



“We Share All Data with Each Other”: Data-Sharing in Peer-to-Peer Relationships

Eva Barlösius¹

Accepted: 9 January 2023 / Published online: 1 February 2023
© The Author(s) 2023

Abstract

Although the topic of data-sharing has boomed in the past few years, practices of datasharing have attracted only scant attention within working groups and scientific cooperation (peer-to-peer data-sharing). To understand these practices, the author draws on Max Weber’s concept of social relationship, conceptualizing data-sharing as social action that takes place within a social relationship. The empirical material consists of interviews with 34 researchers representing five disciplines—linguistics, biology, psychology, computer sciences, and neurosciences. The analysis identifies three social forms of data-sharing in peer-to-peer relationships: (a) closed communal sharing, which is based on a feeling of belonging together; (b) closed associative sharing, in which the participants act on the basis of an agreement; and (c) open associative sharing, which is oriented to “institutional imperatives” (Merton) and to formal regulations. The study shows that far more data-sharing is occurring in scientific practice than seems to be apparent from a concept of open data alone. If the main goal of open-data policy programs is to encourage researchers to increase access to their data, it could be instructive to study the three forms of data-sharing to improve the understanding of why and how scientists make their data accessible to other researchers.

Keywords Data-sharing · Open data · Collaboration · Community · Working group · Social relationship

The topic of data-sharing has boomed in the past few years. One of the key drivers of this trend are the multitude of science-policy programs for enforcing open-data regulations. In 2011, for example, the National Science Foundation started requiring

✉ Eva Barlösius
e.barloesius@ish.uni-hannover.de

¹ Institute for Sociology, Leibniz Universität Hannover, Schneiderberg 50,
D-30167 Hannover, Germany

applicants to meet data management plan requirements and specify how they will share their data with other researchers (National Science Foundation, 2020). The European Commission, too, committed itself to open data, adopting the FAIR Data Principles¹ and building a European Science Cloud to provide an open-data infrastructure for the creation of a European data space (European Commission 2016). In these programs open data has been understood to mean making research data accessible to everyone by, for example, storing them in research infrastructure as archives or data centers.

In addition to these political goals, there are scientific reasons for greater data-sharing, the main one being that science has become very “data-intensive” (Dorta-Gonzalez et al. 2021: 2222), as captured by the search-engine keyword *data-driven research* and even characterized as a new scientific paradigm (Hey et al. 2009). With past ways of handling research data now considered inconsistent with the requirements of data-intensive research, there is a need to adopt new practices, such as those that facilitate the reuse of data by making them open to all who are interested in them.

A look through studies on data-sharing reveals considerable effort invested in analyzing the obstacles to increased open data-sharing. One hurdle is that enforcement of open-data regulations is often inhibited by existing research practices within working groups and scientific cooperation (e.g., Levin & Leonelli 2017). A striking feature of this analytical perspective is that data-sharing seems to be strictly associated with the act of making the data open and accessible to all in anonymous relationships, such as those created by research infrastructures. Peer-to-peer data-sharing practices within working groups and scientific cooperation, the social rules they entail, and the social meaning on which they are based have received scant attention. Little research has been devoted to the joint use of unpublished research data and the ways in which they are shared in personal social relationships such as working groups and scientific cooperation. Yet these kinds of data-sharing occur quite naturally and almost daily in the research community. Some of them, including the communal use of data within a working group, are almost obligatory, whereas others typically do not begin until the researchers have agreed on how they will share the generated research data.

Why and how do these kinds of data-sharing work, which social rules govern them, and how are they justified? These questions address the sociality of personally shared data. To answer these questions, it is crucial to consider the entire research process, not just analyze only the decision for or against open sharing. With whom, in which phase of the research process, and at which step of data-processing do scientists share their data? How do the researchers justify the decision to share or not share their data. Seeking to comprehend these kinds of data-sharing, I apply Max Weber’s (1922/1978) understanding of *social relationship* in order to be open to general social phenomena such as trust and mistrust in social action or the sense of belongingness. I choose such a general theoretical concept because these kinds of data-sharing exist within the scientific field and are shaped by it, but they are ruled by general social meanings and expectations, such as the need for legitimacy and desire for control. The empirical material for my study consists of interviews with 34

¹ FAIR stands for findability, accessibility, interoperability, and reuse.

researchers representing five disciplines—linguistics, biology, psychology, computer sciences, and neurosciences.

This paper begins with a review of the literature on data-sharing literature and focuses on the forms of data-sharing that have been analyzed. I then develop the theoretical concept for the analyses, which are based on Weber's (1922/1978) notion of social relationship. In a further step I describe and justify my empirical material and explain the methodological approach to it. I thereafter describe the social meanings that influence the practices of using research data with other researchers. This groundwork enables me to identify three social forms of data-sharing: closed communal sharing, closed associative sharing, and open associative sharing. I conclude with reflections on what the results mean for the agenda of open data.

What is Known about Different Social Forms of Data-Sharing

The literature on data-sharing has grown enormously in recent years, mainly focusing on the reasons that there are so many barriers to initiatives for making research data open to all. One often studied barrier is the attitude that academics have toward data-sharing: They assert that they would like to share their data but that their colleagues would not reciprocate with their own (Dorta-González et al. 2021; Thoegersen & Borlund 2021). This impediment implies that researchers regard data-sharing as a social relationship that ought to be governed by social reciprocity. Several analyses have confirmed this result and have brought to light additional reasons that scientists are sometimes unwilling to share their data: technical problems, the excessive amount of time needed to render data usable, meager academic recognition, and concerns about misuse and misinterpretation of the data (Bezuidenhout et al. 2017; Fecher et al. 2015; Maienschein et al. 2019; Tedersoo et al. 2021; Velden 2013).

