






# Charles Tapley Hoyt, Ph.D.

cthoht.com ·  linkedin.com/in/cthoht ·  github.com/cthoht ·  @cthoht@scholar.social ·  Q47475003  
 https://orcid.org/0000-0003-4423-4370 ·  PjrpzUIAAAAJ

I'm working towards building a research group in the Institute of Inorganic Chemistry at RWTH Aachen University focused on software development, data standardization, FAIRification, integration, and applications of ML/AI in the chemical, biological, and health sciences. Therefore, I'm interested in **building academic collaborations** that can lead to grant opportunities and **securing project contracts** for unmet business needs addressed by my semantic technologies, capabilities, and experience.

## Fields of Work

Bioinformatics, Pathway Analysis, Machine Learning, Natural Language Processing, Ontology, Knowledge Graph, Proteochemometrics, Drug Repositioning, Systems Biology, Cheminformatics

## Employment

- 2025– **Wissenschaftlicher Mitarbeiter**, *RWTH Aachen University*, Aachen, Germany.
- 2023–24 **Senior Scientist**, *Northeastern University*, Boston, MA, USA (remote).  
Moved with the Gyori Lab from Harvard Medical School
- 2021–23 **Research Fellow**, *Harvard Medical School, Laboratory of Systems Pharmacology*, Boston, MA, USA (remote).  
Biocuration, data standardization, natural language processing, data integration, and automated assembly of knowledge and models (Gyori/INDRA Lab)
- 2020 **Computational Biologist**, *Enveda Biosciences*, Boulder, CO, USA (remote).  
Biomedical data integration, knowledge graph construction, and analysis
- 2018–19 **Lecturer**, *University of Bonn*, Bonn, Germany.
- 2016–19 **Research Fellow**, *Fraunhofer SCAI, Dept. of Bioinformatics*, Sankt Augustin, Germany.  
Systems biology modeling and analysis with the Biological Expression Language
- 2012–15 **Teaching Assistant**, *Northeastern University*, Boston, MA, USA.
- 2014 **in Silico Lead Discovery Co-op**, *Novartis*, Cambridge, MA, USA.  
Mechanism of action deconvolution in high-throughput black-box phenotypic screens
- 2013–14 **Molecular Informatics Internship**, *Pfizer*, Cambridge, MA, USA.  
Software development and database development for cancer genomics data
- 2013 **Post-Selection Chemistry Co-op**, *GlaxoSmithKline*, Waltham, MA, USA.  
Lead optimization of antivirals identified by ultra-high-throughput screening
- 2012 **Research Assistant**, *Pollastri Laboratory, Northeastern University*, Boston, MA, USA.  
Medicinal synthetic chemistry for repurposing PDE4 inhibitors against *T. Brucei*

## Education

- 2018–19 **Doctor of Philosophy**, *Computational Life Sciences*, University of Bonn.  
Thesis: *Generation and Applications of Knowledge Graphs in Systems and Networks Biology*  
Research Advisor: Prof. Dr. Martin Hofmann-Apitius  
GPA: 1,3 (Magna Cum Laude)
- 2015–17 **Master of Science**, *Life Science Informatics*, University of Bonn.  
Thesis: *PyBEL: a Computational Framework for Biological Expression Language*  
Research Advisor: Prof. Dr. Martin Hofmann-Apitius  
GPA: 1,6
- 2011–15 **Bachelor of Science**, *Chemistry*, Northeastern University.  
GPA: 3.93/4.0 (Summa Cum Laude)  
Research Advisor: Prof. Dr. Michael P. Pollastri

## Awards

- 2023 **Excellence in Biocuration Early Career Award**, *International Society of Biocuration*.  
2015 **Bernie Lemire Award**, *Department of Chemistry, Northeastern University*.  
2011 **Presidential Scholarship**, *Northeastern University*.

