Bill Howe

billhowe@uw.edu
310B Mary Gates Hall
University of Washington
Seattle, WA 98105
(404) 514 3003

https://faculty.washington.edu/billhowe
twitter: @billghowe
github: billhowe
linkedin: billghowe
slideshare: billhoweuw

Employment

University of Washington 2016–present
Associate Professor, Information School
Adjunct Associate Professor, Computer Science & Engineering
Adjunct Associate Professor, Electrical Engineering
Founding Program Director and Inaugural Faculty Chair, Data Science Masters Degree
Faculty Director and Co-Founder, Responsibility in AI Systems and Experiences (RAISE)
Senior Data Science Fellow, eScience Institute

University of Washington 2009–2017
Founding Associate Director, eScience Institute
Founding Director, Cascadia Urban Analytics Cooperative and Urbanalytics Group
Affiliate Faculty, Computer Science & Engineering

Oregon Health and Science University 2007–2009
Research Scientist

Portland State University 2001–2006
Graduate Research Assistant

Deloitte, Microsoft, Consultant, Schlumberger, Siebel 1999-2001
Consultant

Education

Portland State University
PhD, Computer Science, 2007.
GridFields: Model-driven Data Transformation in the Physical Sciences, under David Maier.

Georgia Institute of Technology
B.S. in Industrial and Systems Engineering, 1999.

Peer-Reviewed Conference and Journal Articles

1. **The Art of Refusal: A Survey of Abstention in Large Language Models**
   Bingbing Wen, Jihan Yao, Shangbin Feng, Chenjun Xu, Yulia Tsvetkov, Bill Howe, Lucy Lu Wang
   Transactions of the Association for Computational Linguistics (ACL) (2025)
2. **Do Language Models Mirror Human Confidence? Exploring Psychological Insights to Address Overconfidence in LLMs**
   Chenjun Xu, Bingbing Wen, Bin HAN, Robert Wolfe, Lucy Lu Wang, Bill Howe
   Transactions of the Association for Computational Linguistics (ACL) (2025)

3. **Fragments to Facts: Partial-Information Fragment Inference from LLMs**
   Lucas Rosenblatt, Bin Han, Robert Wolfe, Bill Howe
   International Conference on Machine Learning (ICML) (2025)
4. **Epistemic Alignment: A Mediating Framework for User-LLM Knowledge Delivery**
   Nicholas Clark, Hua Shen, Bill Howe, Tanu Mitra
   Conference on Language Models (COLM) (2025)
5. **Can Large Language Models Integrate Spatial Data? Empirical Insights into Reasoning Strengths and Computational Weaknesses**
   Bin Han, Robert Wolfe, Anat Caspi, Bill Howe
   Conference on Language Models (COLM) (2025)
6. **Escaping the SpuriVerse: Can Large Vision-Language Models Generalize Beyond Seen Spurious Correlations?**
   Yiwei Yang, Chung Peng Lee, Shangbin Feng, Dora Zhao, Bingbing Wen, Anthony Z. Liu, Yulia Tsvetkov, Bill Howe
   Neural Information Processing Systems (NeruIPS) (2025) (submitted)
7. **PathwayBench: Assessing Routability of Pedestrian Pathway Networks Inferred from Multi-City Imagery**
   Yuxiang Zhang, Bill Howe, Sachin Mehta, Nicholas J Bolten, Anat Caspi
   Transportation Review Board (TRB) (2025)
8. **Towards Zero-Shot Annotation of the Built Environment with Vision-Language Models**
   Bin Han, Yiwei Yang, Anat Caspi, Bill Howe
   Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL) (2024)
9. **SARN: Structurally-Aware Recurrent Network for Spatio-Temporal Disaggregation**
   Bin Han, Bill Howe
   Proceedings of the 32nd ACM International Conference on Advances in Geographic Information Systems (2024)
10. **Dataset Scale and Societal Consistency Mediate Facial Impression Bias in Vision-Language AI**
    Robert Wolfe, Aayushi Dangol, Alexis Hiniker and Bill Howe
    Conference on Artificial Intelligence, Ethics, and Society (AIES) (2024)
11. **ML-EAT: A Multilevel Embedding Association Test for Interpretable and Transparent Social Science**
    Robert Wolfe, Alexis Hiniker and Bill Howe
    Conference on Artificial Intelligence, Ethics, and Society (AIES) (2024)
12. **Representation Bias of Adolescents in AI: A Bilingual, Bicultural Study**
    Robert Wolfe, Aayushi Dangol, Bill Howe, and Alexis Hiniker
    Conference on Artificial Intelligence, Ethics, and Society (AIES) (2024)
13. **Laboratory-Scale AI: Open-Weight Models are Competitive with ChatGPT Even in Low-Resource Settings**
    Robert Wolfe, Isaac Slaughter, Bin Han, Bingbing Wen, Yiwei Yang, Lucas Rosenblatt, Bernease Herman, Eva Brown, Zening Qu, Nic Weber, Bill Howe
    Conference on Fairness, Accountability, and Transparency (FAccT) (2024)
14. **Label-Efficient Group Robustness via Out-of-Distribution Concept Curation**
    Yiwei Yang, Anthony Z Liu, Robert Wolfe, Aylin Caliskan, Bill Howe
    Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) (2024)
15. **Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy**
    Lucas Rosenblatt, Anastasia Holovenko, Taras Rumezhak, Andrii Stadnik, Bernease Herman, Julia

Stoyanovich, Bill Howe
International Conference on Very Large Databases (VLDB) (2023)
(Best Paper Runner Up, Experiment, Analysis & Benchmark Track; selected for research highlight article, SIGMOD Record)

16. **Contrastive language-vision AI models pretrained on web-scraped multimodal data exhibit sexual objectification bias**
Robert Wolfe, Yiwei Yang, Bill Howe, Aylin Caliskan
Conference on Fairness, Accountability, and Transparency (FAccT) (2023)
17. **Responsible data management**
Julia Stoyanovich, Serge Abiteboul, Bill Howe, HV Jagadish, Sebastion Schelter
Communications of the ACM 65 (6), 64-74 (2022)
18. **Ontologue: Declarative Benchmark Construction for Ontological Multi-Label Classification**
Sean T. Yang, Bernease Herman, Bill Howe
Conference on Neural Information Processing Systems (NeurIPS) (2022)
19. **Surj: Ontological Learning for Fast, Accurate, and Robust Hierarchical Multi-label Classification**
Sean T. Yang, Bill Howe *Companion Proceedings of the Web Conference (WWW)* (2022)
20. **Integrative urban AI to expand coverage, access, and equity of urban data**
Bill Howe, Eva Brown, Bin Han, Bernease Herman, Nic Weber, An Yan, Sean T. Yang, Yiwei Yang
The European Physical Journal Special Topics 231, 1741âĂŞ1752 (2022)
21. **Covid-19 brings data equity challenges to the fore**
HV Jagadish, Julia Stoyanovich, Bill Howe
Digital Government: Research and Practice 2 (2), 1-7 (2021)
22. **EquiTensors: Learning Fair Integrations of Heterogeneous Urban Data**
An Yan, Bill Howe
International Conference on Management of Data (SIGMOD) (2021)
23. **JECL: Joint Embedding and Cluster Learning for Image-Text Pairs**
Sean T. Yang, KH Huang, Bill Howe
International Conference on Pattern Recognition (ICPR) (2021)
24. **SPORES: Sum-Product Optimization via Relational Equality Saturation for Large Scale Linear Algebra**
YR Wang, S Hutchison, J Leang, B Howe, D Suciu
Proceedings of the VLDB Endowment 13 (12), 3474-3488 (VLDB) (2020)
25. **Responsible data management**
Julia Stoyanovich, Bill Howe, HV Jagadish
Proceedings of the VLDB Endowment 13 (12), 3474-3488 (VLDB) (2020)
26. **Fairness-Aware Demand Prediction for New Mobility**
An Yan, Bill Howe
The AAAI Conference on Artificial Intelligence (AAAI) (2020)
27. **Identifying the Central Figure of a Scientific Paper**
Sean T Yang, Po-Shen Lee, Lia Kazakova, Abhishek Joshi, Bum Mook Oh, Jevin D West, Bill Howe
International Conference on Document Analysis and Recognition (ICDAR) (2019)
28. **FairST: Equitable Spatial and Temporal Demand Prediction for New Mobility Systems**
An Yan, Bill Howe
Proceedings of the 27th ACM International Conference on Advances in Geographic Information Systems (SIGSPATIAL) (2019)

