2021

The Strategic Role Of Data In Proposal Competitiveness And Grant Reporting

Tania N. Johnson
Swarthmore College, tjohnso2@swarthmore.edu

The strategic role of data in proposal competitiveness and grant reporting

Tania N. Johnson, MA, CRA
Swarthmore College

Abstract
This reflective inquiry/case study considers the role that research administrators—especially at predominantly undergraduate institutions—can play as strategic partners helping institutions respond to data requests for grant proposals and reports. Funder mandates for greater quantitative validation of program need in the pre-award stage and program success in the post-award stage have increased significantly. The inability to quickly and easily respond to these requests is a systemic problem for grant officers at many predominantly undergraduate institutions. Addressing these challenges—ideally at the beginning of a project—will become a critical factor in grant competitiveness as data-driven decision making by proposal reviewers and program officers becomes institutionalized in an increasingly challenging funding climate. The author discusses her role in a major collaborative effort to plan and galvanize resources to enter and analyze historical data for an assessment of a longstanding undergraduate summer research program.

Keywords: *grants, data, competitiveness, organizational change, strategic*

Introduction

Data are critical components of strategic proposal development and grant impact reporting, both of which responsibilities are essential to positioning sponsored research offices and the faculty they serve for success. This is particularly true at predominantly undergraduate institutions (PUIs), which often have to counter assumptions that faculty at such institutions typically do not make major scientific or research contributions (Rovnyak & Shields, 2017).

Swarthmore College is a small, highly selective liberal-arts college outside Philadelphia with a student body of nearly 1,600, an 8:1 faculty-student ratio, and one of the highest rates of graduates who go on to earn Ph.D. degrees among U.S. baccalaureate institutions. In this paper, I describe what we did at Swarthmore to move from a disjointed data management process to a more integrated one. This improvement occurred in the context of an effort to effectively tell the story of one of our most successful grant-funded STEM initiatives.

Swarthmore's sponsored programs office played an important role in this transformation. At our college, as in many other PUIs, research administrators become deeply involved in strategic initiatives and even serve as agents of change beyond delivering their core services, since PUIs have more limited resources and staff than research universities (Kreidler et al., 2012;

Lowry et al., 2001). At Swarthmore, we consider it essential to reduce the burden of faculty with heavy teaching loads and no graduate assistants, and research administrators are part of the solution.

In this reflective inquiry/case study, I will (1) provide background on how the need for better data management at Swarthmore was determined; (2) outline the problems that served as external catalysts for our data improvement efforts; (3) provide detailed observations of our collaborative process to better integrate our data; (4) connect our experience with prior literature that illuminates the complexity of maturing an institution's data governance culture; and (5) offer lessons learned and recommended solutions related to data for grants.

Background

Even before the project featured in this paper, we had ample opportunity to discover that access to timely, high-quality data would be integral in enabling us to distinguish ourselves from our peers in prestigious grant competitions. However, we were also keenly aware that without effective data management, responding to extensive data requests from funders would require significant investments of staff time.

For example, Swarthmore and 103 other schools faced substantial demands for data when submitting applications in the 2012 competition for the George I. Alden Excellence in Career-Related Education Award. Unfortunately, most of us did not meet expectations. In a 2013 declination letter, the funder indicated that only 15% of applicants "were able to adequately respond to their request for data" (Alden Trust, 2013). This was a clear signal that we needed to make an organizational shift and look at our data in a new way.

Although the inability to respond thoroughly and competently to data requests can have significant consequences, including loss of funding opportunities, the need to improve data management may still not receive sufficient attention. Becoming more equipped to respond to requests from prospective funders competes with many other administrative priorities, most of which are usually more salient to key decision makers in the institution.

Our path to change began when we received a 2014 Biomedical/Biobehavioral Research Administration Development (BRAD) award from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD). Our project goal was to improve sponsored research operations and faculty development support at Swarthmore. As the associate director of sponsored programs and prospective principal investigator (PI) at the time, I wrote the following justification of our BRAD goal related to data management:

> [The] power of analytics is increasingly being leveraged in research administration as well as higher education in general. Swarthmore aims to streamline and enhance its data management systems … to amass case-building data annually in support of proposal writing and grant reporting (e.g., alumni graduate degrees by major and race/gender). Gathering other outcome information, such as career trajectory, is a systemic problem for many colleges and universities. (Johnson, Dinitz, & Kolowrat, 2013).

In our BRAD budget, we requested funds to assist with entering our backlog of historical data for all summer research programs and delivering on faculty's longstanding call for better outcomes data for mentored student research.

Problem Statement

The first initiative that benefited from BRAD funding was Swarthmore's Howard Hughes Medical Institute (HHMI) program. HHMI is a prominent funder based in Maryland that invests significantly in undergraduate education. Swarthmore received HHMI funding from 1988 to 2016. In 2012, it was one of 47 small colleges and universities to receive a four-year, $1 million grant from HHMI's Undergraduate Science Education program. Among this group, the College was designated as one of 11 "Capstone" awardees, joining an esteemed cohort of institutions of varying geographies and types—Barnard, Bryn Mawr, Carleton, Grinnell, Hope, Hunter, Morehouse, Smith, Spelman, and Xavier. In addition to being recognized for the impact of their science education programs, Capstone awardees were required to engage in a summative assessment, sharing our best practices and lessons learned with another subset of grantees identified by HHMI as having particular strengths in offering apprentice-based undergraduate research experiences (AUREs). For Swarthmore, completing this summative assessment would require an evaluation of 28 years of student programming!

The primary problem to overcome in preparing this assessment was that various data elements needed to complete the report had been tracked separately in different departments for different purposes. As a result, considerable effort would be required to reconcile the inconsistencies in data deriving from these competing goals and objectives. We had a foreshadowing of the challenges ahead of us when working with HHMI program staff on assessments and progress reports for previous years. In fact, our expectations of future challenges in responding to HHMI reporting requirements were one of our main reasons for including goals related to grants and data management in our BRAD proposal.

