

Strategies to Overcome the Fear of Scooping Among Researchers in the Era of Open Science

Elina Korhonen

Doctoral Programme in Social Sciences, Faculty of Social Sciences, University of Turku, Turku, Finland Biography

Heidi Laine is a doctoral student in social science history at the University of Helsinki Doctoral School on Humanities and Social Science Doctoral Programme on Political, Social and Regional Change. Her dissertation, which this article is a part of, studies the development of self-regulation in research integrity, using the over two decades old Finnish regulatory framework as a point of departure and comparison. Laine is an advocate of open science and a member of the Open Knowledge Finland Core Team and the secretary of the Finnish Committee for Research Data.

Strategies to Overcome the Fear of Scooping Among Researchers in the Era of Open Science

Abstract

The risk of scooping is often used as a counter argument for open science, especially open data. In this case study I have examined openness strategies, practices and attitudes in two open collaboration research projects created by Finnish researchers, in order to understand what made them resistant to the fear of scooping. The radically open approach of the projects includes open-by-default funding proposals, co-authorship and community membership. Primary sources used are interviews of the projects founding members. The analysis indicates that openness requires trust in close peers, but not necessarily in research community or society at large. Focusing on intrinsic goals, such as new knowledge and bringing about ethical reform, instead of publications, supports openness. Understanding fundamentals of science, such as philosophy and science and research ethics can also have a beneficial effect. Whether there are aspects in open sharing that makes it seem riskier from the point of view of certain demographical groups, such as women, could be worth more studying.

1. Fear of Scooping as a Disincentive to Sharing

'There is always this fear, that someone steals your ideas, or doing the same thing at the same time, and some people fear it more than other people, I think especially younger people, also some older. I think this causes a lot of stress to the scientists, and it has

happened to me. [...] you try not to think about it, you still think that what if someone else is doing the same thing and this is useless work, so then it takes your energy.' (From the interview of research participant NMRLP A)

Scoping is a research community slang term for having someone else claim priority, usually through publishing, to a research idea or result you yourself have been working on. This is considered a problem, because according to a widespread belief, academic journals value novelty and are reluctant to publish results that don't have a high enough novelty factor (Chambers 2016). Journals are the veins through which academic merit flows, so not making it to the high-impact journal of one's dreams could mean losing competitive edge in the race for tenure or funding. Scooping in science is considered to be commonplace, practically an occupational hazard (Rodrigues 1998), even though there is little evidence on its volume or the supposed resulting adverse career effects (Schwarzkopf 2016). The fear persists nonetheless and could even be gaining strength because of another widespread fear, coined in the phrase 'publish or perish'. (Dhand 2002)

In addition to being a source of stress, fear of scooping is seen as a barrier to open science, especially open data (Bishop 2015). A study commissioned by the Knowledge Exchange network reviewed incentives and disincentives for data sharing. Fear of competition, of being scooped and the resulting reduced publication opportunities were at the top of the list. According to the study, these fears affected especially early career researchers, but senior researchers were not entirely immune either. (Van den Eynden & Bishop 2014).

Those wishing to make open data more mainstream should subject these expressions of concern under careful analysis, as they make visible the logic behind fears, thus holding the key to cultural change, as long as these fears are efficiently addressed. In order to achieve that, further empirical research is needed on the incentives and disincentives to sharing.

In the case of fear of scooping this is not an easy task, since the phenomenon combination of two sensitive subjects: misconduct and mistrust. Sometimes it is easier to study and understand a thing through its absence: why someone is not afraid of scooping, or shares despite the fear?

I have interviewed key researchers from two radically open research collaboration projects for my PhD, which deals among other things with open research collaboration and responsible conduct of research. The projects in question have produced almost all of their content openly online in real-time, welcoming outside participation. One of the projects even allocated funding to an online human 'swarm'. The idea for this paper emerged from the existing materials on these cases, as I realized that they could offer insight on the phenomena of fear of scooping in the context of open science, as well as clues to dismantling this particular barrier to sharing.

2. Research question

The aim of this paper is to showcase two pioneering and, especially in relation to scooping, fearless experiments in open science and open collaboration. A second goal is to create understanding about the factors that make sharing possible in the face of a supposed threat of scooping, thus being able to offer not even a recipe, but a partial list of possible ingredients for fostering similar endeavors.

The research task can be broken down to two main levels of inquiry.

1. On a descriptive level the goal is to explain how the open research projects coped with the risk of scooping related misconduct. The questions on the descriptive level deals with the measures and strategies (adopted consciously or unconsciously) used by the projects to tackle research misconduct related problems; what kind of rules were in place; what kind of practices used, how the outputs were made public, and so forth.
2. On an interpretive level the focus is in locating and analyzing factors that allowed the researchers of these projects to ignore or come into terms with the assumed risk of falling victim to scooping. What were the projects' researchers' attitudes and motivations when going in to these projects? Which factors can be traced back to the individual level, which ones have their roots on the social level?

2.1 Social Media for Citizens and Public Sector Collaboration (Somus) Project

The Social Media for Citizens and Public Sector Collaboration (Somus) project ran for two years in 2009 and 2010. The project concept and plan was created by an open membership online community called the Open Research Swarm (ORSi), that emerged on a Finnish microblogging platform Jaiku during 2007 and died down before Somus ended.

The stated aim of the project was to study the dynamics of information, knowledge and citizenship in an open and participative media environment. Both the planning and operating of the project were conducted openly and participatively, in line with the object.

Somus received some 500 0000 euros in funding through the Academy of Finland MOTIVE program funding call, plus in kind support from institutional consortium members University of Jyväskylä, University of Tampere, University of Technology (later part of Aalto University), and the VTT Technical Research Centre of Finland.

