

The Open Science Framework: Improving Science by Making It Open and Accessible

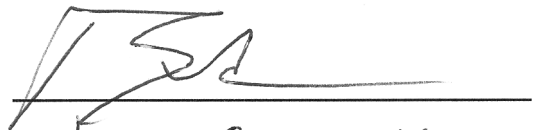
Jeffrey Robert Spies
Bryan, Ohio

M.A. Quantitative Psychology, University of Notre Dame, 2007
B.S. Computer Science, University of Notre Dame, 2004

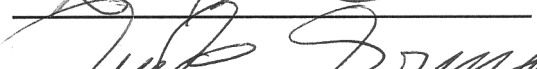
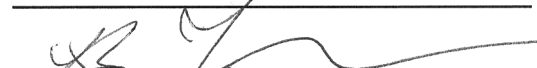
A Dissertation presented to the Graduate Faculty
of the University of Virginia in Candidacy for the Degree of
Doctor of Philosophy

Department of Psychology

University of Virginia
May, 2013



Karen Schmidt



Julie Brown

Abstract

There currently exists a gap between scientific values and scientific practices. This gap is strongly tied to the current incentive structure that rewards publication over accurate science. Other problems associated with this gap include reconstructing exploratory narratives as confirmatory, the file drawer effect, an overall lack of archiving and sharing, and a singular contribution model - publication - through which credit is obtained. A solution to these problems is increased disclosure, transparency, and openness. The Open Science Framework (<http://openscienceframework.org>) is an infrastructure for managing the scientific workflow across the entirety of the scientific process, thus allowing the facilitation and incentivization of openness in a comprehensive manner. The current version of the OSF includes tools for documentation, collaboration, sharing, archiving, registration, and exploration.

Table of Contents

Abstract	ii
Table of Contents	iii
Dedication and Acknowledgement	iv
Chapter 1: Overview of Previous Research	6
Chapter 2: Architecture	26
Chapter 3: Interviews	30
Chapter 4: Survey	42
Chapter 5: Usage	49
Chapter 6: Conclusions and Future directions	53
References	67
Appendices	79
Tables	83

Dedication and Acknowledgements

First off, I would like to acknowledge and thank my advisor, the often emulated, but never replicated, Brian Nosek (not literally—his work has been replicated; for examples, see Nosek & Hansen, 2008; Joy-Gaba & Nosek, 2010; Hawkins & Nosek, 2012). If it was not for his dedication to science, an oddly fortuitous alignment of our interests and skills, and his encouragement that this could, in fact, be more than a side project, I would probably have a very different path ahead of me. I am excited about this path and looking forward to our future partnership in this endeavor.

I would also like to thank Scott Maxwell and Greg Madey, whose mentorship in years past still has considerable influence on my work today. It is highly visible in this project. I'd also like to thank a more recent mentor, Rachel Keen, whose advice and encouragement during my time at UVA I have valued highly.

I have been lucky to work with a number of great scientists, and the members of my committee are no exception. Thank you to Mike Gorman, Karen Schmidt, Bethany Teachman, and Timo von Oertzen. Their feedback has had a significantly positive impact on this project, and I am looking forward to working with them in the future.

I have also received valuable feedback on this project from the community that somehow discovered the Open Science Framework mailing list, volunteered to be beta testers, and self-organized to publish together en masse as the Open Science Collaboration on a variety of projects. The community effort is inspiring.

I would like to thank my friends in the psychology department at UVA, especially the faculty and students in the quantitative area and members of the Nosek Lab. Their

support of this project and support more generally has been appreciated. I would also like to thank my friends outside of UVA for their support—even if it came mostly disguised as jokes about still being in school. A special thanks goes to Rebecca Dore, who has been a significant source of support throughout the entirety this process; it would have been much less pleasant without her.

And last, but certainly not least, I owe a debt of gratitude to my family for their constant support and love and, in particular, to my mother. That constancy is something I cherish and admire, and it is in and of itself motivating. As is the constancy of being asked how one's dissertation is going and when it would be finished by one's mom, aunts, and grandparents every time they call—that's *motivating* as well.

Chapter 1: Background and Overview

Across the sciences there exists a gap between scientific values and scientific practices. Accompanying this gap there also exists behaviors that lead to inefficiencies in conducting science and accumulating knowledge. The problems associated with this gap are well understood as are the solutions: increase disclosure, transparency, and openness. The difficulty arises in implementing these solutions. Scientists do not have incentives to practice them, or the incentives pale in comparison to the incentives for achievement. Scientists have developed workflows to be successful in a system where the primary motivation for reward and advancement is publication. However, some types of findings are more publishable than others. This can put scientists' motivations for achievement at odds with their motivation to be accurate (Nosek, Spies, & Motyl, 2012). It is this consideration that must be taken into account when implementing solutions to change the current, closed culture to a culture that is more open.

The Open Science Framework (OSF) is an infrastructure for managing the scientific workflow across the entirety of the scientific process, and it enables openness in a comprehensive manner. Because a scientists' workflow is the process by which scientists derive success under any incentive structure, workflow is the level at which implemented solutions—specifically, those that facilitate openness—can be incentivized. For this dissertation, the OSF is introduced, the current feature set of the OSF will be enhanced, and the success of these features as solutions to the problems presented below will be evaluated.

The Problems

The gap between scientific values and scientific practices manifests itself in many ways throughout the scientific process. The problems discussed below are those that can be solved with openness.

The Current Incentive Structure

It is well understood that a scientist's professional success is highly dependent on his/her ability to publish—jobs, salary, tenure, promotions, and funding are all dependent on publishing (Mahoney, 1985, pp. 30, Miller & Serzan, 1984; Peters & Ceci, 1982; Martin, 1992). Beyond the individual, publications are used as a criterion for the evaluation and ranking of departments and universities (Ball, 2005; Haslam et al., 2008; Nosek et al., 2010; Ostriker, Holland, Kuh, & Voytuk, 2009; Phillips, 2007). Given the extremely competitive nature of publishing (American Psychological Association, 2011; Zuckerman & Merton, 1971; Hargens, 1988) and the strong tie to professional success, the objective to publish often and quickly is clear (Martin, 1992; Sovacool, 2008).

The fact that publication is valued is not in itself the problem; the problem is that there are well-known publication biases and conforming to these biases increases the chance for publication. For example, novel results are more desirable than replication (Neuliep & Crandal, 1990; Madden, Easley, & Dunn, 1995; Neuliep & Crandall, 1993; Giner-Sorolla, 2012). A number of works have detailed practices that can increase publishability but decrease validity (Nosek, Spies, Motyl, 2012; Fanelli, 2010a; Giner-Sorolla, 2012; Greenwald, 1975; Ioannidis, 2005; John, Loewenstein, & Prelec, 2012; Kerr, 1998; Martinson, Anderson, & Devries, 2005; Rosenthal, 1979; Simmons, Nelson, & Simonsohn, 2011; Sovacool, 2008; Young, Ioannidis, & Al-Ubaydli, 2008). Conflict

occurs when scientists' professional motivations to be successful are pitted against motivations to be accurate (i.e., do good science).

Professionally, it is in a scientist's interest to publish whether or not the findings are accurate (Hackett, 2005; Martin, 1992; Sovacool, 2008). While most scientists would not intentionally commit fraud (Fanelli, 2009; Steen, 2011), strong professional motives produce motivated reasoning to arrive at the desired outcome (Kunda, 1990), especially when practices that will increase the likelihood of this outcome are common, accepted—even encouraged—and, under some circumstances, reasonable. The directional goals for success instantiate motivated reasoning to justify practices as accurate, when, in fact, these practices only increase likelihood of publication (Fanelli, 2010a). When the situation is particularly complex—information is limited or ambiguous or multiple courses of action can be justified—motivated reasoning is most influential (Bersoff, 1999; Boiney, Kennedy, & Nye, 1997; Kunda, 1990). Complexity and uncertainty are common states in cutting-edge science.

Motivated reasoning can occur without intention. Scientists are more likely to readily accept outcomes that conform to their hypotheses and scrutinize those that don't (Bastardi, Uhlmann, & Ross, 2011; Ditto & Lopez, 1992; Lord, Ross, & Lepper, 1979; Pyszczynski & Greenberg, 1987; Trope & Bassok, 1982). Scientists are prone to constructing confirmatory narratives for unexpected results (Fischhoff, 1977; Fischhoff & Beyth, 1975). By applying exploratory techniques to the analysis of data, they are more likely to find the outcome that is most publishable and most conformable to their predictions (Simmons et al., 2011; Wagenmakers, Wetzels, Borsboom, & van der Maas,

2011). The very fact that human memory is limited and research may span great lengths of time invites errors in remembering (Reyna & Brainerd, 1995), providing ample room for remembering the goals, plans, and analysis decisions that look best to be the ones that were present all along. When initial hypotheses are legitimately forgotten, scientists assume that the results they found are what they expected before data collection (Christensen-Szalanski & Willham, 1991; Fischhoff, 1977; Fischhoff & Beyth, 1975).

Reconstructing the Narrative

A narrative: a researcher is analyzing a complex data set. Via a series of linear regressions, the researcher prunes a set of twenty predictors down to two that stand out as highly significant in the same model. And it's perfect—they can easily be explained. But the researcher decides to try one more model. After he cleaned the data, he played with several graphical techniques and stumbled on an interesting plot, and it included another variable. He had already confirmed that the third variable was not be a significant predictor in the model and that made sense theoretically, but the plot suggested an interaction with one of the original two variables. He tests it, it is statistically significant, and one of the other variables falls out of the model. The researcher includes several other variables in the model, tests all possible two-way interaction—even a few three-way interactions—and it's clear: the interaction he predicted is the only one that holds up to scrutiny. It's a guaranteed publication, makes perfect sense, and is completely unexpected in the literature.

This story is only uncommon in that the researcher did good exploratory work to come up with his eventual model. What is common is that this finding would be woven

into a confirmatory narrative, where it would have the highest likelihood of getting published. Scientists are taught to tell a good story by reporting findings as if they were expected (Bem, 2003). The researcher will likely not even recognize that the narrative has changed. For example, it might take some time for the researcher to draft the manuscript, making it difficult to remember the exact process—it was the finding that was so vivid. Without a systematic way to document the process, neither the reader nor the scientist will ever actually know what the series of events that led to the finding were, and, thusly, they will never know how much of the finding was leveraged by chance.

File Drawer Effect

The file drawer effect is well understood (Rosenthal, 1979). Within the dominant testing framework—null hypothesis significance testing (Cohen, 1994; Nickerson, 2000; Rosnow & Rosenthal, 2009; Rozeboom, 1960)—the *de facto* criterion for publishing is $\alpha = .05$. While it is well known that a positive result does not indicate the probability that the hypothesized result is real (Bakan, 1966; Cohen, 1994; Frick, 1996; Oakes, 1986) and a bias toward the positive is detrimental to scientific accuracy and efficiency, most published results are positive (Sterling, 1959; Sterling, Rosenbaum, & Weinkam, 1995; Fanelli, 2010b, 2012). And, the rate of positive results seems to be increasing over time (Fanelli, 2012). Furthermore, Greenwald (1975) showed that behavioral scientists are biased against negative results and are less likely to pursue or report them, and journals are less likely to publish them. Scientists pursue design, analysis, and reporting strategies to this end in order to increase the likelihood of publication (Fanelli, 2010a; Ioannidis,

2005; John et al., 2012; Simmons et al., 2011; Wicherts, Bakker, & Molenaar, 2011; Wong, 1981; Young, et al., 2008).

Given limited journal space and intense competition, not everything can be published and, due to publication biases, what is not being published is not random. It is the science that is being systematically left out that affects the accuracy of literature. These biases not only lead to inaccuracies in the body of public knowledge, but also create inefficiencies in the field. Without any way of knowing what has been tried and produced null or negative results, there is likely a great deal of overlap of work. Furthermore, there are many occasions in which the experimental design is strong enough that the results are of interest whatever the outcome. But, if some outcomes are more publishable than others, systematic biases in the evidence base will persist.