A common feature of many studies on open data-sharing is that they have closely examined the researchers' resistance to making their data accessible to all but have often simply assigned empirically untested attributes to existing data practices within working groups and scientific cooperation (i.e., data-sharing in peer-to-peer relationships). A frequent tendency has been to characterize phenomena as open in contrast with closed data (Wessels et al. 2017), public as opposed to private data (Levin & Leonelli 2017), formal versus informal practices of data-sharing (Stamm 2018), or data as a public good versus research data as personal ownership (Kansa 2014). Generally, the authors of these studies strove to understand why the transition to open data meet as much resistance as they do among researchers. Leonelli & Tempini (2020) were interested in the reasons for the "insistence by the researchers working within their different traditions to tailor their data practices" (p. 4). Wessels et al. (2017) asked why "open access to research data will need to be embedded in current research practices" (p. 113). Compared to the purely personal use of data, open data-sharing has slight impact on a researcher's reputation (Yoon & Kim 2020). So how can that effect be enhanced (Linek et al. 2017)? What anxieties do scientists have about making their data open (Tenopir et al. 2011, 2020)? As this brief overview illustrates, data-sharing practices beyond making data open and accessible to all have been mainly considered with regard to how they hamper the implementation of open-

data programs. The question of how data practices unfold socially in peer-to-peer relationship has scarcely arisen.

The social qualities of existing data practices and the typical social ways of sharing data are seldom the focus of detailed investigation. The limited findings of the few studies that have described data practices such as those within working groups and scientific cooperation are easily summarized. For instance, characteristic social relationships for these kinds of data-sharing are “personal collaboration” (Stamm 2018: 7) or “peer-to-peer” sharing (Belk 2007: 129; Yoon & Kim 2020: 186). They were listed, though not described, by Whyte and Pryor (2011), who distinguished between four types (a) “private management” (p. 205), by which they meant sharing data with members of one’s own research group; (b) “collaborative sharing” (p. 207), that is, the sharing of data within a network; (c) “peer exchange” (p. 207), meaning the granting of access to friendly colleagues; and (d) “community sharing” (p. 207), the term for sharing of data with members of a research community. These kinds of data-sharing are usually governed by a “uniquely high level of trust” and “a social order in which personal [self-collected] data are thought to be safely and transparently managed” (Axelsson & Schroeder 2009: 223). These few words cover the most important messages of these studies, which emphasize the researchers’ understanding of peer-to-peer data-sharing as a social relationship. A systematic study of these sorts of data-sharing practices is lacking, however. Some authors have indicated why it could be useful to know about these practices of data-sharing for improving implementation of open-data programs. Ankeny (2017), for instance, pointed out that the concept of transparency and availability for open data “would not necessarily result in what many would consider to be collaboration or coproduction” (p. 307), for the requirements of open data-sharing suspend the “shared collectivity” (p. 307) of the peer-to-peer and community sharing. Tenopir et al. (2020) underscored “that scientists seem to be more willing to share their data as a direct response to a request made by their peers” (p. 5). These few findings on peer-to-peer and collaborative data-sharing suggest that joint use of data requires a social relationship in which the data are not simply passed on but also made the focus of a scientific exchange.

Despite the prevalence of data-sharing as a research issue, relatively little is known about peer-to-peer data-sharing. In which social relationships does it take place? What social meanings are associated with it? What specific social forms of personal data-sharing exist? These questions revolve around the sociality of data-sharing.

Data-Sharing as a Social Relationship

To understand the sociality of data-sharing in peer-to-peer relationships, it is helpful to use a framework that allows identification of general social meanings and expectations that are not specific to science. Drawing on Max Weber’s (1922/1978) concept of social relationship, I propose that sharing represents a social action that takes place within an established social relationship or from which a social relationship develops.

To Weber (1922/1978), social relationships are regulated forms of social actions. They are regulated in such a manner that actors associate similar *subjective meanings* and *expectations* with them, providing the basis for the way in which the actors react

to each other. This viewpoint raises the question of how the actors can assure that they have the same idea of the social relationship. Weber identified two social mechanisms for rendering the actors' expectations of the social relationship "objectively symmetrical" (p. 27). The first one creates an agreement about the "subjective meaning of the social relationship" (p. 28) and guarantees that the actors will orient "their future behavior" (p. 28) to the agreement, which may be oral or written. The second mechanism for matching the expectations of the social relationship arises from what Weber called "certain empirical uniformities" (p. 29), which result when the actors repeatedly act in a uniform manner, developing the same "subjective meaning" of the social relationship (p. 29). Weber specified two conditions for repeated social actions that result in same subjective meaning of a social relationship. The first condition is that "the practice is based upon long standing" and has thus become a "custom" (p. 29). The second condition is "'determined', insofar as 'the actors' conduct is instrumentally oriented toward identical expectations" (p. 29), for they share the same goal with the relationship.

Another important aspect for understanding sharing as social relationship is that it "may be guided by the belief in the existence of a legitimate order" (p. 31), which authoritatively regulates the meanings of and expectations for social relationships. The legitimacy of such a framework or arrangement may be ensured, for example, by a convention or a law that derives the order's legitimacy from "positive enactment which is believed to be legal" (p. 36).

Weber (1922/1978) focused on two ways of characterizing social relationships. The first relates to the orientations of the social actions. On this point he emphasized conflicts (a social relationship by which the actors assert their own will against that of other actors) and competition (which "consists in a formally peaceful attempt to attain control over opportunities and advantages which are also desired by others," p. 38). The second way of typifying social relationships relates to how the social relationship is established. Weber distinguished between "communal" and "associative" social relationships (p. 40). If the actors have a sense of belonging together, then their social relationship is communally established. If the actors are related to each other by an "adjustment of interests or a similarly motivated agreement" (p. 41), then their social relationship is associatively established. Weber also differentiated between "open and closed relationships." A relationship is open if its social order "does not deny participation to anyone who wishes to join and is actually in a position to do so" (p. 43). A closed relationship excludes the participation of certain persons. The exclusion of persons is justified by the fact that they do not share the same understanding of that relationship. These two differentiations into communal and associative and into open and closed facilitate their combination for the purpose of describing diverse kinds of social relationships, including the one on which this article focuses—peer-to-peer relationships, which are established by researchers themselves.

How can Weber's (1922/1978) concept of social relationship be applied to peer-to-peer data-sharing? First, it facilitates conceptualization of data-sharing as social action that takes place in a social relationship. It also makes it possible to assume that the way the data are shared is linked to the subjective meanings and expectations the researchers have of the social relationship of data-sharing. For instance, they might expect problems, or they might have a sense of belonging together even though they

do not work in the same laboratory. They might know each other for a long time, or they may be members of the same scientific community. The data practices may vary according to whether the relationships are based on customary rules, on instrumental orientations, or on an explicit agreement. It could be that the researchers orient the relationship to a given legitimate order of data-sharing. Lastly, it can be assumed that social relationships of data-sharing are differentiable into communal and associative as well as open or closed social relationships and that combining them may create specific social forms of data-sharing.