29. **Interventional Fairness: Causal Database repair for Algorithmic Fairness**
    Babak Salimi, Luke Rodriguez, Bill Howe, Dan Suciu
    International Conference on Management of Data (SIGMOD) (2019) (Best Paper Award)
30. **Falcon: Balancing Interactive Latency and Resolution Sensitivity for Scalable Linked Visualizations**
    Dominik Moritz, Bill Howe, Jeffrey Heer
    Computer Human Interaction (CHI) (2019)
31. **MobilityMirror: Bias-Adjusted Transportation Datasets**
    Luke Rodriguez, Babak Salimi, Julia Stoyanovich, Bill Howe
    Journal of Information Systems (2019)
    (invited submission, Special Issue on Big Data for Smart Cities)
32. **Database-Agnostic Workload Management**
    Shrainik Jain, Jiaqi Yan, Thierry Cruane, Bill Howe
    Conference on Innovative Database Research (CIDR) (2019)
33. **Beyond Open vs. Closed: Balancing Individual Privacy and Public Accountability in Data Sharing**
    Meg Young, Luke Rodriguez, Emily Keller, Feiyang Sun, Boyang Sa, Jan Whittington, Bill Howe
    Fairness, Accountability, and Transparency (FAT) (2019)*
34. **Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco**
    Dominik Moritz, Chenglong Wang, Greg L. Nelson, Halden Lin, Adam M. Smith,
    Bill Howe, Jeffrey Heer
    IEEE Information Visualization (InfoVis) (2018)
    (Best Paper Award)
35. **EZLearn: Exploiting Organic Supervision in Automated Data Annotation**
    Maxim Grechkin, Hoifung Poon, Bill Howe
    International Joint Conference on Artificial Intelligence (IJCAI) (2018)
36. **Viziometrics: Analyzing visual information in the scientific literature**
    Po-shen Lee, Jevin D. West, and Bill Howe
    IEEE Transactions on Big Data (2018)
37. **Fides: Towards a Platform for Responsible Data Science.**
    Julia Stoyanovich, Bill Howe, Serge Abiteboul, Gerome Miklau, Arnaud Sahuguet, Gerhard Weikum
    ACM Conference on Scientific and Statistical Database Management (SSDBM) (2017)
38. **PhyloParser: A Hybrid Algorithm for Extracting Phylogenies from Dendrograms**
    Poshen Lee, Sean T. Yang, Jevin West, Bill Howe
    Document Analysis and Recognition (ICDAR) (2017)
39. **Scalable and Efficient Flow-Based Community Detection for Large-Scale Graph Analysis.**
    Seung-Hee Bae, Daniel Halperin, Jevin D. West, Martin Rosvall, Bill Howe
    ACM Transactions on Knowledge Discovery from Data 11(3) (TKDD) (2017)
40. **Voyager 2: Augmenting Visual Analysis with Partial View Specifications.**
    Kanit Wongsuphasawat, Zening Qu, Dominik Moritz, Riley Chang, Felix Ouk, Anushka Anand,
    Jock Mackinlay, Bill Howe, Jeffrey Heer
    ACM Human Factors in Computing Systems (CHI) (2017)
41. **The Myria Big Data Management and Analytics System and Cloud Services.**
    Jingjing Wang, Tobin Baker, Magdalena Balazinska, Daniel Halperin, Brandon Haynes, Bill Howe,
    Dylan Hutchison
    Conference for Innovative Data Research (CIDR) (2017)

42. **From NoSQL Accumulo to NewSQL Graphulo: Design and utility of graph algorithms inside a BigTable database**
    Dylan Hutchison, Jeremy Kepner, Vjay Gadepally, Bill Howe
    High Performance Extreme Computing Conference (HPEC) (2016)
43. **SQLShare: Results from a Multi-Year SQL-as-a-Service Experiment**
    Shrainik Jain, Dominik Moritz, Bill Howe, Ed Lazowska
    ACM International Conference on Management of Data (SIGMOD) (2016)
44. **Scalable clustering algorithms for continuous environmental flow cytometry**
    Jeremy Hyrkas, Sophie Clayton, Francois Ribalet, Daniel Halperin,
    E. Virginia Armbrust, Bill Howe
    Bioinformatics 32(3) (2016)
45. **Voyager: Exploratory analysis via faceted browsing of visualization recommendations**
    Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand,
    Jock Mackinlay, Bill Howe, Jeffrey Heer
    InfoVIS & IEEE Transactions on Visualization and Computer Graphics 22(1) (2016)
46. **MusicDB: Relational Approach for Numeric Longitudinal Music Analytics**
    Jeremy Hyrkas, Bill Howe
    International Society on Music Information Retrieval (ISMIR) (2016)
47. **Perfopticon: Visual Query Analysis for Distributed Databases**
    Dominik Moritz, Daniel Halperin, Bill Howe and Jeffrey Heer
    Proceedings of the 17th Eurographics Conference on Visualization (EuroVis) (2015)
48. **GossipMap: a distributed community detection algorithm for billion-edge directed graphs**
    Seung-Hee Bae, Bill Howe
    Supercomputing (2015)
49. **The database group at the University of Washington**
    Magdalena Balazinska, Bill Howe, and Dan Suciu
    ACM SIGMOD Record 43(1) (2014)
50. **Time-Varying Clusters in Large-Scale Flow Cytometry**
    Jeremy Hyrkas, Daniel Halperin, and Bill Howe
    Innovative Applications of Artificial Intelligence (IAAI) (2014)
51. **Helping scientists reconnect their datasets**
    Abdussalam Alawini, David Maier, Kristin Tufte, and Bill Howe
    ACM Scientific and Statistical Database Management Conference (SSDBM) (2014)
52. **Stop that query! The need for managing data use**
    Prasang Upadhyaya, Nick R. Anderson, Magdalena Balazinska,
    Bill Howe, Raghav Kaushik, Ravishankar Ramamurthy, and Dan Suciu
    Conference on Innovative Data Systems Research (CIDR) (2013)
53. **Hadoop's adolescence: an analysis of Hadoop usage in scientific workloads**
    Kai Ren, YongChul Kwon, Magdalena Balazinska, Bill Howe
    Proceedings of the VLDB Endowment 6(10) (PVLDB) (2013)
54. **Real-time Collaborative Analysis with (almost) Pure SQL: A Case Study in Biogeochemical Oceanography**
    Daniel Halperin, Konstantin Weitz, Bill Howe, Francois Ribalet, Mark A. Saito, and E. Virginia Armbrust
    ACM Scientific and Statistical Database Management Conference (SSDBM) (2013)
55. **SQLShare: Scientific Workflow via Relational View Sharing**
    Bill Howe, Francois Ribalet, Daniel Halperin, Sagar Chitnis, E Virgnia Armbrust
    Computing in Science & Engineering Special Issue on Science Data Management 15(2) (2013)

56. **VizDeck: Streamlining Exploratory Visual Analytics of Scientific Data**
    Daniel Perry, Bill Howe, Alicia M. F. Key, Cecilia Aragon
    iConference 2013 Proceedings (iConference) (2013) doi:10.9776/13206
57. **Toward Practical Query Pricing with QueryMarket**
    Paraschos Koutris, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe, and Dan Suciu
    ACM International Conference on Management of Data (SIGMOD) (2013)
58. **Query-based data pricing**
    Paraschos Koutris, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe, and Dan Suciu
    ACM Symposium on Principles of Database Systems (PODS) (2012)
59. **SkewTune: Mitigating Skew in Mapreduce Applications**
    YongChul Kwon, Magdalena Balazinska, Bill Howe, and Jerome A. Rolia
    ACM International Conference on Management of Data (SIGMOD) (2012)
    (currently the most-cited paper from SIGMOD 2012)
60. **Advancing Declarative Query in the Long Tail of Science**
    Bill Howe and Daniel Halperin
    IEEE Data Engineering Bulletin 35(3) (2012)
61. **The HaLoop approach to large-scale iterative data analysis**
    Yingyi Bu, Bill Howe, Magdalena Balazinska, and Michael Ernst
    VLDB Journal, 21(2) (2012)
    (Special Issue: Best of VLDB 2010)
62. **Optimizing Large-Scale Semi-Naive Datalog Evaluation in Hadoop**
    Marianne Shaw, Bill Howe, Paris Koutris, and Dan Suciu
    Datalog 2.0 (2012)
63. **VizDeck: A card game metaphor for fast visual data exploration**
    Bill Howe, Alicia Key, Daniel Perry, and Cecilia R. Aragon
    International Conference on Human Factors in Computing Systems (CHI) (2012)
64. **Towards efficient and precise queries over ten million asteroid trajectory models**
    Yusra AlSayyad, Simon K. Krughoff, Bill Howe, Andrew J Connolly, Magdalena Balazinska, and Lynne Jones
    Scientific and Statistical Database Management Conference (SSDBM) (2011)
65. **Data Markets in the Cloud: An Opportunity for the Database Community**
    Magdalena Balazinska, Bill Howe, and Dan Suciu
    Proceedings of the VLDB Endowment 4(12) (PVLDB) (2011)
66. **Parallel visualization on large clusters using MapReduce**
    Huy Vo, Brian Summa, Joao Comba, Juliana Freire, Bill Howe, Claudio Silva, and Valerio Pascucci
    IEEE Symposium on Large-Scale Data Analysis and Visualization (LDAV) (2011)
67. **Database-as-a-service for Long Tail Science**
    Bill Howe, Garret Cole, Emad Souroush, Paraschos Koutris,
    Alicia Key, Nodira Khoussainova, and Leilani Battle
    Scientific and Statistical Database Management Conference (SSDBM) (2011)
68. **COVE: A Visual Environment for Multidisciplinary Ocean Science Collaboration**
    Keith Grochow, Mark Stoermer, James Fogarty, Charlotte Lee, Bill Howe, and Ed Lazowska
    IEEE Sixth International Conference on eScience (eScience) (2011)
69. **Bioinformatics and data-intensive scientific discovery in the beginning of the 21st century**
    Roger Barga, Bill Howe, David Beck, Stuart Bowers, William Dobyns,
    Winston Haynes, Roger Higdon, Chris Howard, Christian Roth, Elizabeth Stewart
    Omics: A Journal of Integrative Biology, 15(4) (2011)

