

Diversity and Inclusion Activities in EGC – A 2022 Report

Siham Amer-Yahia (CNRS, Univ. Grenoble Alpes), Angela Bonifati (LIRIS, Univ. Lyon 1), Cécile Favre (ERIC, Univ. Lyon 2), Elisa Fromont (IRISA, Univ. Rennes 1), Nicolas Labroche (LIFAT, Univ. Tours), Guy Melançon (LABRI, Univ. Bordeaux), Florence Sèdes (IRIT, Univ. Toulouse 3), Arnaud Soulet (LIFAT, Univ. Tours), Alexandre Termier (IRISA, Univ. Rennes 1)



Figure 1: The authors.

1. D&I@EGC

EGC (“Extraction et Gestion des Connaissances”¹ in French) started in 2001 and is the reference conference for the french community in Knowledge Extraction and Management (equivalent to the French KDD). The topics of EGC include Machine Learning, Knowledge Engineering and Representation, Data and Knowledge Reasoning, Data Mining and Analysis, Information Systems, Databases, Semantic Web and Open Data. The 2022 edition of the EGC conference brought together 219 attendees, 136 of which attended in person in Blois (France), and 83 attended remotely. Among them, there were 64 women (29.2%) and 93 students (42.5%).

We held the first D&I session at EGC. Diversity and Inclusion (D&I) is in line with the 17 UN SDG² (Sustainable Development Goals), and is central 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 [1]. Fostering D&I in individual conferences is important to promote exchanges and address issues related to inclusive interpersonal and small group dynamics, rules and codes of conduct, increasing diversity in under-represented groups and disciplines, and long-term efforts to champion change [5].

This document reports our discussions during the session and the lessons we learned going forward.

Audience polling. An anonymous polling was conducted with all participants (remote and local). 141 attendees par-

¹Website of the association organizing this annual event <http://www.egc.asso.fr/>

²<https://en.unesco.org/sustainabledevelopmentgoals>



Figure 2: The keyword cloud of the 84 answers to the question “How do you interpret D&I?”

anticipated. Figure 2 depicts a keyword cloud summarizing the 84 answers to the question “How do you interpret D&I?”. Table 1 reports some statistics with example answers. A first aspect of this session is that it raised awareness among attendees who either did not know the meaning of D&I (and what it encompasses, 7% as “not aware”) or did not think it was relevant to discuss (7% as “secondary”). The polling revealed a number of complementary ways of interpreting D&I such as equitable access (34%), under-represented group promotion (10%), population representation (8%) or activist effort (6%). This survey also shows that some diversities are more considered than others: gender (24%), social (12%), cultural (13%), handicap (8%) and age (1%).

2. GENDER ANALYSIS

While governments have addressed disability and accessibility, gender imbalance remains a poorly managed issue. Talking about diversity often induces a discussion on under-represented groups. The matter is different with gender gap, since women are not an under-represented group in society. Their under-representation in scientific fields, in particular in STEM (Science, Technology, Engineering, and Mathematics), is indicative of a social phenomenon. The quantification of this phenomenon is of particular interest in scientometrics.

While this section is dedicated to gender, other under-

Keyword	Statistics	Examples
Perception of D&I		
equity	29 (0,345)	“Everyone must have access to the same things and express themselves equally”; “Respect each other and equality for all”
under-representation	9 (0,107)	“Make visible people who are less visible”; “Include all under-represented groups without exception”
representativeness	8 (0,095)	“Better representation of the real world”; “Uniform distribution”
promote change and engagement	6 (0,071)	“Breaking establish practices”; “Sacrificing selection by skill to overcome discriminatory stereotypes”
not aware	6 (0,071)	“I don’t know what it is”; “kezako”
secondary	6 (0,071)	“A questionable problem”; “Something that has nothing to do with EGC. No interest.”
List of diversities		
gender	20 (0,238)	“Women participation, equity and inclusion in the field of technology more specifically in IT”; “Origins, gender identities, disabilities represented equitably”
social	10 (0,119)	“Children of workers (20% of the population) are 12% in bachelor’s degrees, 8% in master’s degrees and only 5% in PhD programs.”
cultural	11 (0,131)	“It is the fact of integrating into the teams more people of different genders (male / female / LGBTQ+...) of different ethnic origins (European, Asian, African-American...) but also to treat them with equity in all circumstances.”
handicap	7 (0,083)	“Inclusion of people with disabilities”; “In the broadest possible sense, i.e. not reduced to the problem of gender but also disability...”
age	1 (0,012)	“Encourage the inclusion of diverse profiles (on several criteria: gender, age, origin and so on)”

Table 1: Summary of the 84 answers to the question “How do you interpret D&I?”

represented groups should not be overlooked and should be explicitly discussed in future events.

