# Edmon Begoli, PhD

Professional and Research Appointments	Distinguished Scientist and Section Head, AI Systems R&D Section National Security Sciences Directorate Oak Ridge National Laboratory (ORNL)	
Academic Appointments	Associated Researcher (Visiting) SKY Computing Lab The Department of Electrical Engineering and Computer Science (EECS) The University of California (UC Berkeley), Berkeley, CA	
	Joint Faculty Professor of Computer Science The Department of Electrical Engineering and The University of Tennessee, Knoxville, TN	Computer Science (EECS)
Contact Information	1 Bethel Valley Rd. Building 5700 Oak Ridge National Laboratory (ORNL) Oak Ridge, TN 37831, USA	Voice: (865) 576-0599 Fax: (865) 576-4368 E-mail: begolie@ornl.gov web: https://www.ornl.gov/staff-profile/edmon-begoli
Research	• Resilient and Robust AI/ML Systems and Architectures	
Interests and Expertise	• High-performance analytic processing	
Education	University of Tennessee, Knoxville, Tennessee USA	
	Ph.D., Computer Science, May 2014	
	<ul> <li>Dissertation Topic: "Procedural-Reasoning Architecture for Applied Behavior Analysis-based Instructions"</li> <li>Area of Concentration: Knowledge Representation, Artificial Intelligence Systems</li> <li>Advisor: Dr. Bruce J. MacLennan</li> </ul>	
	University of Colorado-Boulder, Boulder, Colorado USA	
	<ul> <li>M.S., Computer Science, May 2003</li> <li>Thesis: CARES – Cancer Research Results Exchange System</li> <li>Area of Concentration: Semantic Web Architectures, Artificial Intelligence Systems</li> <li>Advisor: Dr. Kenneth M. Anderson</li> </ul>	
	East Tennessee State University, Johnson City, Tennessee USA	
	B.S., Computer Science, Magna Cum Laude August, 1998	
Honors, Awards	• IEEE Computer Society Distinguished Contributor	
AND RECOGNITION	• 2022 Google Research Innovator	
	• IEEE and ACM Senior Member	
	• Department of Energy, Secretary's Letter of Recognition for Service to the DOE-VA Program	
	• Department of Energy, Deputy Under Secretary for AI, Letter of Commendation for Service as a PI on the DOE-VA MVP CHAMPION program	
	• Heidelberg Laureate Forum - Early Care	er Researcher, 2016. Mentor: Sir Michael Atiyah

- Top 9% Contributor on Stack Overflow
- Nashville Technology Council (NTC) CTO of the Year, finalist, 2014
- White House Big Data Fact Sheet<sup>1</sup> Pg.6, references 'Big Data' project for CMS
- Tennessee Commissioner Mitchel (Department of Safety), A Letter of Commendation, 2011
- OPM e-Government Initiative, NSC Y-12 Team Award, 2005
- Made in Express, Microsoft Global Programming Competition, finalist, 2005
- Lucent Technologies, EAI Implementation, CIO's Letter of Commendation for Outstanding Performance, 2002
- Multiple Undergraduate Honors, Magna Cum Laude, Upsilon Pi Epsilon CS Honors Society

Professional	Oak Ridge National Laboratory, Oak Ridge, Tennessee USA	
Experience	R&D Section Head / Distinguished Scientist	October, 2020 - present

- Research section head, leading research groups within a section focused on the state-of-the-art applications of AI and Machine Learning in cyber and digital domains; managing a 50+ person organization ranging from distinguished and senior staff to post-docs and students,
- Initiative lead for ORNL's "Emerging and Cyber Technologies" Lab Directed Research and Development (LDRD) Initiative
- Principal Investigator (PI) and research lead for two strategic programs focused on highperformance, robust, and reliable analytic processing
- Responsible for the development of a research vision and its execution in the context of national security programs (with focus on cyber and digital domains),

## Oak Ridge National Laboratory, Oak Ridge, Tennessee USA

Director, Scalable Protected Data Facilities (SPDF) / Senior Scientist May, 2016 - September 2020

- Principal Investigator (PI) for the MVP CHAMPION program <sup>2</sup>, a large-scale DOE-VA national program for precision medicine and patient population research.
- Design of large-scale platforms for data analysis ("Big Data") on the major national programs in healthcare, life sciences, and national security.
- Working with a program manager/PI to help translate sponsoring agencies needs into well defined technical requirements, and technical solutions.
- Defining the overall architecture for the projects including detailed definition of the technical tasks and activities.
- Prototyping of the key elements of the architecture.
- Communication and documentation of the technical solutions to the customer and the project team.
- Assisting project manager with appropriate and efficient project staffing and division of tasks.
- Extraction and upkeep of reusable and innovative software solutions and algorithms.

