Check for updates

Received: 16 September 2025 | Revised: 5 March 2026 | Accepted: 17 May 2026

DOI: 10.1002/asi.70094

RESEARCH ARTICLE

JASIST WILEY

Scientific collaboration in data-intensive research networks: PhD fellows' expectations and early experiences

Grischa Fraumann[1,2] | **Haakon Lund**[1] | **Frans van der Sluis**[1] | **Morten Hertzum**[3]

[1]University of Copenhagen, Copenhagen, Denmark

[2]Häme University of Applied Sciences, Hämeenlinna, Finland

[3]Roskilde University, Roskilde, Denmark

Correspondence
Morten Hertzum, Roskilde University, Universitetsvej 1, 4000 Roskilde, Denmark.
Email: mhz@ruc.dk

Funding information
H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 861291

Abstract

Collaboration has become important at all stages of research careers. In data-intensive research fields such as wind energy, many PhD fellows are socialized to such collaboration in networks that train a cohort of PhD fellows. Based on interviews with 23 PhD fellows in four wind-energy training networks, we investigate their expectations and early experiences regarding scientific collaboration. We find that expectations for collaboration are high, but also that the PhD fellows' expectations and early experiences differ. Their experiences are influenced by conditions that make collaboration a balancing act at the organizational, content, process, and personal levels. At the organizational level, learning scientific collaboration involves negotiating interdependencies that are created at the network-proposal stage and presume collaborations about data among the PhD fellows in the network. At the content and process levels, the PhD fellows are dependent on collaboration to collect, analyze, share, validate, and compare data. At the personal level, devising a doable dissertation involves that the individual PhD fellow succeeds in realizing these collaborations. While the studied wind-energy research networks nurture a collaborative environment, the collaboration they presume and the rewards it promises come with the risk of imposing imbalanced collaborations on some PhD fellows.

1 | INTRODUCTION

Doctoral students—hereafter referred to as PhD fellows—are an important part of today's research infrastructure. They are increasingly embedded in data-intensive research networks such as Innovative Training Networks (ITNs), the flagship doctoral program financed by the European Commission (Grant et al., 2022). Through such networks, PhD fellows become socialized to research (Holley, 2020; Johnson et al., 2017). This socialization involves their introduction to and participation in scientific collaboration (Larivière, 2012). Because they are new to both research and data-intensive research networks, PhD fellows start out with little more than expectations toward scientific collaboration. This study investigates these expectations during the period where they begin to transition into experiences. Expectations are important to study because they show how scientific collaboration is perceived by newcomers and because they determine where activities for introducing PhD fellows to scientific collaboration should start.

This is an open access article under the terms of the Creative Commons Attribution-NonCommercial License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited and is not used for commercial purposes.
© 2026 The Author(s). *Journal of the Association for Information Science and Technology* published by Wiley Periodicals LLC on behalf of Association for Information Science and Technology.

Scientific collaboration is the "interaction taking place within a social context among two or more scientists that facilitates the sharing of meaning and completion of tasks with respect to a mutually shared, superordinate goal" (Sonnenwald, 2007, p. 645). In the context of data-intensive research networks, researchers need to pool expertise, resources, and data access from various individuals and organizations to process the large and complex datasets that are captured in experiments or generated by simulations (Hey & Trefethen, 2008; Liew et al., 2017). These demands put collaboration center stage. However, although doctoral socialization and collaboration are well studied, the role of data in structuring doctoral collaboration remains underexamined (Hara et al., 2003; Torka et al., 2022). Prior work on PhD fellows' collaboration often focuses on publication-centered indicators (e.g., co-authorship; Cugmas et al., 2024; Wang et al., 2021), attending less to how collaboration is perceived and initiated by newcomers.

This study investigates four ITNs in the field of wind-energy research, which is a prime example of data-intensive research. Wind-energy research relies on experimental data gathered at wind farms and in wind tunnels as well as on simulations in computer models (Veers et al., 2019). Like other ITNs, the four studied ones are organized by senior researchers from multiple countries and each funds an international group of PhD fellows. PhD fellows in this field need to combine measurement campaigns with computer simulations and to confer with senior researchers, wind-farm developers, the authorities overseeing wind-farm operations, and other stakeholders. On the basis of interviews, we seek to answer the research question: *How do PhD fellows in wind-energy research networks perceive and establish scientific collaboration during the early stages of their fellowships?* In answering this question, we analyze the period during which the interviewed PhD fellows' expectations about scientific collaboration transition into their early experiences with it. While this transition continues beyond early experiences, they are the first reality checks of the expectations and instrumental to PhD fellows' decisions about the planning and setup of their work. The early experiences should not be seen as proxies for end-of-fellowship experiences, but as important in their own right.

The remainder of this article is structured as follows. First, we review related work on PhD fellows' involvement in scientific collaboration in data-intensive research networks. Second, we describe the method used for collecting and analyzing the interview data. Third, we present the findings, which show how the interviewed PhD fellows' expectations and collaborations are shaped by data and dependencies among the PhD fellows. Finally, we discuss the balancing act of making collaboration work in such a setting.

2 | BACKGROUND

This study is positioned at the intersection of research on scientific collaboration, data-intensive research networks, and PhD fellows.

2.1 | Scientific collaboration

Scientific collaboration has multiple benefits but also entails certain costs (Katz & Martin, 1997). The benefits include that researchers may obtain knowledge and skills from others, transfer knowledge and skills to others, get access to data and equipment, generate new knowledge through the cross-fertilization of ideas, enjoy intellectual companionship, increase the visibility of their work, get funding, build social capital, and grow their network of future collaborators (Bukvova, 2010; Dusdal & Powell, 2021; Katz & Martin, 1997; Olson, Zimmermann, & Bos, 2008; Stokols et al., 2008). The combined effect of these benefits is that research is, in principle, performed more effectively. However, Zuccala et al. (2024) find that researchers' collaboration practices often differ from their collaboration preferences, thereby suggesting the presence of costs that overshadow the benefits.

The costs of scientific collaboration include interpersonal conflicts that undermine trust among researchers (Sonnenwald, 2004), leadership styles that allow conflicts to escalate (Stokols et al., 2008), problems with assigning credit for research outputs (Chung et al., 2016), and high coordination costs when the collaborating researchers are geographically distributed (Cummings & Kiesler, 2007). As a possible consequence of the coordination costs, Haman and Hertzum (2019) find that the researchers in a globally distributed research program mostly worked alone on their contributions to their joint work. While the researchers published collaboratively, they restricted their interaction to short bursts in a process otherwise characterized by minimal interaction. This illustrates that scientific collaboration may occur in different ways and at different levels—from complementary to integrative (Hara et al., 2003).

2.2 | Data-intensive research networks

Research in many fields is increasingly dependent on large datasets, infrastructures for managing them, and

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

tools for analyzing them (e.g., Borgman, 2015; Olson, Zimmermann, & Bos, 2008; Stokols et al., 2008). As a result, data-intensive research has emerged as a paradigm alongside theoretical, experimental, and computational research (Hey et al., 2012). It reframes how researchers think about compiling and analyzing data. For example, atmospheric researchers share access to distributed data-acquisition infrastructures (Petzold et al., 2024), neuroscientists crowdsource data annotation (Lacourse et al., 2020), medical scientists publish clinical datasets for others to analyze further (Barros et al., 2022), interdisciplinary researchers rely on open-source software code (Trisovic et al., 2022), and wind-energy researchers negotiate access to continuous data streams from operational wind turbines (Singh & Rizwan, 2022).

Borgman (2015) notes that data sharing is common practice in only a few research fields, though most researchers are, in principle, in favor of sharing. Instead of collaborating through open data sharing, most collaboration around data happens within the projects that originally collect the data. These collaborations become data-intensive research when "the capture, curation, and analysis of large volumes of data is central to the scientific question" (Resnik et al., 2017, p. 344). While admittedly vague, this definition sets data-intensive research apart from the many studies that are based on, among others, theoretical arguments, small-*N* lab experiments, and workplace ethnographies. The dependency on large datasets poses challenges for researchers, who need the skills to work with the data in effective and efficient ways (Scroggins et al., 2020). Often, the required skills are diverse and beyond the skillset of any individual researcher, thereby prompting collaboration (Clifton et al., 2023; Hampton et al., 2017; Petzold et al., 2024).