70. **Data Markets in the Cloud: An Opportunity for the Database Community**
    Magdalena Balazinska, Bill Howe, and Dan Suciu
    Proceedings of the VLDB Endowment PVLDB, 4(12) (2011)
71. **Astronomy in the Cloud: Using MapReduce for Image Coaddition**
    Keith Wiley, Andrew J. Connolly, Jeffrey P. Gardner, K. Simon Krughoff,
    Magdalena Balazinska, Bill Howe, YongChul Kwon, Yingyi Bu
    CoRR, abs/1010.1015. (2011)
72. **Scalable clustering algorithm for N-body simulations in a shared-nothing cluster**
    YongChul Kwon, Dylan Nunley, Jeffrey P Gardner, Magdalena Balazinska, Bill Howe, and Sarah Loebman
    Scientific and Statistical Database Management Conference (SSDBM) (2010)
73. **Skew-resistant parallel processing of feature-extracting scientific user-defined functions**
    YongChul Kwon, Magdalena Balazinska, Bill Howe, and Jerome Rolia
    Proceedings of the ACM Symposium on Cloud Computing (SOCC) (2010)
74. **HaLoop: Efficient iterative data processing on large clusters**
    YingYi Bu, Bill Howe, Magdalena Balazinska, and Michael Ernst
    International Conference on Very Large Databases (VLDB) (2010)
    (Selected for Best of VLDB Issue of VLDB Journal; currently most-cited paper from VLDB 2010)
75. **Client + cloud: Seamless architectures for visual data analytics in the ocean sciences**
    Keith Grochow, Bill Howe, Mark Stoermer, and Ed Lazowska
    Scientific and Statistical Database Management (SSDBM) (2010)
76. **Scientific Mashups: Runtime-Configurable Data Product Ensembles**
    Bill Howe, Harrison Green-Fishback, and David Maier
    Scientific and Statistical Database Management Conference (SSDBM) (2009)
77. **End-to-End eScience: Integrating Workflow, Query, Visualization, and Provenance at an Ocean Observatory**
    Bill Howe, Peter Lawson, Renee Bellinger, Erik Anderson, Emanuele Santos,
    Juliana Freire, Carlos Scheidegger, Antonio Baptista, and and Claudio T. Silva
    4th iEEE international Conference on eScience (eScience) (2008)
78. **Scientific mashups: Runtime-configurable data product ensembles**
    Harrison-Green Fishback and Bill Howe
    IEEE International Conference on e-Science (eScience) (2008)
79. **Scientific exploration in the era of ocean observatories** Antonio Baptista, Bill Howe, Juliana Freire, David Maier, and Claudio T. Silva
    Computing in Science and Engineering 10(3) (2008)
80. **Smoothing the ROI curve for scientific data management applications**
    Bill Howe, David Maier, and Laura Bright
    Conference on Innovative Data Systems Research (CIDR) (2007)
81. **Retrofitting a Data Model to Existing Environmental Data**
    Bill Howe and David Maier
    Scientific and Statistical Database Management Conference (SSDBM) (2005)
82. **Logical and Physical Data Independence for File-based Scientific Applications**
    Bill Howe and David Maier
    IEEE Data Eng. Bull. 27(4) (2004)
83. **Algebraic Manipulation of Scientific Datasets: Extended Results**
    Bill Howe and David Maier
    International Journal On Very Large Data Bases (VLDB Journal) 14(4) (2004)
    (Special Issue: Best of VLDB 2004)

84. **Algebraic Manipulation of Scientific Datasets**
    Bill Howe, and David Maier
    International Conference on Very Large Databases (VLDB) (2004)
    (Selected for Best of VLDB Issue of VLDB Journal)
85. **A language for spatial data manipulation**
    Bill Howe, David Maier, and Antonio Baptista
    Journal of Environmental Informatics 2(2) (2003)
86. **The Forest Portal: a Multidisciplinary Project**
    Marianne Koch, Lois Delcambre, Patricia Toccalino, Eric Landis, Fred Phillips, Tim Tolle, Len Shapiro, Nicole Steckler, David Maier, Mathew Weaver, Shawn Bowers, Balbinder Banga, Jason Brewster, Afrem Gutema, Sudarshan Murthy, Bill Howe, Rupa Tummala, Julia Norman, Kirsten Zillman, David Drake, Craig Palmer, Ashley Burt
    Proceedings of the 2003 annual national conference on Digital government research (dg.o) (2003)
87. **Representing, exploiting, and extracting metadata using metadata++**
    Mathew Weaver, Bill Howe, Lois Delcambre, Tim Tolle, and David Maier
    Proceedings of the 2002 annual national conference on Digital government research (dg.o) (2002)

Book Chapters

1. **Reproducibility, Virtual Appliances, and Cloud Computing**
    Bill Howe
    Chapter 10 in Victoria Stodden, Friedrich Leisch, Roger D. Peng (Ed.)
    Implementing Reproducible Research
    Chapman and Hall/CRC (2014)
2. **A Discussion on Pricing Relational Data**
    Magdalena Balazinska, Bill Howe, Parachos Koutris, Dan Suciu and Prasang Upadhyaya
    In Search of Elegance in the Theory and Practice of Computation (2013)
3. **Transforming Data into the Appropriate Science Context**
    Bill Howe
    Chapter 11 in Terence Critchlow, Kerstin Kleese van Dam (Ed.)
    Data-Intensive Science
    Chapman and Hall/CRC (2013)
4. **Beyond MapReduce: New Requirements for Scalable Data Processing**
    Bill Howe and Magda Balazinska
    Chapter 8 in Ian Gorton and Deborah K. Gracio (Ed.)
    Data-Intensive Computing, Cambridge University Press (2012)

Tech Reports

1. **Reliable, Routable, and Reproducible: Collection of Pedestrian Pathways at Statewide Scale**
    Ricky Zhang, Bill Howe, Anat Caspi
    arXiv preprint arXiv:2410.19762 (2024)
2. **Characterizing LLM Abstention Behavior in Science QA with Context Perturbations**
    Bingbing Wen, Bill Howe, Lucy Lu Wang
    arXiv preprint arXiv:2404.12452 (2024)
3. **Top-down Green-ups: Satellite Sensing and Deep Models to Predict Buffelgrass Phenology**
    Lucas Rosenblatt, Bin Han, Erin Posthumus, Theresa Crimmins, Bill Howe
    arXiv preprint arXiv:2310.00740 (2023)

4. **Adapting to Skew: Imputing Spatiotemporal Urban Data with 3D Partial Convolutions and Biased Masking**
   Bin Han, Bill Howe
   arXiv preprint arXiv:2301.04233 (2023)
5. **Urban Spatiotemporal Data Synthesis via Neural Disaggregation** Bin Han, Bill Howe
   arXiv preprint arXiv:2306.07292 (2023)
6. **Data Management for Causal Algorithmic Fairness**
   Babak Salimi, Bill Howe, Dan Suciu
   arXiv preprint arXiv:1908.07924 (2019)
7. **MultiDEC: Multi-Modal Clustering of Image-Caption Pairs**
   Sean T. Yang, Kuan-Hao Huang, Bill Howe
   arXiv preprint arXiv:1901.01860 (2019)
8. **In Defense of Synthetic Data**
   Luke Rodriguez, Bill Howe
   arXiv preprint arXiv:1905.01351 (2019)
9. **Delineating Knowledge Domains in the Scientific Literature Using Visual Information** Sean Yang, Poshen Lee, Jevin West, Bill Howe
   arXiv preprint arXiv:1908.07465 (2019)
10. **Query2Vec: NLP Meets Databases for Generalized Workload Analytics**
    Shrainik Jain, Bill Howe
    arXiv preprint arXiv:1801.05613 (2019)
11. **Radish: Compiling efficient query plans for distributed shared memory**
    Myers, Brandon, Daniel Halperin, Jacob Nelson, Mark Oskin, Luis Ceze, and Bill Howe
    Technical report, UW CSE Tech Report 14-10-01

