

**SECTION 4: CURRICULUM VITAE AND LIST OF COLLABORATORS****CURRICULUM VITAE**

Hong Cui  
 Professor  
 School of Information, University of Arizona  
 hongcui@email.arizona.edu

**Chronology of Education**

- |                        |        |   |
|------------------------|--------|---|
| 1998-2001<br>2002-2005 | Ph. D. | University of Illinois at Urbana-Champaign<br>Major Field: Library and Information Science<br>Dissertation: "Automating Semantic Markup of Semi-Structured Text via an Induced Knowledge Base: A Case-Study Using Floras."<br>Advisors: Linda C. Smith and Bryan P. Heidorn |
| 2001-2002              | M. S.  | University of Illinois at Urbana-Champaign<br>Major Field: Computer Science   |
| 1994-1997              | M. S.  | Graduate University of Chinese Academy of Sciences<br>Major Fields: Information Science<br>Thesis: "A Study on Citation Behavior of Chinese Scientists."<br>Advisor: Yitai Gong   |
| 1989-1994              | B. S.  | Tongji Medical University (now Tongji Medical College of Huazhong University of Science & Technology)<br>Major Field: Medical Information   |

**Chronology of Employment**

- |                    |  |
|--------------------|--|
| July 2019-present  | Professor, School of Information, University of Arizona  |
| July 2013-Jun 2018 | Associate Professor, School of Information, University of Arizona  |
| July 2007-May 2013 | Assistant Professor, School of Information Resources and Library Science (SIRLS, now the School of Information), University of Arizona   |
| Jan 2005-May 2007  | Assistant Professor, Faculty of Information and Media Studies, University of Western Ontario, Canada                                     |
| Aug 1998-Dec 2004  | Graduate Research and Teaching Assistant, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign |

Aug 1997-Jun 1998 Assistant Librarian, Shanghai Documentation and Information Center of the Chinese Academy of Sciences

### **Honors and Awards**

- 2019 Students' Choice Awards, Spring 2019. Library and Information Student Organization, School of Information, University of Arizona
- 2017 XIX International Botanical Congress 2017 Speaker Award
- 2015 Best Poster Award, SEPEEG 2015  
"Semi-automated extraction of phenomic characters from taxonomic descriptions: natural language processing approaches for the Tree of Life."
- 2011 Outstanding Full-Time Faculty Award, Fall 2011, Library Student Organization, School of Information Resources and Library Science, University of Arizona
- 2011 Outstanding Full-Time Faculty Award, Spring 2011, Library Student Organization, School of Information Resources and Library Science, University of Arizona
- 2009 Outstanding Full-Time Faculty Award, Spring 2009, Library Student Organization, School of Information Resources and Library Science, University of Arizona
- 2009 Most Progressive Full-Time Faculty Award, Spring 2009, Progressive Librarians Guild SIRLS chapter, University of Arizona.
- 2007 Beta Phi Mu International Library and Information Studies Honor Society
- 2005 The Berner Nash Memorial Award for outstanding doctoral dissertation, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign
- 2004 The Incomplete List of Teachers Ranked as Excellent by Their Students, University of Illinois at Urbana-Champaign
- 2004 Jean Tague-Sutcliffe Award, Association for Library and Information Science Education
- 2002 Fellowship, Graduate School of Library and Information Science, University of Illinois at Urbana-Champaign

### **Service/Outreach** (limited to period in current rank)

#### Local/State Outreach

- 2017 Guest lecture for a Master of Science Program's core course: INFO 505 Foundations of Information
- 2016 Guest lecture for INFO 505 Foundations of Information
- 2014 Provided assistance in validating XML files received by University of Arizona Press

#### National/International Outreach

- 2009-present Member, Information Technology Committee of Flora of North America Project
- 2009-present Member, Plazi.org, a Switzerland-based association supporting and promoting the development of persistent and openly accessible digital taxonomic literature
- 2016-present Review Editor, Journal of Association for Information Science and Technology
- 2016-18 Member, Program Committee, iConference, an annual conference for iSchools and information science professionals around the globe.
- 2016-18 Treasurer, Arizona Chapter, the Association for Information Science and Technology
- 2009-18 Ad Hoc Reviewer, Natural Sciences and Engineering Research Council of Canada (2009, 2011, 2012, 2014, 2015, 2018)
- 2018 Organizer and host, Joint ETC-Author Project Meeting, around 20 world-class biologists, biodiversity informaticians, computer and information scientists participated in this NSF funded meeting
- 2017 Member, Proposal Review Panel, National Science Foundation on Big Data
- 2017 Member, Proposal Review Panel, National Science Foundation on Bioinformatics
- 2017 Chair, Program Committee, Ontologizing Botanical Knowledge Symposiums, International Botany Congress 2017
- 2016 Chair, Southwest Chapter, Chinese American Librarian Association

