

Diversity and Inclusion Activities in Database Conferences: A 2021 Report

<https://dbdni.github.io>

Sihem Amer-Yahia, Yael Amsterdamer, Sourav S. Bhowmick, Angela Bonifati, Philippe Bonnet, Renata Borovica-Gajic, Barbara Catania, Tania Cerquitelli, Silvia Chiusano, Panos K. Chrysanthis, Carlo Curino, Jérôme Darmont, Amr El Abbadi, Avrilia Floratou, Juliana Freire, Alekh Jindal, Vana Kalogeraki, Georgia Koutrika, Arun Kumar, Sujaya Maiyya, Alexandra Meliou, Madhulika Mohanty, Felix Naumann, Nele Sina Noack, Fatma Özcan, Liat Peterfreund, Wenny Rahayu, Wang-Chiew Tan, Yuanyuan Tian, Pinar Tözün, Genoveva Vargias-Solar, Neeraja Yadwadkar, Meihui Zhang

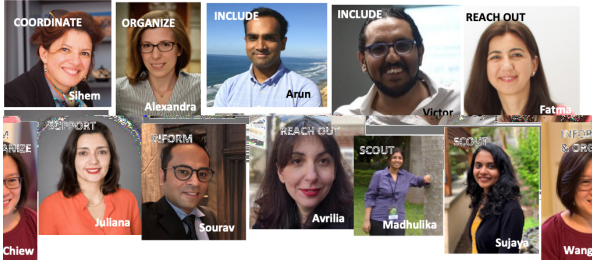


Figure 1: Core Members of the D&I@DB initiative. For details refer to <https://dbdni.github.io/>

1. THE D&I@DB INITIATIVE

Diversity and Inclusion (D&I) are core to fostering innovative thinking. Existing theories demonstrate that to facilitate inclusion, multiple types of exclusionary dynamics, such as self-segregation, communication apprehension, and stereotyping and stigmatizing, must be overcome [11]. A diverse group of people tends to surface different perspectives, which help to understand and address D&I. Fostering D&I in research communities must address issues related to inclusive interpersonal and small group dynamics, rules and codes of conduct, increasing diversity in under-represented groups and disciplines, and organizing D&I events, and long-term efforts to champion change [15].

Why D&I@DB? It is our responsibility as a community to [ensure that attendees of DB conferences feel included irrespective of their scientific perspective and personal background.] The D&I team is dedicated to providing leadership to help our community achieve that goal by [advising D&I chairs at individual DB conferences, serve as a memory and buffer between DB conferences, build an agreed-upon vision, and commit to working together to devise a set of measures for achieving diversity and inclusion.] That is pursued via actions led by our core members (Figure 1): **REACH OUT** collects data and experiences from our community. **INCLUDE** monitors and recommends inclusion efforts. **ORGANIZE** focuses on in-conference organization efforts, such as adopting a code of conduct. **INFORM** communicates through

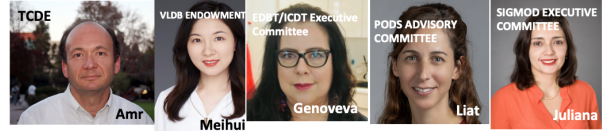


Figure 2: Liaisons members of the D&I@DB initiative. For details refer to <https://dbdni.github.io/>

various channels. **SUPPORT** coordinates D&I support from executive bodies and sponsors. **SCOUT** collates D&I efforts from other communities. **COORDINATE** manages all actions. Everyone in the DB community is welcome to join the initiative.

The D&I@DB initiative also includes liaisons of individual executive bodies all of which are committed to the promotion of D&I in their conferences (Figure 2).

What did we achieve this year? [This year's focus was on raising awareness through the introduction of special D&I sessions and collecting statistics and wishes through surveys. Some active actions were also taken at some of our conferences, i.e., making deliberate efforts to address diversity and gender balance of conference officers, student volunteers, and invited speakers, and drafting and displaying a Code of Conduct.] We also published tips on inclusive writing and on video captioning and developed and deployed tools that facilitate our D&I efforts.

D&I statistics. Table 1 reports the results of a voluntary survey made available to conference attendees. In 2022, we will advertise the survey earlier and link it to registration (as in MDM) to make it available to all.

