This is a preprint of an article published in the Library & Information Science Research: Stvilia, B., Hinnant, C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M. M., & Marty, P. F. (2017). Toward collaborator selection and determination of data ownership and publication authorship in research collaborations. *Library & Information Science Research, 39*(2), 85-97. http://doi.org/10.1016/j.lisr.2017.03.004

Toward collaborator selection and determination of data ownership and publication authorship in research collaborations

Besiki Stvilia[1*], Charles C. Hinnant[1], Shuheng Wu[2], Adam Worrall[3], Dong Joon Lee[4], Kathleen Burnett[1], Gary Burnett[1], Michelle M. Kazmer[1], and Paul F. Marty[1]

[1]School of Information, Florida State University, 142 Collegiate Loop, Tallahassee, FL 32306-2100.

[2]Graduate School of Library and Information Studies, Queens College, 65-30 Kissena Blvd., Queens, NY 11367-1597.

[3]School of Library and Information Studies, University of Alberta, 3-20 Rutherford South, Edmonton, Alberta T6G 2J4, Canada.

[4]University Libraries, Texas A&M University, 5000 TAMU, College Station, Texas 77843-5000.

[*]Corresponding author. E-mail: bstvilia@fsu.edu

Abstract

This study examined factors that might affect researchers' willingness to collaborate with a specific researcher and the priorities given to those factors. In addition, it investigated how researchers determined the ownership of collaborative project data and how they determined the order of authorship on collaborative publications in condensed matter physics. In general, researchers rated their intrinsic motivations the highest, such as the quality of ideas a potential collaborator might have and their satisfaction with a past collaboration, followed by their extrinsic motivations, such as the complementary knowledge, skills, or resources the collaborator could provide. In addition, researchers who had a greater number of collaborative projects and researchers who had served as a project PI or co-PI valued the deep-level, personality-related characteristics of a collaborator higher than did those who had not. Younger researchers were more risk averse and more concerned with a collaborator's reputation and the possible cost of a collaboration decision. Additionally, younger researchers indicated more often than older researchers that they did not know whether their project teams followed any rules or norms or engaged in negotiation to determine the order of authorship on collaborative publications.

1. Introduction

The literature defines scientific collaboration as a set of activities involving multiple researchers, often with complementary knowledge and skills, who share the common objective of creating new knowledge, new products, or both (Hara, Solomon, Kim, & Sonnenwald, 2003; Van Rijnsoever & Hessels, 2011). One can use a cost–benefit analysis framework to explore factors that might affect a researcher's decision to collaborate with another researcher. Some of the benefits of research collaboration identified in the literature include gaining access to a shared pool of ideas, knowledge, skills, techniques, and instrumentation through a division of labor, and interacting with colleagues. The costs of collaboration might include the costs of travel and distributed work coordination, planning, and conflict and dispute resolution and the risk of possible reputation loss from investing one's reputation in a collaboration that might fail (Hara et al., 2003; Katz & Martin, 1997). For a researcher to collaborate with another researcher, the perceived value and importance of the benefits of the collaboration with that researcher must outweigh its perceived costs. Hence, it is essential to gain a better understanding of how researchers prioritize those costs and benefits in order to facilitate research collaborations (e.g., by increasing the benefits and reducing the costs) and make effective recommendations about potential collaborators.

Collaborative research project activities may include, but are not limited to, defining a study's objective and its design; building devices and instruments for experiments; synthesizing sample materials; designing experiments and simulations; collecting and analyzing data; discussing and interpreting results; administering and managing the project; presenting and publishing project results; and submitting patent applications (Jha & Welch, 2010; Katz & Martin, 1997; Stvilia et al, 2015). Researchers may make different kinds contributions and contribute different amounts to different phases of collaborative research projects (Birnholtz, 2006). The types of contributions could be intellectual, financial, data based, infrastructural, manuscript oriented, or presentation based. Each collaborative project might generate multiple types of data and other products throughout its life cycle (e.g., sample materials, instruments, sensor data, PowerPoint presentations, manuscripts; Stvilia et al, 2015). Hence, the information that should be collected about individual contributions to a collaborative project, the means of communicating that information to the community (e.g., through publication authorship) and other stakeholders (e.g., funding and accrediting agencies), and the accurate interpretation of that communication remain significant challenges.

Producing publications and determining who should be listed as the authors of a publication, and in which order, could be often critical to the overall success of a research collaboration. Publication is a major product of a collaboration. Successful, agreeable determination of publication authorship can increase the likelihood of converting a one-time collaboration into a long-lasting research team. The list of authors on a collaborative publication may convey several kinds of information, such as the assignment of credit, the authors' responsibilities, and the ownership of the publication (Birnholtz, 2006; Cronin, 2001). If a publication includes a data set, the list of authors may also indicate the ownership of the data. Because tenure and promotion committees, state and federal governments, and accrediting and funding bodies use authorship to assess the research productivity and impact of individual scientists, departments, and institutions, the accuracy and reliability of mapping between the ordered list of authors and the authors' actual contributions to a publication have been concerns in the literature for some time (Beaver, 2001; Culliton, 1988; "Who'd Want to Work in a Team?," 2003). Some large scientific laboratories have formal authorship policies that define publication procedures that the researchers affiliated with those laboratories must follow (Birnholtz, 2006). Even if a laboratory, scientific society, or publisher has a formal publication authorship policy, that policy rarely specifies a procedure for determining the order of authorship and what that order represents in terms of authors' contributions to a publication or a research project. Researchers often have to rely on discipline- or community-specific norms and conventions or on negotiations to guide a collaborative manuscript-writing process and the allocation of credit through the order of authorship. Hence, it is essential to study the community-specific norms and conventions used to guide the process of collaborative writing and the determination of authorship, as well as various social factors that might affect that process. It would help to construct templates aligned with the existing practices that scientific laboratories could use to guide credit assignment on collaborative publications. In addition, it would help third parties (e.g., promotion and tenure committees and hiring committees) to make a more reliable and accurate determination of a given author's contribution to a publication or a collaborative project.

2. Problem Statement and Research Questions

Selecting a collaborator is an important decision. It not only may determine the success of a collaborative project, but also can affect the researcher's career and reputation. Although a significant body of literature is available on research collaborations, including various social factors that can facilitate or hinder researchers from entering into collaborations and that contribute to the success of those collaborations, there is still a dearth of research on *how researchers select individual collaborators, what factors affect their decision making, and how researchers prioritize those factors.* This research addressed that need by examining researchers' priorities for different individual characteristics of potential collaborators in condensed matter physics (CMP), one of the largest areas of physics.

Also examined in the study were norms and patterns of data ownership and for determining the order of publication authorship in CMP. Having a greater understanding of how researchers prioritize different social factors when selecting collaborators and the norms and rules used in determining data ownership and the order of authorship on publications can help in designing more effective mechanisms for recommending collaborators in research information management systems. In addition, the findings of this study can inform the design of policies and best practice guides for managing research collaborations and can contribute to the development of more accurate and reliable metrics of the productivity and impact of individual researchers, research centers, and institutions. This study was guided by the following research questions:

1. What are some of the factors that may affect CMP researchers' willingness to collaborate with a particular researcher and what are their priorities for those factors?
2. How do CMP researchers determine the ownership of collaborative project data?
3. How do CMP researchers determine the order of authorship on collaborative publications?

3. Literature Review

One can use a cost–benefit analysis framework to explore the factors that might affect a researcher's decision to collaborate with another researcher. That is, one researcher's willingness to collaborate with another researcher can be considered as trade-off function between the potential benefits (both intrinsic and extrinsic) and the costs of collaboration. Katz and Martin (1997) identified five types of potential benefits of collaboration: (1) sharing knowledge, skills, and techniques through division of labor; (2) transfer of knowledge or skills; (3) generation of new perspectives through idea exchange; (4) intellectual companionship, relationship, and network building; and (5) more visible, strong, and impactful research outcomes. Costs of collaboration included (1) travel and communication; (2) preparing funding proposals, planning, and resolving conflicts; (3) negotiating publication authorship; and (4) project administration, including reconciling differences in administrative, disciplinary, and organizational cultures and systems. This section provides a limited review of how some of these and other costs and benefits of collaboration have been discussed by different studies. In addition, it reviews the use of data ownership and authorship norms in scientific collaborations as reported in the literature.