- 2016 Co-chair, Membership Committee, Chinese American Librarian Association
- 2016 Chair, ProQuest ASIST Dissertation Award Jury
- 2016 Trainer, Next Generation Phenomics End User Workshop, Botany 2016 Meeting
- 2016 Member, Program Committee, Semantics and Ontology Symposium, Biodiversity Information Standards 2016 Meeting (formally Taxonomic Database Working Group)
- 2013-15 Project Advisor, the Art of Life project, funded by National Endowment for the Humanities
- 2015 Organizer and co-host, ETC Information Visualization Workshop
- 2015 Co-organizer and host, Next Generation Phenomics Annual Project Meeting
- 2015 Technical Reviewer, Maryland Industrial Partnerships Program
- 2013-14 Project Advisor, Pro-ibiosphere project, funded by the European Union

National and International Service: Ad Hoc Reviewer for Journals/Conferences

*Journal of the American Society for Information Science and Technology*

*International Journal of Library and Information Science*

*Biodiversity Informatics*

*Biodiversity Data Journal*

*Computational Intelligence*

*Bioinformatics*

*IEEE/ACM Transactions on Computational Biology and Bioinformatics*

*iConference*

Invited Participation/Leadership at Professional Meetings

- 2017 Session Chair on “Ontologizing Botanical Knowledge” at XIX International Botany Congress, 2017, Shenzhen, China
- 2016 Session Chair on “Health Informatics” iConference 2016
- 2016 Panel Member, Early Career Colloquium, iConference 2016

Departmental Committees

2015-present	Director of Graduate Studies
2015-present	Chair, Graduate Committee, and Advanced Studies Committee
2015-present	Chair, Financial Aid Committee
2015-present	Member, Executive Committee
2015-present	Advisor, Master of Science Program
2013-present	Member, Curriculum Committee
2017	Chair, Search Committee for Human Communication Interaction/Computational Arts Cluster Assistant/Associate Professor
2017	Alternative Member, Faculty Annual Review Committee
2017	Chair, 3 <sup>rd</sup> -year Review Committee for Dr. X.
2016	Chair, P&T Review Committee for Dr. X
2016	Member, Search Committee for Data Search Assistant Professor
2013-2014	Chair, Master of Science Curriculum Proposal Committee
2013-2014	Graduate Advising Coordinator
2012-2014	School of Information Planning Committee

#### College Committees

2018	Member, SBS University Fellowship Review Panel
2017-present	Member, SBS Research Intuition Board
2016	Member, SBS IT Advisory Board

#### University Committee

2017-present	Graduate Council
--------------	------------------

#### **Publications/Creative Activities (Published or Accepted)**

\* Denotes publication based on work done as graduate student

\*\* Corresponding author

Chapters in Scholarly Books

- Carrine, E. B, **Cui, H.**, Mao, J., Moore, L.R., Thacker, R.W., & Walls, R. L. (in press) Accelerating the Large-Scale Study of Prokaryotic Phenotypes and Niche Space Evolution Through Time Using the Micro Ontology, MicroPIE, and MicroPIEDigester. *Software In Application of Semantic Technologies in Biodiversity Science*, Ed. Anne Thessen.
- Cui, H.**, Sai, D., & Tang, X. (2007). Chapter 3: Information Representation. In Heting Chu & Yin Zhang. (Eds.), *Research Fronts in Library and Information Science in the West*. Beijing: Renmin University Press. (Series on Research Fronts in the Humanities and Social Sciences in the West). [in Chinese]

Refereed Journal Articles, published or accepted in final form

- Cui, H.**, Macklin, J.A., Sachs, J., Reznicek, A., Starr, J., Ford, B., Penev, L., Chen, H.L., (2018) Incentivizing use of structured language in biological descriptions: Author-driven phenotype data and ontology production. *Biodiversity Data Journal* 6: e29616. <https://doi.org/10.3897/BDJ.6.e29616>.
- Endara,L., Thessen, A.E., Cole, H.A., Walls, R., Gkoutos, G., Cao, Y., Chong, S.S., **Cui, H\***(2018). Modifier Ontologies for frequency, certainty, degree, and coverage phenotype modifiers. *Biodiversity Data Journal* 6: e29232. <https://doi.org/10.3897/BDJ.6.e29232>.
- Pender, J., Sachs.J.L., Macklin, J.A., **Cui, H.**, Vallance, A., Lujan-Toro, B., Rodenhausen, T., Belisle-Leclerc, M., Levin, G. (2018). Bringing a Semantic MediaWiki Flora to Life *Biodiversity Information Science and Standards* 2: e25885. <https://doi.org/10.3897/biss.2.25885>
- Dahdul, W., Manda, P., **Cui, H.**, Balhoff, J., Dececchi, A., Ibrahim, N., Lapp, H., Mabee, P., & Vision, T. (2018). Annotating phenotypes using ontological concepts: Inter-curator consistency as a baseline for evaluating the performance of a natural language processing system. *Database (Oxford)*. <https://doi.org/10.1093/database/bay110>. The corpus in the different formats, as well as the ontologies and annotations generated in its production, have been archived at Zenodo (<https://doi.org/10.5281/zenodo.1217594>). The source code for the analysis of inter-curator and SCP consistency based on semantic similarity metrics, as well as the data and ontologies used as input, have been archived separately, also at Zenodo (<https://doi.org/10.5281/zenodo.1218010>). Semantic CharaParser is available in source code from GitHub (<https://github.com/phenoscape/phenoscape-nlp/>) under the MIT license. The version used for this paper is the 0.1.0-goldstandard release (<https://github.com/phenoscape/phenoscape-nlp/releases/tag/v0.1.0-goldstandard>), which is also archived at Zenodo (<https://doi.org/10.5281/zenodo.1246698>).

