Marywood University

Reap College of Professional Studies

PhD Program, Strategic Leadership & Administrative Studies (SLAS)

Factors that Determine the Use of Data Analysis Tools in an Office of Sponsored Research in an Institution of Higher Education within the United States: A Readiness Assessment

By

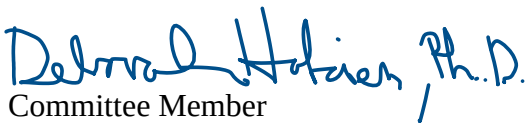
Heather M. Messina (Gazella)

Submitted in Partial Fulfillment of the Requirements for the Degree of Doctorate of Philosophy in Strategic Leadership & Administrative Studies (SLAS)

Date of Approval: March 18, 2024

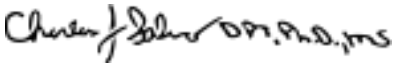
Approved:

Committee Chair



Committee Member

Committee Member



Reader

Reader

Date of Approval:

Ph.D. Program Director

Acknowledgments

I will always remember my years as an undergraduate and graduate student at Marywood University. The early influence of lifelong learning and curiosity began as a student educated by the Sisters of IHM at St. Rose Elementary School and Sacred Heart High School in Carbondale, PA. Thank you to my high school English teacher, who was influential in my career path and would be proud (but not surprised) of this accomplishment. Thank you to my exceptional mentors, past and present supervisors, and amazing colleagues for your friendship, leadership, and mentorship in my research administration career.

To my chair and committee members, Dr. Matthew Caputo, Dr. Deborah Hokien, and Dr. Meagan Mielczarek, your willingness to mentor me throughout this past year has been remarkable. I am sincerely grateful for your unwavering support and invaluable guidance throughout my dissertation journey. Your commitment to academic excellence and mentorship has been a source of inspiration, and I am truly fortunate to have had such a dedicated and knowledgeable committee. Your contributions have enriched my academic experience and played a pivotal role in my personal and professional growth. Thank you for sharing your expertise and for being an integral part of this significant accomplishment.

Thank you to my parents and family, who instilled in me the value of education, hard work, persistence, and grit to accomplish the goals I set for myself. To my husband, Michael, and our children, Cora and Aiden, you were there every step of the way through late nights and long days. Your sacrifices and loving support paved the way for generational accomplishments, transcending the limitations of the past. Your dedication has shaped my journey and laid the foundation for our future generations, creating a legacy of education and opportunity.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

- CAS — Complex Adaptive Systems
- CFR — Code of Federal Regulations
- CQI — Continuous Quality Improvement
- DIDM — Data-informed Decision-making
- D/PU — Doctoral/Professional Universities
- HERD — Higher Education Research & Development Survey
- NCURA — National Council of University Research Administrators
- NSF — National Science Foundation
- OMB — Office of Management and Budget
- OSR — Office of Sponsored Research
- OR — Odds Ratio
- R1 — Doctoral Universities – Very high research activity
- R2 — Doctoral Universities – High research activity
- SRAI — Society of Research Administrators International
- U.S. — United States

Abstract

This cross-sectional study aimed to enhance the understanding of data-informed decision-making in research administration via a quantitative analysis to evaluate the usage of data analytics within the research enterprise at institutions of higher education in the United States (U.S.). A survey developed by the researcher with adaptations from existing tools was administered to 184 director-level participants affiliated with the National Council of University Research Administrators (NCURA) across seven U.S. regions. The survey tool was designed to evaluate respondents' demographics, data analytics tool usage, and skillset offering an institutional readiness assessment and enabling analysis of the types of institutions using data analytics tools for decision-making. Results suggested that factors of participant's age, years of work experience in research administration, institutional administrative structure, salary range, gender identity, proposal and award volume, and role definition were not predictors of the use of data analysis tools within the research enterprise at institutions of higher education in the U.S. A modest correlation was found between institutional expenditures, as reported in the 2021 NSF HERD Survey, and Carnegie Classification of Institutions of Higher Education (2021) R1 (N=86) and R2 (N=17). Regression analysis revealed that institutional research expenditures were the most influential factor for data analysis tool usage, especially for institutions in the top 25% and mid 50% range of expenditures. The study concluded that R2 institutions, comparable to their R1 counterparts, are similarly engaged in data analytics activity. The broad path toward implementation of data analytics tools signals ongoing efforts at institutions of higher education, more specifically the research enterprise, and potential areas further research.

Keywords: *Research administration, sponsored research, institutions of higher education, data analysis tools, data readiness assessment, data analytics, institutional research*

Chapter 1: Introduction

The Problem and Its Setting

Recent studies have suggested significant benefits for institutions of higher education, in particular the research enterprise, to adopt best practices for using data analysis tools to inform decision-making. Institutions of higher education are operating in an increasingly complex and competitive environment. They are under mounting pressure to respond to national and global economic, political and social changes, such as the growing need to increase the proportion and diversity and equity of students in STEM disciplines, increase scientific research capacity and output, and ensure that the quality of learning programs is both nationally and globally relevant (Daniel, 2015; Nguyen et al., 2020; Webber & Zheng, 2020). Institutions of higher education have also been severely affected in recent years by a number of unfavorable happenings, which include but are not limited to, a decline in matriculating high school graduates, an unstable economy, a decrease in federal and state funding, a hyper focus on STEM career training, stricter F-1 visa approvals for international students, and more recently, the COVID-19 pandemic and changes in affirmative action. Institutions have been forced to reevaluate the way strategic decisions are made in order to face the unknowns of the future and the abrupt realities that have collectively reshaped higher education (Chan & Randall, 2021; Ueland et al., 2021; Borgman & Brand, 2022). To navigate these challenges, savvy leaders have an opportunity to leverage data to make informative decisions and steer their institutions through uncertainty.

The question still remains: How many offices of sponsored research, within the research enterprise, at institutions of higher education in the U.S. have assessed their readiness to implement data analytics tools, and therefore, data-informed decision-making? The research administration, institution, management, and employee perspectives highlighted in the literature

review reveal that each has a significant opportunity for impact and to be part of a strategic pathway in influencing data-informed decision-making. Data-informed decision-making can vary depending on the type of institution of higher education. Different types of institutions – from research universities and liberal arts/teaching colleges to community colleges – have different priorities, goals, and challenges. The approach to data-informed decision-making would vary depending on the focus of and influence on the institution and would align with the mission and specific objectives. With large volumes of both academic and scientific research data, institutions of higher education have the data input needed to benefit from the outcomes of using data analytics tools (Daniel, 2015; Nguyen et al., 2020; Borgman & Brand, 2022). For the purposes of this study, data analytics is viewed as a resource to help streamline administrative workflows, identify bottlenecks, and enhance strategic planning using information already collected by the institution of higher education. In research-focused institutions, leveraging data analytics can foster research growth. This involves exploring potential areas of interest, assessing research impacts, and strategically allocating resources for key research initiatives. Proactive measures position research-focused institutions to adapt effectively to evolving funding landscapes and enhance their preparedness for future opportunities.

An Office of Sponsored Research (OSR) serves as the primary campus contact for matters related to the research enterprise including; funding proposal preparation, review and submission; review and acceptance of grant agreements and contracts; furthering relationships with government and industry; establishing and maintaining policies and trainings; and financial management and oversight of projects and programs. Research administrators, the name for professionals in the field, have a breadth and depth of knowledge that goes beyond general administrative and financial skills, and with potential for developing new skillsets, particularly

those in data analysis. The role of a research administrator at an institution of higher education within the U.S. is quickly expanding from pre-award proposal development and post-award financial management to that of an active partner between faculty investigator(s) and institutional leadership to support a path of research growth (Robershaw & Wolf, 2023; Cargill, 2022).

Offices of sponsored research capture a vast amount of data (both quantitative and qualitative) that, with the right data analysis tools, can be used to understand trends and outliers, and forecast future workload volume and infrastructure needs. A requirement of the Office of Management and Budget (OMB) Uniform Guidance 2 CFR 200 (Subpart E), the guidebook for all federal grant regulations, is that in order to effectively manage a grant award, the office of research must capture, in an electronic database, very detailed information about each award (2014, p. 83-127). Information captured by offices of sponsored research includes the principal investigator's name, academic department/center, other key personnel, project title, start and end dates, proposed and awarded amount, indirect cost rate, and sponsoring agency. Ideally, with the right data analysis tools, the information captured (data input) can be synthesized into specific reporting parameters (data output) including, for example, a summary of proposal or award counts and amounts by department, agency, or principal investigator. The data output can be used by research administration leaders to determine office infrastructure needs, justify budget requests, redistribute workload to be more equitable among administrators, or flex between centralized and decentralized models.

One application of the Continuous Quality Improvement (CQI) process aims to enhance the processes of an organization through readiness assessment. The focus on readiness in the context of CQI refers to the organization’s preparedness and ability to effectively implement and sustain improvement initiatives. One of the first steps toward quality improvement is to evaluate the

organization's current state of readiness for implementing quality improvements. This involves assessing the existing infrastructure, resources, and organizational culture to ensure they support the quality improvement initiatives. By focusing on readiness, organizations can enhance their ability to successfully navigate the challenges associated with implementing and sustaining continuous quality improvement initiatives.

Readiness assessments provide information about the specific components of an organization or process that must be improved in order to create or promote a new direction or way of thinking. "Before big data and data analytics tools can be useful to institutions of higher education, there must be a fundamental shift in thinking and repositioning in the mindset of institutional leadership and management. Analytics technology is constantly evolving; it has changed dramatically over the years and is still advancing rapidly today" (Attaran et al., 2018, p. 5).

This study aimed to contribute to the growing body of literature on the use of data analysis tools in research administration, specifically in institutions of higher education within the U.S. For this study, the researcher focused on the research administration enterprise. A survey was conducted with a sample population of research administrators who have active membership in NCURA and employment at institutions of higher education within seven regions across the U.S.

There are three main ways this literature review and subsequent data collection and analysis contributed to current research on this topic. First, understanding how many and what type of institutions, according to the Carnegie Classification of Institutions of Higher Education (2021) have the infrastructure in place (People, Processes, and Data Management) to reach a higher level of data maturity. Second, creating an evaluation tool that can be used by institutions to perform internal assessments of their own research enterprise and adjust as needed. Third,

performing a data readiness assessment combined with other institutional and demographic information to help determine what factors predict data analysis tools usage.

Theoretical Framework

An in-depth literature review revealed four theories of organizational behavior to analyze further the opportunities of using data to support decision-making in institutions of higher education. One of the major theories revealed was Complex Adaptive Systems (CAS) Theory. Institutions of higher education, and the research enterprise that exists within, can be described as meeting the criteria for a CAS because they contain nonlinear relationships and multiple layers of hierarchical administrative and leadership networks that "seem at times to stubbornly resist transformative change while simultaneously adapting and evolving" (Ueland et al., 2021). By embracing a cultural shift in the way decisions are supported, the integration of data analytics tools into higher education institutions holds the potential to substantially enhance strategic decision-making across academic services, scientific research endeavors, and administrative workflows (Daniel, 2015).

Complex Adaptive Systems Theory

CAS Theory provides a framework for understanding the intricate interactions among multiple, independent, and intelligent agents within an umbrellaed and interdependent environment. This theory underscores the cohesive whole that has a distinct identity that transcends its individual components (Ueland et al., 2021). CAS respond to national and global economic, political and social changes, such as the growing need to increase the proportion and diversification of students in certain disciplines, increase scientific research capacity and output, and ensure that the quality of learning programs is both nationally and globally relevant (Daniel, 2015).

A paper written by Kevin Dooley (1997) attempted to forge a unified description of complex adaptive systems from several sources and serves as a basis for explaining what is happening within institutions of higher education and why. Despite being over two decades old, this referenced paper remains relevant, as its concepts anticipated contemporary understanding of data-informed decision-making. One aspect of this paper explored the parallel between human and organizational decision-making, acknowledging that "individuals are limited in their information-processing capabilities, and then so too are organizations" (p. 79). Therefore, relying on human reasoning alone forces decision-making and action based on incomplete information and exploration of a limited number of alternatives (Dooley, 1997; Webber & Zheng, 2020). This underscores the notion that human decision-making has inherent limitations, advocating instead for data-informed decision-making to support organizational leaders. Another aspect of this referenced paper considered the "manner in which organizations adapt to complex, uncertain environments by changing their information-processing capabilities, either by reducing the need for information…or by increasing the capacity for information acquisition, storage, and retrieval (e.g., information systems)" (Dooley, 1997, p. 75).

A key characteristic of complex systems is that they are nonlinear and recognizing that cause and effect are also nonlinear requires a different management approach. Leaders within a complex system are more likely to influence positive change when information is shared openly at all levels across the organization, including faculty, staff, and students. In this way, everyone in the institution has an opportunity to contribute to potential solutions by feeling empowered to engage in free and open dialogue. To address inefficiencies and create opportunities to improve decision-making strategies, management must create space for knowledge sharing about best practices, emerging technologies, and other topics. Influencing and managing change within

institutions of higher education requires capitalizing on moments of disruption, sometimes referred to as the "edge of chaos," to guide the CAS toward a different, more desired state.

Approaching the research enterprise as a CAS is one way to think of how primary factors (or emergent patterns) and secondary factors (or independent agents) are influenced by each other in the achievement of institutional goals. An illustrative representation of complex systems depicted by Matt Seaman (2021) was used to show the theoretical framework of this research study. Seaman defines three distinct components of organizations: emergent patterns, independent agents, and influential interactions.

Documenting emergent patterns or outcomes for an organization will help director-level leadership focus on and strengthen areas of our teams that are essential and pivot away from areas that are non-essential. In many cases there are several independent agents, or people, processes, and tools, at work within an organization. It is critical to be aware of business functions and how much or how little they interact and influence each other, and when we are not certain of all the independent agents at play. A key component of influencing a CAS is identifying major interaction points, and then further defining which ones are beneficial and which ones are not (Seaman, 2021). Adopting a broad perspective on organization change enables a natural evolution, without being too prescriptive and narrow about how the implementation should happen.

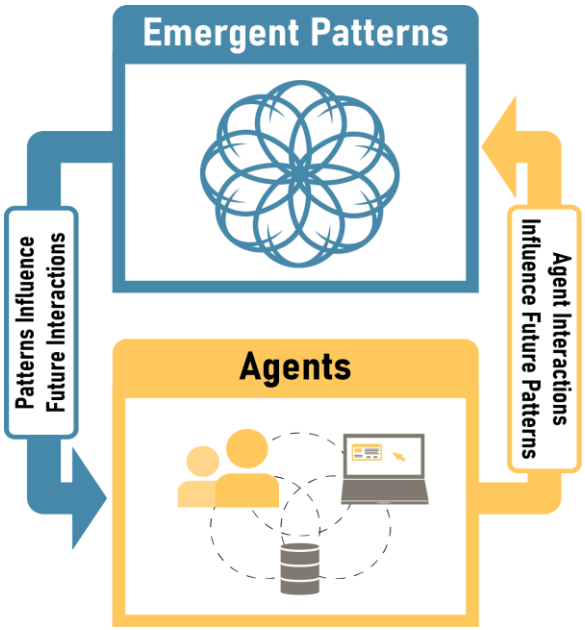


Figure 1. Figure 1 is adapted from an illustration by Seaman, M. (2021) and is the researcher's interpretation of the CAS Theory as it relates to administrative teams. Graphics credit: Emily Creasy of ELC Design & Creative Services.

Conceptual Framework

This study aimed to determine what factors predict the usage of data analysis tools within the research enterprise at institutions of higher education and if there is a correlation with research expenditures. The conceptual framework suggested that - the primary factors (Emergent Patterns) of institutional classification, administrative structure, institutional research expenditures, and proposal and award volume together with secondary factors (Independent Agents) age of participants, years of experience in research administration, gender/gender identity, current salary range, role definition and employee skillset - predict the readiness to use data analytics tools in an office of sponsored research.

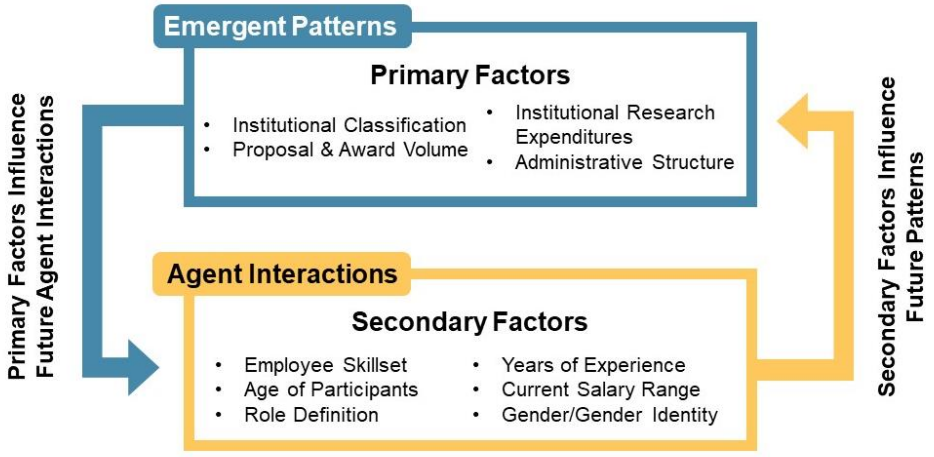


Figure 2. Figure 2 is the researcher's interpretation of CAS Theory as it relates to factors that predict the use of data analysis tools in institutions of higher education. Graphic credit: Emily Creasy of ELC Design & Creative Services.

Figure 2 provides an illustrative link of CAS Theory to this study. The **primary factors** of institutional classification, administrative structure, institutional research expenditures, and proposal and award volume represent emergent patterns or outcomes of the institution (documented emergent patterns). The **secondary factors** of age of participants, years of experience in research administration, employee skillset, role definition, gender/gender identity, and current salary range represent the agent interactions or independent agents at work within the institution. In this model, the primary factors influence future agent interactions, and secondary factors influence future emergent patterns.

Purpose Statement

The purpose of this quantitative correlation survey study was to test the CAS Theory and determine what factors predict the use of data analysis tools in an office of sponsored research in an institution of higher education within the U.S.

Research Question

What factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) predict the use of data analysis tools in an office of sponsored research in institutions of higher education within the United States? The independent variables (predictors) were defined as the age of the participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset. The dependent variable (outcome variable) was defined as the usage of data analysis tools. A secondary analysis looked at whether data analysis tool usage was correlated with institutional research expenditures as reported for the participating institutions in the 2021 NSF HERD Survey.

Subproblems of the Study

- Subproblem 1: Does the number of years of experience (YRS_EXP) in research administration differ between participants that use (USAGE) data analysis tools and participants that do not?
- Subproblem 2: Does skillset (SKILLSET) differ among role definition (ROLE_1 or ROLE_2) controlling for age (AGE) of the participant?
- Subproblem 3: How do the scores (People, Process, Data Management, and Total) differ between R1 and R2 institutional classification (CLASS)?
- Subproblem 4: Is there a difference between proposal volume (PROP_VOL) and award volume (AWARD_VOL) among institutions that use (USAGE) data analysis tools and institutions that do not?