ORSi was named as a consortium member and allocated 15 000 euros in funding. Four subprojects were outsourced through ORSi with these funds. ORSi also had a representative in the projects directing group. At the height of its activity in 2008 the ORSi channel had 110 members, of which approximately 25% were women.

2.2 NMR Lipids (NMRLP) Project

The NMR Lipids Project (NMRLP) is at the time of writing ongoing. It was founded in 2013 by two Finnish nationality post-doctoral researchers, one working at the Aalto University School of Science and the other at Freie Universität Berlin.

The aim of the project is to study the atomistic resolution structures of lipid bilayers and to develop an all-atom force field that will match these data. The project is progressed through the comments in a blog and using a GitHub repository hosting service organization. Everyone is free to join and this includes the co-authorship of articles. Anyone who has written comments, or named in the comments, is automatically added to the list of contributors, which at the time of writing included 30 names. Only one of the contributors is female.

At the time of the interviews (December 2015) the project had not received any direct funding despite applications, and was operating more or less on a voluntary basis.

2.2 Scooping in the context of open science

Being scooped means being left second in an attempt to gain priority over a research idea or discovery. This is a fairly common occurrence and in the majority of the cases no research misconduct is involved. Fields of research often have trendy subjects and 'grand challenges', which can result in people and groups working on the same subject without being aware of each other. (Le 2013)

When considering the fear of scooping in the context of open science the discussion turns to illegitimate scooping; taking someone's work and presenting it as your own. Research misconduct is often divided into three categories: fabrication, falsification and plagiarism (f. e. Steneck 2007). The Finnish definition names also misappropriation as a separate category and defines the two as follows (TENK 2013):

- Plagiarism or unacknowledged borrowing, refers to representing another person's material as one's own without appropriate references. Plagiarism includes direct copying as well as adapted copying.
- Misappropriation refers to the unlawful presentation of another person's result, idea, plan, observation or data as one's own research.

Depending on the case and the regulative framework, scooping is either defined as plagiarism or misappropriation. Misappropriation can be considered as the more serious offence. Plagiarizing affects usually small textual portions of a researcher's work and often targets published texts, so when found out, the victimized researcher can easily prove priority. Research ideas don't usually get published, but can be communicated over coffee breaks, conference cocktail hours or meetings with supervisors. Misappropriation of a research idea can potentially undermine years' worth of work if a line of research gets abandoned due to scooping.

3. Research methods and sources

The preliminary sources of this study are the interviews of four researchers; two from each of the research projects. The interviewees were identified based on their founding role. One of the SOMUS participants is female, all the others are male.

The two research projects were chosen because of their 1) open-by-default strategy towards outputs and collaboration and their 2) affiliation with the Finnish research community, as this study is part of a PhD project using the Finnish RCR mechanism as a point of departure and reference. As far as I am aware, the projects are unique in the above mentioned respects.

The interviews were conducted during December 2015 (for NMRLP) and June and August 2016 (for SOMUS). The participants were first asked to give written answers to a series of background questions about their project. The following oral interviews were conducted face to face with two participants and over a conference call with the two other. Each of the semi-structured interviews lasted from one to two hours. When transcribed, the length of the interviews varied from ca. 7000 to 11 000 words.

Interview questions were designed for an ongoing doctoral research project that studies the mechanisms of research integrity self-regulation in Finland. The main topic of the interviews was responsible conduct of research (RCR) and the ways open research collaboration projects enact and reproduce RCR. The participants were questioned on their values, their knowledge on RCR issues and the practices and nature of the. The concept of this current paper rose from the finished interviews. No complementary questions were asked afterwards for the purpose of this paper, but the participants got a chance to check and comment the interviews before the paper was submitted. No significant new information concerning scooping was disclosed during commenting.

The interviews were coded using qualitative data analysis software. Due to time management issues the material was not re-coded by a third party, as is sometimes customary for added reliability of the results (for example Grand 2016). The research participants have given their consent to depositing the non-anonymized interview records on the Zenodo repository after the PhD project has been finished. They will thus be available upon request for further research and reliability checking.

Even though the research participants are easily identifiable from the open records of the research projects, I will be referring to them in this paper by NMRLP A and B and SOMUS A and B. This is done in order to not draw unnecessary attention to the persons beyond their professional roles and avoid for example individual quotes ending up as results in search-engine searches. This research plan of the PhD project has undergone ethical evaluation by the University of Helsinki Ethical Review Board in the Humanities and Social and Behavioral Sciences. (Laine 2015) The board deemed the plan ethically acceptable, but in its statement raised the question whether 'un-sensitive born' content could be seen as sensitive in some unforeseen online context and recommending careful deliberation on the level of openness.

Other documentation about the two projects has been utilized as secondary sources; for establishing project timelines, checking key dates and cross-checking things discussed during interviews. For NMRLP the main secondary source is the blog nmrlipids.blogspot.fi and for Somus the projects final report (Näkki et al. 2010).