- Xu, D\*\*, Chong S., Rodenhause, T., & **Cui, H\*\***. (2018) Resolving “orphaned” parts using machine learning and natural language processing methods. *Biodiversity Data Journal*. <https://doi.org/10.3897/BDJ.6.e26659>.
- Mao, J. and **Cui, H.\*\*** (2018) Identifying bacterial biotope entities using sequence labeling: performance and feature analysis. *Journal of Association for Information Science and Technology*. <https://doi.org/10.1002/asi.24032>
- Endara, L., Burleigh, J.G., Rodenhause, T. & **Cui, H.** (in press) Natural Language Processing pipeline, a useful tool for the biology, taxonomy or systematics classroom or laboratory. *CourseSource – Teaching tools and strategies*.
- Endara, L., **Cui, H.** & Burleigh, J. G. (in press) A New Approach to Teach Taxonomy and Scientific Research Skills using Natural Language Processing. *CourseSource - Lessons*.
- Endara, L., **Cui, H.** & Burleigh, J. G. (2018) Semiautomatic extraction of phenotypic traits from taxonomic descriptions using a Natural Language Processing approach. *Applications in Plant Sciences*. DOI: 10.1002/aps3.1035
- Endara, L., Cole, H.A., Burleigh, J.G., Nagalingum, N., Macklin, J.A., Liu, J., & **Cui, H\*\***. (2017) Building the “Plant Glossary” — A controlled botanical vocabulary using terms extracted from the Floras of North America and China. *TAXON*. 66(4), 953-966. DOI: 10.12705/664.9
- Mao, J. Moore, L., Blank, C. Wu, E.H-H, Ackerman, M., Ranade, S., & **Cui, H\*\***. (2016). Microbial Phenomics Information Extractor (MicroPIE): A natural language processing tool for the automated acquisition of prokaryotic phenotypic characters from text sources. *BMC Bioinformatics*. 17(1), 528-543. DOI 10.1186/s12859-016-1396-8
- Cui, H.\*\***, Xu, D., Chong, S.S., Ramirez, M.J., Rodenhause, T., Macklin, J.A., Ludascher, B., Morris, R.A., Soto, E. M., & Koch, N.M. (2016). Introducing Explorer of Taxon Concepts with a case study on spider measurement matrix building. *BMC Bioinformatics*. 17(1),471-492. DOI:19.1186/s12859-016-1352-7.
- Blank, C., **Cui, H.**, Moore, L., Ramona, W. (2016). MicroO: an ontology of phenotypic and metabolic characters, assays, and culture media found in prokaryotic taxonomic descriptions. *Journal of Biomedical Semantics*. 7(1), 1-10, DOI: 10.1186/s13326-016-0060-6
- Huang, F, Macklin, J.A., **Cui, H.\*\***, Cole, H.A., & Endara, L. (2015). OTO: Ontology Term Organizer. *BMC Bioinformatics*. 16(1),47-64. DOI:10.1186/s12859-015-0488-1
- Deans AR, Lewis SE, Huala E, Anzaldo SS, Ashburner M, et al [14<sup>th</sup> of 74 authors]. (2015). Finding our way through phenotypes. *PLoS Biology* 13(1): e1002033. DOI:10.1371/journal.pbio.1002033, 9 pages. [perspective paper].

- Dahdul, W.M., **Cui, H.**, Mabee, P. et al. (2014). Nose to tail, roots to shoots: spatial descriptors for phenotypic diversity in the Biological Spatial Ontology. *Journal of Biomedical Semantics*. 5(1), 34-46. DOI:10.1186/2041-1480-5-34.
- Duan, YF., Hei ZZ., Ju, F., & **Cui, H.**(2013). Heuristics based semantic annotation of biodiversity documents in Chinese. *Chinese Journal of Library and Information Science (English)*. 6(2):33-46.  
[http://ir.las.ac.cn/handle/12502/6238?mode=full&submit\\_simple>Show+full+item+record](http://ir.las.ac.cn/handle/12502/6238?mode=full&submit_simple>Show+full+item+record)
- Burleigh, G, et al. [7<sup>th</sup> in 28 authors] (2013). Next generation phenomics for the Tree of Life. *PLOS Currents: Tree of Life*. <http://currents.plos.org/treeoflife/article/next-generation-phenomics-for-the-tree-of-life/>
- Arighi, C.N., et al. [23<sup>rd</sup> of 43 authors] (2013). An overview of the BioCreative 2012 Workshop Track III: interactive text mining task. *Database: The Journal of Biological databases and Curation (Oxford)*. DOI: 10.1093/database/bas056, 18 pages.
- Duan, Y, Hei, Z, Ju, F., **Cui, H.** (2012). Semantic annotation of species description text in Chinese literature by Naive Bayes Classifier. *Journal of the China Society for Scientific and Technical Information (Chinese)*. 31(8), 805-812.
- Duan, Y, Hei, Z, Ju, F., **Cui, H.** (2012). Study on semantic markup of species description text in Chinese based on auto-learned rules. *New Technology of Library and Information Services (Chinese)*. 5, 41-47.
- Thessen, A., **Cui, H.**, & Mozzherin, D. (2012). Applications of natural language processing in biodiversity science. *Advances in Bioinformatics, 2012*, DOI: 10.1155/2012/391574
- Cui, H.** (2012). CharaParser for fine-grained semantic annotation of organism morphological descriptions. *Journal of the American Society for Information Science and Technology*, 63(4), 738-754.
- Cui, H.**, Duan, Y. & Li, F. (2011). Machine learning based semantic markup of biodiversity literature in English. *Document, Information, & Knowledge (in Chinese)*, 2, 73-77.
- Cui, H.** (2010). Competency evaluation of plant character ontologies against domain literature. *Journal of the American Society for Information Science and Technology*, 61(6), 1144-1165.
- Cui, H.\*\***, Boufford, D., & Selden, P. (2010). Semantic annotation of biosystematics literature without training examples. *Journal of the American Society for Information Science and Technology*, 61(3), 522-542.
- Cui, H.** (2010). Semantic annotation of morphological descriptions: An overall strategy. *BMC Bioinformatics*, 11(1): 278-288. DOI:10.1186/1471-2105-11-278.



