

Gregory Todd Williams

+1 (310) 729-6041

greg@kasei.us

Education

- 2007–2013 *Rensselaer Polytechnic Institute* *Ph.D. Computer Science*
Ph.D. in Computer Science, specializing in scalable query answering on the Semantic Web. Designed, implemented, and evaluated a federated query answering system for SPARQL queries able to automatically improve query results through iterative query plan expansion based on the discovery of new, relevant data sources.
- 2006–2007 *University of Maryland* *Ph.D. coursework*
- 2001–2003 *Wheaton College* *Magna cum laude, B.A. Computer Science with Honors, Philosophy Minor*
- 1998–2000 *Santa Monica College*

Experience

- 2020– *Amazon Web Services, Seattle, WA* *Senior Software Development Engineer*
Implemented and maintained the cost-based statistics component of the Neptune graph database, supporting planning and optimization of SPARQL and OpenCypher queries. Designed and implemented cost-based statistics and query-plan components as part of the development of the Neptune Analytics graph database, taking this work from design through product release. Designed extensions to the SPARQL query language to support interoperability with the Property Graph data model.
- 2019–2020 *J. Paul Getty Trust, Los Angeles, CA* *Senior Data Engineer*
Designed and implemented data pipeline for producing semantically enriched art sales, provenance, and publications data from legacy systems. Collaborated with domain experts to ensure both fidelity of data and that the resulting data model would serve research needs. The resulting data conformed to the *linked.art* profile of CIDOC-CRM, and used JSON-LD to allow use by tools supporting a wide range of semantic capability, and supporting research by consumers with varied needs.
- 2015–2018 *Hulu, Santa Monica, CA* *Software Developer*
Designed, implemented, and maintained a meta-query planning tool responsible for processing high-level analytical queries, selecting an appropriate data source, and generating a structured query to fully answer the user's query. The system was designed to work efficiently with complex inputs and over an extensible, wide range of available databases including Hive, Presto, Impala, and MySQL, and generate optimized queries over hundreds of terabytes of data.
- 2014–2015 *Pacific Northwest National Laboratory, Richland, WA* *Consultant*
Designed and implemented a SPARQL query planning and optimization system used in conjunction with existing massively parallel graph database system.
- 2011 *O'Reilly Media, Inc., Sebastopol, CA* *Consultant*
Reviewed existing use of semantic web technology and data, and provided guidance on its continued use in critical business functions. Implemented tools to allow data exchange between existing RDF, relational, and document databases.
- 2005–2006 *Wheaton College Genomics Research Group, Norton, MA* *Consultant*
Consulted on the creation of an introductory programming textbook for biology researchers. Oversaw the technical aspects of authoring Perl code meant for diverse, cross-platform use by novices, and insured that included code used the standards and best practices of the Perl community.
- 2005–2006 *Shopzilla, Inc., Los Angeles, CA* *Software Engineer*
Designed and implemented a site taxonomy server, integrating an existing product taxonomy database with a new REST API for querying and updating an RDF/OWL data model using SKOS. Redesigned and implemented a merchant statistics reporting tool with a focus on scalability, maintainability, and extensibility. Work involved generating, storing and executing complex queries across multiple Sybase ASE and IQ databases using Trasact-SQL and Perl.
- 2003–2004 *BizRate.com, Los Angeles, CA* *Software Engineer*
Implemented "related product" search feature, integrating with an existing mod_perl front-end. Redesigned and implemented database abstraction, localization, business logic and presentation classes for use with consumer search site. Worked with CTO on development of new search technologies requiring scalability and efficiency.

- 2001–2003 *Wheaton College Genomics Research Group, Norton, MA* *Wheaton Research Fellow*
Implemented: C++ and Perl database allowing statistical analysis of DNA sequences across genomes; a search engine to correlate results with existing published literature in PubMed; a framework for researching motif distributions in targeted genomic regions.
- 2001–2003 *Wheaton College, Norton, MA* *Mars Fellow*
Researched surface reconstruction and supporting infrastructure. Designed and implemented an object-oriented, dynamic, surface reconstruction research environment in C++ utilizing OpenGL. Responsible for coordinating the work of two other programmers to produce the final system.
- 1996–2001 *Cnation Inc., Los Angeles, CA* *Senior Software Engineer*
Acted as a project leader for software development projects and was responsible for managing other developers. Designed and implemented a large, open source, mod_perl based web application framework and its associated database abstraction and parsing classes.