- Subproblem 5: What is the relationship between institutional research expenditures (EXPEND_HERD) and those institutions with implementation (TOTAL_SCORE) of data analysis tools?

Hypotheses

Null Hypothesis: The factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) **are not** predictors for the use of data analysis tools **nor** a correlation with research expenditures of an office of sponsored research in an institution of higher education.

Alternative Hypothesis: The factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) **are** predictors for the use of data analysis tools **or** a correlation with research expenditures of an office of sponsored research in an institution of higher education.

Definition of Terms

Higher Education Research and Development (HERD) Survey: Referred to as the 2021 NSF HERD Survey throughout this study, the survey is an annual census of U.S. colleges and universities that expend at least $150,000 in research and development (R&D) expenditures (Gibbons, 2022) commissioned by the National Center for Science and Engineering Statistics (NCSES) and the National Science Foundation (NSF). For the purpose of this study the columns of data used were Institutional Rank and Institutional Expenditures. The last available data at the time of this research study was Fiscal Year 2021.

Office of Sponsored Research (OSR): Serves as the primary campus contact for matters related to the research enterprise including funding proposal preparation, review, and submission; review and acceptance of grant agreements and contracts; furthering relationships with government and industry; establishing and maintaining policies and trainings; and financial management and oversight of projects and programs. Offices of sponsored research can be structured as centralized, decentralized or shared service/combination.

Research Administrator: Is defined as the director-level survey respondent who promotes and facilitates administrative actions related to grant funded activities that support faculty and students involved in research projects. They are part of a group of staff persons at an institution of higher education who are responsible for the preparation, paperwork, maintenance, compliance, review or oversight of externally sponsored research activities from pre-award proposal development through post-award financial management, closeout, and audit.

Research Enterprise: Is defined as the administrative support staff (research administrators) and collective offices (department and college research support, central offices of sponsored research, offices of vice provosts for research and equivalent, and research finance) that support proposal ideation and development, funding negotiation, award acceptance and management, which the complex system together nurtures the creativity and ideation of talented faculty investigators across the university. As described by the National Research Council's Committee on Science, Engineering, and Public Policy, a "highly productive American research enterprise rests on three critical pillars, a talented and interconnected workforce, adequate and dependable resources, and world-class basic research in all major areas of science" (2014).

Data: Refers to the accumulation of quantitative information or data points throughout the lifecycle of a funding proposal and grant award. In this study it refers in particular to the number

of funding proposals submitted to external agencies, the number of awards received by the institution, proposal vs. award success rate, and subsequent institutional research expenditures. Additional data collected in research administration are the number of active research faculty, the number of students supported by grant funds, the rate of indirect cost return, the biggest prime award agencies, and the size of the research infrastructure or enterprise.

Data-informed Decision-making (DIDM): This paper focused on data-informed decision-making and acknowledges both data and human elements to support strategic decisions. For the purpose of this study, the definition of data-informed decision-making was formed by Webber & Zheng (2020) "as the process of organizing data resources, conducting data analysis, and developing data insights to provide the contexts and evidence base for formulating organizational decisions" (p. 8). The foundation of this research study lies in DIDM, which acknowledges the pivotal role of human reasoning and influence in complex, dynamic, and strategic decision-making. Nevertheless, the multifaceted landscape of higher education calls for an inclusion of software and algorithms combined with factors such as politics, human sensitives or limitations, institutional values, and timing as a nuanced process (Webber & Zheng, 2020).

Data Analysis: Refers to looking at raw data sets (with defined columns and rows) and using statistical analysis techniques (descriptive, predictive, or prescriptive) to make conclusions about the information (themes and patterns).

Data Analysis Tools: Is defined as data analysis software programs such as Microsoft Excel and Access, and more advanced software tools such as Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, and Tableau.

Skillset: Is measured in terms of numbers or frequency. Specifically, the skillset is a cumulative sum derived from four survey questions: the number of data analysis tools used, the number of

knowledge acquisition pathways or certifications, the number of reports generated, and the number of target audiences for data output or reporting.

Structure: Is defined as the staff infrastructure or organization structure of the research enterprise at an institution of higher education. A centralized structure for research administration is characterized by a hierarchical framework where decision-making authority is concentrated in one area of the institution of higher education. In this model, key administrative functions such as resource planning and allocation, policy and procedure making, and grant management are managed by a central authority, typically at the university level. Centralized administration is particularly effective in large or high research volume institutions of higher education or for those with complex operations, where a centralized authority can facilitate strategic direction and ensure alignment with overall institutional goals and research policies and procedures.

Research Expenditures: Is defined as the institutional research expenditures from all sources of sponsored research funding. Institutional research expenditures are one measurement of success of a sponsored research program at an institution of higher education. It is used to measure institutional investment and the cost of participating in research.

Delimitations of the study

The study population was recruited from a sample population of NCURA with active membership status as of March 17, 2023. The list of participants was narrowed based on the job title listed in their membership profile as director-level of a research administration office. Participants were delimited to an institution of higher education listed in Region 1 through Region 7 as defined by NCURA. The study population was further delimited by removing retired and emeritus persons, the international region, and anyone not from an institution of higher

education; there was a final potential participant recruitment list of 1,582 active members. The study was delimited to participants who had access to a mobile device or computing device (computer, iPad) that requires internet or accessibility and connectivity.

Assumptions

For the purpose of this study, it was assumed that the participants answered the survey honestly and independently. The researcher also assumed that the individuals who received the survey did not forward the survey link to other individuals to alter the anticipated study population or data results. It was also assumed that participants who had questions contacted the designated persons defined in the supplemental materials and that their questions or concerns were addressed prior to starting the survey. Lastly, the researcher assumed that the study participants understood the survey instructions and questions, which allowed for full attention and participation in the study.

Significance of the Study

The literature, in general, suggests that the implementation of data analytics can have dramatic effects on the improvement of the decision-making process (Guster & Brown, 2012; Ghasemaghaei, 2019; Borgman & Brand, 2022). Recent studies have suggested there are significant benefits for institutions of higher education to adopt best practices for using data analysis tools to support decision-making. To become a data-informed organization, an institution of higher education needs a thoughtfully designed analytics platform that empowers everyone, at all levels of the institution, to make data capture an integrated part of their day-to-day processes. Since the late 2000s, researchers have explored areas of opportunity and challenge of using data analytics to support core university functions, such as teaching, scientific research, and service, in order to deliver a better faculty, staff, and student experience.

The research administration, institution (Data Management), management (Processes), and employee (People) perspectives discussed in the literature review revealed that each has a significant opportunity to impact and be part of a strategic pathway in influencing data-informed decision-making. Therefore, it is suggested that an institutional readiness assessment could be established that informs leadership where there are gaps in implementation so the research enterprise can move toward data-informed decision-making.

This study was significant because it provided a framework and developed a tool for institutions of higher education, in particular offices of sponsored research, to assess their institutional readiness to implement data analytics tools. As a result of this study, the researcher aimed to create an assessment tool that could be used by a research enterprise to assess their own institutional readiness and to determine gaps toward building capacity to use data analytics tools. The research study was also designed to determine which institutions of higher education are using data analysis tools to support decision-making and whether there was a correlation with institutional research expenditures (as reported in the 2021 NSF HERD Survey).

While conducting the literature review, it emerged that the institution perspective (Data Management), management perspective (Processes), and employee perspective (People) were found to be in congruence. Having explored available survey tools, the researcher opted to design a customized survey that included institutional and demographic information with a readiness assessment tool. The readiness assessment section focuses on three (3) key dimensions (People, Processes, and Data Management) that literature has suggested to be vital in determining whether an institution of higher education can effectively engage in data-informed decision-making (Voorhees, 2007).

The findings of this research study are anticipated to captivate the attention of a broad spectrum of higher education professionals engage in the research enterprise, including senior leadership, directors, and staff research administrators. By delving into the current landscape of data analytics implementation in institutions of higher education with the U.S., this research bridges the gap in knowledge, augmenting the theoretical understanding of data-informed decision-making within research administration. Furthermore, the results underscore the pivotal role of fostering data literacy (and the People dimension) as a fundamental starting point in the implementation of data analytics tools.

Chapter 2: Literature Review

The field of research administration collects data on a daily basis. Therefore, it could be particularly advantageous for offices of sponsored research to collect, aggregate, and share data to inform strategic decision-making. There is a growing body of literature that highlights opportunities for institutions of higher education to improve operational efficiency and effectiveness through data-informed decision-making (DIDM). The available literature is new and informative, providing a basis for best practices, recommendations, and future research. The question still remains; How many offices of sponsored research, within the research enterprise, at institutions of higher education in the U.S. have assessed their readiness to implement data analytics tools? In recent months, studies have suggested there are significant benefits for institutions of higher education, in particular offices of sponsored research, to adopt best practices for DIDM. From current literature on the topic of research administration, best practices are focused on general themes or suggestions as a starting point including stakeholder discussions, cost-benefit analysis, analytical skillset and staffing needs, availability and location of quality data, data governance and management, and strategic use cases.

Institutions of higher education are sitting on vast amounts of valuable data stored in different departments and software systems, waiting for the silos to be broken and the dots to be connected. With the adoption of cloud computing infrastructure, higher education can be transformed to make effective and coordinated use of a wealth of data that exists and has the potential to be used to support a wholistic approach to informed decision-making (Attaran et al., 2018; Nguyen et al., 2020; Webber & Zheng, 2020).

The aim of this literature review was to move the field of research administration forward by linking theoretical insights of DIDM with best practices, thus enhancing the understanding of

and approaches to data analytics implementation. The following literature review presents concise information about opportunities for the institution, management, and employees to utilize DIDM in institutions of higher education. The research administration, institution, management, and employee perspectives highlighted in this literature review reveal that each has a significant opportunity to impact and be part of a strategic pathway toward implementation of data analytics tools at institutions of higher education within the U.S. As it concludes, the literature review refers to several opportunities and suggestions for institutional policies and next steps in order to capitalize on data analytics tools in support of better strategic decision-making.

Information referenced throughout the literature review highlights the importance of exploring the opportunities associated with implementing data analytics in institutions of higher education and presents the findings for educators, administrators, and policymakers to consider (Attaran et al., 2018). Data-informed decision-making in research administration involves leveraging insights from data already collected by the institution and using it to improve strategic planning or administrative workflows and processes. Consider a scenario where a research administration team utilizes data-informed decision-making to address inequitable workload distribution. By analyzing historical data on proposal volume, proposal due dates, award volume, team member expertise, number of active faculty investigators, sponsor complexity and requirements, the team could identify the most active departments and therefore imbalances in workload distribution. As a result, the team would strategically revise portfolios or tasks, ensuring a more equitable distribution of responsibilities.

The literature review was conducted through reviewed sources of existing literature, including scholarly journals, book chapters, and reports. Search engines of scholarly databases utilized in this literature review included: Marywood University library website, Lehigh

University library website, EBSCOhost Research Databases, Wiley Online Library, and Google Scholar. Keywords utilized in the literature review included: institutional analytics, research administration, big data, data analytics, data science, institutions of higher education, skills gap, and decision-making.

Research Administration Perspective

An Office of Sponsored Research (OSR) serves as the primary campus contact for matters related to the research enterprise including; funding proposal preparation, review and submission; review and acceptance of grant agreements and contracts; furthering relationships with government and industry; establishing and maintaining policies and trainings; and financial management and oversight of projects and programs. OSR can be structured as centralized or decentralized. A centralized structure suggests one shared staff that manages both pre- and post-award functions that service a defined set of departments or colleges within the institution. A decentralized structure suggests multiple offices on campus housed within departments or colleges that perform similar functions, and report to a centralized office only at the mid-point between proposal submission and notice of award. Little is known about which structure performs the best for supporting faculty in their research endeavors, and it's not unusual for an institution to flex between models over the course of time.

Research administrators, the name for professionals in the field, have a breadth and depth of knowledge that goes beyond general administrative and financial skills, and with potential for developing new skillsets, particularly those in data analysis. The role of a research administrator at an institution of higher education within the U.S. is quickly expanding from pre-award proposal development and post-award financial management to that of an active partner between

faculty investigator(s) and institutional leadership to support research growth (Robershaw & Wolf, 2023; Cargill, 2022).

A systematic literature review was conducted by Robershaw & Wolf in February 2023 that supports research administration and data analytics in higher education and provides evidence-based practices to inform decision-making with regard to institutional research. The results of their literature review support implementation of a data analytics program that improves the operation of an OSR and informs strategic decision-making. Research administrators are experiencing an increase in competitiveness of external and internal sponsored funding, and "it is suggested that in the near future research administrators will play a key role in the planning for strategic alignment of research resources by using advanced data analytics tools" (Robershaw & Wolf, 2023, p. 19). Likewise, their systematic literature review also revealed factors of senior leadership that may hinder the development of an organization's implementation of data analysis tools. For example, it was found that some institutional leaders are often not technology savvy, which has led to the misalignment of readiness or preparedness, and a misunderstanding of existing skillsets among their direct staff. "Analytics in research administration can provide critical insights into the research that is happening across (small and large) universities and suggest novel approaches for how to strategically align resources, support innovative ideas, and create efficient processes" (Wolf et al., 2021, p. 2). Analyzing data that an institution already has, also provides a basis for confirming the effectiveness of decisions as they align with institutional goals.

OSR capture a vast amount of data (both quantitative and qualitative) that, with the right data analysis tools, can be used to understand trends and outliers, and forecast future workload volume and infrastructure needs (Wolf et al., 2021). A requirement of the Office of Management

and Budget (OMB) Uniform Guidance, the guidebook for all federal grant regulations, is that in order to manage a grant award effectively, the office of research must capture, in an electronic database, very detailed information about each award (2014). Information captured as data points includes the principal investigator name, academic department/center, other key personnel, project title, start and end dates, proposed and awarded amount, indirect cost rate, and sponsoring agency. Ideally, with the right data analysis tools, the information captured (data input) can be synthesized into specific reporting parameters (data output) including, for example, a summary of proposal or award counts and amounts by department, agency, or principal investigator. The data output can be used by leadership in the research enterprise to determine infrastructure needs, justify budget requests, redistribute workload to be more equitable among administrators, or flex between centralized and decentralized models (Wolf et al., 2021).

Readiness Assessment

Readiness assessments provide information about the specific components of an organization or process that must be improved in order to create or promote a new direction or way of thinking. Before big data and data analysis tools can be useful to institutions of higher education, there must be a fundamental shift in thinking and repositioning in the mindset of institutional leadership and management. “Analytics technology is constantly evolving, it has changed dramatically over the years and is still advancing rapidly today” (Attaran et al., 2018, p. 5).

An article was published in *Science* in December 2022, by Christine Borgman and Amy Brand, titled “Data blind: Universities lag in capturing and exploiting data.” Findings from this qualitative interview study confirm the researchers’ Qualifying Paper (October 2022) and begins to assess the ways in which institutions of higher education are utilizing data to inform decision-making. The qualitative study sought to “identify sources of tensions and innovative solutions

adopted or under development" at institutions of higher education (Borgman & Brand, 2022, p. 1278). They interviewed twelve (12) university leaders from April through August 2021 that represent a wide-range of perspectives on data management, with roles including vice provost, vice president (or vice chancellor) for research or institutional research, university library, and chief information/technology officer. The sample was diverse by type of institution, public or private; by gender and ethnicity; and by geography, with respondents from east and west coasts of the U.S. Their interview questions addressed various aspects of data management from the participant's role in university data to what key decisions are data-informed at their institution, which led to wide-ranging discussions that addressed data, decisions, strategies, and concerns. The article was written to highlight two parts: urgent challenges and lessons learned.

Of particular interest in the findings of this study, was that Provosts suggested they "could make better strategic hiring and curricular decisions if they had access to comprehensive data on faculty research areas, career interests of prospective students, research funding patterns, higher-education policy trends, and competitive-type information about other institution initiatives" (Borgman & Brand, 2022, p. 1279). Echoed across almost all institutions they interviewed was the difficulty in integrating data and systems, governance, and practices across campus. Some integration issues were technical and limited to the software or systems, while others were conceptual and maligned agreement of leadership on definitions and elements. Almost all of the institutions interviewed indicated that they did not have a coordinated university approach to data management, and that the high cost (and sometimes hidden costs) of software and cloud hosting was a barrier. The study concluded that academic leaders have legitimate concerns about economic constraints and a lack of data expertise, among other challenges. "Even when their institutions are "data rich" they may also be "data poor" in that they are struggling to utilize data

resources to their strategic advantage, or “data-blind” in being reluctant to initiate stakeholder discussions necessary to build consensus for data governance” (Borgman & Brand, 2022, p. 1280; Webber & Zheng, 2020, p. 5).

Similarly, a study completed in two parts by Wolf & Hall (2019 & 2021) aimed to provide critical insights to the use of data analytics in the research enterprise. First, by analyzing results of a quantitative survey conducted in spring of 2019, and then by following up with qualitative interviews to gain additional insights into how various institutions are using analytics tools in their research offices. The outcomes of the research include a list of best practices from both the quantitative and qualitative data collection. The quantitative survey was distributed using a snowball sampling methodology and recruited participants from community forums and a listserv. The survey instrument included items covering a variety of topics and “used the Rasch Measurement Model” (Wolf et al., 2021, p. 5). There was a total of fifty-six (56) respondents who were either central research administration or departmental level staff from a total of forty-eight (48) unique institutions. Twelve (12) follow-up qualitative interviews were conducted, and respondents were selected based on criteria that indicated in the survey that they were already capturing and using basic data elements of research activity to support decision-making. Even though the response rate was low, the groups were geographically representative, having a wide range of roles in the research enterprise, and a varied level of institutional size based on research expenditures. Findings from this two-part study revealed challenges in leveraging data analytics and a difficulty in finding an audience that perceived data to be valuable with regard to decision-making. Another challenge of the two-part study was finding participants with the right knowledge and skillset who perform tasks related to data management and interpretation. The study findings were consistent with those of discussions at professional association meetings of

the Society of Research Administrators International (SRAI) and the National Council of University Research Administrators (NCURA) and the researcher's own Qualifying Paper (October 2022).

Institution Perspective

Institutions of higher education are operating in an increasingly complex and competitive environment. They are under mounting pressure to respond to national and global economic, political and social changes, such as the growing need to increase the proportion and diversification of students in certain disciplines, increase scientific research capacity and output, and ensure that the quality of learning programs is both nationally and globally relevant (Daniel, 2015; Nguyen et al., 2020; Webber & Zheng, 2020). Institutions of higher education have also been severely affected in recent years by a number of unfavorable happenings, which include but are not limited to, a decline in matriculating high school graduates, an unstable economy, a decrease in federal and state funding, a refocus on STEM career training, stricter F-1 visa approvals for international students, and more recently, the COVID-19 pandemic. Institutions have been forced to reevaluate the way strategic decisions are made in order to face the unknowns of the future and the realities that have collectively reshaped higher education (Chan & Randall, 2021; Ueland et al., 2021; Borgman & Brand, 2022). To navigate these challenges, savvy leaders have an opportunity to leverage data to make sound decisions and steer their institutions through uncertainty.