## Research

### Publications

Author lists have been omitted to reduce visual clutter. There are a total of 62 manuscripts (53 peer-reviewed, 8 pre-printed/submitted, and 1 in preparation). The following are noted:

- † signifies first or co-first authorship
- ‡ signifies last or senior authorship

1. †**Assembly and reasoning over semantic mappings at scale for biomedical data integration**. *Bioinformatics*, 2025. doi:10.1093/bioinformatics/btaf542
2. †**Improving reproducibility of cheminformatics workflows with chembl-downloader**. *JOSS*, 2025. doi:10.21105/joss.08844
3. (preprint) †**Causal identification with Y0**. *arXiv*, 2025. doi:10.48550/arxiv.2508.03167
4. (preprint) **The Cell Ontology in the age of single-cell omics**. *arXiv*, 2025. doi:10.48550/arxiv.2506.10037
5. †**Computational tools and data integration to accelerate vaccine development: challenges, opportunities, and future directions**. *Front Immunol*, 2025. doi:10.3389/fimmu.2025.1502484
6. (in preparation) †**Assembly and application of coherent biomedical lexica**. *in preparation for Database*, 2025
7. (submitted) **OmniPath: integrated knowledgebase for multi-omics analysis**. *submitted to Nucleic Acids Research*, 2025. biorxiv:2025.09.11.675512
8. (submitted) **VO: The Vaccine Ontology**. *submitted to Scientific Data*, 2025. biorxiv:2025.08.12.669998
9. (submitted) ‡**More Rigorous Software Engineering Would Improve Reproducibility in Machine Learning Research**. *submitted to TMLR*, 2025. arxiv:2502.00902
10. (in preparation) **Teaching RDM in a smart advanced inorganic lab course and its provision in the DALIA platform**. *in preparation for Nature Computational Science*, 2025
11. **Eliater: A Python package for estimating outcomes of perturbations in biomolecular networks**. *Bioinformatics*, 2024. doi:10.1093/bioinformatics/btae527
12. †**The O3 guidelines: open data, open code, and open infrastructure for sustainable curated scientific resources**. *Scientific Data*, 2024. doi:10.1038/s41597-024-03406-w
13. **An open source knowledge graph ecosystem for the life sciences**. *Scientific Data*, 2024. doi:10.1038/s41597-024-03171-w
14. ‡**ptwt - The PyTorch Wavelet Toolbox**. *JMLR*, 2024
15. **A simple standard for ontological mappings 2023: updates on data model, collaborations and tooling**. *OM 2023*
16. **The Human Phenotype Ontology in 2024: phenotypes around the world**. *NAR*, 2023. doi:10.1093/nar/gkad1005
17. †**Improving reproducibility and reusability in the Journal of Cheminformatics**. *J Cheminform*, 2023. doi:10.1186/s13321-023-00730-y
18. **Democratizing knowledge representation with BioCypher**. *Nature Biotechnology*, 2023. doi:10.1038/s41587-023-01848-y
19. **Optimal adjustment sets for causal query estimation in partially observed biomolecular networks**. *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad270
20. †**Prediction and curation of missing biomedical identifier mappings with Biomappings**. *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad130
21. (preprint) **Experimental design for causal query estimation in partially observed biomolecular networks**.