3.1 Theoretical framework

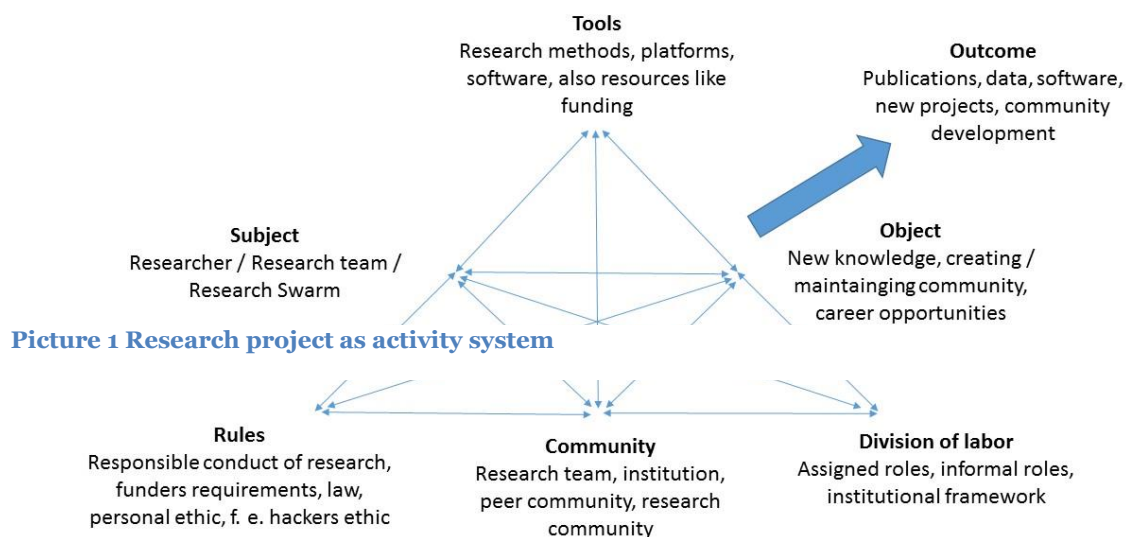
Case study is not a research method in itself. It is a research approach and design suitable for settings where the focus is on understanding and depth of analysis, rather than creating theory or universally applicable laws. This doesn't mean that studies based on individual cases give no room for generalizations. A case study cannot be a representative sample. Instead it can be a 'good fit', in the sense that it matches other situations in which one is interested in. In this case that means other research projects. (Schofield 2009)

The cases in this research are exceptional, but there are still many meaningful points of comparison with more traditional research projects. According to Janet Schofield, case study is a useful method for studying bubbling under trends ('what may be') and ideal scenarios ('what could be'). What's required to make these leading or cutting edge cases 'fitting' in terms of 'what is', is a thick description. (Schofield 2009) Giving a thick description means not just describing human action, but also the context. This is something that today's historians understand especially well (Kalela 2012). According to Thomas D. Cook a 'descriptive theory of science' is only comprehensive, or thick enough, if it does the following; incorporates

attributes of both individuals and the contexts of operation; deals with many contexts on many levels; presumes constant change on all levels as fact; and assumes that individuals influence institutions and are influenced by them. (Cook 1993)

I have used different the levels of action and influence, as suggested by Cook, together with the cultural-historical activity theory (CHAT) activity system model as analytical tools. The NMRLP and Somus researcher interviews have been coded with using the elements of the activity system.

CHAT offers a framework for studying both the social context and structures as well as the individual or social agency. It makes visible interdependencies and tensions between acting subject, mediating artifacts, instruments and elements of the social framework. CHAT recognizes that an activity system, which is a series of actions directed at a certain object, in this case a research project, is in a state of constant motion. It also acknowledges links to other activity systems as important to understanding one system. The system doesn't exist in a vacuum, but is in constant interaction with other activity systems, in this case other research projects, institutions and communities (Engeström 2005).



Picture 1 Research project as activity system

An activity system can be visualized as a triangle, as seen in picture 1. The subject, the acting agent, can be an individual or a group. Object is what the activity is directed towards. With the help of mediating physical and symbolic, external and internal instruments, the object is transformed into outcomes. The community are all the people sharing the same general object, while division of labor refers both to the actual tasks and the division of power and responsibility. Rules are the explicit and implicit regulations, norms and conventions that constrain and direct interactions within the activity system. (CRADLE 2016)

The relationships between elements of the activity system are dynamic and their content fluid and situational. The activity system is not harmonious, as there is often tensions between artefacts. Activity theory offers a framework for understanding problems and challenges through tensions between elements of the activity system. For example, the fear of scooping could be located as a tension between the subject and the community. I argue that neither

the subject nor the community is the original source of tension. Explaining fear of scooping solely by suspicious researchers and untrustworthy or unsupportive communities feels simplistic. Instead, I see tension rising from a discord between the object and other elements of the system, if the object is defined as publication or a career promotion instead of a more intrinsic object, such as discovery or deeper understanding.

4. Case projects' openness practices and strategies

Previous case studies have shown that researchers prefer to view openness as something existing on a continuum, rather than a choice between two opposites; open or closed. (Whyte & Pryor 2011; Grand et al. 2016) This sentiment is also reflected in the answers of Somus and NMRLP researchers.

'I think we used open research [...] as a concept, but I think we didn't fit to the definition from all the aspects that... A lot of the project work we also did very traditionally in the smaller teams, software development and so on, and even if it was openly on the Internet doesn't mean that everyone, if we had our meeting memos online it may be that only the project team participated in editing it, but no one else. So it was also a little bit artificial sometimes that everything must be openly on the internet even if no one reads.' (Somus A)

Whyte and Pryor present in their paper a conceptual tool for typifying and evaluating the openness of a research endeavor consisting of two dimensions; 1. the stage of in the research process where sharing occurs, and 2. the level of engaging the actors involved. For the first dimension they have created a framework of seven stages of the research cycle. The second dimension can be evaluated using a continuum of six degrees of openness, which from the least to most open are; private management; collaborative sharing; peer exchange; transparent governance; community sharing; and public sharing. (Whyte & Pryor 2011)

Table 1 shows how Somus and NMRLP fare on this conceptual ladder. The decisions between degrees of openness are not exact or definite and reflect my subjective weighing of intentions and outcomes. In the following I will highlight some of the actions and choices behind the categorizations.