Lack of Archiving and Sharing

In a letter to Robert Hooke (1676), Isaac Newton said of his own science, “If I have seen further it is by standing on the shoulders of giants.” Science is a cumulative process—one by which past developments inform the present and are extended. For this to occur, sharing and archiving in a properly communicated and organized manner is essential. Given the complexity of science, this is a nontrivial task especially when the ideal is to communicate the entirety of the process. And yet there is no systematic solution for doing so, nor are scientists trained how to do it. Furthermore, there is no shared infrastructure to do so in a comprehensive way to account for the process, and information gets lost.

Perhaps partly as a consequence of the difficulty in obtaining previous materials along with skewed incentives, the scientific culture lacks appreciation for replication. This is particularly surprising given that replication is a defining characteristic of science (Bacon, 1267/1859; al Haytham, 1021, as translated by Sabra, 1989; Jasny, Chin, Chong, & Vignieri, 2011; Kuhn, 1962; Lakatos, 1978; Popper, 1934; Rosenthal, 1991; Schmidt, 2009)—one that differentiates it from other methods of understanding (Braude, 1979; Collins, 1985; Schmidt, 2009). When replications are not being valued, disproportionate credence is given to work that has not been verified and, cyclically, it creates a scientific culture with little respect for sharing and archiving.

Good evidence for a culture of devalued sharing concerns data sharing. Open data sharing policies (National Institutes of Health, 2003; National Science Foundation, 2011) as well as technology to facilitate this type of sharing (Crosas, 2011) are quite progressive, and yet scientists do not share their data even when ethically obligated to do so (Wicherts, Borsboom, Kats, & Molenaar, 2011; Savage & Vickers, 2009). An unwillingness to share data has been demonstrated to be related to the number of errors in the resulting manuscript (Wicherts et al., 2011); that is, the data that need to be examined the closest out of accuracy concerns are the data not being made available.

A Singular Contribution Model

With outcome and publication being so highly incentivized, earning authorship on publications is the scientists' overriding objective. This ultimately devalues other parts of the scientific process. Collaboration is an important part of any difficult enterprise, but if certain roles are not credited with authorship on publications, then those roles will

either go unfilled or will fail to attract appropriate experts who wish to get credit for their contribution. For example, we are entering an age of big data; big data can be defined as data so large as to cause new problems because of size—these problems include analysis, but also cleaning, storing, and accessing. In the current contribution model, there is no obvious credit for contributors to data acquisition, cleaning, and management. A scientist cannot currently define themselves as someone who is good at collecting data, much less cleaning and storing it, and hope to receive meaningful credit for their work. Similarly, the incredibly high value of placement in an authorship list can discourage collaboration as scientists may desire solo authorship or senior authorship. This leads to inefficiency in that optimality is not based on maximizing throughput, but on professional motives. It can also lead to inaccuracy as this limits the potential of experts getting involved and incentivizes those without expertise participating at likely inefficient levels.

The Solution: Openness

The common solution to the problems described above is openness. Openness is an essential part of any definition of science that includes verification and replication—these core aspects of science rely on disclosure and are important to varying approaches to scientific understanding (Gorman, 1992). Disclosure provides evaluators with a basis for computing confidence (Simmons et. al, 2011). Currently, disclosure happens almost entirely through the journal publication, and it is clearly not enough. Buckheit and Donoho (1995) summarize how much of the actual process is available to the reader: “A scientific publication is not the scholarship itself, it is merely advertising of the scholarship.” Disclosure across the scientific process would make plain the narrative to

both the reader and researchers reducing the propensity for reconstructing the narrative, and making available the entire methods and materials behind the research rather than just a summary in the written report. There would not be a file drawer problem as nothing would be hidden or lost.

Beyond the fact that openness is critical to the scientific process at a definition level, openness also creates accountability (Lerner & Tetlock, 1999). Nosek, Spies, and Motyl (2012) argued that the increase in accountability, among other things, make openness the ultimate solution for when professional motivations are in conflict with accuracy motivations. Given the tie between scientific claims and the scientists that make them (Nosek & Bar-Anan, 2012), increased accountability means that scientists' reputations are more closely tied to the accuracy of their claims. Reputation is currently the metric for success in the sciences, but it is tied solely to publication. If reputation were instead tied to accuracy, the conflict between professional motivations and accuracy motivations would not exist.

In support of these notions, the Human Genome Project credits two core policies with the success of the project (Lander et al., 2001). The first was that data would be published within 24 hours of assembly. The second was an absolutely inclusive view as to who could collaborate on the project. If a lack of sharing and disclosure does anything to impede collaboration in the sciences, openness will in turn increase collaboration. If accuracy motivations were aligned with professional motivations and sharing was commonplace, an increased gathering of expertise may occur, especially across disciplines. Not only that, but openness may occur earlier in the research process in order

to increase the likelihood for errors to be identified and corrected before the project is completed. This would not only create greater accuracy, but it would also make the research process more efficient.

Full disclosure of the entirety of the research process would also allow credit be distributed more fairly. Instead of the singular contribution model with authorship on publications, individual contributors can be credited for their unique contributions. Publication is just one part of the process. If contribution were credited across the entire workflow, then collaboration would likely increase—allowing experts to do what they do well and get credit for it. Furthermore, a scientific culture that optimized efficiency and accuracy via collaboration might value curation of the scientific process to maximize information transmission. For example, literate programming (Knuth, 1984) is an approach to programming where documentation and code are embedded together. This approach has been taken up by scientists in order to integrate the analyses they perform with the manuscript they inform (Leisch, 2002).

An increase in openness could further be the precipitant for a more general cultural shift in how the scientific process is viewed, performed, and the motivations that drive it. The open source movement in the field of software development drastically changed how software was developed and the motivations at play in developing software (Lerner & Tirole, 2001; Ye & Kishida, 2003; Bonaccorsi & Rossi, 2003). In contrast to proprietary development, open source software relies on the power of peer review and transparency of process; with source code made available, users can access, use, or modify the code and then redistribute it. Often, however, that code is reintegrated into the

original work, allowing for distributed innovation and development across an entire community. This also allows for community driven error detection leading to more robust and reliable code with little centralized management. Open source software is the basis for the majority of the internet infrastructure and access tools. It has been so successful that it has been adopted by businesses and governments over proprietary development models because of the gains in reliability, quality, and lower cost.

The Open Science Framework

This dissertation implements a solution for promoting openness as a web-based software infrastructure called the Open Science Framework (<http://openscienceframework.org/>). The OSF is an infrastructure for collaborating, documenting, archiving, sharing, and registering scientific materials. The goal of the OSF is to facilitate and incentivize openness throughout the entirety of a scientists' workflow. At its core, the system is a workflow management system, of social networks. The difference from social networks is that in addition to connecting users, it connects projects and materials in projects.

The basic data structures in the system are projects and components. Both projects and components have a contributor list, timestamps, a wiki, tags, logs, and a file repository. On the backend, projects are components with the additional feature that they can contain other components. Because components have the same features as projects, including independent contributor lists, they can be used to organize and curate the project into modular and encapsulated pieces of a scientists' workflow. For example, a project may have the following components: "Background", "Hypotheses", "Materials",

“Data”, “Analyses”, or “Publication”. (From this point on any reference to component will mean project or component unless specifically referring to components that can contain other components.) Components can individually be made public, allowing for users to keep certain aspects of their projects private while making others public. There is no mandate on how projects are organized or publicized. Independence in contributor lists, privacy settings, and content allows users flexibility as to how to conduct and represent their science.

Documentation and Version Control

An accurate narrative of the scientific process is essential in conducting good science. This includes the order of events as well as the state of materials at given times. Some things, like time-stamping the occurrence of certain events, can be done automatically. Others require active curation on the part of the scientist. Because the process is just as important as the individual steps, a common infrastructure for managing the entirety of the scientific workflow is necessary. The infrastructure must be robust to the many workflows of scientists by not mandating a specific organization or method of curation, but providing enough support as to be usable. The OSF meets these requirements via components, multiple documentation tools, and automatic version control.

A broad definition of version control is the management of changes to information. In software development, where version control systems are commonly used to manage source code, the information is generally in the form of files. The core

feature of version control systems is documenting what has changed, how it changed, when it changed, and who changed it.

The OSF provides versioning throughout the system. Every component has a log documenting this information. Projects aggregate logs across the components they contain. In this way, a history of the components and its content is documented and readily available for inspection. Wikis provide online documentation for every component; by definition, they include integrated version control. Files are also version controlled as soon as they are uploaded to the system via the drag-and-drop interface or standard file selection dialog. When a file is uploaded to a component in the system, it is immediately committed to that component's file repository. The repositories are Git repositories. Git is currently one of the most widely used version control systems, and the most popular distributed version control system (Eclipse Foundation, 2012). In the OSF, when a file is uploaded, but already exists in the system, a new version is created representing the latest version. However, the previous versions are still accessible. When those versions are downloaded, they have their upload date appended to the file name.

Not only does the OSF provide a common infrastructure to document the entirety of the scientific process as constructed by the user, it also makes it easy to disclose this documentation to selected individuals or the public. Also, by providing a simplified front-end to a powerful version control system, users may find themselves appreciating the tool and being incentivized to integrate a version control mindset into their workflow. This would improve management of their own projects and files, and also improve discoverability of the development of the research project later by others.

Sharing

Accurate disclosure is a critical component of good science. Evaluators must know the basis of a claim in order to evaluate the claim. Comprehensive sharing of research results would nullify the file-drawer problem. Open data sharing policies as well as technology to facilitate this type of sharing are quite progressive and known to be important, but, again, sharing does not occur even when required. Some aspect of this is probably linked to fear of errors being discovered. Sharing of materials increases the reproducibility of results because some results can be dependent on how the research materials were designed. Re-using the same materials increases the likelihood of reproducing the prior results.

The issues surrounding disclosure and sharing are complex and the core of a movement towards openness. The first step in disentangling these issues is to make sharing trivial—in an age of the internet and rapid communication, there should never be a technical or organizational barrier to sharing. The second step is to incentivize sharing within a scientist's workflow. The major failing of the tools that exist to facilitate the sharing of data or materials is that they give little respect to the scientific workflow. Existing solutions are appended to the process, adding more work to be open on top of the work done to conduct the science in the first place. Scientists are busy and adding more work is not a good way to encourage openness.

The OSF makes sharing trivial. A contributor on a component can make it public, allowing read-only privileges, with a mouse-click. As long as the researcher is using the OSF to manage their files, the step of adding openness is just a matter of desire rather

than requiring additional effort. Researchers also have more fine-grained control over access. If a researcher wants to maintain private status while allowing another user to view or contribute to the work, they can add that individual to the component to gain access without making it public. Because the system covers the entirety of the scientific process, sharing can be easily integrated at the workflow level. With the inclusion of reputation metrics, sharing can further be incentivized in its own right. For example, the OSF documents the number of page views and downloads of projects and components. Public projects will earn more of both. Sharing allows for a scientist's work to be extended, which reflects functional interest and importance. The incentive for sharing can then come from having a quantifiable metric that can be linked to a scientist's reputation.

If sharing were practiced, errors could be detected and corrected at an earlier stage in the research process thus reducing their effect and alleviating the fear of making them in the first place. Collaboration would be valued more highly because it would increase error detection or reduce error creation. A culture may develop in science that is less scared of failing and more driven to succeed, that is accumulating knowledge quickly and accurately.

Forking

Proper credit giving is essential in science given its incremental nature. The goal of being cited drives the publication system. Scientists judge the quality of journals by their citation impact index, a value that should reflect the potential to be cited (it does not; see, Nosek & Bar-Anan, 2012) and, implicitly, quality. Citation as reputation is well

understood, so a natural method to motivate accuracy and good science in this context would be to provide a citation metric more grounded in doing good science.

The OSF promotes a type of citation—and metric—that reflects just that. In software engineering and version control, a fork occurs when a repository is copied into an independent repository, thus creating two distinct projects with a common source. The same concept is applied to the components of the OSF. A user can fork a component—copying that component and all its content as a new component with a contributor list of only the user who did the forking. A link back to the source is maintained and displayed on the page of the forked component.

This citation is very functional in that it describes how many times another scientist wanted that project or component to exist in their own workspace in order to use or extend it. Extension, which can be quantified via differences calculated between the source and the fork, elicits Newton's adage of standing on the shoulders of giants: the number of forks reflects how many scientists are standing upon that components' contributors' shoulders. Credit is immediately granted via the link back to the original displayed on every page. And because pointers exist to source projects, and forks of forks can exist, by traversing the graph to its source, we can quantify how many people high are standing upon one's shoulders. The correlation between these numbers and importance is as apparent—if not more so—as quality derived by citation index and number of citations, but much more strongly embedded in the tenets of science.

