEMs) provide sophisticated mathematical frameworks for flux analysis[22], our ontology addresses the complementary need for standardized experimental metadata representation and semantic data integration. Future development will prioritize alignment with established initiatives, particularly the NIST Industrial Ontology Foundry (IOF) biopharmaceutical manufacturing ontologies built on Basic Formal Ontology (BFO)[23]. We will systematically map our cultivation concepts to IOF Core constructs and investigate integration pathways with industry standards from *BioPhorum Operations Group* (BPOG), *Bio-Process Systems Alliance* (BPSA), and *International Society for Pharmaceutical Engineering* (ISPE). Cross-mappings to external resources including SBO, OBI, CLO, and *Minimum Information About a Proteomics Experiment* (MIAPE)[24] will be formalized through community-standard annotation practices. SKOS definitions will be developed to strengthen semantic rigor. Near-term priorities include comprehensive competency question validation using real mammalian bioprocessing omics datasets, formal evaluation against existing literature, and expert review by industry practitioners, all driven by real use cases as we grow our datahub. We will establish clear pathways for community feedback through academic labs, biofoundries, and industrial partners.

Future iterations will support multi-omics datasets and extend to additional product quality attributes. Importantly, we will explore semantic bridges to existing CHO-GEM metabolic knowledge, potentially providing formal ontological grounding for metabolic model annotations and enabling integrated queries across cultivation conditions and metabolic flux predictions.

All ontology files, schema definitions, example data instances, and form renderers will be available under the permissive MIT license at https://github.com/lewiscelllabs/mcbo where community requests can be made through specialized issue templates and pull requests can be submitted, reviewed, and tracked. Longer-term, we aim to register with identifiers.org, and BioPortal, pursue OBO Foundry acceptance, and establish formal relationships with relevant standards bodies. This community-driven approach ensures the ontology evolves to meet real-world needs while maintaining semantic rigor and interoperability with the broader biomedical ontology ecosystem.

Acknowledgements

NEL, KR were supported in part by funding from the National Institutes for of Health (R35 GM119850).

References

[1] Wurm, F.M.: Production of recombinant protein therapeutics in cultivated mammalian cells. Nat Biotechnol. 22, 1393–1398 (2004).
[2] Walsh, G., Walsh, E.: Biopharmaceutical benchmarks 2022. Nat Biotechnol. 40, 1722–1760 (2022).
[3] Golabgir, A., Gutierrez, J.M., Hefzi, H., Li, S., Palsson, B.O., Herwig, C., Lewis, N.E.: Quantitative feature extraction from the Chinese hamster ovary bioprocess bibliome using a novel meta-analysis workflow. Biotechnol Adv. 34, 621–633 (2016).
[4] Lam, A.N., Elvesæter, B., Martin-Recuerda, F.: Evaluation of a representative selection of SPARQL query engines using wikidata. In: The Semantic Web. pp. 679–696. Springer Nature Switzerland, Cham (2023).
[5] Kim, K., Hong, K., Gulcehre, C., Ailamaki, A.: Optimizing LLM inference for database systems: Cost-aware scheduling for concurrent requests, http://arxiv.org/abs/2411.07447, (2024).
[6] Lysenko, A., Roznovăţ, I.A., Saqi, M., Mazein, A., Rawlings, C.J., Auffray, C.: Representing and querying disease networks using graph databases. BioData Min. 9, 23 (2016).
[7] Mazein, I., Rougny, A., Mazein, A., Henkel, R., Gütebier, L., Michaelis, L., Ostaszewski, M., Schneider, R., Satagopam, V., Jensen, L.J., Waltemath, D., Wodke, J.A.H., Balaur, I.: Graph databases in systems biology: a systematic review. Brief. Bioinform. 25, (2024). https://doi.org/10.1093/bib/bbae561.
[8] The Gene Ontology Consortium: The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Res. 47, D330–D338 (2019).
[9] Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P., Steinbeck, C.: ChEBI in 2016: Improved services and an expanding collection of metabolites. Nucleic Acids Res. 44, D1214-9 (2016).
[10] Sarntivijai, S., Lin, Y., Xiang, Z., Meehan, T.F., Diehl, A.D., Vempati, U.D., Schürer, S.C., Pang, C., Malone, J., Parkinson, H., Liu, Y., Takatsuki, T., Saijo, K., Masuya, H., Nakamura, Y., Brush, M.H., Haendel, M.A., Zheng, J., Stoeckert, C.J., Peters, B., Mungall, C.J., Carey, T.E., States, D.J., Athey, B.D., He, Y.: CLO: The cell line ontology. J. Biomed. Semantics. 5, 37 (2014).
[11] Giusto, P.D., Choi, D.-H., Antonakoudis, A., Duraikannan, V.G., Craveur, P., Cowie, N.L., Ganapathy, T., Ramesh, K., Benavidez-López, S., Orellana, C.A., Jiménez, N.E., Dworkin, L.A., Morrissey, J., Marin de Mas, I., Strain, B., Valdez-Cruz, N.A., Trujillo-Roldán, M.A., Marzluf, J., Martínez, V.S., Zehetner, L., Altamirano, C., Vega-Letter, A.M., Priem, B., Cao, H.C., Hold, M., Ma, J., Hong, Y.F., Gopalakrishnan, S., Enuh, B.M., Tarzi, C., Pang, K.T., Angione, C., Zanghellini, J., Kontoravdi, C., Hefzi, H., Betenbaugh, M.J., Nielsen, L.K., Lakshmanan, M., Lee, D.-Y., Richelle, A., Lewis, N.E.: A community-consensus reconstruction of Chinese Hamster metabolism enables structural systems biology analyses to decipher metabolic rewiring in lactate-free CHO cells, http://dx.doi.org/10.1101/2025.04.10.647063, (2025). https://doi.org/10.1101/2025.04.10.647063.