Conference	# Attendees	# Respondents
EDBT 2021	437	77
ICDE 2021	612	39
SIGMOD 2021	1819	86
VLDB 2021	1045	142
MDM 2021	332	332

Figures 3, 4, 5, and 6 show the distribution of age, gender, race/ethnicity, and professional experience of respondents. We also asked the respondents to tell us

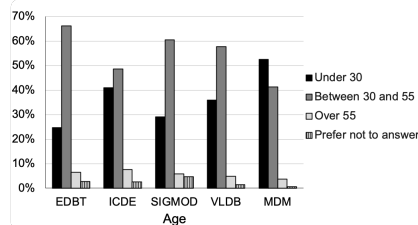


Figure 3: Age at 2021 DB Conferences.

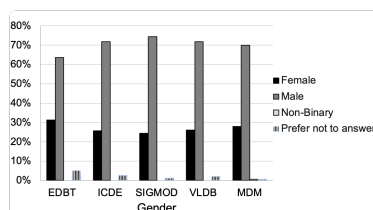


Figure 4: Gender at 2021 DB Conferences.

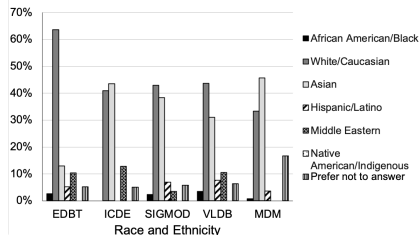


Figure 5: Race and Ethnicity at 2021 DB Conferences.

the top three topics they wanted to see discussed at D&I sessions. Figure 7 shows an overwhelming convergence on five topics, with work-life balance, research (finding topics, setting goals, making progress, determining success), and career paths being selected by the most number of respondents. Other important topics (selected by about 20% of the respondents) included challenges of remote work, writing good research proposals, and inclusive culture in data management. We will continue to encourage participation in the survey, collect more data and use it to track how we are improving on D&I. For instance, PC chairs will be encouraged to present these stats and their evolution from year to year in the same way as paper submission stats. Our aim is to integrate our tools such as the gender analysis tool described in Section 3 to enrich the stats.

2. D&I@DB CONFERENCES 2021

D&I@EDBT/ICDT. EDBT/ICDT 2021 organized the inaugural event of the D&I@DB initiative with a debrief aiming to *Inform and Engage*. The debrief took place in a plenary session and was well-attended. It introduced the general goals, plans, and actions of D&I@DB in general and at EDBT/ICDT 2021 in particular. Two specific actions were piloted, namely pre- and post-conference questionnaires, and a tutorial on preparing inclusive videos. In the debrief, we explained the necessity of gathering data and experiences from the community to drive our actions. We also provided guidelines to session chairs on how to conduct D&I sessions.

D&I@ICDE. The D&I chairs of ICDE 2021 invited Geraldine Fitzpatrick, a Professor of Technology Design and Assessment who leads the Human Computer Interaction group at TU Wien, Austria. The talk entitled “Diversity matters matter” touched upon different notions and perceptions of diversity and bias, and highlighted their complexities. An introduction to the D&I initiative was also given. The talk explained the mission and actions and then focused on the automatic detection of COIs [12] that was performed this year at ICDE.

D&I@MDM. The D&I chairs of MDM 2021 view peer education as an important step towards achieving Diversity and Inclusion at conferences. With this in mind, they organized Education Day that began and ended with D&I events. The first event was a keynote talk

on Federated Learning for Mobile Sensing Data. Cristian Borcea (Associate Dean for Strategic Initiatives, New Jersey Institute of Technology) provided tips for writing inclusive presentations using his own slides as examples. The second event was a D&I debrief, which discussed the two MDM-specific questions on the pre-conference D&I questionnaire, (1) Does research in mobile data management & IoT consider D&I aspects? and (2) How would you assess the current climate for women and under-represented minorities? Two invited talks followed the debriefing and closed the Education Day. The first talk was on Bias and Representation in Sociotechnical Systems (Danaë Metaxa, Stanford University) and the second on Disrupting Models of D&I and Undergraduate Research: Inside the iSchool Inclusion Institute (Kayla Booth, University of Pittsburgh).

The authors were provided with guidelines for preparing inclusive camera-ready papers, talks and videos. To recognize students’ efforts, MDM 2021 granted two *Best D&I Presentation* awards sponsored by the IEEE TCDE. Presentations were judged based on gender neutrality (in narrative and in figures), avoiding exclusionary examples, and illustrating color-blind inclusive graphs.

