Collaboration in a data intensive network of Ph.D-fellows - experiences from Train²Wind

Grischa Fraumann, Frans van der Sluis, Morten Hertzum and Haakon Lund

Ph.D. projects within a specific Innovative Training Network (ITNs) are interconnected, necessitating collaboration across different organizations and fulfilling requirements set by both the host organization and the ITN. Collaboration often occurs during secondments, which are research training periods involving physical mobility. Unlike many other Ph.D. fellowships, MSCA funding supports these intersectoral and international mobility periods, which are considered crucial for every stage of research careers.

With the current development in the digitalization of the wind energy sector [3] Ph.D. fellows are increasingly embedded in data-intensive research environments, which involve processing large volumes of complex data from experiments or simulations. This reliance on data necessitates collaboration due to the varied competencies required, from data acquisition to validation and visualization, often beyond the capacity of individuals. Researchers must pool expertise, resources, and data access from multiple organizations. In wind energy, which relies on experimental data from wind farms and simulations, Ph.D. candidates must combine simulations, measurement campaigns, and data analysis while coordinating with various stakeholders, including wind farm developers and local authorities.

As a general phenomenon, scientific collaboration lacks a standard definition but generally means "joint effort toward a common goal." Sonnenwald defines it as interactions among scientists that facilitate task completion and shared goals [11]. Benefits of such collaboration include informal peer review and the ability to rely on the experience of senior researchers. However, challenges exist, such as demanding mobility requirements, conflicts within research alliances, and dependencies on team members. These challenges can overwhelm Ph.D. candidates, who may feel particularly burdened by the need to rely on others for the completion of their degrees. As part of the Train²Wind project, we have studied collaboration within the project with a specific focus on the particular enablers and challenges for PhDs in wind energy.

These studies illustrate how the complexity and interdisciplinarity of data-intensive research make collaboration essential for the success of Ph.D. students and research networks.

Empirical Studies and Methods

Through three studies we gathered data on collaboration using early-stage interviews, late-stage interviews, and in-between observations at measurement campaigns. To date, there is limited research on data-intensive environments and Ph.D. fellows, particularly regarding how early-stage researchers such as Ph.D. fellows navigate these environments [10]. The

present study therefore provides unique insights into how Ph.D. fellows conceive collaboration in a data-intensive environment.

Based on early-stage interviews with 12 Phd students from the Train^{2}Wind Network we analyzed the expectations of PhD-students of collaboration in a data-intensive research environment. As an added benefit it was possible to interview 11 Ph.D fellows from 3 other similar networks adding more interview data about the expectation and experiences from PhD fellows towards collaboration. The interviews were framed with a background in the theory of remote collaboration (TORSC) [9] and gathered data from informants at various research organizations, and R&D firms in several countries. This study focuses on Ph.D. fellows' forward-looking expectations of collaboration, allowing a detailed analysis of the transition phase into a research environment. The interviews were conducted as remote video interview sessions and the collected audio files were subsequently transcribed and coded by the authors of this paper. The resulting codes were sorted into groups using card-sorting during in-person meetings [8, 6].
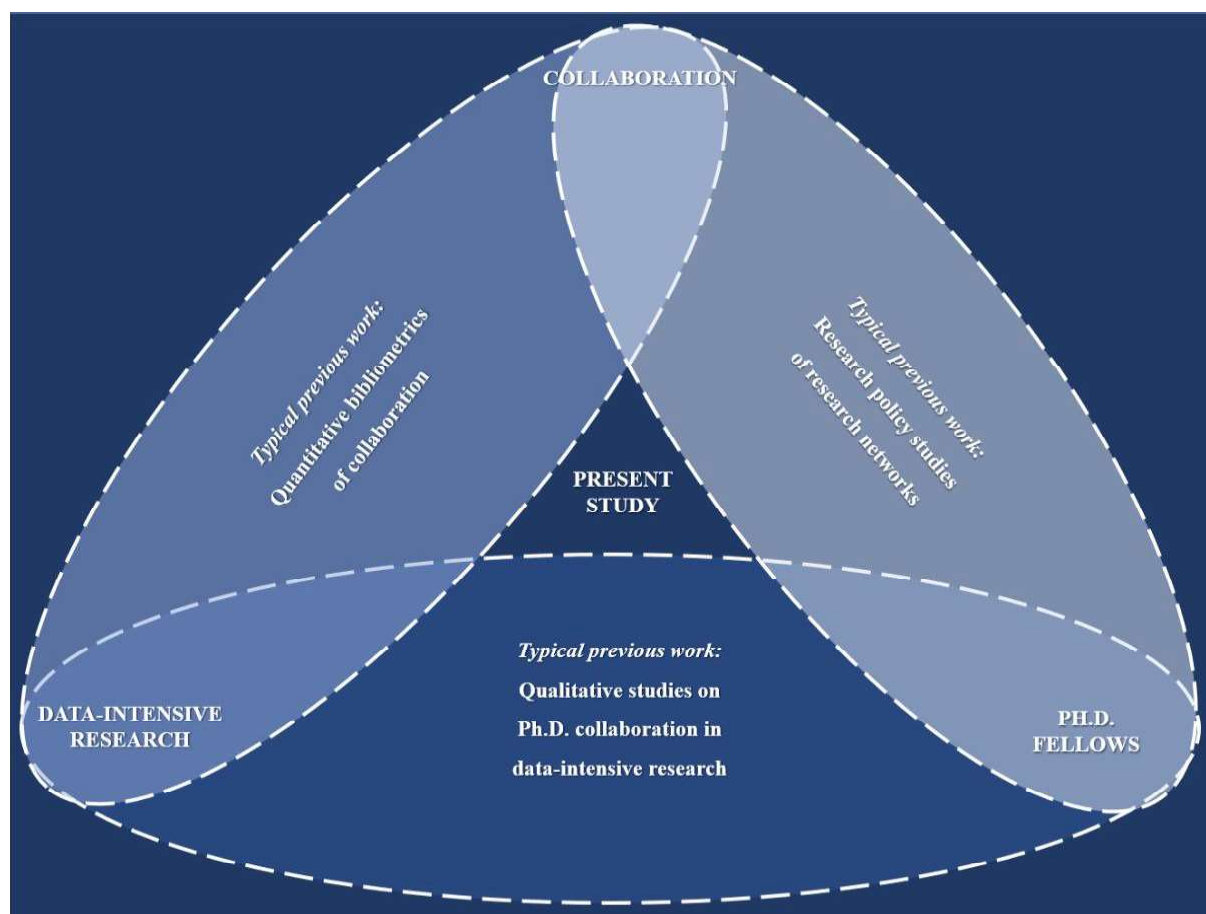