# PYA Analytics, Knoxville, Tennessee USA

Chief Technology Officer

## May, 2013 - May, 2016

• A chief designer of the advanced analytic platform and a product line (Cyclone<sup>TM</sup>) for PYA Analytics;

 $<sup>^1 \</sup>rm https://www.whitehouse.gov/sites/default/files/microsites/ostp/big_data_fact_sheet_final.pdf <math display="inline">^2 \rm https://www.va.gov/opa/pressrel/pressrelease.cfm?id=2810$ 

- Lead analytics architect, Encounters Data Coordination Project (EDCP) for Centers for Medicare and Medicaid Services (CMS), Medicare-Medicaid Coordination Office;
- Lead architect ("Big Data" and analytic solutions) for USMC LOGCOM Master Data Management program;
- technology strategy and architecture development for Patient Master Index (precision health platform) for Sara Cannon Research Institute (SCRI);
- Conceptual designer and implementation lead, Data Quality Framework (DQF) for Hospital Corporation of America (HCA), Clinical Services Group (CSG);
- cloud security architect (AWS, Google Cloud) for in-house Cloud platform;
- Development of the technology strategy; design of the technology organization; recruitment, staffing, and mentoring of data scientists and data engineers.
- Solutions design; strategic business development;
- Supported growth of business from zero to multi-million in annual revenue.

#### Joint Institute for Computational Sciences (JICS), Oak Ridge, Tennessee USA Chief Data Officer

October, 2014 - May, 2016

October, 2007 - April, 2013

Strategy, vision and leadership of data-related research and development initiatives; JICS representative to NSF's regional 'Big Data' innovation hub;

- Principal investigator for HPC "Big Data" research initiatives at JICS;
- JICS representative on a NSF South Big Data Innovation Hub, leadership committee for research;
- Development of the data-related strategy and research agenda;
- Research and development of the data-related algorithms and frameworks; and
- Proposal development; technology selection; partnership development.

#### Oak Ridge National Laboratory (ORNL), Oak Ridge, Tennessee USA

Chief Architect/Researcher

Solutions Architect, Java Developer

- Assessment lead for CMS Integrated Data Repository (IDR), Medicare data warehouse;
- Chief architect for CMS Knowledge Discovery Infrastructure (KDI), a \$35 million dollar program for data analysis improvement at Centers for Medicare and Medicaid Services (CMS);
- Lead engineer and principal investigator (PI) for hardware-based Java Virtual Machine optimizations (DoD/IC);
- Implementation lead and lead architect for FACTS, data analysis solution for Department of Homeland Security (DHS) Chemical Security Assessment Tool (CSAT); and
- Implementation lead and lead architect for SERRI MSSI Suspicious Activities Reports (SAR) information sharing tool.

#### Y-12 National Security Complex, Oak Ridge, Tennessee USA

August, 2004 - September, 2007

- Technical lead for biometric processing of applications for sensitive positions (OPM-FTS);
- Solutions architect for national system for sensitive positions background investigations processing.

Laboratory Corporation of America (LabCorp), Burlington, North Carolina USA Enterprise Architect - Java EE Applications January, 2003 - 2004

- Author of the Java EE technical framework and methodology (LabCorp Unified Process);
- Technology innovation and technology transition (mainframe to Java EE) lead.

Lucent Technologies, Greensboro, North Carolina USA

Software Architect - Java EE Applications, Software developer September, 1998 - January 2003

- Software architect on Lucent Technologies EAI and B2B global solution;
- Java developer for Lucent iBuy, an award winning e-commerce application.

Research	Oak Ridge National Lab. (ORNL), Oak Ridge, Tennessee USA	
EXPERIENCE	Distinguished Scientist	May, 2013 - present
	Analytic platforms for protected data computing, Resilient and reliable A	I/ML platforms.