Research networks are a way of structuring such collaboration. Mulkay et al. (1975, p. 190) define them as "concentrations of interest ties without clear boundaries," thereby emphasizing that they are amorphous groupings. In the case of ITNs, they are funded and distributed project organizations that assemble researchers from established institutions for the purpose of working together on a common research topic. This is particularly beneficial for data-intensive research because it allows tapping into infrastructure, data, and expertise that is not available at researchers' home institutions (Borgman et al., 2012; Mayernik et al., 2013). Wind-energy ITNs also include industry partners such as wind-turbine manufacturers, wind-farm operators, energy firms, and engineering consultancies. Through these partners, the networks foster academia-industry mobility, accelerate technology transfer, and facilitate researcher access to industrial data and facilities (Afewerki & Steen, 2023; Mah & Hills, 2014). This network structure makes wind-energy ITNs well suited for studying collaboration in data-intensive environments.

2.3 | PhD fellows

While some PhD fellows have pre-existing professional networks (Barnacle et al., 2020), others do not, which makes their introduction to scientific collaboration a rite of passage (Hara et al., 2003). As summarized by Torka et al. (2022, p. 118), collaboration is an aspect of PhD programs that "can enrich, but potentially overwhelm, doctoral candidates." In comparison to established scholars, PhD fellows put greater emphasis on the social aspects of collaboration, such as belonging to a student community (Dytham, 2019) and receiving empathetic emotional support (Kemp et al., 2014). That said, the key concern for PhD fellows is to devise a doable dissertation in the midst of a research environment with many possibilities and uncertainties, not least regarding collaboration (Sorgner, 2022).

PhD fellows' collaboration with their supervisor is key to their scholarly socialization—helping them build a network (Wang et al., 2021) and become productive (Paglis et al., 2006). PhD fellows in ITNs also benefit from the access these networks provide to an international group of established researchers committed to the goal of the ITN. For example, these established researchers provide informal peer reviews on the PhD fellows' work (Rigby & Edler, 2005). In addition, Bitsios et al. (2023) find that the PhD fellows in a sample of 550 ITNs commonly co-author publications with one another, thereby showing that their collaborations reach beyond the fellow-supervisor dyad. This co-authorship pattern may reflect the additional dependencies created through the interlinking of one PhD fellow's project with that of another in the setup of the ITN. PhD fellows face the dual requirement of fulfilling their obligations to their host institution and to the ITN with its interlinked PhD projects (Biswas et al., 2021). Furthermore, the PhD fellows in ITNs are exposed to collaboration with the industry partners, for example through secondments, that is, short-term stays. Thereby, the PhD fellows supplement their academic knowledge with business and entrepreneurial knowledge and may develop a researcher role identity as boundary spanners between academia and industry (Gruber et al., 2023).

3 | METHOD

To investigate how PhD fellows perceive scientific collaboration, we conducted an interview study. We chose

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

interviews for data collection because they provide access to perceptions. The perceptions of interest to us were PhD fellows' expectations and early experiences regarding scientific collaboration. The study was approved by our institutional review board.

3.1 | Setting

Formally, an ITN is a research network organized by senior researchers from at least three member states (or associated countries) in the European Union. Each ITN funds a group of PhD fellows and, thereby, functions as a cohort-based doctoral training program in a project structure. Between 2014 and 2021, ITNs have financed a total of 25,000 PhD fellows through the Marie Skłodowska-Curie Actions (European commission, 2021). ITNs—and their successor Doctoral Networks—are similar, among others, to the National Research Training Programs funded by the National Science Foundation in the United States.

The empirical setting of the present study comprised four Horizon 2020 ITNs in the research area of wind energy, each dependent on measurement campaigns for collecting large datasets at wind farms. The 4 ITNs consisted of 28 organizations that hosted a total of 51 PhD fellows. The hosting organizations were located in 11 European countries and, in turn, worked with another 40 partner organizations that provided research data, access to wind farms, hosted PhD fellows for secondments, and so forth. While this empirical setting was restricted to one funding instrument and research field, the diverse set of participating organizations and network members ensured variation in how collaboration was approached.

3.2 | Interview procedure

The interviews were conducted by the first author, a PhD fellow on the topic of scientific collaboration within one of the four ITNs. The other PhD fellows in the ITNs were invited to the interviews via email, after contact to the ITNs had been established through the senior researchers and project managers. Invitations were restricted to PhD fellows in the first half of their 36-month fellowships at the time of the interview, in line with the focus of the study on expectations and early experiences. Across the four ITNs, 23 PhD fellows gave their written informed consent to be interviewed. Interviews took place between May and September 2021. We note that this period coincided with the Covid-19 pandemic and its restrictions on physical meetings. Table 1 shows the interviewee profile.

TABLE 1 Interviewee profile, $N = 23$ PhD fellows.

Characteristic	Count	Percent
Country of workplace (number of different countries)	9	
Nationality (number of different countries)	13	
Workplace organizational type		
PhD fellow at technical university	13	57
PhD fellow at comprehensive university	8	35
PhD fellow at non-university organization	2	9
Gender		
Female	10	43
Male	13	57

Informed by prior work (see Section 2), we developed an interview guide that covered eight topics: (1) criteria for selecting collaborators, (2) role of collaboration in PhD work, (3) perceived barriers to collaboration, (4) role of host institution and secondments, (5) value of knowledge and data sharing, (6) collaboration during the research process, (7) use of websites and software for collaboration, and (8) further comments. Each topic involved several questions (for the full interview guide, see Appendix S1). To test and refine the interview guide, five pilot interviews with other PhD fellows were conducted prior to the interviews.

Due to the geographic spread of the interviewees, all interviews were conducted online. The interviews were audio-recorded, auto-transcribed with the online NVivo transcription application, and manually checked to correct errors in the automatic transcription. Before the analysis, the first author anonymized the transcripts by deleting names, affiliations, genders, nationalities, and the like.

3.3 | Interview coding

We used card sorting (Spencer, 2009) for analyzing the interviews. We chose card sorting because it is a tangible variant of thematic analysis suited to a collaborative analysis process. In preparation for this process, the authors each read one randomly selected transcript to gain a sense of the interview data. The analysis consisted of the following five steps:

First, 12 interviews were selected and coded. To avoid bias in the mix of ITNs, countries, host organizations, and the like, the 12 interviews were randomly selected.

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The coding involved identifying and annotating the interview segments that provided information about the eight topics in the interview guide. For example, one segment was annotated with the codes "openness," "availability," and "easygoing" as criteria for selecting collaborators (the first of the eight topics). To bolster the richness and reliability of the coding, two authors independently coded each interview. In addition, we rotated the researcher pairings (i.e., each of the six different combinations of two authors coded two randomly selected interviews).

Second, the authors met to sort the coded interview segment into groups. Each coded segment was printed on a paper card and then, the authors discussed and sorted these cards into groups of similar content. Discrepancies in the coding were resolved by consensus. As groups started to emerge, they were assigned tentative labels. For example, the cards with the code "openness" entered a group labeled "willingness to collaborate." This card sorting resulted in 25 groups, created in an inductive manner.

Third, after the card sorting of the first 12 interviews had resulted in a shared understanding of their contents, the first author coded the remaining 11 interviews. In this coding, all instances of the 25 groups from the previous step were coded with the group label. In addition, all interview segments that provided new information were identified and annotated. There were 61 such segments.

Fourth, the authors met to incorporate these 61 segments into the card sort, thereby deleting one group, creating one new group, and splitting one group into two. Apart from starting from the 25 existing groups, this additional round of card sorting was performed in the same way as the initial round. The full analysis of the 23 interviews sorted 1007 segments into 26 groups.