Using data to inform decision-making and strategic planning is not new. Business organizations have been storing and analyzing large volumes of data since the arrival of data warehouse systems in the early 1990s. However, the availability of data and the various ways data can now be collected are ever-evolving with the rise of cell phone technology, artificial

intelligence and algorithms, and machine learning, to name a few. The creation and evolution of new technologies bring with them complexity in managing the volume of the collection, storage, privacy, and analysis of these datasets. Since the late 2000s, researchers have explored areas of opportunity (and challenge) in using big data and data analytics to support core university functions and better decision-making. Institutions of higher education within the U.S. should recognize the importance of rapidly adapting and scaling up to "new Fourth Industrial Revolution forms of education to assure the sustainability of our environment and economy, as well as to sustain the relevance of higher education as a responsive and vital component of society" (Penprase, 2018, p. 225).

With large volumes of both academic and scientific research data, institutions of higher education have the data sets needed to benefit from the results of using data analytics tools. Adapting business intelligence techniques from industry can have the potential to alter and significantly improve existing processes of administration, teaching, scientific research, and academic service (Daniel, 2015; Nguyen et al., 2020). Business intelligence tools and data analytics are more than just methods of gathering and analyzing data. They represent a shift in mindset and adopting the role of the experimenter – a willingness to let data inform the strategic decision-making process (Attaran et al., 2018). From an organizational perspective, it is well understood that institutional effectiveness and adaptation to change relies on the analysis of appropriate data, and today's technologies, with advances in sophistication, speed, and accuracy, enable institutions to gain insights from data that was previously unachievable (Daniel, 2015).

The term data analytics refers to looking at raw data and using statistical analysis techniques to make conclusions about the information (themes and patterns). First proposed in the early 1990s and growing over the years, "clustering, association, classification algorithms, regression

models, predictive methods, and factor analysis have become the key approaches for data analysis" (Daniel, 2015, p. 907).

Descriptive analytics is the simplest of the three categories. It allows large amounts of data to be condensed into smaller, more useful pieces of information. Its purpose is to summarize what happened in the past and to uncover patterns that may offer insights into the performance or health of an organization. This enables administrators and managers to monitor their business processes more effectively and efficiently. In descriptive analytics, data modeling, reporting, visualization, and regression are used to collect and store data. Allowing the user access to a dashboard that creates and presents a visual representation of information that can be used to identify trends or patterns in the data (Attaran et al., 2018).

Predictive analytics uses current and historical data to provide insights into what will happen and why it will happen with an acceptable level of reliability. It involves the use of a variety of models and techniques to anticipate future conditions and situations. It does not predict one possible future but rather multiple possible futures based on the decision-maker's actions. Predictive analytics can help organizations analyze historical data and facts to improve their understanding of a wide range of issues (Attaran et al., 2018; Webber & Zheng, 2020). In the book written by Webber & Zheng (2020) they also cite the work of Ekowo & Palmer (2017) regarding predictive analytics where a framework is offered "to examine how predictive analytics ought to be an iterative process whereby as institutional leaders use data and information in new ways, they will need to regularly assess whether ethical standards are maintained and address current data practices. The framework includes five guiding ethical practices for use of predictive analytics" (Webber & Zheng, 2020, p. 71).

Prescriptive analytics goes beyond the descriptive and predictive models and shows the likely outcome of each decision. It goes a step further into the future and attempts to identify what should be done and why. Prescriptive analytics employs techniques such as decision modeling, simulation, and optimization to ascertain actions the organization could take to achieve the desired outcome. The aim is to evaluate the effect of future decisions and to present the best course of action to take in order to adjust decisions before they are made. This is the most valuable category of analytics and usually results in rules and recommendations for the next steps (Attaran et al., 2018).

The decisions required for managing rapid changes in the higher education environment are multifaceted. It is suggested that many important and strategic decisions in higher education in the U.S. are made without access to data dashboards or analytical tools. “These data can play a major part in how we understand the often-contested nature of higher education governance and so ensure that institutions are not only able to respond effectively to changes happening within and outside them but that they also remain pertinent to their purpose in the [societies] that they serve” (Daniel, 2015, p. 904).

A review of current literature suggests that a notable opportunity exists for institutions of higher education to adopt data analytics tools for reporting by first establishing data warehouses and creating data dashboards. These tools would provide an institution with the capability to make timely data-informed decisions across all departments (Daniel, 2015). It is important to acknowledge that the literature review also revealed a number of challenges that exist within institutions of higher education that make adapting new technologies a complex undertaking. To become a data-informed organization, a university needs a “thoughtfully designed analytics platform that empowers everyone to make data an integrated part of their day-to-day processes”

(Attaran et al., 2018, p. 9). And an institution of higher education needs an analytics solution, specifically, that can bring together disparate data in a governed environment that allows users from different departments to model, discover, communicate and distribute information easily. (Attaran et al., 2018; Borgman & Brand, 2022).

Management Perspective

Before advancements in data analytics, strategic decisions in higher education were made according to the experience of leadership and management. Data was manually collected, often in excel spreadsheets, resulting in obvious (or non-obvious) flaws. In the age of big data, and with the accessibility of business intelligence tools, it is now much more efficient to collect, store, mine, and analyze data as well as visualize, monitor, track, and predict outcomes over a time period. The types of data generated by institutions of higher education include teaching, student, scientific research, daily management or programmatic, and other types of service to the university (Wang, 2019).

Similarly, the types of questions being asked by leadership are becoming more complex. They involved not only research activity data, but human resource and other university financial data to answer questions such as “are full professors more or less successful at winning external grant awards than junior faculty,” and “if we increase the number of awards how does that affect research expenditures and staff workload,” and “where should we focus our efforts on building interdisciplinary research teams.” With data requests becoming more complex and dependent on other areas, institutions of higher education are finding opportunities to adopt data analysis tools that facilitate decision-making and strategic planning (Wolf et al., 2021). Recent studies have found that some institutions are creating dashboards that visualize data and make it more accessible to deans, their administrative support teams, and other users across campus.

Institutional leadership at all levels are using data to identify issues and support decision-making (Wolf et al., 2016, p. 2).

The literature, in general, suggests that full implementation of data analytics tools can have dramatic effects on the improvement of the decision-making process for managers (Guster & Brown, 2012; Ghasemaghaei, 2019; Borgman & Brand, 2022). A data governance structure is a critical component of adopting data analysis tools, which includes the establishment of clear ownership of the data/processes that generate the data and a clear understanding of who uses the information and the purpose of its use. These governance principles should be driven by the use of a communication model, not necessarily the technology (Guster & Brown, 2012; Webber & Zheng, 2020). However, the old adage "garbage in, garbage out" still applies when using sophisticated data analysis tools. "The data analytics technology platform should provide a unified and trusted view of the [institution], empowering all employees with insight and aligning with the organization's operational strategy" (Guster & Brown, 2012).

The accumulation of data should not be an end in itself, it should be valued as a means to an end. Therefore, there needs to be an end goal of transforming the data into useful and relevant information that supports the human thoughts, beliefs, and culture of the leadership and the institution. This is done by getting behind the data and bringing managerial experience into the information at hand (D'Auria Stanton & Stanton, 2019). A 2018 research study conducted by Maryam Ghasemaghaei, hypothesized that relationships existed between data analytics use, knowledge sharing, data analytics competency, and decision-making quality. In this way, decision-making quality refers to the correctness and accuracy of decisions for desired outcomes. This study found that decision-making quality improves if the decision-maker has sufficient knowledge about problem variables. However, if the decision-maker does not have the required

knowledge about the relationships among problem variables, the quality of the decision may decrease. As a conclusion of the study, it was suggested that decision quality depends on the inputs (Ghasemaghaei, 2019). Likewise, it is critical to be aware of administrative functions and how much or how little they interact and influence each other, and in cases where we are not certain of all of the independent agents at play (Seaman, 2021).

Ghasemaghaei (2019) found that data analytics significantly improves knowledge sharing, and knowledge sharing increases firm decision-making quality. The study also found that the impact of knowledge sharing on firm decision-making quality is at its highest when employee analytics capability (skillset) is at its highest level. More importantly, the impact of knowledge sharing on decision-making quality is not significant at low levels of analytics capability (Ghasemaghaei, 2019). In general, knowledge sharing within a firm does not necessarily enhance the quality of the decisions made in the firm. One way data analytics competency plays a vital role in the impact of knowledge sharing on decision-making quality is by learning to use data analytics tools (Ghasemaghaei, 2019).

In a book edited by Nancy Gleason (2018), the author uses a lens of the Fourth Industrial Revolution to look at what is happening within institutions of higher education and why (Penprase, 2018). The first three industrial revolutions provided evidence for profound shifts in society, the economy, and education which resulted in the creation of co-curricular innovation and the establishment of new educational institutions (Penprase, 2018). A review of this book suggested that in the Fourth Industrial Revolution the shelf life of any skill in the present-day working environment has become increasingly short, requiring future workers to continuously update their skills and teach themselves about new technologies and new industries that may not have existed while they were being trained for their initial degrees or roles. “One requirement for

management in the Fourth Industrial Revolution would be to include a strong overlay of ethical thinking, intercultural awareness, and critical thinking to enable the thoughtful and informed application of new and developing technologies" (Penprase, 2018, p. 220).

Employees in institutions of higher education need to adapt and respond to the increasing rate of change and the increasing complexity of employment (Penprase, 2018). The study results from Ghasemaghaei (2019) reiterate the importance of training employees and developing sufficient knowledge-sharing mechanisms, and encouraging the adoption and use of data analysis tools (e.g., Tableau, SPSS). "Future jobs that require skills to perform tasks related to data science, such as artificial intelligence (AI), machine learning, robotics, nanotechnology, 3D printing, genetics, and biotechnology, are expected to dominate in the coming decades. Employers and industries are projecting that social skills, which include persuasion, emotional intelligence, and capacity for teaching others, will also be at a premium" (Penprase, 2018, p. 220).

Employee Perspective

Institutions of higher education recognize that their mission is not only focused on preparing students for an economically relevant career but also on helping train existing members of the workforce to gain the skills needed to succeed in a quickly changing workplace. Employees need "opportunities to continue to gain the skills required for a productive career, both for themselves and the institutions for which they work" (Zink et. al., 2022). Many institutions of higher education are experiencing a shift in mindset and are investing in dashboards that visualize data and make it more accessible to deans, their administrative support teams, and other users across campus. As a result, a number of institutions are also exploring new employee recruitment

strategies that include reskilling current staff or bringing in promising candidates and then training them in the upskill needed to perform data capture and analysis.

Closing the skills gap by finding or developing data science talent is a challenge, but not unattainable (D'Auria Stanton & Stanton, 2019). In a future state within the world of higher education, students, administrative staff, and faculty will never be done with their education but instead will continue to engage in learning from their colleagues and outside experts to frequently renew and update their skills (Penprase, 2018; Wolf et al., 2020).

As mentioned in the literature, "functional knowledge is necessary to understand and interpret data accurately… you need someone with data expertise to easily find the themes, patterns, and stories within the data. Programming skills could help maintain and troubleshoot software, and excellent communication and visualization skills could help interpret the data and turn it into actionable items" (Wolf et al., 2021, p. 15). Likewise, although data are invaluable and critical sources of insight for institutions of higher education, they do not and could not replace operational decision-making processes that involve human judgment, political sensitivity, and ethical considerations (Webber & Zheng, 2020).

Employees, especially those who have made a career in higher education, are the best source of knowledge. In fact, knowledge resides within individuals, and in particular, in the employees that generate and apply knowledge in performing their daily tasks (Ghasemaghaei, 2019; Zink, et. al., 2022). While the literature is not specifically focused on retention and voluntary turnover intentions of research administrators within institutions of higher education, the literature reviewed for this paper provides a foundation for the topic. Studies have shown that, in general, higher-skilled employees are more likely to stay at an organization if their job duties include new challenges and opportunities to learn. Additionally, recent research has suggested that

opportunities for career growth and professional development are crucial motivating factors for employee retention (Ghasemaghaei, 2019; Zink, et. al., 2022).

Borrowing an example of a case study from "The Dean's Information Challenge: From Data to Dashboard" written by Wolf et al. (2016), George Washington University is one institution that recognizes that its administrative staff has the ability to act as both artisanal decision support and as data stewards for the university. They define the business terms and appropriate data usage in a data analytics dashboard and help to create the business processes. This helps keep the staff involved in the decision-making process and provides them the opportunity to learn new skills. The data stewards retain some of the benefits of the artisanal decision support method by ensuring the data is used in the right context, with the right business rules, and furthering the institution's mission (Wolf et al., 2016). As a result, the central data warehouse staff spends less time gathering data and more time analyzing results and information.

Institutions of higher education need analytics professionals who have the experience, skills, knowledge, and abilities to comb through seemingly endless data to isolate the most pertinent, actionable findings. Data analytics requires professionals who are good storytellers and can translate their findings into specific strategic recommendations (D'Auria Stanton & Stanton, 2019; Webber & Zheng, 2020; Borgman & Brand, 2022). Among all of the literature was an acknowledgment that data analytics applied to large and complex data sets requires a new type of employee, someone who is fully grounded in an administrative process but who also has a breadth of analytical and interpersonal skills (D'Auria Stanton & Stanton, 2019).

Regardless of the institutional structure (centralized, decentralized, shared service, consulting, functional, or hybrid) and assuming an institution has a clear idea as to where it would like to go with data analytics usage, the key challenge is how best to close the data

analytics skills gap (D'Auria Stanton & Stanton, 2019). While traditional recruitment strategies will continue, the rapidly changing analytics marketplace will challenge organizations to find newer and more creative approaches to talent acquisition. To address this, the literature suggests that a number of institutions have adopted a recruitment strategy that brings in promising candidates and then trains them in the upskill needed to perform data analysis. This creates an opportunity and pathway for candidates to "gain the skills required through non-traditional educational means such as massive open online courses, boot camps, and online training leads to certificates or certifications in data analytics tools and methods" (D'Auria Stanton & Stanton, 2019, p. 179).

Some of the literature suggests that skills needed for data analytics are changing research administration job functions for existing employees within organizations more than any other skill type. The research enterprise within institutions of higher education is viewing analytics and employees who can create and use them as essential for creating value (D'Auria Stanton & Stanton, 2019; Zink, et. al., 2022; Wolf, 2022). Moreover, the assumption to recruit only external candidates for data analytics roles may overlook bright, dedicated, and motivated individuals currently employed by the organization. Internal candidates already know the institution, and the institution knows them. Instead of spending time and resources to search for external candidates, a growing trend is to focus on training, reskilling and upskilling internal employees (D'Auria Stanton & Stanton, 2019).

Upskilling improves the retention of valuable employees, boosts morale, and attracts new talent. Research has shown that employees who receive development opportunities continue to make a commitment to lifelong learning and training. Institutions of higher education are

increasingly finding ways to upskill existing staff to achieve their strategic needs and goals rather than recruiting and training new staff (D'Auria Stanton & Stanton, 2019).

While precise skillsets vary among different industries, organizations, and advertised positions, it is generally agreed that the skillsets for analytics professionals fall into three broad categories: hard skills (technical skills/competencies), soft skills (personal traits or characteristics), and credentials. Hard skills can include software development or programming, the ability to use a variety of analytical and statistical modeling tools, and a general understanding of quantitative datasets. Soft skills include the ability to communicate and present analytical insights in an understandable and compelling way, collaborate and work in teams, be creative and innovative, and the ability to solve complex problems. Credentials that candidates are able to claim at this point in time has a much broader definition, not only degree(s) earned but also work experience, one-day workshops, and online certifications (D'Auria Stanton & Stanton, 2019). Any institution seeking to fill a newly created data analytics role must first ask themselves what complimentary skillsets they are looking for in addition to degrees earned and work experience.

Ethical Implications

Data Security and Privacy (Justice)

The perspectives highlighted in this literature review and the themes revealed throughout show there is a significant opportunity to further the use of data analytics to support decision-making in institutions of higher education within the U.S. However, institutions of higher education have a lot to consider in order to ensure that institutional data (broadly defined) is used responsibly. Institutional leadership, administrators, and even students may be eager for data to be collected and to be used in new ways to support strategic decision-making. Nevertheless,

excitement about new tools and methods should not overshadow the need to make sure predictive tools are deployed in a purposeful and secure manner, the right infrastructure is in place to take ownership and governance, and that it does not further entrench existing institutional inequities (Ekowo & Palmer, 2017; Borgman & Brand, 2022; Webber & Zheng, 2020).

There is an obvious ethical implication of data collection, security, and privacy that should be considered. The collection and storage of large amounts of data pose a serious challenge in protecting individual privacy. Such challenges have emerged because technology has become so advanced that laws, policies, and procedures are still catching up, and databases require continued maintenance and review. “Future work will involve identifying and establishing policies that specify who is accountable for various aspects of institutional data and information, including its accuracy, accessibility, consistency, completeness, and maintenance” (Daniel, 2015, p. 917). Future work in this area will also include establishing contracts and agreements that define legal terminology and risk tolerance for when and how data can be shared or transferred. An institution that uses data to support decision-making must make an active and ongoing commitment to keep data secure and to keep up with new government laws, policies, and best practices as they are established.

A report published in 2016 titled *“The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis”* provides a framework of conversation starters in order to assess whether institutional ethical standards are ready to meet current data privacy regulations. The report also examined how institutions of higher education are using predictive analytics and outlined the challenges they face in ensuring that they are doing so ethically. The framework, with regard to data security and privacy, outlines important reflection topics to consider, which

include; developing a team for planning and establishing a vision for data use and how success will be measured, discussing the possibility of any unintended consequences and how they will be mitigated or addressed, and creating policies and providing regular training to technical support and other staff about how to keep data secure (Ekowo & Palmer, 2016).

Bias in Human Decision-making (Autonomy)

As mentioned in the literature, data-informed decision-making "involves top-down commitment and bottom-up support…with a strong foundation of leaders who support and facilitate organizational programs and procedures that develop a community of analytical talent…in support of the institution's mission and vision for the future" (Webber & Zheng, 2020, p. 25). Research shows that personal biases exist when strategic decisions are made less frequently and without the support of external information such as data collection and analysis. Although personal biases exist, using institutional data dashboards has the potential to lessen the negative effects of decision-making and the actions that follow in carrying out an initiative. In this way, data dashboards and reports used to support decision-making should be carefully reviewed before being acted upon to avoid bias and further entrench existing inequities. Leadership and administrative staff, along with faculty, should be trained on how implicit bias, combined with the limitations of data, can impact how decisions are made. With the proper training, staff will be able to embrace the opportunity to use institutional data (broadly defined) to move their institutions forward (Ekowo & Palmer, 2017). Even when institutional data, predictive models, algorithms, institutional practices, and training are as good as they can be, mistakes can be made when acting on the information (Ekowo & Palmer, 2017).

