Charles Tapley Hoyt, Ph.D.

cthoyt.com · linkedin.com/in/cthoyt · github.com/cthoyt · Q47475003 · https://orcid.org/0000-0003-4423-4370 · PjrpzUIAAAAJ

Fields of Work

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

Education

2018–19 **Doctor of Philosophy**, *Computational Life Sciences*, University of Bonn.
Thesis: *Generation and Applications of Knowledge Graphs in Systems and Networks Biology*
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*
Advisor: Prof. Dr. Martin Hofmann-Apitius
GPA: 1,6

2011–15 **Bachelor of Science**, *Chemistry*, Northeastern University.
GPA: 3.95/4.0 (Summa Cum Laude)

Employment

2021- **Research Fellow**, *Harvard Medical School*, Bonn, Germany.

2020 **Computational Biologist**, *Enveda Biosciences*, Bonn, Germany.

2018-19 **Lecturer**, *University of Bonn*, Bonn, Germany.

2016-19 **Research Fellow**, *Fraunhofer Institute for Algorithms and Scientific Computing*, Sankt Augustin, Germany.

2014 **Intern**, *Novartis*, Cambridge, United States of America.

2013-14 **Intern**, *Pfizer*, Cambridge, United States of America.

2013 **Intern**, *GlaxoSmithKline*, Waltham, United States of America.

2012-15 **Teaching Assistant**, *Northeastern University*, Boston, United States of America.

Awards

2023 **Nominated, 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

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

1. †**Prediction and curation of missing biomedical identifier mappings with Biomappings**. *Bioinformatics*, 2023. doi:10.1093/bioinformatics/btad130
2. (preprint) **Democratising Knowledge Representation with BioCypher**. *arXiv*, 2022. doi:10.48550/arxiv.2212.13543
3. (preprint) **Experimental design for causal query estimation in partially observed biomolecular networks**. *arXiv*, 2022. arxiv:2210.13423
4. †**Unifying the identification of biomedical entities with the Bioregistry**. *Scientific Data*, 2022. doi:10.1038/s41597-022-01807-3

5. **A Simple Standard for Ontological Mappings 2022: Updates of data model and outlook**. *OM 2022*
6. **A review of biomedical datasets relating to drug discovery: a knowledge graph perspective**. *Brief Bioinform*, 2022. doi:10.1093/bib/bbac404
7. **ChemicalX: A Deep Learning Library for Drug Pair Scoring**. *KDD*, 2022. doi:10.1145/3534678.3539023
8. **Integrating multi-omics data reveals function and therapeutic potential of deubiquitinating enzymes**. *eLife*, 2022. doi:10.7554/elife.72879
9. **Understanding the Performance of Knowledge Graph Embeddings in Drug Discovery**. *Artificial Intelligence in the Life Sciences*, 2022. doi:10.1016/j.ailsci.2022.100036
10. **A Simple Standard for Sharing Ontological Mappings (SSSOM)**. *Database*, 2022. doi:10.1093/database/baac035
11. **Gilda: biomedical entity text normalization with machine-learned disambiguation as a service**. *Bioinformatics Advances*, 2022. doi:10.1093/bioadv/vbac034
12. **ProtSTonKGs: A Sophisticated Transformer Trained on Protein Sequences, Text, and Knowledge Graphs**. *SWAT4HCLS 2022*. ceur-ws:3127:13
13. (preprint) **Mondo: Unifying diseases for the world, by the world**. *medRxiv*, 2022. doi:10.1101/2022.04.13.22273750
14. **Do-calculus enables estimation of causal effects in partially observed biomolecular pathways**. *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac251
15. (preprint) †**A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs**. *arXiv*, 2022. arxiv:2203.07544
16. **PyBioPAX: biological pathway exchange in Python**. *JOSS*, 2022. doi:10.21105/joss.04136
17. (preprint) **An Open Challenge for Inductive Link Prediction on Knowledge Graphs**. *arXiv*, 2022. arxiv:2203.01520
18. **STonKGs: A Sophisticated Transformer Trained on Biomedical Text and Knowledge Graphs**. *Bioinformatics*, 2022. doi:10.1093/bioinformatics/btac001
19. **Ontology Development Kit: a toolkit for building, maintaining and standardizing biomedical ontologies**. *Database*, 2022. doi:10.1093/database/baac087
20. **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
21. **The role of metadata in reproducible computational research**. *Patterns*, 2021. doi:10.1016/j.patter.2021.100322
22. ‡**CLEP: a hybrid data- and knowledge-driven framework for generating patient representations**. *Bioinformatics*, 2021. doi:10.1093/bioinformatics/btab340
23. **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
24. †**PyKEEN 1.0: A Python Library for Training and Evaluating Knowledge Graph Embeddings**. *JMLR*, 2021. arxiv:2007.14175
25. **Leveraging Structured Biological Knowledge for Counterfactual Inference: A Case Study of Viral Pathogenesis**. *IEEE TBDATA*, 2021. doi:10.1109/tbdata.2021.3050680
26. (preprint) †**Extension of Roles in the ChEBI Ontology**. *ChemRxiv*, 2020. doi:10.26434/chemrxiv.12591221
27. **The Minimum Information about a Molecular Interaction Causal Statement (MI2CAST)**. *Bioinformatics*, 2020. doi:10.1093/bioinformatics/btaa622
28. **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
29. **PS4DR: a multimodal workflow for identification and prioritization of drugs based on pathway signatures**. *BMC Bioinf.*, 2020. doi:10.1186/s12859-020-03568-5
30. **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
31. **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