- Cui, H.** (2008). Converting taxonomic descriptions to new digital formats. *Biodiversity Informatics*, 5, 20-40.
- \*Cui, H.\*\*** & Heidorn, P.B. (2007). The reusability of induced knowledge for the automatic semantic markup of taxonomic descriptions. *Journal of the American Society for Information Science and Technology*, 58(1), 133-149.
- Cui, H.\*\*** & Nickerson, G. (2007). Use Server2Go to teach IT courses for LIS students. *Journal of Association for Library and Information Science Education*, 48, 261-271.
- Cui, H.** (1998). A clustering analysis on citation motivations of Chinese scientists. *Journal of Information Science (in Chinese)*, 17(2), 68-70.
- Cui, H.** (1998). Analyzing self-citations in Chinese scientists. *Information: Theory and Application (in Chinese)*, 21(3), 153-154, 176.

#### Refereed conference articles

- Endara, L., Burleigh, G., Cooper, L., Jaiswal, P., Laporte, M-A., & **Cui, H.** (2018) A Natural Language Processing Pipeline to Extract Phenotypic Data from Formal Taxonomic Descriptions with a Focus on Flagellate Plants. *Proceedings of the 2018 International Conference on Biomedical Ontologies*. [http://ceur-ws.org/Vol-2285/ICBO\\_2018\\_paper\\_50.pdf](http://ceur-ws.org/Vol-2285/ICBO_2018_paper_50.pdf)
- Zhang, Y. Burkell, J., **Cui, H.**, Mercer, R. (2018). An automated approach for rating the content quality of Web healthcare information: A case study on depression treatment Web pages. *Proceedings of the 2018 International Conference on Health Informatics & Medical Systems*, (eds) Hamid R. Arabnia, Leonidas Deligiannidis, Fernando G. Tinetti, Quoc-Nam Tran, pp 3-8, 2018.
- Zhang, Y. Mercer, R., Burkell, J. & **Cui, H.** (2018). Using shallow semantic analysis to implement automated quality assessment of web health care information. *The 19th International Conference on Computational Linguistics and Intelligent Text Processing, CICLing 2018* March 18 to 24, 2018 in Hanoi, Vietnam, Selected for publication in Springer LNCS.
- Yadav, V., Sadeque, F.Y., Heidorn, B., & **Cui, H.** (2018). Where are iSchools heading? *iConference 2018*.
- Dang, T., **Cui, H.**, & Forbes, A.G. (2016). MultiLayerMatrix: Visualizing large taxonomic datasets. In *Proceedings of the EuroVis Workshop on Visual Analytics (EuroVA)*, Groningen, Netherlands, June 2016.

