

Education

University of Chicago (Chicago, IL)

Ph.D. in Computer Science

Thesis: “Automated metadata extraction can make data swamps more navigable”

Graduated May 2022

GPA 3.89

University of Chicago

M.S. in Computer Science

Graduated Dec 2018

Macalester College (St. Paul, MN)

B.A. in Applied Mathematics and Statistics

Minors in Computer Science and Economics

Graduated May 2016

Experience

Oak Ridge National Lab, U.S. Department of Energy (Oak Ridge, TN)

Jun. 2022 – Present

Associate Research Scientist, Workflows

- Spearhead a comprehensive research portfolio at the intersection of automated metadata extraction, distributed workflow orchestration systems, and science application usability.
- Lead the research and engineering efforts for the Zambeze decentralized workflow orchestration system, which can execute complex machine learning workflows on heterogeneous compute resources. Enable scientists to orchestrate multiple workflows at the ‘push of a button’.
- Design and build novel API for high-performance computing (HPC) resources, leading a team of 10 research and technical staff. Coordinate with other DOE labs for general adoption with their resources.
- Build and help scientists adopt an information retrieval toolkit to enhance their indexing of scientific data, incorporating both off-the-shelf and custom-trained statistical learning algorithms for personal information (PII) detection, file type classification, object recognition, and language processing (NLP).
- Collaborated as co-PI on a grant proposal, successfully securing over \$300,000 in funding for cross-facility workflow orchestration research and development.

University of Chicago (Chicago, IL)

Jun. 2016 – Nov. 2022

Ph.D. Research Assistant

- Created and led research efforts on the first scalable metadata extraction system designed specifically to promote generalizable research data navigation on leadership-scale research repositories. The system scaled to over 8,000 concurrent workers and could intelligently distribute tens of millions of files across a continuum of HPC and edge resources to achieve even higher throughput.
- Devised, trained, and evaluated multi-output machine learning models for file type identification to increase the quality of metadata outputs; this solution had a 35% improvement in F1-score for classifying research data compared to the leading *libmagic* tool.
- Conceived and conducted a comprehensive user study involving national lab scientists to assess the efficacy of automatically extracted metadata in enhancing scientific data navigation. Our findings revealed that, irrespective of the search interface used, such metadata significantly improved the speed, accuracy, and confidence of scientists in navigating their repositories compared to traditional methods.
- Researched and implemented container scheduling strategies and pipelines within the funcX function-as-a-service (FaaS) platform to enable remote execution of extraction functions.
- Executed and evaluated extraction workflows on diverse scientific data sets (e.g., batteries, correlation spectroscopy, climate, materials, personal cloud).

- Mentored 14 undergraduate and high school students in data science and computing systems, leading many to pursue positions, college majors, and research publications in the field.

Principal Lecturer

- Curated and delivered an undergraduate computer science course focusing on foundational concepts such as algorithms, data structures, and memory management. Employed the C programming language as an effective hands-on learning instrument for a cohort of 40 students.
- Contributed as a Teaching Assistant for the professional master's program courses: Cloud Computing (2018) and Distributed Systems (2019-2023). Engaged in student support through weekly office hours, delivered lectures in the professor's absence, and managed grading responsibilities.

MarketRebellion (Chicago, IL)

Jun 2020 – Mar 2022

Data Science Consultant

- Constructed, researched, and maintained financial sentiment workflows for a top a financial education and trading firm.
- Customized the VADER toolkit to improve sentiment prediction accuracy of finance-related social media posts by 55%, as measured by user survey.
- Architected massive-scale cloud systems capable of quickly responding to tens of millions of users' requests per day.
- Scraped live and historical market and sentiment data from many APIs, including Reddit, Twitter, Binance, Voyager, and NASDAQ.
- Generated and maintained low-latency dashboards for CEO, analysts, and social media representatives.

IBM Research (Almaden, CA)

May 2018 – Oct 2018

Graduate Research Intern

- Created and implemented microservice-based genome assembly workflows in IBM's Functional Genomics Platform.
- Exercised working knowledge of genome assembly tools: Trimmomatic, Bowtie2, FLASH, SPAdes, QUAST.
- Constructed full user interface for genome file uploads and pipeline monitoring.
- Communicated with 8 biologists and bioinformaticians regarding system design.
- Programmed and implemented integration tests for continued quality assurance for IBM's clients.

NightWare, Inc. (Hopkins, MN)

Sep. 2015 – Dec 2020

Founder

- Develop a wearable application that leverages statistical learning methods to effectively monitor and treat PTSD survivors' traumatic nightmares, improving their quality of life and well-being.
- Create a feature library and automated model runner to asynchronously train and tune hyperparameters for hundreds of time series forecasting models across millions of feature combinations.
- Assist research team in obtaining FDA clearance in 2021 as a "breakthrough medical device".
- Educate the public about NightWare through various media outlets, including BBC, PBS, and NPR.

Publications and Patents ([student advisees in blue](#))

T. Skluzacek, K. Chard, I. Foster. "Can Automated Metadata Extraction Make Scientific Data More Navigable?" IEEE 19th International Conference on e-Science (e-Science). 2023.

R. Souza, **T. Skluzacek**, S. Wilkinson, M. Ziatdinov, R. Ferreira da Silva. "Towards Lightweight Data Integration using Multi-workflow Provenance and Data Observability". IEEE 19th International Conference on e-Science (e-Science). 2023.

T. Skluzacek, K. Chard, I. Foster. “Automated metadata extraction: challenges and opportunities”. IEEE 18th International Conference on e-Science (e-Science). 2022.

T. Skluzacek, **E. Hsu**, **M. Chen**, K. Chard, I. Foster. “Models and Metrics for Mining Meaningful Metadata”. Proceedings of the International Conference of Computational Science (ICCS). 2022.