Some research has suggested that implicit bias may be heightened with predictive systems because analytics may serve to confirm bias or make implicit bias even more invisible. For

example, using predictive analytics could lead to removing human judgment from decision-making. This may result in decisions that typically require holistic review becoming partly or completely automated using data alone (Ekowo & Palmer, 2017). In other words, leadership and administrators should be aware when decisions need to consider a human aspect that the numbers don't capture.

Policy Recommendations

The research administration, institution, management, and employee perspectives highlighted in this literature review reveal that each has a significant opportunity to impact and be part of a strategic pathway toward implementation. As a result, it is suggested there are several opportunities and suggestions for institutional policies and next steps that capitalize on data analytics tools to support strategic decision-making.

From the perspective of both research administration and the institution, the first recommendation is to create opportunities for discussion and explore the institutional priorities through knowledge sharing. A conversation could begin if leadership makes a commitment to learning more about using data analytics in higher education and then prioritizes sharing the knowledge with other leadership and management. The research administrators, together with the institution, should do a broad review and collect qualitative current state information from key administrative areas on campus in order to determine how decisions are currently made, including an observation of how the leadership or management perceives the effectiveness of the outcomes (without data analysis as support).

The next step would be to establish a committee and perform an institutional data analytics readiness assessment. The readiness assessment tool is organized around key dimensions that literature has suggested to be vital in determining whether an institution of higher education can

effectively implement data analytics tools. The results of the assessment intend to act as a tool for review and discussion and to help institutions determine whether observable practices are helping or hindering the end goal. Developing and pilot testing a data readiness assessment tool with a sample of institutions of higher education provides the next step in filling a gap in the literature. A data readiness assessment could reveal a model for predicting how and when an institution is ready to implement data analytics tools and what is needed in order to be able to move forward in this initiative.

After a data readiness assessment, an institution of higher education needs a thoughtfully designed analytics platform that empowers everyone, at all levels of the institution, to make data capture an integrated part of their day-to-day processes. From the management perspective, the first recommendation is to centralize data capture and develop data capture policies and procedures. A data governance structure is a critical component of adopting data analysis tools, which includes the establishment of clear ownership of the data/processes that generate the data and a clear understanding of who uses the information and the purpose of its use (Guster & Brown, 2012; Borgman & Brand, 2022). Leadership should identify and establish clear guidance (both policies and procedures) that specify who is accountable for various portions or aspects of institutional data and information, including its accuracy, accessibility, consistency, completeness, and maintenance.

Lastly, jobs that require data literacy and skills to perform tasks related to data analysis are expected to grow significantly over the next decade. Moreover, employees need opportunities to gain new skills not only for themselves for a productive career but also for the changing workplace in which they work. Recognizing the complexity of the full implementation of data analytics tools, it is acknowledged that accomplishing this task is no simple feat. Instituting a

successful adoption of data analytics tools in higher education institutions necessitates the assembly of adept leaders, investing in dynamic software and other technologies, formulating transparent data governance, policies, and procedures, and fostering incentives for continuous employee skill development.

Conclusion of the Literature Review

Data analytics can be transformative for the research enterprise within institutions of higher education and support better decision-making in these increasingly competitive times. The objective of this literature review was to advance the field by filling a gap and presenting concise literature around opportunities for the research administrator, institution, management, and employee to use data to inform decision-making in institutions of higher education within the U.S. The perspectives highlighted in the literature review revealed that each has a significant opportunity to impact and be part of a strategic pathway in furthering the use of data analytics to support decision-making. The available literature is new and informative, providing a basis for future research and policy recommendations. Moreover, in general, it is suggested that the implementation of data analytics can have dramatic effects on the improvement of the decision-making process (Guster & Brown, 2012; Ghasemaghaei, 2019; Borgman & Brand, 2022).

There is a crucial need to reposition data analytics within the mindset of research administration leaders responsible for managing day-to-day tasks. Likewise, a robust data governance structure and comprehensive data literacy play a pivotal role in the effective adoption of data analysis tools. This entails establishing clear ownership of the data and the processes generating it, as well as fostering a transparent understanding of the individuals utilizing the information and the intended purpose behind its use.

Institutions of higher education recognize that their mission is not only focused on preparing students for an economically relevant career but also on helping train existing members of the workforce to gain the skills needed to succeed in a quickly changing workplace. “In a future state of higher education, students, administrative staff, and faculty will never be done with their education but instead will continue to learn from their colleagues and outside experts to frequently update their skills” (Penprase, 2018).

The perspectives highlighted in this literature review show there is a significant opportunity to further the use of data analytics to inform decision-making in institutions of higher education within the United States. However, the institution leadership has a lot to consider in order to ensure that institutional data (broadly defined) is used responsibly. There is an obvious ethical implication of data collection, security, and privacy that should be considered. Collecting and storing large amounts of data pose a serious challenge in protecting individual privacy. An institution that uses data to support decision-making must make an ongoing commitment to keep abreast of new government laws, policies, and best practices as they are established.

In conclusion, there are several recommendations for policies and next steps, which are 1) create opportunities for discussion and interest through knowledge sharing, 2) perform a broad review and collect qualitative and quantitative current state information, 3) establish a committee and perform an institutional data readiness assessment, 4) discuss and define the direction of where the organization would like to go in terms of using data analytics tools, 5) centralize data capture and develop data capture policies and procedures, 6) ensure data capture is clean and consistent and develop a data dictionary, and 7) consider and adopt new employee recruitment strategies which include data literacy – reskilling existing employees or making a commitment to upskill potential employees.

While conducting the literature review, it emerged that the institution perspective (Data Management), management perspective (Processes), and employee perspective (People) were found to be in congruence, which led to the design of this study. Therefore, the researcher designed a customized survey tool that combined institutional and demographic information with a readiness assessment tool. The readiness assessment section focuses on three (3) key dimensions (People, Processes, and Data Management) that the literature has suggested to be vital in determining whether an institution of higher education can effectively engage in data-informed decision-making (Voorhees, 2007). The perspectives discussed in the literature review informed the research variables, survey questions, and data collection and analysis.

Chapter 3: Methodology

Introduction

Research into the use of data analytics tools in higher education has made significant strides in recent years, contributing valuable insights to the academic community. However, an examination of the existing literature revealed a noticeable gap in how and when institutions of higher education use data analytics tools to make data-informed decisions. This research study utilized a cross-sectional survey design to collect data using Qualtrics. The survey method was chosen because it was the most efficient way to obtain the proper categories of data and recruit participants from regions in the United States as defined by the National Council of University Research Administrators (NCURA). For this study, data was captured using a survey at a single point in time to minimize internal factors, such as director-level promotions and job changes, and external factors, such as social or global influences, which may have impacted participant responses and final results.

The dependent variable, data usage, was a dichotomous variable that described those who use data analysis tools and those who do not. The independent variables (predictors) were defined as the age of the participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset.

Participants / Study Population

Active NCURA members were selected as the target population to recruit participants for this study. This professional organization had a total membership of 8,200 research administrators (as of July 10, 2023) in seven regions across the U.S. (Appendix H). Region 1 represents the six New England states of Connecticut, Maine, Massachusetts, New Hampshire,

Rhode Island, and Vermont. Region 2 represents the seven Mid-Atlantic states of New York, New Jersey, Pennsylvania, Delaware, Maryland, Washington D.C., and West Virginia. Region 3 represents eleven Southeast states of the U.S. (Alabama, Arkansas, Florida, Georgia, Kentucky, Louisiana, Mississippi, North Carolina, South Carolina, Tennessee, and Virginia), Puerto Rico, and the U.S. Virgin Islands. Region 4 represents twelve Mid-American states including Illinois, Indiana, Iowa, Kansas, Michigan, Minnesota, Missouri, Nebraska, North Dakota, Ohio, South Dakota, and Wisconsin. Region 5 represents two states, Texas and Oklahoma. Region 6 represents the Northwest states of California, Alaska, Hawaii, Nevada, Oregon and Washington, as well as Northern Mariana and Guam. Region 7 represents the Rocky Mountain region of the U.S. including Arizona, Colorado, Idaho, Montana, New Mexico, Utah, and Wyoming. The membership directory captures a person's name, email address, current and past employment information, location, associated region, and job title.

The study population was selected from the NCURA membership directory with active membership status as of March 17, 2023. A purposeful sampling method was used to aggregate a pool of director-level participants to take part in the survey. The list of potential participants was narrowed based on the job title listed in their membership directory profile as director-level (Assistant Director, Associate Director, Director, Interim Director, Senior Director, Senior Associate Director, and Executive Director) of a research administration office. After narrowing the list further by removing retired and emeritus directors, the international region, and anyone not from an institution of higher education, there was a final list of 1,582 members that met the study criteria. An email was sent on March 22, 2023 to the NCURA Membership Director requesting permission to gain the study population email addresses and to send an announcement to the potential participants. Permission was received from the NCURA to contact members of

the professional organization via email and on NCURA Collaborate (Appendix F). It was anticipated that participants in this study would not experience any greater risk than their current daily living activities and no personally identifiable information was collected.

Inclusion criteria for this study scope encompassed individuals who had an active membership with the NCURA as of March 17, 2023 and listed in the membership directory as director-level of an office within the research enterprise. Inclusion criteria also encompassed the participants' current employment at an institution of higher education listed in Region 1 through Region 7. Inclusion criteria allowed for all gender identities, individuals 18 years of age or older, and all races that had registered their email addresses with the NCURA membership directory. The study was limited to participants who had access to a mobile device or computing device (computer, iPad) which requires internet or accessibility and connectivity.

Exclusion criteria were applied to refine the participant pool in this survey both during data collection and data analysis. Individuals with an active listing in the NCURA membership directory but classified as either retired or emeritus were excluded from the recruitment list. Similarly, participants indicating employment at organizations other than institutions of higher education were excluded. Individuals with an active listing in the NCURA membership directory but whose email address bounced or failed to send were also excluded from the final recruitment list. The International region of NCURA was also excluded for the purpose of this study scope. Additionally, during data analysis, records of participants who opened the survey but did not complete any questions were excluded to ensure a focused and meaningful data set.

The Survey Instrument

A 15-question survey was designed and used to collect data on participant demographics, including their institution information related to sponsored research administration and general

participant demographics, their usage of data analytics tools, and the institutions' readiness to use data analytics tools based on a 5-point Likert Scale. The survey instrument was developed from a combination of the researcher's own questions, inferred questions about skillsets from a Parks & Recreation survey tool (Roth, 2016), and three adapted matrix tables from an *Institutional Data Readiness Assessment Tool* (Voorhees, 2007). Qualtrics was used to administer the internet-based survey to the NCURA sample via mass email distribution and a posting on the NCURA Collaborate site, a professional networking platform.

Survey questions 1. – 12. were developed by the researcher and were intended to collect demographic information (Appendix A). The first section included questions about the participants' current employer institution of higher education, institutional classification, current title, primary role in the research enterprise, and institutional proposal and award volume. It also included demographic questions related to the age range of the participant, gender/gender identity, current salary range, and number of years of experience working in research administration. The final question of the first section asked if the participants' current position description or job duties include data aggregation, analysis, or reporting for decision-making.

It is important to note that survey participants who selected "Yes, and it is part of my position description" or "Yes, and it is not part of my position description" to question 12., *"Does your position description or current job duties include data aggregation, analysis, or reporting for interdepartmental (inter-college) decision-making? Data could be used for a number of things internal to your sponsored research office(s) including, but not limited to, strategic planning, infrastructure or staffing decisions, proposal development or center/institute focus,"* were shown five additional sub-questions, and the matrix tables 13. People statements, 14. Processes statements, and 15. Data Management statements. Whereas those who selected "No" or "I'm Not

Sure" to question 12. were shown two additional sub-questions and the matrix tables 13. People statements, 14. Processes statements, and 15. Data Management statements.

The literature review uncovered an article written by Dr. Kevin Roth, Ph.D. (2016) that reported how parks and recreation agencies used survey data to make operational decisions. This informed the conditional sub-questions (question 12.a. – 12.g.) and as a professional courtesy, permission was requested in March 2023 from Dr. Roth to view the survey questions from his study. However, in the absence of a response, the researcher incorporated skillset questions inspired by the survey results from the parks and recreation study and acknowledged the prior research. The conditional sub-questions explored the participants' frequency of using data analytics tools, acquisition of skills, use cases for data analytics reporting, and their main audience for reporting or result output.

A search for a data readiness survey tool revealed one created and validated by Dr. Richard Voorhees (2007) titled *Institutional Data Readiness Assessment Tool*. The last section of the survey was three (3) matrix tables (question 13., question 14., and question 15.) The researcher reviewed the assessment tool and modified the statements for the research administrator audience. From the literature review, it was concluded that three (3) dimensions of People, Processes, and Data Management played integral roles in the effectiveness of data-informed decision-making within institutions of higher education. Each component contributed to creating a robust framework that ensured accurate, reliable, and timely information was leveraged to make informed decisions. Adapted from the assessment tool designed by Dr. Voorhees, the **People** dimension referred to the expertise, receptivity, and commitment to using data among administrative staff. Skilled individuals are essential for interpreting and utilizing data effectively. Decision-makers, in particular those at a director level, need to have a strong

understanding of the context in which decisions are supported, the data that is available, and the implications of different courses of action. Employees in director level roles were selected for this study because they should possess a level of data literacy to comprehend, interpret, and communicate insights effectively. This ensures that decision-makers are not reliant solely on data analytics tools, but also on their professional experience to be able to actively engage with and ask questions of the data.

The **Processes** dimension explores the interaction between people and the procedures necessary to ensure that data are accessible or shared widely and policies are in place to produce information that can be used. Establishing structured processes for collecting and integrating data from various sources is crucial. It ensures that administrative decision-makers have access to and are working with reliable information, reducing the risk of supporting decisions using flawed data. Implementing robust data governance processes ensures that data is managed, protected, and utilized in a compliant and ethical manner. This includes defining roles and responsibilities, setting data standards, and ensuring data security.

The **Data Management** dimension refers to how the institution stores, retrieves, and manages information (Voorhees, 2007). Effective data management involves employing tools and technologies for data analysis and visualization. This allows decision-makers to extract meaningful insights from the data, making it easier to understand complex patterns and trends. As data volumes grow and regulations change, the data management system should be scalable and flexible. This ensures that institutions of higher education can adapt to new data sources and evolving analytical needs.

People bring cognitive and analytical skills, Processes provide the structured framework for data collection and utilization, and Data Management ensures the availability, quality, and

accessibility of data. Together, these three (3) dimensions form a foundation for data-informed decision-making, enabling institutions to respond more effectively to challenges, identify opportunities, and gain a competitive advantage in today's higher education environment.

For this study, the *Institutional Data Readiness Assessment Tool* was adapted specifically for offices of sponsored research, and the framework was used for three matrix tables 13. People (10 statements), 14. Processes (12 statements), and 15. Data Management (13 statements). A 5-point Likert Scale of 0 – 4 was used for each set of statements (0 = No implementation, 1 = Under discussion, 2 = Marginal implementation, 3 = Partial implementation, 4 = Full implementation) whereas No Implementation was interpreted as "Strongly Disagree" and Full Implementation was interpreted as "Strongly Agree." For the purpose of this study, the answers were limited to the existing people, processes, and data management of the research enterprise at the institution of higher education. The matrix table assessment section utilized a scoring system for each statement ranging from 0 – 4, where the highest scores indicated a higher level of readiness for implementation. Subsequently, individual scores were calculated in IBM SPSS Statistics (Version 29.0) for the dimensions of People, Processes, and Data Management. These scores were then aggregated to generate a Total Score, serving as an indicator of the institution of higher education's preparedness for implementing data analysis tools within its research enterprise.

A focus group was a crucial step in the validation process of the customized survey tool. The selected participants were from varied research administration backgrounds and experiences, reflecting the survey's target population and the researcher's network. The focus group was held in September 2023 after Marywood IRB approval was received and was conducted virtually via Zoom, limited to 45 minutes. During the session, the participants engaged in open discussions

shared their thoughts and provided constructive feedback on the survey's questions, format, and overall structure. The researcher carefully observed the discussion and feedback and took notes of any themes or suggestions for improvement. The input collected from the focus group was used to fine-tune the survey tool, ensuring its validity and reliability before the survey population was recruited and the research was implemented. Email invitations were sent to 29 people, with 10 people responding to the test survey and 8 people participating in the virtual Zoom focus group. An amendment to the Marywood University Institutional Review Board application was submitted in line with policies and procedures after revisions were made to the survey tool. Final approval of the revised tool was received in September 2023, and the survey was conducted with the target population.

Data Collection and Procedures

Out of a total population of 8,200 registered NCURA members, a total of 1,582 members were invited to participate in the self-reported web-based survey in late September 2023. Survey data was collected using the online platform Qualtrics, and the survey link was distributed to participants through the email provided in their NCURA membership profile. An email was sent to the participant recruitment list to inform them of the purpose of the study (Appendix A and Appendix C) and invite them to participate, an informed consent document (Appendix B and Appendix D) was also included. The researcher sent a follow-up email within two days of the initial email and then a final reminder email one week later. At the same time, the survey and recruitment materials were posted to NCURA Collaborate, a professional networking platform. The survey was active to participants for a total of 17 calendar days (2 weeks and 3 days). The survey link provided to all potential participants had a disabled Internet Protocol (IP) address so that identifying numbers assigned to network devices would not be collected.

At the conclusion of the survey, participants had the opportunity to provide their email address to receive compensation for their valuable time. As an expression of gratitude for successfully completing the survey the researcher sent a $25 Amazon gift card to the email address provided by the participant for those who opted in to receive compensation. Participants' contact information and the survey data were collected and stored separately, with confidentiality measures in place to ensure that personally identifiable information was not disclosed, shared, or published. The minimum target for a response rate, based on the target sample, was 233 persons with a 95% confidence interval, and 5% margin of error. All information contained within the final data set was stored electronically in Dropbox on a secure laptop.

Immediately following the closure of the survey, data was exported from Qualtrics directly into IBM SPSS Statistics (Version 29.0). Data will be stored for six months following the completion of the study to allow time for data analysis and write-up of results. Following the completion of the researcher's Doctoral Dissertation and official completion of the Ph.D. degree from Marywood University, the data will be destroyed within three to five years after closure and publication of results through deletion of electronic data from Qualtrics application and secure Dropbox records. Throughout the study, the only person(s) with access to the raw data was the researcher and the Dissertation Committee Chair.

Data Preparation and Analysis

Additional data preparation was required before completing further analysis. All cases were examined for normalcy and outliers. The variation in sample sizes (*N* values) within the dataset and subsequent results can be attributed to the intentional design of the survey, which includes skip patterns based on participant responses to certain questions. Participants were guided through question 12. and its sub-questions based on their specific responses, which bypassed

irrelevant questions. The resulting *N* values may also be from the skip patterns of the respondents themselves reflecting the diversity of participant choices when completing the survey.