Fifth, the 26 groups were arranged into five analytic themes. These five themes cut across the eight topics in the interview guide. For example, the first theme merged groups that arose from answers to the second, fifth, and sixth topics, but these topics also contributed to other themes. The five themes provide the structure of the analysis reported in the next section. Figure 1 summarizes the analysis process.

4 | RESULTS

Tables 2–6 present the five themes generated by the card sorting. Each table gives the groups in a theme, the number of interviewees who mentioned these groups in one or several segments, and a sample quote from the segments.

4.1 | Expectations for collaboration were high

Because the interviewees had still not completely planned their PhD projects, their responses included many expectations regarding collaboration. The interviewees were optimistic and expected multiple positive gains and outputs from collaboration. These gains and outputs ranged from "exchange ideas" (Interviewee 17), over "constantly getting feedback" (Interviewee 11), to "a journal paper which showcases your collaboration in final form" (Interviewee 1), thereby spanning the entire research process ("I guess it [i.e., collaboration] is pretty important throughout," Interviewee 4).

As Table 2 shows, the sharing of knowledge and, in particular, data was a benefit expected by many interviewees, including by pooling, comparing, and cross-validating data and analyses. This way, the interviewees expected to "gain from each other's skills" (Interviewee 5) and to be able "to come up with solutions that we wouldn't otherwise come up with" (Interviewee 16). The ITNs were seen as providing rich possibilities for the interviewees to learn from many people in addition to their supervisor, including from the other senior researchers, the other PhD fellows, and the industry partners. All interviewees described their current situation as one characterized by insufficient knowledge and a large need for learning from others. However, the interviewees also mentioned more incidental benefits of collaboration, such as "to be motivated by the progress of each other" (Interviewee 15) and "that I actually can have friends at the end after collaborating" (Interviewee 20). Furthermore, 18 of the

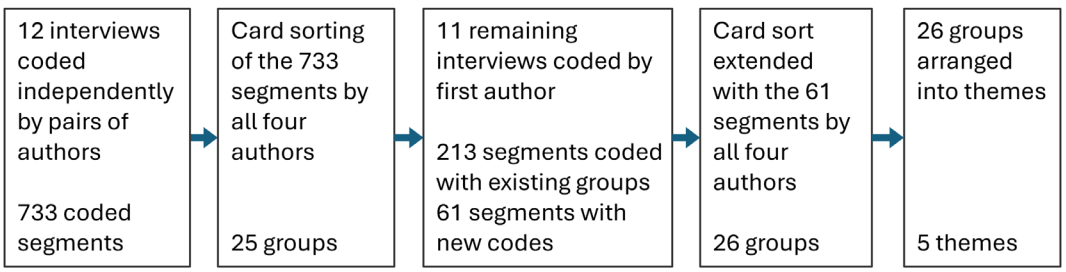


FIGURE 1 Interview coding process.

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2 Expected benefits of collaboration.

Group	Interviewees	Sample quote
Gains and outputs of collaboration	21	"I imagine that at some point we will have a communal pool of data, where people from other universities will have access to our data and we have access to theirs. Then, we can compare and look and maybe offer our opinion." (Interviewee 7)
Pooling data and analyses	19	"We are kind of doing code-to-code comparison of different results ... One team from each country is submitting its results and, in that way, we are collaborating with those people." (Interviewee 17)
Comparing and validating data, models, and instruments	18	"While they're studying the tool itself, I'm expanding this study into a more physical or meteorological point of view. In working together, I would look at some of their data to help them validate their instruments." (Interviewee 21)
Statements that collaboration is important	18	"I think it's very important to share all the important research results or methodology or even research conclusions with your collaborators to create an honest exchange of information between the two parties. It also helps to remove confusion and to identify better ways to treat the research question. So, it's very important that you start sharing your information and results from the very start rather than wait for the end." (Interviewee 1)
Peer-to-peer support and feedback	15	"I don't know if what I'm doing is right and if I'm doing it properly. So, I need feedback from people who have more experience than me, because they can both judge my work and give me tips ... And when I'm stuck, I just need someone that can help me with it, that can grant me vision about things I don't see." (Interviewee 11)
Mutual learning	5	"Let's say there is one PhD student who is working in detail on [models to analyze a wind turbine] which would be really useful for me to implement in my bigger model that works on the whole wind farm. Being able to incorporate her work into mine would be really useful, because it would give more depth to my work, a higher fidelity. And, for her it would be nice to see how her model performs in an actual wind farm rather than in simple benchmark cases." (Interviewee 14)

23 interviewees made statements indicating that—but not how—collaboration was important.

A recurrent argument in the interviewees' descriptions of the benefits of collaboration was a largely implicit expectation of reciprocity. Most individual collaboration instances appeared to be unidirectional, as in: "There is another PhD [in the ITN], who is working on this topic, who will provide me with this data" (Interviewee 15). Because the direction would be reversed in other instances, all parties were expected to benefit from sharing. However, there were also instances in which both parties benefited from the individual instance, thereby showing that the collaboration led to mutual learning (Table 2).

4.2 | Expectations and experiences differed

In addition to expectations, the interviewees also described early collaboration experiences. These experiences differed from their expectations for several reasons (Table 3). Regarding supervision, the interviewees appreciated that their supervisors welcomed their ideas ("I have not felt any kind of dismissal if I have an idea," Interviewee 7), provided pertinent input ("a lot of feedback and a lot of things to read," Interviewee 22), opened doors for them ("The current collaboration I am doing is also an initiative of my supervisor," Interviewee 1), and the like. However, the supervisors were also experienced as busy ("It's difficult to schedule meetings," Interviewee 20); they tended to refer the PhD fellows to others ("always encouraged me to ask other people," Interviewee 22), and the interaction with some of them was quite formal ("It's easier to approach [a postdoc] with any question and even a stupid question," Interviewee 17). The result of these reservations was less collaboration with the supervisor than expected.

Table 3 also shows that dependencies, time, cultural issues, physical distance, and uncertainty and lack of self-efficacy were issues that complicated and detracted from collaboration. An external reason for complications was the Covid-19 pandemic, which restricted the possibilities for physical meetings. The ITNs managed to facilitate collaboration in spite of the pandemic, but many collaboration possibilities were affected by the social distancing introduced to combat the pandemic. The shift to online formats restrained informal exchanges among the interviewed PhD fellows and made meetings more tiring ("It's just very long Zoom meetings every now and then,"

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 3 Conditions influencing collaboration experiences.

Group	Interviewees	Sample quote
Supervision	18	"In the university, we are used to the supervisor having a huge room [that] is only for him or for her. It's difficult to schedule meetings, and we have to be very formal sometimes with our supervisor. Now, my experience in the company [i.e., during the PhD fellow's secondment] is totally different. They are more open, and I feel like they are all my colleagues." (Interviewee 20)
Dependencies	14	"Some of the PhD students in the consortium are not that dependent on other PhD students for data, but for those who are it's very important to get high quality data and on time." (Interviewee 15)
Time-related issues	13	"If you collaborate with other researchers ..., you need to plan your activities very well [because] you will have deadlines, both for the collaboration and for your other projects, and you have to respect those deadlines. ... This will be a challenge at some point." (Interviewee 8)
Cultural issues	11	"The first step is to have everything quite clear from the moment that we reach out to the potential collaborator. ... Because if we don't have that, it gets quite confusing with the communication process, especially when you have a lot of people from different cultural backgrounds, [like in] this European project." (Interviewee 13)
Physical distance, incl. Covid-19 pandemic	11	"I want to do the secondment in person. It will be a great experience to see how the university works and really connect with other students and the professors. Virtual meetings are good because you are used to it, but you cannot connect very well. You don't have like a coffee break ... because when you are in a virtual meeting, you just talk about the topics on the agenda." (Interviewee 22)
Uncertainty and lack of self-efficacy	10	"I am trying to incorporate my past experiences in this PhD as well. But I am not sure how it will proceed in the future, because it's still in its very start and, to be honest, I don't have much experience in collaboration right now." (Interviewee 1)