A number of questions follow from these assumptions. What subjective meanings do the researchers ascribe to the social relationships that are established for data-sharing? How are the subjective meanings related to communal or associative relationships and to closed or open relationships? How do researchers check whether they share the same meanings and expectations about data-sharing? How do researchers assure that the persons with whom they share their data will abide by the promises concerning the process of data-sharing? How is the legitimacy of the data-sharing practices guaranteed?

Empirical Material: Two Research Projects on Data-Sharing

The empirical material for my analyses came from two research projects, “Practical Handling of Research Data: How Do Researchers Protect Their Research Data?” and “On the Relationship between Concepts of Originality and Practical Orientations of Data-Sharing.”² Both projects included researchers’ responses during interviews about projects for which they had received funding from the German Research Foundation (Deutsche Forschungsgemeinschaft, DFG).³ DFG projects were chosen because they guarantee the greatest scope for self-organization and the handling of research data. This decision ensured identical formal conditions for conducting research projects, particularly with regard to the handling research data. Scientists submitting a research proposal to the DFG “explain the nature, scope and documentation of the data” produced in the project and “how they will be stored” for ten years (DFG 2021). In addition, they should “discuss the possibility of subsequent reuse by other researchers” (DFG 2021). There are no stipulations or standards attached, so the scientists can decide for themselves how to share data, if at all. Such projects were selected for which the researchers themselves generated research data, mainly through experiments. These projects were to have advanced to their final phase so that data analysis had already begun. Lastly, they did not fall under private data-protection. Both projects used semistructured qualitative interviews with almost identical structures, except for the concluding items. The interviews pertaining to the first project ended with items about data protection. The second project’s final items inquired into views on scientific originality. Neither concluding block of items was included in the present study.

² The first project has been financed by the Leibniz Center for Science and Society (LCSS); the second, by the German Research Foundation (DFG).

³ I thank Friederike Knoke, Michael Pook-Kolb, and Saskia Rebea Schrade for collecting the data.

In the first research project four to six interviews with postdocs from linguistics, biology, and psychology were conducted because they already had experience with data-sharing. In the second project about six interviews were conducted—in biology, computer sciences, and neurosciences. For this project doctoral students and principal investigators were interviewed, with the latter group generally consisting of professors. What I especially sought to learn from the principal investigators was how they explain and justify the data-sharing practices they use in their working group and who decides on them. I wanted to find out from the doctoral students how they are familiarized with the rules and data-sharing regime.

The two projects entailed 34 transcribed interviews, each lasting from 45 to 90 minutes. The interviewees were requested to explain what they consider research data to be in the selected project and how the data are processed. They were also asked with whom they exchange information about non-published research data, with whom they are less forthcoming, and to whom they tell nothing. The interviewees were then invited to talk about a presentation that might draw on data not yet been published at a conference. The next block of items was designed to elicit the manner in which research data are shared for publications. The respondents were requested to remember one of their publications and explain to whom the data were accessible at which stage of their processing.

Analytical Method and Creation of Categories

Using Weber's (1922/1978) concept of social relationship to identify the social meanings and social forms of data-sharing in peer-to-peer relationships, I proceeded with structured content analysis (Mayering 2008), by which theoretically derived categories are used to analyze the empirical material. From Weber's concept I took three theoretical categories: subjective meaning, legitimate order, and social relationship. After the first reading of the transcripts, I developed a fourth category, data-processing phases, which made it evident that the interviewees, when speaking of sharing data, generally mentioned the stage of processing the data had been in. These four categories were my main ones for coding the interviews, though I remained open to inclusion of others.

In my first main category, subjective meaning, attention focused on interview segments in which the researchers spoke of their own orientations, such as the personal meaning of research data, how they deal with them, and what they expect from other scholars when providing them with data. After analyzing all passages in which the interviewees described the subjective meanings the data have for them and how they care about these meanings, each statement was classified into one of three subcategories: (a) personal value, when researchers emphasized the importance they attach to the data, (b) acknowledgement, when the researchers expressed the subjective meanings of the data by explaining how due credit was to be given to the person sharing them; and (c) protection, when the researchers articulated the subjective meanings by describing how they protect their data against use by other researchers. Analysis of the subcategories of subjective meaning revealed that protection of one's scientific achievements is what orients the sharing of data.

I proceeded similarly with the second main category, legitimate order. I combed through the interviews for clues to the rules, and requirements that the respondents followed as they explained and justified their declared reasons for sharing or not sharing data. The answers revealed five main reasons, which I applied as subcategories: (a) scientific community, when the respondents referred to the validity or nonvalidity of rules within their discipline; (b) trust, provided they mentioned trust or lack thereof as a motive for sharing or not sharing; (c) concurrence and scooping⁴ (Bezuidenhout et al. 2017: 470), when the respondents justified their data-sharing practices on the grounds of scientific competitiveness; (d) control of disposal, when they justified their data-sharing practices by indicating that they have no right to decide on what happens with the data, and (e) monitoring, when the respondents legitimized their checking up on what happens with the data.

To record the responses to items relating to my third main category, social relationships, I marked the survey passages in which the interviewees characterized the social relationships in which they share data. I analyzed these texts for those relationships that involved use of research data with other scholars. I then summarized the text passages about different kinds of relationships and groups of actors. The fourth and final main category, phases of data-processing, was analyzed for the steps through which the data progressed in the research process and for the manner in which the researchers described the steps.

To identify the social forms of peer-to-peer data-sharing, I next analyzed which relations exist between the four main categories. Both the manner in which the researchers protected their scientific performances and the explanations and legitimations to which they referred turned out to be closely interconnected with the reported social relationships and the different phases of data-processing. Three social forms of data-sharing among researchers emerged: closed communal sharing, closed associative sharing, and open associative sharing. The following sections present the four main categories and describe the three social forms of data-sharing. The subjective meanings and the forms of data-sharing diverged very little across the disciplines represented in this study, and no gender-specific differences surfaced. Disparity in academic status did become apparent, however, for the doctoral students, as expected, were new to data-sharing. Before presenting data at a conference, they were told by the postdocs and professors which data were allowed to be imparted and how.