When constructing this section of the BRAD proposal, the director of sponsored programs outlined four main categories of challenges related to data management, based on obstacles faced when collecting data for past proposals and reports for funders like HHMI and Alden: "*technical* (integration of data from different systems, elimination of 'shadow systems,'

and historical data in different formats with varying degrees of integrity); *regulatory* (limits on sharing of student data posed by the Family Educational Rights and Privacy Act, or FERPA); *time* (coordination of data entry by multiple administrators with competing demands); and *campus culture* (shifting from departmental ownership of data to a culture of standardized information-sharing)." We encountered all these issues when attempting to clean and analyze our HHMI data. Despite having an administrative system of record that was intended to serve as our authoritative source of institutional data, we found the data in the system to be notably deficient in this instance and had to integrate disparate data originally captured for other purposes to fulfill our reporting needs.

Our main data sources were Banner (the College's primary administrative information system) and the following additional sources outside our administrative system: an Excel spreadsheet tracking outcomes of HHMI-funded students, departmental records, student survey results, old College publications, and external data sources. While helpful for the purposes of day-to-day program administration, the Excel spreadsheet was a "shadow system" that did not align with the best practice of having our administrative system be the "one source of the truth" with regard to College data. In accordance with FERPA regulations, access to demographic data had to be requested from the Registrar's Office. We discovered that only 32% of the HHMI awardees since 1994 had been coded in the College's administrative system at the start of the project (D. Thompson, personal communication, June 10, 2014). This was due to modest staffing and the lack of consistent methods of recording program data over the years,

In fall 2015, the College faced the task of producing a summative assessment of its HHMI program activities six months earlier than expected, in the midst of full workloads across departments. Although this was a very challenging requirement, the HHMI report served as a pivotal external catalyst that pushed the boundaries of our institutional data culture.

Observations

Completing the historical record of 20+ years of data on HHMI-funded students as part of the HHMI Capstone awardee cohort and integrating it with other institutional data were described by HHMI program staff as an "ideal pilot project" for allocating BRAD sponsored research capacity-building funds that were earmarked for grants data management efforts (E. Vallen, personal communication, June 5, 2014). This challenge was tackled by the Summer Opportunities and Research (SOAR) coding project. The director of sponsored programs developed this coding project in collaboration with the office's data coordinator, who leveraged experience as a data steward in various capacities to develop the project plan. An internal employee—a member of the Alumni and Gift Records department—was hired as our limited-term BRAD program assistant. This position was created per the BRAD capacity-building plan to help with various BRAD program initiatives.

The first task was to ensure that all HHMI awardees were coded in our administrative information system, a foundational step in our efforts to improve case-building for grant proposals and reports. Our program assistant worked 10 hours per week during this period to accelerate our progress in coding, with approximately 50% of that time devoted to this project. In addition to developing the SOAR blueprint, the data coordinator supervised the program assistant and ensured successful integration of the newly coded data into our Argos reporting system. As the BRAD PI, I monitored progress in meeting project milestones, leveraged BRAD funding to support the initiative, and liaised with key stakeholders.

The SOAR project was an important step in addressing the incomplete historical electronic record of summer research experiences, as evidenced by the fact that most of our summer research data at the time represented only the previous 12 of the total project's 20-plus years. The first phase of the project focused on making the summer research awardee data from 2002 to 2014 complete, which required extensive investigation to address missing data elements. For instance, a subset of the records for awardees lacked an award year, which would render moot any attempt at a longitudinal analysis. The awardees were segmented into four categories: (1) current awardees receiving stipends from a fund with an established code in Banner; (2) recent awardees from 2013 and 2014 receiving stipends from a fund without an established code; (3) historical records, about one-third of which were not coded; and (4) current awards with incorrect or incomplete information. It was not possible to simply fill in the missing data; for instance, where a code was missing, a new one might have to be created by the advancement systems department. Another critical task was to link each awardee record to the corresponding student ID, the linchpin required to connect disparate datasets and thereby unlock key insights concerning the impact of summer research on student outcomes. The award types included interdisciplinary awards, mentorship awards for students in their first or second year (especially underrepresented students), and off-campus awards.

As a result of the coding of student IDs and the subsequent linkage to the student demographic and academic data elements shown in Table 1, we were able to provide more detailed snapshots of students in the first phase of the HHMI capstone analysis than in the past:

(Insert Table 1 Here)

Figures 1 to 3 provide an overview of student demographic data on our HHMI-funded students by three main independent variables (ethnicity, gender, and first-generation status) as examples of analyses that could not have been performed with accuracy before the SOAR project.

(Insert Figure 1 Here)

(Insert Figure 2 Here)

(Insert Figure 3 Here)

During the second phase of the SOAR project, the goal was to code the student IDs and activity codes (indicating funding sources) for a larger dataset covering all Natural Science and Engineering (NSE) student summer research (including but not limited to HHMI-funded research) going back to 2002. This would enable Swarthmore's Institutional Research (IR) office to conduct further analyses that were not previously possible due to gaps in data, including career outcomes for alumni who received HHMI summer research funding while at Swarthmore. The College's advancement systems team combined the Swarthmore HHMI program's detailed Excel spreadsheet of historical career outcomes data with a newly created database using the Argos reporting tool. The new database would allow the program to assess the impact of summer research experiences "parsed by a range of independent variables, including funding source, age of student when they received the award, 1st-gen or URM [underrepresented minority] status, etc." as well as to "compare 'summer fellowship awardees' to the subset of NSE majors who never received college-administered support for summer research" (K. Siwicki, personal communication, August 11, 2015).

More than 250 additional hours were spent coding our backlog of data during this phase in preparation for later analysis by IR. Significant time was expended in reviewing old documents, including old course catalogues, hard-copy files, and reports on fellowships and prizes (tasks largely completed by our Sponsored Programs student worker). Once again, BRAD funds helped increase our capacity through hiring another temporary worker to research and input student IDs and activity codes. This infusion of additional resources proved timely when the College was notified that the Capstone report deadline had been moved up to six months earlier. A satisfactory evaluation of our report was a condition for authorization to spend down a significant amount of remaining grant funds.