Research cycle stage	Somus	NMRLP
	Outputs	
Conceptualizing, networking	<i>Public sharing</i> on Jaiku sevice.	<i>Community sharing</i> on ArXiv and direct e-mail.
Proposal writing and design	<i>Public sharing</i> on Jaiku and wiki.	Initial setup <i>private management</i> or <i>collaborative sharing</i> , after that <i>public sharing</i> on Blogspot and GitHub
Collecting and analyzing	<i>Community sharing</i> : research data shared among research group	<i>Public sharing</i> on Blogspot and GitHub.

Documenting and sharing	<i>Public sharing</i> on Etherpad and wiki.	<i>Public sharing</i> on the blog.
Publishing and reporting	<i>Community sharing</i> through standard toll-access articles, public sharing of the final report.	<i>Community sharing</i> : drafts of articles on GitHub, preprints on ArXiv, finished product in traditional journals. Subject matter unintelligible to most laypersons.
Engaging and translating	<i>Public sharing</i> through the social media applications created for case studies	<i>Public sharing</i> through blog posts and public appearances to promote open science
Infrastructuring	<i>Private management</i> : social media applications co-produced with citizen participants, but all rights remain with consortia institutions, also one proprietary application funded	<i>Public sharing</i> : data and code shared and licensed openly (GitHub), although not understandable for laypersons.

Table 1 Degree of openness at different research cycle stages

Perhaps the most radical action concerning openness and the possible threat of scooping was the Open Research Swarms (ORSi) decision to draft openly online the funding application that resulted in the establishment of the Somus project.

'[...] even in Norway they were very confused in the research community: "What are you doing, someone can steal your idea!" But I have always, also in hacker time, thought so that the idea is not a big deal. For good, very good or excellent idea, you actually need thousand ideas. The main point is how to operate and technically do that thing actually around this idea.' (Somus B)

Instead of having their ideas stolen, the ORSi proposal received contributions from their competition.

'SOMUS was very exceptional there that we also but the project plans openly. It was very funny to see that also people who had their own applications for the same call were contributing in our plan, [...].' (Somus A)

The competitors found the openness of the ORSi proposal helpful, but not for copying or modelling the content. Instead it helped them in focusing their own applications.

'It became clear through discussions with other applicants, however, that the open draft actually helped them focus their projects on different research themes to avoid direct competition with Somus. When a text is documented in a wiki, it is easy to find an author and time stamp for the text, making it uncomplicated to solve authorship questions' (Kronkvist 2011)

The open project planning was justified in hindsight in the final Somus report by arguing that early openness helps to prove priority. The interviews reflect a much less calculated attitude towards open planning and ownership of ideas.

'[...] for me it was like, well if someone would find out plans for new project, who cares.' (Somus A)

'I started for example microblog in Jaiku about ideas and concepts I may research. [...] That's the main point of sharing ideas, that someone recycles them.' (Somus B)

The NMRLP researchers saw openness as an antidote to scooping and especially the fear of scooping. This reasoning is interesting, because it is the exact opposite to the fear of scooping as a barrier to sharing line of reasoning.

'Q: So you were in a way protecting your own work by publishing it openly?

A: Yes, but the most important thing is that I wanted to get rid of this [stress]. Even though you try not to think about it, you still think that what if someone else is doing the same thing and this is useless work, so then it takes your energy.' (NMRLP A)

Both of the projects experiment with open collaborative writing. Publishing is at the core of accumulating academic merit and authorship related disputes are among the most common reasons for blowing the whistle on research misconduct in Finland (Niemelä 2015). The NMRLP researchers were aware that article authorship is both a source of motivation and potential conflict, so they define clear rules for authorship.

'But this authorship thing was [important], because that had to be clear from the beginning, because you can't change it. We thought that this has to be done as simple as possible, in a way that we don't have to change it, we don't have to think, we don't have to do any decisions.' (NMRLP A)

The NMRLP rules determine that all of the authors are equal and whether a contribution is sufficient for a co-authorship is left to the contributor to decide. The rules are much laxer than for example the often cited ICMJE guidelines (ICMJE 2016), allowing at least in theory something akin to honorary or gift authorship. Due to the transparency of contributions and subtle pressure from the community, this risk has not actualized.

'Also, [from] other projects I know, that this part, when you start putting names on the paper, it's the most annoying part and that can lead to fights. So it's best to define it clearly from the beginning. [...] it looked for a while that there will be people who actually, for the first paper, would have the names there who didn't actually contribute, but when these people read the article they decided they don't want to have their name there, 'cause they don't have a contribution. [...] maybe in this issue of the first paper there was a very mildly implicitly pointing out, that you didn't actually participate with the first paper.' (NMRLP B)

In the collaboratively written Somus paper (Heiskanen et al. 2008) one of the co-authors came on board at the very last minute for language revision and was granted co-authorship.

As neither project saw scooping as a risk, no clearly identifiable measures against it were taken. Using an open attribution license and identifying clearly the conditions under which open content can be used could be seen as a preventive measure against accidental scooping. The NMRLP content on GitHub is licensed using the GNU General Public License according to the researchers. The NMRLP blog is not licensed, nor were any of the Somus outputs licensed.

Overall I would position NMRLP in the sixth, most open degree in terms of Whyte's and Pryor's typology (2011); all of the projects resources are technically available for access by any member of the public, the data and metadata is designed to be reused by a designated research community without an embargo with an open attribution license. There is, however, one aspect where the project falls short of the broadest possible openness, as there is not much effort made to make the data and results understandable for a lay audience. However, the interviewed researchers have been vocal about openness in various science and research policy discussions and promoted openness as an efficient research method both in terms of scientific excellence and research integrity.