Figure 1: Position of the study. The present study combines three research topics (i.e., collaboration, Ph.D. fellows, and data-intensive research networks).

Framework for studying collaboration

A well established framework for studying collaboration is the Theory of Remote Scientific Collaboration (TORSC). TORSC was preferred as the theoretical framework because of its focus on collaboration in remote and technical environments [1, 7] Train^{2}Wind network can be characterized as relying heavily on remote collaboration due to its distributed structure with partners in several countries and it is to a high degree embedded in a highly technical environment. To guide the interviews with participating ESR, TORSC was used as a framework for developing an interview guide focusing on the following five key elements [9]:

Nature of the Work: Characteristics of the research, which in data-intensive environments can create communication bottlenecks due to the different expertise contributed by team members.

Common Ground: Mutual knowledge and shared understanding, which can be hindered by a lack of shared interests and common vocabulary.

Collaboration Readiness: Individual characteristics and motivations, which may be less developed in early-stage Ph.D. fellows, as their researcher role identity evolves over time within the network.

Organization and Management: The structure and management of collaboration, with known challenges in distributed collaborations and clearly defined roles in EU-funded projects.

Technology Readiness: Availability and proficiency in using the right technologies, which is generally less of an issue in data-intensive networks due to the high level of technical competence.

We derived four themes from our analysis of the collected data, each theme illustrating aspects of collaboration and the challenges experienced by the participating young researchers.

Collaboration is a balancing act

Research collaboration is inherently complex, requiring researchers to balance diverse priorities, collaborators, cultures, and interests [4]. This balancing act involves aligning different cognitive mindsets, managing multiple interdependent teams in a multiteam system, and reconciling various R&D goals between universities and firms [12, 2].

Ph.D. candidates in data-intensive research networks face specific balancing acts. They must balance the expectations and goals of their host university with those of the larger research network, often requiring the use of shared infrastructure and data. For instance, one interviewee highlighted the challenge of integrating knowledge from different facilities, while another emphasized the importance of cultural sensitivity in European projects.

Effective communication is crucial in these collaborations, especially given the reliance on remote interactions. Ph.D. candidates must navigate between local research group require-

ments and the broader network's objectives, balancing their focus on core Ph.D. elements with enriching opportunities from collaboration. This balancing act is further complicated by the need to evaluate the costs and benefits of collaborations in terms of time and resources, a significant challenge at the start of their Ph.D. journey.

Quotes from interviewees illustrate these challenges. One described how collaboration could add to their workload and distract from their primary research goals, while another stressed the importance of careful planning to meet deadlines for both collaborations and other projects. Time constraints, especially within the typical three-year Ph.D. duration, add to the complexity of producing high-quality research outputs while engaging with external collaborators.