In the next phase of the project, our temporary worker (a local college student majoring in computer science) entered data for all of our NSE student data going back to 1994. Deciding how far to go back chronologically was itself a major question that required considering several factors, including the availability of data, the data format, the presence or absence of student IDs and whether the data would have to be manually entered or could be electronically transferred. The earlier data were quite spotty—at the beginning of this phase, we had no identifiable records on research experience for nearly three-quarters of the students.

Unforeseen data quality issues were addressed by drawing on institutional knowledge across campus to devise creative solutions. For instance, inconsistent data collection had led to the discovery of false negatives for summer research participation. A faculty member suggested

using our records for student participation in Swarthmore's chapter of the Sigma XI scientific research society to identify missing student researchers. Accordingly, we cross-checked Sigma Xi information in our database against information in archived hard-copy bulletins. Additionally, alumni who were awardees of the National Science Foundation Graduate Research Fellowship Program (GRFP) were coded simply as NSF recipients, since this fit the purposes for which the data was originally collected. We contacted the NSF GRFP program to request a list of recipients who received their undergraduate degree from Swarthmore, but our request could not be fulfilled. Fortunately, our Fellowships and Prizes office was able to identify GRFP recipients for many of the years in the reporting period.

Another data issue concerned the difference in how career outcomes were collected in the HHMI program's database as opposed to the College's Banner database. The career outcomes collected in Excel were captured solely in free-form text fields. However, Banner used standardized codes to classify career outcomes in addition to text comments. The three categories of occupational codes were primary occupation (e.g., Researcher [P24]), employer's primary area of work (e.g., Research—Other Scientific [RSOS], and specialty codes within an occupation (e.g., Biomedical [BU]).

The various cleanup efforts performed by many dedicated staff members resulted in a mass upload of data by our advancement systems office. This empowered our IR office to conduct extensive analyses of career and educational outcomes for alumni who received HHMI summer research fellowships, as well as for NSE students who conducted summer research, in comparison to those who had no summer research experience. Data from the National Student Clearinghouse, the premier organization for educational reporting and data exchange, supplemented the College's information on Ph.D. matriculation and attainment. We were able to navigate previous obstacles to obtain more complete data on students in faculty or teaching careers. However, given our time constraints, we were unable to look at other career outcomes due to the extensive data cleanup that would have been required.

The culmination of our efforts appeared in the College's 2016 capstone report to HHMI. In the report, the Swarthmore faculty member directing HHMI-funded activities acknowledged this herculean collaborative effort, stating, "Identifying [HHMI award subgroups and demographic contrast groups] and integrating with outcomes records was a major undertaking for the College, engaging institutional research and administrative staff with the HHMI grant program in new and significant ways" (Siwicki, 2016). Based on the satisfactory completion of the summative assessment, the HHMI program authorized Swarthmore to spend down the remaining grant funds. Swarthmore joined other Capstone awardees in sharing a summary of its findings on the "Supporting STEM Success in a Liberal Arts Context" website (serc.carleton.edu/liberalarts).

In addition to satisfying the HHMI reporting requirements, the data that we collected reinforced the College's sense of the value of summer research programs and were used in communications to other internal and external stakeholders. As anticipated, the Capstone analysis indicated that students' experience of conducting summer research was correlated with positive educational outcomes, and this information was shared with the Swarthmore community in June 2016 (Myers, 2016). Although the identified correlation does not prove a cause-effect relationship, it is strongly suggestive of the positive value of summer research experiences. The generalizability and statistical significance of the results are limited, since creating a control group of students who do *not* receive research experience does not align with our institutional values. However, Swarthmore science majors who conducted summer research enrolled in educational programs after graduation at higher rates than students who had not participated in a summer research experience. The difference was especially apparent with regard to Ph.D. attainment, as those with summer research experience earned doctorates at twice the rate of other students. The findings also included very favorable outcomes for first-generation and underrepresented students who had conducted summer research; among graduates from 1999 to 2005, this subgroup of science majors and minors achieved the highest rate of completing graduate programs of any subgroup of NSE students. Over 80% of these students had received summer research fellowships through Swarthmore's HHMI grant funding.

Survey of Related Literature

The HHMI data project raised our organization's awareness that data collection conducted for the purpose of daily operations of administrative departments may not serve our institutional need to demonstrate program outcomes and accountability to funders, which would necessitate institutional change. For instance, the absence of student IDs connected to individual HHMI student records did not pose a problem in the course of administering HHMI initiatives but proved a fatal flaw once an aggregate analysis of the program was required. Studies of organizational dynamics indicate that systemic change rarely happens in a linear fashion (Meyer et al., 2005) and that it is often catalyzed by external forces and characterized by incremental progress. The increasing "volume, quantity, and complexity" of external reporting, and the corresponding demands on offices across campuses, have recently been discussed in the institutional research literature (see Ma et al., 2016, p. 35). As numerous as the obstacles are, it is imperative for institutions of higher education to grapple with the issue of data so as to remain competitive in today's environment of limited funding, skepticism about the value of postsecondary education, and calls for accountability.

The trend of institutions having to respond to more burdensome requests for data has been increasing for at least the past two decades. Wells et al. (1999) discussed how "the pressure to respond to these mandates has grown as key funding streams are increasingly linked to performance-based formulas and threaten the financial health of our institutions" (p. 29). The

authors cited federal funders, such as the National Institutes of Health (NIH) and National Science Foundation (NSF), that were (and continue to be) keenly interested in data-driven initiatives to increase the number of students of color who enter and persist in science education. Morest (2009) highlighted the work of the Lumina Foundation, which has been at the vanguard of private funders actively engaging their awardees in "building a culture of evidence." For instance, Lumina's "Achieving the Dream" community college initiative seeks to improve student outcomes for underrepresented students by engendering institutional change to allow administrators to "use data to identify problems, set goals, establish institutional priorities, allocate resources and measure progress" (Morest, 2009, p. 20). Barriers identified among its grantees include limited informational technology and access to data, as well as inadequate institutional research staffing (pp. 20–21).