3.1 Intellectual satisfaction

Interesting research ideas a potential collaborator might have serve as intrinsic motivation to bring about collaboration, determine its overall success and impact, and provide intellectual satisfaction to researchers and joy from the collaboration (Melin, 2000). The similarity of research interests—the intellectual distance—between a pair of researchers may determine intellectual companionship and compatibility and their predisposition to collaboration (Hara et al., 2003; Katz & Martin, 1997; Kraut, Egido, & Galegher, 1988).

3.2 Personality, knowledge, skills, and resources

Researchers' personalities, including their credibility and generosity, are critical to collaboration success (Hara et al., 2003; Melin, 2000; "Who'd Want to Work in a Team?," 2003). The quality of team members and the prevention or successful resolution of conflicts impacts the quality and success of a collaborative project. Quality criteria—e.g. researcher competence, reliability, trustworthiness, and generosity—are frequently mentioned in collaboration literature. For example, Hara et al. (2003) found personality compatibility—including personal friendships and trust—is a requirement for highly integrated collaborative projects. Researchers' personalities can make collaborations either enjoyable or burdensome. Having fun or gaining pleasure is a benefit of a successful research collaboration (Beaver, 2001; Katz & Martin, 1997; Melin, 2000).

With increased specialization in the sciences, having access to specialized instrumentation, knowledge, and skills becomes critical to completing complex research projects and serves as a main benefit and motivation for collaborating (Bozeman & Gaughan, 2011; Finholt, 1999; Hara et al., 2003; Katz & Martin, 1997; Melin, 2000; Stevens & Campion, 1994). Researchers with complementary skills, knowledge, work styles, priorities, and characteristics can form productive, repeated collaborations (Hara et al., 2003).

3.3 Reputation

Research reputation refers to one's position in the merit-based hierarchy of a particular research community. Researchers are influenced by both intrinsic motivation—research idea quality, pleasure or joy of intellectual companionship—and extrinsic motivation—higher reputation or greater visibility by collaborating with researchers with high reputation and good community standing (Beaver, 2001; Katz & Martin, 1997). Researchers' reputation has often been used to predict the quality of their research and outcomes. A research partner with a high reputation or prominent community standing can positively affect the success and acceptance of joint grant proposals and collaborative publications (Hara et al., 2003; Melin, 2000). Melin (2000) found that established senior researchers often may feel a duty to support younger researchers' careers and help them acquire funding (Melin, 2000).

Alternatively, possible reputation loss is a risk from investing in a collaboration that might fail (Hara et al., 2003). Researchers' careers, research impact, and future collaboration opportunities may suffer if collaborators commit fraud or have papers retracted (Mongeon & Larivière, 2016). In addition, declining an offer of collaboration may cause negative effects on researchers relationships and their standing in the community.

3.4 Seniority

Collaborating with or being part of a research team led by a senior researcher could lead to higher productivity. Martín-Sempere et al. (2008) found senior researchers are more often associated with larger scientific research teams with high levels of consolidation and integration, and larger teams often lead to greater research productivity.

Too many senior members on a team could also be disadvantageous. The publication patterns of a CMP community indicated researchers' seniority levels may have mixed effects on the productivity and impact of research collaborations (Hinnant et al., 2012; Stvilia et al., 2011). Collaborations with more homogeneous seniority levels and fewer senior members were more productive; those with lower average seniority levels had higher publication impact. However, having a more senior member as the first author was associated with a higher publication impact (Hinnant et al., 2012).

3.5 Institutional, disciplinary, and cultural affiliation

Different disciplines might have different propensities for interdisciplinary collaboration. Van Rijnsoever and Hessels (2011) found researchers from disciplines dominated by basic, fundamental research (e.g., mathematics) engaged in fewer interdisciplinary collaborations than researchers from more applied disciplines (e.g., medicine). Given these differences, Haythornthwaite (2006) argued, successful interdisciplinary or cross-community distributed collaborations require attention to invisible practices. Differences in team members' knowledge, practices, and physical locations, based on their institutional associations, their disciplinary ties, and their cultural outlook, must be bridged to facilitate and increase data, information, and knowledge sharing. Such bridging is often invisible work (Star & Strauss, 1999) requiring "learning about others' fields and practices, and developing new practices" (Haythornthwaite, 2006, p. 763). Norms, values, and other characteristics of disciplines, communities, organizations, and cultures are important bridging factors that influence collaboration (Burnett et al., 2014).

High levels of internal diversity in institutional associations has been associated with increased difficulty in intrateam coordination (Cummings & Kiesler, 2005). Although diversity in institutional associations may inhibit teams' coordination, diversity within the discipline may lead a team to possess broader knowledge and skills and increased productivity (Cummings & Kiesler, 2005; Porac et al., 2004). Greater interdisciplinarity in a collaborative team can lead to as or more positive research outcomes (Cummings & Kiesler, 2005) and higher publication productivity (Porac et al., 2004; Stvilia et al., 2011).

3.6 Gender and culture

Gender and cultural diversity may have mixed effects on collaborations. Greater gender and cultural diversity may lead to higher quality outcomes, greater network reach and access to resources, and greater opportunities for sharing and unique perspectives (Katz & Martin, 1997). Increased intracollaboration conflict may occur because of collaborators' diverse backgrounds and experiences, lack of shared understandings, and different communication practices (Pelled, 1993; Pelled et al., 1999). Childbirth and child care may disproportionately disrupt female researchers' careers and productivity (Kyvik & Teigen, 1996). Gender may play a role in the establishment of personal network ties (Brass, 1985). Van

Rijnsoever and Hessels (2011) found female researchers were more likely to engage in interdisciplinary collaborations than males. Female faculty at a large Canadian university had more females in and received more psychological and career support from their professional networks than did male faculty (Rothstein & Davey, 1995). In addition, collaborations of female researchers may be more locally oriented than those of male researchers (Larivière, Ni, Gingras, Cronin, & Sugimoto, 2013).

3.7 Shared academic genealogy

Research collaboration can be a teaching and mentoring mechanism for students. Collaboration success can serve as a “rite of passage” or a sign of acceptance to a scientific community (Hara et al., 2003). One would expect faculty advisors to be one of the first collaborators for students; however, Hara et al. (2003) found some faculty may not consider students equal partners, emphasizing teaching and mentoring objectives in student collaborations. Students may serve as bridges for their faculty advisors to collaborate with other faculty, termed “collaboration through students” by Hara et al. (2003, p. 958). Alternatively, participation in collaborative projects through advisors can give students and postdoctoral researchers opportunities to expand their networks and identify new collaboration opportunities. Existing social relationships, including academic genealogy-based networks, are one of the main sources of collaboration (Crane, 1972; Katz & Martin, 1997; Melin, 2000; Sugimoto, 2014).

3.8 Geographic distance.

Collaborative decision making can be influenced by potential collaborators’ physical and social proximity. Kraut et al. (1988) found organizational and locational proximity and research interest similarity increase two researchers’ chances of collaborating. Being in close proximity increases the chances of informal and formal communication and of becoming aware of others’ complementary research interests, knowledge, and skills, which can facilitate collaborations (Hara et al., 2003; Katz & Martin, 1997). Closeness of location can lead to “lower transaction costs, ease of coordination, shared organizational culture, swift trust formation, preferential attachment, and sunk cost of investment in co-authors” (Cronin, 2008, p. 1005). These benefits may outweigh those perceived of more distant collaborations based solely on scientific congruency.

Remote communication, even with modern technologies, can be information lossy and less rich than face-to-face communication. Distance communication technology may not convey completely the context of a remote classroom or laboratory, possibly leading to inaccurate interpretations of the content or intent of the communication. Lack of shared local context (e.g., time zone differences) may be another limitation of remote collaboration (Olson & Olson, 2000). Remote communications may not be as effective as face-to-face communication for establishing trust and friendship—two requirements for repeated successful collaborations (Hara et al., 2003)—between researchers.