Interviewee 7). A prevalent ITN-internal reason for complications concerned the availability and sharing of data, which were often managed by industry partners or supervisors' collaborators. Access to these datasets required negotiation. Rather than being shielded from these negotiations, the interviewed PhD fellows found themselves immersed in them:

> "I have encountered some problems in this regard, basically. So, right now in the wind energy research field, there are many datasets which are not available to the public and they are basically being restricted to only a number of researchers in a group or a specific research project. Those datasets could be ... very beneficial for my research project, but I'm still finding it difficult to get access." (Interviewee 1)

Access to datasets was often restricted to the ITN or project that collected the data because the data were a valuable research resource and because their collection involved resource-demanding measurement campaigns. In other cases, industry partners provided access to data, but their business culture entailed commercial interests that limited the use of the data:

> "With the private sector, we have limitations on what we can do with our data. We have limitations on what we can release to the public [i.e., publish] ... It is difficult to form these collaborations because it is so internal." (Interviewee 21)

The difficulties included that the negotiations over the use of some data and tools were formalized in contracts. At times, these contracts were a long time in the making, or they could only be made with researchers who had reached a certain level of proficiency, measured in months of employment. While such precautions served to protect critical tools from misapplication, they also introduced collaboration bottlenecks that were seen by the interviewees as delaying their progress:

> "I'm having problems to use [a tool from another organization], because I don't have a signed contract with me. I would just have that after one year and a half. That's a bit annoying because I will have the tools at some point, but I cannot have them now because I don't have the contract signed now. If I had the tools right now, I could perhaps go much further in my project than I

TABLE 4 Requirements for collaboration to happen.

Group	Interviewees	Sample quote
Prior experience with collaboration	19	"When I was doing my master's, I was in a group that our supervisor encouraged to do a lot of teamwork. They didn't like us to work individually. If I had a question and he noted somebody else had some knowledge, he wouldn't answer me. He would just direct me to that person or say 'You should ask that person'. It helped me to learn how to communicate with people in a team and get to know their knowledge." (Interviewee 23)
Complementary perspectives	15	"The first thing I will search for is another researcher that has abilities that I don't have and that are related to my topics. ... So, people that can share things I really lack and that can give me tools to overcome some challenges I'm not able to overcome right now." (Interviewee 11)
Similar topics	14	"For me, it would be useful to collaborate with people who come from the modeling side of things, like I do. ... We should, kind of, be producing similar results, and it's really useful to compare whether we actually do as well as [to compare] our different approaches and stuff." (Interviewee 4)
Willingness to collaborate	14	"Everyone is focusing on their project and when you try to suggest something, the other person usually knows everything about it already. But, I'm still like trying to ask around and suggest to other people. I'm very open to have things suggested to me too. I would actually like that, even though there are not a lot of people who do what I do, but it's really helpful for me." (Interviewee 7)
Personality that is open to sharing and communication	12	"In terms of collaboration from colleagues or other PhD students in the same project, I think it's very important to be a communicator." (Interviewee 15)
Professionality	8	"The qualities that I value the most are to be sincere, to be honest—like if you don't know something, please just tell me you don't know. ... And, of course, respect of the timeline. So, punctuality. If you don't have time, it is better to just say that to me." (Interviewee 8)

TABLE 5 Research infrastructure and data.

Group	Interviewees	Sample quote
Use of apps and repositories for collaboration	23	"I'm really lucky that we're already working with Gitlab, [but] I think I can accommodate to any other software that everybody else use." (Interviewee 10)
Sharing data models and measurement devices	21	"We will need to combine these different ways of getting data and later process it together. This is an example of using different techniques and models to compare. With the other institutions that work in modeling, I will use their results to compare with what we have." (Interviewee 9)
Industry collaboration	5	"I can achieve some data from the industrial partner of my PhD, also their expertise in this field could help me. In other parts [of my work], I have this collaboration with some code developers for [modeling] and so on." (Interviewee 12)

am going without the aid of these tools that we have available." (Interviewee 13)

Prior experience from research or industrial settings helped some interviewees form more realistic expectations. The absence of such experience added to the uncertainty experienced by the other interviewees about reconciling their expectations with their possibilities ("I am not sure how I will proceed with this," Interviewee 1).

4.3 | Collaboration was a balancing act

In negotiating the conditions that influenced their collaboration experiences, the interviewees drew on their experience from past collaborations and gravitated toward collaborators with specific characteristics, see Table 4. Prior experience with collaboration made the interviewees appreciate the importance of communication skills to succeeding as a PhD fellow ("The key point in any collaboration is the communication part,"

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/09/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 6 Facilitating collaboration.

Group	Interviewees	Sample quote
Organizational support	21	"They [i.e., the senor researchers in the ITN] are also organizing a lot of training events for us to get to know each other and to connect with each other and within the ITN itself." (Interviewee 17)
Advance planning	13	"The first thing to do is planning. This month, I'm planning my work. For example, tomorrow I have a meeting with my collaborators in order to define what I have to do: the schedule, the time, the rhythm, etcetera." (Interviewee 11)
Clarity of communication	9	"I think the key point in any collaboration is the communication part. When we decide on a given collaboration, we should meet more often, try not to hide anything, follow all the ethical research guidelines, stay professional in our work, and always state our objectives." (Interviewee 5)
Hospitable environment	4	"Collaboration is also an opportunity to socialize, to create a good working environment, which in turn makes us happy. If we are happy at work, the chances of succeeding are higher. So, I think every PhD student should collaborate and be open, because I think that's the way to go." (Interviewee 16)
Being spontaneous	1	"I just feel that the things happened. There wasn't really any planning; things just evolved in the direction of collaboration. ... One of the most important things definitely is to get along with the other person. If that happens, maybe planning won't be needed at all. ... People can plan periodic meetings to discuss their projects, but it's also important to include a spontaneous component." (Interviewee 16)

Interviewee 5). The interviewees also emphasized that for a collaboration to be worthwhile the collaborator should have a perspective that complemented their own and a topic similar to their own. The ITNs made such collaborators readily available because the other PhD fellows' work was also aligned with the ITN goal ("There are two other ESRs [early-stage researchers, i.e., PhD fellows] that are working in my [work package]," Interviewee 18). However, the interviewees did not take for granted that other PhD fellows and researchers would be open to collaboration. Instead, they emphasized that it required willingness ("If you want a collaboration, you have to share something, right?," Interviewee 6), a communicative personality ("It's very important to be a communicator," Interviewee 15), and a professional approach, as in someone who "sticks to the topic and works hard on it" (Interviewee 10).

The conditions influencing collaboration (Table 3) and the requirements for it to be attractive (Table 4) made collaboration a balancing act, at multiple levels. At the organizational level, the interviewees had to negotiate collaborations within and across the boundaries set by their ITN and host university. Some interviewees' projects had cross-organizational collaboration as a prerequisite: "The whole idea [of my PhD] is to use the knowhow and knowledge of two different facilities" (Interviewee 13).

At the content level, the interviewees were seeking complementary perspectives and similar topics to arrive at collaborations that balanced the effort required from each collaborator against the benefit to that collaborator. Imbalanced collaborations would divert other PhD fellows from their goals, and that was a situation the interviewees wanted to avoid: "That is something I really don't want to end up doing because each person has their own goals, and I don't want to be a disturbance or a distraction from their own goals" (Interviewee 10).