Sharing Data Peer to Peer

Protection of One's Scientific Achievement

When asked why research data have personal meaning, the respondents explained that many of their own ideas inform the data's generation and result from their scholarly

⁴ Scooping means the possibility that one's data will be taken and published by others in work on the same topic.

achievements. They referred to their “own thinking” (Bio09, postdoc, f, 296)⁵ and a “great deal of their own intellectual work” (Psycho01, postdoc, m, 141) that had been invested. Underscoring the personal meaning involved, some respondents stated that the research data had become “an important part of me,” and others said that the research data were “of course, sort of like the scientists’ children” (Neuro05, prof, m, 65).

The respondents found it important that the scientific achievement represented by the data be recognized. They said that they make sure they are cited as a coauthor or mentioned in the acknowledgements when the data they have generated are passed to other researchers. The access they allow to their data is something they see as a social relationship. They orient themselves to the way in which they can profit academically from sharing this resource as well as to the purpose for which their data is to be used by those who receive them. For example, a postdoc from the neurosciences reported that he negotiates the benefit to him when asked to share his research data, for publications and citations are the currency in science and academics (Neuro06, postdoc, m, 79). Becoming a coauthor is considered appropriate if the idea underlying the data is crucial to their further use.

Naturally, we have a vested interest in defending the fact that these are data we have developed from our own ideas, and you don’t want to just hand over the idea so that someone else can turn around and publish with it. (Neuro04, prof, m, 73)

The respondents take care to ensure that they receive credit for their research ideas by being named a coauthor. If they have “no intellectual input to [the article], then they would say, ‘If you can clearly explain why you want to have the data, take them. ... [and] mention me in the acknowledgements’” (Bio04, prof, m, 157). Other respondents, too, were satisfied with an acknowledgement if the data were to be used for treatment of a research question having little or nothing to do with their own.

The respondents were particularly guarded about their own scientific achievements when including unpublished data in a presentation if it is not socially binding that the data be attributed to them as authors. The researchers stated that they weigh what they offer because they understand presenting as “making open” in the sense that they share their data with individuals they do not know personally. As one neuroscientist put it,

At that moment [i.e., the lecture] I give it freely. There are thirty people there making live video recordings or taking photos. At that moment I have to figure

⁵ The original German quotations from the interviews have undergone a bit of language editing, but their messages have not been changed. The sources of the quotations are coded by discipline, interview, academic status, gender, and paragraph: Bio=biology, Com=computer sciences, Ling=linguistics, Neuro=neurosciences, and Psych=psychology. The first two digits identify the interview; the subsequent digits, the paragraph in the transcript. Because this study draws on interviews in biology from both projects, the two coded references to those from the first project are capitalized (BIO). Academic status is indicated by abbreviations: prof=professor, postdoc=postdoctoral candidate, doc=doctoral candidate, and gender is identified as either “f” (female) or “m” (male).

that the data will be published. Depending on who the audience is, I have to consider at what stage of maturity the data are. If they are really far advanced, then I can show them to a professional audience in cases where I think I am quick enough to publish them. If I don't think so, then I have to consider whether I'm really going to share them or rather wait until next year. (Neuro03, postdoc, m, 59)

When presenting to an audience with people unknown to them, the interviewed researchers essentially protect their future publications, their research ideas for follow-up projects, and their complicated experimental setups and experiments that they have “gotten to work.” They are careful to present their data in such a way that other scientists cannot leapfrog them. One defensive strategy that was described in many variations is to avoid giving precise information. The protein is renamed, or named with only one letter; the gene name is not shown; only a low-resolution jpg image is viewed; a graph excludes the raw data; or a figure is drawn, but no individual values are given.

Legitimate Order: Explanations and Justifications

The interviewees explained the conditions under which they could imagine sharing data and justified the conditions that rule it out in their minds. When referring to their scientific community, they justified their practices mainly by arguing that it was common to exchange and support each other within that circle. A biologist stated, for example, “We're plant people, we make up a small subject, so we help each other. We tell each other we have the vector, and then the other colleague says, ‘Oh, I'd like to have that’, and then we do it—yes, even if they are not published” (Bio02, doc, f, 91). Notably, the interviewees did not mention the scientific community when justifying why they do not share data. They tended instead to point to something else: “You are in competition with other research groups that are doing similar things. I have to make sure I don't feed my competition with data. I would be relatively bad at business [if I did share]” (Com07, postdoc, m, 71).

Another reason that respondents gave for not sharing unpublished data was their concern about scooping others. Recounting an instance of being scooped, one biologist said, “We've also had bad experiences.. .. [After] we presented data from a doctoral student, we were scooped, and it really burns” (Bio02, doc, f, 93). The respondents frequently referred to trust as a condition for data-sharing before publication. They stressed, however, that trust develops only through peer-to-peer interactions. In general, they cited shared research experiences as the basis for trust, or the framing of data-sharing as research collaboration. “When I trust someone,” said one neuroscientist, “I know. .. they respect my work, [and] I respect theirs” (Neuro03, postdoc, m, 73). Lack of power to decide for oneself whether to share the data in the first place (control of disposal) gave rise to the next justification for not doing so. This explanation is heard mainly from doctoral students, sometimes by postdocs. They explained that they must ask their superior because that person decides on how the data are to be handled.

The justifications discussed so far invoke the binding nature of social rules or their nonbinding nature and rule-breaking. Articulating a justification for not sharing data, the scientists argued that those who generated the data have the prerogative to control to whom they make the data available. It would be the obligation of the data's originators to determine whether the research purpose for which the data are being used is scientifically meaningful. To this end, they asked what the data would be used for. If they judged it to be epistemically appropriate, they shared the data: "I would say, if it makes sense for the research. .. then I would just share the data" (Com08, doc, m, 75). The explanations and justifications predominantly involve confirmation that the scientists to whom the respondents make their data available have the same notion of the legitimate order about the data transfer as they do and will act accordingly.

Social Relationships

The analysis of the interview passages in which the researchers talked about research interactions resulted in three different kinds of social relationships: group relationships, project cooperation, and step cooperation. Group relationships pertain between the members of the same research group, mostly members of the same laboratory. The respondents generally spoke of "my group" or "our group," "my laboratory" or "our laboratory," and "us." Other respondents speak of the "unit" or the "research unit" (Com01, prof, m, 163; Bio07, prof, m, 138). All the research work, from data generation to publication, is described as a joint process, though not everyone participates equally.