In the data cleaning process, missing responses were systematically recoded as 99 or -99 to facilitate a standardized representation of the data (*N*=145). Additionally, categories with a limited number of responses were judiciously combined, ensuring a more robust and manageable dataset for subsequent analysis while preserving the integrity of the information received from the survey. Instances where participants did not provide a response were treated as missing data and addressed uniformly across all relevant questions. Likewise, "No response provided" was added as a category to all questions where relevant.

The Age category of 18 – 24 was excluded from the analysis due to a lack of responses, and the age groups 55 – 64 and 65+ were merged into a consolidated category labeled 55+. Regarding gender identity, the category "prefer not to answer" was excluded from the analysis due to no responses, and "no response provided" was added as a category because there was a valid number. Concerning the variable for "Years of experience in research administration," responses indicating less than one year of experience were combined with those reporting 1 – 5 years of experience, resulting in a total of 19 responses. For the Salary variable, there were no responses within the $0 - $30,000 range, and only two responses in the $31,000 - $60,000 range, as a result these categories were combined into a broader category labeled "Less than $90,000." The researcher also combined "No answer provided" for the salary range question into a single category denoted as "Prefer not to answer."

Several variables were combined to form new variables, Appendix I explains the scored and recoded variables in detail. In terms of data analytics, the skillset is often measured in terms of numbers or frequency of use. Specifically, the skillset terminology used in this study is a

cumulative sum derived from four survey questions: the number of data analysis tools used, the number of knowledge acquisition pathways or certifications, the number of reports generated, and the number of target audiences for data output or reporting. The conditional questions of question 12c. “the number of software tools used;” question 12e. “the number of courses taken or ways learned;” question 12f. “the number of different reports generated;” and question 12g. “the number of audiences with whom reports are shared” were scored to form a new variable (SKILLSET). The higher the sum was interpreted as the respondent’s proficiency in data analysis or advanced skills. The variable for skillset was then recoded into a categorical variable with three levels, “Foundational skills, Intermediate skills, and Advanced skills,” for further analysis.

Additionally, question 12 was transformed to form a new dichotomous variable (USAGE) to indicate Tools Used or Tools Not Used. Statements from the three matrix tables were scored to their respective total sum for each case, question 13. (PEOPLE), question 14. (PROCESSES), and question 15. (DATA MANAGEMENT), and then a cumulative sum formed a new variable (TOTAL_SCORE). The higher the sum was interpreted as being toward Full Implementation of data analytics tools at the institution. Based on the institution name provided by the participant, institution rank (RANK_HERD) and institution research expenditures (EXPEND_HERD) from the 2021 NSF HERD Survey as raw data were added to the dataset post-survey. Institutional rank and institutional expenditures were grouped into quartiles to form a new version of each of these variables for additional analysis: “Top 25%”, “50%”, “75%”, and “Bottom 25%.”

Additional data cleaning was performed in preparation for the Binary Logistic Regression with the main research question. Questions falling within categories with fewer than 5 responses were either excluded or reclassified in a subsequent manner. Regarding gender identity, “non-

binary/third gender" was recoded as missing in the regression analysis because there were only 2 responses. Regarding salary, "prefer not to answer" was recoded as missing because there were fewer than 5 responses. Regarding the primary role in research administration (ROLE_1), 1 post-award non-financial response was recoded to pre-award and post-award non-finance. Regarding where your role is within the research enterprise structure (ROLE_2), financial administration and research and grant development were recoded as missing because they each had less than 5 responses. Regarding the question of the respondent's role, including data management, the answer of "no" was recoded as missing because there were only 2 responses.

Statistical Methodology

All variables were analyzed with descriptive statistics (mean, median, standard deviation) to evaluate the responses to each of these variables for preparation of analysis. Differences between the independent variables and the dependent variable (USAGE) were examined using the chi-square test. The analysis included Independent Samples t-test, binary logistic regression, and ANOVA. Independent Samples t-tests were used to compare the means of two groups and explore the potential between mediating and moderating variables by examining demographic variables such as years of experience in research administration, role definition, and skillset. Additionally, subgroup analyses were conducted based on relevant characteristics to gain insights into the nature of the group differences. Further investigation into outliers or influential cases was performed to assess their impact on the t-test results.

In conjunction with binary logistic regression, several supplemental analyses were conducted to enhance the understanding of the relationships between variables. These included examining the assumptions of linearity, normality, and homoscedasticity. Interaction effects between predictor variables were explored to determine if the relationship between the dependent variable

and independent variables differs across different scores of data implementation and data analytics tool usage. Diagnostic plots were examined to evaluate the model's adequacy and identify potential issues.

After conducting an analysis of variance (ANOVA) to compare means across multiple groups, post hoc tests were used to determine which specific group differences were significant. These tests, such as Tukey's HSD and Levine's Test, provided a more detailed understanding of the pairwise differences between and within groups. Furthermore, effect size measures, such as eta-squared, were calculated to assess the practical significance of the observed group differences. Exploratory analyses, such as examining the interaction between categorical variables provided valuable insights into how different factors might interact to influence the dependent variable. Additionally, diagnostic checks for assumptions, such as normality and homogeneity of variances were performed to ensure the validity of the ANOVA results.

In order to answer the main research question *"What **factors** (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) **predict** the use of data analysis tools in an office of sponsored research in institutions of higher education within the United States?"* a chi-square test was performed for each independent variable as a crosstab with the dichotomous dependent variable (USAGE) to determine which variables were significant. To perform further analysis of the significant variables, a Binary Logistic Regression was used to determine which variables were most significant to be included in the model.

A Mann-Whitney *U* test was calculated to examine subproblem one to determine if there was a significant difference among years of experience in research administration and data analysis

tool usage. In order to assess subproblem two a One-way ANOVA was used to compare the mean or skillset among age groups. Additional analysis was conducted using a One-way ANOVA to examine the skillset of research administrators among the roles reported in (ROLE_1) and (ROLE_2). Additional analysis was conducted for subproblem two using a One-way ANOVA to examine the skillset of research administrators among the structure of the research enterprise at institutions of higher education. For the purpose of this study, the structure was defined as centralized administration, decentralized administration, and shared service administration.

In order to answer subproblem three, separate Independent Samples t-tests were performed to compare the means. During data preparation, the numerical answers to the questions were scored in a sum to get the People, Processes, and Data Management scores. A fourth Total Score was derived by adding the total scores from each category People, Processes, and Data Management. D/PU: Doctoral Professional Universities and missing values were excluded from the first Independent Samples t-test. Additional supplemental analysis was conducted to determine if there is a significant difference among all institution classifications (Doctoral Universities – Very high research activity (R1), Doctoral Universities – High research activity (R2), D/PU: Doctoral/Professional Universities) related to each score (People, Processes, and Data Management). A One-way ANOVA was calculated to compare the people score, processes score, and data management score among the three institution classifications.

In order to answer subproblem four, separate Independent Samples t-tests comparing the means of the two groups were performed for (PROP_VOL) and (USAGE) and award volume (AWARD_VOL) and (USAGE) of data analysis tools. In order to answer subproblem five, a Pearson correlation coefficient was calculated for the relationship between research expenditures

as reported in the 2021 NSF HERD Survey, and Total Score (People, Processes, and Data Management combined). Additional supplemental analysis was conducted to determine if there is significance between (RANK_HERD) and (TOTAL_SCORE). A Pearson correlation coefficient was calculated for the relationship between institutional rank as reported in the 2021 NSF HERD Survey and Total Score (People, Processes, and Data Management combined).

This research study was approved by the Marywood University Institutional Review Board. The null hypothesis was tested based on statistical significance criteria of a pre-established (*a priori*) probability alpha (a) level of $\alpha = .05$. Having outlined the research methodology; the focus now shifts to the Results chapter, where key findings and insights derived from the research will be presented and discussed.

Chapter 4: Results of the Study

This chapter presents the survey findings and the statistical analysis associated with the research questions from Chapter One. The study surveyed director-level research administrators with membership in the National Council of University Research Administrations (NCURA) from institutions of higher education across the United States. Participants were located across seven regions in the United States as defined by NCURA.

The purpose of the study was to determine what factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) predict the use of data analysis tools in an office of sponsored research in institutions of higher education within the United States. As this chapter concludes, a synopsis of the research questions and associated hypothesis and general findings are discussed.

Data was collected starting September 25, 2023 and closed October 11, 2023. The total sample population ended up being 1,406 out of 1,582 due to inactive emails at the time of survey deployment. Upon initial review of the data collected, 184 out of 1,406 potential participants opened and began the survey. Out of 184 possible respondents, 145 selected at least one answer throughout the survey and became the data analysis sample. Cases where participants did not answer at least one question (39 cases) were removed from the dataset and data cleaning and analysis were performed. While the response rate for this study was relatively low at 10.3%, these findings align with previous surveys targeting research administrators as the sample population.

After preparing the data for analysis, descriptive statistics and frequency tables were run for all variables and the following tables were of interest to the researcher for the purpose of this study.

Demographics and Descriptive Statistics

Table 4.1

Response Rate

Description	*N* Study Participants
Total NCURA membership as of March 17, 2023	8,200
Total members who met the study criteria	1,582
Total deliverable direct recruit or valid email addresses	1,406
Total number of participants who opened the survey	184
Total number of participants who answered at least one question and were included in the analyses	145

Table 4.2

Frequency distribution (N=145) for participants' reported age range, gender identity, years of experience in research administration, and current salary range as a director-level research administrator.

Description	*N*	Valid Percent
Age Range		
35 – 44 years	48	33.1%
45 – 54 years	35	24.1%
55+ years	30	20.7%
No response provided	19	13.1%
25 – 34 years	13	9.0%
Gender Identity		
Female	102	70.3%

Male	22	15.2%
No response provided	19	13.1%
Non-binary / Third gender*	2	1.4%
Years of Experience in Research Administration		
More than 15 years	55	37.9%
11 – 15 years	33	22.8%
No response provided	20	13.8%
Less than 5 years	19	13.1%
6 – 10 years	18	12.4%
Salary Range, Annual (USD)		
$91,000 - $120,000	43	29.6%
Less than $90,000	35	24.1%
More than $150,000	23	15.9%
Prefer not to answer	23	15.9%
$121,000 - $150,000	21	14.5%

**Gender identity was combined later to perform logistic regression*

The majority of participants indicated that they were in the age range of 35 – 44 years (*N*=48), 45 – 54 years (*N*=35), and 55+ years (*N*=30) which indicates a more seasoned and generally more experienced population of the workforce. This data aligns with general trends and recent work surveys in the field of research administration. Research administrators can enter the field with diverse educational and work experiences. Some may start in entry-level positions after completing an undergraduate degree, while others enter the field after completing advanced degrees or other work experiences inside or outside higher education.

Most participants indicated that they have more than 15 years of experience (*N*=55) in research administration. The number of years of experience required to become a director-level research administrator can vary widely depending on several factors, including the organization,

the complexity of the growth and the growth of the institution, and the individual's career path. There is no standard career path, and the progression to a director-level position is influenced by factors such as education, skills, performance, and leadership qualities. In general, the criteria for this research study were those that typically need several years of progressively responsible experience, often ranging from 5 to 15 years or more.

This finding underscores the alignment between most participants' reported current salary range ($91,000 - $120,000, *N*=43) and the anticipated salary range for individuals in director-level research administration roles. Although not a focus of this study, it remains an observation that the survey respondents were overwhelmingly female (*N*=102) within the age range of 35 to 44 (*N*=48). This gender distribution emphasizes the need for equity and signifies positive strides in fostering inclusive career growth and advancement within research administration in recent years.

Table 4.3

Frequency distribution (N=145) for participants' reported institution classification, NCURA Region, reported structure of the institution, and participants' institutional proposal vs. grant award yield rate within the last 3 years (most recently completed FY20 - FY22).

Description	*N*	Valid Percent
Institution Classification		
Doctoral Universities - Very High Research Activity (R1)	86	59.3%
D/PU: Doctoral/Professional Universities	29	20.0%
Doctoral Universities - High Research Activity (R2)	17	11.7%
No response provided	13	9.0%
NCURA Region		
Region 4 – Mid-America	37	25.5%

Region 3 – Southeastern	33	22.8%
Region 2 – Mid-Atlantic	24	16.6%
Region 1 – New England	13	9.0%
Region 6 – Western	12	8.3%
Region 7 – Rocky Mountain	11	7.6%
Region 5 – Southwestern	10	6.9%
No response provided	5	3.4%
Structure of the Research Enterprise		
Centralized administration	85	58.6%
Decentralized administration	35	24.1%
Shared service administration	25	17.2%
Institution Yield Rate		
I don't have access to this information	56	38.6%
More than 50%	21	14.5%
No response provided	19	13.1%
21% - 30%	17	11.7%
41% - 50%	12	8.3%
31% - 40%	10	6.9%
Less than 20%	10	6.9%

A majority of the participants indicated that their current employer is a Doctoral University – Very High Research Activity, also known as (R1) (*N*=86) according to the Carnegie Classification of Institutions of Higher Education (2021). The lowest number of participants indicated that their current employer is a Doctoral University – High Research Activity, also known as (R2) (*N*=17). Despite a substantial difference in respondent numbers between R1 and R2 institutions, the results of this study were largely nonsignificant related to data analytics usage. This result suggests that, despite the variance in sample size, comparable levels of activity

related to data analysis within the research enterprise exist at both R1 and R2 institutions. The limited response from R2 institutions doesn't hinder the conclusion, emphasizing a uniformity in data analytics activities across these two institution types. The data suggests that R2 institutions, despite their smaller representation in the sample, are actively engaged in data analytics activities comparable to their R1 counterparts.

The majority of participants reported that their institution has a Centralized administration (*N*=85) or a Decentralized administration (*N*=35) which are the two most common structures. Although it has advantages, the centralized structure may face challenges related to flexibility and responsiveness at the college or department level, prompting some institutions of higher education to adopt a hybrid, shared, or decentralized model for specific functions of the grant lifecycle.

Most participants indicated they do not have access to the yield rate data for their institution of higher education (*N*=56) and a number of people (*N*=19) did not provide a response to this question. The yield rate is a calculated percentage from the total number of proposals submitted to the total number of awards received, also known as the success rate.

Table 4.4

Frequency distribution (N=145) for participants reported primary role within research administration and within the research enterprise at their institution, and whether their position description or duties include data analysis, and the dichotomous variable usage.

Description	*N*	Valid Percent
Participant Primary Role		
Pre-award and Post-award (non-financial)	51	35.2%
Pre-award and Post-award finance administration	37	25.5%
Pre-award proposal development	27	18.6%

Post-award finance administration	16	11.0%
University research compliance	14	9.7%
Role within Research Enterprise		
Central office of research	76	52.4%
Department administration	22	15.2%
College administration	22	15.2%
Provost or other leadership office	13	9.0%
No response provided	7	4.8%
University research compliance	5	3.4%
Position Includes Data Analysis		
Yes, and it is part of my position description	52	35.9%
Yes, and it is not part of my position description	43	29.7%
No	32	22.1%
*No response provided	18	12.4%
Data Analysis Tools Usage		
Tools Used	95	74.8%
Tools Not Used/No Response	32	25.2%

**Not included in the dichotomous variable tools usage*

Most of the participants indicated that their primary role is either pre-award and post-award (non-financial) (*N*=51) or pre-award and post-award (financial) administration (*N*=37) which means they handle all of the administrative tasks of the grant lifecycle from funding proposal development through grant award financial management. Research administrators in this type of role are cornerstones of the research enterprise at their institutions of higher education. They often coordinate tasks with faculty members and business managers to oversee the entire grant lifecycle for a defined portfolio or department.

The majority of participants indicated they are from the central office of research (*N*=76). The Central office of research often serves as the Authorized Organization Representative or signature authority for the institution of higher education. This office is often the central location where proposals are submitted to external sponsor agencies, policies and procedures are developed, and new initiatives and training are implemented.

The majority of participants indicated that their position description does include a key responsibility for data aggregation, analysis, or reporting (*N*=52). However, an almost equal number of participants noted that their job descriptions do not explicitly include data aggregation, analysis, or reporting (*N*=43). Despite this, they recognize that engaging in data analysis and understanding how to use data is or is becoming a significant aspect of their role.

The majority of respondents indicated they use data analysis tools (*N*=95). Most participants demonstrate a high frequency of engagement with data analysis tools, utilizing them on a daily basis or at least once per week. This group of participants exemplifies a strong commitment to integrating data-informed decision-making methodologies into their routine practices. Their daily utilization underscores a consistent reliance on analytical and reporting tools and a proactive approach to data-informed decision-making. Additionally, the subset of participants who engage with these tools at least once per week indicates a broader commitment to data-related activities.

Most participants shared that they acquired data analysis skills independently, describing themselves as self-taught or self-learners (*N*=69). Many participants reported acquiring their skills in data analysis at their current employer institution (*N*=42) and/or previous employer (*N*=27). A few participants indicated that they acquired their data analysis skills through coursework in an advanced degree program (*N*=21).

Table 4.5

Mean and standard deviation for the raw score sums of Tools Score, Learn Score, Use Case, Audience Score, and Skillset Score (N=145)

	Mean	Standard Deviation
Tools Score	1.90	1.084
Learn Score	1.90	1.162
Use Case Score	6.05	2.946
Audience Score	2.85	1.327
Skillset Score	12.71	4.652
*No response provided		

**52 participants did not provide a response to the conditional survey sub-questions of 12.c., 12.e., 12.f., and 12.g.*

Participant Profiles

A notable feature of this dataset was its capacity to construct participant profiles for specific use cases. For instance, the attributes of Case #132 are described:

Doctoral University-Very high research activity (R1)	*Yield rate 31% - 40%*
Public institution of higher education	*Use data tools once or twice per week*
Decentralized structure	*It is not part of their position description*
Central office of sponsored research	*Use 6 different data analysis tools*
Role is Pre- and Post-Award Non-financial	*Acquired skills, advanced training*

As another example, the attributes of Case #20 are described:

Doctoral University-Very high research activity (R1)	*Yield rate more than 50%*
Private institution of higher education	*Use data tools everyday*

Shared service administrative structure	*It is part of their position description*
Senior leadership office	*Use 6 different data analysis tools*
Role is research data analytics	*Acquired skills, advanced training*
They have the most use cases, outputs	*The broadest audience to share data*

Likewise, Case #59, Case #26, and Case #39 reported full implementation of all areas including People, Processes, and Data Management. Their institutions were Doctoral Universities – Very High Research Activity (R1), and all indicated that data is part of their role and current position description.

An initial review of the dataset suggested that a few institutions have invested in specific areas of implementation such as Data Management, and little or no investment in the other areas of People and Processes. For example, the scores from Case #78 indicated there is no implementation in People and Processes and marginal implementation in Data Management. Their 2021 NSF HERD Survey rank is in the 800s with $318,000 in research expenditures, a reasonably low yield rate for proposals vs. awards 21% - 30%, and uses data analytics tools only annually.