The research administration profession can learn from the experiences described in the literature about the evolving role of the institutional research office. Research administrators are akin to institutional researchers who, according to Wells et al., are "uniquely positioned to design the linkages among data, information, and decision makers" (1999, p. 31). Wells et al. cited Claggett (1996) and Terenzini (1991), who "encourage institutional researchers to assume a proactive role in shaping the organizational intelligence of their institutions" (Wells et al., 1999, p. 31). Claggett and Terenzini specifically described how the "institutional researcher is [increasingly] becoming involved in institutional advocacy" (quoted in Wells et al., 1999, p. 30). They asserted that to serve as effective knowledge brokers, institutional researchers must "provide contextual understanding, broadening office expertise into new areas … and emphasizing issue-oriented integration and synthesis of findings from multiple projects" (quoted in Wells et al., 1999, p. 30). This requires three types of organizational intelligence: technical/analytical intelligence, issues intelligence, and contextual intelligence.

In recent years, this vision of the institutional researcher as a proactive advocate has started to become institutionalized in the culture of IR and affiliated professions. In its "Statement of Aspirational Practice for Institutional Research," the Association for Institutional Research called for institutional researchers to work with more constituents within their institutions, develop the data skills of staff, and inform the strategic use of data at the executive level (Johnson & Gagliardi, 2019, p. 5). Swing and Ross (2016) proposed a new oversight model for institutional research that supports constituents as they use analytical tools to shed light on issues they are facing while continuing to fulfill traditional reporting obligations (Blake & Wyn, 2019, p. 49). This shift would be of particular benefit to grants professionals, as the current allocation of IR time for fundraising is only 3% according to one estimate (Parnell, 2019, p. 18). The IR office is a key actor in telling the story of the institution (Fingerson & Troutman, 2019, p. 43). Small colleges can be especially fruitful environments for mutually beneficial collaboration between institutional researchers and campus constituents beyond the typical users of institutional data, since specialization within the IR office may not be possible (Blake & Wyn,

2019, p. 50). The National Association of College and University Business Officers (NACUBO) and EDUCAUSE (the higher education IT association) have joined AIR in advocating for new ways of working together to leverage data as an institutional asset (Taylor & Martineau, 2019, p. 53).

The process of responding to external data requests, such as those from grant funders, provides an important opportunity for organizations to examine their institutional data in new ways. In doing so, they often unearth emerging data management issues by shifting the focus from individual transactions to the reporting of aggregate data and overall trends (Kirby & Floyd, 2016, p. 50). Research administrators have the requisite organizational intelligence to view institutional data differently from other administrators who rely on the data solely to execute day-to-day operations. They can overcome differing perspectives on the merits of devoting institutional resources to time-consuming data collection for funders by demonstrating how this effort can simultaneously serve the needs of internal constituents (Ma et al., 2016). Given the high stakes often involved in such requests, research administrators are especially incentivized to leverage their broad institutional knowledge and experience to initiate or join in efforts that bring together campus partners (e.g., faculty, information technology, institutional research, advancement, alumni relations, etc.) to move such initiatives forward.

The stakeholders involved in the HHMI data project were not systematically following a particular methodology for project conception and implementation. Nevertheless, our efforts often aligned with best practices (and pitfalls) in this area as outlined in the literature on data governance. The phenomena that the director of sponsored programs was observing in relation to institutional data resembled those described in an EDUCAUSE case study entitled "DataMASTER: Success and Failure on a Journey to Business Intelligence" (Blanton, 2012). The DataMASTER (Management and Analytics for Strategic, Timely Education Reporting) project at Portland State University was created to address the inability of the institution's legacy reporting system to produce the data needed to answer key strategic questions. This required an organizational shift (similar to the one we undertook at Swarthmore) from owning data to stewarding data that belong to the institution, in order to break down silos and unearth the many shadow systems that existed.

Reflection and Recommended Solutions

The key observations of the HHMI data project provide an instructive case study regarding the use of data to demonstrate program effectiveness as well as the imperative to mature from purely transactional to strategic uses of data. Colleges and universities must have institutionalized practices that ensure the gathering of complete and accurate data in a flexible manner while minimizing response times and administrative burdens for time-sensitive funder requests. Given the scale of the societal issues that prominent funders are endeavoring to

ameliorate, they no longer want to invest their financial resources in inputs without compelling evidence of positive outcomes. Accordingly, grantees must take their enhanced data reporting responsibility seriously to rise to this new level of accountability. The foundation for such an evolution is the creation of a data culture that encompasses multiple stakeholders. Our experience with the HHMI project yielded the following lessons that will help any school improve their data management.

One lesson from the HHMI data project is to rely less heavily on self-reporting of student and alumni outcomes (such as those found in the Survey of Undergraduate Research Experiences [SURE] and the Consortium on Financing Higher Education [COFHE] alumni survey), since alumni data are often incomplete or labor-intensive to track down. For instance, our IR office used the National Student Clearinghouse—a U.S. nonprofit organization and the premier provider of student data, collected from data exchanges with more than 3,600 educational institutions—to secure Ph.D. enrollment data on our alumni to complement the alumni surveys that have been a staple of our advancement division's work. Our Records office is also increasingly using the Clearinghouse and is piloting a new commercial solution that cleans and updates data, starting with employment information that can be uploaded into our database. These innovations are making data integrity efforts more proactive and not solely dependent on busy alumni to share information about themselves or respond to alumni surveys. As Swarthmore finalizes its BRAD grant reporting, access to more robust employment information about our NIH-funded students will allow Sponsored Programs to tell a fuller story about the impact of the NIH research grants our faculty have secured during the BRAD project period.