At the process level, collaborations added new deadlines that had to be aligned with existing ones to resolve interdependencies and balance the workload over the project period. As a result, the interviewees were time conscious ("there are time limitations on everything," Interviewee 4). In coping with this temporal pressure, some interviewees became hesitant toward communication with people they were not already collaborating with, thereby narrowing their openness to collaboration:

> "We are already quite full with work, and then adding the fact that we have to meet and talk with people who are not in our direct workflow. This takes up more of our time and would be quite a pressure." (Interviewee 7)

At the personal level, all the other levels had to come together in a series of collaborations and individual efforts that would produce a PhD dissertation for the individual interviewee. The balancing between group collaboration and individual dissertation pinpointed the stakes for the interviewees. For some interviewees, the dissertation appeared to determine their overall view on the role of collaboration in their projects:

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

> "I think most of the work will be done by me individually, but then the parts in which other PhD students are involved, there will be some collaboration there. And, of course, some input from the supervisors." (Interviewee 15).

4.4 | Data shaped the collaborations

The interviewees' PhD projects were interlinked through the overall structure of their ITN. This interlinking meant that data dependencies were created at the network-proposal stage, thereby presuming collaborations among PhD fellows who had not yet been recruited. In this way, data shaped the interviewees' collaborations:

> "[The ITN] has been designed in such a way that I have to use data from other PhD students and then give my data to some other PhD students. So, it's a part of the structure of the PhD." (Interviewee 15)

While the interviewees were conscious of not introducing imbalanced collaborations, they did not perceive the collaborations that were built into their projects as restrictions. On the contrary, they saw the built-in collaborative structure of the ITNs as a way of ensuring their access to data and opening up collaborations that were necessary for them to complete their PhD projects. An interviewee elaborated how collaborations were shaped by data and indispensable for progress:

> "[From my analysis], I have received some results—figures and plots about the wind wakes and the deficit wind percentiles in some clusters [of wind farms] for specific time periods. ... But I cannot say whether this is good or bad. Therefore, I have to look for other PhD researchers who are modelling this phenomenon in order to see: 'In this period of time, I have received, for example, an eighth percentile wind deficit. Is this correct?' ... I need to compare my results with other PhD researchers. Then we can say 'Yes, I have received a good result' or, if it's far off, then we have to think [about whether] the problem is in my side or in his side. So, the collaboration is significant to my PhD project." (Interviewee 2)

To manage datasets, models, and analysis tools in ways that made them available for sharing, the ITNs needed a technical research infrastructure that cut across the individual PhD projects. This infrastructure consisted of multiple, loosely integrated apps and repositories that collectively supported collaboration within, and beyond, the ITNs. All 23 interviewees used such apps and repositories for collaboration (Table 5). Typical examples were GitLab for program code, SURFdrive for data files, Overleaf for writing papers, Teams for online meetings, and Slack for informal messaging. Through these apps and repositories, nearly all interviewees were involved in sharing not just data but also data models, measurement devices, and the like. Industry partners were also a valued source of data and knowledge ("I can achieve some data from the industrial partner of my PhD, [and their] expertise in this field could help me," Interviewee 12). They were, however, mentioned less frequently (Table 5), possibly indicating that the interviewees expected industry collaboration to be a somewhat more distant possibility than collaboration with other PhD fellows.

4.5 | Network events nurtured a collaborative environment

In addition to presuming collaboration, the interviewees also described how the ITNs helped nurture it, see Table 6. For example, the ITNs ran 1-week training schools for all the PhD fellows in an ITN to learn about cross-cutting topics and discuss with one another. There were also meeting series to discuss literature and other issues:

> "We have, for example, in our consortium once a month a meeting with all the PhD students. It's always amazing, because we always get to learn something new. It's very free, so we can ask each other a lot [of questions] – not just about technical stuff." (Interviewee 15)

These events created links among the PhD fellows. As a result, several interviewees expressed that they experienced their ITN as a hospitable environment (Table 6), thereby lowering the barriers toward future collaboration. An interviewee appreciated how the ITN and its events allowed for collaboration to emerge spontaneously, rather than always having to be set up (Table 6). Another interviewee explicitly contrasted the ITNs with doing a PhD without such a network, especially when doing the PhD abroad:

> "Many people perceive it [i.e., doing a PhD] as a lonely path towards achieving something,

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

> while having this kind of program that brings many people together is a nice way to create some more community and teamwork that I think is useful for people in general and also for people that come usually from different places." (Interviewee 14)

Apart from the ITN and its events, the interviewees reiterated the importance of clear communication and advance planning in facilitating collaboration (Table 6). These issues could be highlighted at the training schools but had to be practiced by the PhD fellows in their collaborations.

5 | DISCUSSION

In a data-intensive research field such as wind energy, everybody in an ITN is dependent on collaboration and sharing. PhD fellows' socialization into this collaborative approach to research is essential to the field but not easy for the PhD fellows. In the following, we discuss the balancing act they engage in to learn and practice scientific collaboration. That is, we discuss how early collaboration experiences differ from expectations, how collaborations revolve around data, and how collaboration is part of devising a doable dissertation.

5.1 | Between expectations and experiences

Olson, Hofer, et al. (2008) make the point that collaboration is more likely to succeed if the involved actors are ready to collaborate and that collaboration among distributed actors is more likely to succeed if they are also ready to use collaboration technology. The interviewees are ready. They have high expectations regarding collaboration, and they embrace apps and repositories for collaboration. When their early collaboration experiences fail to meet their expectations, it is because collaboration is challenging, comes with dependencies, and requires negotiation. The PhD fellows in the studied ITNs are not shielded from such dependencies and negotiations, which are, on the contrary, inscribed in the network structure as part of the research funding proposal. For the networks to succeed in meeting their research goals, the PhD fellows must succeed in navigating the challenges and dependencies of collaboration.

A key collaborator for the PhD fellows is their supervisor, who has taken part in setting up the ITNs, is knowledgeable about wind-energy research, and has a network of contacts. On the one hand, PhD fellows benefit from their supervisors' knowledge and networks (Paglis et al., 2006; Wang et al., 2021). On the other hand, the interviewees experience the supervisors as busy and somewhat formal. Vähämäki et al. (2021) find that the fellow-supervisor relation can range from hostile to mutual understanding, thereby emphasizing the importance of supervisors' interpersonal and leadership skills in addition to their scientific skills. To meet PhD fellows' expectations, supervisors must take on the responsibility that comes with the PhD fellows' apprentice position (Laudel & Gläser, 2008). While supervisors provide a learning environment by setting up an ITN, their responsibility for PhD fellows' learning and progression also includes facilitating them in initiating, negotiating, and running collaborations within the ITN.

In addition to their supervisor, the interviewees are dependent on collaborations with other PhD fellows, other senior researchers, and industry partners. Scientific collaboration entails a strong expectation of reciprocity (Zuccala et al., 2024), thereby making prior collaborations an important social capital in negotiating new collaborations. The interviewees share the norm of reciprocity, but many of them start out with no social capital in terms of prior collaborations. Therefore, they take care not to impose imbalanced collaborations on others—especially other PhD fellows—who may experience the extra work as a disturbance. That said, the collaborative environment in the ITNs also promotes generalized reciprocity (Putnam, 2000)—the norm that everybody should contribute to some collaborations with some people but may benefit from other collaborations with other people. Generalized reciprocity underpins team science (Stokols et al., 2008) by combining altruism and self-interest in a belief in sharing. This belief may foster expectations that sometimes cannot be realized in a busy workday with many obligations.

The collaborations expected and experienced by the interviewees spanned a range of issues, including collaboration with their supervisor about formulating research questions, with everybody in the ITN about digesting relevant literature, with industry partners about getting access to data, with other PhD fellows about validating results against independent analyses, and with some of all these people about writing up the resulting publications. That is, their collaborations are spread across the stages of the research process (Bukvova, 2009) to learn about diverse issues of importance to their PhD projects. Yet, at the same time as they are collaborating to learn, they are also learning to collaborate. For example, the interviewees repeatedly mention the need to become skilled in communication. For some PhD fellows, the introduction to scientific collaboration is an overwhelming rite of passage (Hara et al., 2003; Torka et al., 2022). To navigate the complexities of collaboration, researchers

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

at every career stage need to balance multiple collaborators, cultures, organizational priorities, research interests, and so forth (Carter et al., 2019; Dusdal & Powell, 2021). The interviewees face this balancing act from the very beginning of their fellowships and must learn to handle it as they go along, thereby adding to an already steep learning curve.