*arXiv*, 2022. arxiv:2210.13423

22. †**Unifying the identification of biomedical entities with the Bioregistry.** *Scientific Data*, 2022. doi:10.1038/s41597-022-01807-3
23. **A Simple Standard for Ontological Mappings 2022: Updates of data model and outlook.** *OM 2022*
24. **A Review of Biomedical Datasets Relating to Drug Discovery: A Knowledge Graph Perspective.** *Brief Bioinform*, 2022. doi:10.1093/bib/bbac404
25. **ChemicalX: A Deep Learning Library for Drug Pair Scoring.** *KDD*, 2022. doi:10.1145/3534678.3539023
26. **Integrating multi-omics data reveals function and therapeutic potential of deubiquitinating enzymes.** *eLife*, 2022. doi:10.7554/elife.72879
27. **Understanding the Performance of Knowledge Graph Embeddings in Drug Discovery.** *Artificial Intelligence in the Life Sciences*, 2022. doi:10.1016/j.ailsci.2022.100036
28. **A Simple Standard for Sharing Ontological Mappings (SSSOM).** *Database*, 2022. doi:10.1093/database/baac035
29. **Gilda: biomedical entity text normalization with machine-learned disambiguation as a service.** *Bioinformatics Advances*, 2022. doi:10.1093/bioadv/vbac034
30. **ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs.** *SWAT4HCLS 2022*. ceur-ws:3127:13
31. (preprint) **Mondo: Unifying diseases for the world, by the world.** *medRxiv*, 2022. doi:10.1101/2022.04.13.22273750
32. **Do-calculus enables estimation of causal effects in partially observed biomolecular pathways.** *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac251
33. (preprint) †**A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs.** *arXiv*, 2022. arxiv:2203.07544
34. **PyBioPAX: biological pathway exchange in Python.** *JOSS*, 2022. doi:10.21105/joss.04136
35. (preprint) **An Open Challenge for Inductive Link Prediction on Knowledge Graphs.** *arXiv*, 2022. arxiv:2203.01520
36. **STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs.** *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac001
37. **Ontology Development Kit: a toolkit for building, maintaining, and standardising biomedical ontologies.** *Database*, 2022. doi:10.1093/database/baac087
38. **Bringing Light Into the Dark: A Large-scale Evaluation of Knowledge Graph Embedding Models Under a Unified Framework.** *TPAMI*, 2021. doi:10.1109/tpami.2021.3124805
39. **The role of metadata in reproducible computational research.** *Patterns*, 2021. doi:10.1016/j.patter.2021.100322
40. †**CLEP: a hybrid data- and knowledge-driven framework for generating patient representations.** *Bioinformatics*, 2021. doi:10.1093/bioinformatics/btab340
41. **A Systems Biology Approach for Hypothesizing the Effect of Genetic Variants on Neuroimaging Features in Alzheimer's Disease.** *JAD*, 2021. doi:10.3233/jad-201397
42. †**PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings.** *JMLR*, 2021. arxiv:2007.14175
43. **Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis.** *IEEE TBDATA*, 2021. doi:10.1109/tbdata.2021.3050680
44. (preprint) †**Extension of Roles in the ChEBI Ontology.** *ChemRxiv*, 2020. doi:10.26434/chemrxiv.12591221
45. **The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST).** *Bioinformatics*, 2020. doi:10.1093/bioinformatics/btaa622
46. **GuiltyTargets: Prioritization of Novel Therapeutic Targets with Deep Network Representation Learning.** *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2020. doi:10.1109/tcbb.2020.3003830
47. **PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures.** *BMC Bioinf.*, 2020. doi:10.1186/s12859-020-03568-5

48. **Identifying the parametric occurrence of multiple steady states for some biological networks.** *Journal of Symbolic Computation*, 2020. doi:10.1016/j.jsc.2019.07.008
49. **A Computational Approach for Mapping Heme Biology in the Context of Hemolytic Disorders.** *Frontiers in Bioengineering and Biotechnology*, 2020. doi:10.3389/fbioe.2020.00074
50. **The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling.** *Frontiers in Genetics*, 2019. doi:10.3389/fgene.2019.01203
51. **Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm.** *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-3101-1
52. **The KEEN Universe - An Ecosystem for Knowledge Graph Embeddings with a Focus on Reproducibility and Transferability.** *ISWC 2019*. doi:10.1007/978-3-030-30796-7\_1
53. **Predicting Missing Links Using PyKEEN.** *ISWC 2019*. ceur-ws:2456:64
54. **RatVec: A General Approach for Low-dimensional Distributed Vector Representations via Domain-specific Rational Kernels.** *LWDA 2019*
55. **BioKEEN: a library for learning and evaluating biological knowledge graph embeddings.** *Bioinformatics*, 2019. doi:10.1093/bioinformatics/btz117
56. **PathMe: merging and exploring mechanistic pathway knowledge.** *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-2863-9
57. (preprint) **†Integration of Structured Biological Data Sources using Biological Expression Language.** *bioRxiv*, 2019. doi:10.1101/631812 bioRxiv:631812
58. **ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases.** *NPJ Syst Biol Appl.*, 2019. doi:10.1038/s41540-018-0078-8
59. **†Re-curation and rational enrichment of knowledge graphs in Biological Expression Language.** *Database*, 2019. doi:10.1093/database/baz068
60. **Challenges of Integrative Disease Modeling in Alzheimer's Disease.** *Frontiers in molecular biosciences*, 2019. doi:10.3389/fmolb.2019.00158
61. **†BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language.** *Database*, 2018. doi:10.1093/database/bay126
62. **BEL2ABM: agent-based simulation of static models in Biological Expression Language.** *Bioinformatics*, 2018. doi:10.1093/bioinformatics/bty107
63. **†PyBEL: a Computational Framework for Biological Expression Language.** *Bioinformatics*, 2018. doi:10.1093/bioinformatics/btx660
64. **†A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities.** *Database*, 2018. doi:10.1093/database/bay050
65. **A Case Study on the Parametric Occurrence of Multiple Steady States.** *ISSAC 2017*. doi:10.1145/3087604.3087622
66. **Repurposing human PDE4 inhibitors for neglected tropical diseases: design, synthesis and evaluation of cilomilast analogues as Trypanosoma brucei PDEB1 inhibitors.** *Bioorg Med Chem Lett*, 2014. doi:10.1016/j.bmcl.2014.07.063