- Cui, H.\*\***, Dahdul, W., Dececchi, A., Ibrahim, N., Mabee, P., Balhoff, J., Gopalakrishnan, H. (2015). CharaPaser+EQ: Performance evaluation without gold standard. *Annual Meeting of the Association for Information Science and Technology*, Nov 6-10, St Louis, Missouri, 2015. (acceptance rate: 36%)  
<https://www.asist.org/files/meetings/am15/proceedings/openpage15.html>
- Zhang, Y., **Cui, H.**, Burkell, & J. Mercer, R.E. (2014) A machine learning approach for rating the quality of depression treatment web pages. *iConference*, 2014.  
<http://hdl.handle.net/2142/47314>
- Cui, H.**, Balhoff J., Dahdul W., Lapp H., Mabee P., Vision T., & Chang, Z. (2012). PCS for Phylogenetic Systematic Literature Curation. *Proceedings of the BioCreative workshop 2012* (pp.137-144).
- Cui, H.\*\***, Singaram, S., & Janning, A. (2011). Combine unsupervised learning and heuristic rules to annotate morphological characters. *Proceedings of Annual Meeting of American Society of Information Science and Technology 2011*
- Cui, H.\*\***, Sanyal P.P., & Yu C. (2010). Tools for semantic annotation of taxonomic descriptions. In R. Setchi et al. (Eds.): *Proceedings of 14th International Conference on Knowledge-Based and Intelligent Information and Engineering Systems, Part IV, Lecture Notes in Artificial Intelligence*, 6279, 506-516.
- Cui, H.** (2010). Linking corpus characteristics to performance of semantic annotation systems for biosystematic descriptions. *Proceedings of the 2nd International Conference on Bioinformatics and Biomedical Technology* (pp. 92-96).
- Cui, H.** (2008). Approaches to semantic mark up for natural heritage literature. *Proceedings of the iConference 2008*. [http://www.ischools.org/oc/conference08/pc/PA5-2\\_iconf08.doc](http://www.ischools.org/oc/conference08/pc/PA5-2_iconf08.doc)
- Cui, H.** (2008). Unsupervised semantic markup of literature for biodiversity digital libraries. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 25-28). [short paper]
- \***Cui, H.** (2005). MARTT: Using induced knowledge base to automatically mark up plant taxonomic descriptions with XML. *Proceedings of the Annual Meeting of the American Society for Information and Technology*. DOI: 10.1002/meet.1450420170
- \***Cui, H.** (2005). MARTT: A general approach to automatic markup of taxonomic descriptions with XML. *Proceedings of the Annual Conference of Canadian Association for Information Science 2005*. [http://www.cais-acsi.ca/proceedings/2005/cui\\_2005.pdf](http://www.cais-acsi.ca/proceedings/2005/cui_2005.pdf)
- \***Cui, H.**, Heidorn, P.B., & Zhang, H. (2002). An approach to automatic classification for information retrieval. *Proceedings of the Joint Conference on Digital Libraries 2002* (pp. 96-97).

\*Heidorn, P.B. & **Cui, H.** (2000). The interaction of result set display dimensionality and cognitive factors in information retrieval systems. *Proceedings of the Annual Meeting of the American Society for Information Science* (pp. 258-270).

Refereed conference posters, system demos and abstracts

Chong, S.S., Xu, D., Rodenhause, T., & **Cui, H\*** (2018) Resolving Orphaned Parts in Taxonomic Descriptions with Machine Learning and Natural Language Processing Methods. *The Second Annual Digital Data in Biodiversity Research Conference*, Jun 4-6, 2018. UC Berkeley.

Cao, Y., Franz, N., Macklin, J., Mao, J. & **Cui, H\*\*** (2018) Resolving Taxonomic Names using Evidence Extracted from Text. *iConference 2018*. [poster]

**Cui, H.** (2017) Bottom-up ontology building *IBC XIX*, July, 2017. Shen Zhen, China. [abstract]

Dahdul, W., Manda, P., **Cui, H.**, Balhoff, J. Dececchi, A., Ibrahim, N., Mabee, P., Vision, T. (2017). Gold standard evaluation of machine- and human-generated annotations of biodiverse phenotypes. *Biocuration 2017*. [abstract]

Joseph, C., Stadelmann, K., Endara, L., & **Cui, H.** (2016). The New Language of Science: A novel approach to generating a phenomic matrix of gymnosperms. *Botany 2016*, Savannah, GA. [poster]

Endara, L., Burleigh, J.B., & **Cui, H.** (2016). Using natural language processing tools to facilitate learning taxonomy, nomenclature, and botany in the classroom. *Botany 2016*. July 30-Aug 3, Savannah, Georgia. 2016. [abstract].

Endara, L., **Cui, H.**, Burleigh, J.B., & Nagalingum, N. (2016). Constructing morphological character matrices from taxonomic descriptions using new software approaches. *Botany 2016*. July 30-Aug 3, Savannah, Georgia. 2016.[abstract and oral presentation]

**Cui, H\*\*.**, Macklin, J.A., Sachs, J., Rodenhause, T., Chong, S., & Xu, D. (2016) Bottom up, pattern-based ontology building for phenotype characters. *TDWG 2016 Conference*. Dec 3-10, 2016. Costa Rica. [abstract]

Mao, J., Moore, L., Blank, C. & **Cui, H\*\*.** (2016). MicroPIE *iConference 2016*. [poster]

Endara, L., Burleigh, G., & **Cui, H.** (2015). Semi-automated extraction of phenomic characters from taxonomic descriptions: natural language processing approaches for the Tree of Life. *SEPEG 2015*. Eatonton, Georgia, USA [Best Poster Award]

Blank, C. E., Wu, H., **Cui, H.**, Moore, L.R., Burleigh, J.G., Liu, J., & Gasparich GE. (2014). AVAToL microbial phenomics: an ontology and natural language processing tools to facilitate trait evolution studies for the archaeal domain of life. *Evolution 2014 meeting*, Raleigh, NC, June 20-24. [abstract]