The next two types of social relationships—project cooperation and step cooperation—represent two poles between which a wide range of social interactions take place, depending on the scope of joint research.

There are just projects that we plan together from the beginning, with everyone doing their part. We sit down together once a month. Then there are projects on which I cooperate from time to time because I need a certain piece of information about a certain system. (Bio02, doc, f, 35)

Project cooperation consists of relationships that span the entire research process. The researchers plan and conduct a project together from start to finish. Step cooperation consists of research interactions that entail only one or two steps of a research process. Project cooperation is characterized by common interests, the only condition through which "cooperation really comes to life" (Neuro05, prof, m, 73). As a rule, the cooperating scientists have known each other for some time and have a personal relationship. In many instances they have previously engaged in "step cooperation" because "if a connection does not already exist" (Neuro01, prof, f, 65), it must be established before an exchange takes place. This point was emphasized by a biologist as well: "I have to be sure beforehand that we are in the same boat. Cooperation develops best if I know the scientists, say, through a conference, a personal meeting, or a fairly long period of correspondence" (Bio05, postdoc, f, 78).

In step cooperation, scientists cooperate for one or two steps of a project because, for example, they lack the necessary experimental setups or equipment.

In collaborations it is often the case that someone has particular expertise, some technique that those investigating something haven't mastered. There are then two choices. Either they spend a lot of time learning the technique, or they consult the person who knows the technique. (BIO08, postdoc, m, 85)

In both types of collaboration, the scientists agree on how their working relationship will proceed, such as who will provide what services and how they will be scientifically compensated. The social relations are driven by common interests, as Weber (1922/1978) described for associative social forms.

Data-processing Phases

When the respondents talked about how they share data, they usually mentioned distinct phases of data-processing. To understand which data the researchers share with whom, it is important to identify the different phases of data-processing and to know the steps relating to them. The respondents distinguished mainly between four phases, with fluid transitions between them. In the first phase they work with the raw data stemming directly from the experiments: "The raw data are very often image-based, that is, images as they come out of the microscope" (BIO02, postdoc, m, 79). Raw data could be microcopy images, audios, photos, or completed questionnaires, for example. In general, raw data are data before any treatment.

The second phase involves what is known as data preparation, and most of the respondents called this phase prepared data. For example, a biologist explained that the team members check whether the raw data "were OK, whether there are any doubts about these data, whether we need to repeat the experiment" (Bio05, postdoc, f, 53). She added that "dubious" raw data are removed, "are not included in the analysis," and "simply don't belong there. What is left are the data I call cleaned" (Bio05, postdoc, f, 55). Scientists in computer science and psychology described this processing step similarly: "You take out the data sets that are obviously nonsense" (Com02, prof, m, 102). The raw data are checked "for plausibility," "whether the data can be coherent in the sense of whether the values are in the normal range" (Psych04, postdoc, m, 145).

The third phase entails the data analyses. The data are processed for testing the research hypothesis (processed data). To conduct this phase, the researchers asked themselves "under which aspects [the data] should be examined" (Com04, prof, m, 57). It is often associated with selecting the data that relate to the research hypothesis. The respondents referred to these data as "processed," "analyzed," or "evaluated." One psychologist said that the data are "summarized in a meaningful way" during this stage, "so that we can do a meaningful analysis with it" (Psych09, postdoc, m, 110). Characteristically, the data are extracted with an eye to "specific areas of interest" (Ling09, postdoc, f, 108).

The fourth and last phase consists of the analyzed data that many interviewees called final data, which in most cases are identical with those published in articles.

For publication the data are often presented in diagrams, graphs, or pictures. “The final step is then to create data diagrams,” which are “transformed into finished graphs by evaluation software” (Neuro03, postdoc, m, 47).

Three Social Forms of Data-Sharing in Peer-to-Peer Relationships

In this section I describe three social forms of data-sharing that emerged from the analysis of the relationships between the four main categories. They are presented in an ideal-type way to distinguish them clearly. Of course, each form varies, and there are transitions from one to another. The three forms differ according to whether they involve communalizing or socializing relationships, and whether they are open or closed. The difference between the relationships is linked to the subjective meanings and to the legitimate order the respondents associate with sharing. This combination explains which data are made available: raw, prepared, processed, or final.

Closed Communal Sharing—Feeling of Belonging Together

Closed communal sharing is determined by social relationships that arise from belonging to a working group. People who work together share what Weber (1922/1978) described as “relatively permanent social relationships between the same persons” (p. 41). It is not simply the duration of their work together that shapes the social ties necessary for a group relationship; it is the feeling of belonging together. This feeling also arises from the fact that those people share the scientific results and success. In a working group the members usually take care of each other, helping each other with experiments, sharing authorship, and ensuring that the junior scientists can complete their doctorates. This interaction includes the sharing of all the research data at all four steps along the way.

The sharing of data in the working group is not an independent social phenomenon. It results from the way the members of the working group conduct research with each other: It is determined by affiliation. Almost all the interviews contained some variant of the phrase, “In our working group we share all data with each other.” The fact that the scientists perceived themselves as belonging to a social group is particularly apparent in the subjective meanings they associate with closed communal sharing. The repeated courses of action within the group assured them that they shared the same subjective meaning of sharing data. For example, when asked who receives his data, a computer scientist answered, “My coworkers, of course. They get them anyway; they are like me” (Com03, prof, m, 153). His phrase “they are like me” expresses that he and his team members attach the same subjective meanings to data-sharing, so he does not have to negotiate with them specifically about the use of the data. The importance of this process of repeated social actions and the importance of the relatively permanent social relationships becomes clear in the following excerpt:

I have been working together with my colleagues in the field for a very long time, since the beginning of our bachelor studies. We have also been friends the whole time. We mutually accept the idea that we’re developing [this topic]

together, and somehow it becomes clear that this person has taken it now because the idea came from the person. He is legitimized to follow through on the project, even though we have exchanged ideas about it. (Com09, doc, m, 17)

These scientists were sure that their scientific achievements are protected in and recognized by the group, so they shared everything.