## Invited Presentations (Selected)

1. **Modern software development practice with Python.** *May Institute* (May 7, 2024)
2. **Assembly and inference over semantic mappings to support the NFDI Terminology Service.** *TS4NFDI Community Workshop* (April 9, 2024)
3. **Assembly of Domain Knowledge at Scale in Biomedicine and Beyond.** *Harvard Medical School - Laboratory of Systems Pharmacology Meeting* (January 19, 2024)
4. **Machine-assisted integration of data and knowledge at scale to support biomedical discovery.** *NIH BISTI Seminar* (December 7, 2023)
5. **Democratizing Biocuration, or, How I Learned to Love the Drive-by Curation.** *International Society of Biocuration Annual General Meeting* (October 18, 2023)

6. **Improving ontology interoperability with Biomappings.** *OBO Academy - Monarch Training Series* (September 19, 2023)
7. **Modern prefix management with the Bioregistry and 'curies'.** *OBO Academy - Monarch Training Series* (September 5, 2023)
8. **Axiomatizing Chemical Roles.** *Ontologies4Chem Workshop 2022* (September 7, 2022)
9. **Modern Scientific Software Development Practice in Python.** *May Institute* (May 20, 2022)
10. **Knowledge Graph Embedding with PyKEEN in 2022.** *Knowledge Graph Conference (KGC 2022)* (May 5, 2022)
11. **The Biopragnatics Stack: Biomedical and Chemical Semantics for Humans.** *Machine-Actionable Data Interoperability for Chemical Sciences (MADICES)* (February 8, 2022)
12. **Reusable Science in Python.** *May Institute* (July 28, 2021)
13. **Current Issues in Theory, Reproducibility, and Utility of Graph Machine Learning in the Life Sciences.** *Graph Machine Learning in Industry* (September 23, 2021)
14. **Perspectives on Knowledge Graph Embedding Models in/out of Biomedicine.** *AstraZeneca* (April 6, 2021)
15. **Introduction to the Biological Expression Language and the Rational Enrichment Workflow.** *CoronaWhy* (May 6, 2020)
16. **Applications of Knowledge Graphs in Drug Discovery.** *Computational Drug Discovery Group, University of Leiden* (November 5, 2019)
17. **Generation and Application of Biomedical Knowledge Graphs.** *Harvard Medical School* (July 19, 2019)