3.9 Data ownership and authorship norms

Activities are mediated by communities’ rules, norms, and conventions. Identifying these and contradictions among activities’ components is essential for guiding new instances of activities, predicting and understanding activities’ outcomes, and devising effective interventions (Engestrom, 2000). In research collaborations, determinations of data ownership and credit allocation through publication authorship are mediated by the rules, norms, and conventions of a laboratory, institution, or community. Wallis and Borgman (2011) found the phrase *data authorship* was confusing to and had a negative connotation (of falsifying data) among respondents at a National Science Foundation (NSF)-funded interdisciplinary research center; the term *data ownership* was seen as more straightforward and less controversial. A contributor was named the owner of data more often than PIs or the institution. Credit for collaborative publications is often allocated through authorship order (Beaver, 2001). Authors may be listed alphabetically or through the first and last place conventions: the most important contributors are listed as the first or last authors (Beaver, 2001).

In addition, team members with less research experience and more educational or teaching-based roles within and beyond the team make less use of technology, use others' data less, and seek data and information with greater serendipity than members focused on and experienced in research (Borgman, 2006). Their skills in data management and data use are usually lower than those who have more established research practices.

4. Study Design

This article reports on the data ownership and collaboration sections of a comprehensive survey of the data and collaboration practices of a CMP community gathered around the National High Magnetic Field Laboratory (NHMFL). The NHMFL is a unique interdisciplinary scientific center, one of the largest of its kind, collaboratively operated by Florida State University, the University of Florida, and Los Alamos National Laboratory. It provides scientists with free access to its facilities for research involving magnetic fields, superconducting magnetometry, magnetic resonance imaging, and magnetic spectroscopy. To better understand the data and collaboration practices, issues, and problems of the community and to develop a survey instrument, the study first conducted 12 semistructured interviews with representatives of different groups from the community, including sample material growers, experimentalists, theorists, visiting scientists, local scientists, administrators, senior scientists, junior scientists, postdoctoral researchers, and students at NHMFL. The study used concepts and relationships from the literature to develop questions for the interview protocol. The audio recordings of the interviews were transcribed and content analyzed. The study then used the interview findings to expand on and refine the set of interview questions and develop a survey instrument. The survey instrument was pretested with nine participants from the CMP community for readability and validity. The finalized survey was distributed online to 672 scientists in the fall of 2012 using Qualtrics survey software. The scientists were invited via their e-mail addresses, which were obtained from the NHMFL database of researchers who had conducted experiments using the laboratory facilities between 2008 and 2011. Only scientists who had indicated CMP as their discipline were selected. In addition, the study used two follow-up/reminder emails with two week intervals to increase the survey's response rate. The survey consisted of seven sections and 89 questions[1]. Although participants completed early sections of the survey at higher rates, 160 participants completed all the questions, resulting in an overall response rate of 24%.

The present study looks specifically at the results of the collaboration and data ownership sections of the survey, which were completed by 162 participants. The largest number of them identified themselves as

[1] The survey instrument is available upon request

White (54%) and Asian (32%). Nine percent declined to specify their race. 82% reported their gender as male, 13% were female, and 5% declined to answer the question. The median age was 38.

Before participating in the interview or completing the online survey, participants were given a consent form approved by the Human Subjects Committee of Florida State University. The form contained information about the project, including information about potential risks associated with participation in the data collection. Participants who completed the interview or the survey were e-mailed a $50 Amazon gift card.

5. Findings

5.1 Factors affecting researchers' willingness to collaborate with a particular researcher

The survey asked participants to rate, on a 7-point Likert scale ranging from *extremely unimportant (1)* to *extremely important (7)*, the importance of 21 reasons or factors that might influence their decision regarding whether to collaborate with a particular researcher. The reasons or factors were identified based on findings from the semistructured interviews and a literature analysis. In total, 162 participants responded to the questions regarding the reasons for entering into a research collaboration. Exploratory factor analysis was used to determine the underlying structure of those reasons.. The analysis treated each reason as a variable. The measure of sampling adequacy (MSA) of each of the variables was greater than 0.61, with an overall MSA of 0.72, and the Bartlett test of sphericity was significant at the 0.0001 level.

Principal components analysis was used in the study to extract the factors. Factors with eigenvalues above 1 were selected for inclusion in the factor model. The components analysis factor matrix was rotated using the Varimax rotation algorithm with Kaiser normalization. As suggested by the literature (Hair, Black, Babin, Anderson, & Tatham, 2005), factor loadings of 0.45 or greater were identified as significant. The first round of factor analysis indicated that the shared academic genealogy and the researcher's administrative position were loaded significantly on more than one factor. Because of the cross-loadings, those variables were deleted one by one from the model and the loadings were recalculated after each deletion. The resultant factor model had six factors with eigenvalues greater than 1 but without any significant cross-loadings (see Table 1). The MSA of each of the variables was still greater than 0.61, with an overall MSA of 0.70, and the Bartlett test of sphericity was significant at the 0.0001 level. The six factors captured 66% of the total variance of the data. The mean importance ratings for each aspect of collaboration are shown in Table 2. Researchers' reliability in complying with commitments and deadlines, and their genuine interest and intellectual engagement in the collaborative project were some of the other characteristics participants identified that might influence their decision regarding whether to collaborate with a particular researcher.

Table 1. Factor loadings of the reasons for collaboration, a final model

	Component					
	1	2	3	4	5	6

Researcher is from the same or a different organization	.280	–.085	.034	***.762***	–.040	.205
Researcher is from the same or a different academic discipline	–.038	.136	.027	***.821***	.029	.049
Researcher belongs to the same or a different sex	.095	.031	–.032	.073	.017	***.882***
Researcher has a similar or different cultural background	.118	.050	–.036	.136	.121	***.866***
Researcher has complementary or similar knowledge	–.017	***.840***	.217	.136	.032	.045
Researcher has complementary or similar skills	–.023	***.873***	.150	.143	–.074	–.042
Researcher has or lacks access to important resources	.029	***.668***	.032	–.086	.083	.060
Researcher’s seniority level	.294	.171	–.153	.277	***.518***	.141
Researcher’s research reputation	–.073	–.042	.253	–.013	***.840***	–.061
Reputation of researcher’s home institution	.265	.041	.090	.135	***.730***	.165
Researcher’s community affiliation	.148	.065	–.032	***.654***	.346	.011
Your satisfaction from past collaborations	.099	.154	***.618***	–.184	.237	–.129
Researcher’s personality	.162	–.090	***.797***	.129	.066	.126
Researcher has interesting research ideas	–.075	.231	***.683***	–.020	.094	–.260
Researcher has similar or different research interests	–.039	.273	***.716***	.079	–.068	.109
Possible effect of your decision on your standing in the community	***.914***	–.016	.018	.135	.089	.054
Possible effect of your decision on your standing in the organization	***.903***	–.038	.004	.078	.066	.102
Researcher’s geographic proximity	.401	.048	.120	.265	.221	.243
Your availability or lack of resources for a new project	***.457***	.435	.290	–.008	.254	.022

Note. Significant relationships are in boldface italics.
[a]Extraction method: principal components analysis. Rotation method: Varimax with Kaiser normalization.