Peer-Reviewed Workshop Papers

1. **Escaping the SpuriVerse: Can Large Vision-Language Models Generalize Beyond Seen Spurious Correlations?** Yiwei Yang, Chung Peng Lee, Shangbin Feng, Dora Zhao, Bingbing Wen, Anthony Zhe Liu, Yulia Tsvetkov, Bill Howe ICML 2025 Workshop on Reliable and Responsible Foundation Models (ICML R2-FM) (2025)
2. **Mitigating Overconfidence in Large Language Models: A Behavioral Lens on Confidence Estimation and Calibration**
   Bingbing Wen, Chenjun Xu, Bin HAN, Robert Wolfe, Lucy Lu Wang, Bill Howe
   NeurIPS 2024 Workshop on Behavioral ML (2024)
3. **Regularizing Model Gradients with Concepts to Improve Robustness to Spurious Correlations**
   Yiwei Yang, Andrew Liu, Robert Wolfe, Aylin Caliskan, Bill Howe
   ICML Workshop on Spurious correlations, Invariance, and Stability 2023
4. **Top-down Green-ups: Satellite Sensing and Deep Models to Predict Buffelgrass Phenology**
   Lucas Rosenblatt, Bin Han, Erin Posthumus, Theresa Crimmins, Bill Howe
   NeurIPS Workshop on Tackling Climate Change with Machine Learning: Blending New and Existing Knowledge Systems (2023)
5. **Aunt Lily Can Say Her Visualizations: Directing Analysis, Design, and Storytelling in Natural Language**
   Zening Qu, Fan Du, Ryan Rossi, Bill Howe
   VIS Workshops 2021

6. **The many facets of data equity**
   HV Jagadish, Julia Stoyanovich, Bill Howe
   EDBT/ICDT Workshops 2021
7. **MobilityMirror: Bias-Adjusted Transportation Datasets**
   Luke Rodriguez, Babak Salimi, Julia Stoyanovich, Bill Howe
   Workshop on Big Social Data and Urban Computing (BiDU) 2018
   (co-located with VLDB)
8. **EZLearn: Exploiting Organic Supervision in Large-Scale Data Annotation**
   Maxim Grechkin, Hoifung Poon, Bill Howe
   Learning with Limited Labeled Data: Weak Supervision and Beyond (LLD) (2017)
   (Best Paper award, co-located with NIPS 2017)
9. **Classifying digitized art type and time period**
   Sean Yang, Bum Mook Oh, Daniel Merchant, Bill Howe, Jevin West
   Workshop on Data Science for Digital Art History
   (co-located with KDD 2018)
10. **Synthetic Data for Social Good**
    Bill Howe, Julia Stoyanovich, Haoyue Ping, Bernease Herman, Matt Gee
    Bloomberg Data for Good Exchange (2017)
11. **Deep Mapping of the Visual Literature**
    Bill Howe, Po-shen Lee, Maxim Grechkin, Sean T. Yang, and Jevin D. West
    Workshop on Big Scholarly Data (BigScholar) (2017)
    (co-located with WWW)
12. **Profiling a GPU database implementation: a holistic view of GPU resource utilization on TPC-H queries.**
    Emily Furst, Mark Oskin, and Bill Howe
    International Workshop on Data Management on New Hardware (DaMON) (2017)
    (co-located with SIGMOD)
13. **LaraDB: A Minimalist Kernel for Linear and Relational Algebra Computation**
    Dylan Hutchison, Bill Howe, and Dan Suciu
    Workshop on Algorithms and Systems for MapReduce and Beyond (BeyondMR) (2017)
    doi: https://doi.org/10.1145/3070607.3070608
    (co-located with SIGMOD)
14. **Towards a general-purpose query language for visualization recommendation**
    Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand,
    Jock Mackinlay, Bill Howe, and Jeffrey Heer
    Workshop on Human-In-the-Loop Data Analytics (HILDA) (2016)
    doi: https://doi.org/10.1145/2939502.2939506
15. **High-variety cloud databases**
    Shrainik Jain, Dominik Moritz, Bill Howe
    Workshop on Quality in Databases (QDB) (2016)
    (co-located with VLDB)
16. **Towards a general-purpose query language for visualization recommendation**
    Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand, Jock Mackinlay, Bill Howe, Jeff Heer
    Workshop on Human-In-the-Loop Data Analytics (HILDA) (2016)
    (co-located with SIGMOD)
17. **High Variety Cloud Databases**
    Shrainik Jain, Dominik Moritz, Bill Howe
    Workshop on Cloud Data Management (CloudDM) (2016)

(invited keynote)
(co-located with ICDE)

18. **VizioMetrix: A Platform for Analyzing the Visual Information in Big Scholarly Data**
    Poshen Lee, Jevin West, Bill Howe
    Workshop on Big Scholarly Data (BigScholar) (2016)
    (co-located with WWW)
19. **Massive-Scale Cyber Traffic Analysis: A Driver for Graph Database Research**
    Cliff Joslyn, Sutanay Choudhury, David Haglin, Bill Howe, Bill Nickless, Bryan Olsen
    Workshop on Graph Data Management Experiences and Systems (GRADES) (2013)
    (co-located with SIGMOD)
20. **Scalable flow-based community detection for large-scale network analysis**
    Seung-Hee Bae, Daniel Halperin, Jevin West, Martin Rosvall, Bill Howe
    International Conference on Data Mining Workshops (ICDMW) (2013)
21. **Compiled plans for in-memory path-counting queries**
    Brandon Myers, Jeremy Hyrkas, Daniel Halperin, and Bill Howe
    International workshop on In-Memory Data Management and Analytics (IMDM) (2013)
    (co-located with VLDB)
22. **SQL is dead; long live SQL: Lightweight query services for ad hoc research data**
    Bill Howe and Garret Cole
    4th Microsoft eScience Workshop (2010)
    (co-located with eScience)
23. **Analyzing massive astrophysical datasets: Can Pig/Hadoop or a relational DBMS help?**
    Sarah Loebman, Dylan Nunley, YongChul Kwon, Bill Howe, Magdalena Balazinska, Jeffrey P Gardner
    IEEE International Conference on Cluster Computing (CLUSTER) (2009)
24. **Embracing Uncertainty in Large-Scale Computational Astrophysics**
    Dan Suciu, Andrew J Connolly, and Bill Howe
    Workshop on Management of Uncertain Data (MUD) (2009)
    (co-located with VLDB)
25. **Query-driven visualization in the cloud with mapReduce**
    Bill Howe, Huy Vo, Claudio Silva, and Juliana Freire
    Workshop on Ultrascale Visualization (2009)
    (co-located with Supercomputing)
26. **Quarrying dataspaces: Schemaless profiling of unfamiliar information sources**
    Bill Howe, David Maier, Nicolas Rayner, and James Rucker
    Workshop on Information Integration Methods, Architectures, and Systems (IIMAS) (2008)
    (co-located with ICDE)
27. **The eScience Appliance: Provisioning an Inexpensive Bottom-Up Cyberinfrastructure**
    Bill Howe, and Roger Barga
    Microsoft eScience workshop (2008) (co-located with eScience)
28. **Managing the forecast factory**
    Laura Bright, David Maier, and Bill Howe
    Workshop on Workflow and Data Flow for Scientific Applications (SciFlow) (2006)
29. **Emergent semantics: Towards self-organizing scienitifc metadata**
    Bill Howe, Kuldeep Tanna, Paul Turner, and David Maier
    Workshop on Semantics for a Networked World: Semantics for Grid Databases (2004)
    (co-located with VLDB)

30. **Modeling data product generation**
    Bill Howe and David Maier
    Workshop on Data Derivation and Provenance (2002)

Peer-Reviewed Demonstration Papers

1. **Mithralabel: Flexible dataset nutritional labels for responsible data science**
   Chenkai Sun, Abolfazl Asudeh, HV Jagadish, Bill Howe, Julia Stoyanovich
   Proceedings of the 28th ACM International Conference on Information and Knowledge Management (SIGMOD) (2019)
2. **GraviTIE: Exploratory Analysis of Large-Scale Heterogeneous Image Collections**
   Sean T. Yang, Luke Rodriguez, Jevin D. West, Bill Howe
   The World Wide Web Conference (2019)
3. **A Nutritional Label for Rankings**
   Ke Yang, Julia Stoyanovich, Abolfazl Asudeh, Bill Howe, H.V. Jagadish, Gerome Miklau
   ACM Conference on Management of Data (SIGMOD) (2018)
4. **DataSynthesizer: Privacy-Preserving Synthetic Datasets**
   Haoyue Ping, Julia Stoyanovich, and Bill Howe
   ACM Scientific and Statistical Database Management Conference (SSDBM) (2017)
5. **Demonstration of the Myria big data management service**
   Daniel Halperin, Victor Teixeira de Almeida, Lee Lee Choo, Shumo Chu,
   Parachos Koutris, Dominik Moritz, Jennifer Ortiz, Vaspol Ruamviboonsuk, Jingjing Wang,
   Jingjing and Andrew Whitaker, Shengliang Xu, Magdalena Balazinska, Bill Howe, Dan Suciu,
   ACM Conference on Management of Data (SIGMOD) (2014)
6. **The power of data use management in action**
   Prasang Upadhyaya, Nick R. Anderson, Magdalena Balazinska,
   Bill Howe, Raghav Kaushik, Ravishankar Ramamurthy, and Dan Suciu *ACM International Conference on Management of Data* (SIGMOD) (2013)
7. **VizDeck: Self-organizing dashboards for visual analytics**
   Alicia Key, Bill Howe, Daniel Perry, and Cecilia R. Aragon
   ACM Conference on Management of Data (SIGMOD) (2012)
8. **Automatic Example Queries for Ad Hoc Databases**
   Bill Howe, Garret Cole, Nodira Khoussainova, Leilani Battle
   ACM Conference on Management of Data (SIGMOD) (2011)
9. **Querying and Visualizing Gridded Datasets for eScience**
   Bill Howe, and David Maier
   International conference on Data Engineering (ICDE) (2005)