5.2 | Data-centered collaboration

Like other data-intensive research, research in wind energy implies the management and manipulation of large datasets. Ihde (2009) points out that these data become available to human perception only through technological mediations that visualize certain aspects of the data, such as wind wakes. The data are inseparable from the technological tools developed for analyzing them, and the insights obtained from the data are, in turn, inseparable from how these tools visualize the data. The development of such tools, in the form of computer models and analysis software, is central to the interviewees' research, which as a result interweaves data, tools, and insights. This interwovenness means that their collaborations involve elaborating how the developed tools make the data visible and how these technological mediations influence the kinds of conclusions that can be drawn from the data (Mayernik et al., 2013), thereby emphasizing that their collaborations amount to much more than the simple exchange of a set of data points.

Data-intensive research relies on empirical experimentation for collecting data, such as measurements at operational wind farms or experiments in wind tunnels. While research on information seeking tends to suggest that experimentation is an alternative to seeking information from other people (Hertzum, 2023), this study shows that experimentation (in wind-energy research) involves working with other people. The data needed for this kind of research can only be compiled and made available to human perception through team efforts. As a result, the interviewees' collaborations are organized around data. They negotiate access to data, share data, analyze data, compare data, discuss data, validate data, and so forth. The focal role of data in the PhD fellows' collaborations may pertain to their early career stage, which implies that they have many responsibilities directly related to data compilation and analysis. However, a similar focus on collaborative data work is present in other studies of data-intensive research. For example, Shankar et al. (2021) emphasize the importance of collective data curation in responding to the Covid-19 pandemic, Olson, Zimmermann, and Bos (2008) analyze extensive collaborations around scarce equipment for collecting and analyzing datasets in the natural sciences, and Borgman (2015) extends the argument about collaborative data work to the social sciences and humanities. Collaboration around data precedes the publication of the resulting research. In this sense, co-authored publications are the final output showcasing successful collaborations around data.

5.3 | Devising doable dissertations

The many interdependencies among the collaborators in ITNs create uncertainties and risks for the PhD fellows in addition to possibilities and benefits. If their collaborators fail to deliver, the interviewees may be unable to complete their PhD dissertation. Conversely, if they deliver, the dissertation will attain a level beyond what the individual interviewee could accomplish alone—a high-risk, high-reward situation. To balance these risks and rewards, the interviewees must accept the ensuing uncertainty and show confidence in their collaborators. Without prior experience in scientific collaboration, PhD fellows often find this challenging (Delamont & Atkinson, 2001) or even overwhelming (Torka et al., 2022). Being in the first half of their fellowships, the interviewees were hopeful and still in the process of forming collaborations. Their end goal was to devise dissertations they would be able to complete within their 36-month fellowships. Biswas et al. (2021) find that the interdependencies in ITNs tend to make this goal harder to attain, thereby pointing to a tension between contextual conditions and individual success as a PhD fellow (Sorgner, 2022). Devising a doable dissertation is a very real issue for the interviewees; collaboration is a necessary and complex part of that process.

The interviewees included experimentalists as well as modelers. While the experimentalists built and used equipment for collecting wind-energy data, the modelers built and validated simulation models of wind-energy phenomena. Experimentalists and modelers complemented each other: Experimentalists would bring data to the collaboration and need models for analyzing them; modelers would bring models and need data for validating them. These complementary perspectives promised high rewards but also involved considerable interdependencies. In contrast, experimentalists shared similar topics that motivated collaboration among experimentalists, and similarly among modelers. The similar topics—as opposed to complementary perspectives—led to different collaborations, including more everyday support. Regarding the possible impact of the Covid-19 pandemic on the interviewees' collaborations, we note that research during the pandemic continued its trend of increasing

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

collaboration, when measured by the number of coauthors on publications (Cunningham et al., 2021). However, the conditions for collaboration changed. The common ground required to bridge across the complementary perspectives in experimentalist-modeler collaborations is more difficult to establish at a distance (Olson & Olson, 2000) and, thus, complicated by social distancing. The lab is a frequent setting for collaboration among experimentalists but less accessible during pandemic-induced restrictions on the number of simultaneous lab users. Though the PhD fellows talked little about the pandemic during the interviews, we speculate that it influenced their image of a doable dissertation.

5.4 | Limitations

Four limitations should be remembered in interpreting the results of this study. First, we investigated one data-intensive research field, wind energy. Because of the climate crisis, this field attracts substantial research and industry activity and enjoys considerable research funding. Future studies should validate our results in data-intensive research fields other than wind energy and for funding instruments other than ITNs. Second, the interviewees' outlook was restricted to expectations and early experiences because they were in the first half of their PhD fellowships. While the differences between these expectations and experiences are important, collaboration experiences may evolve—positively as well as negatively—during the remainder of their fellowships. Future studies should investigate end-of-fellowship experiences. Third, in addition to the PhD fellows, research networks such as ITNs also involve senior researchers. They were not included in this study but organize the networks and take part in the collaborations. Future studies could investigate the management of data-intensive ITNs by interviewing supervisors and other senior ITN contributors. Fourth, it should be remembered that the interviews were conducted during the Covid-19 pandemic. Validation studies in post-pandemic settings would strengthen the results of the present study.

6 | CONCLUSION

In the field of wind-energy research, PhD fellows' training is often organized in ITNs or similar networks. Through these networks, the PhD fellows are socialized to research, including scientific collaboration. On the basis of interviews with 23 PhD fellows in four wind-energy ITNs, we find that their expectations for collaboration are high, that expectations and early experiences differ, that collaboration is a balancing act, that data shape the collaborations, and that network events nurture a collaborative environment. Unmet expectations call for action, including from supervisors, who can assist in creating collaborations and in expectation management. Learning scientific collaboration involves navigating interdependencies that must be balanced against one another. Some of these interdependencies were created at the stage of writing the research funding proposal, but the interviewed PhD fellows find themselves immersed in resolving them in ways that avoid introducing imbalanced collaborations. The PhD fellows are dependent on collaboration to collect, analyze, share, validate, and compare data. To devise a doable dissertation, the individual PhD fellow in wind energy—and probably also in other data-intensive research fields—must learn and practice data-centered collaboration.

AUTHOR CONTRIBUTIONS

Grischa Fraumann: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Writing – original draft preparation, Writing – review & editing. **Haakon Lund**: Conceptualization, Formal analysis, Funding acquisition, Methodology, Supervision, Writing – review & editing. **Frans van der Sluis**: Conceptualization, Formal analysis, Methodology, Supervision, Writing – original draft preparation, Writing – review & editing. **Morten Hertzum**: Conceptualization, Formal analysis, Funding acquisition, Methodology, Writing – original draft preparation, Writing – review & editing.

ACKNOWLEDGMENTS

We are grateful to the interviewees for sharing their expectations and early experiences regarding scientific collaboration. We thank Camilla Civardi, Anna Maria Sempreviva, Elzbieta Jolanta Wójcik-Leese, and other colleagues in Copenhagen and Lund for commenting on drafts and early presentations of this work.

FUNDING INFORMATION

This study was part of the Train2Wind Innovative Training Network, funded by the European Commission under the H2020-MSCA-ITN program (861291).

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The interviewees were assured that the full interview transcripts would not be shared in publications.

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ETHICS STATEMENT

The collection and processing of the data were approved by the review board of the Faculty of Humanities, University of Copenhagen (file number 514–0081/21–4000).