Table 4.6

Mean and standard deviation for the raw score sums of People Score, Processes Score, Data Management Score, and Total Score (N=145)

	Mean	Standard Deviation
People Score	18.97	8.236
Processes Score	21.19	9.526
Data Management Score	23.30	11.498

Total Score	63.46	26.870

*No response provided

**33 in each category were coded as Missing due to non-response*

The cumulative raw scores across the three matrix tables were examined; dimensions of People, Processes, and Data Management. The findings of this study suggested that institutions that reached a total score of 30 or more in the People dimension are progressing toward full implementation of data analysis in terms of personnel resources and training at their institution of higher education. The findings of this study suggested that institutions that reached a total score of 30 or more in the Processes dimension are progressing toward full implementation of data analysis in terms of standardized policies and practices at their institution of higher education related to data analytics tools. Similarly, the data suggested that institutions that reached a total score of 35 or more in the Data Management dimension are progressing toward full implementation of data analytics tools. The findings of this study suggested that institutions that reached a Total Score of 87 (86.25) and are progressing toward meeting all of the criteria needed to successfully implement data analytics tools and data-informed decision making. The frequencies associated with the three dimensions provided valuable insights into the varied and intricate states of readiness across different institutions, underscoring the nuanced nature of implementing data analytics and aligning with the principles of Complex Adaptive Systems (CAS) theory.

The absence of any institution indicating full implementation in any dimension emphasizing the complexity of achieving comprehensive readiness. The lack of full implementation signals ongoing efforts and potential areas for improvement and further research, encouraging the

continuous development of strategies for maximizing the benefits of data analytics in higher education and the research enterprise.

Research Question Analysis

In order to determine what factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) predict the use of data analysis tools in an office of sponsored research in institutions of higher education within the U.S., a Binary Logistic Regression was performed. Prior to the regression analysis, supplemental analysis using the chi-square test was performed for each independent variable as a crosstab with the dichotomous dependent variable (USAGE) to determine which variables were significant. The variable for (SKILLSET) was intentionally excluded from the analysis of the overall research question because the questions contributing to this continuous variable were conditional. Only participants who answered "Yes" to question 12 provided input for these conditional questions, leading to the decision to omit the skillset variable from the model. Results of the chi-square test suggest that only the variable of EXPEND_HERD (df (145) = 8.918, p = .030 was significantly associated with whether participants used data analysis tools or not (USAGE).

Table 4.7

Comparison of independent variables to the dichotomous dependent variable data analytics tools usage (N=145)

Description	*N*	Tools Used (%)	*p* value
Rank_HERD			.053
Top 25% of Rank	26	23.0%	
50% of Rank	29	25.7%	

75% of Rank	28	24.8%	
Bottom 25% of Rank	30	26.5%	
Expend_HERD**			.030
Top 25% of Expenditures	23	20.4%	
50% of Expenditures	30	26.5%	
75% of Expenditures	32	28.3%	
Bottom 25% of Expenditures	28	24.8%	
Institution Classification			.239
Doctoral Universities (R1)	66	64.7%	
Doctoral Universities (R2)	11	10.8%	
D/PU: Doctoral/Professional Universities	25	24.5%	
Age (Years)			.835
25 – 34 years	10	10.6%	
35 – 44 years	34	36.2%	
45 – 54 years	26	27.7%	
55+ years	24	25.5%	
Gender/Gender Identity			.716
Male	17	18.5%	
Female	75	81.5%	
Years of Experience			.639
Less than 5 years	13	14.0%	
6 – 10 years	13	14.0%	
11 – 15 years	23	24.7%	
More than 15 years	44	47.3%	
Salary Range			.231
Less than $90,000	22	24.4%	
$91,000 - $120,000	32	35.6%	
$121,000 - $150,000	16	17.8%	
$151,000 +	20	22.2%	
Primary role in research administration			.072
Pre-award proposal development	24	21.2%	

Pre-award and Post-award (non-financial)	35	31.0%	
Post-award finance admin	12	10.6%	
Pre-award and Post-award finance admin	28	24.8%	
University research compliance	14	12.4%	
Role within the research enterprise			.415
Central office of research	59	54.1%	
College administration	19	17.4%	
Department administration	15	13.8%	
Provost or other leadership	11	10.1%	
University research compliance	5	4.6%	
Structure of the research enterprise			.421
Centralized administration	68	60.2%	
Decentralized administration	28	24.8%	
Shared service administration	17	15.0%	
Yield rate (proposal/award, FY20-FY22)			.520
Less than 20%	8	8.5%	
21% - 30%	10	10.6%	
31% - 40%	9	9.6%	
41% - 50%	10	10.6%	
More than 50%	16	17.0%	
I don’t have access to this information	41	43.6%	
Total implementation score			.360
Full implementation	24	27.6%	
Partial implementation	21	24.1%	
Under discussion	22	25.3%	
No implementation	20	23.0%	

***Statistical significance found at the 0.05 level*

Based on the significance of the individual predictors found, a Binary Logistic Regression was used to determine whether Expenditures, separated into quartiles from the 2021 NSF HERD Survey, was a predictor of the usage of data analysis tools. Regression results

indicated that the overall model fit of the predictor (EXPEND_HERD) was adequate (-2 Log Likelihood = 136.673) and was statistically reliable in distinguishing between participants that use data analysis tools and those that do not (USAGE) (X^2 (6) = 20.602, $p < .026$). The model fit was further supported by the Hosmer and Lemeshow Test formula (X^2 (6) = 13.400, $p = .037$). The model correctly classified 74.8% of the cases. Regression coefficients are presented in Table 4.8. *Wald* statistics indicated that the classification of Expenditure into quartiles significantly predicted whether the use of data analysis tools was reported as a "Yes." Participants were about 21 times more likely to report the use of data analysis tools if they worked for an institution in which research expenditures fall within the Top 25% of Expenditures and 15 times more likely for an institution with 50% of institutional expenditures, compared to institutions in the Bottom 25% expenditures category, as reported in the 2021 NSF HERD Survey.

Table 4.8

Binary Logistic Regression for Institutional Expenditure and Data Usage (N=145)

	Exp(B) O.R.	95% C.I. for Exp(B)	
		Lower	Upper
Bottom 25% Expenditures	Reference		
75% of Expenditures	1.514	.399	5.738
50% of Expenditures	15.130	1.267	180.710
Top 25% of Expenditures	20.602	1.441	294.556

Subproblem One

In order to determine if years of experience in research administration is a factor in whether people are using data analysis tools a Mann-Whitney *U* test was used. To answer subproblem one *(Does the number of years of experience (YRS_EXP) in research administration differ between participants that use data analysis tools and participants that do not (USAGE)?)* a

Mann-Whitney *U* test was calculated to examine if there is a significant difference among years of experience in research administration and data analysis tool use. Results suggest there was no significant difference ($U = 1292.500$, $p > .05$) in whether or not data analysis tools are used by research administrators with varying years of experience. Length of experience in research administration does not appear to be a factor for whether or not data analysis tools are used. The data from this study suggests that years of experience in research administration is not a factor in determining whether someone uses data analysis tools or not.

Table 4.9

Two-Tailed Mann-Whitney U Test for Data Analysis Tool Usage and Years of Experience in Research Administration

	Mean Rank				
Variable	Group B	Group A	*U*	*Z*	*p*
Usage	Tools Not Used/ Not reported	Tools Used	1292.500	-1.172	.241

Subproblem Two

In order to answer subproblem two *(Does skillset differ between the age range of the participants, role definition of primary role and role within the research enterprise, and research enterprise structure?)* a One-way ANOVA was conducted to examine the mean of skillset across age groups. No significant difference was found (F (3,91) = 1.086, $p = .359$), indicating that the age of the respondent was not related to skillset.

Additional supplemental analysis was performed to determine if skillset differs among roles (ROLE_1) and (ROLE_2). ROLE_1 was defined as the participants' primary role within the research enterprise at their institution of higher education. Most of the participants indicated that

their primary day-to-day role is either pre-award and post-award (non-financial) or pre-award and post-award (financial) administration, which means they handle all of the administrative tasks of the grant lifecycle from funding proposal development through grant award financial management. A One-way ANOVA was conducted to examine the impact of the skillset of research administrators among the roles reported in (ROLE_1). A significant difference was found (F (4,92) = 3.840, $p < .05$) which suggests there is a significant difference between skillset and a person's primary role in research administration. The most significant difference was found with participants whose primary role is Research Compliance ($p < .05$).

Table 4.10

Subproblem Two; ANOVA Statistics for Skillset and Primary Role

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	259.890	4	73.972	3.840	.006
Within Groups	1695.271	88	19.264		
Total	1991.161	92			

Additional analysis was conducted using a One-way ANOVA to examine the skillset of research administrators among the roles reported in (ROLE_2). ROLE_2 was defined as the unit within the research enterprise at the institution of higher education. The results suggested a significant difference (F (4,89) = 4.833, $p < .05$) between skillset and where the participants' role was within the research enterprise at the institution of higher education. The most significant difference was found among participants who reported their role was in the Provost Office/Other leadership offices ($p = .001$).

Table 4.11

Subproblem Two; ANOVA Statistics for Skillset and Role within the Research Enterprise

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	363.027	4	90.757	4.833	.001
Within Groups	1596.095	85	18.778		
Total	1959.122	89			

Additional analysis was conducted using a One-way ANOVA to examine the skillset of research administrators within the structure of the research enterprise at institutions of higher education. For the purpose of this study, structure was defined as centralized administration, decentralized administration, or shared service administration. The results suggested no significant difference (F (2,92) = 1.945, p = .149) between skillset and the structure of the research enterprise at their institution of higher education. The data from this study suggests that the institutional structure of the research enterprise is not significantly related to a person's skillset for using data analysis tools.

Table 4.12

Subproblem Two; ANOVA Statistics for Skillset and Structure

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	82.481	2	41.241	1.945	.149
Within Groups	1908.680	90	21.208		
Total	1991.161	92			

Subproblem Three

In order to answer subproblem three *(How do the scores (People Score, Processes Score, Data Management Score, and Total Score) differ between R1 and R2 institutional classification?)* separate Independent Samples t-tests were performed to compare the means. During data preparation the numerical answers to the questions were scored in a sum to get the

People score (*N*=112), Processes score (*N*=112), and Data Management score (*N*=112). A fourth Total Score (*N*=112) was derived by adding the total scores from each category People, Processes, and Data Management. For the purpose of this study, the higher or highest scores were interpreted as "Full Implementation" of data analytics, and the lower or lowest scores were interpreted as "No Implementation." This interpretation implied that the higher the score the more likely a research enterprise met the criteria for people, processes, and data management. D/PU: Doctoral Professional Universities and missing values were excluded from the first Independent Samples t-test. Raw scores were used to perform the separate Independent Samples t-tests and results suggest that no significant differences were found between the means of any of the scores (People, Processes, and Data Management) nor between the two groups (Doctoral Universities – Very high research activity (R1) and Doctoral Universities – High research activity (R2). The results were reported as People: t (34.508) = .844, p = .404 Processes: t (33.695) = 1.399, p = .171 and Data Management: t (76) = 1.289, p = .201 (equal variances assumed) and Total Score: t (32.951) = 1.423, p = .164

Table 4.13

Independent Samples t-test for People Score, Processes Score, Data Management Score, and Total Score

Levine's Test for Equality of Variances					
	F	Sig.	*t*	*df*	Two-sided *p*
People Score	4.216	.043	.844	34.508	.404
Processes Score	6.521	.013	1.399	33.695	.171
Data Management Score	.256	.614	1.289	76	.201
Total Score	3.823	.054	1.423	32.951	.164

Additional supplemental analysis was conducted to determine if there is a significant difference among all institution classifications (Doctoral Universities – Very high research activity (R1), Doctoral Universities – High research activity (R2), D/PU: Doctoral/Professional Universities) related to each score (People, Processes, and Data Management). A One-way ANOVA was calculated to compare the People Score among the three institution classifications. A significant difference was found (F (2,101) = 11.184, $p < .001$. A large effect size for this Analysis of variance model was calculated $\eta^2 = 0.181$. Additional *Tukey HSD* post-hoc analysis revealed the nature of these differences. The mean People Score recorded for the D/PU: Doctoral/Professional Universities classification (M = 13.23, sd = 6.17) was significantly lower (toward No Implementation) than both the Doctoral Universities – Very high research activity (R1) (M = 21.59, sd = 8.55) and Doctoral Universities – High research activity (R2) (M = 20.13, sd = 5.21). There was no significant difference between the (R1) and (R2) classified institutions of higher education in terms of the People Score.

Table 4.14

Subproblem Three; ANOVA Statistics for People Score

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	1297.728	2	648.864	11.184	< .001
Within Groups	5859.619	101	58.016		
Total	7157.346	103			

A One-way ANOVA was calculated to compare the Processes Score among the institution classifications. A significant difference was found (F (2,101) = 9.194, $p < .001$. A large effect size for this Analysis of variance model was calculated $\eta^2 = 0.154$. Additional *Tukey HSD* post-hoc analysis revealed the nature of these differences. The mean Processes Score recorded for the

D/PU: Doctoral/Professional Universities classification (*M* = 15.88, *sd* = 8.16) was significantly lower (toward No Implementation) than both the Doctoral Universities – Very high research activity (R1) (*M* = 24.43, *sd* = 9.21) and Doctoral Universities – High research activity (R2) (*M* = 21.80, *sd* = 5.72) institutions. There was no significant difference between the (R1) and (R2) classified institutions of higher education in terms of the Processes Score.

Table 4.15

Subproblem Three; ANOVA Statistics for Processes

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	1343.739	2	671.869	9.194	< .001
Within Groups	7380.482	101	73.074		
Total	8724.221	103			

A One-way ANOVA was then calculated to compare the Data Management Score among the institution classifications. A significant difference was found (F (2,101) = 5.253, p = .007. A small effect size for this Analysis of variance model was calculated $\eta^2 = 0.094$. Additional *Tukey HSD* post-hoc analysis revealed the nature of these differences. The mean Data Management score recorded for the D/PU: Doctoral/Professional Universities classification (*M* = 18.62, *sd* = 10.339) was significantly lower than both the Doctoral Universities – Very high research activity (R1) (*M* = 26.67, *sd* = 11.304) and Doctoral Universities – High research activity (R2) (*M* = 22.60, *sd* = 9.402) institutions. There was no significant difference between the (R1) and (R2) classified institutions of higher education regarding the Data Management Score.

Table 4.16

Subproblem Three; ANOVA Statistics for Data Management

	Sum of Squares	*df*	Mean Square	*F*	Sig.

Between Groups	1230.775	2	615.388	5.253	.007
Within Groups	11831.754	101	117.146		
Total	13062.529	103			

A One-way ANOVA was then calculated to compare the Total Score (Implementation) among the institution classifications. A significant difference was found (F (2,101) = 9.619, p < .001. A large effect size for this Analysis of variance model was calculated $\eta^2 = 0.166$. Additional *Tukey HSD* post-hoc analysis revealed the nature of these differences. The mean Total Score (Implementation) recorded for the D/PU: Doctoral/Professional Universities classification (M = 47.73, sd = 21.743) was significantly lower than both the Doctoral Universities – Very high research activity (R1) (M = 72.68, sd = 26.693) and Doctoral Universities – High research activity (R2) (M = 64.53, sd = 17.349) institutions. There was no significant difference between the (R1) and (R2) classified institutions of higher education in terms of the Total Score (Implementation).

Table 4.17

Subproblem Three; ANOVA Statistics for Total Score (Implementation)

	Sum of Squares	*df*	Mean Square	*F*	Sig.
Between Groups	11467.962	2	5733.981	9.619	< .001
Within Groups	60208.500	101	596.124		
Total	71676.462	103			

Subproblem Four

In order to answer subproblem four *(Is there a difference between proposal volume and award volume among institutions that use data analysis tools and institutions that do not?)* separate Independent Samples t-tests comparing the means of the two groups were performed.

The results of the Independent Samples t-tests show that Levine's Test is significant, which means equal variances are not assumed. No significant difference between the means of the two groups was found; proposal volume: t (17.236) = 1.460, $p > .05$) and award volume: t (18.041) = 1.150, $p > .05$). The data suggests that for the two groups; participants who reported using data analysis tools and participants who reported not using data analysis tools, there was no significant difference in the proposal volume or award volume of the institution as reported by the participants. Responses from participants at Doctoral Universities – High research activity (R2) could potentially enhance the results. The absence of a substantial distinction in institutional outcomes regarding sponsored research activity suggested that users of data analysis tools had not experienced a notable impact on their proposal or award statistics at the time of this study.

Table 4.18

Independent Samples t-test for Proposal Volume and Award Volume

Levine's Test for Equality of Variances					
	F	Sig.	*t*	*df*	Two-sided *p*
Award Volume	6.278	.014	1.150	18.041	.265
Proposal Volume	8.793	.004	1.460	17.236	.162

Subproblem Five

In order to answer subproblem five *(What is the relationship between institutional research expenditures and those institutions with implementation of data analysis tools)* a Pearson correlation coefficient was calculated for the relationship between research expenditures as reported in the 2021 NSF HERD survey, and Total Score (People Score, Processes Score, and Data Management Score combined). A significant but weak positive relationship was found (r (109) = 0.241, $p < .012$), indicating that as the recorded research expenditures increased, the

Total Score of data analytics also appeared to increase. Additional supplemental analysis was conducted to determine if there is significance between institutional rank, as reported in the 2021 NSF HERD Survey, and Total Score (People Score, Processes Score, and Data Management Score combined). A Pearson correlation coefficient was calculated and the analysis suggests a significant but weak negative relationship (r (109) = -0.347 p < .001), indicating that as the Total Score increased, the rank of the institution tended to be numerically lower.

Table 4.19

Pearson Correlation Coefficient Between 2021 NSF HERD Survey Institutional Rank, Institutional Expenditures, People Score, Processes Score, Data Management Score, and Total Score

		Total Implementation Score
Rank_HERD	Pearson Correlation	**-.347
	Sig. (2-tailed)	< .001
	N	109
Expend_HERD	Pearson Correlation	*.241
	Sig. (2-tailed)	.012
	N	109
People Score	Pearson Correlation	**.909
	Sig. (2-tailed)	< .001
	N	112
Process Score	Pearson Correlation	**.945
	Sig. (2-tailed)	< .001
	N	112
Data Management Score	Pearson Correlation	**.900
	Sig. (2-tailed)	< .001
	N	112

**Correlation is significant at the 0.05 level (2-tailed)*
***Correlation is significant at the 0.01 level (2-tailed)*

Chapter 5: Discussion & Conclusion

This cross-sectional quantitative study examined what factors (age of participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal and award volume, institutional research expenditures, institution classification, role definition, and skillset) predict the use of data analysis tools in an office of sponsored research in institutions of higher education within the United States. A secondary question looked at whether data analysis tool usage is correlated with respondents' reported institution and the institutional research expenditures as reported in the 2021 NSF HERD Survey. A growing body of literature highlighted opportunities for higher education institutions to improve operational efficiency and effectiveness through data-informed decision-making. The available literature is new and informative, providing a basis for best practices, recommendations, and future research.