Table 2. Mean importance ratings of the reasons for collaboration

	Mean	*SD*	Analysis *N*
Researcher has interesting research ideas	5.96	0.99	162
Your satisfaction from past collaborations	5.95	1.03	162
Researcher has similar or different research interests	5.51	1.24	162
Your availability or lack of resources for a new project	5.38	1.24	162
Researcher has or lacks access to important resources	5.26	1.64	162
Researcher has complementary or similar knowledge	5.12	1.54	162
Researcher has complementary or similar skills	5.10	1.58	162
Researcher’s personality	5.06	1.43	162
Researcher’s research reputation	5.00	1.54	162
Researcher is from the same or a different academic discipline	3.86	1.82	162
Possible effect of your decision on your standing in the community	3.86	1.68	162

Possible effect of your decision on your standing in the organization	3.76	1.63	162
Reputation of researcher's home institution	3.72	1.70	162
Researcher's community affiliation	3.24	1.69	162
Researcher's geographic proximity	3.19	1.71	162
Researcher's seniority level	3.18	1.70	162
Researcher is from the same or a different organization	2.96	1.77	162
Researcher has a similar or different cultural background	1.77	1.29	162
Researcher belongs to the same or a different sex	1.30	0.86	162

On the basis of the significant loadings, the six factors were labeled Personality, Resources, Costs, Reputation, Affiliation, and Culture. The authors then evaluated the internal consistency of the factor constructs with Cronbach's alpha. The alpha values of the constructs were 0.71, 0.76, 0.76, 0.64, 0.70, and 0.74, respectively, suggesting adequate internal consistency. The alpha value for the Reputation construct was slightly lower than the lower limit for consensus (0.70) but is still acceptable in exploratory research (Hair et al., 2005).

The reasons that loaded significantly on each factor were then used to develop summated scales. Seven summated scales were developed by averaging the scores of reasons assigned to each factor (see Table 3). The Personality scale had the highest average importance score, followed by the Resources and Costs scales. The Culture scale had the lowest average importance score. The scores for the summated scales were added to the rest of the data and used in examining the relationships among the priorities for collaborator characteristics and other aspects of collaboration practices and researchers' demographic characteristics.

Table 3. Mean importance scores for the collaboration scales.

Data quality scale	Mean rating
1. Personality Your satisfaction from past collaboration(s) Researcher's personality Researcher has interesting research ideas Researcher has similar or different research interests	5.62
2. Resources Researcher has complementary or similar knowledge Researcher has complementary or similar skills Researcher has or lacks access to important resources	5.16
3. Costs Possible effect of your decision on your standing in the community Possible effect of your decision on your standing in the organization Your availability or lack of resources for a new project	4.33
4. Reputation Researcher's seniority level Researcher's research reputation Reputation of researcher's home institution	3.97
5. Affiliation Researcher is from the same or a different organization Researcher is from the same or a different academic discipline Researcher's community affiliation	3.35
6. Culture Researcher has a similar or different cultural background Researcher belongs to the same or a different sex	1.53

The median number of collaborative projects that survey participants had in a typical year was 3, and the median project team size was 4. A nonparametric Spearman correlation test showed that the number of collaborative research projects a researcher had in a typical year was positively correlated with the Personality scale (0.18, p = 0.05) and negatively correlated with the Costs and Culture scales (–0.19, p = 0.05). The analysis also indicated that the researcher's age was negatively correlated with the Costs (–0.3, p = 0.001) and Reputation (–0.19, p = 0.02) scales. In addition, the researcher's age was positively correlated with the number of collaborative projects she or he worked on in a typical year (0.28, 0.001) and the typical size of the researcher's project teams (0.16, p = 0.05; see Figure 1).

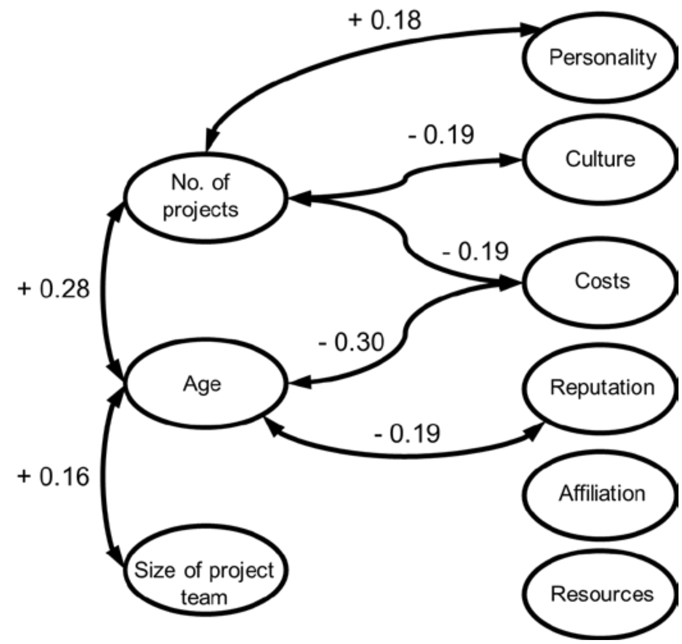

Figure 1. The Spearman correlation test of participant characteristics on the collaboration scales (only significant relationships are included).

The Kruskal–Wallis test of the dependence of the collaboration scales on participant characteristics indicated several significant relationships (see Table 4). The analysis revealed significant differences for the Personality scale scores in the primary roles participants played in funded projects in general. Principal investigators and co-principal investigators (co-PIs) had higher mean rank scores for the construct than did postdoctoral researchers and students. In addition, researchers whose project teams followed norms or policies to determine the order of authorship on publications had a higher mean rank for the Resources construct than did participants who did not. Similarly, the Kruskal–Wallis test indicated a significant dependence of the Personality and Resources scale score distributions on whether participants negotiated the authorship order on their publications. Scientists who negotiated the order of authorship on publications with the members of their project teams had higher mean ranks for those scales than did scientists who did not. Finally, non-White participants had higher mean ranks for the Culture scale than did White participants.

The Kruskal–Wallis test also indicated a significant relationship between participants' race, participants' primary roles in funded projects, and age. White participants had a higher mean rank for the researcher's age variable than did non-White participants ($\chi^2 = 13.69$, $p = 0.02$). In addition, as it was expected, PIs, co-PIs and senior investigators were older than postdoctoral associates and student research assistants. The former group of project roles had higher mean rank scores for age than the later ($\chi^2 = 96.22$, $p = 0.001$).

Table 4. Kruskal–Wallis test of the dependence of the collaboration scales on participant characteristics

		Personality	Resources	Costs	Reputation	Affiliation	Culture
Primary role in funded research projects	χ^2	***12.49***	2.39	10.52	5.70	10.64	10.70
	df	***5***	5	5	5	5	5
	sig	***0.03***	0.79	0.06	0.34	0.06	0.06
Were there any norms, rules, or policies that your project team(s) followed to determine the order of authorship for publications?	χ^2	0.25	***5.86***	0.97	0.31	0.85	1.30
	df	2	***2***	2	2	2	2
	sig	0.88	***0.05***	0.61	0.86	0.66	0.52
Have you ever negotiated the order of authorship for publications with the members of your project team?	χ^2	***9.60***	***7.21***	.09	3.42	2.10	.08
	df	***2***	***2***	2	2	2	2
	sig	***0.01***	***0.03***	0.96	0.18	0.35	0.96
Race	χ^2	6.95	10.03	5.46	2.83	5.13	***14.81***
	df	5	5	5	5	5	***5***
	sig	0.22	0.07	0.36	0.73	0.40	***0.01***

Note. Significant relationships are in boldface italics.

5.2 Determining the ownership of collaborative project data

Participants reported that most of the time, the data generated by the research projects they worked on were owned by a project PI, followed by collective ownership by a project team (see Table 5). Only 11% of participants indicated that their project team followed a norm or rule to determine the ownership of project data. An even smaller number (7%) stated that they had had the experience of negotiating the ownership of project data (see Table 6). The majority of participants indicated they had negotiated project data ownership at the beginning of the project. Half of the respondents indicated that they had negotiated data ownership at the publication writing stage (see Table 7).

Table 5. Who owns the data generated by the research projects you worked on?

	No. of responses	%
Principal investigator (PI)	107	66
Another team member (different from the PI)	28	17
The project team collectively	80	49

The research institution you work for	40	25
Funding agency	18	11
Other (please specify):	2	1
Don't know	12	7

The Kruskal–Wallis test indicated a significant relationship between the use of norms and negotiations in determining data ownership and the researcher's age. Researchers who reported that they did not know whether their project teams followed any norms or policies to determine the ownership of project data had a lower mean rank score for the researcher's age variable than did participants who reported that their teams did or did not follow a norm or policy in determining data ownership ($\chi^2 = 6.66$, $p = 0.04$).

