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
Program Director and Faculty Chair, Data Science Masters Degree

UNIVERSITY OF WASHINGTON	2009–2017

Founding Associate Director, eScience Institute
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. **Fides: Towards a Platform for Responsible Data Science.**
   Julia Stoyanovich, Bill Howe, Serge Abiteboul,
   Gerome Miklau, Arnaud Sahuguet, and Gerhard Weikum
   ACM Conference on Scientific and Statistical Database Management (SSDBM) (2017)
2. **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)
3. **Viziometrics: Analyzing visual information in the scientific literature**
   Po-shen Lee, Jevin D. West, and Bill Howe
   IEEE Transactions on Big Data (2017)

4. **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)

5. **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)

6. **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)

7. **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)

8. **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)

9. **Voyager: Exploratory analysis via faceted browsing of visualization recommendations**
   Kanit Wongsuphasawat, Dominik Moritz, Anushka Anand,
   Jock Mackinlay, Bill Howe, Jeffrey Heer
   IEEE Transactions on Visualization and Computer Graphics 22(1) (2016)

10. **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

11. **MusicDB: Relational Approach for Numeric Longitudinal Music Analytics**
    Jeremy Hyrkas, Bill Howe
    International Society on Music Information Retrieval (ISMIR) (2016)

12. **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)

13. **GossipMap: a distributed community detection algorithm for billion-edge directed graphs**
    Seung-Hee Bae, Bill Howe
    Supercomputing (2015)

14. **The database group at the University of Washington**
    Magdalena Balazinska, Bill Howe, and Dan Suciu
    ACM SIGMOD Record 43(1) (2014)

15. **Time-Varying Clusters in Large-Scale Flow Cytometry**
    Jeremy Hyrkas, Daniel Halperin, and Bill Howe
    Innovative Applications of Artificial Intelligence (IAAI) (2014)

16. **Helping scientists reconnect their datasets**
    Abdussalam Alawini, David Maier, Kristin Tufte, and Bill Howe
    ACM Scientific and Statistical Database Management Conference (SSDBM) (2014)

17. **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)

18. **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)

19. **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)

20. **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)

21. **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

22. **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)

23. **Query-based data pricing**
    Paraschos Koutris, Prasang Upadhyaya, Magdalena Balazinska, Bill Howe, and Dan Suciu
    ACM Symposium on Principles of Database Systems (PODS) (2012)

24. **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)

25. **Advancing Declarative Query in the Long Tail of Science**
    Bill Howe and Daniel Halperin
    IEEE Data Engineering Bulletin 35(3) (2012)

26. **The HaLoop approach to large-scale iterative data analysis**
    Yingyi Bu, Bill Howe, Magdalena Balazinska, and Michael Ernst
    Special Issue: Best of VLDB 2010 VLDB Journal, 21(2) (2012)

27. **Optimizing Large-Scale Semi-Naive Datalog Evaluation in Hadoop**
    Marianne Shaw, Bill Howe, Paris Koutris, and Dan Suciu
    Datalog 2.0 (2012)

28. **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)

29. **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)

30. **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)

31. **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)

32. **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)
33. **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)
34. **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)
35. **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)
36. **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)
37. **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)
38. **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)
39. **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)
40. **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)
41. **Scientific Mashups: Runtime-Configurable Data Product Ensembles**
    Bill Howe, Harrison Green-Fishback, and David Maier
    Scientific and Statistical Database Management Conference (SSDBM) (2009)
42. **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)
43. **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)
44. **Smoothing the ROI curve for scientific data management applications**
    Bill Howe, David Maier, and Laura Bright
    Conference on Innovative Data Systems Research (CIDR) (2007)
45. **Retrofitting a Data Model to Existing Environmental Data**
    Bill Howe and David Maier
    Scientific and Statistical Database Management Conference (SSDBM) (2005)

46. **Logical and Physical Data Independence for File-based Scientific Applications**
    Bill Howe and David Maier
    IEEE Data Eng. Bull. 27(4) (2004)
47. **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)
48. **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)
49. **A language for spatial data manipulation**
    Bill Howe, David Maier, and Antonio Baptista
    Journal of Environmental Informatics 2(2) (2003)
50. **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)
51. **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. **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. **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)
   (co-located with NIPS 2017)
2. **Synthetic Data for Social Good**
   Bill Howe, Julia Stoyanovich, Haoyue Ping, Bernease Herman, Matt Gee
   Bloomberg Data for Good Exchange (2017)
3. **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)
4. **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)
5. **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)
6. **High-variety cloud databases**
   Shrainik Jain, Dominik Moritz, Bill Howe
   Workshop on Quality in Databases (QDB) (2016)
   (co-located with VLDB)
7. **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)
8. **High Variety Cloud Databases**
   Shrainik Jain, Dominik Moritz, Bill Howe
   Workshop on Cloud Data Management (CloudDM) (2016)
   (invited keynote)
   (co-located with ICDE)
9. **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)
10. **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)
11. **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)