The University of California (UC Berkel	ey) EECS, Berkeley, CA
Visiting Researcher Scholar	February, 2019 - February 2022

University of Tennessee, Knoxville, Tennessee USAJoint Faculty Professor of Computer ScienceOctober, 2017 - Present

## Oak Ridge National Lab. (ORNL), Oak Ridge, Tennessee USA

 Research Staff
 October, 2007 - April, 2013

 Massively parallel processing databases; performance of hardware-based virtual machines.

GRANTS AND FUNDING AWARDS

- 2018-present, Principal Investigator (PI), HPC/AI RADV Program, Center for Medicare and Medicaid Services (CMS), \$14.5M
- 2016-present, Principal Investigator (PI), MVP CHAMPION, VICTOR Program, The Department of Veterans Affairs (VA), \$20M
- 2019-present, Google Cloud Platform Teaching Credits
- 2019-present, Google Cloud Platform Research Grant, \$7500
- 2016, NSF Full Travel Grant for 2016 Heidelberg Laureate Forum attendance. \$2200
- 2016, Data Accelerated Neuroimaging Library (DANIL), 2016 Intel Parallel Computing Center (Intel PCC) Grant, \$125,000 annually<sup>3</sup>
- 2013, National Science Foundation, Travel Grant, Artificial Intelligence in Education Conference, Doctoral consortium
- 2012, Co-PI, Laboratory Directed Research and Development (LDRD), Hybridizing HPC: Modern Cray Architectures for Big Data Problems, \$650,000
- 2010, Principal Investigator (PI), South Eastern Region Research Initiative (SERRI), Multistate Information Sharing Initiative, PI, \$450,000
- 2010, Principal Investigator (PI), South Eastern Region Research Initiative (SERRI), CLAERA: Collaborative, Location Aware Emergency Response Application, \$450,000
- 2009, Co-PI, Laboratory Directed Research and Development (LDRD) ID: 32112302, Inferring and Predicting the Social Dynamics of Groups via PsychoTextual and Communications Flow Analysis, Co-PI, \$650,000

 $<sup>^3\</sup>mathrm{grant}$  transferred to JICS PI L. Crosbie because of a move to ORNL.

Teaching Experience and Classes Taught	• ECE 692 Natural Language Processing, Fall 2020, EECS, The University of Tennessee, Knoxville
	• ECE 691 Scalable and Resilient AI/ML Systems, Spring 2020, EECS, The University of Tennessee, Knoxville
	• ECE 691 Scalable Data Analytics, Fall 2019, EECS, The University of Tennessee, Knoxville
Academic Advising and Research Mentoring	Maria Mahbub, 2019-ongoing, Data Science, PhD, Bredesen Center, UTK Ioana Danciu, 2018-ongoing, Biomedical Informatics, PhD, Vanderbilt University Eduardo Ponce Mojica, 2016-ongoing, Computer Science, PhD, University of Tennessee, EECS Sudarshan Srinivas, PhD, 2019, Computer Science, PhD, University of Tennessee, EECS Kris Brown, 2018, Computer Science, MS, University of Tennessee, EECS Michael Bowie, 2016-ongoing, Computer Science, BS/MS, UTK EECS / Oak Ridge National Lab- oratory Jeevith Bopaiah, Prashant Chandrasekar, Ashwin Kumar Vajantri, Kunwar Deep Singh Toor, Sum- mer 2017, Computer Science, graduate, Oak Ridge National Laboratory Jordan Chapian, 2014, Computer Science, graduate Linnet Vacha, Fall 2012, Mathematics, undergraduate, Oak Ridge National Laboratory Chelsey Dunnivan, Summer 2012, Computer Science, undergraduate, Oak Ridge National Labora- tory
Conference Papers	Begoli, Edmon, Sudarshan Srinivasan, and Maria Mahbub. "Improving Efficiency and Robustness of Transformer-based Information Retrieval Systems." In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, pp. 3433-3435. 2022.
	Begoli E., Goethert I., Knight K., A Lakehouse Architecture for the Management and Analysis of Heterogeneous Data for BiomedicalResearch and Mega-biobanks, 2021 IEEE International Confer- ence on Big Data, 4th Special Session on HealthCare Data
	Akidau T., Begoli E., Chernyak S., Hueske F., Knight K., Knowles K., Mills D., Sotolongo D., Watermarks in Stream Processing Systems:Semantics and Comparative Analysis of Apache Flink and Google Cloud Dataflow, 2021 Very Large Databases (VLDB)
	Knight K., Honerlaw J., Danciu I., Linares F., Ho Y., Gagnon D., Rush E., Gaziano M., Cho K., Cho K. , Begoli E., Standardized Architecture for a Mega-Biobank Phenomic Library: The Million Veteran Program (MVP), AMIA Bioinformatics Conference 2020, p. 326-334
	Rush E., Danciu I., Ostrouchov G., Cho K., Mayer B., Ho Y., Honerlaw J., Costa L., Linares F., Begoli E., JSONize: A Scalable Machine Learning Pipeline to Model Medical Notes as Semi-structured Documents, p. 533-541
	Begoli E., Akidau T., Hueske F., Hyde J., Knight K., Knowles K., One SQL to Rule Them All: An Efficient and Syntactically Idiomatic Approach to Management of Streams and Tables: An Industrial Paper. In 2019 International Conference on Management of Data (SIGMOD 19), June 30-July 5, 2019, Amsterdam, Netherlands. ACM, New York, NY, USA, 16 pages.
	Begoli E., Brown K., Srinivas S., and Tamang S., (December, 2018) SynthNotes: A Generator Framework for High-volume, Highfidelity Synthetic Mental Health Notes, Proceedings of the 2018 International Conference on Management of Data (Big Data), Seattle, Washington, USA.
	Begoli E., Mior M., Camacho-Rodríguez J., Hyde J., Lemire D. (2018, June) Apache Calcite: A Foundational Framework for Optimized Query Processing Over Heterogeneous Data Sources, Pro-
	5