D&I@SIGMOD. SIGMOD 2021 D&I efforts were centered around educating authors and reviewers on inclusive writing, and hearing the perspectives of under-represented groups. The D&I chairs provided detailed instructions [6] on how to create inclusive papers and presentations and asked PC members to evaluate whether submitted papers followed inclusive writing instructions.

We organized round tables targeting under-represented groups: (1) a student/post-doc panel where the new generation of DB researchers provided their perspectives on academia, challenges during PhD and experiences at DB conferences, (2) a Women-in-DB panel where women shared their opinions on how to make conferences more inclusive, (3) two round-tables featuring senior and junior members of our community from both academia and industry discussing their experiences with failures and impostor syndrome, and (4) a panel discussing the D&I@DB initiative.

D&I@VLDB. VLDB 2021 D&I events were as follows. The first was a keynote titled “What relationship between an author’s ethics and their work?” by Gisèle Sapiro, research director at Centre national de la recherche scientifique (CNRS) and professor at the École des hautes études en sciences sociales. Gisèle provided an in-depth

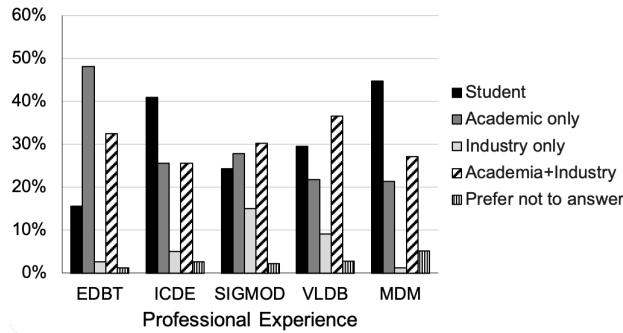


Figure 6: Professional experience at 2021 DB Conferences.

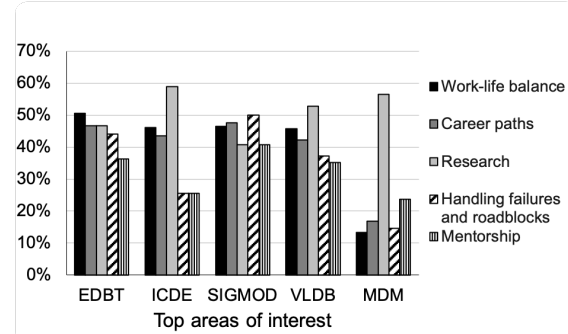


Figure 7: Top areas of interest at 2021 DB conferences.



Figure 8: D&I Chairs of 2021-2023 DB Conferences

analysis of the controversy following the Ullman Turing Award, discussed some cases of scandals, from Heidegger to Polanski and Handke, and recommended drafting a Code of Ethics for our community.

The second event was a panel titled “Are we D&I?”. The panelists debated the question of conducting research on D&I. The audience was polled during the panel with three questions: (1) In your opinion, what is and what is not D&I? Answers were Inclusive paper writing, Video captioning, Fund under-represented communities, Research on data ethics; (2) Do you think we are more/less D&I than other research communities? to which the majority answered “less”, and (3) How can we be more D&I? Answers included Session on personal experiences (good and not so good ones), Actions toward students, and Modifying review forms.

The third event was a “Women in Databases: What does it mean?” round table. The coffee style session invited participants to discuss anything from the impact of Covid-19 on our lives, to how to choose a promising research topic or establish a fruitful collaboration.

D&I@ADBS. ADBIS 2021 promoted the D&I Code of Conduct [5] as part of the effort to bring awareness to the goals and actions of the D&I@DB initiative. The conference has been organized as a hybrid event and, to make it more inclusive, the organizers provided a room for participants with children.

D&I@WISE. WISE 2021 first action was to enforce, whenever possible, a balance of gender representation

as well as country of origin in conference officers: PC selection, session chairs, student volunteers, and award sessions. The approach taken was being more inclusive rather than quota-based, such that to ensure that we remain fair in the process. This was a deliberate effort made throughout the whole planning and preparation of the conference by the organising team, and it would be a continuous improvement process where diversity and gender balance will play a key role in the organisation of upcoming WISE conferences. The program included a special session on raising awareness of D&I, and invited suggestions for the D&I agenda at future editions. One specific outcome is the need for a plenary speaker to discuss the D&I in our research community and the role every researcher can play to promote D&I.

D&I@SoCC. A key action at SoCC’21 was the involvement of the D&I chair in all conference-related decisions, including selecting PC chair, PC members, and keynote speakers with an effort to select conference officers from under-represented groups, e.g. females. The program for SoCC also worked towards ensuring that keynotes, and paper talks are scheduled at the best feasible times for as many attendees as possible.