[12] Kulvatunyou, B., Drobnjakovic, M., Ameri, F., Will, C., Smith, B.: The Industrial Ontologies Foundry (IOF) Core Ontology. Boonserm Kulvatunyou, Milos Drobnjakovic, Farhad Ameri, Chris Will, Barry Smith. (2022).
[13] Drobnjakovic, M., Kulvatunyou, B., Frechette, S., Srinivasan, V.: Recent developments in ontology standards and their applicability to biomanufacturing. In: Volume 2: 43rd Computers and Information in Engineering Conference (CIE). p. V002T02A058. American Society of Mechanical Engineers (2023).
[14] Information technology - Top-level ontologies (TLO) - Part 2: Basic Formal Ontology (BFO).
[15] Bernabé, C.H., Queralt-Rosinach, N., Silva Souza, V.E., Bonino da Silva Santos, L.O., Mons, B., Jacobsen, A., Roos, M.: The use of foundational ontologies in biomedical research. J. Biomed. Semantics. 14, 21 (2023).
[16] Courtot, M., Juty, N., Knüpfer, C., Waltemath, D., Zhukova, A., Dräger, A., Dumontier, M., Finney, A., Golebiewski, M., Hastings, J., Hoops, S., Keating, S., Kell, D.B., Kerrien, S., Lawson, J., Lister, A., Lu, J., Machne, R., Mendes, P., Pocock, M., Rodriguez, N., Villeger, A., Wilkinson, D.J., Wimalaratne, S., Laibe, C., Hucka, M., Le Novère, N.: Controlled vocabularies and semantics in systems biology. Mol. Syst. Biol. 7, 543 (2011).
[17] Bandrowski, A., Brinkman, R., Brochhausen, M., Brush, M.H., Bug, B., Chibucos, M.C., Clancy, K., Courtot, M., Derom, D., Dumontier, M., Fan, L., Fostel, J., Fragoso, G., Gibson, F., Gonzalez-Beltran, A., Haendel, M.A., He, Y., Heiskanen, M., Hernandez-Boussard, T., Jensen, M., Lin, Y., Lister, A.L., Lord, P., Malone, J., Manduchi, E., McGee, M., Morrison, N., Overton, J.A., Parkinson, H., Peters, B., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R.H., Schober, D., Smith, B., Soldatova, L.N., Stoeckert, C.J., Jr, Taylor, C.F., Torniai, C., Turner, J.A., Vita, R., Whetzel, P.L., Zheng, J.: The Ontology for Biomedical Investigations. PLoS One. 11, e0154556 (2016).
[18] Gkoutos, G.V., Schofield, P.N., Hoehndorf, R.: The Units Ontology: a tool for integrating units of measurement in science. Database (Oxford). 2012, bas033 (2012).