Somus resides somewhere between the fifth and sixth degree. Some aspects of the project exceed the openness of NMRLP, as the projects outputs translate also to non-experts. A major short-coming is the lack of long-term preservation and accessibility to the various outputs. This is partly due to bad luck, with the seized development of the Jaiku platform and a memory stick containing archives on their way to a new storage location being somehow wiped empty, partly to insufficient planning.

5. Researchers' motivations

Motivation is a key concept when talking about human action, especially if the aim is to influence that action, as is the case with the open science movement. Motivation is the process of energizing, directing and sustaining activity towards a goal. It is affected among other things by outcome expectations, values and social comparisons. (Schunk 2012) Motivations as well as their hierarchical relations can change in the course of the activity (Tikhomirov 1999), so when explaining past motives, the narratives people construct tell not just about the past, but the present as well.

In the activity system model motivation is part of the object. Collective activity is connected to object and motive, of which the individual subjects are often not consciously aware. Individual action is connected to a more or less conscious goal. (CRADLE 2016) In the cases of Somus and NMRLP differentiating the individual motives from the collective ones was challenging, as all of the four researchers were very aware of the societal context of their actions and strongly motivated by being able to bring about a change and reform in that context. The clearly individual goals that were recognizable are:

- career goals: being able to financially continue work as a professional researcher,
- personal wellbeing goals: relieving a stress caused by the fear of scooping (see section four on openness practices and strategies),
- curiosity driven goals: doing something new and exciting, and
- social goals: helping out a friend and being a member of a community.

One reason why fear of scooping might be so widespread and resistant to lack of evidence, is that it hits some of the core motivations of researchers. Studies suggest that more than monetary gain, researchers are motivated by intellectual stimulation and desire for original contribution and novelty. (Shadish et al. 1994)

'[...] then obviously if you are the first one doing something like that, it will benefit your career, because then you will get associated with that. This is the guy who does openly, which is of course a good thing.' (NMRLP B)

'I think the feeling of we are doing something totally new, that it's cool to be in this kind of group that is changing everything, as we thought.' (Somus A)

A major collective and individual level motivation that the interviewees shared was addressing ethical and integrity related problems in the research community, despite the lack of formal RCR training and low level of familiarity with RCR guidelines. Instead they all professed a strong intrinsic understanding of RCR issues and had all proactively sought education on philosophy of science.

'So I was thinking that I would open the possibility to people who have done mistakes or produced results which are wrong. I would give them the opportunity to participate in fixing them, because in science the problem is not that you make things wrong, especially if you fix them yourself.' (NMRLP A)

'I had been interested in the idea of open science, especially frustrated with the way that science has been published since the late, [...].' (NMRLP B)

'[...] [ORSi members] want to change the world and change the way research is done. New practices for everything, people who are criticizing the common way of doing research, that it doesn't anymore fit the social media era.' (Somus A)

Despite the ethically driven goals, openness was mostly seen as a means to an end, rather than an object on its own. For Somus openness and its potential for added value was an object of study. In the NMRLP's case the open method had primarily instrumental value.

'I had these results that showed problems in the field. I saw that this will affect the conclusions. I saw that there are results which are wrong, but people think that they are right. There is a research line which is wrong. [...] so I thought a way which would be the fastest and most effective way to fix this [...]' (NMRLP A)

What is interesting in terms of the threat of scooping, is that none of the four researchers directly or indirectly named scientific novelty, break-throughs and publications as an object of activity. NMRLP used publications as a way of making participating attractive and rewarding, but they didn't believe in co-authorship as a primary motive. They thought that colleagues would participate out of altruistic (helping a colleague), utilitarian (advancing knowledge) and/or moral (doing the 'right thing') reasons. Publications were needed as a form of compensation to the employing institutions and funders, which had a claim on the researchers' time and efforts.

'People need those because of funding reasons and that was one of the ideas to attract people: how do you get people to put their time to it, they need articles because of this thing. That's the way things work nowadays. [...] I think the main motivation is that they feel that this is the right thing to do, but they can rationalize it for themselves and also their boss and institute by saying that I get this paper.' (NMRLP A)

The case study results show that open sharing as a concept and a practice doesn't necessarily challenge the idea of ownership and credit and researchers who engage in sharing can be very cognisant of other's contributions. NMRLP rules of authorship state for example that

'[...] it is the responsibility of the people communicating via this blog to make sure that in case they write something in which other people have had a scientific contribution (such as supervisors writing comments based on discussions with their students), the identities of all contributors should be made known. These secondary contributors will also be offered coauthorship in the manuscript.' (NMRLP XX)

In similar fashion, participant Somus B who described his values as 'hacker ethic' and didn't think ideas were a 'big deal' and not worth hiding, recognised misappropriating a research idea as questionable behaviour:

'[...] in normal daily life in universities it is not, I can't say common, but not rare, that professors and high standing older researchers are copying some ideas or concepts to study from the graduate papers. But it is also a bit gray area, because the ideas are free. But maybe professors and other older researchers should talk a little bit more with graduates, like "wow, you have a great idea, maybe we can make a project of it together", or something like that.'

6. Research culture and community

In this chapter I will examine the cultural and social contexts of the case projects. Three main levels of social identification, interaction and engagement are recognizable, in line with the levels proposed by Cook (1994). The community working on the case project, meaning community in the activity system sense, is the primary social level (see picture 1). On this level the existence of trust between individuals is raised as an important condition for collaboration.

'[...] because I'm a Finn, we have that some kind of trust in people, personality stuff. In Finland if people agree on something, maybe with a handshake or not, that is considered done, meaning that in most cases when Finnish people agree something they can suppose that the other people will make those things we have agreed.'