A 15-question survey was designed and used to collect data on participant demographics, including their institution information related to research administration and general participant demographics, their usage of data analytics tools, and the institutions' readiness to use data analytics tools based on a 5-point Likert Scale. Data was collected from September 25, 2023 to October 11, 2023. The total sample population ended up being 1,406 out of 1,582 due to inactive emails at the time of survey deployment. Upon initial review of the data collected, 184 out of 1,406 potential participants opened and began the survey. Out of 184 possible respondents, 145 selected at least one answer throughout the survey and became the data analysis sample. Cases where participants did not answer at least one question (39 cases) were removed from the dataset.

Interpretation of Results: Descriptive Statistics

It was hypothesized that factors such as the age of the participant, years of experience in research administration, administrative structure, current salary range, gender/gender identity, proposal, and award volume, role definition, institutional research expenditures and institutional rank predicted the use of data analysis tools within the research enterprise at institutions of higher education in the United States. The results of the binary logistic regression failed to reject the entire null hypothesis because it was suggested that, of the independent variables examined, only institutional research expenditures was determined to be a significant predictor for the use of data analysis tools. Moreover, the results of this study suggested that a correlation between institutional research expenditures and the use of data analytics tools existed. The correlation observed in this study does not imply causation.

Regression results indicated that the overall model fit of one factor (EXPEND_HERD) was adequate and was statistically reliable in distinguishing between participants who use data analysis tools and those who do not. Additionally, the odds ratio for this variable indicates that respondents' institutions were 21 and 15 times more likely to use data analytics tools if their reported institution had institutional expenditures in the top 25 % and mid 50%, respectively, as reported in the 2021 NSF HERD Survey. This result indicates a dose-response pattern between institutional expenditures and higher rates of data analysis tools usage. Participants were most likely to use data analysis tools if their reported institutional expenditures were in the top 25% as reported in the 2021 NSF HERD Survey.

A few notable results were discussed prior to the statistical analysis of the subproblems. First, demographically the participants in this study were majority female (N=102) in the age range of 35 – 44 (N=48), and they indicated to have more than 15 years of experience (N=55) in research administration. These demographics, along with a response rate of 145 participants, are

consistent with previous studies cited in the literature review where research administrators were the sample population. Encouraging participation in quantitative survey data collection can be challenging, even if the survey is short and focused, with a clear purpose, and offers an incentive. One suggestion for future research would be to give professionals within research administration the opportunity to participate in the survey design and development on a broader scale or to seek endorsements from collaborators or professional organizations. Another suggestion would be, if time allows, to keep the survey open longer, perhaps for a period of a few weeks to increase the number of participants.

Second, the majority of the participants indicated their current employer was a Doctoral University – Very High Research Activity (R1) (*N*=86) according to the Carnegie Classification of Institutions of Higher Education (2021). Whereas, the lowest number of participants indicated that their current employer was a Doctoral University-High Research Activity (R2) (*N*=17). This research study suggested that (R2) institutions, despite their smaller representation in the sample, are actively engaged in data analytics activities comparable to their (R1) counterparts. Despite the variance in sample size, comparable levels of activity related to data analysis within the research enterprise exist at both (R1) and (R2) institutions. The limited response from (R2) institutions doesn't hinder the results of this study, emphasizing a uniformity in data analytics activities across these two institution types. A greater number of responses from participants at (R2) institutions would have made the sample and data analysis more robust.

Third, most participants reported that their institution has a centralized administrative structure (*N*=85). A centralized structure for research administration is characterized by a hierarchical framework where decision-making authority is concentrated in one area of the institution. In this model, key administrative functions such as resource planning and allocation,

policy and procedure making, and grant management are managed by a central authority, typically at the university level.

Centralized administration is highly effective in large or research-intensive higher education institutions, offering a scalable framework adaptable to evolving needs. Of the total 145 responses considered for this study, 68 participants (60.2%) were in a centralized administration office and use data analysis tools. This structure enables consolidated reporting and data visualization through unified dashboards, aiding decision-makers with clear insights into key performance indicators and trends. Centralization allows control over data infrastructure, facilitating integration from various resources for a comprehensive organizational overview. The centralized structure suggested better aligned data analytics initiatives with broader strategic goals, ensuring efforts address organizational challenges and support objectives for improved overall performance.

Fourth, a surprising result of this study was that most participants indicated they do not have access to the yield rate data for their institution of higher education (N=56) and a number of people (N=19) did not provide a response to this question. The yield rate is a calculated percentage from the total number of proposals submitted to the total number of awards received, also known as the success rate. The lack of access to the institutional yield rate for proposals vs awards can present several challenges. Research administrators could struggle to make informed decisions regarding resource allocation, program development, and strategic planning without knowledge of the yield rate. Understanding how many proposals submitted to external sponsoring agencies ultimately get awarded is critical for optimizing research program development initiatives at the university and college/school levels.

To effectively address the data-sharing challenge, research administration leadership should initiate an annual calculation of the institutional yield rate and disseminate this information to director level research administrators. Taking a proactive approach, the institution can explore the implementation of a structured data access request system. The system would allow individuals to submit requests for specific data access aligned with their roles and responsibilities. Involving key stakeholders within the institution such as IT professionals and end-users in the discussion around data and reporting needs can enhance the development of robust data access and policies. This collaborative approach ensures that the university embraces diverse perspectives and considers varied needs, promoting a well-rounded and inclusive framework for shaping its data access policies.

Fifth, a number of participants noted their job descriptions do not explicitly include data aggregation, analysis, or reporting (N=43). Despite this, they seemed to recognize (based on answers to subsequent survey questions) that engaging in data analysis and understanding how to use data is or is becoming a significant aspect of their role. This discrepancy suggested a potential misalignment between stated responsibilities in their position descriptions and the actual expectations of their positions. This observation in the data raised a point about the need to review and potentially update position descriptions to actively reflect the dynamic requirements of current roles, fostering a more transparent and aligned understanding between employer and employee regarding the essential skills and responsibilities associated with their positions. The acknowledgment of data analysis as a component of their current role also supports recent studies as noted in the literature review of the evolving nature of roles of research administrators, in particular at the director level.

Next, the majority of respondents indicated they use data analysis every day (*N*=40) or once or twice a week (*N*=23). Most participants demonstrated a high frequency of engagement with data analysis tools, utilizing them on a daily basis and at least once per week. This group of participants exemplifies a strong commitment to integrating data-informed decision-making methodologies into their routine practices. The frequency of use highlighted a consistent reliance on data analysis and reporting tools, showcasing a proactive approach to extracting insights and making informed decisions. Additionally, the subset of participants who use these tools at least once per week indicated a broader commitment to engage in regular data-related activities. Most participants reported they acquired data analysis skills independently, describing themselves as self-taught or self-learners. This finding highlighted the participants' proactive initiative in cultivating additional skills to perform their jobs, a self-directed approach to skills development. The predominant self-learning pattern suggests a resourceful and motivated cohort of research administrators. This insight sheds light on the importance of data literacy, upskilling or reskilling existing employees, and the importance of learning opportunities within the participant group. This might also suggest a broader trend of autonomy of professionals within the field of research administration and a desire for continuous skill development – in particular related to data analysis – as it becomes more widely used in higher education.

Finally, the absence of any participant indicating full implementation at their institution in any dimension (People, Processes, Data Management) underscored the complexity of achieving comprehensive readiness and supported the theory of higher education as a Complex Adaptive System (CAS). The data suggested that institutions are at different stages of their data-informed decision-making journey. The lack of full implementation signals ongoing efforts and potential areas for improvement and further research, encouraging the continuous development of

strategies for maximizing the benefits of data analytics in the research enterprise within institutions of higher education.

Interpretation of Results: Subproblems

The results of the first subproblem sought to explore the relationship between the number of years of experience in research administration and the participant's use of data analysis tools. The results of this study suggest that no significant difference ($p > .05$) among research administrators with varying years of experience and whether or not data analysis tools are used. Length of experience in research administration does not appear to be a factor for whether or not data analysis tools are used.

The results of the second subproblem sought to explore the relationship between age and skillset. No correlation was found ($p = .377$) indicating no significant relationship between these two variables. A supplemental analysis was conducted to determine if skillset is related to a person's primary role in research administration. Results of a One-way ANOVA suggested there is a significant difference between skillset and a person's primary role in research administration, with the most significant difference among participants who reported their primary role was Research Compliance. A review of the descriptive and frequencies for primary role revealed that all (N=14) respondents to Research Compliance as their primary role also indicated using data analysis tools (N=14). Supplemental analysis was conducted to determine if a participant's unit within the research enterprise was a factor in the participant's skillset. Results of a One-way ANOVA suggested there was a significant difference between skillset and the participant's unit within the research enterprise, with the most significant difference among participants who reported their unit was the Provost Office/Other leadership offices. The results of this subproblem are consistent with current literature that suggests Institutional Research/Provost

Offices are the first to adopt data analytics tools in institutions of higher education. In addition, the results from the supplemental analyses of subproblem two suggested that role definition and unit within the research enterprise are correlated to a person's skillset. Specifically, the skillset was a cumulative sum derived from four survey questions: the number of data analysis tools used, the number of knowledge acquisition pathways or certifications, the number of reports generated, and the number of target audiences for data output or reporting. A non-significant result in skillset and institutional structure suggests a level of consistency or uniformity within the research structure (Centralized, Decentralized, Shared service) in institutions of higher education. In some cases, employees may naturally align with the organizational norms, resulting in a lack of statistically significant differences in skillsets among research administrators as reported at the time of this study.

The results of the third subproblem sought to explore the relationship between implementation scores (People Score, Process Score, Data Management Score, and Total Score) and the three classifications of institutions (Doctoral Universities – Very High Research Activity (R1), Doctoral Universities – High Research Activity (R2), and D/PU: Doctoral/Professional Universities). Three Independent Samples t-tests were performed to compare the means. The first Independent Samples t-test excluded D/PU: Doctoral/Professional Universities. The results of the first t-test suggest that no significant differences were found between the means of any of the scores (People, Processes, and Data Management) between the two groups (Doctoral Universities – Very high research activity (R1) and Doctoral Universities – High research activity (R2). Additional supplemental analysis was conducted to determine if there was a significant difference among all institution classifications (Doctoral Universities – Very high research activity (R1), Doctoral Universities – High research activity (R2), D/PU:

Doctoral/Professional Universities) related to each score (People Score, Processes Score, and Data Management Score, and Total Score). A One-way ANOVA was calculated to compare each of the scores among the three institution classifications. A significant difference was found with a large effect size for this analysis of variance. The mean score recorded for the D/PU: Doctoral/Professional Universities classification was significantly lower (toward No Implementation) than both the Doctoral Universities – Very High Research Activity (R1) and Doctoral Universities – High Research Activity (R2). There was no significant difference between the (R1) and (R2) classified institutions of higher education for each score.

The results of this subproblem suggest much lower activity in terms of implementation of data analytics tools and in all dimensions of People, Processes, and Data Management for participants that reported their institution is D/PU: Doctoral/Professional Universities. It is possible that D/PU institutions focus their data analytics strategy on other areas such as academic strategies, where an emphasis is placed on teaching evaluation and learning processes rather than research administration. Additional studies, perhaps that are qualitative in nature, are needed to further explore the results of this subproblem.

The results of the fourth subproblem sought to determine if there is a difference between proposal volume and award volume among institutions that use data analysis tools and institutions that do not. The results of the Independent Samples t-tests suggest that for the two groups, participants who reported using data analysis tools and participants who reported not using data analysis tools, there was no significant difference in the reported proposal volume or award volume of the institution. The absence of a substantial distinction in institutional outcomes regarding sponsored research activity implies that users of data analysis tools had not experienced a notable impact on their proposal or award statistics at the time of this study. This

underscores the need for greater participation from research administrators at Doctoral Universities – High research activity (R2), to ensure a comprehensive understanding of the potential impacts of data analysis tool usage on institutional performance metrics. Expanded insights from diverse participants can contribute to a more robust data analysis.

The fifth subproblem sought to explore the relationship between institutional research expenditures and institutional rank, as reported in the 2021 NSF HERD Survey, and the Total Score (People, Processes, and Data Management combined). A significant but weak positive relationship indicated that as the recorded research expenditures increased, the Total Score of data analytics also appeared to increase. Additionally, the relationship between institutional rank, as reported in the 2021 NSF HERD Survey, and the Total Score (People, Processes, and Data Management combined) was explored. A significant but weak negative relationship was found, indicating that as the Total Score (Implementation) increased, the rank of the institution tended to be numerically lower. However, this result does not imply causation as to whether or not the use of data analysis tools, as indicated by Total Score (Implementation), leads to higher research expenditures or vice versa.

Limitations of the study

The researcher recognized several limitations of this study. A limitation of the study is the low response rate (N=145) to the survey and, therefore, generalizability to and beyond the population studied. A limitation of the study was the timing of the survey deployment as it coincided with a major funding proposal deadline for the National Institutes of Health and, separately, regional meetings that required the attention of research administrators.

A self-reported survey introduces limitations because respondents may provide answers based on their perceptions, opinions, or subjective experiences. Individuals may unintentionally

(or intentionally) present information in a manner that aligns with personal biases, affecting the accuracy and reliability of the data. Another limitation of the study was determined to be the survey design. All participants had the opportunity to respond to questions 1. through 12. and the matrix tables 13., 14., and 15. in sequence, with conditional sub-questions related to skillset dependent on their responses to question 12. While this approach was adopted for the current research study, a future study might benefit from an alternative survey design where participants were offered to complete every question. Consideration could be given to having participants complete the survey in a comprehensive manner rather than segmenting the questions based on a response to question 12.

The body of literature on the use of data analytics in higher education is growing, and therefore, a number of books, magazine articles, and publications have been made publicly available since the beginning of this dissertation paper. It is possible there is additional literature not captured due to the timing and completion of this research.

Conclusion

This section will conclude the study by summarizing the key research findings in relation to the aims of this study, contributions to the field of research administration, and suggestions for future research. This study aimed to determine what factors predict the usage of data analysis tools within the research enterprise at institutions of higher education and if there is a correlation with research expenditures. The conceptual CAS framework suggested that - the primary factors (Emergent Patterns) of institutional classification, institutional research expenditures, proposal and award volume, and administrative structure together with secondary factors (Independent Agents) age of participants, years of experience in research administration, role definition,

current salary range, gender/gender identity and employee skillset - predict the readiness for offices of sponsored research to use data analytics tools.

The results of the study suggested that all secondary factors and most primary factors are not significant predictors for the use of data analysis tools. The results of the analysis of this study suggested that one Emergent Pattern, a primary factor of institutional research expenditures was a significant predictor of data analysis tools usage. Additional research is needed to evaluate the context of this finding.

Regression results indicated that the overall model fit of one predictor (EXPEND_HERD) was adequate and was statistically reliable in distinguishing between participants who use data analysis tools and those who do not. The model correctly classified 74.8% of the cases. Participants were 21 times more likely to report the use of data analysis tools if they worked for an institution in which research expenditures fall within the Top 25% of Expenditures and 15 times more likely for an institution with 50% of institutional expenditures, compared to the Bottom 25% expenditures category, as reported in the 2021 NSF HERD Survey.

When statistically analyzing at a few of the factors individually, there was no significant difference in the age of the participant, the reported proposal volume or award volume of the institution, and the number of years of experience in research administration reported by the participant also did not appear to be a factor for whether or not data analysis tools were used. However, the analysis results suggested there is a significant difference between skillset and the participants reported primary role in research administration, and the participants reported unit within the research enterprise. Despite a small sample size, the most significant difference was among participants who reported their primary role was Research Compliance (N=14) and participants who reported their unit was the Provost Office/Other leadership offices (N=13).

A surprising result of the assessment of dimensions of this study suggested a much lower mean Total Score (Implementation) in all dimensions of People, Processes, and Data Management for participants who reported their institution was D/PU: Doctoral/Professional University. It is possible that D/PU institutions focus their data analytics strategy on other areas such as academic strategies, where an emphasis is placed on teaching evaluation and learning processes rather than research administration.

The majority of participants indicated their current employer is a Doctoral University – Very High Research Activity (R1) (*N*=86) according to the Carnegie Classification of Institutions of Higher Education (2021). The lowest number of participants indicated that their current employer is a Doctoral University-High Research Activity (R2) (*N*=17). The results of the subproblems that looked at institutional classification and data analysis tool usage and implementation scores across all means found no significance between (R1) and (R2) classifications. These findings suggested that, despite the variance in sample size, comparable levels of activity related to data analysis within the research enterprise exist at both (R1) and (R2) institutions. A greater number of responses from participants at (R2) institutions would have made the sample and statistical analysis results more robust.

Another surprising result of this study was that most participants indicated they do not have access to the yield rate data for their institution of higher education (*N*=56) and a number of people (*N*=19) did not provide a response to this question. Director-level research administrators could struggle to make informed decisions regarding resource allocation, program development, and strategic planning without knowledge of the yield rate. Understanding how many proposals submitted to external sponsoring agencies ultimately get awarded is critical for optimizing research program development initiatives at the university and college/school levels.

The overall results of this research study underscore the need to start simple in relation to adopting data analytics tools within the research enterprise at institutions of higher education within the United States, perhaps with informed discussions, knowledge sharing, and data literacy training at all levels of the research enterprise.

Contributions to Research Administration

While conducting the literature review, it emerged that the institution perspective (Data Management), management perspective (Processes), and employee perspective (People) were found to be in congruence. Therefore, the researcher designed a customized survey tool that combined institutional and demographic information with a readiness assessment tool. The readiness assessment section focuses on three (3) key dimensions (People, Processes, and Data Management) that literature has suggested to be vital in determining whether an institution of higher education can effectively engage in data-informed decision-making (Voorhees, 2007). The survey tool will be made available to conduct further research.

The researcher plans to continue exploring this topic in their role as Associate Director of Research Administration in the College of Health at Lehigh University and as an active member of the NCURA and collaborator of the SRAI.

To put research into practice, the researcher was awarded a “Future Makers” grant in January 2024 in the amount of $6,800 (internal seed funding) to lead a team to develop a data literacy training presentation for the research enterprise at Lehigh University. The researcher also wrote an article for the NCURA magazine’s March/April 2024 issue titled *“Enhancing Skill Development: Three Approaches for Upskilling Department Research Administrators,”* emphasizing the potential for skill development within the context of capacity building at the college or department level. A future interest for the researcher will be to explore generative AI

as a way to leverage research administration knowledge management and sharing as a potential way to enhance skill development, performance, and research capacity-building. In the future, the researcher intends to repeat this survey data collection with the director level research administrators who are members of the SRAI.

