DataEd'23 - 2nd International Workshop on Data Systems Education
Bridging Education Practice with Education Research

Aivaloglou, Efthimia; Fletcher, George; Miedema, Daphne

Citation (APA)
Aivaloglou, E., Fletcher, G., & Miedema, D. (2023). DataEd'23 - 2nd International Workshop on Data Systems Education: Bridging Education Practice with Education Research. In *SIGMOD 2023 - Companion of the 2023 ACM/SIGMOD International Conference on Management of Data* (pp. 313-314). (Proceedings of the ACM SIGMOD International Conference on Management of Data). Association for Computing Machinery (ACM). https://doi.org/10.1145/3555041.3590823

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

DataEd'23 - 2nd International Workshop on Data Systems Education: Bridging Education Practice with Education Research

Efthimia Aivaloglou
e.aivaloglou@tudelft.nl
Delft University of Technology
Center for Education and Learning
Delft, the Netherlands

George Fletcher
g.h.l.fletcher@tue.nl
Eindhoven University of Technology
Artificial Intelligence and Data
Engineering (AIDE) Lab
Eindhoven, the Netherlands

Daphne Miedema
d.e.miedema@tue.nl
Eindhoven University of Technology
Artificial Intelligence and Data
Engineering (AIDE) Lab
Eindhoven, the Netherlands

ABSTRACT

Interest in data systems education is increasing, especially with the rise in demand for well-trained and re-trained data scientists. The database and the computing education research communities have complementary perspectives and experiences to share with each other. The DataEd workshop is organized as a dedicated venue for these communities to come together to share findings, cross-pollinate perspectives and methods, and shed light on opportunities for mutual progress in data systems education. In the DataEd workshop, we will present and discuss data management systems education experiences and research via keynotes and paper and poster presentations.

CCS CONCEPTS

• **Social and professional topics** → **Computing education**; • **Information systems** → **Structured Query Language**.

KEYWORDS

Data Systems, Databases, Education

1 MOTIVATION

In this abstract we describe the second edition of the DataEd workshop to be held in conjunction with SIGMOD/PODS 2023 in Seattle, Washington (USA) on 23 June 2023. The DataEd workshop took place for the first time at SIGMOD 2022 [2, 6]. The focus of DataEd is on data systems education, which is foundational in a variety of programs such as computer science, data science, and information systems and science. A continual focus since the 1970s in the database research community is the place in curricula and best practices for teaching data systems concepts. This important conversation is particularly lively in recent years given the rise of data science [3–5, 7–9]. Open questions here include: What counts as data systems education? Which student audiences (data science majors, CS majors, business majors, pre-college, professional, etc.) get exposure to particular components of data systems education?

There is also a long tradition in the CS Education community on research into how students learn data systems concepts. With the increasing focus on data in the past decade, there is renewed focus on data systems in education research. Both the DB and CS Education communities, and adjacent communities, e.g., in Statistics Education, have complementary perspectives and experiences to share with each other, yet there has been relatively little interaction between these communities. There is much to be gained by bringing the communities more closely together: to share findings, to cross-pollinate perspectives and methods, and to shed light on opportunities for mutual progress. To our knowledge, there currently does not exist any other dedicated venue for the presentation and discussion of data systems education research.

With the DataEd workshop series, we aim to take an important step towards building these bridges between the DB community and the broader computer and data science education research communities. The goal is to start conversations toward shared research agendas in data education.

2 EXPERIENCES FROM THE FIRST DATAED

DataEd 2022 was a successful workshop, attended by a total of 41 participants, that had a strong focus on encouraging interaction among the participants. It took place as a full-day workshop consisting of:

- Two keynote talks, the first by Kathi Fisler (Brown University) on Data-Centricity: Rethinking Introductory Computing to Support Data Science, and the second by Julia Stoyanovich (NYU Tandon School of Engineering) on Teaching Responsible Data Science,
- ten paper presentations & discussions,
- an industry panel discussion on industry perspectives on education and training for emerging roles in data organized by Juan Sequeda (data.world), with panelists Sarah Krasnik (independent) and Emilie Schario (Amplify), and
- a discussion session on topics prioritized by the attendees, including curriculum placement & content of data systems topics and assessment types.

The workshop gave participants insight into data systems course and curriculum design, learning instruments, tools, and practices,

ethics and responsibility, formative and summative assessment, and industry perspectives on data management knowledge and skills. The workshop's main results, observations, and emerging research directions have been reported in detail [1].

3 SCOPE AND TOPICS OF INTEREST FOR THE SECOND DATAED

DataEd explores the broad area of data systems education: the teaching and learning of databases/data management/data systems topics, ranging across the whole field, from classical topics (such as physical design, query optimization, data modeling, data integration, visual analytics, and query languages) to contemporary topics (such as ML & AI for data management systems, data management for ML & AI, very large data science applications/pipelines, and responsible data management). Topics of interest include:

- teaching practices and instructional approaches supporting data systems education and training in any context; e.g., higher education, non-CS majors, K-12, informal learning, lifelong and distance learning, professional training
- learnability of query languages and tools
- data systems course and curriculum design, and knowledge and skills requirements on data systems
- learning sciences work in the data systems content domain, psychology of querying and data modeling
- measurement instruments, evaluation, and assessment in data systems education
- technology enhanced learning, intelligent tutoring in data systems, and learning technologies and tools that support data systems education
- integration of data management topics in other subjects in the curriculum (such as software engineering and ML & AI)
- equity, diversity, inclusion in data systems education
- ethical and responsible data management course and curriculum design.