Peer-Reviewed Invited Papers and Extended Abstracts

1. **Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy**
   Lucas Rosenblatt, Bernease Herman, Anastasia Holovenko, Wonkwon Lee, Joshua Loftus, Elizabeth McKinnie, Taras Rumezhak, Andrii Stadnik, Bill Howe, Julia Stoyanovich
   ACM SIGMOD Record, Volume 53, Issue 1, Pages 65âĂŞ74 (2024)

2. **Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy**
   Lucas Rosenblatt, Bernease Herman, Anastasia Holovenko, Wonkwon Lee, Joshua Loftus, Elizabeth McKinnie, Taras Rumezhak, Andrii Stadnik, Bill Howe, Julia Stoyanovich
   ACM SIGMOD Record 53 (1), 65-74 (2024)

3. **Does a Fair Model Produce Fair Explanations? Relating Distributive and Procedural Fairness**
   Yiwei Yang, Bill Howe
   Hawaii International Conference on System Sciences (HICSS) (2024)
4. **Urban Spatiotemporal Data Synthesis via Neural Disaggregation**
   Bin Han, Bill Howe
   Hawaii International Conference on System Sciences (2024)
5. **Technical perspective: Visualization search: from sketching to natural language**
   Bill Howe
   Communications of the ACM 65 (7) (2022) (invited paper review)
6. **Technical Perspective: Imperative or Functional Control Flow Handling: Why not the Best of Both Worlds?**
   Bill Howe
   ACM SIGMOD Record 51 (1), 59-59 (2022) (invited paper review)
7. **Technical Perspective: From Sketching to Natural Language: Expressive Visual Querying for Accelerating Insight**
   Bill Howe
   SIGMOD Record 2021 (invited paper review)
8. **Digital Government: Research and Practice**
   Julia Stoyanovich, Bill Howe, HV Jagadish
   Digital Government: Research and Practice (2020)
9. **Database Repair Meets Algorithmic Fairness**
   Babak Salimi, Bill Howe, Dan Suciu
   ACM SIGMOD Record 49 (1), 34-41 (2020)
10. **Fairness in Practice: A Survey on Equity in Urban Mobility**
    An Yan, Bill Howe
    IEEE Data Engineering Bulletin 42(3) (2019)
11. **Nutritional Labels for Data and Models**
    Julia Stoyanovich, Bill Howe
    IEEE Data Engineering Bulletin 42(3) (2019)
12. **Data Management for Causal Algorithmic Fairness**
    Babak Salimi, Bill Howe, Dan Suciu
    IEEE Data Engineering Bulletin 42(3) (2019)
13. **The Principles of Tomorrow's University**
    Daniel S. Katz, Gabrielle Allen1, Lorena A. Barba, Devin R. Berg, Holly Bik, Carl Boettiger, Christine L. Borgman, C. Titus Brown, Stuart Buck, Randy Burd, Anita de Waard, Martin Paul Eve, Brian E. Granger, Josh Greenberg, Adina Howe, Bill Howe, May Khanna, Timothy L. Killeen, Matthew Mayernik, Erin McKiernan, Chris Mentzel, Nirav Merchant, Kyle E. Niemeyer, Laura Noren, Sarah M. Nusser, Daniel A. Reed, Edward Seidel, MacKenzie Smith, Jeffrey R. Spies, Matt Turk, John D. Van Horn, Jay Walsh
    F1000 Research (2019)
14. **Special Session: A Technical Research Agenda in Data Ethics and Responsible Data Management**
    Julia Stoyanovich, Bill Howe, H.V. Jagadish
    Panel event, SIGMOD (2018)
15. **Lightweight Data Systems in the Cloud: Costs, Benefits and Best Practices**
    Rob Fatland, Anthony A Arendt, Bill Howe, Nancy J Hess, Joe Futrelle
    American Geophysical Union, 2015

16. **SQLShare: DB-as-a-Service for Science**
    Bill Howe
    American Geophysical Union Fall Meeting, 2013
17. **Hadoop's Adolescence; A Comparative Workloads Analysis from Three Research Clusters**
    Kai Ren, Garth Gibson, YongChul Kwon, Magda Balazinska and Bill Howe
    High Performance Computing, Networking, Storage and Analysis (SCC) 2012
18. **Efficient Prediction of Asteroid Positions from Solar System Models**
    Yusra AlSayyad, Simon Krughoff, Andrew J. Connolly, L. Jones, Tomas Budavari, and Bill Howe
    Bulletin of the American Astronomical Society, 43 (2011)
19. **The Ocean Appliance: Complete Platform Provisioning for Low-Cost Data Sharing**
    Bill Howe and Nicholas Hagerty and Ethan Van Matre and David Maier and Antonio Baptista and Charles Seaton, Paul Turner
    IEEE OCEANS 2007 (OCEANS) (2007)

Research Grants

1. Co-PI, **USDOT** Transportation Data Equity Initiative, 2021-2025, $11,500,000, with PI Anat Caspi (UW)
2. PI, **Cisco** Ethical AI in the Public Sector: Towards A Semi-Synthetic Data Fabric for AI Evaluation, 2021-2023, $200,000
3. Co-PI, **NSF Grant** FIDES: Framework for Integrative Data Equity Systems, 2019-2021, $2,000,000, with Julia Stoyanovich (NYU), H. V. Jagadish (University of Michigan)
4. PI, **NSF Grant** FORDS: Workshop on Foundations of Responsible Data Science, 2019-2020, $100k
5. PI, **Elsevier** Large-Scale Privacy-Preserving Synthetic Datasets, 2019-2021, $137k
6. Co-PI, **NSF Grant** BIGDATA: F: Collaborative Research: Foundations of Responsible Data Management, 2017-2021, $1,500,000, with Julia Stoyanovich (Drexel), Gerome Miklau (UMass Amherst), H. V. Jagadish (University of Michigan)
7. Co-PI **Microsoft** Transportation Data Collaborative, $150k, with Jan Whittington (UW Urban Planning) and Mark Hallenbeck (TRAC)
8. Lead PI, **Bill and Melinda Gates Foundation** Advance Data Analytic Support for Strategic PNW Partners, A126396, 2017-2022, $750,000, with Ariel Rokem (UW) and Bryna Hazelton (UW)
9. Lead PI, **Microsoft** Cascadia Urban Analytics Collaborative, 2017-2019, $500,000
10. Lead PI, **Department of Defense and PNNL** Polystore Data System, 2015-2016, $500,000
11. Lead PI, **Macarthur Foundation** Advancing the Seattle MetroLab, 2016-2017, $250,000
12. Co-PI **NSF Big Data Hub, Western Region**, 2015-2018, $1,200,000, with Ed Lazowska (UW), Ariel Rokem (UW), Mike Franklin (UC Berkeley), and Mike Norman (SDSC)
13. Lead PI, **Department of Defense and PNNL** Unified Services for Federated Big Data Environment, 2015-2016, $500,000
14. Lead PI, **Macarthur Foundation** Towards a Seattle Metro Consortium for Urban Data Science, 2014-2016, $300,000
15. Co-PI **UW Provost's Office** Urban@UW, 2015-2017, $750,000, with Thaisa Way (UW), Vikram Jandalaya (UW), Thom Dunning (PNNL)
16. Co-PI, **Gordon and Betty Moore Foundation and the Alfred P. Sloan Foundation**, Data Science Environments, 2014-2019, $32.8M, with NYU and UC Berkeley
17. Co-PI, **Intel Science and Technology Center for Big Data**, 2014-2016, $125,000, one part of large distributed group across MIT, Brown, UW, and Portland State
18. Lead PI, **Department of Defense and Pacific Northwest National Labs** Hybrid Architectures and Languages for Graph Query, 2013-2015, $300,000