32. **The Impact of Pathway Database Choice on Statistical Enrichment Analysis and Predictive Modeling**. *Frontiers in Genetics*, 2019. doi:10.3389/fgene.2019.01203
33. **Quantifying mechanisms in neurodegenerative diseases (NDDs) using candidate mechanism perturbation amplitude (CMPA) algorithm**. *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-3101-1
34. **The KEEN Universe**. *ISWC 2019*. doi:10.1007/978-3-030-30796-7$_1$
35. **Predicting Missing Links Using PyKEEN**. *ISWC 2019*. ceur-ws:2456:64
36. **RatVec: A General Approach for Low-dimensional Distributed Vector Representations via Domain-specific Rational Kernels**. *LWDA 2019*
37. **BioKEEN: a library for learning and evaluating biological knowledge graph embeddings**. *Bioinformatics*, 2019. doi:10.1093/bioinformatics/btz117
38. **PathMe: merging and exploring mechanistic pathway knowledge**. *BMC Bioinf.*, 2019. doi:10.1186/s12859-019-2863-9
39. (preprint) †**Integration of Structured Biological Data Sources using Biological Expression Language**. *bioRxiv*, 2019. doi:10.1101/631812 biorxiv:631812
40. **ComPath: an ecosystem for exploring, analyzing, and curating mappings across pathway databases**. *NPJ Syst Biol Appl.*, 2019. doi:10.1038/s41540-018-0078-8
41. †**Re-curation and rational enrichment of knowledge graphs in Biological Expression Language**. *Database*, 2019. doi:10.1093/database/baz068
42. **Challenges of Integrative Disease Modeling in Alzheimer's Disease**. *Frontiers in molecular biosciences*, 2019. doi:10.3389/fmolb.2019.00158
43. †**BEL Commons: an environment for exploration and analysis of networks encoded in Biological Expression Language**. *Database*, 2018. doi:10.1093/database/bay126
44. **BEL2ABM: agent-based simulation of static models in Biological Expression Language**. *Bioinformatics*, 2018. doi:10.1093/bioinformatics/bty107
45. †**PyBEL: a Computational Framework for Biological Expression Language**. *Bioinformatics*, 2018. doi:10.1093/bioinformatics/btx66
46. †**A systematic approach for identifying shared mechanisms in epilepsy and its comorbidities**. *Database*, 2018. doi:10.1093/database/bay050
47. **A Case Study on the Parametric Occurrence of Multiple Steady States**. *ISSAC 2017*. doi:10.1145/3087604.3087622
48. **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