Institutions reflecting on their data practices should also weigh the costs and benefits of each data request and determine the appropriate level of staff time to invest in responding to it. The urgency of the HHMI data project and the potential loss of funding if we failed to provide satisfactory data forced us to find ways to protect staff time for data requests that are of the highest consequence to the institution. At the department level, Sponsored Programs changed its approach to responding to faculty data requests in order to preserve administrators' bandwidth for such large-scale, mandatory data requests. Instead of reflexively providing the exact data requested for routine proposals, we now first ask if the faculty member might consider using existing data from the "Outcomes" section of the College's Fact Book or other IR reports. For example, data comparing Swarthmore medical school acceptance rates to the national average can serve as a proxy for student caliber, which would otherwise have to be demonstrated by providing department-level data on medical school students and graduates. Although the substitution is not perfect, it is adequate for most grant proposals and avoids additional time expenditure in data gathering. It is especially important for grants offices at smaller colleges to be mindful of the number of special data requests made, so that staff with these responsibilities have more time to attend to important data integrity issues and strategic considerations rather than being overloaded with day-to-day fulfillment requests. Our grants office now views itself as

one of many guardians of data quality, tracking the sources of data used in institutional boilerplate more consistently and ensuring that the most up-to-date institutional data are used in our grant proposals.

The director of sponsored programs codified these lessons learned in a departmental policy for the use of data in grant proposals. While acknowledging that data are important in creating a compelling proposal narrative and increasing the competitiveness of grant applications, the new policy also clearly delineated what the office can and cannot provide. Institutional boilerplate and data for "Facilities and Resources" documents and impact statements are regularly updated. We also offer to craft new language using readily available IR data so as to make a special data request unnecessary. If these options do not meet the faculty member's needs, we next ask faculty to consider working with our Alumni and Gift Records department before approaching IR.

Institutions may consider replicating the "Alumni Data Request Form for Grants and Departmental Reviews" created by Sponsored Programs and Alumni and Gift Records (hereafter "Records"). The form asks for the deadline, the purpose of the request (e.g., grant proposal, grant report), whether the information is required by the funder, and what data points (columns) should be included in the Excel spreadsheet. The possible selections of data points are major, minor, Honors status, scholarships/prizes/fellowships, highest degree, institution awarding that degree, occupation, industry, job title, employer, and other. Raw data are provided to the faculty member, with the expectation that they or departmental administrators will handle the analysis. This process ensures that only the most complex requests are directed to IR. The corresponding policy concludes with an appeal to faculty to share alumni information with the Records department, which will enable better responses to both internal and external inquiries.

To effectively sustain change in the institution's data culture and allay concerns, it is imperative to educate key constituents and stakeholders on the high stakes associated with these issues. Following the HHMI report experience, Swarthmore's director of sponsored programs, the director of institutional relations, and a lead faculty member on another massive data initiative presented a faculty lunch talk that discussed data issues faced by both STEM and humanities faculty (Kuharski, Foreman, & Johnson, 2016). The presentation highlighted grantmaking trends, including data-driven decision making and the key role of data in telling a good story about one's institution and project. Our discussion of systemic data issues resonated with our faculty and prompted a collective recognition of the need to move beyond spreadsheets to a coordinated institutional response.

Beyond educating the community, it also became clear that we needed to engage with our senior leadership to develop a sustainable institutional response to our systemic data dilemmas. To strengthen our case as we sought to encourage senior leaders to support more robust data

governance, the director of sponsored programs participated in the June 2017 "Data Lift Off" Northeast Regional Computing Program (NERCOMP) conference. Best practices were brought back and included with the memo prepared for senior leadership. A summary of key takeaways is provided in Table 2. The discussions sparked by this memo led to the creation of a Swarthmore Data Governance Committee, co-chaired by the Chief Information Technology Officer and the Assistant Vice President for Institutional Effectiveness and Assessment. The committee has a public website outlining its charge, philosophy, and working group structure at https://www.swarthmore.edu/data-governance.

(Insert Table 2 Here)

Conclusion

The HHMI Capstone project and its reporting requirements represented uncharted territory for Swarthmore, compelling a concerted effort to revamp how we collected and managed data. Our BRAD-funded project to improve our data management contributed to critical awareness building and laid the groundwork for collaboration between our sponsored programs office, IR office, records and systems staff, and faculty. It also created the momentum that eventually led to a more robust data governance initiative at the institution. This case study demonstrates the value of considering data analysis at the *beginning* of a project and provides lessons learned about navigating the organizational and technical challenges faced when responding to funders' data requests. It also demonstrates how grant officers at predominantly undergraduate institutions often serve as strategic collaborators in areas that extend beyond the traditional mandate of the sponsored research office yet are essential to serving faculty. As these types of initiatives evolve, institutions should consider encouraging faculty to share data with administrators, in order to derive greater institutional benefit from information and respond more effectively to high-level issues and opportunities.

Author's Note

This paper is the author's sole responsibility and does not claim to represent the views of Swarthmore College, the National Institutes of Health, or the Howard Hughes Medical Institute. The author explored these themes in a previous discussion group at the National Council of University Research Administrators (NCURA) annual conference in August 2016.

Acknowledgments

This article summarizes collaborative work carried out by the author and at least 35 colleagues as part of their institutional duties. Special thanks are due to Don Cooney, Kathy Siwicki, Beth Svenson, Debbie Thompson, Theresa Rodriguez, Robin Shores, Barbara Mann, Mimi Weiler,

Liz Vallen, Melissa Mandos, Jesse Cason, Joe Mulray, and many others for their invaluable contributions. The data work described in this article was supported by the Eunice Kennedy Shriver National Institute of Child Health and Human Development at the National Institutes of Health (NIH) under Award Number G11HD080231. The science education programs referenced in this article were supported by grants from the Howard Hughes Medical Institute (HHMI), most recently a Capstone Award from the Undergraduate Science Education program (Grant #52007561). I also wish to thank the JRA Author Fellowship program and, in particular, Dr. Cliff Studman of St. John's University in Dodoma, Tanzania for his guidance and support throughout the writing process.