(Somus B)

'In Finland this lipids simulation community is very small, so effectively everyone who is participating from Finland I know. [...] I think it's also important in the beginning that those people who participated from Finland knew us, so that they trusted us, that we will do the work well and reliably.' (NMRLP B)

The participants were also asked directly whether they trust their colleagues. The Somus researchers, who mostly operated in the national level and in the fields of social science and engineering, described their communities as supportive and signaled a high level of trust:

'Actually I've never thought that about that someone would lie. Basically I think that all researchers speak the truth.' (Somus A)

NMRLP researchers were also mainly positive, especially concerning close colleagues and other early career researchers. However, they expressed a concern that the integrity of more

senior researchers in their field was sometimes compromised due to financial obligations alienating them from science:

'The origin of the things is that these people who have the most power, who are deciding things, who are deciding about the money and who are the bosses, they use most of their energy to get money and funding and promoting and stuff, they are not doing the science themselves and they don't really understand it.' (NMRLP A)

Participant Somus B also reflected on the field specific differences concerning trust and integrity:

'I personally think that in nature and mathematics and engineering, physics, chemistry, biology area the funding is much bigger and the competition globally is much bigger and there it might be more tempting or alluring to not play right, than in the social sciences, sociology and media and communication and that kind of thing. This is just my personal feeling, not a fact.' (Somus B)

The core communities were described as mainly heterogeneous both demographically and in terms of values. Somus A described the group as young, enthusiastic and idealistic. Somus B described them as early adapters and compared the atmosphere to Silicon Valley and the start-up business culture. NMRLP B described the group as mostly young, but with also a few more senior participants. He mentioned being especially pleased by the encouragement received from many older colleagues who were close to retiring or already retired and therefore didn't want to get involved, but saw the project as something very positive.

Both projects were male dominated in terms of numbers: approximately 25% of ORSi members were women, while the Somus research group was more balanced. For NMRLP only one of the 30 contributors is a female. This was despite the communities being at least formally open to anyone. For example the ORSi changed its original name 'Researcher Swarm' into 'Research Swarm' in order to not create an image of restricted membership.

'I mean there's not many in our field, but I would say that the number of people participating is even less than in general. But this might be just your random fluctuation, cause the sample is one. But I have talked about this with more women than who participate. [...] When I talked about this, especially in the early stages, I think that these few women were not interested because in principle they see it as cool, but too risky. It somehow seems like not a good idea, that it's too risky. [...] But again the sample size is two, so it's not really... I mean it's a general problem in the field, its maybe that 10% are female.' (NMRLP B)

The community of peers is the secondary level, more specifically the field of research and especially other researchers in the same career stage. In the case of NMRLP, the fields named were biophysical chemistry (NMRLP A), computational molecular biophysics (NMRLP B), user centered design (Somus A) and digital sociology (Somus B).

The attention the projects received from their peer communities was either encouraging and/or mildly wary in terms of the risk some colleagues saw the participants were subjecting themselves to, but discouraging or otherwise negative comments were not reported by the interviewees.

From the point of view of fear of scooping, just as important and informative as social context is the data context, in which the case projects exist. Data and practices around it are becoming ever more central to science, insofar that they provide an important window to research; tell me what is your data, and I'll tell you your field, even research question. Or to be more precise: tell me when is your data. The latter question comes from Christine Borgman, who has developed the concept of data scholarship to frame the set of complex relationships between research and data. According to her, data is 'an extremely useful environment in which to view radical shifts in scholarly practice and how they intersect with technology, education, and policy'. (Borgman 2015)

In her book 'Big data, little data, no data' Borgman presents data scholarship case studies. Some of the challenges raised in the social science case study apply also to the case of Somus: confidentiality concerns limiting release of data, datasets remaining with the researchers and the objects of study and the tools to study evolving too quickly for effective sharing. (ibid) In a way all of this shields from scooping, but is also in the way of open and effective reuse.

'Collaboration tools, editing tools we tested numerous or innumerable [...] Actually we changed tools all the time during 2 years and when we think there would be better, we changed. [...] We thought we would have time to copy and paste everything and make backups. That's one reason why we do not have so much documentation. For example, I have much more primary sources on my laptop and backups, [then there are] on the whole internet about Somus.' (Somus B)

NMRLP represents a more mature data culture, largely due to the less complicated nature of the data itself. Since the data is computational producing it doesn't require expensive infrastructure or extensive human resources. The relatively low human and monetary price tag on data can be seen as something that lowers the threshold on sharing and as the data is less scarce it could also make illegitimate scooping less appealing.

Third level is the wider research community, research funders and also society at large. Somus researchers didn't mention which benchmarks they had, but in terms of culture and ethos made references to the open source movement, start-up businesses and hackers. NMRLP interviews bring up Linux and the Polymath Project as models. Interestingly, even though Finland is a small country and often jokingly referred to as a 'club', the two projects researchers have not been aware of each other prior to this case study. This lack of awareness is a testament to the importance of communication as an aspect of openness. It also tells us of the importance of the global field specific community as a primary point of identification instead of the national research community and to the prevalence of the divide between natural and human sciences.

All of the participants saw science as vital for positive societal development. As mentioned before, the interviewed researchers were aware and even worried about societal and science policy developments in Finland and worldwide. They didn't draw a line between science and society, but generally saw them as parts of a whole.