ceedings of the 2018 International Conference on Management of Data (SIGMOD), Houston, Texas, USA.

Bowie M., Begoli E., Park B., Bopaiah J. (2017, October) Towards the LSTM-based Approach for Detection of Temporally Anomalous Data in Medical Datasets, In 2017 22nd MIT International Conference on Information Quality (ICIQ), Little Rock, Arkansas, USA.

Begoli E., Frasure C., Dunning T. (2016, March) Real-Time Discovery Services over Heterogeneous, Schema-less, and Complex Healthcare Datasets, In 2016 IEEE Big Data Services, Exeter College, Oxford UK

Baer T., Peltz P., Yin Q., Begoli E., (2015, May) Integrating Apache Spark Into PBS-Based HPC Environments. In proceedings of 2015 The Extreme Science and Engineering Discovery Environment (XSEDE) conference.

Bottles, K., & Begoli, E. (2014). Understanding the Pros and Cons of Big Data Analytics. Physician executive, 40(4), 6.

Begoli, E. (2014, March). Procedural Reasoning System (PRS) architecture for agent-mediated behavioral interventions. In SoutheastCon 2014, IEEE (pp. 1-8). IEEE.

Lim, S. H., Horey, J., Yao, Y., Begoli, E., & Cao, Q. (2013, May). Performance Implications from Sizing a VM on Multi-core Systems: A Data Analytic Application's View. In Parallel and Distributed Processing Symposium Workshops & PhD Forum (IPDPSW), 2013 IEEE 27th International (pp. 1001-1008). IEEE.

Begoli, E., Chila, T. F., & Inmon, W. H. (2013, April). Scenario-driven architecture assessment methodology for large data analysis systems. In Systems Conference (SysCon), 2013 IEEE International (pp. 51-55). IEEE.

Begoli, E., Ogle, C. L., Cihak, D. F., & MacLennan, B. J. (2013, January). Towards an Integrative Computational Foundation for Applied Behavior Analysis in Early Autism Interventions. In Artificial Intelligence in Education (pp. 888-891). Springer Berlin Heidelberg.

Begoli, E., & Horey, J. (2012, August). Design principles for effective knowledge discovery from big data. In Software Architecture (WICSA) and European Conference on Software Architecture (ECSA), 2012 Joint Working IEEE/IFIP Conference on (pp. 215-218). IEEE.

Horey, J., Begoli, E., Gunasekaran, R., Lim, S. H., & Nutaro, J. (2012, January). Big data platforms as a service: challenges and approach. In Proceedings of the 4th USENIX conference on Hot Topics in Cloud Computing (pp. 16-16). USENIX Association.

BOOKS AND	Begoli, E., Fusheng, W., Gang, L. (Eds.) Data Management and Analytics for Medicine and Health-
Proceedings	care, Springer Lecture Notes in Computer Science (LNCS), Proceedings of the Third International
	Workshop, DMAH 2017, Held at VLDB 2017, Munich, Germany, September 1, 2017.