Tania N. Johnson, MA, CRA
Director of Sponsored Programs
Swarthmore College
500 College Avenue
Swarthmore, PA 19081
United States of America
Telephone: (610) 690-5713
tjohnso2@swarthmore.edu

Correspondence concerning this article should be addressed to Tania Johnson, 500 College Avenue, Swarthmore, PA 19081, tjohnso2@swarthmore.edu.

Tania Johnson, MA, CRA is director of sponsored programs at Swarthmore College. Since 2005, she has served in roles of increasing responsibility in support of the College's efforts to strengthen sponsored research, including the role of lead PI on Swarthmore's NIH Biomedical/Biobehavioral Research Administration Development (BRAD) grant. Her partnerships with faculty across departments and divisions have translated into increased funding for research, instrumentation, and curriculum development during her tenure. She has participated in two noteworthy leadership programs in higher education administration: the HERS Bryn Mawr Summer Institute in 2010 and the Harvard Management Development Program in 2012. Ms. Johnson received her undergraduate and master's degrees from the University of Pennsylvania and studied at the Institut d'Etudes Politiques de Paris (Sciences Po). She received the Certified Research Administrator (CRA) designation in 2013.

References

Blake, L. P., & Wynne, T. C. (2019). An Integrated View of Student Success at Small Colleges. *New Directions for Institutional Research*. 2019(184). DOI: 10.1002/ir.20321

Blanton, S. (2012). DataMASTER: Success and Failure on a Journey to Business Intelligence.

EDUCAUSE Review Online. (July 2012). https://er.educause.edu/articles/2012/7/datamaster-success-and-failure-on-a-journey-to-business-intelligence

Claggett, C. A. (1996) "Maximum Impact: Research Strategies for Effective Institutional Management, Assessment, and Advocacy." Presentation at the Annual Summer Institute on Institutional Effectiveness and Student Success, Charleston, S.C., June 1996.

"Data Governance" (2021). Retrieved from Swarthmore College website: https://www.swarthmore.edu/data-governance

"Discussion of George I. Alden Excellence in Career-Related Education Award—Executive Summary" (2013). Retrieved from Internet Archive website: https://web.archive.org/web/20190306155101/http://www.aldentrust.org/observations/aldenexcellenceaward.html

Fingerson, L., & Troutman, D. R. (2019). Measuring and Enhancing Student Success. *New Directions for Institutional Research.* 2019(184). DOI: 10.1002/ir.20320

"HHMI Grant Program" (2012). Retrieved from Swarthmore College website: https://www.swarthmore.edu/hhmi-grant-program

Johnson, G., & Gagliardi, J. S. (2019). Editors' Notes. *New Directions for Institutional Research.* 2019(184). DOI: 10.1002/ir.20317

Johnson, T., Dinitz, K., & Kolowrat, N. (2013). Strengthening Sponsored Research Infrastructure and Faculty Development at Swarthmore College. Proposal to the Biomedical/Biobehavioral Research Administration Development (BRAD) program, Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD).

Kirby, Y. K., & Floyd, N. D. (2016). Maximizing Institutional Research Impact Through Building Relationships and Collaborating Within the Institution. *New Directions for Institutional Research.* 2015(166). DOI: 10.1002/ir.20130

Kreidler, M. & and Southwick, S.J. (2012, January-February). Research Development at PUIs: It's What We Do. *NCURA Magazine*. Vol. XLIV, No. 1, 29-30, Retrieved from: ncura.edu

Kuharski, A., Johnson, T., & Foreman, D., (2016). CHOPIN WITHOUT PIANO and Beyond: Reflections and Opportunities for Faculty and Institutional Project Grants, Faculty Lunch Presentation, Swarthmore College.

Lowry, P.S. & Hansen, S. (2001). Reconsidering Research Administration at Predominantly Undergraduate Colleges and Universities. *Research Management Review*. 12(1), 11-15.

Ma, Y., Kulich, J., & Hu, S. (2016). Burden in the Eye of the Beholder: Why Should Administration Care? *New Directions for Institutional Research*. 2015(166), 35-46. DOI: 10.1002/ir.20129

Meyer, A. D., Gaba, V., & Colwell, K. A. (2005). Organizing Far from Equilibrium: Nonlinear Change in Organizational Fields. *Organization Science.* 16(5). https://www.jstor.org/stable/25145986

Morest, V.S. (2009). Accountability, accreditation, and continuous improvement: Building a culture of evidence. *New Directions for Institutional Research.* 2009(143), 17–27. DOI: 10.1002/ir.302.

Myers, G. (2016, June 23). Swarthmore Recognized for Sustained Science Education Excellence. Retrieved from Swarthmore College website: https://www.swarthmore.edu/news-events/swarthmore-recognized-sustained-science-education-excellence

Parnell, Amelia R. (2019). The Journey of Analytics. *New Directions for Institutional Research.* 2019(184). DOI: 10.1002/ir.20318.

Rovnyak, D. S., & Shields, G. C. (2017, July 7). How Undergraduate Research Drives Science Forward. *Inside Higher Ed*. Retrieved from: https://www.insidehighered.com/views/2017/07/07/undervaluation-role-undergraduate-research-advancement-scientific-knowledge-essay

Siwicki, K.S. (2016). Year 4 Narrative Report for 2012 Capstones: Swarthmore College.

"Supporting STEM Success in a Liberal Arts Context" (2016). Retrieved from SERC website: serc.carleton.edu/liberalarts

Talend.com. Data Lake vs. Data Warehouse (2021). Retrieved from: https://www.talend.com/resources/data-lake-vs-data-warehouse/#:~:text=Data%20lakes%20and%20data%20warehouses,processed%20for%20a%20specific%20purpose.