1. **Using dashboards to monitor ontology standardisation and community activity**. *Ontology Summit 2023* (February 15, 2023)
2. **Introduction to WPCI 2022**. *2022 Workshop on Prefixes, CURIEs, and IRIs* (December 5, 2022)
3. **The Bioregistry, CURIEs, and OBO Community Health**. *International Conference on Biomedical Ontology (ICBO)* (September 26, 2022)
4. **Axiomatizing Chemical Roles**. *Ontologies4Chem Workshop 2022* (September 7, 2022)
5. **Knowledge Graph Embedding with PyKEEN in 2022**. *Knowledge Graph Conference (KGC 2022)* (May 5, 2022)
6. **The Biopragmatics Stack: Biomedical and Chemical Semantics for Humans**. *Machine-Actionable Data Interoperability for Chemical Sciences (MADICES)* (February 8, 2022)
7. **Introduction to WPCI 2021**. *2021 Workshop on Prefixes, CURIEs, and IRIs* (October 29, 2021)
8. **Current Issues in Theory, Reproducibility, and Utility of Graph Machine Learning in the Life Sciences**. *Graph Machine Learning in Industry* (September 23, 2021)
9. **Perspectives on Knowledge Graph Embedding Models in/out of Biomedicine**. *AstraZeneca* (April 6, 2021)

10. **Future Directions for WikiPathway Meta-curation**. *WikiPathways Developers Conference Call* (January 6, 2021)
11. **The Biological Expression Language and PyBEL in 2020**. *COVID-19 Disease Map Community Meeting* (July 10, 2020)
12. **Introduction to the Biological Expression Language and the Rational Enrichment Workflow**. *CoronaWhy* (May 6, 2020)
13. **Applications of Knowledge Graphs in Drug Discovery**. *Computational Drug Discovery Group, University of Leiden* (November 5, 2019)
14. **Generation and Application of Biomedical Knowledge Graphs**. *Harvard Medical School* (July 19, 2019)
15. **The PyBEL Ecosystem in 2018**. *OpenBEL Community Meeting* (May 14, 2018)

Talks and Posters

1. **Promoting the longevity of curated scientific resources through open code, open data, and public infrastructure**. *Biocuration 2023* (April 26, 2023)
2. **A Unified Framework for Rank-based Evaluation Metrics for Link Prediction in Knowledge Graphs**. *Graph Learning Benchmarks (GLB 2022)* (April 26, 2022)
3. **Biomappings: Community Curation of Mappings between Biomedical Entities**. *4th Session of the International Society of Biocuration 2021 Virtual Conference* (October 5, 2021; poster)
4. **The Bioregistry: A Metaregistry for Biomedical Entities**. *12th International Conference on Biomedical Ontologies* (September 17, 2021)
5. **Maintenance and Enrichment of Disease Maps in Biological Expression Language**. *4th Disease Maps Community Meeting* (October 4, 2019; poster)
6. **Identifying Drug Repositioning Candidates using Representation Learning on Heterogeneous Networks**. *The Eighth Joint Sheffield Conference on Chemoinformatics* (June 19, 2019; poster)
7. **From Knowledge Assembly to Hypothesis Generation**. *Bio-IT World* (April 22, 2018)
8. **Knowledge Assembly in Systems and Networks Biology**. *Bio-IT World* (April 23, 2018; poster)
9. **The Human Brain Pharmacome: An Overview**. *3rd European Conference on Translational Bioinformatics* (April 17, 2018; poster)
10. **Gene Set Analysis using Phenotypic Screening Data**. *Research, Innovation and Scholarship Expo 2015* (April 9, 2015; poster)

Research Software

1. **PyBEL** — pybel/pybel
   A compiler for the Biological Expression Language (BEL)
2. **Bio2BEL** — bio2bel/bio2bel
   A framework for reproducible data integration in BEL and knowledge graph construction
3. **BEL Commons** — bel-commons/bel-commons
   A web application for the interactive exploration of networks encoded in BEL
4. **PyKEEN** — pykeen/pykeen
   The most expansive knowledge graph embedding framework to date
5. **SeMRA** — biopragmatics/semra
   Automated assembly and inference of semantic mappings
6. **Bioontologies** — biopragmatics/bioontologies
   Access and processing of ontologies on top of ROBOT and OBO Graphs
7. **PyOBO** — pyobo/pyobo
   Harmonization of biological ontologies and controlled vocabularies