- Endara, L., Liu, J., Burleigh, J.G., Nagalingum, N., & **Cui, H.** (2014). Transformation of taxonomic literature to a taxon-character matrix for phylogenetics. *Evolution 2014 Meeting*. [abstract]
- Macklin, J. A., Cole, H., **Cui, H.**, & Sachs, J. (2014). Prototyping a botanical knowledge portal. Canadian Botanical Association Annual Meeting. Montreal, QC. June 16, 2014. [poster]
- Rodenhausen, T., **Cui, H.**, Huang, F., Ludäscher, B., Macklin, J., Morris, R., Yu, S. (2014). ETC: From description to matrix and beyond in a web-based toolbox. *TWDG 2014 Meeting*. [software demo].
- Blank, C.E., Moore, L.R., **Cui, H.**, Wu, H., Burleigh, G., Liu, J., Slonczewski, J.L., Barich, D., & Gasparich, G.E.(2014). AVAToL microbial phenomics: developing a microbial ontology and natural language processing tools to automate the study of the evolution of microbial traits. *Joint Aquatic Sciences Meeting*, Portland, OR, May 18-23. [poster].
- Moore, L. R., Blank, C., Burleigh, J. G., **Cui, H.**, Gasparich, G., Liu, J. & Ranade, S. (2013). Next generation microbial phenomics, Australian Association for Microbiology Annual Scientific Meeting [poster].
- Cui, H\*\*.**, Huang, F., & Rodenhausen, T. (2013). Semantic annotation of organism morphological descriptions using CharaParser along with a web-based ontology term organizer (OTO). *IEvoBio 2013*. [software demo].
- Ranade, S., **Cui, H\*\***, Moore, L. et al. (2013). A preliminary analysis of application of Stanford Parser and OSCAR4 for parsing and annotating microbial descriptions. *IEvoBio 2013*. [abstract].
- Walls, R., **Cui, H.**, Macklin J.A., Mungall, C., Cooper, L., Stevenson, D.W., & Jaiswal, P.(2012). Mapping of glossary terms from the Flora of North America to the Plant Ontology enhances both resources. *Proceedings of the 3rd International Conference on Biomedical Ontology*, June 2012 Austria. [extended abstract]
- Huang, F.Q., Macklin, J., Morris, P., Sanyal, P.P., Morris, R.A. , Cole, H. & **Cui, H\*\*** (2012). OTO: Ontology Term Organizer. *Annual Conference of American Society for Information Science and Technology*. [poster]
- Cui, H.\*\***, Dusenbery, A., Morris, R.A., Macklin, J., & Huang, F (2012). Semantic annotation, ontology building, and interactive key generation from morphological descriptions. *TDWG Annual Meeting 2012*, Beijing, Oct 22-26, 2012. [abstract]
- Chang Z., Balhoff J., Dahdul W., Lapp H., Mabee P., Vision T., & **Cui, H\*\***. (2012). Workflow of CharaParser and Phenex: Turning character descriptions to EQ statements. *BioCreative workshop 2012*, Washington, D.C. [poster and software demo]

- Janning, A. & **Cui, H\*\***. (2012). Evaluating the botanical coverage of PATO using an unsupervised learning algorithm. *Proceedings of the 2012 iConference* (pp. 505-505). [poster]
- Macklin, J., **Cui, H.**, Morris, R., & Morris, P. (2011). Floras in the 21st Century: The Flora of North America. Panelist in *Creating Next Generation Floras Symposium, XVIII International Botanical Congress*. [abstract]
- Cui, H.** (2011). Fine-grained semantic markup of descriptive data. Panelist in *Informatics Tools for the Semantic Enhancement of Taxonomic Literature Symposium, XVIII International Botanical Congress* (on USB). [abstract]
- Cui, H.**, Jiang, Y., & Sanyal, P.P. (2010). From text to RDF triple store: An application for biodiversity literature. *Proceedings of the Annual Meeting of American Society of Information Science and Technology*. [software demo]
- Cui, H.** (2010). Unsupervised extraction of text segments from heterogeneous document collections. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology 2010*. Oct. 2010. [poster]
- Cui, H.\*\***, & Sanyal, P. P. (2009). "Fine-Grained Semantic Annotation of Descriptive Data for Knowledge Application in Biodiversity". *Taxonomic Database Working Group 2009 Annual Meeting*. Paris, France. [software demo]
- Cui, H.**, Yu, C., & Macklin, J. (2009). Application of semantic annotation for quality insurance in biosystematics publishing. *Proceedings of the Annual Meeting of the American Society for Information Science and Technology* (on CD). [poster]
- Cui, H.** (2008). An application for semantic markup of biodiversity documents. *Proceedings of the 8th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp.421). [software demo]
- McCourt, R.M., **Cui, H.**, Guiry, M. & Feist, M. (2006). Using machine learning environments to extract taxonomic information from text: An example from print and online texts on algae. *Phycological Society of America Annual Meeting 2006*. Juneau, AK. [poster]
- Hirokawa, S. & **Cui, H.** (2006). Automatic generation of hierarchy for plant identification terminology. *Proceedings of the 2nd International Digital Curation Conference*. Glasgow, UK. [software demo]
- Cui, H.**, McCourt, R. M., & Feist, M. (2006). Unsupervised structure discovery for biodiversity information. *Proceedings of the 6th ACM/IEEE-CS Joint Conference on Digital Libraries* (pp. 382). [software demo]
- Cui, H.**, McCourt, R. M., & Feist, M. (2006). Automated concept discovery in corpora of morphological descriptions. *Proceedings of the Annual Meeting of the American Society for Information and Technology* (on CD-ROM).[poster]