Taylor, J. L., & Martineau, M. (2019). Creating IR and Faculty Partnerships Toward the Shared Goal of Student Success. *New Directions for Institutional Research.* 2019(184). DOI: doi.org/10.1002/ir.20322

Terenzini, P. T. (1991) "On the Nature of Institutional Research and the Knowledge and Skills It Requires." Paper presented at the thirty-first forum of the Association for Institutional Research, San Francisco, May 1991.

Wells, J., Silk, E., & Torres, D. (1999). Accountability, Technology, and External Access to Information: Implications for IR. *New Directions for Institutional Research.* 1999(103), 23-39. DOI: 10.1002/ir.10302

Table 1. Data coded for HHMI awardees

Data elements for all HHMI awardees (Awards I-VII)	
Independent Variables	
Demographic	Gender Ethnicity *(HHMI categories aligned with pre-2010 federal reporting standards for consistency with historical data)* American Indian or Native Alaskan Asian Black (not Hispanic origin) Hispanic/Latino Native Hawaiian or other Pacific Islander Not reported Other minority White (not Hispanic origin) First generation (Y/N)
Student Engagement Outcomes	
Academic	Majors Honors
Activities	Sigma Xi

Figures 1-3 represent analyses that could not have been easily produced prior to the HHMI data project.