We posted D&I guidelines in writing and presentations as part of camera-ready instructions. We also posted the standard ACM policy against harassment and ACM processes for reporting unacceptable behavior. We enabled closed captioning on the client side with support from Microsoft Teams. To help students feel more connected and included, we ran a mentorship program offering them explicit help to interact with senior attendees. Additionally, the D&I chair executed the students travel scholarship program with the support from ACM. Lastly, the D&I chair organized and moderated a panel about “Research environment and Diversity in Cloud Computing” that focused on the topics that were ranked the highest in the D&I survey taken by the attendees at registration. The topics ranged from finding research problems, choosing career paths, handling failures and roadblocks to women in cloud computing.

3. TOOLS

There is a continuous ongoing effort towards building data-driven tools with the broad aim of improving con-

ference attendance and review quality for our venues. We described two tools deployed this year.

COI tool. Our professional societies, such as ACM, clearly dictate that conflicts of interest (COI) in the publication process are to be avoided in our venues [9]. Analysis of review data over multiple venues reveals that a reviewer who has COI with an author (referred to as *COI reviewer* and *COI author*, respectively) of a submission often tends to ignore or gloss over in his/her review critical issues with the submission that may be raised by his/her non-COI counterparts. Clearly, this adversely impacts review quality. Furthermore, our analysis shows that some COI reviewers accept *only* submissions of COI authors assigned to him/her while rejecting all other submissions. In particular, when COI authors are from the same country or ethnicity as the COI reviewer and the authors of rejected submissions have high diversity w.r.t these attributes, such reviewing behavior may collide with our D&I goals.

Double-blind reviewing is not designed to address these concerns. The goal of double-blind is to mitigate bias that may arise due to the ‘stature’ of authors or institutions, not to address COI [17]. Additionally, anecdotal evidence from PC chairs [16] indicates the existence of collusion rings that subvert the double-blindness by creating covert communication channels between COI reviewers and COI authors.

To facilitate detection and management of COIs during a review process, we have built CLOSET to automatically detects unreported/overlooked COIs [12]. CLOSET has been deployed for SIGMOD, VLDB, ICDE since 2020. Our initial analysis reveals that more than 20% of the submissions (conservatively) in some venues have one or more unreported COIs. That is, the problem of unreported COI is not insignificant in our venues.

Gender analysis tool. As an extension of the CMTStat tool, which automatically analyzes and visualizes conference statistics, such as acceptance rates and author country distribution [1], we are developing tools to import gender annotations of authors and perform analytics including growth in diversity for different Computer Science fields, acceptance/rejection rates based on gender, and cross-gender collaboration frequency. We have initial results on published papers across CS conferences and communities [13]. With GAP (Gender Analysis for Publications), we are currently leveraging the gender annotation on Gender API [4], a commercial tool already used for gender distribution in CS [18], and integrating manual annotations. The aim is to empower PC chairs to present enriched statistics at conferences [3].

4. GOING FORWARD

A number of important matters are being framed. We are currently discussing our bylaws with the aim of involving as many people as possible in this initiative. We also need to position D&I with respect to DBCares [2] a joint effort by the VLDB Endowment and the SIGMOD Executive Committee whose purpose is to foster zero tolerance for abuse, discrimination, or harassment. ACM has a policy, including a definition of abuse, discrimination and harassment, a mechanism for reporting

grievances, and consequences for validated grievances. Both SIGMOD and VLDB follow this policy. DBCares provides advice from respected and approachable members of the community with the goal of making reporting easier and less intimidating.

We have been scouting for Codes of Ethics from venues in other fields [10, 14] to identify unique aspects from each of them to integrate into our community. Additionally, we are looking into organizing committees in other conferences [7, 8]—particularly those that are targeted towards D&I efforts such as Accessibility Committee, Family Committee, and Sustainability Committee.

Reaching out to traditionally under-represented groups is an essential step in strengthening the influence of our community and attracting young scientists to leading conferences. Giving access to courses, material, as planned by SIGMOD, will be an efficient way to connect with universities and institutions that do not usually participate to major DB conferences. Our initial analysis showed that the Hispanic and Latin American community is particularly under-represented both as participants and authors. We have started to identify agencies in Latin American who can relay calls and funding opportunities for attending conferences. We will enhance these efforts by reaching out to (i) the Hispanic and Latin American community in the US and Canada through local associations; (ii) education managers of DB companies that have access to talented assets; (iii) under-represented countries in other regions, for example, the Caribbean, Asia Pacific and Africa.