'That people are ignoring facts to easier [get] the publication in high impact journals. This is not possible, this is something that caused a lot of stress to me when I realized that people are really doing this and it's not a marginal phenomenon, and it is

really affecting the universities and the scientific community and the society. I think it's really affecting the society also.' (NMRLP A)

Finland is considered a research and science friendly society. The Finnish society has embraced digitalization and openness both in innovation and governance (Tieteen tiedotus 2013). The governmentally funded Open Science and Research Initiative (2015-2017) aims at mainstreaming open science in Finland (OKM 2015). It can be argued that on the level of attitudes the atmosphere in the society and research community alike has been in the worst case neutral and in the best case favorable for open research experiments like Somus and NMRLP. At the same time the fact that the two cases are the only ones of their kind gives an indication that a vaguely favorable atmosphere without significant supporting structures or mechanisms only helps so far.

Even though none of the participants saw open sharing as a career risk, at the time of the interviews only one of them had a permanent salaried position as a researcher. All expressed worries about the direction of science policy and research funding in Finland. The NMRLP researchers had reason to worry about the continuity of funding also on a personal level. Finland was seen as no safe haven for open research or article based metrics.

'[...] it applies especially in Finland. Also abroad, but especially in Finland. Currently it seems that this is a strong trend. It's even in the very high governmental level, they say it almost directly. They use Finnish word "vaikuttavuus", but I don't know what they mean with it. If you translate it in English it's impact, then it's very [sure], even though I don't think they mean that, but it's very [sure] this high impact, it's the same word.' (NMRLP A)

7. Conclusions

Incentives and disincentives are external, motivation internal. Only after we understand what motivates researchers we can construct incentives to match. Similarly, only when we understand the risks researchers perceive, we can recognize disincentives to openness and remove them.

Getting scooped, having your research idea or results published by someone else, is a common fear among researchers. It can be a major stress factor and an energy drain. The risk of scooping is often used as a counter argument for open science, especially open data. All scooping isn't illegitimate. Most often scooping is accidental and becomes misconduct only when idea or content used was taken from another researcher, without giving them credit.

In this case study I have examined the openness strategies and practices of two open collaboration research projects created by Finnish researchers in order to understand what made them resistant to the fear of scooping.

The Somus project was created by an open online collective called the Open Research Swarm (ORSi). It operated on a Finnish microblogging service Jaiku, which has since ceased to exist. Somus combined methods from engineering sciences and social sciences, co-creating and testing applications and social media platforms with citizen stakeholders.

NMR Lipids (NMRLP) project is an ongoing open scientific collaboration project in the field of computational biophysics, aiming at understanding the atomistic resolution structures of lipid bilayers. The discussions happen on a blog, while manuscripts and data are developed in an open repository.

The primary sources of this study were the interviews of two key researchers from each project. Analytical tools came social psychology of science and the cultural-historical activity theory (CHAT).

The openness strategies of the two projects could be described as 'open by default', with no conscious measures taken to prohibit scooping. The Somus project proposal was drafted completely openly. Instead of scooping them competitors aiming for the same funding call ended up actively avoiding overlaps in their proposals, even contributed to the proposal. The tragedy of Somus was that the idealistic enthusiasm resulted in a lack of data management. Today only a PDF report remains openly available online.

The NMRLP participants are expected to credit ideas received even in informal discussions. Anyone who has commented on the project blog is eligible for co-authorship. It remains for each participant to decide for themselves whether their contribution is sufficient. Due to the transparency and peer pressure, there have been no cases of unmerited co-authorship, something that is actually rare in field in case.

For one of the NMRLP researchers, the fear of scooping was a major stress factor and openness a way of removing it. The rationale was that after your work is published online, it is easier to prove priority. All of the interviewed researchers named the unfairness of the academic publishing model as a motivation for creating alternative ways of disseminating research. The interviewees had a high level of trust in their immediate community, but some mistrust for the academic establishment.

All interviewees named scientific curiosity as a source of fulfilment and motivation. I argue that one cause for the fear of scooping comes from making publications the primary object of research activity. They should be returned to the place and purpose they suit best: outcome of a research and one communicational tool among others, instead of the object of research

The interviews show a link between understanding and recognising research integrity and research ethics issues and the willingness to share. Research integrity training for researchers is already a policy priority at least in Europe and this conclusion only adds to its importance.

The open collaboration projects show a worrying lack of engagement from women. The level of female participation was in both cases below of what can be considered typical for the fields: about 25% for SOMUS and 1/30 for NMRLP. The gender-specific barriers in the way of sharing should be further studied.

Acknowledgements

The author would like to thank the four research participants from NMRLP and SOMUS, her supervisors Doctor, University Lecturer Antti Häkkinen (University of Helsinki) and Doctor, Vice-Rector Erika Löfström (University of Tallinn, University of Helsinki) for guidance, Tiina and Antti Herlin Foundation for continuous financial support to her PhD research, which this

study is one part of, the Finnish Committee for Research Data for making attending the SciDataCon 2016 possible, and the Finnish Advisory Board on Research Integrity for in kind support during fall 2015.

Competing Interests

The author declares that she has no competing interests.

Notes

- 1 In this paper I use the term 'science' in the broadest possible sense, meaning all scholarly activity from natural science to human and engineering science.
- 2 They were asked whether they were familiar with the following documents: Responsible Conduct of Research and Handling Allegations of Misconduct in Finland 2012, the Singapore Statement on Research Integrity and the Montreal Statement on Cross-Boundary Research Collaboration. One of the participants had read the Finnish guideline, none of them had read the other two.

References

Alatarvas, R 2013 *Summary of the Finnish Science Barometer 2013*. Helsinki: Tieteentiedostus ry. Available at www.tieteentiedotus.fi/files/Sciencebarometer_2013_netsummary.pdf

Bishop, D 2015 Who's afraid of Open Data: Scientists' objections to data sharing don't stand up to scrutiny, 16 December 2015. Available at <http://blogs.lse.ac.uk/impactofsocialsciences/2015/12/16/whos-afraid-of-open-data-dorothy-bishop/> [Last accessed on 31 October 2016].