BOOK CHAPTERS Begoli E., DeFalco, J. & Ogle C. (2016). Applications of Virtual and Augmented Reality Technologies in Education of Individuals with Autism Spectrum Disorder (ASD), in Supporting the Education of Children with Autism Spectrum Disorders, Advances in Early Childhood and K-12 Education (AECKE), IGI Publishing.

#### JOURNAL PAPERS Ferolito, Brian, Italo Faria do Valle, Hanna Gerlovin, Lauren Costa, Juan P. Casas, J. Michael Gaziano, David R. Gagnon, Edmon Begoli, Albert-Lszl Barabsi, and Kelly Cho. "Visualizing novel

connections and genetic similarities across diseases using a network-medicine based approach." Scientific Reports 12, no. 1 (2022): 1-12.

Xu, Yang, Edmon Begoli, and Rachel Patton McCord. "scican: Single-cell chromatin accessibility and gene expression data integration via cycle-consistent adversarial network." NPJ systems biology and applications 8, no. 1 (2022): 1-10.

Mahbub, Maria, Sudarshan Srinivasan, Ioana Danciu, Alina Peluso, Edmon Begoli, Suzanne Tamang, and Gregory D. Peterson. "Unstructured clinical notes within the 24 hours since admission predict short, mid & long-term mortality in adult ICU patients." PLOS one 17, no. 1 (2022): e0262182.

Hong, C., Rush, E., Liu, M. et al. Clinical knowledge extraction via sparse embedding regression (KESER) with multi-center large scale electronic health record data. Nature Digit. Med. 4, 151 (2021). https://doi.org/10.1038/s41746-021-00519-z

Begoli, E., Bridges, R.A., Oesch, S. and Knight, K.E., What Clinical Trials Can Teach Us About the Development of More Resilient AI for Cybersecurity. Editorial Board, European Cybersecurity Journal, 2021, p.22.

Hanna Gerlovin, Daniel C Posner, Yuk-Lam Ho, Christopher T Rentsch, Janet P Tate, Joseph T King, Jr., Katherine E Kurgansky, Ioana Danciu, Lauren Costa, Franciel A Linares, Ian D Goethert, Daniel A Jacobson, Matthew S Freiberg, Edmon Begoli, Sumitra Muralidhar, Rachel B Ramoni, Georgia Tourassi, J Michael Gaziano, Amy C Justice, David R Gagnon, Kelly Cho, Pharmacoepidemiology, Machine Learning, and COVID-19: An Intent-to-Treat Analysis of Hydroxychloroquine, With or Without Azithromycin, and COVID-19 Outcomes Among Hospitalized US Veterans, American Journal of Epidemiology, 2021;, kwab183, https://doi.org/10.1093/aje/kwab183

Baldoni, John, Edmon Begoli, Dimitri Kusnezov, and John MacWilliams. Solving Hard Problems with AI: Dramatically Accelerating Drug Discovery Through A Unique Public-Private Partnership. In Journal of Commercial Biotechnology 25, no. 4 (2020).

UMLS resources improve sieve-based generation and BERT-based ranking for concept normalization, Dongfang Xu, Manoj Gopale, Jiacheng Zhang, Kris Brown, Edmon Begoli, Steven Bethard, Journal of American Medical Informatics Association (JAMIA), 2020.

Rich, A.S., Rudin, C., Jacoby, D.M.P. et al. AI reflections in 2019. Nat Mach Intell 2, 29 (2020). https://doi.org/10.1038/s42256-019-0141-1

Artificial intelligence's essential role in the process of drug discovery Edmon Begoli and Dimitri Kusnezov, Future Drug Discovery, October 2019 1:2, DOI: https://doi.org/10.4155/fdd-2019-0026

Begoli, E., Bhattacharya, T., & Kusnezov, D. (2019). The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence. Nature Machine Intelligencevolume 1, 2023, DOI: https://doi.org/10.1038/s42256-018-0004-1

van Wyk, Franco, Anahita Khojandi, Akram Mohammed, Edmon Begoli, Robert L. Davis, & Rishikesan Kamaleswaran. (2018). A Minimal Set of Physiomarkers in High Frequency Real-Time Physiological Data Streams Predict Adult Sepsis Onset Earlier. International Journal of Medical Informatics.