Registration

In confirmatory research, disclosure is critical to increased confidence in the order with which the process was conducted. For replication, a thorough understanding of the exact state of a project at certain stages is critical (e.g., the materials used for data collection, the data used in analyses). To combat the file-drawer effect, providing knowledge that a study exists before the outcome is realized is necessary. In each example, a priori registration increases transparency and encourages confidence.

Registration in the OSF is the act of copying a component and freezing it such that its content can no longer be edited. This new component can either be private or public, but, no matter the privacy setting, it is listed on the page of the registered component. Users are also given the option of filling out a survey, selected from a list of options of registration templates, whose data will be attached to the frozen component. These registration templates can be created by users using a scripting language the supports loops and conditionals. For example, the template could dynamically generate a survey: for every hypothesis in a project, ask three questions; if the answer to the first question has a particular response, ask one additional question.

The OSF has particular advantage in this domain as timing is so important to registration. Registration is not appended onto the process of doing science; it is built into the system that manages workflow. The documented timestamps that occur with every action have added meaning within the workflow because the system that generates them is the same for every other action, including the registrations themselves. This time information that is contained by every project in the form of logs creates a context that support the claims made by registration. The logs are frozen along with all of the other

content. The registered component(s) might have hypotheses in the wiki, or materials in the file repository, or pre-written analyses scripts followed by another registration containing a data component. Additionally, users do not have to worry about documenting each step to ensure understandability. The components are what they are. Would benefits of curation exist in order to improve understandability and thus confidence? Of course, but when an entire workflow is preserved exactly, any curation just serves to improve how quickly another person can understand the narrative.

Searching and Exploration

Citation relies on work being discovered. Fortunately, the problems related to search and discovery have been greatly solved with the information age, the internet, and companies like Google, Microsoft, and Yahoo. Work that is discoverable will be cited more. And yet scientists willingly publish their work to closed-access journals that confine the work with a paywall. This is one of the greatest ironies of the current incentive system.

The OSF combines analysis of wiki content as well as tags to create a simple search engine for the OSF's current release. Furthermore, all pages are rendered to be maximally indexed by search engines like Google. In a future version of the OSF, using the indices generated for search, heuristics, analysis of the component-contributor-forking network, and user content, like tags, users will be provided with interactive tools to not only explore the scientific network of their projects and other public projects, but also receive recommendations on projects or people they may be interested in viewing. This type of technology can help aggregate meta-data across studies allowing for

increased automation in meta-analyses (i.e., increasing identification of and reducing type-II errors).

Reputation Metrics

The current metric used to evaluate the quality of a researcher's work is citation count and the citation index of the publication where it was published. The OSF provides real-time analytics as an array of alternative metrics concerning viewership of content or files (number of views, downloads, etc.). Other sites provide similar metrics. Number of links pointing to a given URL as collected by search engines like Google can be used a modern citation count. By concentrating on workflow, and incorporating forking as a feature of the OSF, the number of forks a project or component has provides a new type of citation.

Overview of Chapters 2 through 6

Chapter 1 described the feature set for the 1.0 target release of the OSF. All of the listed features are operational in the current public beta release. The public beta was developed for this dissertation. Transition to Version 1.0 will occur in 2013 as this project is moved from a dissertation to being the core infrastructure supported by the Center for Open Science. This non-profit venture will support the further development of this software by employing a team of developers to enhance the feature set as well as facilitating the growth of the open-source community contributing to the project.

The OSF is being developed in such a way as to maximally promote open source contribution, both by the Center's staff and by the open-source community. These design decisions will be discussed in Chapter 2. In order to inform OSF's design decisions with

respect to supporting the scientific workflow, I conducted a semi-structured interview with four scientists and surveyed a wider audience about how using the OSF compared to their alternative workflows. These are discussed in Chapters 3 and 4 respectively. Current usage statistics as compared to predictions made on statistics as of 4 months ago are examined in Chapter 5. Chapter 6 discusses the future direction of the OSF.

Chapter 2: Architecture

The OSF is being architected to maximize modularity and flexibility of its components. This will maximize the potential for open source community contribution from programmers of all experience levels. For example, the majority of the technology stack is Python-based. Python is open source and has a design philosophy that emphasizes readability making it an easy language to understand and learn (Python Software Foundation, 2013). These attributes have fostered the emergence of an active and passionate development community--a community that the OSF would like to engage. Another design decision was in the web framework; the OSF uses a “micro-framework” called Flask. The micro-framework has fewer dependencies than a traditional framework, allowing for more flexibility when choosing libraries that support the OSF-specific content. It can also take longer to get started because less content works right out-of-the-box. As the web application grows in size, customized libraries are often required--those out-of-the-box features require editing. However, there is a benefit to this lack of dependence on existing libraries. When editing software with strong dependencies, the number of interactions increases with each change. This increases the complexity of the codebase and results in a higher barrier of entry for community involvement. Microframeworks have fewer of these issues.

A goal of the current design is to allow a developer to work on one aspect of the OSF without requiring knowledge or interaction with many others. For example, a student learning web technologies could download the wiki code and have only the wiki running without needing code for the version control system or registrations. This is done

by creating controllers (logic that connects the model or data representation to what the user sees) that have abstract methods or hooks that can be overwritten. In this way, the OSF's wiki could become a stand-alone wiki by overwriting the authentication methods. Instead of asking the OSF's database for the current user and their permission, developers could write their own code that connects to some other database. The hook is always the same (e.g. "get_user"), but the implementation can be easily changed. This has three effects: (1) the codebase is easier to understand allowing less experienced programmers a point of entry, (2) the codebase can easily be integrated into other projects thus incentivizing developer involvement, and (3) community efforts can be efficiently directed to the functionality of the given module without distracting or overwhelming contributors.

I chose a NoSQL database, specifically MongoDB--a highly scalable database engine--for similar reasons. NoSQL databases differ from SQL or relational databases in that they are schema-less. Schemas represent database structure and organization. They define integrity constraints for the database in order to ensure accuracy and consistency in the data stored. However, these constraints make rapid prototyping difficult, as a migration process must occur for each change. With that said, schemas do have value in that they provide integrity assurances. In order to take advantage of the flexibility and robustness of MongoDB while providing integrity checks, I wrote a custom ODM (Object Data Mapper). This allows for the data in the database to be treated like an object. This makes programming with data from the database feel like one is programming with local data structures. It also allows for functions to be attached to variable types in order

to ensure integrity or deal with it if a constraint has been violated. It also allows for a lazy migration process to occur should a change impact previous data; that is, as soon as the data is accessed and a new ODM schema has been recognized, the data will upgrade itself using definitions from the developer. Lazy, in this case, is a good thing.

The ODM also supports a variety of backends, of which MongoDB is only of one. For example, using the same code with a minor configuration change, a developer could interact with MongoDB or, for example, a local file-system database. This allows for community developers to write modules without needing to install and maintain a complicated database engine for development purposes, but allow the exact same code to be integrated in large projects, like the OSF.

The OSF's file handling is built on top of the Git version control system. While Git has many desirable features that make it an easy choice for version control implementation, the fact that it is currently one of the most widely used version control systems and the most popular distributed version control system is very attractive. The OSF will eventually support other version control protocols, but choosing to implement the most popular one first allows us to rapidly connect with users that already use the system, as well as encourage open source contribution from an already active development community.

Lastly, design decisions have been made in order to not preclude Application Programming Interface (API) development and integration. An API is a protocol that defines a programmatic method for interacting with a service. For example, Github is a Git repository host with tools aimed at developers. While the OSF offers similar services,

if the Github is a service that is part of OSF users' workflow, the OSF can programmatically interact with the Github via their API, documenting and perhaps backing up what occurs on Github. The goal is to decrease perceived competition and encourage collaboration. It also allows for easy connection with the service's community of users and developers. An API will be available for other services to connect to the OSF and make use of for their own offerings. Because the OSF has no profit-motivation, new services can be built on top of the OSF as a backend, allowing the science that takes place at those services to be documented and easily made available, while allowing other services to innovate in terms of user interface, complimentary service provision, etc.

In short, developers can start working with the OSF software with a relatively small initial investment, many contributors can be working on different parts of the code with less concern of interference with each other's work, and the OSF-developed components can integrate with other software packages easily without refactoring each code base, further incentivizing participation. These strategies are expected to have significant payoff in the future development of the OSF.

Chapter 3: Interviews

Interviews Overview

Semi-structured interviews were conducted with four scientists in order to understand the major components of their scientific workflow, the role technology plays in each component, how efficient each component is, how materials are shared with collaborators, and whether or not the OSF can be integrated into that workflow.

Interviews Method

Participants.

Four psychological scientists with connections to the University of Virginia were invited to take part in the interviews. Three of the participants were graduate students at various points in their academic career, and one was a faculty member.

Procedure.

Participants were first asked to describe the major components of their workflow. These components were added to the following set if they were unique: project conceptualization, hypothesis generation, material generation, data collection, data analysis, manuscript generation, and project archival. Then, for each component, the participants were asked to describe what technologies were involved, how collaboration and sharing occurs, how efficient they felt their process is, and how the OSF could be integrated into their workflow.

Interview Summaries

Participant A

Participant *A* is a mid-stage graduate student in social psychology. *A* has a number of studies hosted on the OSF and has integrated it into her workflow. *A*'s ideas come from anywhere--reading a paper or seeing a talk--but most projects are non-creative. Her advisor tells her what to do, or the project is a logical extension of prior work, for instance in a new domain. Project conceptualization is more about selecting good questions that are interesting or motivating. *A* puts some of her ideas in a paper notebooks or Google Docs, but most ideas are not very good. If she circles similar questions and they come up more than once, then she studies that topic. She is part of a strong mentorship model--research is handed down to her as a list of projects: her job is to turn their raw ideas into a research project. Some ideas arise at conferences.

A wishes that her peers would read more, as she feels a lot of work is repeated due to ignorance of the literature. It may be a communication issue as ideas from many disparate places; it is difficult to separate the signal from the noise in tools like Google Scholar, for example. Research is badly packaged when it comes to titles and abstracts. *A* is surprised more attention is not paid to these in peer review given it is the only view of the project most readers will see. The lack of reading leaves individuals stuck in their own sub-areas and is an impediment to better theories. *A* believes that components in the OSF can help make connections between what is being tested conceptually and then analysis plan. A Google Doc- like interface would be more simple and user friendly than the current wiki.

With use of the OSF, *A* logs hypotheses more explicitly as compared to the hypotheses only existing in her mind. The OSF forces *A* to think "out loud". Hypotheses

are such a core piece of the research process, and explicit statement ahead of time has value. With that said, if it's an idea that is just going to fizzle, she doesn't want to write it down. Her plan: don't put a hypothesis on the OSF until there is a plan to collect data.

In terms of material generation, *A* runs many web and/or reaction time studies. She keeps the code to be able to run again. Undergraduates might collaborate in this process by filling out templates. That process could be more efficient. *A* lot of time is spent educating research assistants to do this, and there is no savings when new assistants are trained. Communication on testing and editing comes via email or in person. *A* does not store materials on the OSF, but in Dropbox, email, and the lab server. She uploads the materials to the OSF post-publication.

Data collection is time consuming in that *A* has to train RAs on how to conduct in-lab studies. She writes procedural scripts in order to reduce this time as well as reduce experimenter variability. On-the-fly changes, however, do not end up on this script and are almost never documented. *A* uses email to give access to Google Docs to her collaborators. Most people are already signed up for Google Docs, the interface auto-saves, has version control, has better formatting options, and requires less clicks to edit. Google Docs also inspires more frequent checking as it is part of her Google account. *A* anticipates making data available to the public, but only after publication for fear of being scooped. A significant amount of data cleaning happens at run-time via scripting, so the actual dataset used in analyses is never readily available.