- Cui, H.** (2005). A machine learning environment for automatic markup of taxonomic descriptions with XML. *Taxonomic Database Working Group 2005 Annual Meeting* (pp.16). [software demo]
- Cui, H.** (2004). Knowledge-based semantic markup of plant descriptions. *Doctoral Students SIG Research Poster Session, Annual Conference of the Association for Library and Information Science Education 2004*, San Diego, CA. [poster, Jean Tague-Sutcliffe Award]
- Heidorn, P.B., **Cui, H.**, Yu, B. Wu, J., & Zhang, H. (2002). Taxonomic description creation, search and display in XML. *Botany 2002*. Madison, WI.[abstract]
- Cui, H.** (2002). Automatic/semi-automatic parse of Flora of North America records into XML format using machine-learning techniques. *Doctoral Students SIG Research Poster Session, Annual Conference of the Association for Library and Information Science Education*, New Orleans, LA. [poster]

#### Other creative products

- ETC Toolkit: Five tools are provided for rapidly converting taxonomic descriptions into character matrices for systematics and evolutionary studies. <http://etc.cs.umb.edu/etcsite/>
- OTO: Ontology Term Organizer: Conflict-exposing and consensus-promoting multi-user controlled vocabulary building web application, treating *is\_a*, *part\_of*, and *order* relationships critical in biodiversity domain. <http://biosemantics.arizona.edu/OTO/>

#### Github code repositories

- <https://github.com/biosemantics>  
<https://github.com/hongcui/>  
<https://github.com/phenoscape/phenoscape-nlp>

#### **Work in Progress**

- Cao, Y., **Cui, H\*\***, Mao, J. et al. (in prep). User study of the ETC ontology building tool. *BMC Bioinformatics*
- Cao, Y., Franz, N., Macklin, J., Mao, J. & **Cui, H.\*\*** (in prep) Character evidence-based taxon concept comparison. *TAXON*

#### **Conferences/Scholarly Presentations** (limited to period in current rank)

#### Colloquia

- 2017            iSchool, University of Arizona. Talk title: Authors in the driver's seat: new paradigm for phenotype data and ontology production

- 2016 iSchool, Wuhan University, China. Talk title: Constructing and using ontologies: case studies in biodiversity domain [invited]
- 2016 School of Management Information Systems, Central China Technology University, China. Talk title: Constructing and using ontologies: case studies in biodiversity domain [invited]
- 2015 iSchool, University of Illinois at Urbana-Champaign. Talk title: Developing controlled vocabularies for organism traits [invited]

### Symposia and Workshops

- 2018 Forthcoming *Next Generation Phenomics Tools for the Tree of Life Workshop*. University of South Maine [invited].
- 2016 *Next Generation Phenomics Tools for the Tree of Life Workshop*. University of South Maine. Talk title: Explorer of Taxon Concepts web toolkit [invited]
- 2016 *Phenotype Ontology Research Collaboration Network Meeting*. Tucson, AZ. Talk title: OTO – Ontology Term Organizer [invited]
- 2015 *ETC Information Visualization Workshop*. Urbana, Illinois. Talk title: ETC user task challenges [organizer]
- 2015 *Next Generation Phonemics Project Annual Meeting* Oracle, AZ. Talk title: From text to matrices: Using Natural Language Processing (NLP) to facilitate assembling taxon-character matrices [co-organizer and host]
- 2014 *Phenotype Ontology Research Collaboration Network Meeting*. Tucson, AZ, Talk title: Explorer of Taxon Concepts project and initial steps [invited]
- 2013 *pro-iBiosphere workshop I*, Leiden, the Netherlands. Talk title: CharaParser and associated software [invited]
- 2013 *pro-iBiosphere workshop II*, Leiden, the Netherlands. Talk title: Parsing morphological descriptions to support semantic-based access [invited]
- 2013 *pro-iBiosphere workshop III*, Leiden, the Netherlands. Talk title: Markup tools: CharaParser [invited]

### Conferences

- 2018 2018 International Conference on Biomedical Ontologies. Eugene, Oregon. Talk

title: ETC Toolkits

- 2018 iConference 2018, Sheffield, UK. Talk title: Where are iSchools heading?
- 2018 *Phenomes 2018*, Tucson AZ. Talk title: From text blobs to computable data: challenges in mining phenotypical data from text [invited]
- 2017 *Evolution 2017*, Portland Oregon. Talk title: Authors in the driver's seat.
- 2017 XIX International Botanical Congress, Shenzhen, China. Talk title: Bottom-up ontology building
- 2016 *Botany 2016*, Savannah, GA. Talk title: Using natural language processing tools to facilitate learning taxonomy, nomenclature, and botany in the classroom
- 2016 *Biodiversity Information Standards Annual Meeting*. Costa Rica. Talk title: Bottom up, pattern-based ontology building for phenotype characters.
- 2015 *Annual Meeting of the Association for Information Science and Technology*, St Louis, Missouri. Talk title: CharaPaser+EQ: Performance evaluation without gold standard
- 2013 *IEvoBio 2013*, Snowbird Utah. Talk title: A preliminary analysis of application of Stanford Parser and OSCAR4 for parsing and annotating microbial descriptions.