## Talks and Posters (Selected)

1. **Bioregistry Workshop at Biocuration 2025.** *Bioregistry Workshop @ Biocuration 2025* (April 6, 2025)
2. **Bioregistry and the NFDI in 2024.** *3rd Ontologies4Chem Workshop* (December 9, 2024)
3. **Assembly and Reasoning over Semantic Mappings at Scale.** *Biocuration 2024* (March 7, 2024)
4. **Introduction to WPCI 2023.** *Winter 2023 Workshop on Prefixes, CURIEs, and IRIs* (November 27, 2023)
5. **Standardization of chemical prefixes, CURIEs, URIs, and semantic mappings.** *Ontologies4Chem Workshop 2023* (October 12, 2023)
6. **Improving the reproducibility of cheminformatics workflows with chembl-downloader.** *RDKit User Group Meeting 2023* (September 21, 2023)
7. **Promoting the longevity of curated scientific resources through open code, open data, and public infrastructure.** *Biocuration 2023* (April 26, 2023)
8. **Using dashboards to monitor ontology standardisation and community activity.** *Ontology Summit 2023* (February 15, 2023)
9. **Introduction to WPCI 2022.** *2022 Workshop on Prefixes, CURIEs, and IRIs* (December 5, 2022)
10. **The Bioregistry, CURIEs, and OBO Community Health.** *International Conference on Biomedical Ontology (ICBO)* (September 26, 2022)
11. **Closing the Semantic Gap: Identifying Missing Mappings and Merging Equivalent Concepts to Support Knowledge Graph Assembly.** *Harvard Medical School - Sorger Lab Meeting* (August 1, 2022)
12. **A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs.** *Graph Learning Benchmarks (GLB 2022)* (April 26, 2022)
13. **Introduction to WPCI 2021.** *2021 Workshop on Prefixes, CURIEs, and IRIs* (October 29, 2021)
14. **Biomappings: Community Curation of Mappings between Biomedical Entities.** *4th Session of the International Society of Biocuration 2021 Virtual Conference* (October 5, 2021; poster)
15. **The Bioregistry: A Metaregistry for Biomedical Entities.** *12th International Conference on Biomedical Ontologies* (September 17, 2021)











16. **Future Directions for WikiPathway Meta-curation.** *WikiPathways Developers Conference Call* (January 6, 2021)
17. **The Biological Expression Language and PyBEL in 2020.** *COVID-19 Disease Map Community Meeting* (July 10, 2020)
18. **Maintenance and Enrichment of Disease Maps in Biological Expression Language.** *4th Disease Maps Community Meeting* (October 4, 2019; poster)
19. **Identifying Drug Repositioning Candidates using Representation Learning on Heterogeneous Networks.** *The Eighth Joint Sheffield Conference on Chemoinformatics* (June 19, 2019; poster)
20. **The PyBEL Ecosystem in 2018.** *OpenBEL Community Meeting* (May 14, 2018)
21. **From Knowledge Assembly to Hypothesis Generation.** *Bio-IT World* (April 22, 2018)
22. **Knowledge Assembly in Systems and Networks Biology.** *Bio-IT World* (April 23, 2018; poster)
23. **The Human Brain Pharmacome: An Overview.** *3rd European Conference on Translational Bioinformatics* (April 17, 2018; poster)
24. **Gene Set Analysis using Phenotypic Screening Data.** *Research, Innovation and Scholarship Expo 2015* (April 9, 2015; poster)

## Research Software (Selected)










Achieving reproducibility and reusability in the computational sciences heavily depends on code that is developed, tested, documented, and distributed according to best software development practices. I aim to make as much of my research as possible available and reusable in the form of open source software. Accordingly, I have pushed tens of thousands of commits to GitHub across my own and external projects to which I have contributed. I have also written extensively on teaching best practices to scientists on my blog.

- |  |   |
|--|---|
| 1. <b>PyBEL</b><br>A compiler for the Biological Expression Language (BEL)   |  <a href="#">pybel/pybel</a>                 |
| 2. <b>Bio2BEL</b><br>A framework for reproducible data integration in BEL and knowledge graph construction           |  <a href="#">bio2bel/bio2bel</a>             |
| 3. <b>BEL Commons</b><br>A web application for the interactive exploration of networks encoded in BEL                |  <a href="#">bel-commons/bel-commons</a>     |
| 4. <b>PyKEEN</b><br>The most expansive knowledge graph embedding framework to date                                   |  <a href="#">pykeen/pykeen</a>               |
| 5. <b>SeMRA</b><br>Automated assembly and inference of semantic mappings   |  <a href="#">biopragnatics/semra</a>         |
| 6. <b>Bioontologies</b><br>Access and processing of ontologies on top of ROBOT and OBO Graphs                        |  <a href="#">biopragnatics/bioontologies</a> |
| 7. <b>PyOBO</b><br>Harmonization of biological ontologies and controlled vocabularies                                |  <a href="#">pyobo/pyobo</a>                 |
| 8. <b>curies</b><br>Idiomatic conversion between URIs and compact URIs (CURIEs)                                      |  <a href="#">biopragnatics/curies</a>        |
| 9. <b>GuiltyTargets</b><br>Target prioritization framework using gene expression and network representation learning |  <a href="#">guiltytargets/guiltytargets</a> |
| 10. <b>RatVec</b><br>Sequence-based representation learning  |  <a href="#">ratvec/ratvec</a>               |
| 11. <b>PS4DR</b><br>Drug repositioning based on bioactivity pattern matching and GWAS                                |  <a href="#">ps4dr/ps4dr</a>                 |
| 12. <b>SeffNet</b>   |  <a href="#">seffnet/seffnet</a>             |