4 ORGANIZATION AND PROGRAM

DataEd 2023 takes place as a full-day workshop consisting of (1) keynote talks, (2) paper presentations & discussion, and (3) a poster session. Information on the program and the accepted papers can be found at the workshop website [6].

4.1 Invited keynote speakers

Two keynote speakers will introduce data systems education, each from the perspective of their research discipline.

Sourav Bhowmick is an Associate Professor at the School of Computer Science and Engineering (SCSE), Nanyang Technological University, Singapore. His core research expertise is in data management, human-data interaction, and data analytics. His research has appeared in premium venues such as ACM SIGMOD, VLDB, and VLDB Journal. He is a co-recipient of Best Paper Awards in ACM CIKM 2004, ACM BCB 2011, and VLDB 2021. He is also a co-recipient of the 2021 ACM SIGMOD Research Highlights Award. Sourav is serving as a member of the SIGMOD Executive Committee, a regular member of the PVLDB advisory board, and a co-lead in the committee for Diversity and Inclusion in Database Conference Venues. He is a co-recipient of several service awards including

VLDB Service Award in 2018, Distinguished AE Award in SIGMOD 2021 and VLDB 2022, and Distinguished Reviewer Award in 2020. He is the inventor of CLOSET. Sourav was inducted into Distinguished Members of the ACM in 2020. He is a strong advocate of research that directly or indirectly impacts end users.

Toni Taipalus is a teacher and a researcher at the University of Jyväskylä, Finland. He completed his Ph.D. in information systems, focusing on query language education. Currently, he is bridging the gap between database management systems and human-computer interaction with the goal of facilitating data systems education.

4.2 Workshop organization

DataEd is organized by *Fenia Aivaloglou* (Delft University of Technology), *George Fletcher* (Eindhoven University of Technology), and *Daphne Miedema* (Eindhoven University of Technology). The workshop's advisory board members supported setting the scope of the workshop and provided valuable feedback in all stages of its preparation: *Michelle Hoda Wilkerson* (Graduate School of Education, University of California, Berkeley), *Zachary Ives* (University of Pennsylvania), *Shriram Krishnamurthi* (Brown University), *Juan Sequeda* (data.world), and *Julia Stoyanovich* (NYU Tandon School of Engineering).

We appreciate the efforts of the program committee in reviewing the submissions, and thank them for their valuable support: *Alireza Ahadi* (University of Technology Sydney), *Abdussalam Alawini* (University of Illinois Urbana-Champaign), *Matthew Beckman* (Penn State University), *Susan Davidson* (University of Pennsylvania), *Karen Davis* (Miami University), *Lorena Etcheverry* (Universidad de la Republica), *Alan Fekete* (University of Sydney), *Paul Groth* (University of Amsterdam), *Lukas Höper* (Paderborn University), *Bill Howe* (University of Washington), *HV Jagadish* (University of Michigan), *Shriram Krishnamurthi* (Brown University), *Hui Li* (Xidian University), *Michael Liut* (University of Toronto Mississauga), *George Obaido* (University of California, Berkeley), *Rachel Pottinger* (University of British Columbia), *Alexandra Poulovasillis* (University of London), *Stefanie Scherzinger* (University of Passau), *Monique Snoeck* (KU Leuven), *Daniele Traversaro* (Universita degli Studi di Genova), and *Thomas Zeume* (TU Dortmund),

REFERENCES

[1] E. Aivaloglou, G. Fletcher, M. Liut, and D. Miedema. 2023. Report on the First International Workshop on Data Systems Education (DataEd'22). *ACM SIGMOD Record* 51, 4 (2023), 49–53.
[2] E. Aivaloglou, G. Fletcher, and D. Miedema (Eds.). 2022. *Proceedings of the 1st ACM SIGMOD International Workshop on Data Systems Education: Bridging Education Practice with Education Research, DataEd@SIGMOD 2022, 17 June 2022, Philadelphia, PA, USA.* ACM. https://doi.org/10.1145/3531072
[3] A. Fekete, J. Kay, and U. Röhm. 2021. A Data-Centric Computing Curriculum for a Data Science Major. In *Proceedings of SIGCSE '21.* ACM, 865–871.
[4] M. J. Franklin and D. L. Nicolae. 2021. A Perspective on Designing Undergraduate Education in the New Discipline of Data Science. *Harvard Data Science Review* 3, 2 (7 6 2021). https://hdsr.mitpress.mit.edu/pub/jbpym4zp
[5] B. Howe, M. Franklin, L. Haas, T. Kraska, and J. Ullman. 2017. Data Science Education: We're Missing the Boat, Again. In *2017 IEEE ICDE.* 1473–1474.
[6] DataEd Initiative. [n. d.]. http://dataedinitiative.github.io.
[7] Z. G. Ives, R. Pottinger, A. Kumar, J. Gehrke, and J. Giceva. 2021. The Future of Data(Base) Education: Is the "Cow Book" Dead? *Proc. VLDB Endow.* 14, 12 (2021).
[8] A. Lewis and J. Stoyanovich. 2021. Teaching responsible data science: Charting new pedagogical territory. *International Journal of Artificial Intelligence in Education* (2021), 1–25.
[9] A. F. Wise. 2020. Educating data scientists and data literate citizens for a new generation of data. *Journal of the Learning Sciences* 29, 1 (2020), 165–181.