The experience of agreeing on the same subjective meanings attached to the social relationship of data-sharing also surfaces in the expectations underlying a legitimate order of data-sharing. The interviewees explained closed communal sharing with the fact that only “the highest level of trust exists within the working group” (Neuro03, postdoc, m, 61). Trust that the members orient their actions to the group’s legitimate order of data-sharing was the most salient justification for closed communal sharing. It was central to the collaborative framework of data-sharing that research data are treated as a common good that belongs to the working group. As a biologist starkly put it, “I am already aware that my data belong to the institute, to this working group” (Bio07, prof, m, 138). He presumed that, if he leaves the group, he may not take with him the data he helped generate.

Data are generated by a social group, which does not mean that all members have equal part in generating the data. The generation of data is understood as a task of the community. Hence, the research data are scientifically exploited by a community, although that framework does not mean that all members can have equal part in “harvesting” all the data, that is, in transforming them into scientific achievements. Instead, care is taken to recognize the main generators of the data for their achievement, especially through primary authorship.

Closed Associative Sharing—Participation by Agreement

In this kind of sharing, research data are made available to persons outside the working group. The access to prepared or processed data, not raw data, is usually given: “[R]aw data do not go out. If [anything], then processed data go out” (Bio09, postdoc, f, 275–276). Generally, the justification for making the data available to other scientists lies in the research itself, such as the desire to address further research questions or to gain access to necessary methods. In many cases the social relationships for sharing the data have to be established first. An exception is project cooperation, which, as noted above, usually rests on established research relationships and thus comes with the social prerequisites for sharing data. This kind of sharing, which is based on closed associative relationships, typically involves gaining agreement on the subjective meanings of their data-sharing. This agreement produces the associative character of the social relationships, which arises from the adjustment of the researchers’ interests in sharing the data and thus has instrumental underpinnings. These relationships are closed because they include only the researchers with whom the agreement was made.

The path to an agreement on data-sharing was described by the respondents as a social process: “You’ve agreed, but I have to be sure that the person on the other end works just as fairly as I do on my end. And that’s the way you coordinate

with each other” (Bio07, prof, m, 101). The agreement process transitions an open relationship into a closed one because the agreement about data access applies only to the scientists sharing them.

The agreement comprises mutual recognition of the scientific expertise brought into and developed during the collaboration. If joint publications are agreed on, then the data providers generally discuss in advance “what our share in the publication might actually be. It is conceivable that [the recipients] have their own research question and use our data for it with our consent, and we create a joint publication out of it” (Neuro01, prof, f, 63). To this end, the recipients would have to “pledge that these data will not be misused, that is, utilized for their own purposes without involving me” (Neuro04, prof, m, 77). If no joint publication is sought, then it is usually specified that the provider of the data should be cited. To protect one’s own scientific achievements, it is also agreed that access to the data may be given only to the parties to the agreement, the purpose being to prevent the data from “being shared” outside the closed associative relationship.

The respondents explained this social form of data-sharing by citing mainly three requirements of a legitimate order. First, associative closed sharing, like communal closed sharing, involves trust as an indispensable condition, the difference being that trust in associative closed sharing is not taken for granted; it must be created: “Trust builds during the interaction”. It is not like a button is pressed and out comes the data. There is a bit of back-and-forth instead” (Bio06, postdoc, m, 127). Second, closed associative sharing is rooted in the assumption that researchers are not competing scientifically. When data are made available to scientists doing research in a completely different field, agreement about the use of the data is less restrictive. For example, one biologist related that if she were approached by a scientist “doing something completely different that doesn’t affect me at all, I would say, ‘here, take it’” (Bio02, doc, f, 146). Third, respondents justified the right, and sometimes the duty, to control the way in which the data may be used in further research, saying that they have generated them. These scientists make their data available for other research only if they agree with the research purposes: “I want to know what will be done with the data and how they will be used.. .. That is an essential point in order to decide whether the data will be used appropriately” (Com07, postdoc, m, 101). A few respondents made further stipulations for closed associative sharing, referring to the small scientific community in which they operate and in which there is a high degree of social control: “I have no reservations about sharing my ideas in that context. So far nothing negative has happened. It is so frowned upon in the community” (Com02, postdoc, m, 45). A biologist who does research in a small, very delimited research area explained that they know each other, “that people support each other; they already [share data]. That’s already the case in my field” (Bio07, prof, m, 129).

Open Associative Sharing—Oriented to “Institutional Imperatives” (Merton) and to Formal Regulations

This kind of sharing is preceded by the publication of the data, making it potentially accessible to everyone who formally asks for the data according to formal regulations. Publication occurs through presentations or posters at conferences and through

publication, which cover the final data. “To us, sharing data means going to scientific conferences and showing the data or writing publications” (Neuro03, postdoc, m, 53). For lectures, generally binding formal regulations do normally not exist. Some respondents reported that that setting often also lacks a commitment to the scientific community’s “institutional imperative” (Merton, 1973), which forbids appropriation of the data of others researchers. They pointed out that the open nature of the lecture makes for tricky scientific exchange because listeners can exploit the presented research data without attributing them in the same way as in written publication. “If I present stuff there that is unpublished and that can be directly replicated by others, then I’m damaging my own collaborators” (Bio07, prof, m, 33). The protection of one’s scientific achievement becomes relevant at this point. The respondents stated that they protect their contributions by refraining from delivering a lecture or presenting data in a way that allows them to be used without permission. Depending on whether the researchers think a listener will abide by the commitment to cite the presented data in the proper manner, they give access to that material when personally asked.

Consider now the sharing of data through publication in a scholarly paper. As noted by Merton (1973), publication means that the knowledge generated by researchers passes into the domain of general knowledge. He identified this transition as the institutional imperative of science and called it “communism” (p. 273). The aspect lies in the fact that scientists transform their newly generated knowledge into common property through publication. A linguist among the respondents vividly described this transformation:

When it is published, it is completely accessible to the public. It is then accessible to everyone. That’s why it has a different status for me. Outsiders do not get to see my data, which I have directly in my computer. (Ling03, postdoc, f, 73)

She stated that she did not perceive the published data as hers; she keeps the raw data to herself. A computer scientist similarly described the difference between the data that pass into a publication and the data that remain with him. Someone asking him for data receives “the version that corresponds to the publication, not all the data series generated in between, which we also keep, of course. [The latter] are internal data, not the results that have been published” (Com04, prof, m, 73). The respondents in my interviews viewed the sharing of the published data as a duty stemming from the binding character of a formally guaranteed framework of sharing. “We are obliged to make what we publish available to the world as well” (Bio07, prof, m, 93). The data become known to a greater public through lectures and publications if the data are made open. However, researchers share even the final data quite differently. They maintain peer-to-peer data-sharing of their presented data because they perceived inadequate commitment to “institutional imperatives,” whereas referring to formal regulations they make final data accessible to everyone.