19. Co-PI, **NSF Grant** IIS-1247469, BIGDATA:A Formal Foundation for Big Data Management, 2013-2016, $2,966,667, with Dan Suciu (UW) and Magda Balazinska (UW)
20. Co-PI, **NSF Grant** ACI-1216879, Collaborative Research: Conceptualizing An Institute for Empowering Long Tail Research, 2012-2014, $500,000, with Ian Foster (UChicago), Bryan Heidorn (UArizona), Christine Borgman (UCLA), Carl Kesselman (USC)
21. Lead PI, **NSF Grant** IIS-1064505, Collaborative Research: Database-As-A-Service for Long Tail Science 2011-2013, $800,000, with Mike Cafarella (UMich), David Maier (Portland State), and Dan Suciu (UW).
22. Co-PI, **EMC Grant** Large-Scale Visual Analytics, 2012-2014, $200,000, with Magda Balazinska (UW)
23. Co-PI, **NSF Grant** CCF-1047815, Relational Data Markets in the Cloud, 2011-2013, $370,000, with Magda Balazinska (UW) and Dan Suciu (UW)
24. Lead PI, **NSF Grant** CCF-1060213, EAGER: Scalable Algebraic Visualization in the Cloud, 2010-2011 $117,567
25. Co-Lead, **Gordon and Betty Moore Foundation**, Data-Intensive Science in the Long Tail, 2010-2013, $750,000, with Ed Lazowska
26. Lead PI, **NSF Grant** IIS-0844572, Collaborative Research: Where the Ocean Meets the Cloud: Ad Hoc Longitudinal Analysis and Collaboration Over Massive Mesh Data, 2009-2011, $500,000, with Claudio Silva (NYU) and Juliana Freire (NYU).

Student and Postdoc Advising

Doctoral Students

Robert Wolfe, Information School (with Alexis Hiniker and Tanu Mitra)	**2022–present**
Isaac Slaughter, Information School	**2023–present**
Bingbing Wen, Information School	**2022–present**
Bernease Herman, Information School	**2021–present**
Eva Maxfield Brown, Information School (with Nic Weber)	2020–present
Yiwei Wang, Information School	**2020–present**
Bin Han, Information School	**2020–present**
Zening Qu, Information School	**2019–present**
Sean Yang, Electrical Engineering *now Research Scientist at Yahoo Research*	2016–2022
An Yan, Information School *now Research Engineer at Facebook, on the ML team*	2018–2020
Luke Rodriguez, Information School (Masters) *now at Microsoft*	2017–2019
Shrainik Jain, Allen School of Computer Science & Eng. *now at Snowflake*	2014–2019
Shana Hutchison, Allen School of Computer Science & Eng. *now at SpaceX*	2016–2019
Dominik Moritz, Allen School of Computer Science & Eng. (with Jeff Heer) *now at CMU*	2016–2019
Maxim Grechkin, Allen School of Computer Science & Eng. *now Research Scientist at Facebook, on their Misinformation team*	

Kanit Wongsuphasawat, Allen School of Computer Science & Eng. (with Jeff Heer) 2015–2018
now Researcher at Apple

Brandon Myers, Computer Science & Engineering (with Mark Oskin) 2013–2016
now faculty at University of Iowa

Poshen Lee, Electrical Engineering 2014–2017
now Data Scientist at ExtraHop

Jeremy Hyrkas, Computer Science & Engineering 2013–2016
now at Microsoft

Yongchul Kwon, Computer Science & Engineering (with Magda Balazinska) 2012–2015
now at Microsoft

Yingyi Bu, Computer Science & Engineering (with Magda Balazinska) 2011–2013
moved to UC Irvine

Postdoctoral Fellows and Research Scientists

- **Sophie Clayton**, Oceanography (with Ginger Armbrust) 2016–2018
- **Dan Halperin**, Computer Science & Engineering 2013–2016
  now at Google
- **Andrew Whitaker**, Computer Science & Engineering 2012–2015
  now at Amazon
- **Seung-Hee Bae**, Computer Science & Engineering 2012–2015
  now faculty at Western Michigan University
- **Marianne Shaw**, Computer Science & Engineering 2013–2016
  now at Tableau

Teaching

- IMT 598 Epistemological Foundations of AI, Spring 2024
- INFO 430 Advanced Database Systems, Winter 2024
- Course Refresh/Designer: IMT 575 Data Science III: Scaling, Applications, and Ethics, Fall 2023 (Added Module on Generative AI and LLMs, replaced all assignments
- Course Lead/Advisor: IMT 575 Data Science III: Scaling, Applications, and Ethics, Winter/Spring 2023 (on sabbatical in 2023)
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Spring 2022
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Winter 2022
- INFO 330 Relational Database Management Systems, Fall 2021
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Spring 2021
- INFO 340 Relational Database Management Systems, Fall 2020
- INFO 330 Relational Database Management Systems, Fall 2020
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Spring 2020
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Spring 2019
- INFO 340 Relational Database Management Systems, Fall 2019
- IMT 563 Advanced Relational and Non-Relational Databases, Fall 2018
- IMT 575 Data Science III: Scaling, Applications, and Ethics, Fall 2018
- INFO 340 Relational Database Management Systems, Spring 2018

- IMT 575 Data Science III: Scaling, Applications, and Ethics, Spring 2018
- INFX 575 Data Science III: Scaling, Applications, and Ethics, Fall 2017
- INFX 575 Data Science III: Scaling, Applications, and Ethics, Spring 2017
- *Introduction to Data Science* (Coursera), Summer 2014 (100,000+ students)
- *Introduction to Data Science* (Coursera), Spring 2013 (100,000+ students)
- *Introduction to Data Programming* (with Mike Ernst), Winter 2013.
- *Introduction to Data Science* (PCE Certificate), Fall 2012.
- *CSE 190p Introduction to Data Programming with Applications*, Summer 2012.
- *Data-Intensive Computing in the Cloud* (PCE Certificate), Fall 2012.
- *CSE599c Scientific Data Management*, Winter 2011.
- *CSE344 Introduction to Data Management*, Fall 2010

Press

- *Techxplore*, April 2022, "Using AI to expand the quality and fairness of urban data" **research feature**
- *Seattle Met*, September 2021, "Could Coders Automate Away Their Own Work?," Stefan Milne, **quoted**
- *Government Technology*, July 2021 ,"Project Aims to Better Use Municipal Open Data, Boost Equity," Josh Schacht **research feature**
- *Networked World*, June 2021, "AI tackles data-center workload management," John Edwards, **quoted**
- *Slate*, September 2020, "Fake Data Could Help Solve Machine LearningâĂŹs Bias ProblemâĂŤif We Let It," Todd Feathers, **quoted**
- *GeekWire*, August 2019, "How tech keeps SeattleâĂŹs transit system running âĂŤ and why more innovation could be coming," Aria Thaker, **quoted**
- *Seattle Times*, March 2019, "AmazonâĂŹs role in co-sponsoring research on fairness in AI draws mixed reaction," Benjamin Romano, **quoted**
- *UW News*, January 2019, "Seattle bike share programs show infrequent helmet use, little disparity in access to bikes among neighborhoods," Jackson Holtz, **research feature**
- *Seattle Times*, "Opinion: Protect the public from bias in automated decision systems", **invited piece**
- *King 5 KUOW*, April, 2018"UW students learn lessons from Facebook data breach," *television interview*, on privacy and data reuse related to Cambridge Analytica
- *GovTech.com*, November 2017, "University Researchers use 'Fake' data for social good," Ben Levine, **research feature**, with Julia Stoyanovich.
- *NPR*, September 2017, "Future City," Wes Moore, **interview** on regional-scale smart cities efforts
- *Geekwire*, September 2017, "How to detect fake news: UW professor teaching 'Calling Bulls**t' class offers tips to spot misinformation," Taylor Soper, piece on panel I chaired with participants Ryan Calo, Jevin West, Kate Starbird
- *Geekwire*, August 2017, "Why Seattle is poised to be a leader in 'smart city' technology and regulations," Taylor Soper, **quoted** heavily on Seattle's smart cities initiatives.
- *King5 News*, August 2017, "One-third of Seattle drivers 'cruising' for parking, rides, study finds," local television piece on the results from one of our student team's Data Science for Social Good project.