8. **curies** cthoyt/curies
   Idiomatic conversion between URIs and compact URIs (CURIEs)
9. **GuiltyTargets** guiltytargets/guiltytargets
   Target prioritization framework using gene expression and network representation learning
10. **RatVec** ratvec/ratvec
    Sequence-based representation learning
11. **PS4DR** ps4dr/ps4dr
    Drug repositioning based on bioactivity pattern matching and GWAS
12. **SeffNet** seffnet/seffnet
    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. **INDRA** sorgerlab/indra
    Automated knowledge assembly and modeling in biomedicine
19. **INDRA CoGEx** bgyori/indra_cogex
    A 10^8 relation-scale knowledge graph extending on causal knowledge from INDRA
20. **MIRA** indralab/mira
    Machine-assisted scientific modeling using meta-model templates and domain knowledge graphs
21. **Gilda** indralab/gilda
    Biomedical named entity recognition and grounding using machine-learned disambiguation

Databases

1. **Bioregistry** bioregistry/bioregistry , `https://bioregistry.io`
   An integrative meta-registry of biological databases, ontologies, and nomenclatures
2. **Biomappings** biopragmatics/biomappings , `https://biopragmatics.github.io/biomappings`
   Predicted and curated mappings between named biological entities
3. **Biolookup** biopragmatics/biolookup , `http://biolookup.io`
   Comprehensive database of identifiers, names, synonyms, cross-references, properties, and relations for biomedical entities
4. **Bioversions** biopragmatics/bioversions
   Automated tracking of the current version for each biological database?
5. **OBO Database Ingest** biopragmatics/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

External Contributions

EHDAA2, BFO, GO, ECO, OMO, FYPO, COB, RO, HSAPDV, MMUSDV, MP, AGRO, CIDO, EUPATH, HANCESTRO, OBA, PDUMDV, GEO, HTN, TTO, MOD, CL, CDAO, AISM, HAO, ONS, PCL, SWO, TO, TAXRANK, WBBT, WBLS, OLATDV, PECO, UPHENO, XCO, IAO, PO, OGMS, OHD, OBI, MIAPA, MONDO, LEPAO, EXO, BSPO, DDPHENO, UBERON, FBBT, ENVO, SYMP, CHMO, OAE, HP, MAXO, COLAO, SSSOM

Funding

1. **Rapid Assessment of Platform Technologies to Expedite Response**
   Defense Threat Reduction Agency (DTRA), (FP00012844), $623K (subcontract) / $15M overall, 2022-2024 (18 months), Grant: HDTRA1242031
   Role: Key Person. Co-wrote proposal with PNNL collaborators. Responsible for design and implementation of research plan.

2. **MIRA: Modeling with an Intelligent Reasoning Assistant**
   DARPA ASKEM Program, (HR00112220036), $2.1M, 2022-2026 (42 months)
   Role: Participant

3. **Young Faculty Award / DARPA Director's Fellowship Award**
   DARPA, (W911NF2010255), $750K, 2020-2023 (36 months)
   Role: Participant

4. (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)
   Role: Participant

5. (past) **PYBEL2NDEX**
   University of California, San Diego, $28K 2018 (6 months)
   Role: Key person. Wrote proposal. Responsible for design and implementation of research plan.

Community

Professional Affiliations

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

Service to the Community

Scholarly Journals

1. Scientific article reviewer in:
   - Bioinformatics
   - Database
   - BMC Bioinformatics
   - Journal of Cheminformatics
   - Journal of Biomedical Semantics
   - eLife
   - MATCH Communications in Mathematical and in Computer Chemistry
2. Reproducibility Editor, Journal of Cheminformatics (pending)

Conference Organizing Committees

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

Teaching

Courses Taught

University of Bonn

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

Northeastern University

1. Drug Discovery and Development (Lecture; Summer II 2015)
2. Organic Chemistry II for Majors (Lecture; Spring 2015)
3. Organic Chemistry I for Majors (Lecture; Fall 2014)
4. Organic Chemistry II for Majors (Lecture; Spring 2014)
5. Organic Chemistry I for Majors (Lecture; Fall 2013)

Supervision

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)