Figure 1. Ethnic diversity among HHMI students

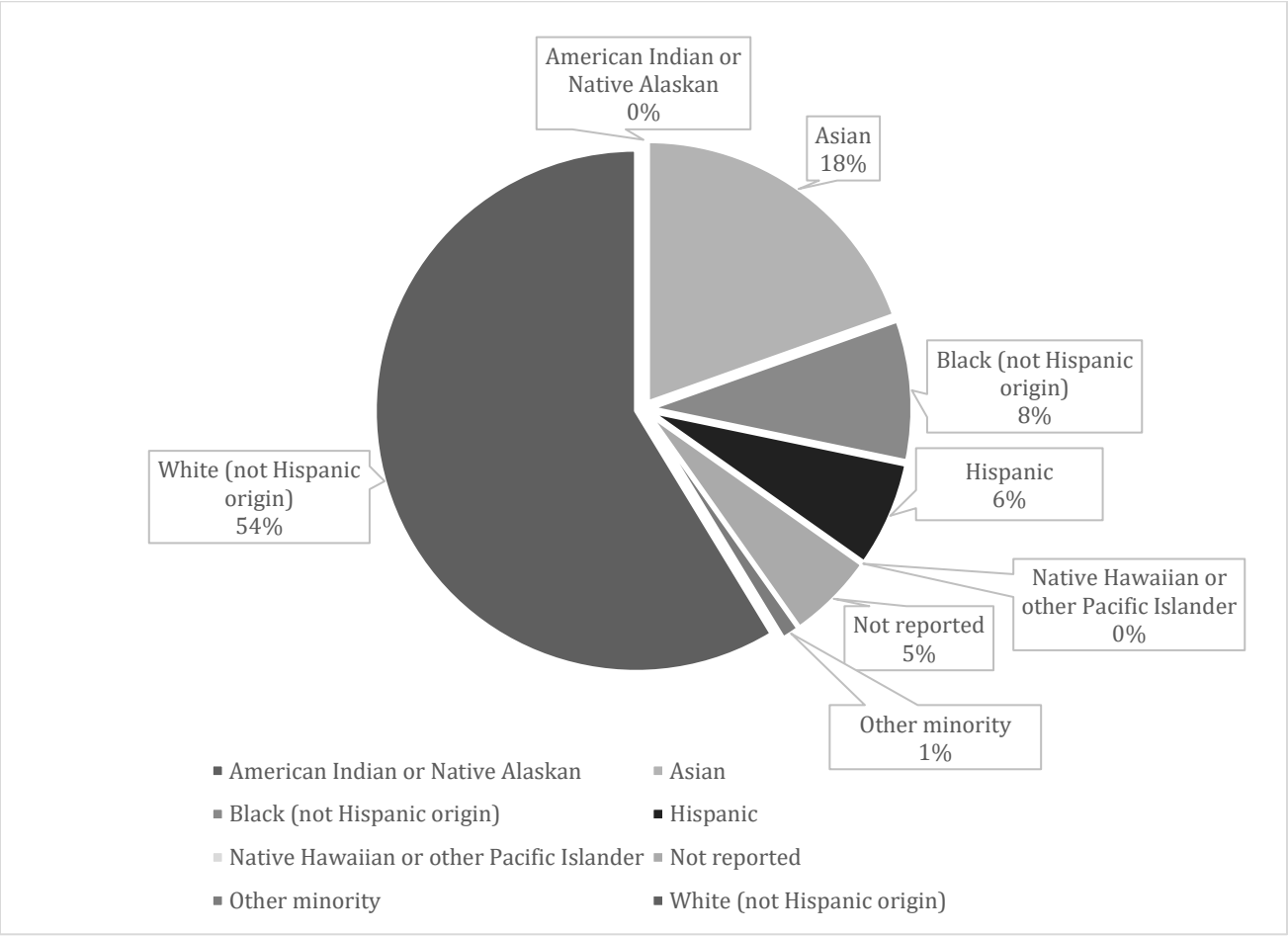

Figure 2. Gender breakdown of HHMI students

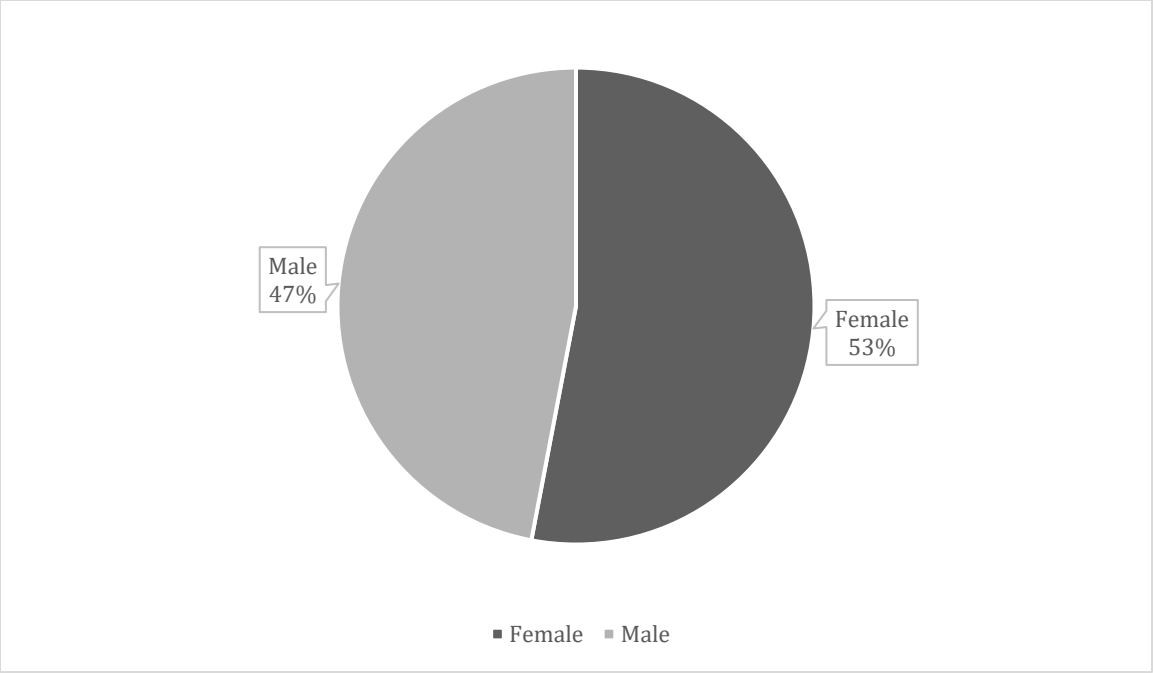

Figure 3. First-generation access among HHMI students (Awards II-VII)

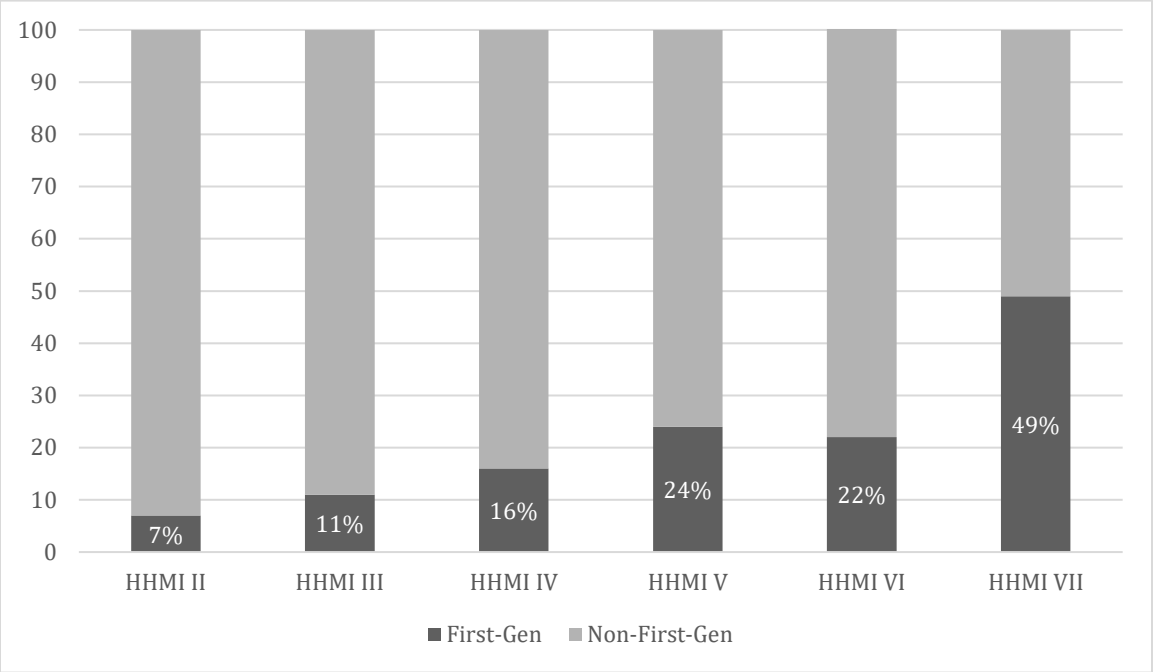

Table 2. NERCOMP "Data Lift Off" Key Takeaways

Leverage Data for Student Success
• Dashboards for president, faculty, deans, chief diversity officer, etc. • Predictive analytics (e.g., enrollments) • Inclusivity dashboards (e.g., graduation/retention, study away, etc. by race/gender) • Target students who meet criteria for fellowships • Use postgraduate outcome data for assessment
Data Governance: People Before Systems
• Determine leader of the initiative (someone outside IT, though IT is a key partner) • Need sponsors and champions for project (e.g., advisory committee) • Bring in functional users and focus on their needs • Understand the business process to understand the data model • Understand each other's languages and professional cultures (e.g., IR and IT) • Data silos are not a technology problem: "Gatekeeper model can go for only so long"
Policy
• Create guiding principles for security/access • Develop data dictionaries and documentation (and training) • Decide who should see certain data points • Resolve conflicting definitions and strive for one definition of truth (with flexibility) • Formalize behavior around use of data • Determine acceptable level of data quality (e.g., 90% accurate) • Empower users as well as data trustees
Systems
• Make "plumbing connections" between systems (e.g., Chief Integration Officer) • Data lake (vast pool of data with undefined purpose) vs. Data warehouse (structured repository of data with a purpose) • Erroneous assumptions built into solutions (e.g., all first-gen students at-risk) • Different standards for transactional systems vs. reporting systems • Mitigate impact of enterprise system change on data governance plans

Note: Descriptions of data lake vs. data warehouse retrieved from Talend.com (see References)