A uses a variety of tools to analyze her data including SPSS, SAS, R, and Excel. *A* uses SPSS because of its point-and-click user interface, which is especially advantageous

when trying new analyses or data cleaning methods. Excel has the advantage of being easy to use to create graphs. *A* will often put analyses in Word, Powerpoint, or Excel to share with collaborators. *A* will often ask peers for scripts that might be relevant to her analyses. *A* feels this step of the workflow is inefficient. There is a lot of interactivity in the process--realizing something else needs cleaned, redoing analyses, fixing an error, redoing analyses, etc. There is a lot of starting over and writing syntax can be time-consuming. *A* believes the value in making materials at this stage public comes in when posting cleaning scripts. The analyses change so much that it is not worth her time to document and share them.

Word is the primary tool *A* uses for manuscript generation, although she does use Google Docs when working with particular collaborators. Typically *A* uses Word with Track Changes turned on and emails drafts between collaborators. *A* has recently begun including a page at the beginning of the document detailing the version history of the document. This most recent action has made things more efficient, however, the complexity begins when having to go back to data analyses and cleaning--this is where a lot of checking and rechecking occur. *A* feels that manuscript generation is highly intertwined with analyses. *A* prefers Word with Track Changes or Google Docs over the OSF.

For project archival and curation, data cleaning scripts are well documented--*A* runs them every time an analysis is conducted rather than output and storing the data. The syntax is simply not as readable when using SPSS as compared to R (and to some degree SAS). *A* feels that the scripts used for analyses should be recreatable and, because of their

dynamic nature, believes dealing with redundant or obsolete versions is not worth her time. Archived data does not have much perceived value: if there are too many variables, and they are not well documented or labelled, *A* immediately disregards the datasets value--she will not read documentation. The OSF has value in this arena in that materials can be made available and links to them can be included in manuscripts. *A* is unsure about how effective this will be and says that the domain name is too long.

A would like to see templates for components with a checklist for each or be able to use copies of previous components as a starting point.

Participant B

Participant *B* is an early-career professor in social psychology who advises students and runs a lab. Most of her projects begin with a question and the hope to generate tests to answer that question. *B* evaluates the possible answers to that question and usually accepts one answer as the hypothesis. *B* then formulates the design needed to test that hypothesis. By herself or with a colleague, *B* creates a Word document that describes plans for the first step, summarizes the question, and includes a few ideas for the studies. During this stage, the Word document is used with Track Changes turned on and shared via email or Dropbox. *B* feels that, at this stage in particular, better organization would make the process more efficient. She thought Google Wave, a defunct Google product for collaborating and communicating, would be the technology to change this. It would have documented the conversation around the process and merge everything together in the same stream, and there would be a history of the work. It did not work. Doing this is difficult in Word because, after a few rounds, it is hard to see the

history. At this point, *B* uses email to share these Word documents as a method of tracking conversation. *B* also tried Evernote, but it did not work well. At the idea stage, *B* would like everything to be linked and visible to get a holistic view of the process and its history. *B* values being able to upload files to the OSF and can see the possibility of a system that has features similar to what she thought were going to exist in Google Wave.

B follows the idea stage by adding the purpose and procedure to the Word document, either herself or having a student collaborator do it. This includes methods, materials, and conditions, but is not always as detailed as she would like. She does not pressure students to make it more detailed. Sometimes this will be turned into an abstract for an imaginary paper with predicted results--sometimes it will be multiple hypothetical abstracts. This helps *B* evaluate whether the project is important enough; it also helps *B* generate new ideas or realize something else needs to be in the study.

When the experiment is running, *B* will often analyze the data of approximately 1-2 subjects in each condition to check for errors. Sometimes the data is analyzed again before all the planned data have been collected, and if the effect is negative (i.e., does not exist) then the data collection might be stopped to avoid wasting time and money. When data collection is finished, *B* analyzes the data independently, in conjunction with a student, or allows capable students to analyze on their own. There are occasions when analyses will be conducted in an iterative process. *B* pastes output from analyses into a new Word document and might add a sentence of interpretation. *B* also adds the purpose and procedure, sometimes pulled from the previous Word document, along with a short conclusion. For the manuscript, analyses are conducted once more, but usually without

looking at the previously created Word document; that document is used only if the manuscript is never written-up (and that second analysis never conducted) or if there was a delay in publication.

Across all steps, *B* is uploading Word documents and associated files to the OSF. It is helpful that all of the relevant files are visible in one location. *B* also perceives sharing on the OSF to be an advantage. All of the materials are documented and students can always look at past work. *B* feels that all projects, published or not, should have a summary statement for curation purposes. However, *B* feels that documenting on the OSF via the wiki is not convenient, and *B* would rather Google Doc functionality more like Word. *B* would also like offline access to materials.

Participant C

Participant *C* is quantitative psychology Ph.D. candidate transitioning to a professorship. *C* does a considerable amount of consulting, does not collect data, but instead acquires previously collected data and applies advanced modeling techniques to generate and explore questions that are unable to be asked with simpler analytic techniques. She does not use the Open Science Framework. *C*'s work is highly dependent upon the wishes of her collaborators. When working in a group, *C* documents ideas in order to stake claim on a question. She may write ideas in a Word document or use email to share ideas with collaborators. Her current workflow is "good enough", but could be more efficient. She believes that if there were a general method everyone knew how to use (even Word document templates), the system would be improved. *C* believes that the

OSF could be such a system, but is concerned about the learning curve. Once over that, however, the process could be much more efficient.

When working on previously collected data, *C* usually receives the data in anonymized or deidentified form by email. Occasionally, collaborators will use Dropbox or exchange data in-person using a flash drive. *C* believes that the OSF could be particularly useful here as uploading large files to Dropbox can be a hassle, in person transfers are not always possible, and e-mail is not as secure as *C* would prefer.

C's data analysis consists of specialized software for her area of expertise as well as R, Mplus, and EQS--they are all script based. In this domain, collaboration style varies greatly by collaborator. When working with research assistants, for example, *C* has found that Dropbox is the most convenient as it provides a centralized place for students to post the analysis they were working on that she can keep track of in order to provide feedback. While Dropbox works well with RAs if they are conducting structured tasks that require review, email works well with others. She has encountered space limits on Dropbox when doing simulation research. If the OSF did not have similar space constraints, it would easily replace Dropbox. *C* believes that having one system for all projects would be more efficient. There would only need to be one introduction for all workflow components rather than introductions to each method for a given step in the workflow; it would also reduce time spent choosing one method over another--with API connections to other research components, individuals could work the way they choose, but still have everything accessible on the OSF.

When it comes to manuscript generation, *C* works mostly in Word--she would prefer LaTeX, but few collaborators use it. The track changes feature in Word is also compelling. Exchanging drafts occurs via email. Dealing with drafts can be frustrating as individuals have different preferences on when to consider a draft a new version, but otherwise a naming system is efficient enough. *C* believes that the OSF's automatic version control will improve this and eliminate the time spent discussing how to deal with drafts. Having version history may also help avoid the difficulty of back tracking on revisions when an author changes his or her mind or when multiple authors reopen a previously concluded debate. The OSF could document the history of these decisions.

C currently does not engage in project archival, other than making sure documents are backed up on external hard drives, Dropbox, and/or email. She believes she would be more likely to make things public if she were using the OSF--one click would be sufficiently easy. Making things available would have advantages in the scientific community as, now, knowledge of another's unpublished work happens via word of mouth. The OSF could make this transmission much faster.

Overall *C* feels that a major source of inefficiency has to do with switching methods depending upon who is on the project. The OSF would provide a means to get everyone on one system for every stage of the research process.

Participant D

Participant *D* is a late-stage Ph.D. candidate in clinical psychology. *D* collects her own data and works mostly closely with her advisor. *D* has not used the Open Science Framework. *D* has a formal method for organizing her research. She begins with a Word

document for logging project ideas that she generates or are generated by her weekly advisor meetings. While she does use Dropbox and Google Docs for working with larger groups, most of her collaboration occurs during in-person meetings and via email.

Once an idea turns into an actual project, *D* creates a subfolder in her “Papers in Progress” folder. In that folder, each step of her process has its own Word document. *D* sees value in the Project/Component structure of the OSF as it maps nicely on her current workflow.

D creates experiment generation scripts using tools like Media Lab, Inquisit, and ePrime. They are not organized and versions are difficult to track causing error and inefficiency. *D* believes the automatic version control of the OSF would solve this problem.

For analyses, *D* uses SPSS and AMOS. While using SPSS, *D* uses the “Paste” function to extract scripts that can be annotated. This annotation is, essentially, her results section. Collaboration does not occur with the scripts, but instead over the results themselves. By annotating the scripts in this way, it makes it easy to retrace her steps.

When it comes time to generate the manuscript, *D* merges each of the Word file components as well as the analytic annotations into a first draft. She uses a system with her advisor of changing the date and adding initials in the file name as well as using Track Changes. Sometimes drafts with multiple initials can be confusing, but by storing them in the same folder, she is able to trace the history of the files fairly easily. When working on larger collaborations, this system can become complex, and *D* desires a more efficient way to handle this. *D* sees value in the OSF being centralized and having the

version control happening automatically such that the most recent update is always available. *D* continues using this Word file as component system post-submission for dealing with reviewer response letters.

D does not publically archive her projects, but marks the final draft as such in her papers folder. She does not make her projects public because of the time it takes, the time required for learning how to do it, and the fact that there is no credit for doing so. *D* claims that using the OSF for her workflow would make her much more likely because of the ease by which it is to make her materials public. There still exists a lack of perceived return value; it would be too easy not to do it.

D considers each step an iterative process, and, as such, believes each project is a stepping stone to another project. She does not see a logical process for organizing and making that explicit. By having everything together in the OSF--no hunting for files, efficient organization system--she thinks it would be easier to build off of previous work as opposed to starting over again. Forking projects/components would be allow for this and leave a trail back to the project that inspired it.

Interview Results

Based on the interviews, it is clear that workflow style is highly variable between scientists. One point of consistency is the importance of efficiency. If a practices causes inefficiency with perceivably minor differences in quality, it has a decreased likelihood of adoption.

The people with whom the participants collaborate dictate some variation in workflow. Future development should place heavy focus on reducing the OSF's barrier to

entry, decreasing learning curve, and creating connections between a variety of tools in order to maximize collaboration among multiple teams even using different tools.

It is clear that the use of technologies including email, Google Docs, Dropbox, Word, Word's Track Changes functionality, and file-system based organization play a large role across the participants and within their collaborator networks. The features of each will be a focus of future development and/or integration of the tool in some form.

Chapter 4: Survey

Survey Overview

The benefits and barriers of use of the OSF were examined with a usability survey distributed to a sample of OSF users. The goals of this survey were to understand the relationship between the solutions offered by the OSF as compared to a user's current solutions, the ease of use of the system, the likelihood of using the solutions offered by the system, and the perceived barriers and advantage of using the OSF for the following features: documenting, archiving, sharing with collaborators, sharing content publically, and registration. The survey can be found in Appendix A.

Survey Method

Participants. Participants were recruited via an email announcement to registered Open Science Framework users, the Open Science Framework/Collaboration mailing list, the UVA Psychology Graduate Student mailing list, and announcements on the author's personal Twitter account, the Open Science Framework Twitter account, and accounts of those individuals who retweeted the announcement. The announcement included a link to a three-minute video covering Open Science Framework features as well as a link to the online questionnaire. Of the sample who started the survey (N=107), 72% completed it (N=77). Of these 77 participants, 6% were regular users of the OSF, 9% occasional users, 34% just starting to work with it, 46% with a general impression of it, and 5% with no experience at all.

Procedure. Participants completed the online questionnaire consisting of eight pages: informed consent, a more in-depth introduction to the study, five pages of

questions covering the features listed in the following section, and page asking for additional comments and voluntary email address. After submission, participants were thanked for their participation.

Measures. As described in Appendix A, participants were asked a set of questions on the following activities: “document your research”, “archive your research materials”, “share research materials with collaborators”, “share research materials with non-collaborators”, and “register study designs and/or analysis plans in advance of data collection or analysis”. Before these questions, participants were asked about their experience with the OSF and, afterwards, if the user would volunteer their email address for follow-feedback and if they had other comments not addressed in the previous set of questions.