- *Seattle Times*, August 2017, "Getting a job in big data," Lizz Giordano, **interview** on data science training and careers
- *Nature News*, June 2017, "Text-mining tool seeks out hidden data," Dalmeet Singh Chawla, **research feature**
- *Phys.org*, June 2017, "Wide-Open accelerates release of scientific data by identifying overdue datasets," **research feature**
- *Xinhua News Agency*, June 2017, "Researchers develop new tool to help advance open science," **research feature**
- *USA Today*, May 2017, "Here's what you need to land America's best jobs," Marco della Cava and Eli Blumenthal, **interview**
- *TechCrunch*, August 2016, "Student projects leapfrog governments and industry in 'Data Science for Social Good' program," Devin Coldewey, piece on the Data Science for Social Good program I founded
- *Crosscut.com*, February 2017, "Could Data Solve Seattle's Biggest Problems?," Samantha Larson, **quoted**, Urbanalytics group and Cascadia Urban Analytics Cooperative described
- *The Economist*, June 2016, "A scientific study of the importance of diagrams to science," **research feature** on viziometrics project
- *MIT Review*, May 2016, "The First Visual Search Engine for Scientific Diagrams," Emerging Technology from the arXiv, **research feature**
- *GovTech.com*, March 2016, "'Array of Things' Expands to Cities with Research Partnerships," Ben Miller, **quoted** heavily on our Seattle Array of Things project.
- *GeekWire*, December 2015, "University of Washington adds data science masterâĂŹs program to meet demand in job market," James Risley, **quoted** as program founder
- *International Business Times*, June 2015, "Big Data Helps Identify Potential Link Between Common Heartburn Medicines And Risk Of Heart Attack," Amy Nordrum, **quoted** as external data expert
- *Dataconomy.com*, September 2014, "10 Online Data Science Courses," my Coursera data science course included in the list.
- *Forbes*, May 2013, "What MOOCs Can Offer Social Entrepreneurs," Coursera data science course described.
- *Sciencemag.org*, May 2013, "When all Science Becomes Data Science," Vijaysree VenkatramanMay, **quoted** on data science careers (article title adapted from my quote).
- *New York Times*, April 2013, "Data Science: The Numbers of Our Lives," Claire Cain Miller, **quoted** on data science education.
- *Science*, May, 2013, "When All Science Becomes Data Science," Vijaysree Venkatraman **quoted**
- *New York Times*, April 2013, "Geek Appeal: New York vs. Seattle," Claire Cain Miller, **quoted**
- *King 5 News*, September 2011, Television **interview** on cloud computing.
- *New York Times*, November 2009, "Supercomputing for the Masses," Ashlee Vance, **quoted** on cloud computing
- *ReadWrite.com*, April 2009, "Web as Platform For Research on Oceans, Galaxies," Richard Macmanus, **quoted** on data-intensive science

Honors and Awards

1. **Information Schools Strategic Research Initiative Awardee**, "Laboratory-scale AI: Rigorous Evaluation and Reliable Deployment of Small Trainable Adaptors for Generative AI"
2. **Best Paper Runner Up, Experiment, Analysis & Benchmark Track, VLDB 2023**, "Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy"
3. **Selected for Research Highlight article, SIGMOD Record, 2023**, "Epistemic Parity: Reproducibility as an Evaluation Metric for Differential Privacy"
4. **Best Paper, SIGMOD 2019**, "Interventional fairness: Causal database repair for algorithmic fairness"
5. **Best Paper, InfoVis 2019**, "Formalizing Visualization Design Knowledge as Constraints: Actionable and Extensible Models in Draco"
6. **Best Paper, LLDM Workshop 2017** "EZLearn: Exploiting Organic Supervision in Large-Scale Data Annotation," Learning with Limited Labeled Data Workshop (co-located with NeurIPS) 2017
7. **Best of VLDB 2010**, selected for publication in VLDB Journal special issue, for "Haloop: Efficient iterative data processing on large clusters." (currently the most cited paper from VLDB 2010)
8. **Microsoft Jim Gray Seed Grant** 2010, $25,000
9. **Microsoft Jim Gray Seed Grant** 2007, $40,000
10. **Department Dissertation Award**, Maseeh College of Engineering and Computer Science, Portland State University, May 2007
11. **Best of VLDB 2004**, selected for publication in VLDB Journal special issue, for "Algebraic manipulation of scientific datasets"

Selected Talks and Panels

1. Invited Talk, 'How is Big Data Constructed, Where is it Housed, and How is it Accessed and Analyzed?," National Courts and Sciences Institute (NCSI) Data Science & Artificial Intelligence Initiative, online seminar "Demystifying Data Science and Artificial Intelligence Technologies"
2. Keynote, "Data Curation as Programming," Alberto Mendelzon International Workshop on Foundations of Data Management, Peru, April 2023
3. Invited Talk, "Applied AI in High-Expertise Settings, or Curation as Programming," Engineering IDBE Seminar Series, NYU Abu Dhabi, January 2023
4. Invited Talk, "Applied AI in High-Expertise Settings, or Curation as Programming," AI2, July 2022
5. Invited Talk, "Responsible AI Systems and Experiences," Discovering AI@UW, May 2022
6. Invited Talk, "Ethical AI in the Public Sector: Towards A Semi-Synthetic Data Fabric for AI Evaluation" Cisco, January 2022
7. Invited Talk, "Data-Centric AI: Reuse, Integration, and Synthesis of Weakly Structured Data," Northeastern, December 2021
8. Keynote, "Introspection and Inerventions in Data Equity Systems," Provenance and Visualisation workshop, 2021
9. Invited talk, "Equitensors: Learning Fair Integration of Urban Mobility Data," Berkeley Institute for Transportation Studies, February 2021
10. Keynote, "Data Equity Systems," Northwest Database Society, 2021
11. Keynote, "Data Equity Systems," DEEM Workshop, 2020

12. Invited Talk, “Public interest research in data management and machine learning,” NYU, February 2020
13. Distinguished Colloquium, “The Next Decade of Data Science,” University of Maryland, College Park, December 2017
14. Invited Talk, “Epistemic Issues in Data Science,” University of Massachusetts, Amherst, December 2017
15. Invited Speaker, Data Analysis and Visualization Workshop, Dagstuhl, Germany, 2017
16. Keynote, ”Responsible Urban Data Science,” UrbanGIS Workshop, Redondo Beach, CA, November 2017
17. Invited Talk, “Viziometrics: Mining the Visual Literature,” Scientific Computing Institute, University of Utah, November 2017
18. Invited Talk, “Deep Curation,” New York University, 2017
19. Invited Talk, “Fake Data for Social Good,” Bloomberg Data for Good Exchange, 2017
20. Panelist, Bias and Ethics in City Services Data Science, Bloomberg Data for Good Exchange, 2017
21. Invited Talk, “Big Data + Big Sim: Query Processing over Unstructured CFD Models,” ISIM Workshop, Durham, UK
22. Invited Talk, “Data, Responsibly: The Next Decade of Data Science,” iSchool Founding Board, University of Washington, 2017
23. Invited Speaker, Data Responsibly Workshop, Dagstuhl, Germany, 2016
24. Invited Talk, “Data Science, Data Curation, and Human-Data Interaction,” Data Science Summit, San Francisco, 2016
25. Keynote, Cloud Data Management Workshop (with ICDE), 2016, Helsinki
26. Invited Talk, “Urban Data Science @ UWâĂİ, NIAC Workshop on Smart Cities, 2016
27. Invited Talk, “Big Data Talent in Academic and Industry R&D,” 2015
28. Panelist, “Data Science Education,” NSF BigData PI Meeting, 2017
29. Invited Talk, “High Productivity ComputingâĂİ, Los Alamos National Lab, November 2015
30. Invited Talk, “High Productivity ComputingâĂİ, Chesapeake Large-Scale analytics Conference, October 2015
31. Invited Talk, “High Productivity ComputingâĂİ, Neurofutures Symposium, Portland, August 2015
32. Invited Speaker, “Lightweight Data Systems in the Cloud: Costs, Benefits and Best Practices,” American Geophysical Union, 2015
33. Invited Talk, ITIC and Macarthur workshop “Partnership for Sustainable Smart Cities of the Future”
34. Invited Talk, National Research Council Meeting on “Training the Big Data Workforce”
35. Keynote at eXtremely Large Databases, South America, 2014
36. Invited Talk at JSM panel on Big Data Education: Big Data Curricula at the UW eScience Institute.
37. Invited Talk on Myria and scalable graph clustering at the Mining of Massive Datasets workshop at UC Berkeley, 2014.
38. Invited Speaker, Cloud Computing Workshop, Qatar Computing Research Institute, Qatar, 2014
39. Keynote at eResearch New Zealand, “eScience and Data science at the University of Washington’s eScience Institute,” 2013
40. Invited Talk, “Myria: Analytics-as-a-Service for (Data) Scientists,” High Performance Transaction Processing, 2013
41. Invited Panelist, “Big Data Curricula at the UW,” Joint Statistical Meeting, 2013

42. Invited Talk, "Enabling Collaborative Research Data Management with SQLShare," Internet2 Symposium, 2013
43. Invited Talk, "Conservative Regridding and Algebraic Manipulation of Meshes with GridFields," HPEC Conference, 2012
44. Invited Talk, "Democratizing Data Science," UC Berkeley, 2012
45. Invited Talk, "Iterative Computation for Big Data," University of Indiana, 2012
46. Invited Talk, "Virtual Appliances, Cloud Computing, and Reproducible Research," AMP Workshop on Reproducible Research, 2011
47. Invited Talk, "A New Partnership for Cross-Scale, Cross-Domain eScience," Carnegie Mellon University, 2010
48. Invited Talk, "End-to-End eScience: Integrating Query, Workflow, and Visualization at an Ocean Observatory," Microsoft Research, 2009