Table 1 *Social Forms of Data-Sharing in Peer-to-Peer Relationships*

Categories	Closed communal sharing	Closed associatiive sharing	Open associatiive sharing
Meanings	Sense of belonging together ensuing from the protection of the working groups' scientific achievement	Shared instrumental orientations ensuing from an adjustment of interests and ensuring the scientific achievement of the cooperation partners	Institutional imperatives and formal regulations ensuing from belief or nonbelief in the authority of these orders
Legitimate order	Establishment of a high level of trust, location of power of disposal in the director	Creation of trust, elimination of competition, retention of control over the data	Acknowledgment of authorship and obligation to give proper source citation
Social relationships	Permanent social relationships	Cooperation for a limited time	All interested scientists
Data-processing phases	All phases of data-processing	Prepared and processed data	Final data

Discussion: The Sociality of Peer-to-Peer Data-Sharing

The three social forms of data-sharing that have been identified in this article clearly differ according to the social relations underlying them, the subjective meanings that are regarded as protecting the individual's own scientific contributions, and the accepted and practiced legitimate order (see Table 1).

The starting assumption of the analysis is that data-sharing is a social action embedded in a social relationship. The three forms of data-sharing make it evident that very different social relationships are involved in making data available to others. The assumption implies that the social relationship derives from an awareness of sharing something social with each other, which lays the foundation for sharing data in the first place.

In closed communal sharing the members of the research group share a sense of belonging together, which nurtures mutual trust. This form of data-sharing also includes recognition of the institutionalized legitimate order of the work in the research group, such as acceptance that the power of disposal over the data lies with the head of the group. Sociality is thus determined by the shared sense of belonging. This feeling constitutes the basis for sharing data.

Closed associative sharing is a social relationship whose rationale stems from a finite period of joint research. This phase can vary in length and scope. The decisive factor is that sharing be oriented to joint research as the purpose and that its sociality issues from this project. The sociality of the cooperation follows upon coordination of research interests, that is, upon the participants' recognition of each other's scientific achievements. The scientists thus enter into an agreement on data-sharing, although they can have quite different interests. The central point is that they have agreed to respect each other's interests, and this social consent constitutes the scientific basis for the sharing of the data.

In open associative sharing there is sociality guiding the scientists before they enter into a specific social relationship and ask someone for their data or make them available. The sociality is predicated by institutional imperatives and formal regulations because it is based on rules, which should be operative before researchers make their data accessible. The implication is that the legitimate order is not produced and cannot be controlled by the scientists who are sharing data. This relation between the legitimate order and data-sharing scientists could explain why many of the interviewees did not have confidence that the rule of acknowledging their scientific achievements would be applied to conference presentations and feared their data would be scooped. In open associative sharing, social obligations external to the social relationship of data-sharing constitute the basis for deciding which data are shared and how.

All three forms of data-sharing follow on the creation or existence of a certain kind of sociality. The specifics of this sociality determine why, how, and what data are shared. It is remarkable that the sociality of data-sharing in peer-to-peer relationships, like the sociality of other social relationships, is essentially characterized by general social phenomena such as trust and mistrust, the feeling of belonging together, agreements on social relationships. The three forms typically come to bear on different stages of the research process, so data from different phases of processing are shared. Although closed communal sharing encompasses all four phases of data-processing, it is particularly typical of the production of research data. Closed associative sharing takes place in the intermediate parts of the research process, whereas open associative sharing is not usually practiced until the research process has largely been completed. It is therefore important not to talk about data-sharing in general but rather to look at which data are shared in which phase of the research.

More is Shared Than is Apparent

Drawing on Max Weber (1922/1978) in this study, I have conceptualized data-sharing as a social action that takes place within a social relationship. The empirical material consisted of interviews with doctoral students, postdocs, and professors, totaling 34 researchers from five disciplines. I have characterized the different social relationships in which the respondents share or make available the data they generate. For this purpose I elaborate on the subjective meanings, which are essentially oriented to the protection of the researcher's own scientific input, and on the legitimate order, which explains and justifies sharing or not sharing. This analysis has identified three

social forms of data-sharing in peer-to-peer relationships: closed communal sharing, closed associative sharing, and open associative sharing. They rest on different kinds of sociality. The specifics of each set forth why, how, and which data are shared. The three forms typically come into play at different stages of the research process, so the data being shared or made available have undergone unlike degrees of processing. The three forms also differ in their subjective meanings, the nature of the social relationships they entail, and the data shared in each case, and the legitimate order with which the interviewees reason their data-sharing practices.

Overall, this study reveals that far more data-sharing is happening in scientific practice than seems to be the case if one works only with the concept of open data. In terms of the sociality that open data predicate, they represent a variant of data-sharing clearly different from the forms elaborated in this article. It is based on anonymous and abstract social relationships and is not the subject of this study. The interviewees only sporadically commented on this form. From their perspective open data represent a form of sharing entertainable only after completion of the publication process and only if final data are meant. “The most I could imagine is publishing the data set, that is, what you have in your statistical program at the end” (Ling03, postdoc, f, 131). Data would be uploaded only after publication—“for reasons of reputation and originality” (Com03, prof, m, 127). If the main goals of open-data policy programs are to encourage researchers to increase access to their data, intensify scientific cooperation, and improve data quality, it could be instructive to study the three forms of data-sharing in peer-to-peer relationships to improve understanding of why and how scientists make their data accessible to other researchers.