Survey Results

Participants were asked how the OSF compares to their current solution on a seven point Likert scale from “Much Worse” (-3) to “Much Better” (3). Using a t-test to compare responses to 0, or no difference between the current solution and OSF, participants responded that their solution was significantly worse than the OSF for documenting, $t(68)=-3.52$, $p=0.001$, but not for archiving, $t(67)=-1.74$, $p=0.087$, sharing with collaborators, $t(64)=-1.46$, $p=0.148$, sharing with non-collaborators, $t(55)=-1.21$, $p=0.231$, or registering materials, $t(57)=-0.64$, $p=0.525$. When asked if participants could perform the activity on the OSF the way they would like, 86% said yes for documenting, 81% for archiving, 78% for sharing with collaborators, 96% for sharing with non-collaborators, 96%, and 92% for registering materials.

Participants were asked how likely they were to use the OSF from “Not at all likely” (-1.5) to “Very likely” (1.5) on a 4-point Likert scale with no explicit 0. *t*-tests were used to test a difference from 0, or neither likely nor unlikely to use the OSF. Participants were significantly likely to use it for sharing with non-collaborators, $t(54)=2.44$, $p=0.018$, but not documenting, $t(73)=1.97$, $p=0.053$, archiving, $t(62)=1.61$, $p=0.113$, sharing with collaborators, $t(61)=1.96$, $p=0.055$, and registering materials, $t(52)=1.74$, $p=0.087$. Participants were asked to endorse a set of items for how likely they are to use the OSF for each activity. Possible response options were: I will always use it; I haven't tried it, but I plan to; I will use it when my collaborators use it; I will use it when it is easier than my current solution; I will use it when it is faster than my current solution; I will use it when I'm mandated to use it; I will use it if others demonstrate its value in the long term; and, I will not use it at all. Percentage endorsement is listed in Table 1. To summarize, endorsement of “I will not use it at all” was minimal across all activities, while “I haven't tried it, but I plan to” was endorsed the most frequently across activities. Endorsement percentages were the lowest for sharing with non-collaborators and registering materials and highest for documenting materials.

Participants were asked how easy it is to perform a given activity on the OSF from “Impossible” (-3.5) to “Trivial” (3.5) on an eight point Likert scale with no explicit 0. *T*-tests compared the average rating against 0, or no bias towards easiness or difficulty. The OSF was significantly easy for all tasks: documenting, $t(64)=3.84$, $p<0.001$, archiving, $t(55)=4.84$, $p<0.001$, sharing with collaborators, $t(55)=4.03$, $p<0.001$, sharing

with non-collaborators, $t(46)=4.40$ $p<0.001$, and registering materials, $t(47)=4.80$, $p<0.001$.

When asked how much time the OSF saves as compared to their current solution on the short-term from “Adds Much More Time” (-3) to “Saves Much More Time” (3) on a seven point Likert scale, as compared to 0 (no difference) participants said it added significantly more time for documenting, $t(65)=-5.53$, $p<0.001$, and archiving, $t(60)=-5.7$, $p<0.001$, but neither significantly added nor saved time for sharing with collaborators, $t(59)=-0.72$, $p=0.471$, sharing with non collaborators, $t(49)=0.25$, $p=0.805$, or registering materials, $t(48)=-1.31$, $p=0.197$. When asked the same question but instead over the long-term, participants said it would significantly save time for documenting, $t(65)=3.27$, $p=0.002$, archiving, $t(59)=2.44$, $p=0.018$, sharing with collaborators, $t(59)=3.21$, $p=0.002$, and sharing with non-collaborators, $t(49)=3.56$, $p=0.001$, but neither significantly added nor saved time for registering, $t(49)=1.52$, $p=0.135$.

Participants were also asked questions with open-ended response formats. When asked to describe how they currently document their research, the majority used files and folders to organize their research, often consisting of Word files and Dropbox syncing. The responses were similar when asked to describe how they archive their research. Responses were biased toward organization and backup rather than making materials public. For sharing materials with collaborators and non-collaborators, the vast majority of participants use email, Dropbox, and Google Docs/Drive. Again, for non-collaborator sharing, there was little mention of a public presence. In describing how they currently

register their materials, the majority do not engage in this activity. Others use the OSF, are mandated by their IRB, or do so informally.

Participants were asked for open-ended feedback on how the OSF could be made more intuitive for each activity. For documenting research, many participants mentioned making obvious a hierarchical-like folder structure, documenting current features to make the learning process more efficient, and allowing for desktop-syncing with the OSF and/or Dropbox for offline storage. The same issues appeared when asked about project archival. When asked about sharing with collaborators, syncing with Dropbox appeared often, along with feature documentation and email notification. There was less feedback when asked about sharing with non-collaborators and registering materials, although the need for more documentation was mentioned several times.

Participants were asked to describe how the OSF was better or worse than their current solution for each activity. The feedback was overwhelmingly positive about the OSF being used for documenting because of the organization, version control, logging of who did what and when, and centralization. For those that said documenting worse on the OSF, they mentioned the need to be on the internet, interface issues, the lack of assurance that the OSF will always exist, and the need to get collaborators using the system. Similarly, for archiving, the majority of positive comments had to do with OSF as a backup source, version control, logging, and registration (taking snapshots rather than explicitly timestamping an event). Negatives had to do with lack of assurance as to the longevity of the OSF and the ease of services like Dropbox. For sharing with collaborators, participants felt transparency, organization of material history, and

centralization were advantageous, especially as compared to an email-based-workflow. The requirement for collaborators to have an account and the lack of usability as compared to Dropbox were weaknesses. Responses were similar for sharing with non-collaborators, although the positive of it requiring less time to share was mentioned several times. Participants were split as to whether or not the OSF was better or worse for registering materials. The vast majority thought the OSF would be better because there lacked a comparison or it was their primary method for registrations; some saw it as one of the primary features of the OSF. Others argued that registration has little value in the first place, while a several participants were cautious about registrations whether or not they are valuable.

Participants were asked to describe barriers to using the OSF for each of the activities. The most common barriers listed for using it for documenting research were convincing collaborators to adopt/use it, the time costs in learning or switching to it, and security--both for sensitive information as well as longevity. The same issues were raised for using the OSF to archive and sharing with collaborators (although there was more of an emphasis on convincing collaborators to adopt/use the OSF). Responses were similar for sharing with non-collaborators, although several participants emphasized their lack of comfort with sharing their research with non-collaborators. The barriers for using the OSF for registering materials had much more to do with not seeing the value in registering or having to convince collaborators and/or not being in a position to do so.

Survey Discussion

Survey responses suggest that users are not yet convinced that the OSF will save time for their workflow in the short-term, but are more optimistic about its benefits in the long-term. This view appears to translate into some uncertainty about the extent that they will use the OSF for assisting with their workflow.

Given the lack of current tools to support registering materials, the quantitative data is not encouraging in terms of future use of the OSF for registration. Most participants should plan on using the OSF because it is one of the few tools for this task. However, when paired with the qualitative responses, it seems that other issues are at play--likely due to the current culture: strict respect for the confirmatory/exploratory designation and adherence to an accurate account of process is currently undervalued and not incentivized.

It is clear that participants recognize the long-term benefits of integrating the OSF into their current workflow. However, there is room for refinement in terms of providing tools to make activities more efficient in order to incentivize use on the short term. In particular, based on this feedback, high priorities for near-term development are the collaborative editing features of Google Docs, the offline access and automatic local file system synchronization of Dropbox, and the hierarchical folder file organization of the file system and systems like Google Drive.

At the very least, very few users said that they would not use OSF at all in the future, and no user endorsed not using it at all for more than one item. This is further evidence that the users are recognizing value or potential in the OSF.

Chapter 5: Usage

Usage Overview

A prerequisite of a shift towards openness is using technology that facilitates it. As such, an obvious indicator of OSF success is whether scientists use it. A successful system will solve problems or improve the scientist's workflow. If scientists are not convinced that the system is doing that, then they will not use it. The variables documenting usage of the OSF are as follows:

Users. Acceptance of the OSF and its ability to facilitate openness requires users. The number of users, and rate at which users create accounts are the most direct indicators. For this and all of the other usage indicators, a higher number of users and an accelerating rate of growth are key indicators of success.

Projects. The number of projects created, and the rate of growth of projects over time go beyond a count of the user base and reflect actual use. Are the users using the system to document their science?

Public Projects. Making projects public is possibly the most direct metric of whether the OSF is facilitating openness. The number of public projects and the rate at which projects are made public reflect the use of the OSF as a system to share and archive content in a transparent way.

Forks. Forking provides a formal method of using and extending the work of others. The number of forks in the system as well as the rate of forking will reflect adoption of this new form of citation and acceptance as a systematic way to expand on previous science as part of the workflow.

Downloads. Similar to number of forks, number of downloads and the rate of downloads reflects interest and use of a given component and its files. Downloads is a proxy for increased discovery and the value of disclosing information publicly. If materials are not made public, this metric is constrained to the number of occasions that authors receive and respond to requests for research materials from a given study.

Views. Like downloads, the number of views is a simple metric for reflecting interest in a component, likely as an indicator of increased discovery. The number of views and the rate of viewing projects are additional indicators of the effect of making materials available openly.

Registrations. The number of total registrations and the rate of registrations are indicators of capturing the state of a research project at a particular point in time. The most obvious occasion to do this is at the point of initiating data collection. But, there are many other occasions during which users may wish to create a frozen copy.

Contributors Per Project. The mean number of contributors per project and rate of change of that mean are indicators that the OSF is facilitating collaboration. The mean number of contributors listed on a given project should reach the mean number of authors in the traditional publication model. Getting to this point will reflect adoption of the OSF as a collaborative tool. However, it is expected that the number of authors would eventually increase beyond the traditional model as the OSF supports variable contribution models and specificity via components as to who does what.

Usage Results

The variables indicating usage are listed in Table 2. The table includes current usage statistics as compared to estimates based upon usages statistics from four months prior and two models of growth. The linear model of growth is relatively conservative, while the exponential model, based upon Facebook's growth, is optimistic. Numbers between the two were expected.

As expected, number of users, downloads, project registrations, and views fell between the conservative and optimistic predictions. The number of contributors per project only had one estimate, and it was as predicted. The number of projects made public was below even the conservative estimates, but only by a small margin (134 vs. 140). The number of projects and forks were above the optimistic estimates.

Usage Discussion

Site usage overall is showing effective growth, particularly for a beta release. Scientists are creating accounts, viewing site content, downloading files, and registering their projects at an increasing rate. To a greater extent, they are hosting their projects on the site as well as using the fork functionality. The number of projects made public did increase, but not at the rate it had increased previous to this evaluation.

Together this suggests that exposure and adoption of the OSF as a collaborative workspace for organizing scientific materials is going well. It also suggests that greater focus be placed on incentivizing openness or decreasing the fears associated with making projects public. Making materials public will require a cultural shift, and it is expected that this--of all the variables--will be adopted the most slowly. Table 3 shows additional

predictions from one to five years in the future. Given the data seen thus far, it will be a useful benchmark for setting realistic milestones.

Chapter 6: Conclusions and Future Directions

Success of the Open Science Framework can be demonstrated in several ways: Is the OSF facilitating openness? Is that openness reducing the gap between scientific values and scientific practices? Is that openness making science more efficient? Beyond development of the OSF, the goals of this dissertation were to evaluate the current success of the OSF and inform future evaluation of adoption, growth, and success. Evaluation was performed in three ways: semi-structured interviews of four scientists, a usability survey distributed to a wider audience of scientists, and an examination of usage metrics.

Some findings suggest that the goals increasing openness across the entire workflow are not shared universally by the early adopters of the OSF. For example, a relatively common qualitative comment expressed skepticism about the value of registering study designs and analysis plans to affirm confirmatory research designs. The fact that there is diversity in this belief is not surprising, particularly considering that such registration is very non-normative in present research practices. A virtue of the OSF is that it provides the means of doing open practices - such as making materials public or registering analysis plans - but does not require them. Thus, users who do not embrace these practices can still gain value from the OSF for management of their own research workflow. And, if the norms change - such as with the OSF features that incentivize openness - then researchers' behavior can be updated easily by taking advantage of the new features.

This suggests that one primary implication for the OSF is to assure that it meets researchers' needs for their present practices and values, and simultaneously enhance the incentives for openness to facilitate the evolution of values and practices. Simultaneously, there are improvements and innovations that must occur in the technology in order to meet this goal and the broader goals for the OSF. The qualitative and quantitative results provide clear suggestions for improving the technology and documentation. The remainder of this chapter outlines the next steps for the project based on the results.

Technology Direction

Organization.