Of the participants who reported the use of norms to determine the ownership of project data, 43% indicated that the norm their project teams followed was to own data collaboratively, compared with 29% who reported the norm was that the researcher(s) who collected the data owned it. Fewer than 14% of participants mentioned that the data were owned by the PI or by the institution or laboratory.

5.3 Determining the order of authorship on collaborative publications

In contrast to the responses on determining data ownership, the percentages of participants who reported that they used a norm or used negotiation to determine the order of authorship on collaborative publications were much higher, 50% and 54%, respectively (see Table 6). Furthermore, although data ownership was negotiated most often at the beginning of the project, authorship order was primarily negotiated at the paper-writing stage (see Table 7).

Table 6. Use of norms and negotiations to determine the ownership of project data and the order of authorship on publications

Answer	Do your project teams follow any norms regarding the ownership of project data?	Have you ever negotiated the ownership of data?	Were there any norms, rules, or policies that your project team(s) followed to determine the order of authorship for publications?	Have you ever negotiated the order of authorship for publications with the members of your project teams?
Yes, no.(%)	18(11)	12(7)	80(50)	88(54)
No, no.(%)	99(61)	141(87)	49(30)	69(43)
Don't know, no.(%)	45(28)	9(6)	33(20)	5(3)

Table 7. When did scientists negotiate data ownership and the order of authorship for publications?

No.	Project phase	Negotiated data ownership, no.(%)	Negotiated authorship order, no.(%)
1	At the start of a project	8(67)	19(22)
2	During research design	2(17)	9(10)
3	During data management planning	0(0)	1(1)
4	During data collection	2(17)	6(7)
5	During data analysis	4(33)	11(13)
6	For the presentation of findings at a conference	3(25)	31(35)
7	When writing a paper	6(50)	74(84)
8	When publishing a paper in a preprint archive (e.g., arXiv.org)	2(17)	28(32)
9	When publishing a paper in a peer-reviewed journal	2(17)	35(40)
10	At the end of the project	3(25)	0(0)
11	When preparing data for preservation	1(8)	
12	When depositing data in an institutional or subject data repository	0(0)	
13	Other (please specify):	0(0)	0(0)

All the publication authorship norms reported by participants referenced the degree of contribution and leadership as the main criteria used to determine who should be included as the authors of a publication. Six participants referenced the guidelines developed by the American Physical Society, NSF, or Nature Publishing Group to guide their decision making.

Only one participant mentioned that her team listed authors in alphabetical order. The rest reported that the relative degrees of contribution and leadership provided to a research project determined who should be listed as the lead authors of a publication. In seven cases, participants reported the use of multiple criteria to determine how to order the authors of a publication. In particular, they mentioned the career phase of an author, the number of teams participating in the collaboration, the ownership of data, and the potential effect on the prestige of a publication as additional criteria used to determine the order of publication authorship. As one participant noted, “Multiple methods [are used]: (1) Order of contribution. (2) Order that would further one member’s career when not detrimental to others. (3) Order that will raise prestige of work.”

The Kruskal–Wallis test of dependence showed that researchers who reported that their teams did or did not follow a norm or policy to determine the order of publication authorship had higher mean ranks for the researcher’s age and the number collaborative projects than did participants who reported that they did not know. In addition, researchers who had negotiated the order of publication authorship had higher mean rank scores for the researcher’s age and the number of collaborative projects than did participants who did not (see Table 8).

Table 8. Kruskal–Wallis test of dependence of participant characteristics on the practices used in determining authorship order

		How many collaborative research projects do you work on in a typical year?	What is the most common size of your project teams?	What was your age at your last birthday?
Primary role in funded research projects	χ^2	***27.05***	7.39	***96.72***
	df	***6***	6	***6***
	sig	***0.001***	0.29	***0.001***
Were there any norms, rules, or policies that your project team(s) followed to determine the order of authorship for publications?	χ^2	***9.78***	3.63	***13.42***
	df	***2***	2	***2***
	sig	***0.008***	0.163	***0.001***
Have you ever negotiated the order of authorship for publications with the members of your project team?	χ^2	***11.96***	1.82	***14.70***
	df	***2***	2	***1***
	sig	***0.003***	0.40	***0.001***

Note. Significant relationships are in boldface italics.

6. Discussion

6.1 Factors affecting CMP researchers' willingness to collaborate with a particular researcher

In the first research question, the study sought to identify some of the factors that could affect researchers' willingness to collaborate with a particular researcher. The factor analysis identified six factors, which were designated Personality, Resources, Costs, Reputation, Affiliation, and Culture (see Table 3). These six constructs are referred to hereafter in this section as the CMP collaboration reasons model.

One could further group these factors into intrinsic and extrinsic motivations that might affect a researcher's decision to collaborate with another researcher. According to Ryan and Deci's (2000) review of self-determination theory, intrinsic motivations are autonomous and self-determined because the person finds the activity she or he performs interesting or pleasant. Extrinsic motivations, on the other hand, are externally induced through rewards or punishments. The reasons grouped under the Personality factor could also be labeled as the expected satisfaction or enjoyment of collaborating with a particular researcher. Hence, they can be considered intrinsic motivations. Reasons grouped under the Resources and Costs factors can be considered extrinsic motivations. They refer to specific types of rewards or punishments associated with different outcomes of decision making about possible collaboration with a particular researcher. If a researcher decides to collaborate with another researcher, that decision may bring benefits in the form of access to complementary knowledge, skills, or unique instrumentation. Gaining access to some of the most powerful magnets in the world at the NHMFL is an example of such a reward. On the other hand, declining a collaboration request might have negative repercussions for a researcher's career or standing in a community, especially if the requesting researcher has higher seniority or holds a position in the community's power hierarchy.

Reputation is usually used as an indirect means of assessing quality (Stvilia et al., 2007). In particular, it might be used to predict the credibility of a person—her or his trustworthiness and expertise (Choi & Stvilia, 2015). For instance, the high reputation of a potential collaborator might signal to a researcher that the collaborator can bring valuable expertise, a good work ethic, or both to the collaboration. In addition to using indirect assessments, the reasons grouped under the Reputation factor can provide direct rewards and penalties. For example, a junior researcher might increase her or his chances of receiving research funding or gaining access to a higher impact publication venue by collaborating with someone with an established reputation in the community. Alternatively, an established senior researcher serving as a mentor to or collaborating with a junior researcher might be driven by the altruistic, intrinsic reward received from helping the junior researcher begin or advance her or his career (Bozeman & Corley, 2004). Finally, a researcher could be discouraged to collaborate with a another researcher by the risk to her or his reputation and career if the collaboration fails, the collaborator commits a fraud, or their paper is retracted (Hara et al., 2003; Mongeon & Larivière, 2016).

Close to the present study is that of Bozeman and Corley (2004). They used a survey of scientists and engineers working at NSF- and U.S. Department of Energy-funded centers to examine the importance of 13 reasons for researchers to collaborate: relationship length, administration request, helping junior colleagues, helping graduate students, strong science reputation, complementary skills, previous high quality collaborations, fun or entertaining personality, common fluency in language, common nationality, strong work ethic, adherence to schedules, and knowing how to assign credit. They used factor analysis to identify six underlying factors of the reasons and named the factors as follows: Taskmaster, Nationalist, Mentor, Follower, Buddy, and Tactician (see Table 9). A comparison of the CMP collaboration reasons model with Bozeman and Corley' factor model shows overlap as well as differences. The latter has a more detailed structure for past experiences with a potential collaborator, whereas the former has a single item, satisfaction from past collaboration, that covers those characteristics. The Bozeman and Corley model has two factors related to satisfaction from a past collaboration. The Taskmaster factor includes characteristics related to work ethic, whereas the Buddy factor includes an item on the quality of the prior collaborations themselves. In addition, the Bozeman and Corley model has a more detailed structure for seniority-related aspects. It contains one construct related to the mentoring intent of the collaboration when collaborators are of unequal seniority. Another construct includes items that refer to the reputational and administrative status of a collaborator. The CMP collaboration reasons model, on the other hand, has a more detailed structure for the extrinsic rewards of collaboration, such as access to complementary knowledge and resources, and for constructs related to collaboration costs. The cost-related constructs (i.e., Costs, Affiliation, Culture) include costs associated with cultural, affiliation, and disciplinary distances as well as the costs of decisions (see Table 9). More important, the CMP model has more detailed coverage of intrinsic motivations, which participants ranked highest; this includes a researcher's interest in the research idea, her or his satisfaction from a past collaboration, and the similarity or difference in their ideas. Bozeman and Corley (2004) did not report rankings for individual items or factors in their model. Hence, it is unknown what their participants' priorities were for the reasons for collaboration included in their model and how those priorities compare with values in the CMP community for the same reasons.