12. **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)
13. **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)
14. **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)
15. **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)
16. **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)
17. **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)
18. **The eScience Appliance: Provisioning an Inexpensive Bottom-Up Cyberinfrastructure**
    Bill Howe, and Roger Barga
    Microsoft eScience workshop (2008) (co-located with eScience)
19. **Managing the forecast factory**
    Laura Bright, David Maier, and Bill Howe
    Workshop on Workflow and Data Flow for Scientific Applications (SciFlow) (2006)
20. **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)
21. **Modeling data product generation**
    Bill Howe and David Maier
    Workshop on Data Derivation and Provenance (2002)

Peer-Reviewed Demonstration Papers

1. **DataSynthesizer: Privacy-Preserving Synthetic Datasets**
   Haoyue Ping, Julia Stoyanovich, and Bill Howe
   ACM Scientific and Statistical Database Management Conference (SSDBM) (2017)
2. **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)

3. **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)
4. **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)
5. **Automatic Example Queries for Ad Hoc Databases**
   Bill Howe, Garret Cole, Nodira Khoussainova, Leilani Battle
   ACM Conference on Management of Data (SIGMOD) (2011)
6. **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. **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
2. **SQLShare: DB-as-a-Service for Science**
   Bill Howe
   American Geophysical Union Fall Meeting, 2013
3. **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
4. **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)
5. **Scientific mashups: Runtime-configurable data product ensembles**
   Harrison-Green Fishback and Bill Howe
   IEEE International conference on e-Science (eScience) (2008)
6. **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, **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)
2. Co-PI **Microsoft** Transportation Data Collaborative, $150k, with Jan Whittington (UW Urban Planning) and Mark Hallenbeck (TRAC)
3. 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)
4. Lead PI, **Microsoft** Cascadia Urban Analytics Collaborative, 2017-2019, $500,000
5. Lead PI, **Department of Defense and PNNL** Polystore Data System, 2015-2016, $500,000
6. Lead PI, **Macarthur Foundation** Advancing the Seattle MetroLab, 2016-2017, $250,000

7. 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)
8. Lead PI, **Department of Defense and PNNL** Unified Services for Federated Big Data Environment, 2015-2016, $500,000
9. Lead PI, **Macarthur Foundation** Towards a Seattle Metro Consortium for Urban Data Science, 2014-2016, $300,000
10. Co-PI **UW Provost's Office** Urban@UW, 2015-2017, $750,000, with Thaisa Way (UW), Vikram Jandalaya (UW), Thom Dunning (PNNL)
11. 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
12. 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
13. Lead PI, **PNNL Subcontract** Hybrid Architectures for Graph Query, 2014-2015, $200,000
14. 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)
15. Lead PI, **PNNL Subcontract** Hybrid Languages for Graph Query, 2013-2014, $100,000
16. 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)
17. 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).
18. Co-PI, **EMC Grant** Large-Scale Visual Analytics, 2012-2014, $200,000, with Magda Balazinska (UW)
19. Co-PI, **NSF Grant** CCF-1047815, Relational Data Markets in the Cloud, 2011-2013, $370,000, with Magda Balazinska (UW) and Dan Suciu (UW)
20. Lead PI, **NSF Grant** CCF-1060213, EAGER: Scalable Algebraic Visualization in the Cloud, 2010-2011 $117,567
21. Co-Lead, **Gordon and Betty Moore Foundation**, Data-Intensive Science in the Long Tail, 2010-2013, $750,000, with Ed Lazowska
22. 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

- **Luke Rodriguez**, Information School   2017–present
- **Sean Yang**, Electrical Engineering   2016–present
- **Emily Furst**, Computer Science & Engineering (with Mark Oskin)   2016-present
- **Shrainik Jain**, Computer Science & Engineering   2014–present
- **Dylan Hutchison**, Computer Science & Engineering   2016–present
- **Dominik Moritz**, Computer Science & Engineering (with Jeff Heer)   2016–present
- **Brandon Myers**, Computer Science & Engineering (with Mark Oskin)   2013–2016
  now faculty at University of Iowa
- **Poshen Lee**, Electrical Engineering   2014–2017
  now postdoc at UW iSchool
- **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–present
- **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