While there were many positive comments about the current organizational capabilities of the OSF, there is a lot to improve in the user interface. Users desire a stronger representation of the familiar hierarchical folder structure. A more intuitive dashboard might include showing more critical information such as files in folders, rather than emphasizing logging information as strongly. Providing the ability to move components and folders around like one does on a personal computer may provide a more intuitive interface for interacting within and between projects.

Desktop/Offline Syncing

Desire for Dropbox and Dropbox-like services came up over and over in the interviews and survey responses. Such services would allow users to work with files on their local file system, using their desired tools, and then upload automatically to centralized server that allows for easy sharing. The OSF will offer this functionality in

the form of a cross-platform compatible desktop application that uses operating-system level hooks to monitor changes to selected folders and files. The OSF will also integrate with services like Dropbox via API calls to manage synchronization between the OSF and the storage service. This feature should decrease the learning curve and improve integration with existing workflows dramatically.

Collaborative editing

Many surveyed use Google Docs in their existing workflow. It provides many of the features of local word processing and spreadsheet software with the added capability of easy sharing and collaboration. The OSF may pursue two avenues for the integration of collaborative editing. First, OSF will allow Google Doc “file types” that would be monitored with OSF via API calls. And, second, we could implement a version of collaborative editing tools for the OSF. There is considerable complexity in implementing collaborative editing because of the need for operational transformation algorithms and asynchronous server technology. Nonetheless, a prototype has been implemented and development will continue in the future. The potential power of a large-scale open-source effort could make such a system competitive with Google’s services by allowing the innovation that can occur in open environments..

Word is often the least common denominator when writing in a collaborative environment: if one member does not know how to use documentation tools like LaTeX-- a documentation markup language with particular benefits in the sciences--everyone must use Word. That is not to say that Word lacks value: its low barrier to entry and track changes functionality are highly attractive features, but other formats are more conducive

to scientific writing. A project is currently being developed to parse the Docx file format, including track changes markup. The package has been written in such a way as to allow users the ability to easily write parsers to and from their desired format. Users can also extend the parsers provided, including HTML, LaTeX, reStructuredText, and Markdown. Support will be made available to translate to and from the Word document format.

When a user uploads a Word document to the OSF, links will be available to download it in a variety of formats. The same goes for other formats that can be parsed to Word.

Version Control

While not as pervasive in the scientist's workflow, access to the Git repositories via Git tools could be a useful feature. Also, implementing this would increase the reputation of the OSF in the developer community, an important consideration for broadening the contributors to the software. There are some technical complications related to user authentications and the storage of large, binary data files for implementing this feature. However, the developer community has started to address these issues. It is conceivable that these features could be integrated into the version 1.0 release.

Commenting and Discussion

The ability to facilitate and document conversation about projects, components, and materials was another commonly requested feature. Integrated public discourse is a major advantage of open access/open science and one that will be supported in the version 1.0 release of the OSF. The current plan is to allow meta-data to be attached to any object in the system (e.g., projects, components, logs, and files), and one type of meta-data would be a comment thread. These threads could then be viewed like

comments typically seen on blogs or aggregated to appear to users as forums or archived mailing lists. With a strong commenting system, a longer-term development possibility is full support of journal submission, review, publication, and post-publication peer review through the OSF.

Notification

With the introduction of commenting to the OSF, the need for a notification system will be even more relevant than it is in the current state of the system. This would include the ability to be notified via email when an event occurs related to a user's projects, components, or materials. For example, users of Google Docs can assess very rapidly when shared documents have been edited by (a) filenames being presented in bold in the dashboard, (b) an option to sort by date of last editing, and (c) email notifications. All of these can be integrated easily into the OSF. Further, for conversation-based notifications, we will implement a reply-by-email feature such that users with proper authentication can interact with the system by email, thus reducing the learning curve and integrating well with email-heavy workflows.

Development Paradigm

Further effort will be placed on pursuing the development paradigm as described in Chapter 2. The open source development community has already started to get involved with the OSF. For example, at PyCon 2013, eleven developers donated as much as four full days in a sprint for the OSF. Sprints are programming marathons dedicated to a specific project. Making the OSF code base accessible will encourage investment by developers of all levels. The continued use of the modular paradigm is expected to

further increase open source development because developers will be able to identify problems that they can solve, and then see the results of their effort with a low barrier to entry.

Templating Engine

The templating system allows users to create registration or other forms for data collection. A new version of the templating language has been written that will provide flexibility for registration templates, and, eventually, free-form meta-data attachments. Eventually, we could create a user interface on top of this engine in order to provide users an intuitive, yet flexible method of creating dynamic forms. Such forms could be used for documenting projects, registration, or even behavioral research.

Automatic CV Generation

In order to incentivize use of OSF-based metrics and components of the contribution model, automatic CV generation is a feature that will likely follow the version 1 release. It will allow users to automatically generate a customized CV specifically designed to emphasize the contribution of the users whether it was for data collection, data analysis, writing the report, or any other component of the research process. Presenting the number of forks next to articles or graphs of downloads for datasets may facilitate recognition and adoption of alternative metrics to complement the traditional considerations of # of articles, impact factor of publishing journals, and individual article citation counts.

Integration with Existing Technologies

API development will enable connection and integration with many other existing services. For example, through API calls, users could work with analytic platforms like R and IPython, data visualization tools, data collection engines, and data repositories through the OSF interface. Similarly, those services could use OSF in their back-end to improve their own service. Further, IRBs and funding bodies could integrate their submission systems with OSF so that the researcher can use the same base of materials for applying for grants, submitting to IRB, and working with collaborators on the final study designs. Giving users access to a broad range of services that they already use will decrease integration/adoption times and also reduce development efforts. If another organization does something well, we would rather collaborate than complete--connect rather than reimplement.

Exploration and Search

The version 1 release will likely contain a variety of features related to project exploration and search in the OSF. We will enhance the search engine and create new methods of viewing projects. This will include project listing by tag, order by views, forks, and other metrics, and network visualizations of the underlying collaborator and forking graphs. A recommendation engine will automatically recommend projects and users that contain features similar to the users' own projects or profile. This may increase cross-disciplinary discovery and collaboration.

Demonstration and Documentation

Increased exploration and discovery features will allow the community to demonstrate the value of openness simply by doing their science. For those not wholly

convinced by the merits of openness, we expect more rapid adoption if peers and respected members of the scientists' community actively demonstrate the positive aspects and lack of negative outcomes related to conducting science openly. For example, we will create 1-3 minute video clips showing how a particular feature can be used effectively. Also, example projects illustrating best practices will help users imagine uses that they may not have considered. Some of this demonstration will occur via education--specifically documentation--and we can make those demonstrations more vivid via technology.

Standard Documentation

As mentioned in the survey responses, there is a substantial lack of documentation of current features. Version 1.0 will include a help system, tutorials, frequently asked questions, and user guides on how to use the OSF and integrate it with one's own workflow.

Development of Best Practices

Another form of documentation that will be developed is a set of best practices on using the OSF to conduct science openly. Some users may have a misunderstanding what is involved in openness--overestimating the amount of time, effort, and resources that must be devoted to it. This would be an easy way to make clear what is involved and how the OSF greatly reduces that workload.

Templates

Component templates were requested in both the interviews and surveys. Like best practices, this would make clear what might be expected in an open workflow as

well as provide a starting point for OSF usage. Again, if misunderstandings are an issue in adoption of the system, templates may be an easy way to provide clarity to users.

Interactivity

For each of these documentation extensions, we can use wizard-like systems to guide users in initial involvement with the system so as to reduce the actual or perceived time and effort required to adopt an open workflow. For example, a user could launch a “New Project wizard.” An overlay with animations would appear and guide users to click on buttons or enter text and provide advice at important decision points.

Future Evaluation and Metascience

Evaluation of the OSF will be an ongoing process throughout its development and increased usage stemming from the evaluations presented in this document. For example, a rigorous examination of the measures could help create a more efficient evaluation mechanism. There are also interesting additions to the evaluations that might offer insight on how science is conducted. For example, we could examine personality with regards to openness. What are the personality characteristics of those who align themselves closely with open workflows? Could openness be a personality trait? Does the trait change over time? Is it more accurately deemed a state?

Once there is an established regular user base, increased metascience research possibilities unfold. For example, a sample of OSF users will be contacted asking for materials from two different projects—one that is on the OSF and one that is not. The amount of time needed to respond to the request will be examined. The expectation is that material hosted on the OSF will require significantly less time to share than material

that is not hosted. This would indicate an increase in efficiency based on its archiving and sharing features. Over time, a larger number of projects will be amassed on the system—some of which are published—and comparisons can be made between those that used the OSF and those that did not. Evaluation will be an ongoing feature of OSF development into the foreseeable future.

The database of public projects and collaborators and their connections to one another will be made available for public exploration. Another OSF-related project is underway aggregating data from a variety of sources in order to create a similar dataset for traditional publications and their authors, but with citations rather than forks creating connections. These two datasets will be able to be compared, asking questions such as: Is multidisciplinary collaboration more frequent in OSF data? The hypothesis is that openness creates increase collaboration possibilities by increasing the likelihood of overlapping interests and distributing credit in order to properly incentivize the collaboration.

One recently funded project led by Jonathan Schooler at UCSB will randomly assign researchers to either register their study designs and analysis plans with OSF or not for a prospective replication project. Researchers would conduct two studies in their lab, one registered one not, and then the results of those studies would be replicated by one or more other labs. It is expected hypothesized that registering analysis plans in advance will (a) decrease the frequency of reporting positive results, and (b) increase the likelihood that the observed positive results will replicate. These would indicate material benefits of using a system like the OSF for managing the research workflow.

Limitations

Science is an institution, and like most institutions, change will occur slowly, even under the best of conditions. As demonstrated by the interviews and survey results, while there are commonalities in workflow, there also exists considerable variability. Developing technology robust enough to account for that level of variability is challenging. While the heavy focus on API development and flexible, robust infrastructure components is advantageous, we will never be able to satisfy every user's desired workflow.

Security issues were brought up a number of times in the survey responses. Is my private data secure? It is an important issue, but there are significant complexities in creating systems that are simultaneously secure, accessible by multiple users, and recoverable. Also, some level of security requirements will likely not be met. For example, we will have little flexibility to address requirements about the physical security at the data storage site. While we will encourage API collaborations with services that provide advanced security guarantees, we may not be able to fully appease users in how tightly those services can be integrated with the OSF. Security requirements for some data are simply too high for using any connection to a networked, collaborative system. Such data will continue to be stored on idiosyncratic, private, highly secure systems.

Questions about the longevity of the project were also raised in the survey responses. How do I know that this system will last? While we have specific strategies for creating sustainability, we will never be able to provide absolute guarantees. And while there are assurances that come with open source software and accessible public

data, it is difficult to promise end-of-production failover plans for private data without compromising that privacy. Regardless, we will need to develop a strong sustainability plan that includes a relationship with another society willing to take over support of the data should we cease to exist. An obvious target for this support is <http://archive.org/>, one of the best guarantees of persistence for public-good data in existence.

As the amount of data increases on the OSF, so will the problem of dealing with differentiating signal from noise. Careful attention will need to be paid towards creating search, sort, and filter functionality that allows users to efficiently find information. Furthermore, the systems on the backend that aggregate across meta-data will need to be created in such a way as to make use of these parameters to support a flexible user interface as well as allow others to create innovative search functionality atop the OSF.

A voiced concern about a move toward widespread openness is that it could have unintended consequences on the things that science does well presently—pursue new ideas, innovate, and proceed down an intriguing road of inquiry even if there is only a bit of suggestive evidence. Might openness stifle such things because of an increased concern for failing publicly? Openness actually helps alleviate some of this concern: when a workflow is open, the community can recognize problems at an earlier stage of the process where errors have presumably less of an impact than in the current system where error-detection often occurs post-publication, if it is possible to occur at all. In the longer term, openness may create a cultural understanding, if not appreciation, for failure and the important role it plays in learning and innovation. And the definition of failure will likely change; the issues tackled in the sciences are complex—the unexpected is

expected. With that said, cultural change can be slow, and the initial concern is valid. Short term incentivizes, like self-nominated “most surprising finding” awards, or encouraged use of registration templates that specify tests of multiple competing hypotheses, such that even failure is interesting, may help motivate this change in the perception of failure.