The CMP researchers prioritized the intrinsic motivations and reasons grouped under the Personality factor the highest, such as the quality of ideas and their satisfaction from a past collaboration; the next highest were extrinsic motivations, such as complementary knowledge, skills, or resources a potential collaborator could provide. Overall, they assigned higher priorities to the potential benefits of a collaboration than to the costs (see Table 4). This finding is in agreement with prior studies. Novel, innovative research ideas and unique research expertise, techniques, or data (e.g., new physical materials) are the main determinants of the success of a research project, its scholarly impact, or its potential for commercialization (Zucker & Darby, 1996). Hence, it is not surprising that these researcher characteristics may serve as the main attractors of potential collaborators. The literature shows that in many disciplines, the number of collaborators and the number of papers by individual scientists are distributed according to power law properties. A few star scientists are highly productive and attract many collaborators, whereas the overwhelming majority of scientists have only a few collaborators or have published only a few papers (Barabási et al., 2002; Newman, 2001).

The analyses showed that older researchers were less sensitive to the reputation or seniority of a potential collaborator. In addition, researchers who had a greater number of projects or who had served as PIs or co-PIs in a typical year cared more about the quality of ideas and the personality of a collaborator and less about social characteristics of the collaborator, such as her or his sex or culture. This finding is in agreement with the literature. The organizational literature divides personal characteristics into two groups: surface-level and deep-level characteristics. Surface-level characteristics refer to demographic characteristics that are easily observable, such as age, sex, or race. Deep-level characteristics include less observable, more psychological characteristics, such as personality, values, and attitudes (Harrison, Price, Gavin, & Florey, 2002; Jackson et al., 1995; Pelled, 1996). The literature shows that if individuals are interdependent in achieving a shared goal, their impressions of each other are more nuanced, more focused on individual personalities, and less influenced by social categories (e.g., age, race, or sex; Fiske 2000). As the history of collaboration among members of a team lengthens, the effects of perceived surface-level differences on the team's social integration may diminish and the effects of deep-level, behavior-based differences may increase (Harrison et al., 2002; Pelled et al., 1999). Hence, one would expect that researchers who have had more experience with collaboration and the management of collaborative projects would pay more attention to a potential collaborator's personality and less attention to the collaborator's surface-level characteristics compared with younger researchers who have not had that level of collaboration experience.

This study found that a potential collaborator's culture or sex was more important to non-White participants than to White participants. One may hypothesize that this too could be caused by possible differences in collaboration experiences between the two groups. Non-White participants were younger than White participants. Hence, they might have had less experience with research collaboration and might be more sensitive than older White participants to collaborators' surface-level characteristics. Future qualitative research could provide a deeper account of and explanations for the effects of cultural and gender diversity on the selection of a research collaborator and on the collaborations.

Similarly, participants who were older and who had perhaps already established themselves in the community cared less about the possible costs of their collaboration decisions on their standing in the organization or community than did younger researchers who were perhaps nearer the beginning of their careers. This could have been caused by younger researchers being less independent in their decision making about collaborations. Often, they work under the direct supervision of senior researchers. Hence, they might not be able to decline a collaboration offer from a senior researcher without incurring a negative impact on their work status or future careers. Similarly, being near the beginning of their careers and perhaps not having a similar level of job security as older, more experienced researchers (e.g., tenure), younger researchers might have to be more strategic, choosing collaborations that are less risky and that will help advance their careers in a predictable way (e.g., achieve a career milestone).

Table 9. Comparison of the CMP collaboration reasons model with the Bozeman and Corley (2004) model

	Bozeman and Corley model					
CMP collaboration reasons model	Taskmaster • Collaborator has strong ethics • Collaborator sticks to the schedule	Nationalist • Collaborator is fluent in respondent's language • Respondent and collaborator are of the same nationality	Mentor • Collaborate to help junior colleagues • Collaborate to help graduate students	Follower • Someone in administration requested the collaboration • Collaborator has a strong science reputation	Buddy • Length of time the respondent has known a person • Quality of previous collaborations with a person • Collaborator is fun or entertaining	Tactician • Respondent and collaborator have complementary skills
Personality • Your satisfaction from past collaboration(s) • Researcher's personality • Researcher has interesting research ideas • Researcher has similar or different research interests	X				X	
Resources • Researcher has complementary or similar knowledge • Researcher has complementary or similar skills • Researcher has or lacks access to important resources						X
Costs • Possible effect of your decision on your standing in the community • Possible effect of your decision on your standing in the organization • Your availability or lack of resources for a new project						
Reputation • Researcher's seniority level • Researcher's research reputation • Reputation of researcher's home institution			X	X		
Affiliation • Researcher is from the same or a different organization • Researcher is from the same or a different academic discipline • Researcher's community affiliation						
Culture • Researcher has a similar or different cultural background • Researcher belongs to the same or a different sex		X				

6.2 Determining the ownership of collaborative project data

The second research question focused on identifying how CMP researchers assigned the ownership of collaborative project data. Participants reported that most often, the data generated by the research projects they worked on were owned by the PI, followed by collective ownership by a project team. Few of them indicated that their teams followed a norm or used negotiation to determine the ownership of project data. Although a majority of those who negotiated data ownership indicated they determined data ownership at the beginning of a project, half of the participants reported that they negotiated data ownership at the publication-writing stage, and more than a third negotiated data ownership at the analysis stage (see Table 8). This could have been caused by multiple types of data being generated throughout the life cycle of a typical CMP project; the ownership of those different data could have been negotiated or determined at different stages of the project and owned by different members of the project team (Stvilia et al, 2015). Indeed, as one participant noted, "The team who measures owns the data. This is informal, common sense, and not declared in writing . . . Each collaborator owns and is responsible for the data generated by the collaborator."

Of the participants who reported the use of norms to determine the ownership of project data, 43% indicated that the norm their project teams followed was to own data collaboratively. Fewer than 14% of participants mentioned that data were owned by the PI. This finding somewhat contradicts participants' responses to the question asking who owned the data generated by the projects they worked on. On that question, participants reported that most of the time, the data were owned by the PI. It seems that if no rule or norm was used to determine the ownership of data of a collaborative research project, the default assumption was that the data were owned by the PI.

Indeed, the analysis showed younger participants stated more often than older participants that they did not know whether their teams followed a norm(s) to determine the ownership of data. This result suggests that younger CMP researchers, such as graduate students or postdoctoral researchers, were perhaps less familiar with the norms and rules used in research data management or were not in a position to make decisions about the ownership of data.

A certain amount of variance was found in the norms used to determine data ownership, which is in agreement with the literature (Wallis & Borgman, 2011). Responses included collective ownership of data by a research team, data being owned by the researcher who measured or generated it, data being owned by the institution or funding agency, or a combination of the above. In some cases, the same data could have multiple owners based on team and institutional relationships and power structures. Furthermore, data could have formal and "active" ownership. As one respondent noted, "[The] grad student or postdoc who takes it 'owns' it. [The] PI 'owns' the research and therefore ultimately owns the data. Funding agencies may own it by agreement, but they wouldn't know what to do with it."