ORCID

Frans van der Sluis https://orcid.org/0000-0002-3638-0784

REFERENCES

Afewerki, S., & Steen, M. (2023). Gaining lead firm position in an emerging industry: A global production networks analysis of two Scandinavian energy firms in offshore wind power. *Competition & Change*, *27*(3–4), 551–574. https://doi.org/10.1177/10245294221103072

Barnacle, R., Cuthbert, D., Schmidt, C., & Batty, C. (2020). Vectors of knowledge exchange: The value of industry engagement to HASS PhDs. *Higher Education*, *80*(5), 973–987. https://doi.org/10.1007/s10734-020-00528-z

Barros, J. M., Widmer, L. A., Baillie, M., & Wandel, S. (2022). Rethinking clinical study data: Why we should respect analysis results as data. *Scientific Data*, *9*(1), 686. https://doi.org/10.1038/s41597-022-01789-2

Biswas, R., Schiller, A., Casolani, C., Daoud, E., Dode, A., Genitsaridi, E., Jacquemin, L., Liyanage, N., Lourenco, M., Makani, P., Parameshwarappa, V., Riha, C., L Santacruz, J., Shabbir, M., Simoes, J., Trpchevska, N., & Schoisswohl, S. (2021). Doctoral studies as part of an innovative training network (ITN): Early stage researcher (ESR) experiences. *Open Research Europe*, *1*, 13094.2. https://doi.org/10.12688/openreseurope.13094.2

Bitsios, I., Martone, F., Ricci, R., & Arfi, A. (2023). The innovative dimension of the research training programmes under H2020-MSCA-ITN: A methodological approach to track, measure and analyse innovative aspects and provide policy-feedback conclusions. *F1000Research*, *12*, 1020. https://doi.org/10.12688/f1000research.138482.2

Borgman, C. L. (2015). *Big data, little data, no data: Scholarship in the networked world*. MIT Press.

Borgman, C. L., Wallis, J. C., & Mayernik, M. S. (2012). Who's got the data? Interdependencies in science and technology collaborations. *Computer Supported Cooperative Work*, *21*(6), 485–523. https://doi.org/10.1007/s10606-012-9169-z

Bukvova, H. (2009). Research as a process: A comparison between different research approaches. *Sprouts: Working Papers on Information Systems*, *9*(29), 1–20. https://aisel.aisnet.org/sprouts_all/283

Bukvova, H. (2010). Studying research collaboration: A literature review. *Sprouts: Working Papers on Information Systems*, *10*(3), 1–17. https://aisel.aisnet.org/sprouts_all/326

Carter, D. R., Asencio, R., Trainer, H. M., DeChurch, L. A., Kanfer, R., & Zaccaro, S. J. (2019). Best practices for researchers working in multiteam systems. In K. L. Hall, A. L. Vogel, & R. T. Croyle (Eds.), *Strategies for team science success* (pp. 391–400). Springer. https://doi.org/10.1007/978-3-030-20992-6_29

Chung, E., Kwon, N., & Lee, J. (2016). Understanding scientific collaboration in the research life cycle: Bio- and nanoscientists' motivations, information-sharing and communication practices, and barriers to collaboration. *Journal of the Association for Information Science and Technology*, *67*(8), 1836–1848. https://doi.org/10.1002/asi.23520

Clifton, A., Barber, S., Bray, A., Enevoldsen, P., Fields, J., Sempreviva, A. M., Williams, L., Quick, J., Purdue, M., Totaro, P., & Ding, Y. (2023). Grand challenges in the digitalisation of wind energy. *Wind Energy Science*, *8*(6), 947–974. https://doi.org/10.5194/wes-8-947-2023

Cugmas, M., Mali, F., & Kronegger, L. (2024). Longitudinal patterns of scientific collaboration in doctoral studies. *Scientometrics*, *129*(2), 1055–1077. https://doi.org/10.1007/s11192-023-04900-5

Cummings, J. N., & Kiesler, S. (2007). Coordination costs and project outcomes in multi-university collaborations. *Research Policy*, *36*(10), 1620–1634. https://doi.org/10.1016/j.respol.2007.09.001

Cunningham, E., Smyth, B., & Greene, D. (2021). Collaboration in the time of COVID: A scientometric analysis of multidisciplinary SARS-CoV-2 research. *Humanities and Social Sciences Communications*, *8*(1), 240. https://doi.org/10.1057/s41599-021-00922-7

Delamont, S., & Atkinson, P. (2001). Doctoring uncertainty: Mastering craft knowledge. *Social Studies of Science*, *31*(1), 87–107. https://doi.org/10.1177/030631201031001005

Dusdal, J., & Powell, J. J. W. (2021). Benefits, motivations, and challenges of international collaborative research: A sociology of science case study. *Science and Public Policy*, *48*(2), 235–245. https://doi.org/10.1093/scipol/scab010

Dytham, S. (2019). A framework of postgraduate collaboration: Postgraduate collaborative space in a UK university. *Studies in Higher Education*, *44*(3), 446–458. https://doi.org/10.1080/03075079.2017.1371688

European commission. (2021). Marie Skłodowska-Curie actions: Over 25 years of European support for researchers' work.

Grant, B., Nerad, M., Balaban, C., Deem, R., Grund, M., Herman, C., Mrčela, A. K., Porter, S., Rutledge, J., & Strugnell, R. (2022). The doctoral-education context in the twenty-first century: Change at every level. In M. Nerad, D. Bogle, U. Kohl, C. O'Carroll, C. Peters, & B. Scholz (Eds.), *Towards a global core value system in doctoral education* (pp. 18–42). UCL Press.

Gruber, M., Crispeels, T., & D'Este, P. (2023). Who am I? The influence of knowledge networks on PhD students' formation of a researcher role identity. *Minerva*, *61*(4), 521–552. https://doi.org/10.1007/s11024-023-09492-1

Haman, M., & Hertzum, M. (2019). Collaboration in a distributed research program: Islands of intensity in a sea of minimal interaction. *Journal of Documentation*, *75*(2), 334–348. https://doi.org/10.1108/JD-05-2018-0078

Hampton, S. E., Jones, M. B., Wasser, L. A., Schildhauer, M. P., Supp, S. R., Brun, J., Hernandez, R. R., Boettiger, C., Collins, S. L., Gross, L. J., Fernández, D. S., Budden, A., White, E. P., Teal, T. K., Labou, S. G., & Aukema, J. E. (2017). Skills and knowledge for data-intensive environmental research. *Bioscience*, *67*(6), 546–557. https://doi.org/10.1093/biosci/bix025

Hara, N., Solomon, P., Kim, S.-L., & Sonnenwald, D. H. (2003). An emerging view of scientific collaboration: Scientists' perspectives on collaboration and factors that impact collaboration. *Journal of the American Society for Information Science and Technology*, *54*(10), 952–965. https://doi.org/10.1002/asi.10291

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Hertzum, M. (2023). Information seeking by experimentation: Trying something out to discover what happens. *Journal of the Association for Information Science and Technology*, *74*(4), 383–387. https://doi.org/10.1002/asi.24740

Hey, T., Gannon, D., & Pinkelman, J. (2012). The future of data-intensive science. *IEEE Computer*, *45*(5), 81–82. https://doi.org/10.1109/MC.2012.181

Hey, T., & Trefethen, A. (2008). E-science, cyberinfrastructure, and scholarly communication. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the internet* (pp. 15–31). MIT Press.

Holley, K. A. (2020). Interdisciplinarity and doctoral education: Socialization, process, and outcomes. In J. C. Weidman & L. DeAngelo (Eds.), *Socialization in higher education and the early career: Theory, research and application* (pp. 269–284). Springer. https://doi.org/10.1007/978-3-030-33350-8_15

Ihde, D. (2009). *Postphenomenology and technoscience*. State University of New York Press.