#### **Awarded Grants and Contracts (limit to period in current rank)**

- 2017-2020 \$641,763.00. PI. Project title: Collaborative Research: ABI innovation: Authors in the driver's seat: fast, consistent, computable phenotype data and ontology production. Source: National Science Foundation. Co-PIs: James Macklin and Joe Sachs (Both Canadian researchers, no budget).
- 2017 \$60,000. PI. USM Next Generation Penomics Project Subcontract. Source: University of South Maine. Effort=100%
- 2016 \$100,000. PI. Subcontract to 1147273 Exploring Taxon Concepts (ETC) through Analyzing Fine-grained Semantic Markup of Descriptive Literature. Source: University of Illinois at Urbana-Champaign.
- 2016-2019 \$53,000. PI. Project title: Collaborative Research: Building a Comprehensive Evolutionary History of Flagellate Plants. Source: National Science Foundation.
- 2015-2017 \$117,000. PI. Special Creativity Extension to "Collaborative Research: AVAToL - Next Generation Phenomics for the Tree of Life". Source: National Science Foundation.



- 2014-2017 \$7,500. PI. Research Experience for Undergraduate Students (REU) supplement to "Collaborative Research: AVAToL - Next Generation Phenomics for the Tree of Life". Source National Science Foundation.
- 2012-2015 \$173,000. Senior Personnel. Project title: La SCALA: Latino Scholars Cambio Leadership Academy. Source: Institute for Museum and Library Services
- 2012-2018 \$1,095,946. PI. Project title: BCSP: Collaborative Research: ABI Development: Exploring Taxon Concepts (ETC) through analyzing fine-grained semantic markup of descriptive literature. Source: National Science Foundation.
- 2012-2015 \$335,000. PI. Project title: Collaborative Research: Next Generation Phenomics for the Tree of Life. Source: National Science Foundation.

### **List of Collaborators and their Organizational Affiliations**

#### Collaborators on Grants and Publications

Arighi, Cecilia (University of Delaware)  
 Ackerman, Marcia (University of South Maine)  
 Balhoff, Jim (Duke University)  
 Blank, Carrine (University of Montana)  
 Burkell, Jacquie (Western University, Canada)  
 Burleigh, Gordon (University of Florida)  
 Chong, Steven (National Center for Ecological Analysis and Synthesis)  
 Cole, Heather (Agriculture Agri-Food Canada)  
 Dahdul, Wasila (University of South Dakota)  
 Daly, Mary Megan (Ohio State University)  
 Deans, Andrew (Ohio University)  
 Dietterich, Thomas (Oregon State University)  
 Duan, Yufeng (East China Normal University)  
 Endara, Lorena (University of Florida)  
 Franz, Nico (Arizona State University)  
 Haendel, Melissa (Oregon Health and Science)  
 He Zhenzhen (East China Normal University)  
 Huang, Feiqiong (University of Arizona)  
 Ibrahim, Nizar (University of Chicago)  
 Jaiswal, Pankaj (Oregon State University)  
 Ju, Fei (East China Normal University)  
 Koch, Nicolás (Yale University)  
 Lapp, Hilmar (Duke University)  
 Liu, Jing (Wuhan University, China)  
 Ludascher, Bertram (University of Illinois at Urbana-Champaign)  
 Mabee, Puala (University of South Dakota)  
 Macklin, James (Agriculture Agri-Food, Canada)

Mao, Jin (Wuhan University, China)  
Mercer, Robert (Western University, Canada)  
Moore, Lisa (University of South Maine)  
Morris, Robert(Retired)  
Nagalingum, Nathalie (University of California)  
O'Leary, Maureen (Stony Brook University)  
Ranade, Sonali (University of Arizona)  
Ramirez, Martin (Museo Argentino de Ciencias, Naturales, CONICET, Argentina.)  
Rodenhausen, Thomas(Arizona)  
Sacks, Joel (Agriculture Agri-Food, Canada)  
Simmons, Nancy (American Museum of Natural History)  
Soto, Eduardo (Yale University)  
Stevenson, Dennis (New York Botanical Garden)  
Thacker, Robert (Stony Brook University)  
Theriot, Edward (University of Texas)  
Thessen, Anne (independent researcher)  
Todorovic, Sinisa (Oregon State University)  
Vision, Todd(University of North Carolina)  
Walls, Ramona (University of Arizona)  
Wu, Evis (University of Arizona)  
Yu, Shizhuo (University of California at Davis)

Graduate, Postdoctoral, Thesis Advisors or Sponsors

Boufford, E. David (Harvard University),  
Doan, Anhai (Illinois at Urbana-Champaign),  
Heidorn, P. Bryan (University of Arizona),  
Smith, C. Linda (University of Illinois at Urbana-Champaign).