Begoli, E., Schlegel, V., Atiyah, M., Adeyemo, P., & Baarslag, T. (2017). The Heidelberg Laureate Forum on the moving frontier between mathematics and computer science. XRDS 23, 3 (April 2017), 46-49. DOI: https://doi.org/10.1145/3055143

Begoli, E., Patel, P., & Christian, B. (2016). Storage and Read-Optimized Data Placement Structures for High Performance Analysis, Tutorials in Operations Research, INFORMS 2016, DOI: https://doi.org/10.1287/educ.2016.0143

Bottles, K., & Begoli, E. (2013). Understanding the pros and cons of big data analytics. Physician executive, 40(4), 6-10.

WORKSHOP PAPERS Begoli E., Lim SH, S., Performance Profile of Transformer Fine-Tuning in Multi-GPU Cloud Environments, IEEE BigData 2021 workshop BPOD 2021

Srinivasan, S., Begoli, E., Mahbub, M. (2021). *Nomen est Omen* – The Role of Signatures in Ascribing Email Author Identity with Transformer Neural Networks, 2021 Symposium on Advances in Digital Forensics Engineering (SADFE), Co-located with IEEE Security and Privacy (S&P).

Begoli, E., Srinivasan S., and Mahbub, M. The Transformers for Polystores – The Next Frontier for Polystore Research. In Heterogeneous Data Management, Polystores, and Analytics for Healthcare: VLDB Workshops, Poly 2020 and DMAH 2020, Virtual Event, August 31 and September 4, 2020, Revised Selected Papers 6, pp. 72-77. Springer International Publishing, 2021.

Mayer, Benjamin, Joshua Arnold, Edmon Begoli, Everett Rush, Michael Drewry, Kris Brown, Eduardo Ponce, and Sudarshan Srinivas. "Evaluating text analytic frameworks for mental health surveillance." In 2018 IEEE 34th International Conference on Data Engineering Workshops (ICDEW), pp. 39-47. IEEE, 2018.

Bowie, Michael, Edmon Begoli, and Byung H. Park. "Improving Quality of Observational Streaming Medical Data by Using Long Short-Term Memory Networks (LSTMs)." In Proceedings of 2018 International Conference on Data Engineering (ICDE), Paris, France, April 16-20, 2018. Workshop Data Management and Analytics for Precision Medicine (DEPM).

Vajantri, Ashwin Kumar, Kunwar Deep Singh Toor, Edmon Begoli, and Jack Bates. "An apache calcite-based polystore variation for federated querying of heterogeneous healthcare sources." In Big Data (Big Data), 2017 IEEE International Conference on, pp. 3221-3227. IEEE, 2017.

Begoli, E., Bates, J., Kistler, D. (2016, December). Towards a Heterogeneous, Polystore-like Data Architecture for the US Department of Veteran Affairs (VA) Enterprise Analytics. In Methods to Manage Heterogeneous Big Data and Polystore Databases Workshop collocated with 2016 IEEE International Conference on Big Data.

Burkhardt, F., Becker-Asano, C., Begoli, E., Cowie, R., Fobe, G., Gebhard, P. & Llewellyn, T. (2014, May). Application of EmotionML. In 5th International Workshop on Emotion, Social Signals, Sentiment & Linked Open Data ( $ES^3LOD$  2014).

Begoli E. (2012, September). Knowledge Discovery Fabric: data systems architecture in support of broad, comprehensive and flexible data analysis functions in national healthcare space. In Smoky Mountains Computational Sciences and Engineering Conference.

Begoli, E. (2012, August). A short survey on the state of the art in architectures and platforms for large scale data analysis and knowledge discovery from data. In Proceedings of the WICSA/ECSA 2012 Companion Volume (pp. 177-183). ACM.