Johnson, C. M., Ward, K. A., & Gardner, S. K. (2017). Doctoral student socialization. In J. Shin & P. Teixeira (Eds.), *Encyclopedia of international higher education systems and institutions* (pp. 1–4). Springer. https://doi.org/10.1007/978-94-017-9553-1_296-1

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy*, *26*(1), 1–18. https://doi.org/10.1016/S0048-7333(96)00917-1

Kemp, M. W., Lazarus, B. M., Perron, G. G., Hanage, W. P., & Chapman, E. (2014). Biomedical Ph.D. students enrolled in two elite universities in the United Kingdom and the United States report adopting multiple learning relationships. *PLoS One*, *9*(7), e103075. https://doi.org/10.1371/journal.pone.0103075

Lacourse, K., Yetton, B., Mednick, S., & Warby, S. C. (2020). Massive online data annotation, crowdsourcing to generate high quality sleep spindle annotations from EEG data. *Scientific Data*, *7*(1), 190. https://doi.org/10.1038/s41597-020-0533-4

Larivière, V. (2012). On the shoulders of students? The contribution of PhD students to the advancement of knowledge. *Scientometrics*, *90*(2), 463–481. https://doi.org/10.1007/s11192-011-0495-6

Laudel, G., & Gläser, J. (2008). From apprentice to colleague: The metamorphosis of early career researchers. *Higher Education*, *55*(3), 387–406. https://doi.org/10.1007/s10734-007-9063-7

Liew, C. S., Atkinson, M. P., Galea, M., Ang, T. F., Martin, P., & Van Hemert, J. I. (2017). Scientific workflows: Moving across paradigms. *ACM Computing Surveys*, *49*(4), 1–39. https://doi.org/10.1145/3012429

Mah, D. N.-Y., & Hills, P. (2014). Collaborative governance for technological innovation: A comparative case study of wind energy in Xinjiang, Shanghai, and Guangdong. *Environment and Planning C: Government and Policy*, *32*(3), 509–529. https://doi.org/10.1068/c11101

Mayernik, M. S., Wallis, J. C., & Borgman, C. L. (2013). Unearthing the infrastructure: Humans and sensors in field-based scientific research. *Computer Supported Cooperative Work*, *22*(1), 65–101. https://doi.org/10.1007/s10606-012-9178-y

Mulkay, M. J., Gilbert, G. N., & Woolgar, S. (1975). Problem areas and research networks in science. *Sociology*, *9*(2), 187–203. https://doi.org/10.1177/003803857500900201

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human-Computer Interaction*, *15*(2 & 3), 139–178. https://doi.org/10.1207/S15327051HCI1523_4

Olson, G. M., Zimmermann, A., & Bos, N. (Eds.). (2008). *Scientific collaboration on the internet*. MIT Press. https://doi.org/10.7551/mitpress/9780262151207.001.0001

Olson, J. S., Hofer, E. C., Bos, N., Zimmerman, A., Olson, G. M., Cooney, D., & Faniel, I. (2008). A theory of remote scientific collaboration. In G. M. Olson, A. Zimmerman, & N. Bos (Eds.), *Scientific collaboration on the internet* (pp. 73–97). MIT Press.

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451–476. https://doi.org/10.1007/s11162-005-9003-2

Petzold, A., Bundke, U., Hienola, A., Laj, P., Lund Myhre, C., Vermeulen, A., Adamaki, A., Kutsch, W., Thouret, V., Boulanger, D., Fiebig, M., Stocker, M., Zhao, Z., & Asmi, A. (2024). Opinion: New directions in atmospheric research offered by research infrastructures combined with open and data-intensive science. *Atmospheric Chemistry and Physics*, *24*(9), 5369–5388. https://doi.org/10.5194/acp-24-5369-2024

Putnam, R. D. (2000). *Bowling alone: The collapse and revival of American community*. Simon & Schuster.

Resnik, D. B., Elliott, K. C., Soranno, P. A., & Smith, E. M. (2017). Data-intensive science and research integrity. *Accountability in Research*, *24*(6), 344–358. https://doi.org/10.1080/08989621.2017.1327813

Rigby, J., & Edler, J. (2005). Peering inside research networks: Some observations on the effect of the intensity of collaboration on the variability of research quality. *Research Policy*, *34*(6), 784–794. https://doi.org/10.1016/j.respol.2005.02.004

Scroggins, M. J., Pasquetto, I. V., Geiger, R. S., Boscoe, B. M., Darch, P. T., Cabasse-Mazel, C., Thompson, C., Golshan, M. S., & Borgman, C. L. (2020). Thorny problems in data (−intensive) science. *Communications of the ACM*, *63*(8), 30–32. https://doi.org/10.1145/3408047

Shankar, K., Jeng, W., Thomer, A., Weber, N., & Yoon, A. (2021). Data curation as collective action during COVID-19. *Journal of the Association for Information Science and Technology*, *72*(3), 280–284. https://doi.org/10.1002/asi.24406

Singh, U., & Rizwan, M. (2022). SCADA system dataset exploration and machine learning based forecast for wind turbines. *Results in Engineering*, *16*, 100640. https://doi.org/10.1016/j.rineng.2022.100640

Sonnenwald, D. H. (2004). Managing cognitive and affective trust in the conceptual R&D organization. In M.-L. Huotari & M. Iivonen (Eds.), *Trust in knowledge management and systems in organizations* (pp. 82–106). IGI Publishing. https://doi.org/10.4018/978-1-59140-126-1.ch004

Sonnenwald, D. H. (2007). Scientific collaboration. *Annual Review of Information Science and Technology*, *41*(1), 643–681. https://doi.org/10.1002/aris.2007.1440410121

Sorgner, H. (2022). Constructing 'doable' dissertations in collaborative research: Alignment work and distinction in experimental high-energy physics settings. *Science & Technology Studies*, *35*(4), 38–57. https://doi.org/10.23987/sts.109709

Spencer, D. (2009). *Card sorting: Designing usable categories*. Rosenfeld.

Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science. Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine*,

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

35(2 supplement), S77–S89. https://doi.org/10.1016/j.amepre.2008.05.002

Torka, M., Kohl, U., Mahoney, W. M., & Working Group 2 at the Hannover Conference. (2022). Funding doctoral education and research. In M. Nerad, D. Bogle, U. Kohl, C. O'Carroll, C. Peters, & B. Acholz (Eds.), *Towards a global core value system in doctoral education* (pp. 110–132). UCL Press.

Trisovic, A., Lau, M. K., Pasquier, T., & Crosas, M. (2022). A large-scale study on research code quality and execution. *Scientific Data*, *9*(1), 60. https://doi.org/10.1038/s41597-022-01143-6

Vähämäki, M., Saru, E., & Palmunen, L.-M. (2021). Doctoral supervision as an academic practice and leader–member relationship: A critical approach to relationship dynamics. *The International Journal of Management Education*, *19*(3), 100510. https://doi.org/10.1016/j.ijme.2021.100510

Veers, P., Dykes, K., Lantz, E., Barth, S., Bottasso, V. L., Carlson, O., Clifton, A., Green, J., Green, P., Holttinen, H., Laird, D., Lehtomäki, V., Lundquist, J. K., Manwell, J., Marquis, M., Meneveau, C., Moriarty, P., Munduate, X., Muskulus, M., ... Wiser, R. (2019). Grand challenges in the science of wind energy. *Science*, *366*(6464), eaau2027. https://doi.org/10.1126/science.aau2027

Wang, C., Guo, F., & Wu, Q. (2021). The influence of academic advisors on academic network of physics doctoral students: Empirical evidence based on scientometrics analysis. *Scientometrics*, *126*(6), 4899–4925. https://doi.org/10.1007/s11192-021-03974-3

Zuccala, A. A., Jensen, M. S., Wærn, E. K., & Hertzum, M. (2024). Social capital in academia: Measuring researchers' collaboration habits versus preferences. *Research Evaluation*, *33*, rvae028. https://doi.org/10.1093/reseval/rvae028

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fraumann, G., Lund, H., van der Sluis, F., & Hertzum, M. (2026). Scientific collaboration in data-intensive research networks: PhD fellows' expectations and early experiences. *Journal of the Association for Information Science and Technology*, *77*(9), 1153–1168. https://doi.org/10.1002/asi.70094

23301643, 2026, 9, Downloaded from https://asistdl.onlinelibrary.wiley.com/doi/10.1002/asi.70094 by Morten Hertzum - Royal Danish Library , Wiley Online Library on [14/08/2026]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License