Drug repositioning framework based on network representation learning

13. **CLEP**  [hybrid-kg/clep](#)  
Patient stratification framework based on network representation learning
14. **BEL2SCM**  [bel2scm/bel2scm](#)  
Generation of structural causal models (SCMs) from BEL
15. **y0 Causal Inference Engine**  [y0-causal-inference/y0](#)  
Representation and manipulating probabilistic expressions
16. **STonKGs**  [stonkgs/stonkgs](#)  
Multimodal Transformers for biomedical text and Knowledge Graph data
17. **ChemicalX**  [AstraZeneca/chemicalx](#)  
A deep learning library for drug-drug interaction, polypharmacy side effect, and synergy prediction
18. **PathMe**  [PathwayMerger/PathMe](#)  
Web application for exploration of pathway databases
19. **INDRA**  [sorgerlab/indra](#)  
Automated knowledge assembly and modeling in biomedicine
20. **INDRA CoGEx**  [bgyori/indra\\_cogex](#)  
A 10<sup>8</sup> relation-scale knowledge graph extending on causal knowledge from INDRA
21. **MIRA**  [indralab/mira](#)  
Machine-assisted scientific modeling using meta-model templates and domain knowledge graphs
22. **Gilda**  [indralab/gilda](#)  
Biomedical named entity recognition and grounding using machine-learned disambiguation

## Databases (Selected)

1. **Bioregistry**  [bioregistry/bioregistry](#) , <https://bioregistry.io>  
An integrative meta-registry of biological databases, ontologies, and nomenclatures
2. **Biomappings**  [biopragnatics/biomappings](#) , <https://biopragnatics.github.io/biomappings>  
Predicted and curated mappings between named biological entities
3. **Biolookup**  [biopragnatics/biolookup](#) , <http://biolookup.io>  
Comprehensive database of identifiers, names, synonyms, cross-references, properties, and relations for biomedical entities
4. **Bioversions**  [biopragnatics/bioversions](#)  
Automated tracking of the current version for each biological database?
5. **OBO Database Ingest**  [biopragnatics/obo-db-ingest](#)  
Conversion of biomedical databases into ontologies
6. **Chemical Roles Graph**  [chemical-roles/chemical-roles](#)  
Connecting roles in the ChEBI ontology to their targets
7. **CONSO**  [pharmacome/conso](#)  
Ontology of phenomena related to neurodegeneration
8. **CONIB**  [pharmacome/conib](#)  
Curated knowledge graphs describing neurodegeneration in BEL
9. **ComPath**  [ComPath/ComPath](#) , <https://compath.scai.fraunhofer.de/>  
A data integration pipeline, manually curated dataset of mappings between pathway databases, and web application for integrative analysis of pathways

## External Contributions

Manually and semi-automatically curated resources benefit greatly from open governance models that enable external contribution. Instead of creating siloed or single use improvements to resources I consume, I aim to make as many upstream contributions as possible. Accordingly, I have made hundreds of small contributions across 50+ ontologies in the OBO Foundry and other resources including:



EHDA2, SSSOM, MIAPA, DDPHENO, BFO, ECO, OMO, GO, FYPO, OAE, HSAPDV, COB, RO, MMUSDV, MP, AGRO, CIDO, EUPATH, GEO, TTO, OGMS, BSPO, OLATDV, FBBT, PECO, XCO, UPHENO, CL, CDAO, AISM, HAO, HANCESTRO, ONS, PCL, SWO, OBA, PDUMDV, HTN, MOD, PO, OHD, OBI, LEPAO, EXO, MONDO, UBERON, ENVO, SYMP, IAO, CHMO, TO, TAXRANK, WBBT, WBL, HP, MAXO, COLAO