Further, the D&I chairs of EDBT 2022 wish to promote a global perspective and understanding of Diversity and Inclusion, which includes promoting discussions on the variations in underrepresented groups in different countries and regions. Towards this end, the D&I chairs are planning a panel on “Diversity and Inclusion Perspectives Around the World” at EDBT/ICDT 2022.

The response from the community in various events in the 2021 conferences has been extremely positive and experiences shared across DB conferences has already enabled the scaling up of D&I activities in 2022 conferences. D&I@VLDB 2022 will focus on reaching out to other scientific communities (Web and IR, Distributed Systems, CHI) and organizing an event on what we learned so far and how we can do better at conducting D&I research. D&I@SIGMOD 2022, MDM 2022, AD-BIS 2022 and WISE 2022 plan to focus on mentoring activities that will extend beyond the main conferences.

We will encourage the wider adoption of automatic tools, such as for COI and Gender Analysis, along with extensive consent-based data collection to be combined with these tools¹. CLOSET is currently being extended to support data-driven PC formation with the goal of optimizing diversity, topic-related expertise, and COI-related risks. CLOSET was recently deployed in ACM/IEEE LICS 2022 (a premium venue in theoretical computer science community). There is also ongoing discussions with ACM on further development and deployment in venues beyond DB conferences.

¹Data collection has to comply with data protection rules, such as the European GDPR and alike.

5. REFERENCES

- [1] CMTStat Tool. https://github.com/HPI-Information-Systems/cmt_statistics_tool.
- [2] DB Cares. <https://sigmod.org/sigmod-policies/dbcares-policy>.
- [3] GAP - Gender Analysis for Publications. <https://github.com/HPI-Information-Systems/GenderAnalysis>.
- [4] Gender API. <https://gender-api.com/>.
- [5] D&I Code of Conduct. <https://dbdni.github.io/pages/codeofconduct.html>, 2021.
- [6] SIGMOD 2021 Paper Writing Instructions. https://2021.sigmod.org/calls_papers_inclusion_and_diversity.shtml, 2021.
- [7] NeurIPS 2021. 2021 organizing committee. <https://nips.cc/Conferences/2021/Committees>.
- [8] CHI 2022. Equity, justice, and access commitments. <https://chi2022.acm.org/2021/06/23/equity-justice-and-access-commitments>.
- [9] ACM. Code of ethics. <https://www.acm.org/publications/policies/conflict-of-interest>.
- [10] ACM. Code of ethics. <https://www.acm.org/code-of-ethics>.
- [11] Ruth Sessler Bernstein, Morgan Bulger, Paul Salipante, and Judith Y. Weisinger. From diversity to inclusion to equity: A theory of generative interactions. *Journal of Business Ethics*, 167(3):395–410, 2020.
- [12] Sourav S Bhowmick. Closet: Conflict of interest detection & management system. 2019. <https://personal.ntu.edu.sg/assourav/research/DARE/closet.html>.
- [13] Angela Bonifati, Michael J. Mior, Felix Naumann, and Nele Noack. How inclusive are we? An analysis of gender diversity in database venues. *SIGMOD Record*, 50(4), 2021.
- [14] NeurIPS Foundation. Ethics guidelines. <https://nips.cc/public/EthicsGuidelines>.
- [15] R.Stuart Geiger, Orianna DeMasi, Aaron Culich, Andreas Zoglauer, Diya Das, Fernando Guardia, Kellie Ottoboni, Marsha Fenner, Nelle Varoquaux, Rebecca Barter, Richard Barnes, Sara Stoudt, Stacey Dorton, and Stéfan van der Walt. Best practices for fostering diversity and inclusion in data science: Report from the bids best practices in data science series. 02 2019.
- [16] Michael L. Littman. Collusion rings threaten the integrity of computer science research. *Commun. ACM*, 64(6):43–44, 2021.
- [17] Richard T. Snodgrass. Single-versus double-blind reviewing: an analysis of the literature. *SIGMOD Rec.*, 35(3):8–21, 2006.
- [18] Lucy Lu Wang, Gabriel Stanovsky, Luca Weihs, and Oren Etzioni. Gender trends in computer science authorship. *Communications of the ACM*, 64(3):78–84, 2021.