An understanding of these balancing acts can feed into a better network organization and is crucial for successful training networks, given their specific focus on PhDs and collaboration.

Overall, Ph.D. candidates in data-intensive networks must balance project-network goals, focus versus enrichment, and timing versus benefits. This intricate process requires managing expectations from both their home institution and the broader network, making collaboration in such settings a true balancing act.

Experiences and expectations differ

Research collaboration, particularly in data-intensive research networks, involves navigating a complex balance of expectations and experiences. Ph.D. candidates often enter these collaborations with optimistic expectations, anticipating numerous opportunities to enhance their research. For example, respondents noted the supportive and encouraging nature of their professors and supervisors, who play a pivotal role in fostering collaborations by leveraging their networks and providing guidance. This optimism is rooted in the early stages of a Ph.D., where the focus is on planning and leveraging existing contacts and infrastructure facilitated by supervisors.

The positive outlook on collaborations is illustrated by interviewees' remarks on the beneficial role of supervisors and the encouragement to engage with external collaborators. Supervisors' involvement is crucial, especially in fields like wind-energy research, where access to infrastructure, data, and industry expertise is often managed through these established networks. For instance, respondents emphasized how supervisors initiate and support collaborations, ensuring that Ph.D. candidates can access necessary resources and expertise.

However, the reality of university-industry collaborations often presents challenges that temper initial expectations. Issues such as data sharing restrictions and limitations imposed by industrial partners can hinder the smooth execution of collaborative projects. Respondents highlighted difficulties in accessing valuable data sets that are not publicly available, which can significantly impact their research progress. These experiences underscore the

need for realistic expectations when engaging in industry collaborations, where proprietary data and internal restrictions can complicate the collaborative process.

The difference between expectations and actual experiences in collaborations is also influenced by prior experience in industry R&D settings. Ph.D. candidates with such backgrounds tend to have more pragmatic views on the availability and accessibility of data and infrastructure. Their firsthand experience with the constraints of industry collaborations helps them set more realistic goals and navigate potential challenges more effectively.

Overall, our analysis suggests that while Ph.D. candidates start with high expectations for collaborations, their experiences gradually shape a more balanced perspective. Supervision plays a critical role in managing these expectations, and prior industry experience helps in formulating realistic collaboration goals. As candidates progress through their Ph.D., they learn to balance their initial optimism with the practical realities of collaboration in a data-intensive research environment. This ongoing adjustment highlights the dynamic nature of research collaboration, where expectations and experiences continuously interact and evolve

Data shape collaboration

Networks can predefine research agendas, making collaboration a necessity rather than an option. This enforced collaboration is embedded in the network's structure, where roles, resources, and tasks are divided among members, as seen in respondents' reliance on data from other Ph.D. students for their projects. While such dependencies can create uncertainties, they also integrate diverse knowledge and perspectives, enriching the Ph.D. experience.

Ph.D. projects within ITNs (Innovative Training Networks) are often interlinked, with data dependencies set at the project proposal stage [5]. This structure, while potentially limiting individual research freedom, ensures a collaborative space that facilitates shared objectives and co-authored publications. For example, some respondents noted that their research questions depended on results from other Ph.D. students, highlighting the collaborative nature of their work.

Despite the predefined aspects, respondents viewed these structures as opportunities rather than constraints. The network's organizational structure and planned activities, such as meetings and secondments, provide additional collaboration opportunities, making the Ph.D. experience more enriching. The necessity of collaboration within the network, enforced through dependencies and predefined structures, ultimately supports a collaborative space that benefits the Ph.D. candidates by offering access to specialized knowledge and facilitating a structured yet enriching research environment.

Social events glue the network together

The COVID-19 pandemic significantly impacted collaborations within research networks by preventing most physical meetings, which are crucial for effective interaction among Ph.D. fellows. The enforced distance disrupted planned activities, such as physical meetings and common fieldwork, which were integral to the projects. This limitation was highlighted by our informants, who emphasized the need for frequent contact and social events to maintain a functional network.

The absence of in-person meetings created challenges for spontaneous interactions and informal problem-solving sessions, as expressed by respondents who missed the ease of face-to-face meetings. For example, one respondent noted, "it is much harder to get in contact with my colleagues in my [organization], and I really miss the possibility of this half hour of face-to-face meeting to solve some questions. [...] COVID is one of the problems that I think that still today we have to deal with" (Interviewee 19). This sentiment underscores the disruption caused by the pandemic, making it difficult to maintain the frequent contact necessary for successful collaboration.