2.1 Preamble on gender and statistics

Quantified statistics are useful to evaluate the representation of subgroups. However, producing statistics to examine diversity is not such a simple task to implement due to lack of data or to legal aspects related to data protection. For instance, assigning a gender label is a prerequisite for constructing statistics. This raises the question of this label and its modalities, beyond the gender binarity to ensure more inclusiveness. This label assignment is often based on the first names and analyzes in terms of gender in science are more specifically focused in analyzing the place of women.

2.2 Gender analysis in EGC

EGC celebrated its 20th anniversary in 2020. A challenge was launched by making available a certain amount of data from the conference: “20 years of history for what future?”. The winner proposed to adopt a social point of view of the conference data [8]. The contribution was to explore the place of women in the community, by adding a gender label to data. Usually, the gender analysis concerns the publications due to the availability of bibliographic data. Here, it was possible to focus also on other aspects of the community by gathering manually various data from the website of the association.

- **Authorship:** Among the authors of EGC, there was 27% of women on all editions until 2019. This roughly matches the overall percentage of women in the discipline. Among the people who contributed the most to the conference (more than 10 papers), there is also an equal percentage of women (12 women out of 48 people).
- **Program committee:** In the first years of the conference, there was a significant under-representation of

women. The percentage of women has increased and was around 34% during the last years, slightly more than the percentage of women in the community. Thus women are not “forgotten” for the task of evaluating papers.

- **Keynote speakers:** EGC invites between 3 and 5 keynote speakers at every edition. Between 2003 and 2019, there were 16 women out of 65 invited keynote speakers (25%). This corresponds to the overall percentage of women. However, it can be noted that in terms of diversity, there were 4 editions where there were no women at all. In 2020, there was 1 woman and 5 men. In 2021, it was 2 women and 3 men. In 2022, 5 women were invited to talk as invited keynote speakers!
- **Presidencies:** EGC features three types of presidencies: chair of the organizing committee, program committee chair and the honorary presidency. Figure 3 reports gender representation according to the types of presidency. We note that women involvement as chairs of organizing committees increased in relation to their representation in the community (33% of women). The representation of women as program committee chairs corresponds to their representation in computer science as a whole (around 24%). However, women were less chosen as honorary president (only 17%). In 2021, a woman was chosen as a honorary president.

As a result, the analysis of the EGC community allowed to examine other aspects than authorship. This was possible because of the restricted perimeter of this community. Analyzing larger communities is done more from the authorship point of view as we will see in the next section.

2.3 Gender analysis across Computer Science communities

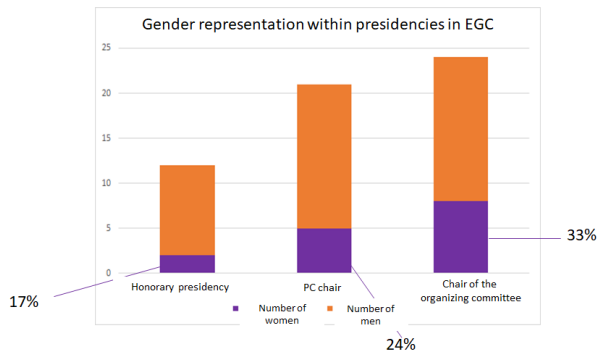


Figure 3: Gender representation according to the different types of presidency in EGC conferences.

We discussed a gender analysis of computer science publications across HCI, AI, Algorithms, Databases, Networking and Operating Systems. This empirical analysis has been carried out primarily by focusing on conferences and journals in the Data Management field [2]. The results show that the number of FAF (First Author Female) papers has increased steadily in the last 40 years (from below 5% to 20%). HCI has witnessed a growth to 35% while having a more diverse composition compared to other communities. Our analysis shows that despite this growth, the number of FAF papers is still low compared to all-male papers. Much work needs to be done to make our conferences and publication venues more inclusive for female colleagues.

3. D&I: CAN WE GO TOO FAR?

The audience was polled again. Table 2 shows the results of the 56 answers to the question “How to be more D&I?”. Responses ranged from taking action (20%) to reaffirming important principles to guide the community (20%). Information (14%) and education (11%) emerged as the primary suggestions. Some attendees praised taking action (11%) while others denied it (5%).