Borgman, C 2015 *Big Data, little data, no data*. Cambridge, Massachusetts / London, England: The MIT Press.

Chambers, C 2016 *How to deal with being 'scooped': The vast majority of science is a process of derivative, incremental advance*, 19 April 2016. Available at <http://blogs.lse.ac.uk/impactofsocialsciences/2016/04/19/so-youve-been-scooped/> [Last accessed on 31 October 2016].

CRADLE Center for Research on Activity, Development and Learning 2016 *The Activity System*, publishing date unknown. Available at <http://www.helsinki.fi/cradle/activitysystem.htm> / [Last accessed on 31 October 2016].

Dhand, R 2002 Does research misconduct extend beyond biomedicine? In: White, C (ed.) 2002 *The COPE Report 2002 - Annual Report of the Committee on Publication Ethics*. London: BMJ Books.

Engeström, Y 2005 Developmental Work Research – Expanding activity theory in practice. International Cultural-historical Human Sciences, Volume 12. Berlin: Lehmanns Media. pp. 17-36.

Grand, A, Wilkinson, C, Bultitude, K and Winfield, A F 2016 Mapping the hinterland: Data issues in open science. *Public Understanding of Science*, 25 (1). pp. 88-103.
<http://dx.doi.org/10.1177/0963662514530374>

Heiskanen, T, Hintikka, K A, Hintsala, T, Kokkonen, J, Kola, P and Näkki, P 2008 Tutkimusparvi - The Open Research Swarm in Finland. *MindTrek '08 Proceedings of the 12th international conference on Entertainment and media in the ubiquitous era*. New York: ACM. pp. 154-158. doi:[10.1145/1457199.1457233](https://doi.org/10.1145/1457199.1457233)

ICMJE International Committee of Medical Journal Editors 2016 *Defining the Role of Authors and Contributors* [http, publishing date unknown](http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html). Available at [://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html](http://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html) [Last accessed on 31 October 2016]

Kalela, J 2012 *Making history – the historian and uses of the past*. London: Plagrave Macmillan.

Kronkvist, J 2011 Somus as an attempt at a new paradigm. In: Näkki et al *Social media for citizen participation - Report on the Somus project*. VTT Publication 755. Espoo, Finland: VTT Technical Research Centre of Finland.

Laine, H 2015 *Ethical evaluation: passed, 11 November 2015*. Available at <https://thehonestbrokerblog.org/2015/11/11/ethical-evaluation-passed/> [Last accessed on 31 October 2016]

Lahti, L, Lehto, A, Miettinen, M and Ollila, S 2015 Beyond Open Access - The Changing Culture of Producing and Disseminating Scientific Knowledge. *Proceedings of the 19th International Academic Mindtrek Conference*. New York: ACM. pp. 202-205 DOI: <http://dx.doi.org/10.1145/2818187.2818282>

Le, M 2013 *Avoiding the scoop, 30 April 2013*. Available at <http://www.impressmagazine.com/avoiding-the-scoop/> [Last accessed on 31 October 2016].

Näkki, P, Bäck, A, Ropponen, T, Kronkvist, J, Hintikka, K A, Harju, Pöyhtäri, R and Kola, P 2011 *Social media for citizen participation. Report on the Somus project*. VTT Publications 755. Espoo: VTT Technical Research Centre of Finland.

OKM Ministry of Education and Culture 2015 *Open science and research leads to surprising discoveries and creative insights - Open science and research roadmap 2014–2017*. Reports of the Ministry of Education and Culture, Finland 2014:21.

Rodrigues, L C 1998 Having your research idea stolen. *BMJ*, 1998 Sep 26; 317(7162): 857. doi: <http://dx.doi.org/10.1136/bmj.317.7162.857>

Schunk, D and Usher, E L 2012 Social Cognitive Theory and Motivation. In: Ryan, R M (ed.) 2012 *Oxford Handbook on Human Motivation*. Oxford: Oxford Handbooks Online. DOI: 10.1093/oxfordhb/9780195399820.013.0002

Schwarzkopf, S 2016 *It's not the end of the world if your research gets 'scooped'*, 9 April 2016. Available at <https://www.timeshighereducation.com/blog/its-not-end-world-if-your-research-gets-scooped> [Last accessed on 31 October 2016].

Schofield, J W 2009 Increasing the Generalizability of Qualitative Research. In: Gomm, R, Hammersley, M and Foster, P 2009 *Case Study Method*. SAGE Publications Ltd. pp. 69-97. DOI: <http://dx.doi.org/10.4135/9780857024367>

Steneck, N H 2007 *ORI Introduction to the Responsible Conduct of Research*. Updated ed. Washington: ORI.

TENK Finnish Advisory Board on Research Integrity 2013 *Responsible conduct of research and procedures for handling allegations of misconduct in Finland*. Helsinki: TENK.

Tikhomirov, O K 1999 The theory of activity changed by information technology. In: Engeström, Y, Miettinen, R and Punamäki, R-L 1999 *Perspectives on activity theory*. Cambridge: Cambridge University Press. pp. 350-351.

Van den Eynden, V and Bishop, L 2014 *Incentives and motivations for sharing research data, a researcher's perspective*. Knowledge Exchange. Available at <http://www.knowledge-exchange.info/event/sowing-the-seed> [Last accessed 30 May 2016].

Whyte, A and Pryor, G 2011 Open Science in Practice: Researcher Perspectives and Participation. *The International Journal of Digital Curation*. Issue 1, Vol. 6, No. 1, pp. 199-213. doi:10.2218/ijdc.v6i1.182