Professional Service

- Associate Editor, SIGMOD Conference Program Committee, 2023-2024
- Program Committee, Data Education Workshop, co-located with SIGMOD, 2023
- Program Committee, ACM Conference on Fairness, Accountability, and Transparency (FAccT), 2023
- Steering Committee, ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2019-present
- Review Process Working Group, ACM Conference on Fairness, Accountability, and Transparency (FAccT) 2022-2023
- Program Committee, ACM International Conference on Management of Data, (SIGMOD) 2022
- Associate Editor, ACM Transactions on Data Science (TDS), 2019-present
- Metrolab Data Governance Task Force, Privacy, AI, and Governance working groups, 2022-2023
- Reviewer, Machine Learning and Knowledge Extraction, 2021
- Sponsorship Chair, ACM International Conference on Management of Data, (SIGMOD) 2020, 2021
- Steering Committee, Theory and Practice of Provenance, 2017-2022
- National Academy of Sciences Roundtable on Postsecondary Data Science Education, 2016 - 2019
- Program Committee, International Conference on Data Engineering (ICDE), 2019
- Core Program / Area Chair, ACM International Conference on Management of Data, (SIGMOD) 2019
- Program Committee, BigVis Workshop, 2018, (co-located with EDBT)
- Program Committee, International Conference on Data Engineering (ICDE), 2018
- General Chair, TAPP 2017
- Area Chair, ACM International Conference on Management of Data, (SIGMOD) 2018
- Program Committee, ACM International Conference on Management of Data, (SIGMOD) 2017
- Program Chair, eScience 2016
- Analysis in Motion Advisory Board, Pacific Northwest National Labs, 2015 - 2017
- Program Committee, HotCloud 2014
- Program Committee, eScience 2014
- Program Committee, International Conference on Scientific and Statistical Databases (SSDBM), 2013
- Program Committee, Workshop on Large-Scale Data Analysis and Visualization, 2013
- Chair, Workshop on HPC meets Databases, co-located with Supercomputing 2012

- Program Committee, International Conference on Data Engineering (ICDE), 2013
- Program Committee, Proceedings of Very Large Databases, 2012-2013
- Demo Co-chair, International Conference on Scientific and Statistical Databases (SSDBM), 2012
- Program Committee, ScienceCloud 2012
- Chair, Workshop on HPC meets Databases, co-located with Supercomputing 2011
- Editorial Board, Journal of Data Semantics
- Organizing Committee, Workshop on eXtremely Large Databases, 2011
- Program Committee, Workshop on Large-Scale Data Analysis and Visualization, 2011
- Program Committee, ScienceCloud 2011
- Co-Chair, Workshop on Array Databases
- Registration Chair, International Conference on Scientific and Statistical Databases (SSDBM), 2011
- Program Committee, International Conference on Scientific and Statistical Databases (SSDBM), 2011
- Demonstrations Program Committee, ACM International Conference on Management of Data, (SIGMOD), 2011
- Program Committee, International Conference on Extending Database Technology (EDBT), 2010
- Program Committee, International Conference on Scientific and Statistical Databases (SSDBM), 2010
- Program Committee, Workshop on Information Integration Methods, Architectures, and Systems (IIMAS), 2008
- Reviewer, VLDB Journal, 2007
- Program Committee, dg.o 2006
- Program Committee, dg.o 2005
- Demonstrations Program Committee, ACM International Conference on Management of Data, 2005
- Student Session Program Committee, dg.o 2004
- Referee

  Science Advances (Open Access expansion of Science Magazine), *Journal of Computer Science and Technology*, *Computing in Science & Engineering*, *Academic Journal 5*, *Academic Journal 6*, *Academic Journal 7*

Staff Supervision

- **Emily Keller**, Program Manager, Urbanalytics 2017–2022
- **Sarah Stone**, Ph.D. Oceanography, Executive Director, eScience 2012–2017
- **Micaela Parker**, Ph.D. Oceanography, Executive Director, eScience 2012–2017
- **Rob Fatland**, Ph.D. Physics, Director of Cloud and Data Solutions 2016–2017
- **Jacob Vanderplas**, Ph.D. Astronomy, eScience Data Scientist 2012–2017
- **Ariel Rokem**, Ph.D. Neuroscience, eScience Data Scientist 2012–2017
- **Bernease Herman**, eScience Data Scientist 2012–2017
- **Valentina Staneva**, Ph.D. Applied Math, eScience Data Scientist 2012–2017

Diversity, Equity, Inclusion

- Co-PI and Lead Engineer for $11.5M USDOT grant on Data Equity for Urban Mobility emphasizing disability equity
- Multiple NSF Grants on Data Equity Systems and Responsible Data Science
- Developed Diversity Sponsorship Program at SIGMOD 2021 (as Sponsorship Chair)
- Revised Curricula for Data Science and Database courses to include material on technology reinforcing systemic discrimination

National Leadership

- Chair, Workshop on Foundations of Responsible Data Science, 2020
- Member, National Academy of Sciences Roundtable on Data Science Education, 2016-2019
- Co-Organizer, *Special Session: A Technical Research Agenda in Data Ethics and Responsible Data Management* , Julia Stoyanovich, Bill Howe, H.V. Jagadish, 2018
- Co-PI, NSF Western Big Data Hub, 2016–present
- Co-Organizer, NSF Translational Data Science Workshop, Chicago, 2017
- Chair, *Data Science Education: We're Missing the Boat, Again*, Panel, International Conference on Data Engineering, 2017
- Organizer, MetroLab Workshop on Big Data and Human Services, 2017
- Invited Participant, White House Frontiers Conference, Pittsburgh, 2016
- Represented Seattle and the University of Washington at the White House Smart Cities Forum, White House Office of Science and Technology Policy, Washington DC, 2015
- Invited Member, SHRP 2 Naturalistic Driving Study Business Plan Expert Advisory Board, 2016-2019
- Invited Participant, Data Science and Privacy Roundtable, Macarthur Foundation, Chicago, 2015
- Invited to participate in National Research Council meeting on "Training the Big Data Workforce," 2014
- Invited Participant, Cloud Computing Roundtable, DARPA, Seattle, 2014
- Invited Participant, White House Data to Knowledge to Action, 2014
- Chair, *Teaching Data science and Databases*, Panel, ACM SIGMOD international conference on Management of data.
- Co-Organizer, *Designing good algorithms for MapReduce and beyond*, Proceedings of the Third ACM Symposium on Cloud Computing, 2014.
- Co-Organizer, *Education and career paths for data scientists*, Panel, Magdalena Balazinska, Susan B Davidson, Bill Howe, Alexandros Labrinidis, Proceedings of the 25th International Conference on Scientific and Statistical Database Management, 2013
- Chair, High-Performance Computing Meets Databases, 2011
- Chair, High-Performance Computing Meets Databases, 2010

Campus Leadership

- Graduate Council Member, 2023-present
- Co-Founder, Responsibility in AI Systems and Experiences (RAISE), 2021-present
- Steering Committee, UW eScience Institute, 2017-present

- Steering Committee, Center for Statistics in the Social Sciences, 2017-2023
- Faculty Senate Representative, 2021-2023
- Elected Faculty Council Chair, Information School, 2019-2022
- Founding Associate Director of the eScience Institute, responsible for management of the staff, operations, and program development.
- Program Director and Faculty Chair, Data Science Masters Degree. Led curriculum design, program design, hiring
- Founder, UW Data Science for Social Good Program
- Co-Founder, Urban@UW, campus-wide organization to facilitate interaction between technical and qualitative research in urban science
- Director, Urbanalytics Group
- UW Director, Cascadia Urban Analytics Collaborative (with University of British Columbia and Portland State)
- Director, Seattle MetroLab (with Michael Mattmiller, CTO of City of Seattle)
- Steering Committee, Center for Statistics in the Social Sciences, 2015-present
- Designed and filled a new UW-IT position Director of Cloud and Data Solutions (Rob Fatland, previously of MSR) to organize cloud efforts campus-wide
- Lead, Software Working Group, eScience Institute, 2012-2016
- Launched a free campus-wide service based on our research data management system SQLShare, in partnership with UW-IT
- Lead, Twitter Data Grant Program providing access to Twitter data for 10+ projects across five departments at the University of Washington.
- Founder, UW eScience Incubation Program for Data-Intensive Science Projects
- Founder, University Education Outreach Certificate in Data Science, University of Washington, 2013
- Founder, University Education Outreach Certificate in Cloud Computing, University of Washington, 2013
- Designed and operated *Introduction to Data Science* online course on Coursera, attracting 200,000+ students across two offerings

Local Engagement

- Invited Speaker, Workshop on Science and Technology for Washington State: Advising the Legislature, October 2017
- Moderator, "Fake News, Privacy, and Big Data," Public Forum, Seattle, August 2017
- Moderator, "Big Data Visualization," Public Forum, Seattle, 2016
- Invited Speaker, City of Seattle Data Camp, 2016
- Moderator, "Big Data Systems," Public Forum, Dato, Seattle 2014