[19] Kol, S., Ley, D., Wulff, T., Decker, M., Arnsdorf, J., Schoffelen, S., Hansen, A.H., Jensen, T.L., Gutierrez, J.M., Chiang, A.W.T., Masson, H.O., Palsson, B.O., Voldborg, B.G., Pedersen, L.E., Kildegaard, H.F., Lee, G.M., Lewis, N.E.: Multiplex secretome engineering enhances recombinant protein production and purity. Nat Commun. 11, 1908 (2020).
[20] Hefzi, H., Martínez-Monge, I., Marin de Mas, I., Cowie, N.L., Toledo, A.G., Noh, S.M., Karottki, K.J. la C., Decker, M., Arnsdorf, J., Camacho-Zaragoza, J.M., Kol, S., Schoffelen, S., Pristovšek, N., Hansen, A.H., Miguez, A.A., Bjørn, S.P., Brøndum, K.K., Javidi, E.M., Jensen, K.L., Stangl, L., Kreidl, E., Kallehauge, T.B., Ley, D., Ménard, P., Petersen, H.M., Sukhova, Z., Bauer, A., Casanova, E., Barron, N., Malmström, J., Nielsen, L.K., Lee, G.M., Kildegaard, H.F., Voldborg, B.G., Lewis, N.E.: Multiplex genome editing eliminates lactate production without impacting growth rate in mammalian cells. Nat Metab. 7, 212–227 (2025).
[21] Chiang, A.W.T., Li, S., Kellman, B.P., Chattopadhyay, G., Zhang, Y., Kuo, C.-C., Gutierrez, J.M., Ghazi, F., Schmeisser, H., Ménard, P., Bjørn, S.P., Voldborg, B.G., Rosenberg, A.S., Puig, M., Lewis, N.E.: Combating viral contaminants in CHO cells by engineering innate immunity. Sci Rep. 9, 8827 (2019).
[22] Park, S.-Y., Choi, D.-H., Song, J., Lakshmanan, M., Richelle, A., Yoon, S., Kontoravdi, C., Lewis, N.E., Lee, D.-Y.: Driving towards digital biomanufacturing by CHO genome-scale models. Trends Biotechnol. 42, 1192–1203 (2024).
[23] Šormaz, D., Seeharit, S., Kulvatunyou, B., Drobnjaković, M.: A Basic Formal Ontology-based ontological modeling for plan and occurrence, a biomanufacturing process

verification use case. In: Volume 2B: 44th Computers and Information in Engineering Conference (CIE). American Society of Mechanical Engineers (2024). https://doi.org/10.1115/detc2024-143710.

[24] Taylor, C.F., Paton, N.W., Lilley, K.S., Binz, P.-A., Julian, R.K., Jr, Jones, A.R., Zhu, W., Apweiler, R., Aebersold, R., Deutsch, E.W., Dunn, M.J., Heck, A.J.R., Leitner, A., Macht, M., Mann, M., Martens, L., Neubert, T.A., Patterson, S.D., Ping, P., Seymour, S.L., Souda, P., Tsugita, A., Vandekerckhove, J., Vondriska, T.M., Whitelegge, J.P., Wilkins, M.R., Xenarios, I., Yates, J.R., 3rd, Hermjakob, H.: The minimum information about a proteomics experiment (MIAPE). Nat. Biotechnol. 25, 887–893 (2007).