While we agree that D&I environments are richer, fairer, and more aligned with our moral values, pro-D&I measures come with strong downsides, especially for under-represented groups. The panel discussed possible negative side-effects of pro-D&I measures.

3.1 Impostor Syndrome

The impostor syndrome is a psychological pattern in which an individual doubts their skills, talents, or accomplishments and has a persistent internalized fear of being exposed as a “fraud” [6]. In principle, everybody could suffer from this. However, as the measures taken to increase D&I favor under-represented groups, they may increase their chance to suffer from such a syndrome, especially for people in the group with advanced career positions (producing the idea of “the quota person” for instance). This is probably inevitable, but avoiding the words such as “positive discrimination” (use instead “positive action”) or making sure that “being a member of an under-represented group” is merely a tie breaker when choosing between two excellent persons, usually helps tone down this effect.

3.2 Gender equality paradox, glass ceiling and sticky floor

“Gender equality paradox” names the observation that the under-representation of women in scientific fields (in particular those related to STEM) is stronger in the most developed countries. As one says: “We are open. The door is just very heavy”. The door is not closed but women must get a foot in the door to be invited, integrated, accepted, and hired. Women fight to break the glass ceiling in many professions. At the same time, as they can encounter obstacles, the sticky floor restricts them to relative non strategic positions and prevents their scientific career from really taking off.

3.3 Under-Representation Everywhere

One of the easiest methods to promote D&I is to impose *quotas* on the presence of under-represented groups. One of the nastier drawbacks of quotas is the over-solicitation of under-represented groups. This has two very direct consequences: 1) they spend a ridiculous amount of time attending councils, committees, science popularization events, D&I working groups... 2) they, individually, tend to have a huge influence on the carrier of everyone in their community. This first point might naturally fade while increasing the number of people in all under-represented groups to balance the workload. So, there seems to be, again, an inevitable phase-transition that under-represented groups have to bare now for the future greater good. The second point though, is quite difficult to avoid except by limiting (at the national scale) the number of administrative tasks somebody in an under-represented group could undertake.

3.4 Tightly Selected Majorities

When designing pro-D&I measures, it is important to consider their possible unwelcome side effects. Consider for instance the selection of candidates for advanced career positions, with a strongly unbalanced pool of candidates, as is the case of gender in CS. Due to usually larger pool sizes, members of the majority who will pass “despite pro-D&I measures” will necessarily be scientifically stronger than members of under-represented groups. This may have an adverse effect on under-represented members by (unfairly) reinforcing their impostor syndrome. However, if that may be true when taking a snapshot at a particular moment in time, one might (rightfully) argue that under-represented groups are socially discriminated against from a much earlier stage than majorities. This means that the tighter selection in under-represented groups happens earlier but leads to the same result, which is that the (scientific) level of all people selected for advanced carrier positions can, in fact, be considered equal.

3.5 Who pays for D&I?

As said earlier, women may pay a toll on D&I measures by being solicited for time consuming tasks (committees, etc.), a time that is freed for men to pursue more career-rewarding scientific endeavors. D&I is generally implemented as new measures by the scientific management, to show that it is “doing something” about D&I. The actual something is performed by the lower levels of the hierarchy (ordinary researchers, professors and engineers), who have to comply with new rules. This is perfectly acceptable given the vital importance of building a D&I environment. However,

Keyword	Statistics	Examples
take action	11 (0,196)	“More anonymization ; Allow specifying an identity of use at the registration: first name of use, pronouns” ; “Make color-blind slides...”
principles	11 (0,196)	“Fostering women networks” ; “Accepting to address unequal situations”
inform	8 (0,143)	“Awareness-raising and engagement at all levels, especially from childhood” ; “Brief young colleagues (women but not only) on career management”
positive action	6 (0,107)	“Accept that there is positive action when addressing unequal situations” ; “Impose quotas for under-represented groups”
educate	6 (0,107)	“Critical thinking education, especially for the youngest to avoid cognitive biases” ; “Promote bridges between university pathways to facilitate access to IT from other courses”
do nothing	3 (0,054)	“Why do we need to become more D&I?????????????” ; “Refusing someone’s application because you want to recruit someone of the other gender (even less competent), will necessarily be a waste and irresponsible”
role models	1 (0,018)	“Have a woman as president of the EGC association”
penalize	1 (0,018)	“Denouncing toxic behaviors and policies”

Table 2: Summary of the 56 answers to the question “How to be more D&I?”

it can be unnerving to see that when real problems such as sexual harassment arise at higher responsibility levels, there is still a tendency to cover them up, leading to a “double standard”. Perhaps the steady increase of members of under-represented groups in the CS community will allow smoother progress in that direction.