Conclusion

Based on the favorable feedback received in the interviews and the significant growth trends extracted from usage statistics, I am optimistic that these are early indicators of a successful future. Progress in the last eighteen months has been encouraging and prospects for the next eighteen months are similarly exciting. With the OSF as our flagship project, Brian Nosek and I were awarded a starter grant of \$5,250,000 by the Laura and John Arnold Foundation in order to found the Center for Open Science (COS; Center for Open Science, 2013). The COS has three main goals: (1) continue development of the OSF and other technologies in order to facilitate openness in the sciences, (2) engage the open source/open science community in an effort to bring the two communities closer together, and (3) conduct meta-science, or science of science, research. Other positive signs for the future of the OSF include an anonymous donation of \$400,000 to the COS, another open science service hiring a developer dedicated to API development in order to connect that service to the OSF, a number of promising collaborations with journals, societies, and technical organizations starting to take shape, and—as of March 8, 2013—the first published article using OSF’s documentation, archival, and registration systems (LeBel & Campbell, In Press). Every revolution begins

with a first step. In this dissertation, I have described the goals and first step for a possible revolution in openness. Hopefully, in time, what may feel like revolution will simply be thought of as evolution--progress toward a higher state of science because the features of openness are selected for being more successful in attaining scientific accuracy and efficiency.

References

- American Psychological Association (2011). *Publication Manual of the American Psychological Association (Sixth ed.)* Washington, DC: Author.
- Bacon, R. (1267/1859). *The Chronicles and Memorials of Great Britain and Ireland during the Middle Ages*. Longman, Green, Longman and Roberts. Retrieved from: <http://books.google.com/books?id=wMUKAAAAYAAJ>
- Bakan, D. (1966). The test of significance in psychological research. *Psychological Bulletin*, 66, 423-437. doi: 10.1037/h0020412
- Bakker, M., & Wicherts, J. M. (2011). The (mis)reporting of statistical results in psychology journals. *Behavior Research*, 43(3), 666-678. doi: 10.3758/s13428-011-0089-5
- Ball, P. (2005). Index aims for fair ranking of scientists. *Nature*, 436, 900. doi:10.1038/436900a
- Bastardi, A., Uhlmann, E. L., & Ross, L. (2011). Wishful thinking: Beliefs, desire, and the motivated evaluation of scientific evidence. *Psychological Science*, 22, 731-732. doi: 10.1177/0956797611406447
- Bem, D. J. (2003). Writing the empirical journal article. In J. M. Darley, M. P. Zanna, & H. L. Roediger III (Eds.), *The Compleat Academic: A Career Guide* (171–201). Washington, DC: American Psychological Association.
- Bersoff, D. M. (1999). Why good people sometimes do bad things: Motivated reasoning and unethical behavior. *Personality and Social Psychology Bulletin*, 25, 28-39. doi:10.1038/436900a

- Boiney, L. G., Kennedy, J., & Nye, P. (1997). Instrumental bias in motivated reasoning: More when more is needed. *Organizational Behavior and Human Decision Processes*, 72, 1-24. doi: 10.1006/obhd.1997.2729
- Bonaccorsi, A., & Rossi, C. (2003). Why Open Source software can succeed. *Research Policy*, 32, 1243–1258. doi:10.1016/S0048-7333(03)00051-9
- Braude, S. E. (1979). ESP and Psychokinesis. A Philosophical Examination. Philadelphia: Temple University Press.
- Buckheit, J. B., & Donoho, D. L. (1995). WaveLab and Reproducible Research. Department of Statistics, Stanford University, Technical Report 474. Retrieved from: http://www-stat.stanford.edu/~wavelab/Wavelab_850/wavelab.pdf.
- Center for Open Science. (2013). Center for Open Science to provide revolutionary approach to scientific communication. Retrieved from: <http://centerforopenscience.org/pr/2013-03-05/>.
- Christensen-Szalanski, J. J. J., & Willham, C. F. (1991). The hindsight bias: A meta-analysis. *Organizational Behavior and Human Decision Processes*, 48, 147-168. doi: 10.1016/0749-5978(91)90010-Q,
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49, 997-1003. doi: 10.1037/0003-066X.49.12.997
- Crosas, M. 2011. The Dataverse Network: An Open-source Application for Sharing, Discovering and Preserving Data. *D-Lib Magazine*, 17(1/2), doi:10.1045/january2011-crosas

Collins, H. M. (1985). *Changing order: Replication and Induction in Scientific Practice*.

London: Sage.

Ditto, P. H., & Lopez, D. F. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, *63*, 568-584. doi:10.1037/0022-3514.63.4.568

Eclipse Foundation. (2012). Eclipse Community Survey 2012. Retrieved April 9, 2013 from: http://eclipse.org/org/community_survey/Survey_Final_Results_2012.xls

Fanelli, D. (2009). How many scientists fabricate and falsify research? A systematic review and meta-analysis of survey data. *PLoS ONE*, *4*(5), 1-11.

doi:10.1371/journal.pone.0005738

Fanelli, D. (2010a). "Positive" results increase down the hierarchy of the sciences. *PLoS ONE*, *5*(4), e10068. doi:10.1371/journal.pone.0010068

Fanelli, D. (2010b). Do pressures to publish increase scientists' bias? An Empirical Support from US States Data. *PLoS ONE*, *5*(4), e10271.

doi:10.1371/journal.pone.0010271

Fanelli, D. (2012). Negative results are disappearing from most disciplines and countries.

Scientometrics, *90*, 891-904. doi: 10.1007/s11192-011-0494-7

Fischhoff, B. (1977). Perceived informativeness of facts. *Journal of Experimental*

Psychology: Human Perception and Performance, *3*, 349-358. doi: 10.1037/0096-1523.3.2.349

- Fischhoff, B., & Beyth, R. (1975). "I knew it would happen" Remembered probabilities of once-future things. *Organizational Behaviour and Human Performance*, *13*, 1-16. doi: 10.1016/0030-5073(75)90002-1
- Frick, R. W. (1996). The appropriate use of null hypothesis testing. *Psychological Methods*, *1*, 379-390. doi: 10.1037/1082-989X.1.4.379
- Gorman, M. E. (1992). *Simulating Science: Heuristics, Mental Models, and Technoscientific Thinking*. Bloomington, IN: Indiana University Press.
- Gottfredson, S. D. (1978). Evaluating psychological research reports: Dimensions, reliability, and correlates of quality judgments. *American Psychologist*, *33*, 920-934. doi: 10.1037/0003-066X.33.10.920
- Greenwald, A. G. (1975). Consequences of prejudice against the null hypothesis. *Psychological Bulletin*, *82*, 1-20. doi: 10.1037/h0076157
- Hackett, B. (2005). Essential tensions: Identity, control, and risk in research. *Social Studies of Science*, *35*(5), 787-826. doi:10.1177/0306312705056045.
- Hargens, L. L. (1988). Scholarly consensus and journal rejection rates. *American Sociological Review*, *53*, 139-151. Retrieved from: <http://www.jstor.org/stable/2095739>
- Haslam, N., Ban, L., Kaufmann, L., Loughnan, S., Peters, K., Whelan, J., & Wilson, S. (2008). What makes an article influential? Predicting impact in social and personality psychology. *Scientometrics*, *76*, 169-185. doi: 10.1007/s11192-007-1892-8

- Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, 2, e124. doi:10.1371/journal.pmed.0020124
- Jasny, B. R., Chin, G., Chong, L., & Vignieri, S. (2011). Again, and again, and again... *Science*, 334, 1225. doi: 10.1126/science.334.6060.1225
- John, L., Loewenstein, G., & Prelec, D. (2012). Measuring the prevalence of questionable research practices with incentives for truth-telling. *Psychological Science*, 23, 524-532. doi: 10.1177/0956797611430953
- Kerr, N. L. (1998). HARKing: Hypothesizing after the results are known. *Personality and Social Psychology Review*, 2, 196–217. doi: 10.1207/s15327957pspr0203_4
- Kuhn, T.S. (1962). *The Structure of Scientific Revolutions*. Chicago, IL: University of Chicago Press.
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108, 480-498. doi:10.1037/0033-2909.108.3.480
- Lakatos, I. (1978). *The Methodology of Scientific Research Programmes: Philosophical Papers*. Cambridge: Cambridge University Press.
- Lander, E. S, Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., et al. (2001). Initial sequencing and analysis of the human genome. *Nature*, 409, 860-921. doi:10.1038/35057062
- LeBel, E. P., & Campbell, L. (In Press). Heightened sensitivity to temperature cues in highly anxiously attached individuals: Real or elusive phenomenon? *Psychological Science*.

- Leisch, F. (2002) Sweave: Dynamic generation of statistical reports using literate data analysis. In Härdle, W. and Rönz, B. (Eds.), *Compstat 2002 - Proceedings in Computational Statistics* (575-580). Heidelberg: Physica Verlag.
- Lerner, J., & Tetlock, P. E. (1999). Accounting for the effects of accountability. *Psychological Bulletin*, *125*, 255-275. doi: 10.1037/0033-2909.125.2.255
- Lerner, J., & Tirole, J. (2001). The open source movement: Key research questions. *European Economic Review*, *45*(4-6), 819–826. doi: 10.1016/S0014-2921(01)00124-6
- Lord, C. G., Ross, L., & Lepper, M. R. (1979). Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. *Journal of Personality and Social Psychology*, *37*, 2098-2109. doi:10.1037/0022-3514.37.11.2098
- Madden, C. S., Easley, R. W., & Dunn, M. G. (1995). How journal editors view replication research. *Journal of Advertising*, *24*, 78–87. Retrieved from: <http://www.jstor.org/stable/4188990>
- Mahoney, M. J. (1985). Open exchange and epistemic process. *American Psychologist*, *40*, 29–39. doi: 10.1037/0003-066X.40.1.29
- Martin, B. (1992). Scientific fraud and the power structure of science. *Prometheus*, *10*(1), 83–98. doi:10.1080/08109029208629515
- Martinson, B. C., Anderson, M. S., & Devries, R. (2005). Scientists behaving badly. *Nature*, *435*, 737-738. doi:10.1038/435737a

- Miller, A. C., & Serzan, S. L. (1984). Criteria for identifying a refereed journal. *The Journal of Higher Education*, 55, 673–699. Retrieved from: <http://www.jstor.org/stable/1981508>
- National Institutes of Health (2003). Final NIH Statement on Sharing Research Data. In *NIH Data Sharing Policy*. Retrieved from <http://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>.
- National Science Foundation (2011). Dissemination and Sharing of Research Results. In *Grant Proposal Guide*. Retrieved from http://www.nsf.gov/pubs/policydocs/pappguide/nsf11001/aag_6.jsp
- Neuliep, J. W., & Crandall, R. (1990). Editorial bias against replication research. *Journal of Social Behavior and Personality*, 5, 85–90. doi: 1991-00044-001
- Neuliep, J. W., & Crandall, R. (1993). Reviewer bias against replication research. *Journal of Social Behavior and Personality*, 8, 21–29. doi: 1994-15187-001
- Newton, I. (1676, February 5). [Letter from Isaac Newton to Robert Hooke] In Maury, J.P. (1992). *Newton: Understanding the Cosmos*, Thames & Hudson.
- Nickerson, R. S. (2000). Null hypothesis significance testing: A review of an old and continuing controversy. *Psychological Methods*, 5, 241-301. doi: 10.1037/1082-989X.5.2.241
- Nosek, B. A. & Bar-Anan, Y. (2012). Scientific Utopia: I. Opening scientific communication. *Psychological Inquiry*, 23(3), 217-243. doi:10.1080/1047840X.2012.692215

- Nosek, B. A., Graham, J., Lindner, N. M., Kesebir, S., Hawkins, C. B., Hahn, C., Schmidt, K., Motyl, M., Joy-Gaba, J., Frazier, R., & Tenney, E. R. (2010). Cumulative and career-stage citation impact of social-personality programs and their members. *Personality and Social Psychology Bulletin*, *36*, 1283-1300. doi: 10.1177/0146167210378111
- Oakes, M. (1986). *Statistical inference: A commentary for the social and behavioral sciences*. New York: Wiley.
- Ostriker, J. P., Holland, P. W., Kuh, C. V., & Voytuk, J. A. (2009). *A Guide to the Methodology of the National Research Council Assessment of Doctorate Programs*. Washington, DC: National Academic Press.
- Peters, D. P., & Ceci, S. J. (1982). Peer-review practices of psychological journals: The fate of published articles, submitted again. *Behavioral and Brain Sciences*, *5*, 187–255. doi: 10.1017/S0140525X00011183
- Phillips, N. (2007). Citation counts, prestige measurement, and graduate training in social psychology. *Dialogue*, *22*(2), 24-26. Retrieved from: <http://94.23.146.173/ficheros/27cce4ffdec57b71e642ea8538dcd635.pdf>
- Piwowar, H. A., Day, R. S., & Fridsma, D. B. (2007). Sharing detailed research data is associated with increased citation rate. *PLoS ONE*, *2*, e308. doi:10.1371/journal.pone.0000308
- Popper, K. (1934/1992). *The Logic of Scientific Discovery*. New York: Routledge.