- 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

- *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.
- *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. **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)
2. **Microsoft Jim Gray Seed Grant** 2010, $25,000
3. **Microsoft Jim Gray Seed Grant** 2007, $40,000
4. **Department Dissertation Award**, Maseeh College of Engineering and Computer Science, Portland State University, May 2007
5. **Best of VLDB 2004**, selected for publication in VLDB Journal special issue, for "Algebraic manipulation of scientific datasets"

Selected Talks and Panels

1. Distinguished Colloquium, "The Next Decade of Data Science," University of Maryland, College Park, December 2017
2. Invited Talk, "Epistemic Issues in Data Science," University of Massachusetts, Amherst, December 2017
3. Invited Speaker, Data Analysis and Visualization Workshop, Dagstuhl, Germany, 2017
4. Keynote, "Responsible Urban Data Science," Redondo Beach, CA, November 2017
5. Invited Talk, "Viziometrics: Mining the Visual Literature," Scientific Computing Institute, University of Utah, November 2017
6. Invited Talk, "Deep Curation," New York University, 2017
7. Invited Talk, "Fake Data for Social Good," Bloomberg Data for Good Exchange, 2017
8. Panelist, Bias and Ethics in City Services Data Science, Bloomberg Data for Good Exchange, 2017
9. Invited Talk, "Big Data + Big Sim: Query Processing over Unstructured CFD Models," ISIM Workshop, Durham, UK
10. Invited Talk, "Data, Responsibly: The Next Decade of Data Science," iSchool Founding Board, University of Washington, 2017
11. Invited Speaker, Data Responsibly Workshop, Dagstuhl, Germany, 2016
12. Invited Talk, "Data Science, Data Curation, and Human-Data Interaction," Data Science Summit, San Francisco, 2016
13. Keynote, Cloud Data Management Workshop (with ICDE), 2016, Helsinki
14. Invited Talk, "Urban Data Science @ UWâĂİ, NIAC Workshop on Smart Cities, 2016
15. Invited Talk, "Big Data Talent in Academic and Industry R&D," 2015
16. Panelist, "Data Science Education," NSF BigData PI Meeting, 2017
17. Invited Talk, "High Productivity ComputingâĂİ, Los Alamos National Lab, November 2015
18. Invited Talk, "High Productivity ComputingâĂİ, Chesapeake Large-Scale analytics Conference, October 2015
19. Invited Talk, "High Productivity ComputingâĂİ, Neurofutures Symposium, Portland, August 2015
20. Invited Speaker, "Lightweight Data Systems in the Cloud: Costs, Benefits and Best Practices," American Geophysical Union, 2015
21. Invited Talk, ITIC and Macarthur workshop "Partnership for Sustainable Smart Cities of the Future"
22. Invited Talk, National Research Council Meeting on "Training the Big Data Workforce"
23. Keynote at eXtremely Large Databases, South America, 2014
24. Invited Talk at JSM panel on Big Data Education: Big Data Curricula at the UW eScience Institute.
25. Invited Talk on Myria and scalable graph clustering at the Mining of Massive Datasets workshop at UC Berkeley, 2014.
26. Invited Speaker, Cloud Computing Workshop, Qatar Computing Research Institute, Qatar, 2014
27. Keynote at eResearch New Zealand, "eScience and Data science at the University of Washington's eScience Institute," 2013
28. Invited Talk, "Myria: Analytics-as-a-Service for (Data) Scientists," High Performance Transaction Processing, 2013
29. Invited Panelist, "Big Data Curricula at the UW," Joint Statistical Meeting, 2013
30. Invited Talk, "Enabling Collaborative Research Data Management with SQLShare," Internet2 Symposium, 2013
31. Invited Talk, "Conservative Regridding and Algebraic Manipulation of Meshes with GridFields," HPEC Conference, 2012
32. Invited Talk, "Democratizing Data Science," UC Berkeley, 2012
33. Invited Talk, "Iterative Computation for Big Data," University of Indiana, 2012
34. Invited Talk, "Virtual Appliances, Cloud Computing, and Reproducible Research," AMP Workshop on Reproducible Research, 2011

35. Invited Talk, "A New Partnership for Cross-Scale, Cross-Domain eScience," Carnegie Mellon University, 2010
36. Invited Talk, "End-to-End eScience: Integrating Query, Workflow, and Visualization at an Ocean Observatory," Microsoft Research, 2009

Professional Service

- National Academy of Sciences Roundtable on Postsecondary Data Science Education, 2016 - 2019
- Program Committee, BigVis Workshop, 2018, (co-located with EDBT)
- Program Committee, International Conference on Data Engineering (ICDE), 2018
- eScience Conference Steering Committee, 2017-2019
- 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–present
- **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

National Leadership

- Member, National Academy of Sciences Roundtable on Data Science Education, 2016-2019
- 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

- 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