INVITED PAPERS Begoli, E., Christian, J., Gadepally V., Papandoupulos, S. (September, 2017) An Emerging Role for Polystores in Precision Medicine. In Third International Workshop on Data Management and Analytics for Medicine and Healthcare, to be held at Munich, Germany, on September 1, 2017, in

	conjunction with the 43rd International Conference on Very Large Data Bases (VLDB).
Posters	One SQL to Rule them All, SIGMOD 2019
	Knowledge Discovery Architecture, ORNL Computing and Computational Sciences Directorate, National Advisory Committee, 2011
Invited Talks and Presentations	<ul> <li>2022 National Academies of Sciences, ICSB</li> <li>2021 Space Resilience Summit, Keynote Speaker</li> <li>2021 Lyber Forensics Summit, Speaker</li> <li>2021 Lab Research Technical Exchange, Adversarial AI, Keynote Speaker</li> <li>2020 BioData World Congress, Keynote speaker</li> <li>2020 Future Labs, panelist</li> <li>2020 AMIA Clinical Informatics Conference, Panelist</li> <li>2020 Health Datapalooza, Panelist</li> <li>2019 BioData World Congress, Basel, Switzerland</li> <li>2019 BioData World Congress, Basel, Switzerland</li> <li>2019 BioT World, Boston, Massachusets, USA</li> <li>2018 BioData World Congress, Basel, Switzerland</li> <li>2018 Stanford University, The School of Public Health, Stanford, California</li> <li>2018 UC Berkeley, RISE Lab, Berkeley, California</li> <li>2015 Data Quality Techniques and Methodologies, Chief Data Officer Summit, December 2015, Financial District, New York City, New York USA</li> <li>2014 Survey of Techniques for Record Linkage and Entity Resolution, Chief Data Officer Summit, New York City, New York USA</li> <li>2014 Sig Data - Hype and Reality, Business Analytics Society, Department of Business Analytics, University of Tennessee, Knoxville, Tennessee USA</li> <li>2014 Emerging Technologies - Panelist, Global Action Platform Summit, Nashville, TN USA</li> <li>2013 Inside the Data and Analytics-Driven Organization, Stanford Graduate School of Business, Stanford, California USA</li> <li>2012 Building Effective Knowledge Discovery Architecture, Presentation, National Health and Welfare Institute, Helsinki, Finland</li> <li>2010 Emerging Trends Impacting Information Sharing, Panel, AFCEA Symposium, George Mason University, Virginia USA</li> </ul>
Technical Reports	<ul> <li>Caddy, C., Begoli, E., Chanowski, S., Gates, A., Stockton, P., Wright, V., Cybersecurity and Digital Components – Supply Chain Deep Dive Assessment, U.S. Department of Energy Response to Executive Order 14017, "Americas Supply Chains", February 24, 2022</li> <li>Pleszkoch, M. G., Klasky, H. B., Advani, A., Begoli, E., Boone, A., and Fihn, S. D. (2018). Game Theoretic Approach for Understanding and Modeling Clinical Pathways (Stable Ischemic Heart Disease) (No. ORNL/TM-2018/980). Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).</li> <li>Park, B. H., Laska, J. A., Klasky, H. B., Boone, A., Ozmen, O., Karthik, R., and Begoli, E. (2018). Advanced Analytics Studies Applied to US Department of Veterans Affairs' Corporate Data Warehouse (No. ORNL/TM-2018/1034). Oak Ridge National Lab.(ORNL), Oak Ridge, TN (United States).</li> </ul>