## Funding

1. **Advancing Data Integration and Discovery in Biomedicine with the Bioregistry**  
Chan Zuckerberg Initiative (CZI), (2023-329850), \$250K overall, 2023-2025 (24 months), Grant: 2023-329850  
PI: Benjamin M. Gyori  
Role: Key person. Co-wrote proposal, responsible for design and implementation of research plan.
2. (past) **Rapid Assessment of Platform Technologies to Expedite Response**  
Defense Threat Reduction Agency (DTRA), (FP00012844), \$623K (subcontract) / \$15M overall, 2022-2024 (18 months), Grant: HDTRA1242031  
PI: Benjamin M. Gyori  
Role: Participant. Co-wrote proposal with PNNL collaborators. Responsible for design and implementation of research plan. Lead developer of semantic and analytical technologies. Note: left position before project completion.
3. (past) **MIRA: Modeling with an Intelligent Reasoning Assistant**  
DARPA ASKEM Program, (HR00112220036), \$2.1M, 2022-2026 (42 months)  
PI: Benjamin M. Gyori  
Role: Participant. Note: left position before project completion.
4. (past) **Young Faculty Award / DARPA Director's Fellowship Award**  
DARPA, (W911NF2010255), \$750K, 2020-2023 (36 months)  
PI: Benjamin M. Gyori  
Role: Participant
5. (past) **Ecosystem of Machine-maintained Models with Automated Assembly**  
DARPA Automating Scientific Knowledge Extraction (ASKE) / Artificial Intelligence Exploration (AIE) program, (HR00111990009), \$2M, 2018-2020 (18 months + extension)  
PI: Benjamin M. Gyori  
Role: Participant. Note: Joined in 2021
6. (past) **Interdisciplinary Research Award**  
Fraunhofer Center for Machine Learning, (Internal), \$40K 2019 (3 months)  
PI: Martin Hofmann-Apitius  
Role: Key person. Co-wrote proposal. Responsible for design and implementation of research plan including applications of PyKEEN in biomedicine.
7. (past) **Fraunhofer Service Contract**  
Philip Morris International, (Contract), €80K 2019-2020 (4 months)  
PI: Martin Hofmann-Apitius  
Role: Key Person. Co-responsible for design and implementation of visual analytics software for the Biological Expression Language.
8. (past) **Human Brain Pharmacome**  
Fraunhofer, under the MAVO Scheme, 2018-2021 (4 years)  
PI: Martin Hofmann-Apitius  
Role: Participant. Lead project on Fraunhofer SCAI team. Note: left position before project completion.
9. (past) **PYBEL2NDEX**  
University of California, San Diego, (Contract), \$28K 2018 (6 months)  
PI: Martin Hofmann-Apitius  
Role: Key person. Wrote proposal. Responsible for design and implementation of research plan for implementing interoperability between PyBEL and NDEx.
10. (past) **AETIONOMY**



European Union (EU), European Federation of Pharmaceutical Industries and Associations (EFPIA), and Innovative Health Initiative (IMI), (115568), €17.8M 2014-2018 (4 years)

PI: Martin Hofmann-Apitius

Role: Participant. Note: Joined in 2016.

# Community

## Professional Affiliations

- International Society of Biocuration (2021 -)
- CoronaWhy (2020 - 2021)
- OpenBEL Consortium (2017 -)
- Erasmus Student Network Bonn (2016 -)
- American Chemical Society (2011 - 2022)

## Service to the Community

### Scholarly Journals

1. Scientific article reviewer (*ad hoc*) in:
  - Bioinformatics
  - Database
  - BMC Bioinformatics
  - Journal of Cheminformatics
  - Journal of Biomedical Semantics
  - eLife
  - MATCH Communications in Mathematical and in Computer Chemistry
  - PeerJ
2. Guest Reproducibility Editor, Journal of Cheminformatics (2023-)
3. Grant reviewer for:
  - CZI Essential Open Source Software (EOSS) Cycle 6 (2023)