6.3 Determining the order of authorship on collaborative publications

The third research question investigated how CMP researchers determine the order of authorship on collaborative publications. An overwhelming majority of the norms specified by participants used the degree of contribution and leadership provided to a study to determine the order of authorship on publications. In general, students and postdoctoral researchers were given priority, whereas PIs and team leaders were listed last. However, listing the order of authors by the degree of their contributions could be different. Students would be listed in decreasing order of their contributions, whereas senior researchers would be listed in increasing order. In some cases, more nuanced or complex decision making could be involved to determine the authorship order, and some collaborations could be more complex than others; for example, they could include more than one research team from more than one institution:

> "The first author should normally be the graduate student or postdoc that has carried out the work; others by contribution; in collaborative cases according to the subject of the publication, authors from certain institutions would be listed first before the other institution(s)." (s70)

The analysis showed that a much higher percentage of participants used norms and negotiations to determine the order of authorship on publications than they did to determine the ownership of data. This result could reflect a practice in which most of the time, the ownership of data is predetermined and any data produced by a collaborative project is assumed to be owned by the PI (see Table 6). This finding also could reflect the "precompetitive" (Curry, Freitas, & O'Riáin, 2010) nature of most types of data in CMP. Although some data can eventually lead to high-impact publications, opportunities for collaboration, or valuable patents, until the data are analyzed and the findings are presented or published, outside parties might not use the ownership of data to evaluate a researcher.

Only 22% of the participants who indicated that they negotiated the order of authorship on a collaborative publication reported that they did so at the beginning of the project. Negotiating the order of publication authorship when writing a paper was much more frequent, being done by 84% of participants. More than 30% of the participants reported negotiating the order of authorship when presenting findings at a conference or publishing a paper in a peer-reviewed journal. These findings might point to CMP projects producing multiple publications at different stages of their life cycles, and consequently having a need for multiple negotiations. Indeed, in addition to peer-reviewed papers, conference or workshop presentations are important mechanisms for gaining visibility, particularly for young researchers who may need to establish their research reputations in the community. Hence, a project team might need to negotiate to whom those opportunities should be given (Birnholtz, 2006; Knorr Cetina, 1999).

The analysis indicated that researchers whose team(s) used norms or negotiations to determine the order of authorship of a publication had higher mean ranks for the Resources scale than did those who did not. Furthermore, researchers who negotiated the order of authorship of publications ranked the Personality and Resources scales higher than did those who did not. In addition, the researchers were older and had a larger number of collaborative projects in a typical year. These findings again might point to the importance of the quality of research ideas, complementary knowledge and skills, and researchers' personalities in the success of collaborations. Likewise, researchers who have more collaborative projects and work with larger project teams might have more need for and experience in the use of norms and negotiations to determine publication authorship. Hence, it is important that researchers who plan to engage in a collaborative project, especially with a large team, are properly trained in the social and ethical aspects of collaborative research, including how to determine the authorship order on collaborative publications and the responsibilities associated with that authorship (Nature, 2017).

The findings of this study suggest that younger researchers might be less involved than older researchers in decision making related to collaborative projects and that the process might remain opaque to them. This result echoes the findings by Hara et al. (2003) that some senior researchers might not consider students working with them on joint projects as collaborators and that they might see themselves in those projects more as teachers than as collaborators. Future research could examine in greater detail the present practices around student participation in decision making on collaborative research projects. In particular, it could investigate how the power asymmetries in academia might affect that participation, as well as students' knowledge and sense making of the management of research collaborations.

The present study found that different collaborations used different norms to determine the order of authorship. However, few collaborations used norms to determine the ownership of data. The variance and lack of consistency in determining the order of authorship on collaborative publications may lower the quality of interpretations of citation data for the purposes of determining research productivity and evaluating impact. With the growing interest in collecting and aggregating more complete information on researchers' productivity and impact from different sources and in developing more nuanced metrics (i.e., other than a raw citation count or h-index) to measure research productivity and impact, the ability to interpret citation data in a meaningful and reliable way becomes increasingly important (National Information Standards Organization, 2016; Priem, Piwowar, & Hemminger, 2012). Scholarly communities, libraries, publishers, and data aggregators promote the use of persistent, unique identifiers for authors, publications, and data sets to enhance the accuracy of bibliographic entity disambiguation, data linking, and aggregation (Lee & Stvilia, 2014). Standardizing the norms used to determine the order of authorship on collaborative publications into templates and making those templates machine readable and referenceable with persistent identifiers from a designated registry would further enhance the quality of research productivity and impact assessment. Future related research could investigate the data infrastructure needed for such a registry.

7. Conclusion

This exploratory study extended the literature on scientific collaboration by identifying additional factors that may affect how researchers select individual collaborators, and researchers' priorities for those factors. In addition, the study extended the data curation literature by analyzing the practices of determining data ownership and the order of publication authorship in CMP. In particular, the study found that researchers rated their intrinsic motivations the highest, such as the quality of ideas a potential collaborator might have and their satisfaction with a past collaboration, followed by their extrinsic motivations, such as the complementary knowledge, skills, or resources the collaborator could provide. Characteristics related to a potential collaborator's reputation, affiliation, culture, and sex were rated lower. In addition, researchers who had a greater number of collaborative projects and researchers who had served as a project PI or co-PI valued the deep-level, personality-related characteristics of a collaborator more than did researchers who had not. Younger researchers were more risk averse and concerned about a collaborator's reputation and the possible cost of a collaboration decision. Furthermore, younger researchers indicated more often than older researchers that they did not know whether their project teams followed any rules or norms or whether the order of authorship on collaborative publications was determined through negotiation. The findings of this study can inform the design of best practice guides, policies, and training modules used to manage research collaborations in the CMP community and other related communities.

8. Acknowledgements

The authors thank two anonymous reviewers for their helpful comments and suggestions on an earlier version of this manuscript. This research was supported in part by the Florida State University Office of Research and by the NSF under Grant OCI-0942855. The article reflects the findings and conclusions of the authors, and do not necessarily reflect the views of Florida State University, the NSF, or the NHMFL.

9. References

Barabási, A. L., Jeong, H., Néda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A: Statistical Mechanics and Its Applications, 311*(3), 590–614.

Beaver, D. D. (2001). Reflections on scientific collaboration (and its study): Past, present, and future. *Scientometrics, 52*, 365–377.

Birnholtz, J. P. (2006). What does it mean to be an author? The intersection of credit, contribution, and collaboration in science. *Journal of the American Society for Information Science and Technology, 57,* 1758–1770.

Borgman, C. L. (2006). What can studies of e-learning teach us about collaboration in e-research? Some findings from digital library studies. *Computer Supported Cooperative Work, 15*(4), 359–383. doi:10.1007/s10606-006-9024-1

Bozeman, B., & Corley, E. (2004). Scientists' collaboration strategies: Implications for scientific and technical human capital. *Research Policy, 33*, 599–616.

Bozeman, B., & Gaughan, M. (2011). How do men and women differ in research collaborations? An analysis of the collaborative motives and strategies of academic researchers. *Research Policy, 40*, 1393–1402.

Brass, D. J. (1985). Men's and women's networks: A study of interaction patterns and influence in an organization. *Academy of Management Journal, 28*(2), 327–343.

Burnett, G., Burnett, K., Kazmer, M. M., Marty, P. F., Worrall, A., Knop, B., … Wu, S. (2014). Don't tap on the glass, you'll anger the fish! The information worlds of distributed scientific teams. In P. Fichman & H. Rosenbaum (Eds.), *Social informatics: Past, present, and future* (pp. 118–134). Newcastle, UK: Cambridge Scholars.

Choi, W., & Stvilia, B. (2015). Web credibility assessment: Conceptualization, operationalization, variability, and models. *Journal of the Association for Information Science and Technology, 66*, 2399–2414.

Crane, D. (1972). *Invisible colleges: Diffusion of knowledge in scientific communities*. Chicago, IL: University of Chicago Press.