The enforced physical distance stifled many collaborations, particularly within ITNs that started during the pandemic. Respondents reported that the shift to virtual meetings and online conferences hindered the natural flow of collaboration. For instance, one respondent shared, "[all of these ITNs started] during or under corona. [...] For a lot of these collaborations, I think it really stifled that, right? All of these conferences have been online. I've never met another [name of ITN] member and most of the ones that I do know are the ones from here at [name of organization]. I think that does block some of these collaborations" (Interviewee 21). This illustrates how the lack of physical interaction obstructed collaborative efforts.

Respondents also expressed a preference for in-person meetings over virtual ones, highlighting the human element and the ability to better gauge reactions and emotions face-to-face. One respondent mentioned, "[…] it's just very long Zoom meetings every now and then. [...] I mean, it's good. But anyway, I like in-person meetings more. I'm more comfortable with seeing the person, see how they feel, see how they react. It's more human for me" (Interviewee 7). The virtual format, while functional, failed to replicate the dynamics of in-person interactions.

Despite these challenges, respondents adapted by leveraging technologies to maintain continuous communication and share research data. Skills in using collaborative tools became essential, with respondents listing various websites and software used in their daily research practices. However, the extraordinary meeting restrictions during the pandemic added an extra layer of complexity to an already challenging environment, as respondents had to find alternative ways to communicate and collaborate over a distance.

Overall, the pandemic underscored the importance of social events and spontaneous interactions in gluing the network together. While technologies facilitated remote collaboration, the absence of physical meetings created barriers that affected the spontaneity and effec-

tiveness of interactions. This experience highlighted the crucial role of in-person meetings and social events in fostering a collaborative and cohesive research environment.

Perspective

The perspectives on collaboration evolve as candidates adjust their initial optimism to the realities of collaborative research, influenced by prior experiences and ongoing interactions within the network. Supervisors play a critical role in managing expectations and providing guidance, particularly in navigating the complexities of university-industry collaborations where data sharing restrictions can pose significant obstacles.

Despite the challenges, collaboration remains central to the Ph.D. experience, with networks often predefining research agendas and creating a structured space for shared objectives. While the pandemic disrupted physical meetings, highlighting the importance of social events for spontaneous interactions, candidates adapted by leveraging digital tools to maintain collaboration. Ultimately, successful collaboration in these networks requires a nuanced approach to balancing competing demands and leveraging the structured yet flexible environment provided by the network.

References

[1] M. J. Bietz, S. Abrams, D. Cooper, K. R. Stevens, F. Puga, D. Patel, G. M. Olson, and J. S. Olson. "Improving the Odds Through the Collaboration Success Wizard". In: *Translational behavioral medicine* 2.4 (2012), pp. 480–486.

[2] Gabrielle Cederholm. "Success Factors in University-industry Collaborations. A comparison of a research and development project". MA thesis. University of Gothenburg, 2015.

[3] A. Clifton, S. Barber, A. Bray, P. Enevoldsen, J. Fields, A. M. Sempreviva, L. Williams, and et al. "Grand challenges in the digitalisation of wind energy". In: *Wind Energy Science Discussions* 8.6 (2022), pp. 1–42.

[4] Jennifer Dusdal and Justin J W Powell. "Benefits, Motivations, and Challenges of International Collaborative Research: A Sociology of Science Case Study". In: *Science and Public Policy* 48.2 (2021), pp. 235–245.

[5] Mavrina E, Kimble L, Waury K, Gogishvili D, Gómez de San José N, Das S, and et al. "Multi-Omics Interdisciplinary Research Integration to Accelerate Dementia Biomarker Development (MIRIADE)". In: *Frontiers Neurologoy* (2022).

[6] Uwe Flick. *An introduction to qualitative research*. SAGE Publications Ltd, 2022.

[7] Magdelena Haman and Morten Hertzum. "Collaboration in a distributed research program: Islands of intensity in a sea of minimal interaction". In: *Journal of Documentation* 75.2 (2019), pp. 334–48.

[8] Conrad LY and Tucker VM. "Making it tangible: hybrid card sorting within qualitative interviews". In: *Journal of Documentation* 75.2 (2019), pp. 397–416.

[9] Judith S. Olson and Gary M. Olson. *Working Together Apart: Collaboration over the Internet*. Morgan and Claypool Publishers, 2014.

[10]   Biswas R, Schiller A, Casolani C, Daoud E, Dode A, and et al. Genitsaridi E. "Doctoral Studies as part of an Innovative Training Network (ITN): Early Stage Researcher (ESR) experiences". In: *Open Research Europe [Internet]* (2021).

[11]   Diane H. Sonnenwald. "Scientific collaboration". In: *Annual Review of Information Science and Technology* 41 (2007), pp. 643–81.

[12]   Wanzenböck I Wuestman M Frenken K. "A genealogical approach to academic success". In: *PLOS ONE* 15.12 (2020).