T. Skluzacek, **R. Wong**, Z. Li, R. Chard, K. Chard, I. Foster. “A Serverless Framework for Distributed Bulk Metadata Extraction”. Proceedings of High-Performance Parallel and Distributed Computing (HPDC). 2021.

Z. Li, R. Chard, L. Ward, K. Chard, **T. Skluzacek**, ..., M. Franklin, I. Foster. “DLHub: Simplifying publication, discovery, and use of machine learning models in science”. Journal of Parallel and Distributed Computing (JPDC). 2020.

R. Chard, Y. Babuji, Z. Li, **T. Skluzacek**, . . . , K. Chard, I. Foster. “funcX: a Federated Function Serving Fabric for Science”. Proceedings of High-Performance Parallel and Distributed Computing (HPDC). 2020.

T. Skluzacek, R. Chard, **R. Wong**, Z. Li, Y. Babuji, L. Ward, B. Blaiszik, K. Chard, I. Foster. “Serverless Workflows for Indexing Large Scientific Data”. 5th International Workshop on Serverless Computing (WoSC). 2019.

T. Skluzacek. “Dredging a Data Lake: Decentralized Metadata Extraction”. International Middleware Conference Doctoral Symposium. 2019.

R. Chard*, **T. Skluzacek***, et al., K. Chard, I. Foster. “Serverless Supercomputing: High Performance Function as a Service for Science”. ArXiv Preprint: <https://arxiv.org/abs/1908.04907>. 2019. (*: co- first authors)

T. Skluzacek, **R. Kumar**, R. Chard, G. Harrison, **P. Beckman**, K. Chard, I. Foster. “Skluma: An extensible metadata extraction pipeline for disorganized data”. IEEE 14th International Conference on e-Science (e-Science). 2018.

P. Beckman, **T. Skluzacek**, K. Chard, I. Foster “Skluma: A Statistical Learning Pipeline for Taming Unkempt Data Repositories”. International Conference on Scientific and Statistical Database Management (SSDBM). 2017.

T. Skluzacek. Full Patent. “Method for Interrupting Night Terrors During Sleep”. 2017. PTO#20200376231

T. Skluzacek, K. Chard, I. Foster. “Klimatic: A Virtual Data Lake for Harvesting and Distribution of Geospatial Data”. In 1st Joint International Workshop on Parallel Data Storage and data Intensive Scalable Computing Systems (PDSW-DISCS). 2016.

Selected Presentations

ParslFest 2023 <i>“Parsl in Zambeze: rafting down your distributed science river.”</i>	Oct. 2023
World Usability Day – Cloudberry Creative; New York City, New York <i>“PTSD and Digital Therapeutics: designing wearable applications for vulnerable populations.”</i>	Nov. 2022
Oak Ridge National Lab --- National Center for Computational Science <i>“Automated metadata extraction to make data swamps more navigable.”</i>	Mar. 2022
Sandia National Labs Invited Talk <i>“Metadata Extraction for the Battery Modeling Community.”</i>	Feb. 2022
UChicago Data Science Institute Rising Stars Summit (Chicago, IL) <i>“Enabling Data Utility Across the Sciences.”</i>	Nov. 2021

Parsl and funcX Fest (Virtual) "A Serverless Framework for Distributed Bulk Metadata Extraction."	Oct. 2020
CERES Center for Unstoppable Computing Annual Summit (Virtual and Chicago, IL) "Draining the Data Swamp -- Funded Project Updates."	Mar. 2020
Data Science Institute Summer Research Lab "Making the most of your research summer."	May 2019
SC19 Lightning Talk Series (Dallas, TX; Denver, CO) "Container Orchestration for Genome Sequencing Pipelines." "Serverless Metadata Extraction Workflows for Better Science."	Nov. 2019
IOTFuse "NightWare: Transformative Research on the Human Body."	Sep. 2017

Awards

Data Science Institute (University of Chicago) Rising Star Awardee	2021
Illinois Fifty for the Future	2018
Outstanding Graduate Research Award – CERES Center for Unstoppable Computing	2017
Nominet Trust 100 Award (myBivy)	2016
Minnesota Rising Technology Star	2015
MentorMate MobDemo 2016 Champion	2015
HackDC: mobile solutions for PTSD Grand Champion	2015

Service

Data4All Curriculum Development 2020 – 2022

- Designed and delivered a data science curriculum based on COVID-19 and 19th-century London cholera epidemic. Ensured that lessons are entertaining, beginner-friendly, and teach state-of-the-art data science tools and practices.
- Taught south side Chicago high school students via Python, Jupyter notebooks, and Kepler.

Program committee

- SC (2023), HPDC (2023), ICPP (2023), PEARC (2023), WORKS (2022—2023), SciPy (2021)

Research paper reviewer (*outside of program committee responsibilities*)

- Cluster, eScience, EuroPar, FGCS, Gateways, HPCC

SC Lead Student Volunteer 2018—2019

Judge for DOE Science Bowl 2019

Funding

Co-PI: “Multi-Workflow Orchestration and Lightweight Integrated Data Analysis” 2023
Oak Ridge National Lab Directed Research and Development Award \$325,000 over 3 years

Selected Media

Henry Ford’s Innovation Nation with Mo Rocca on CBS: 2022
[“NightWare.”](#)

Wired Magazine 2020
[“How a vibrating smartwatch could be used to stop nightmares.”](#)

NPR: 2020
[“FDA Approves Smart Watch Application To Help Those With PTSD.”](#)

CNN: 2020
[“The son of an Iraq War Veteran designed an app to stop his dad’s PTSD nightmares.”](#)

PBS Almanac: 2016
[“An app that combats PTSD”](#)