Cronin, B. (2001). Hyperauthorship: A postmodern perversion or evidence of a structural shift in scholarly communication practices? *Journal of the American Society for Information Science and Technology, 52*, 558–569.

Cronin, B. (2008). On the epistemic significance of place. *Journal of the American Society for Information Science and Technology, 59,* 1002–1006.

Culliton, B. J. (1988). Authorship, data ownership. *Science, 242,* 658.

Cummings, J. N., & Kiesler, S. (2005). Collaborative research across disciplinary and organizational boundaries. *Social Studies of Science, 35*(5), 703–722.

Curry, E., Freitas, A., & O'Riáin, S. (2010). The role of community-driven data curation for enterprises. In D. Wood (Ed.), *Linking enterprise data* (pp. 25–47). New York, NY: Springer. doi:10.1007/978-1-4419-7665-9_2

Engestrom, Y. (2000). Activity theory as a framework for analyzing and redesigning work. *Ergonomics, 43*, 960–974.

Finholt, T. (1999). Collaboratory life: Challenges of Internet-mediated science for chemists. In National Research Council (Ed.), *Impact of advances in computing and communications technologies on chemical science and technology: Report of a workshop* (pp. 97–108). Washington, DC: National Academies Press.

Fiske, S. T. (2000). Interdependence and the reduction of prejudice. In S. Oskamp (Ed.), *Reducing prejudice and discrimination* (pp. 115–135). Mahwah, NJ: Erlbaum.

Hair, J., Black, B., Babin, B., Anderson, R., & Tatham, R. (2005). *Multivariate data analysis*. Upper Saddle River, NJ: Prentice-Hall.

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

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface- and deep-level diversity on group functioning. *Academy of Management Journal, 45*, 1029–1045.

Haythornthwaite, C. (2006). Articulating divides in distributed knowledge practice. *Information, Communication and Society, 9*, 761–780.

Hinnant, C., Stvilia, B., Wu, S., Worrall, A., Burnett, G., Burnett, K., …. Marty, P. (2012). Author team diversity and the impact of scientific publications: Evidence from physics research at a national science lab. *Library & Information Science Research, 34*, 249–257.

Jackson, S., May, K., & Whitney, K. (1995) Understanding the dynamics of diversity in decision-making teams. In: Guzzo, R. and Salas, E. (Eds.), *Team Effectiveness and Decision Making in Organizations*. San Francisco, CA: Jossey-Bass. pp. 204–261.

Jha, Y., & Welch, E. W. (2010). Relational mechanisms governing multifaceted collaborative behavior of academic scientists in six fields of science and engineering. *Research Policy, 39*, 1174–1184.

Katz, J. S., & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26*(1), 1–18.

Knorr Cetina, K. (1999). *Epistemic cultures: How the sciences make knowledge.* Cambridge, MA: Harvard University Press.

Kraut, R., Egido, C., & Galegher, J. (1988). Patterns of contact and communication in scientific research collaboration. In I. Greif (Chair), *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*, Portland, OR, September 26-28, 1988 (pp. 1–12). New York, NY: ACM Press.

Kyvik, S., & Teigen, M. (1996). Child care, research collaboration, and gender differences in scientific productivity. *Science, Technology & Human Values, 21*(1), 54–71.

Larivière, V., Ni, C., Gingras, Y., Cronin, B., & Sugimoto, C. (2013). Bibliometrics: Global gender disparities in science. *Nature, 504*(7479), 211–213.

Lee, D. J., & Stvilia, B. (2014). Developing a data identifier taxonomy. *Cataloging & Classification Quarterly, 52*, 303–336.

Martín-Sempere, M.J., Garzón-García, B., & Rey-Rocha, J. (2008). Team consolidation, social integration and scientists' research performance: An empirical study in the biology and biomedicine field. *Scientometrics, 76,* 457–482

Melin, G. (2000). Pragmatism and self-organization: Research collaboration on the individual level. *Research Policy, 29,* 31–40.

Mongeon, P., & Larivière, V. (2016). Costly collaborations: The impact of scientific fraud on co-authors' careers. *Journal of the Association for Information Science and Technology, 67,* 535–542.

National Information Standards Organization. (2016). *NISO alternative assessment metrics (altmetrics) initiative*. Retrieved from http://www.niso.org/topics/tl/altmetrics_initiative/

Nature. (2017). *Authorship.* Retrieved March 27, 2017, from http://www.nature.com/authors/policies/authorship.html

Newman, M. E. (2001). The structure of scientific collaboration networks. *Proceedings of the National Academy of Sciences, 98*(2), 404–409.

Olson, G. M., & Olson, J. S. (2000). Distance matters. *Human–Computer Interaction, 15*(2), 139–178.

Pelled, L. H. (1993). *Work group diversity and its consequences: The role of substantive and affective conflict* (Unpublished doctoral dissertation). Stanford University, Stanford, CA.

Pelled, L. H. (1996). Demographic diversity, conflict, and work group outcomes: An intervening process theory. *Organization Science, 7*, 615–631.

Pelled, L. H., Eisenhardt, K. M., & Xin, K. R. (1999). Exploring the black box: An analysis of work group diversity, conflict, and performance. *Administrative Science Quarterly, 44*(1), 1–3.

Porac, J. F., Wade, J. B., Fischer, H. M., Brown, J., Kanfer, A., & Bowker, G. (2004). Human capital heterogeneity, collaborative relationships, and publication patterns in a multidisciplinary scientific alliance: A comparative case study of two scientific teams. *Research Policy, 33,* 661–678.

Priem, J., Piwowar, H. A., & Hemminger, B. M. (2012). *Altmetrics in the wild: Using social media to explore scholarly impact* (arXiv preprint). arXiv:1203.4745.

Rothstein, M. G., & Davey, L. M. (1995). Gender differences in network relationships in academia. *Women in Management Review, 10*(6), 20–25.

Ryan, R. M., & Deci, E. L. (2000). Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *American Psychologist, 55*(1), 68.

Star, S. L., & Strauss, A. (1999). Layers of silence, arenas of voice: The ecology of visible and invisible work. *Computer Supported Cooperative Work, 8*(1–2), 9–30.

Stevens, M.J., & Campion, M.A. (1994). The knowledge, skill, and ability requirement for teamwork: Implications for human resource management. *Journal of Management, 20,* 503–530.

Stvilia, B., Gasser, L., Twidale M., B., Smith L. C. (2007). A framework for information quality Assessment. *Journal of the American Society for Information Science and Technology, 58,* 1720-1733.

Stvilia, B., Hinnant, C., Schindler, K., Worrall, A., Burnett, G., Burnett, K., … Marty, P. (2011). Composition of scientific teams and publication productivity at a national science lab. *Journal of the American Society for Information Science and Technology, 62*, 270–283.

Stvilia, B., Hinnant, C., Wu, S., Worrall, A., Lee, D. J., Burnett, K., Burnett, G., Kazmer, M. M., & Marty, P. F. (2015). Research project tasks, data, and perceptions of data quality in a condensed matter physics community. *Journal of the Association for Information Science and Technology, 66*(2), 246-263.

Sugimoto, C. R. (2014). Academic genealogy. In B. Cronin and C. R. Sugimoto (Eds.), *Beyond bibliometrics: Harnessing multidimensional indicators of scholarly impact* (pp. 365–382). Cambridge, MA: MIT Press.

Van Rijnsoever, F. J., & Hessels, L. K. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy, 40*, 463–472.

Wallis, J. C., & Borgman, C. L. (2011). Who is responsible for data? An exploratory study of data authorship, ownership, and responsibility. *Proceedings of the American Society for Information Science and Technology, 48*. doi:10.1002/meet.2011.14504801188

Who'd want to work in a team? (2003). *Nature, 424*(6944), 1.

Zucker, L. G., & Darby, M. R. (1996). Star scientists and institutional transformation: Patterns of invention and innovation in the formation of the biotechnology industry. *Proceedings of the National Academy of Sciences, 93*(23), 12709–12716.