- Pyszczynski, T., & Greenberg, J. (1987). Perspectives on social inference: A biased hypothesis-testing model. In L. Berkowitz (Ed.), *Advances in Experimental Social Psychology*, 20, 297-340. doi: 10.1016/S0065-2601(08)60417-7
- Python Software Foundation. (2013). About Python. Retrieved April 9, 2013 from: <http://python.org/about/>.
- Reyna, V.F., & Brainerd, C.J. (1995). Fuzzy trace theory: An interim synthesis. *Learning and Individual Differences*, 7, 1–75. doi: 10.1016/1041-6080(95)90031-4,
- Rosenthal, R. (1979). The file drawer problem and tolerance for null results. *Psychological Bulletin*, 86, 638-641. doi: 10.1037/0033-2909.86.3.638
- Rosenthal, R. (1991). Replication in behavioral research. In J. W. Neuliep (Ed.), *Replication Research in the Social Sciences* (1–39). Newbury Park: Sage.
- Rosnow, R. L., & Rosenthal, R. (2009). Effect sizes: Why, when, and how to use them. *Journal of Psychology*, 217, 6-14. doi: 10.1027/0044-3409.217.1.6
- Rozeboom, W. W. (1960). The fallacy of the null-hypothesis significance test. *Psychological Bulletin*, 57, 416-428. doi: 10.1037/h0042040
- Sabra, A. I., trans. (1021/1989), *The Optics of Ibn al-Haytham. Books I–II–III: On Direct Vision*. English Translation and Commentary. London: The Warburg Institute, University of London.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13, 90-100. doi: 10.1037/a0015108

Schroter, S., Black, N., Evans, S., Godlee, F., Osorio, L., & Smith, R. (2008). What errors do peer reviewers detect, and does training improve their ability to detect them? *Journal of the Royal Society of Medicine*, *101*, 507-514.

doi: 10.1258/jrsm.2008.080062

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-positive psychology: Undisclosed flexibility in data collection and analysis allows presenting anything as significant. *Psychological Science*, *22*, 1359-1366.

doi:10.1177/0956797611417632

Sovacool, B. K. (2008). Exploring scientific misconduct: Isolated individuals, impure institutions, or an inevitable idiom of modern science? *Journal of Bioethical Inquiry*, *5*, 271-282. doi: 10.1007/s11673-008-9113-6

Steen, R.G. (2011). Misinformation in the medical literature: What role do error and fraud play? *Journal of Medical Ethics*, *37*, 498-503.

doi:10.1136/jme.2010.041830

Sterling, T. D. (1959). Publication decisions and their possible effects on inferences drawn from tests of significance—or vice versa. *Journal of the American Statistical Association*, *54*, 30-34. doi: 10.1080/01621459.1959.10501497

Sterling, T. D., Rosenbaum, W. L., & Weinkam, J. J. (1995). Publication decisions revisited: The effect of the outcome of statistical tests on the decision to publish and vice versa. *The American Statistician*, *49*, 108-112. doi:

10.1080/00031305.1995.10476125

- Trope, Y., & Bassok, M. (1982). Confirmatory and diagnosing strategies in social information gathering. *Journal of Personality and Social Psychology*, *43*, 22-34. doi:10.1037/0022-3514.43.1.22
- Wagenmakers, E. J., Wetzels, R., Borsboom, D., & Van der Maas, H. (2011). Why psychologists must change the way they analyze their data: The case of psi. *Journal of Personality and Social Psychology*, *100*, 426–432. doi: 10.1037/a0022790
- Wicherts, J. M., Bakker, M., & Molenaar, D. (2011). Willingness to share research data is related to the strength of the evidence and the quality of reporting of statistical results. *PLoS One*, *6*, e26828. doi:10.1371/journal.pone.0026828
- Wicherts, J. M., Borsboom, D., Kats, J., & Molenaar, D. (2006). The poor availability of psychological research data for reanalysis. *American Psychologist*, *61*, 726–728. doi:10.1037/0003-066X.61.7.726
- Wong, P. T. (1981). Implicit editorial policies and the integrity of psychology as an empirical science. *American Psychologist*, *36*, 690–691. doi: 10.1037/0003-066X.36.6.690
- Ye, Y., & Kishida, K. (2003, May). *Toward an Understanding of the Motivation of Open Source Software Developers*. Presentation at the Proceedings of the International Conference on Software Engineering, Portland, OR.
- Young, N. S., Ioannidis, J. P. A., & Al-Ubaydli, O. (2008). Why current publication practices may distort science. *PLoS Medicine*, *5*, 1418–1422. doi:10.1371/journal.pmed.0050201

Zuckerman, H., & Merton, R. K. (1971). Patterns of evaluation in science:

Institutionalization, structure and functions of the referee system. *Minerva*, 9, 66-100.

Appendix

Usability Questionnaire

The usability questionnaire addresses four topics of interest to evaluating and improving the OSF: [1] What are researchers current solutions for managing their workflow and how do they compare with doing it with the OSF; [2] How easy is it to use the OSF; [3] How likely is it that the OSF will be used; and [4] what are the resource costs to integrating the OSF into the daily research workflow. I will examine each of these questions separately for five primary features of the OSF: documenting, archiving, sharing with collaborators, sharing with non-collaborators, and registration.

Each item is asked 5 times with “X” being replaced with: [1] document your research materials, [2] archive your research materials, [3] share research materials with collaborators, [4] share research materials with non-collaborators, and [5] register study designs and/or analysis plans in advance of data collection or analysis.

Survey

Instructions

Thank you for agreeing to complete this survey about your research practices. The questions address how you manage your research workflow, particularly your research materials, and your experience with using the Open Science Framework (OSF). Research materials refers to your surveys, measures, experimental designs, IRB proposals, analysis scripts, data, reports, and any other components of a research project. Your direct and honest answers will be very helpful for assessing variation in research workflows and how the OSF compares to your other practices.

Please answer your questions based on your current impression of OSF, even if you have not used it much yet. Your impressions will be valuable for our evaluation of its intuitiveness, ease of use, and perceived potential.

You will receive five pages of the same questions, each time about a different part of your workflow:

- documenting your research
- archiving you research materials
- sharing your materials with collaborators
- sharing your materials with non-collaborators
- registering your hypotheses, designs, or analysis plans

Answers can be concise. Thank you!

Repeated Questions

Current Solution

- a Describe how you currently X while a project is active and after it is complete or becomes inactive for other reasons. [open-ended response]
- b How does your current solution compare to using the OSF to X? [My current solution is: Much Better, Better, Slightly Better, No Difference, Slightly Worse, Worse, Much Worse]
- c Describe how using the OSF to X is better or worse than your current solution. [open-ended response]

Ease of Use

- a Can you X the way you'd like with the OSF? [Yes, No, Don't know]

- b Describe how using the OSF to X could be made more intuitive. [open-ended response]
- c How easy is it to X on the OSF? [Impossible, Very Difficult, Difficult, Pretty Difficult, Pretty Easy, Easy, Very Easy, Trivial]

Likelihood of Use

- a How likely are you to use the OSF to perform X? [Not at All Likely, Slightly Likely, Moderately Likely, Very Likely]
- b Will you use the OSF in the future? [Yes, No, Not sure]
- c How likely are you to use the OSF to X (select all that apply): [I will always use it; I haven't tried it, but I plan to; I will use it when my collaborators use it; I will use it when it is easier than my current solution; I will use it when it is faster than my current solution; I will use it when I'm mandated to use it; I will use it if others demonstrate its value in the long term; I will not use it at all]
- d Other than what you have answered above, what barriers are there to using the OSF to X? [open-ended response]

Resource costs

- a Compared to your current solution, how much time does the OSF add or save you to X in the short-term (i.e., the initial effort required during the standard workflow of the project)? [OSF adds much more time, OSF adds time, OSF adds slightly more time, No difference, OSF saves slightly more time, OSF saves time, OSF saves much more time]

- b** Compared to your current solution, how much time does the OSF add or save you to X in the long term (i.e., across the lifespan of interest in the project)? [OSF adds much more time, OSF adds time, OSF adds slightly more time, No difference, OSF saves slightly more time, OSF saves time, OSF saves much more time]

Non-repeated

- a** Enter your email address if you are open to follow-up regarding your survey responses.
- b** Any other comments about the OSF that the prior items did not assess?
[open-ended item]

Table 1
Percentage endorsement for each of the items (rows) across activity types (columns)

Item	Document	Archive	Share Collab.	Share Non.	Register
I will always use it	10%	9%	9%	6%	16%
I haven't tried it, but I plan to	40%	38%	39%	34%	32%
I will use it when my collaborators use it	29%	30%	35%	16%	23%
I will use it when it is easier than my current solution	40%	34%	31%	21%	12%
I will use it when it is faster than my current solution	25%	21%	19%	14%	10%
I will use it when I'm mandated to use it	16%	14%	12%	10%	16%
I will use it if others demonstrate its value in the long term	30%	26%	25%	17%	22%
I will not use it at all	0%	0%	1%	4%	3%

Note: Share Collab. = share with collaborators, Share Non. = share with non-collaborators

Table 2

Actual usage as compared to predicted usage

Attribute	-4 Months Actual	Predicted Conservative	Predicted Optimistic	Current Actual
Users	342	595	1034	746
Projects	172	445	520	508
Public Projects	60	140	181	134
Forks	3	12	20	55
Downloads	1256	2149	3798	2397
Views	50065	97192	151406	133616
Registrations	3	26	44	39
Contributors Per Project	2	2		2

Note. The conservative prediction assumes linear growth based on our rate of growth since the beginning of open beta. The optimistic prediction assumes the exponential growth that Facebook demonstrated in its initial years of operation, estimated with a nonlinear regression model as $2.19 * Usage * \exp (.84 * Year)$. No growth information was available for Forks or Registrations, so proportional predicted rate of growth of users was applied to the baseline plus an estimated constant.

Table 3

Predictions five years out

Attribute	-4 Months	Current Actual	Year 1	Year 2	Year 3	Year 4	Year 5
Users [con]	342	746	1113	1883	2654	3424	4195
Users [opt]			1755	4105	9605	22473	52579
Projects [con]	172	508	1003	1835	2666	3498	4329
Projects [opt]			833	2065	4831	11302	26443
Public Projects [con]	60	134	303	547	790	1033	1277
Public Projects [opt]			308	720	1685	3943	9224
Forks [con]	3	55	23	39	54	70	85
Forks [opt]			36	84	197	460	1076
Downloads [con]	1256	2397	3973	6690	9408	12125	14842
Downloads [opt]			6444	15077	35276	82533	193099
Views [con]	50065	133616	193409	336752	480096	623439	766783
Views [opt]			256876	600999	1406125	3289835	7697952
Registrations [con]	3	39	49	83	116	150	184
Registrations [opt]			77	180	421	986	2306
Contributors Per Project	2	2	3	4	5	5	5

Note. Attribute: [con] = conservative estimate; [opt] = optimistic estimate. Conservative estimates indicate linear growth based on our rate of growth since the beginning of open beta. Optimistic estimates are based on a rate equal to the exponential growth rate that Facebook has demonstrated, estimated with a nonlinear regression model as $2.19 * \text{Usage}$

* $\exp (.84 * \text{Year})$. These predictions also assume that the OSF will not reach the point of exponential decay towards an upper threshold within five years. No growth information was available for Forks or Registrations, so proportional predicted rate of growth of Users was applied to the baseline plus an estimated constant.