Publications

- 2024 Michael Schmidt, Brad Bebee, Willem Broekema, Mohamed Elzare, Carlos Manuel Lopez Enriquez, Marcin Neyman, Florian Schmedding, Andreas Steigmiller, Bryan Thompson, Geo Varkey, Gregory Todd Williams and Amanda Xiang. (2024). openCypher over RDF: Connecting Two Worlds. *Proceedings of the ISWC Poster and Demos Track, Baltimore, MD, November 11 2024*.
- Willem Broekema, Mohamed Elzare, Ora Lassila, Carlos Manuel Lopez Enriquez, Marcin Neyman, Florian Schmedding, Michael Schmidt, Andreas Steigmiller, Geo Varkey, Gregory Todd Williams and Amanda Xiang. (2024). openCypher Queries over Combined RDF and LPG Data in Amazon Neptune *Proceedings of the ISWC Industry Track, Baltimore, MD, November 11 2024*.
- Olaf Hartig, Gregory Todd Williams, Michael Schmidt, Ora Lassila, Carlos Manuel Lopez Enriquez and Bryan Thompson. (2024). Datatypes for Lists and Maps in RDF Literals. *Proceedings of the Extended Semantic Web Conference Poster and Demos Track, Hersonissos, Greece, May 26 2024*. (Best Poster Award.)
- 2015 Morari, A., Castellana, V. G., Villa, O., Weaver, J., Williams, G. T., Haglin, D. J., Tumeo, A. and Feo, J. (2015). GEMS: Graph Database Engine for Multithreaded Systems in K. Li, H. Jiang, L. T. Yang and A. Cuzzocrea (Eds.), *Big Data: Algorithms, Analytics, and Applications* (pp. 139–156). New York, NY: Chapman and Hall/CRC.
- 2014 Gregory Todd Williams and Kjetil Kjernsmo. (2014). Pushing complexity down the stack. *Proceedings of the ISWC Developers Workshop 2014, Riva del Garda, Italy, October 19 2014*.
- 2013 Gregory Todd Williams. (2013). Planning and Evaluation of Federated Queries on the Web. *Ph.D. thesis, Rensselaer Polytechnic Institute, Troy, New York, March 2013*.
- 2011 Gregory Todd Williams and Jesse Weaver. (2011). Enabling fine-grained HTTP caching of SPARQL query results. *Proceedings of the 10th International Conference on The Semantic Web (ISWC), Bonn, Germany, October 23 2011*.
- 2010 Gregory T. Williams, Jesse Weaver, Medha Atre, and James A. Hendler. (2010). Scalable Reduction of Large Datasets to Interesting Subsets. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*.
- Sibel Adalı, Robert Escrivá, Mark K. Goldberg, Mykola Hayvanovych, Malik Magdon-Ismael, Boleslaw K. Szymanski, William A. Wallace and Gregory T. Williams. (2010). Measuring Behavioral Trust in Social Networks. *Proceedings of IEEE International Conference on Intelligence and Security Informatics (ISI 2010)*.
- 2009 Gregory Todd Williams, Jesse Weaver, Medha Atre and James A. Hendler. (2009). Scalable Reduction of Large Datasets to Interesting Subsets. *Proceedings of the 8th International Semantic Web Conference (Billion Triples Challenge), Chantilly, Virginia, October 27 2009*.
- Jesse Weaver and Gregory Todd Williams. (2009). Scalable RDF Query Processing on Clusters and Supercomputers. *Proceedings of the 5th International Workshop on Scalable Semantic Web Knowledge Base Systems (SSWS), Chantilly, Virginia, October 26 2009*.
- 2008 Gregory Todd Williams. (2008). Supporting Identity Reasoning in SPARQL Using Bloom Filters. *Proceedings of Workshop on Advancing Reasoning on the Web, Tenerife, Spain, June 2 2008*.
- 2007 Gregory Todd Williams. (2007). Extensible SPARQL Functions With Embedded Javascript. *Proceedings of 3rd Workshop SFSW'07, Innsbruck, Austria, June 6 2007*.

- 2005 Gregory Todd Williams. (2005). MT-Redland: An RDF Storage Backend for Movable Type. *Proceedings of 1st Workshop SFSW'05, Hersonissos, Greece, May 30 2005*, CEUR Workshop Proceedings, ISSN 1613-0073.
- 2004 Betsey D. Dyer, Mark D. LeBlanc, Stephen Benz, Peter Cahalan, Brian Donorfio, Patrick Sagui, Adam Villa and Gregory Williams (2004). A DNA motif lexicon: cataloguing and annotating sequences. *In Silico Biology* 4,0039(2004).
- 2003 Gousie, M. B. Williams, G., Agnitti, T., and Doolittle, N. (2003). CompSurf: An Environment for Exploring Surface Reconstruction Methods on a Grid. *Computers & Geosciences* 29,9(2003).
- 2002 Williams, G., Doolittle, N. and Agnitti, T. (2002). A surface reconstruction research environment. *The Journal of Computing in Small Colleges*,v17(6), 301–302. Presented at the 2002 Northeastern Conference on Computing in Small Colleges, Worcester, MA, April 2002.
- LeBlanc, M., Baron, M., Christoforou, A., Doolittle, N., Kimball, M., Villa, A., Williams, G. and Dyer, B. (2002). The DNA Motif Lexicon – cataloguing and annotating genomes. In *Proceedings of the 14th International Genome Sequencing and Analysis Conference, October 2–5, 2002, Boston, MA*.

Program Committees

- 2017–2020 International Semantic Web Conference (ISWC)
- 2019–2020, 2022 European Semantic Web Conference (ESWC)
- 2010–2018 Linked Data on the Web (LDOW)
- 2013–2016 European Semantic Web Conference (ESWC)
- 2012 Joint Workshop on Scalable and High-Performance Semantic Web Systems (SSWS+HPCSW)
- 2011 High-Performance Computing for the Semantic Web (HPCSW)
- 2010 International Semantic Web Conference (ISWC)
- 2008–2010 Scripting for the Semantic Web (SFSW) (Program Chair, 2010)

Standards Groups

- 2022– W3C RDF-star Working Group
- 2009–2013 W3C SPARQL Working Group. Editor, SPARQL 1.1 Service Description, SPARQL 1.1 Protocol.

Projects

- 2016– Designed and implemented the open source **Kineo** SPARQL system for Swift. Based on trait-based design and lessons learned in the **Attean** project, Kineo is a research platform for exploring new approaches to implementing and extending SPARQL.
- 2004– Designed and implemented the open source **Attean**, **RDF::Trine** and **RDF::Query** RDF frameworks and query engines for Perl. RDF::Query was one of the initial 14 implementations of SPARQL as it was published as a W3C Recommendation in 2008, and was one of the first system to fully support the SPARQL 1.1 Query, Update, Service Description, and Protocol standards in 2013.
- 2009–2013 Designed and implemented a system as part of the **data.gov** project to convert open governmental data from tabular formats to RDF, allowing iterative enhancement of data including schema mapping and recording of maximal data provenance.