Like any other study, this one has limitations. First, the sample of interviews included only five disciplines; classic subjects in the humanities were completely absent. Most research in the disciplines covered in this study takes place in groups that practice closed communal data-sharing. Projects in the humanities are often conducted by individual scientists, so that form of data-sharing is probably less common. How does this likelihood affect the two other forms of data-sharing—closed and open associative sharing? In compiling the sample, I selected respondents who generate experimental data. It would therefore be interesting to see whether the results also apply to other kinds of data, such as those collected through surveys. Second, although my methodological approach was geared to ensuring as clear a link as possible between the interviewees and projects, the descriptions in the preceding pages are nevertheless not practices I actually observed. Lastly, there are limitations concerning the research context. Academia in the German science system has a distinct status hierarchy, a characteristic that may explain why some doctoral interviewees stated that they do not have the right to decide about data-sharing. With the DFG hardly regulating modes of data-sharing, this analysis has had the advantage of making it possible to work out how researchers informally share data. However, this framework, too, constitutes a limitation, for more the practices of data-sharing would presumably look different under more formal conditions.

Beyond these limitations there were two most remarkable findings. First, the respondents had a large repertoire of social rules about how they organize their sharing of data, particularly those from unpublished research. Second, the interviews suggest that data-sharing is a generally viable part of everyday research: the subjective

meanings attached to it are mutually respected, and the individual researcher's scientific achievements are protected.

Funding Acknowledgement This work was supported by the German Science Foundation [grant number: 347305329], and the Leibniz Center of Science and Society of the Leibniz Universität Hannover. Open Access funding enabled and organized by Projekt DEAL.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>

References

- Ankeny, Rachel A. 2017. Bringing data out of the shadows. *Science, Technology, & Human Values* 42(2): 306–310.
- Axelsson, Ann-Sofie, and Ralph Schroeder. 2009. Making it open and keeping it safe: e-enabled data-sharing in Sweden. *Acta Sociologica* 52(3): 213–226.
- Belk, Russell. 2007. Why not share rather than own? *Annals of the American Academy of Political and Social Science* 611: 126–140.
- Bezuidenhout, Louise, Sabina Leonelli, and Ann H. Kelly. 2017. Beyond the digital divide: Towards a situated approach to open data. *Science and Public Policy* 44(4): 464–475.
- Deutsche Forschungsgemeinschaft (DFG). 2021. *DFG form 54.01–04/21 Proposal Preparation Instructions: Project Proposals*. https://www.dfg.de/formulare/54_01/54_01_en.pdf.
- Dorta-González, Pablo, Sara M. González-Betancor, and María I. Dorta-González. 2021. To what extent is researchers' data-sharing motivated by formal mechanisms of recognition and credit? *Scientometrics* 126: 2209–2225.
- European Commission. 2016. *European Cloud Initiative - Building a competitive data and knowledge economy in Europe*. COM(2016) 178 final.
- Fecher, Benedikt, Sascha Friesike, Marcel Hebing, Stephanie Linek, and Armin Sauermann. 2015. Reputation economy: results from an empirical survey on academic data-sharing. *Discussion Paper des Deutschen Instituts für Wirtschaftsforschung*, no. 1454. www.diw.de/documents/publikationen/73/diw_01.c.497416.de/dp1454.pdf (8 July 2015).
- Hey, Tony, Stewart Tansley, and Kristin Tole. 2009. *The fourth paradigm: Data-intensive scientific discovery*. http://research.microsoft.com/en-us/collaboration/fourthparadigm/4th_paradigm_book_complete_lr.pdf (7 July 2016).
- Kansa, Eric C. 2014. The need to humanize open science. In *Issues in open research data*, ed. Samuel A. Moore, 31–76. London: Ubiquity Press.
- Leonelli, Sabina, and Nicolò Tempini, eds. 2020. *Data journeys in the sciences*. Cham: Springer.
- Levin, Nadine, and Sabina Leonelli. 2017. How does one “open” science? Questions of value in biological research. *Science, Technology, & Human Values* 42(2): 280–305.
- Linek, Stephanie B., Benedikt Fecher, Sascha Friesike, and Marcel Hebing. 2017. Data-sharing as social dilemma: Influence of the researcher's personality. *PLoS ONE* 12 (8): 1–21.
- Maienschein, Jane, John N. Parker, Manfred Laubichler, and Edward J. Hackett. 2019. Data management and data-sharing in science and technology studies. *Science Technology & Human Values* 44 (1): 143–160.
- Mayering, Philipp. 2008. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Merton, Robert K. 1973. *The sociology of science: Theoretical and empirical investigations*. Chicago: University of Chicago Press. (Original work published 1942).

- National Science Foundation. 2020. *Dissemination and sharing of research results*. <https://www.nsf.gov/bfa/dias/policy/dmp.jsp> (8 April 2020).
- Stamm, Isabell. 2018. Organized communities as a hybrid form of data-sharing: Experiences from the Global STEP Project. *FQS* 19(1), Article 16, <https://www.qualitative-research.net/index.php/fqs/article/view/2885> (10 May 2022).
- Tedersoo, Leho, et al. 2021. Data-sharing practices and data availability upon request differ across scientific disciplines. *Scientific Data* 8(192): 1–11.
- Tenopir, Carol, et al. 2011. Data-sharing by scientists: Practices and perceptions. *PLoS ONE* 6(6): 1–21.
- Tenopir, Carol, et al. 2020. Data-sharing, management, use, and reuse: Practices and perceptions of scientists worldwide. *PLoS ONE* 15(3): 1–26.
- Thoegersen, Jennifer L., and Pia Borlund. 2021. Researcher attitudes toward data-sharing in public data repositories: A meta-evaluation of studies on researcher data-sharing. *Journal of Documentation* 0022–0418: 1–17.
- Velden, Theresa. 2013. Explaining field differences in openness and sharing in scientific communities. *Proceedings of the 2013 conference on computer-supported cooperative work*, 445–458. San Antonio: ACM Press.
- Weber, Max. 1978. *Economy and society: An outline of interpretive sociology*. 2 vols. (eds. Guenter Roth, et al., Trans.) Berkeley: University of California Press. (Original work published 1922).
- Wessels, Bridgette, et al. 2017. Mobilising data: Scientific disciplines, scientific practice and making research data open. In *Open data and the knowledge society*, eds. Bridgette Wessels, et al., 103–120. Amsterdam: Amsterdam University Press.
- Whyte, Angus, and Graham Pryor. 2011. Open science in practice: Researcher perspectives and participation. *The International Journal of Digital Curation* 1(6): 199–213.
- Yoon, Ayoung, and Youngseok Kim. 2020. The role of data-reuse experience in biological scientist's data-sharing: An empirical analysis. *The Electronic Library* 38(1): 186–208.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.