4. FROM DIVERSITY & INCLUSION, TO THE SENSE OF BELONGING AND ENGAGEMENT

While gender crystallizes most of the discussions, D&I is not limited to supporting people or under-represented groups based on visible diversity criteria such as gender, age or physical disabilities. Many other, sometimes invisible, aspects of diversity should be taken into account such as psychological attitude (see Cain [3] or Cathey [4] on being an introvert), experience, religion or beliefs, marital or family status among many others.³

As remarkably theorized by Pat Wadors [9], a transverse solution to D&I could be to consider the notion of *belonging* as essential to ensure *engagement*. As she puts it: “D&I is necessary, but not sufficient.” Rather, it is an “intellectual argument” – not about “the heart and the humanity of it.”[7]. Belonging is related to the sense that persons are accepted in a group for who they are, and not just invited to join because of a D&I rule. As explained by Ramirez [7], “Belonging is not a feeling that should come from an individual, but rather something that can be instilled by others”, i.e. the community has to be active to nurture a sense of belonging.

Indeed, the “power of belonging” is supported by many research works and according to Greg Walton, a psychologist at Stanford University, reinforcing belonging could help oneself reduce “stress levels, consequently improving physical health, emotional well-being, and performance”. This is exacerbated by the proliferation of hybrid conferences within which community members, in particular juniors and students, may suffer more from not feeling a sense of belonging. Dedicated D&I approaches can help in the short-term horizon, while nurturing belonging and engagement may take longer to become effective, but should be more rewarding

³See the Jaw dropping session at the Grace Hopper Conference 2019: <http://signage.showprg.com/ghc19/d2026874-050d-4ccb-aa5d-6906540a66b0-95683-Janice-Wilson.pdf>

in the long term. Hence, D&I should not only be tackled as a set of rules, but rather as a way of life within a scientific community. This should also generalize to our research. How can we imagine working for the social good, welfare and well-being of our fellow citizens if our algorithms and models of knowledge neglect, or even worse discard, even a single point of view.

Although algorithmic biases are common, used properly, AI could have the power to eradicate discrimination against certain social groups. Integrating women into development teams could allow for the inclusion of women-specific data and perspectives due to the effects of “group membership” on emotions and behaviors (given their socialization). The more diverse the “AI coders”, in terms of ethnicity, culture, level or type of study, the more diverse the perspectives. Thus, gender diversity and team diversity in the broad sense could lead to a more inclusive and egalitarian AI, in addition to the known benefits of team gender diversity, better performances and increased productivity and creativity.⁴ The diversity in a scientific community is based on the attractiveness of this discipline. If we consider, for example, the low representation of women, this notably reflects a socialization that keeps girls away. Role models can help support engagement in a discipline based on identification. Understanding that in the context of our research communities is an essential first step toward a more inclusive science.

5. REFERENCES

- [1] 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.
- [2] 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.
- [3] Susan Cain, Gregory Mone, and Erica Moroz. *Quiet power: the secret strengths of introverts*. Penguin, 2016.
- [4] Glen Cathey. A Manifesto for Introversion in Diversity, Inclusion, and Belonging. <https://www.linkedin.com/pulse/manifesto->

⁴<https://bit.ly/3IFfFig>

- introversion-diversity-inclusion-belonging-glen-cathey/, 2019. [Online; accessed 10-February-2022].
- [5] 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.
- [6] Joe Langford and Pauline Rose Clance. The impostor phenomenon: recent research findings regarding dynamics, personality and family patterns and their implications for treatment. *Psychotherapy: Theory, Research, Practice, Training.*, 30 (3):95–50, 1993.
- [7] J. Cook Ramirez. Is belonging the new D&I ? Human Resource Executive. <https://hrexecutive.com/is-belonging-the-new-di/>, 2022. [Online; accessed 02-February-2022].
- [8] Sébastien Valat and Cécile Favre. Regards d’actualité au prisme des enjeux sociétaux sur les données historisées d’egc. In *Extraction et Gestion des Connaissances, EGC 2020, Brussels, Belgium, January 27-31, 2020*, volume E-36 of *RNTI*, pages 205–216. Éditions RNTI, 2020.
- [9] P. Wadors. Diversity efforts fall short unless employees feel that they belong. Harvard Business Review Digital Articles. <https://hbr.org/2016/08/diversity-efforts-fall-short-unless-employees-feel-that-they-belong>, 2016. [Online; accessed 02-February-2022].