	Vacha, L., Schryver, J., Shankar, M., & Begoli, E. (2013). Hierarchical Cluster Analysis of Service Usage and Geographic Variation in Medicare Spending (No. ORNL/TM-2013/70). Oak Ridge National Laboratory (ORNL).
	Begoli, E., Boehmann, B., & DeNap, F. A. (2012). Final Report: Multi-State Sharing Initiative (No. ORNL/TM-2012/82). Oak Ridge National Laboratory (ORNL).
	Begoli E., & Tomkins-Tinch C. (2011). Final Report: Collaborative, Location Aware, Emergency Response Application, South East Research Initiative (SERRI), SERRI Report 89930-01
	Schryver, J. C., Begoli, E., Jose, A. C., & Griffin, C. (2011). Inferring Group Processes From Computer-Mediated Affective Text Analysis. Tech. Rep. ORNL/TM-2010/277, Oak Ridge National Laboratory, Oak Ridge, TN.
Technical Articles and Media Presence	The Data Exchange with Ben Lorica, June 2022, February 2020 WBIR 2019, local TV feature on ORNL's work on suicide prevention, https://bit.ly/2PF96SL 3 technical articles on advanced Java EE and Python topics, Dr. Dobbs Journal 4 technical articles on functional programming, Python and Go, Developer.com Interview for Software Development Times (SD Times), Hadoop 2 article, 2014 Interview for Oracle Magazine, Pioneering work with web services, 2004
Workshop and	2023 NeurIPS, Workshop on Synthetic Data, Program Committee
Conference Organization and Service	2021 Systematic Approaches to Digital Forensics Engineering (SADFE) Symposium, Collocated with 2021 IEEE Security and Privacy, co-chair
	DMAH 2020, International Workshop on Data Management and Analytics for Medicine and Health- care (DMAH), VLDB 2020, Program Committee
	Poly 2020, Towards Polystores that manage multiple Databases, Privacy, Security and/or Policy Issues for Heterogeneous Data, VLDB 2019, Program Committee
	DMAH 2019, International Workshop on Data Management and Analytics for Medicine and Health- care (DMAH), VLDB 2020, Program Committee
	Reviewer, Healthcare Information and Management Systems Society (HIMSS) 2019 Conference.
	Co-chair and organizer of the Workshop on Data Engineering for Precision Medicine (DEPM), held at Paris, France, on April 16th, 2018, in conjunction with the 34th International Conference on Data Engineering (ICDE).
	Program committee on 2nd Workshop on Methods to Manage Heterogeneous Big Data and Polystore Databases, Co-located with the 2017 IEEE International Conference on Big Data.
	Co-chair and co-organizer of the Third International Workshop on Data Management and Analytics for Medicine and Healthcare, held at Munich, Germany, on September 1, 2017, in conjunction with the 43rd International Conference on Very Large Data Bases (VLDB).
	Co-organizer of the invite-only Summit on the Applications of HPC Techniques for Geo-Spatial Intelligence Analysis. Geo-spatial Intelligence Agency (NGA). Springfield, VA, 2017.
	Co-organizer of the invite-only Workshop on the Applications of HPC Techniques for Geo-Spatial

	Intelligence Analysis. Geo-spatial Intelligence Agency (NGA). Springfield, VA, 2016.
	"A Moving Frontier between Mathematics and Computer Science" co-organized with Sir Michael Atiyah and Vincent Schlegel, Heidelberg Laureate Forum Workshops, 2016.
	Session Chair - Enterprise Systems Engineering I, IEEE 2013 International Systems Conference (SysCon). Orlando, FL, USA.
	Organizer, International Workshop at WICSA/ECSA 2012: Architectures and platforms for knowl- edge discovery from data (APKDD 2012). Helsinki, Finland.
DISSERTATION	Begoli, E. (2014). Procedural-Reasoning Architecture for Applied Behavior Analysis-based Instructions.
Publications Review	<ul> <li>SIGMOD Record, 2021 - present</li> <li>Nature Communications, 2021-present</li> <li>Nature Machine Intelligence, 2020-present</li> <li>Heliyon, Elsevier, 2019</li> <li>American Medical Informatics Association (AMIA), 2019</li> <li>Journal Sensor, MDPI, 2016</li> <li>Journal IEEE Transactions on Services Computing, 2015</li> <li>IEEE Southeast East Conference 2014, 2015, 2016</li> <li>Elsevier Science and Technology, Book proposals, "Data Fusion for Intelligent Vehicles"</li> <li>Manning Technical Publications</li> </ul>
Programming Skills	Pyhton, Java, Clojure, SQL, C/C++, Linux shell scripting, ${\rm L\!AT}_{\rm E\!X}$
Technical Skills	Object Oriented Design, Functional Programming, Relational and Massively Parallel Processing (MPP) Databases, Cloud Architectures, Linux system administration, database administration.
CERTIFICATIONS	$ICS^2$ HealthCare Information Security and Privacy Practitioner (HCISPP) $ICS^2$ Certified Authorization Professional (CAP) WebMethods Developer Java Certified Programmer (2000-2003)
Professional Activities and Memberships	New York Academy of Sciences INCITS DM32.2 Task Group on Database, Database Languages, 2018-2020 Apache Software Foundation, 2018-Present NSF Regional 'Big Data' Innovation Hub, BD Hub South, 2015-2017 Global Action Platform - Strategic Technology Advisor, 2014-present NIST "Big Data" Working Group - Member, 2013-2016 W3C Multimodal Interaction Group - Invited Expert 2010-2012 The Association for Computing Machinery (ACM) The Institute of Electrical and Electronics Engineers (IEEE), Senior Member Private Sector Technology Group - Medicaid, Technical Advisory Group, 2013-2016