### Conference Organizing Committees

1. ICBO 2024 (Program Committee)
2. 2023 Workshop on Prefixes, CURIEs, and IRIs (Organizer)
3. Biocuration 2024 (Organizing Committee)
4. Biocuration 2023 (Co-chair)
5. 2022 Workshop on Prefixes, CURIEs, and IRIs (Organizer)
6. ICBO 2022 Workshop on Ontology Tools (Co-organizer)
7. ICBO 2022 (Program Committee)
8. ISMB 2022 (Bio-Ontologies/BOSC joint session) (Program Committee)
9. Biocuration 2022 (Organizing Committee)
10. 2021 Workshop on Prefixes, CURIEs, and IRIs (Organizer)

### Other Service

1. International Society of Biocuration (Executive Committee)

# Teaching

## Courses Taught

I have given 11 lectures, 1 seminars, and 7 practicals / labs as either the primary or a guest instructor.

### Northeastern University



1. Modern software development practice with Python (Practical; Instructor; Summer 2024)
2. Modern Scientific Software Development Practice in Python (Practical; Instructor; Summer 2022)
3. Reusable Science in Python (Practical; Instructor; Summer 2021)
4. Drug Discovery and Development (Lecture; Teaching Assistant; Summer II 2015)
5. Organic Chemistry II for Majors (Lecture; Teaching Assistant; Spring 2015)
6. Organic Chemistry I for Majors (Lecture; Teaching Assistant; Fall 2014)
7. Organic Chemistry II for Majors (Lecture; Teaching Assistant; Spring 2014)
8. Organic Chemistry I for Majors (Lecture; Teaching Assistant; Fall 2013)

### University of Bonn


1. Mechanism Enrichment Using NeuroMMSig (Practical; Guest Lecturer; Winter 2020-2021)
2. Mechanism Enrichment Using NeuroMMSig (Practical; Guest Lecturer; Winter 2019-2020)
3. Mathematics Meets Life Sciences (Lecture; Guest Lecturer; Winter 2019-2020)
4. Enzyme Technology Internship (Practical; Instructor; Summer 2019)
5. Life Sciences Knowledge Discovery (Lecture; Guest Lecturer; Summer 2019)
6. Knowledge Assembly, Data Integration, and Modeling in Systems and Networks Biology (Seminar; Instructor; Winter 2018-2019)
7. Biological Databases (Lecture; Guest Lecturer; Winter 2018-2019)
8. Life Sciences Knowledge Discovery (Lecture; Guest Lecturer; Summer 2018)
9. Biological Databases (Lecture; Guest Lecturer; Winter 2017-2018)
10. Life Sciences Knowledge Discovery (Lecture; Guest Lecturer; Summer 2017)
11. Biomedical Database Lab (Practical; Teaching Assistant; Winter 2016-2017)

## Supervision

### Northeastern University

- Aria Agarwal  Undergraduate Research Assistant (January - May 2024)
- Sara Taheri  Doctoral Student (July 2023 - April 2024)





### CoronaWhy

- Aman Choudhri  Student Research Assistant (June - October 2020)

### Fraunhofer

- Lauren Nicole DeLong  Student Research Assistant (September - December 2019)
- Vinay Bharadhwaj  Student Research Assistant (July - December 2019)
- Yojana Gadiya  Student Research Assistant (April - May 2019)
- Trusha Adeshara  Student Research Assistant (April - May 2019)
- Rana Aldisi  Student Research Assistant (July 2018 - March 2019)
- Lingling Xu  Student Research Assistant (July 2018 - March 2019)
- Kristian Kolpeja  Student Research Assistant (July - November 2018)
- Esther Wollert  Student Research Assistant (July 2018 - August 2019)
- Sandra Spalek  Student Research Assistant (July 2018 - August 2019)
- Keerthika Lohanadan Student Research Assistant (July - September 2018)
- Colin Birkenbihl  Student Research Assistant (July - October 2017)
- Aram Grigoryan Student Research Assistant (July - December 2017)

### University of Bonn

- Mauricio Pio de Lacerda  Master's Student (March - December 2019)
- Rana Aldisi  Master's Student (March - December 2019)
- Lingling Xu  Master's Student (March - December 2019)
- Özlem Muslu  Master's Student (May - December 2018)