Suggestions for Future Research

To enhance the depth of understanding and explore nuanced insights gained from this study, future research should consider qualitative approaches to data analytics tool use, such as focus group discussions or individual interviews. Specifically, conducting a follow-up qualitative study and recruiting participants who are at institutions defined as D/PU: Doctoral or Professional Universities by the Carnegie Classification of Institutions of Higher Education (2021) could reveal intricate challenges, experiences, and perspectives not fully captured by this study. Similarly, conducting a follow-up qualitative study and recruiting participants who identify their primary role within research administration as Research Compliance or within the Provost Office/Other leadership offices would enhance our understanding of their data analysis tools usage and skillset.

To replicate the data analysis from this study, the researcher can include in the dataset Institutional Rank and Institutional Expenditures as reported in the 2022 NSF HERD Survey (released November 30, 2023) as a comparative and supplementary analysis to investigate whether Institutional Expenditures remained a significant predictor of data analysis tools usage in 2022 compared to the findings in 2021. Additionally, replicating this study with a targeted focus on the recruitment of director-level members of the SRAI could provide an enhanced or comparison perspective from additional research administration leadership.

Furthermore, additional quantitative analysis within the existing dataset could uncover other hidden patterns or correlations, providing a richer understanding of the variables. This approach would involve designing additional research questions and employing more advanced statistical methods to extract deeper insights from the available data. Overall, adopting a multi-faceted approach to furthering the research on this topic by conducting qualitative follow-up studies, targeted replications of this study, and using advanced statistical analysis on the existing dataset would enhance the understanding of data analysis tool usage within the research enterprise in institutions of higher education in the United States.

References

Attaran, M., Stark, B., Stotler, D. (2018) “Opportunities and Challenges for Big Data Analytics in American Higher Education – A conceptual model for implementation.” *Industry and Higher Education.* pp. 1-15 Accessed 9/12/2022

Borgman, C., Brand, A. (2022) “Data Blind: Universities Lag in Capturing and Exploiting Data” *Science*. Vol. 378, Iss. 6626 pp. 1278-1281 ISSN: 0036-8075

Celeste, R.F., Griswold, A., Straf, M.L. (2014) “Understanding the Research Enterprise as a Complex System” *Committee on Assessing the Value of Research in Advancing National Goals; Division of Behavioral and Social Sciences and Education; National Research Council*; Furthering America's Research Enterprise. Washington (DC): National Academies Press (US); Available from: https://www.ncbi.nlm.nih.gov/books/NBK253891/

Daniel, B., Butson, R. (2017) “Chapter 5: The Rise of Big Data and Analytics in Higher Education.” *The Analytics Process.* Imprint Auerbach Publications. 14 pages. Accessed 9/12/2022

Daniel, B., Butson, R. (2014) “Foundations of Big Data and Analytics in Higher Education” *The Proceedings of the International Conference on Analytics Driven Solutions* pp. 39-47 E-Book ISBN: 978-1-910309-50-6

Daniel, B. (2015) “Big Data and Analytics in Higher Education: Opportunities and Challenges.” *British Journal of Educational Technology*. Vol. 46, Iss. 5 pp. 904-920

D’Auria Stanton, A., & Stanton, W. W. (2019). Closing the skills gap: Finding skilled analytics professionals for a dynamically changing data-driven environment. *Applied Marketing Analytics 5*(2), 170-184

Dooley, K., (1997) “A Complex Adaptive Systems Model of Organization Change,” *Nonlinear Dynamics, Psychology, and Life Sciences* Vol. 1, Iss. 1 pp. 69-97

Ekowo, E., Palmer, I., (2017) “Predictive Analytics in Higher Education: Five Guiding Practices for Ethical Use,” *New America, Education Policy*

Ekowo, E., Palmer, I., (2016) “The Promise and Peril of Predictive Analytics in Higher Education: A Landscape Analysis,” *New America, Education Policy*

Falzon, K. (2019) “The Future is Now: Closing the Data Analytics Skills Gap, Town Hall,” available at: https://workingnation.com/wharton/ Accessed 9/20/2022

Ghasemaghaei, M. (2019) “Does data analytics use improve firm decision making quality? The role of knowledge sharing and data analytics competency” *Decision Support Systems, 120,* 14-24. https://doi.org/10.1016/j.dss.2019.03.004

Gibbons, Michael T. (2022) National Center for Science and Engineering Statistics (NCSES). *2021 National Science Foundation, Higher Education Research & Development Survey. NSF* 23-303. Available at https://ncses.nsf.gov/pubs/nsf23303

Guster, D., Brown, C. (2012) “The Application of Business Intelligence to Higher Education: Technical and Managerial Perspectives,” *Journal of Informational Technology Management* Vol. 23 Iss. 2 pp. 42-62

ITA Group (2022) “How Upskilling your Workforce Benefits your Organization” Retrieved from http://itagroup.com/insights/how-upskilling-your-workforce-benefits-your-organization

Knyazeva, H. (2020) “System Theory Approach as a Basis of Strategic Management,” *Foresight and STI Governance*. Vol. 14, Iss. 4 pp. 6-8

Morgenthaler, S. (2009) “Exploratory Data Analysis,” *WIREs Computational Statistics*. Vol. 1

pp. 33-44. Accessed 9/18/2022

Nguyen A., Gardner, L., Sheridan, D. (2020) “Data Analysis in Higher Education: An Integrated View,” *Journal of Information Systems Education*. Vol. 31 Iss. 1

Penprase, B.E. (2018). The fourth industrial revolution and higher education. In Gleason, N. (Eds.), *Higher education in the era of the fourth industrial revolution* (pp. 207-229). Palgrave Macmillan. https://doi.org/10.1007/978-981-13-0194-0_9

Rao, A., Vardhan, B., Shaik, H. (2021) “Role of Exploratory Data Analysis in Data Science” *Proceedings of the 6th International Conference on Communication and Electronics Systems* (ICCES-2021) ISBN: 978-0-7381-1405-7

Robershaw, K., Wolf, B., (2023) “Research Analytics: A Systematic Literature Review,” *Available at SSRN:* https://ssrn.com/abstract=4363262 or http://dx.doi.org/10.2139/ssrn.4363262. Accessed 4/4/2023

Roth, K. (2016) “Embracing Data in Operational Decision-making,” *Parks & Recreation Magazine* National Recreation and Park Association (NRPA) Ashburn, VA

Seaman, M. (2021) “Influencing Change in a Complex Adaptive System” *Consortium for Service Innovation*. Accessed 3/25/2023 from https://www.serviceinnovation.org/influencing-change-complex-adaptive-system/

Sebag-Montefiore, L. (2022) “Upskilling the Workforce,” *The Undercover Recruiter*. Accessed from http://theundercoverrecruiter.com/author/liz-sebag-montefiore Accessed 9/18/2022

Szulawski M., Kaźmierczak I., Prusik M. (2021) “Is Self-determination Good for Your Effectiveness? A Study of Factors which Influence Performance within Self-determination Theory,” *PLoS ONE* pp. 1-17 https://doi.org/10.1371/journal.pone.0256558

Ueland, J., Hinds, T., & Floyd, N. (2021) "Equity at the Edge of Chaos: Applying Complex Adaptive Systems Theory to Higher Education," *New Directions for Institutional Research*. pp. 121-138 https://doi.org/10.1002/ir.20356

Voorhees, R. A. (2007) Institutional Data Readiness Assessment Tool

Wang, L. (2019) "Analysis of Application of Big Data in College Education Management," *Journal of Physics: Conference Series* doi:10.1088/1742-6596/1314/1/012220 pp. 1-6 Accessed 9/9/2022

Webber, K., Zheng, H. (2020) "Big Data on Campus: Data Analytics and Decision Making in Higher Education" *Johns Hopkins University Press*. ISBN: 978-1-4214-3903-7

Wolf, B. (2022) "Hot Topics in Research Administration—Baron Wolf" *Podcast*. https://soundcloud.com/srai-shop-talk/hot-topics-in-research-administration-episode-2-baronwolf

Wolf, B., Hall, T., and Robershaw, K. (2021) "Best Practices for Research Analytics & Business Intelligence within the Research Domain." *Research Management Review*, Vol. 25 Iss. 1 pp. 1-37

Wolf, B., Hall, T., and **Messina, H.** (2020) "Data Analytics and Research Administration, Results of a National Data Collection" *iCatalyst. Society of Research Administrators International.* Accessed 3/1/2020.

Wolf, M., Taimuty, M., Patel, M., and Meteyer, J. (2016) "The Dean's Information Challenge: From Data to Dashboard," *EDUCAUSE Review*. Accessed 9/9/2022

Yu, S., Zhang, F., Nunes, L., & Levesque-Bristol, C. (2021) "Doing Well vs. Doing Better: Preliminary Evidence for the Differentiation of the Statis and Incremental Aspects of the Need for Competence," *Journal of Happiness Studies* pp. 1121-1140 https://doi.org/10.1007/s10902-021-00442-w

Zhu, G., Ma, X., Liu, B. (2018) “Research on the Mechanism of Knowledge and Intelligence Sharing in Enterprises under Big Data: From the Perspective of Status-striving Motivation and Opinion Dynamics,” *Expert Systems*. https://doi.org/10.11-11/exsy.12313

Zink, H. R., Hughes, D., & Vanderford, N. L. (2022). Reconfiguring the research administration workforce: A qualitative study explaining the increasingly diverse professional roles in research administration. *Journal of Research Administration, 53*(2), 119-140. https://www.srainternational.org/blogs/srai-jra2/2022/09/28/reconfiguring-the-research-administration-workforc

Appendix A

Advertising Templates for Recruitment of Human Subjects
Email Recruitment Message – Focus Group

Email Subject Line: You Are Invited: Test Survey Tool and Virtual Focus Group

Dear Colleagues,

As you might know, I am a Ph.D. Candidate at Marywood University and writing my dissertation. I am conducting a research study. Its purpose is to advance the understanding of data-informed decision-making in research administration at institutions of higher education in the United States. The study aims to determine what factors predict readiness to use data analysis tools and if there is a correlation with research expenditures.

You are invited to review, and provide feedback and validation for the survey tool by taking a test survey at this link via Qualtrics. In August, I will be hosting a 45-minute virtual focus group via Zoom. The only purpose of the focus group is to provide feedback on the questions, relevance, and format in support of validating the survey tool. Please use this link When2Meet to indicate your interest and availability to participate in the virtual focus group.

This study has been approved by Marywood University's Institutional Review Board (IRB). Approval number: #2079272-3.

Sincerely,

Heather M. Messina, MPA CRA (PhD Candidate)
hmmessina@m.marywood.edu
(610) 758-3347

Appendix B

IRB Informed Consent Form – Focus Group

Factors that Determine the Use of Data Analysis Tools in an Office of Sponsored Research in an Institution of Higher Education within the United States: A Readiness Assessment

Principal Investigator (PI): Heather M. Messina, MPA CRA, PhD Candidate at Marywood University
Principal Investigator Contact Information: hmmessina@m.marywood.edu, 610-758-3347

Research Advisor: Dr. Deborah Hokien, PhD, Palm Beach State College (former Marywood)
Research Advisor Contact Information: dhokien@hotmail.com, (570) 947-1269

Invitation for a Research Study

You are invited to participate in a pilot study, focus group to help support validation of a survey tool for a research study. You were chosen to participate in the feedback focus group because you are part of the researcher's own network of research administrators and other professionals inside or outside of higher education.

The purpose of the research study, for which you are providing feedback, is to assess the readiness of an office of sponsored research within an institution of higher education in the United States to use data-informed decision-making and data analytics tools. You are part of the pilot, and therefore not part of the real study. Please read this form. Ask any questions you may have before agreeing to take part in this focus group.

Purpose – About the Study

The only purpose of the focus group is to support the validity of the survey tool. The primary aim is to assess the survey tool's clarity, relevance and effectiveness in capturing the intended data. During the session you will engage in open discussions, share your thoughts, and provide constructive feedback on the survey's design, questions, format, and overall structure.

The purpose of the research study is to advance the understanding of data-informed decision-making in research administration at institutions of higher education in the United States. The study aims to determine what factors predict readiness to use data analysis tools and if there is a correlation with research expenditures.

Procedures - What You Will Do

You will be asked to take a test survey at this link via Qualtrics, and reflect on the survey content, layout, format, and overall structure, including the instructions. Next, you will be asked

to indicate your availability for a 45-minute virtual focus group via zoom at this link When2Meet. Your answers to the survey are part of the pilot, and therefore not part of the real study.

You will participate in a 45-minute virtual focus group via zoom. During the focus group the researcher will ask you to review and provide feedback on the survey instructions, each question of the survey tool 1-11, question 12 and its sub-questions, and an assessment section with three (3) Likert Scale questions 13-15. If you have little knowledge about data-informed decision-making or research administration, but you are familiar with completing surveys, your feedback during this focus group would also be valuable. The researcher will type notes during the session and the session will not be recorded.

Risks and Benefits

It is anticipated that participants in this study will not experience any greater risk than their current daily living activities and no personally identifiable information will be collected.

A benefit may be that you will contribute valuable insights into the use of data analytics tools at an institution of higher education in an effort to improve the way decisions are made in support of research capacity building, and you will contribute valuable information that could ultimately improve the way research administrators work in our professional field.

Payment or Other Rewards

You will not receive a payment or reward.

Confidentiality

No web-based action is perfectly secure. However, reasonable efforts will be made to protect your transmission from third-party access. The records of this study will be kept private. Information used to fine-tune the survey tool will not make it possible to identify you. Only the Principal Investigator will have access to the research records. Records will be kept in a locked or password-protected file. Records will be kept for a period of at least three years, then the electronic format(s) will be deleted from the hard drive, and the web-based survey platform.

Taking Part is Voluntary

Participation is Voluntary. Your decision whether or not to participate will not affect your current or future relationship with the investigator. Participation will not affect your relationship with Marywood University. To withdraw from the virtual focus group, simply exit the zoom meeting at any time.

Contacts and Questions

If you have questions about this virtual focus group at any time, contact the principal investigator or the advisor. Their contact information appears at the top of this form.

If you have questions related to the rights of research participants or research-related injuries (where applicable), please contact the Institutional Review Board at (570) 961-4782 or irbhelp@marywood.edu.

You may save or print a copy of this form for your records.

Statement of Consent

By joining the virtual focus group via zoom:

- You understand what the study involves.
- You have asked questions if you had them.
- You agree to participate in the study.

Appendix C

Advertising Templates for Recruitment of Human Subjects
Email Recruitment Message - Survey

Email Subject Line: Invitation Survey and Readiness Assessment for Data-Informed Decisions

Dear Colleagues in Research Administration,

Happy Research Administrator Day (belated, 9/25)! and I hope this email finds you well.

My name is Heather Messina, and I am a Ph.D. candidate at Marywood University in Scranton, Pennsylvania. I am also the Assistant Director of Research Administration in the College of Health at Lehigh University in Bethlehem, Pennsylvania.

I am conducting a research study in fulfillment of the requirements for my Ph.D. program and I would greatly value your participation. The purpose of the quantitative survey is to advance the understanding of data-informed decision-making in research administration at institutions of higher education in the United States. The study aims to determine what factors predict readiness to use data analysis tools and if there is a correlation with research expenditures.

You are invited to participate if you qualify. **To qualify, you must meet the following criteria:**

- Registered member of the National Council of University Research Administrators (NCURA)
- Employed at an institution of higher education within the United States (Region 1 through Region 7)
- Director-level including titles such as Assistant Director, Associate Director, Director, Interim Director, Senior Director, Senior Associate Director, and Executive Director
- In an office of sponsored research or other administrative office within your institution's research enterprise

The research will take place online through the survey platform Qualtrics. Click here to access the survey and it will take about 10-15 minutes to complete.

A benefit may be that you will contribute valuable insights into the use of data analytics tools at your institution of higher education in an effort to improve the way we make decisions in support of research capacity building, and you will contribute valuable information that could ultimately improve the way we work in our professional field.

At the conclusion of the survey, you will have the opportunity to provide your email address to receive compensation for your valuable time. As an expression of gratitude for successfully completing the survey the researcher will send a $25 Amazon gift card to the email address you provide. Your contact information and the survey data will be stored separately, with confidentiality measures in place to ensure personally identifiable information is not disclosed, shared, or published.

This study has been approved by Marywood University's Institutional Review Board (IRB) approval number #2079272-3 dated 8/7/2023 and 9/15/2023. A focus group was held to validate the survey tool. Please find attached an advertisement and informed consent.

Warmly,

Heather Messina, MPA CRA *(PhD Candidate '24, Marywood University)*
Assistant Director of Research Administration
College of Health, Lehigh University
Email for this purpose: hmmessina@m.marywood.edu

Appendix D

IRB Informed Consent Form - Survey

Factors that Determine the Use of Data Analysis Tools in an Office of Sponsored Research in an Institution of Higher Education within the United States: A Readiness Assessment

Principal Investigator (PI): Heather M. Messina, MPA CRA, PhD Candidate at Marywood University
Principal Investigator Contact Information: hmmessina@m.marywood.edu, 610-758-3347

Research Advisor: Dr. Deborah Hokien, PhD, Palm Beach State College (former Marywood)
Research Advisor Contact Information: dhokien@hotmail.com, (570) 947-1269

Invitation for a Research Study

You are invited to participate in a research study that will assess the readiness of your institution to use data-informed decision-making and data analytics tools. You were chosen because you are a registered member of the National Council of University Research Administrators (NCURA) and your member profile indicates a Director-level position as of March 17, 2023 (including titles such as Assistant Director, Associate Director, Director, Interim Director, Senior Director, Senior Associate Director, and Executive Director). This survey includes Director-level roles within an office of sponsored research or other administrative office within your institution's research enterprise. The survey is also limited to those employed at an institution of higher education in the United States, listed in Region 1 through Region 7 as defined by NCURA. Please read this form. Ask any questions you may have before agreeing to take part in this study.

Purpose – About the Study

The purpose of this quantitative research study is to advance the understanding of data-informed decision-making in research administration at institutions of higher education in the United States. The study aims to determine what factors predict readiness to use data analysis tools and if there is a correlation with research expenditures.

Procedures - What You Will Do

You will participate in a 10-15 minute online survey, and answer questions about your organization and its use of data and data analysis tools. If your organization does not use data-informed decision-making or if you are unsure about the extent to which data is used, this information is also valuable for the purpose of this study.

Risks and Benefits

It is anticipated that participants in this study will not experience any greater risk than their current daily living activities and no personally identifiable information will be collected.

A benefit may be that you will contribute valuable insights into the use of data analytics tools at your institution of higher education in an effort to improve the way we make decisions in support of research capacity building, and you will contribute valuable information that could ultimately improve the way we work in our professional field.

Payment or Other Rewards

At the conclusion of the survey, you will have the opportunity to provide your email address to receive compensation for your valuable time. As an expression of gratitude for successfully completing the survey the researcher will send a $25 Amazon gift card to the email address you provide. Your contact information and the survey data will be stored separately, with confidentiality measures in place to ensure personally identifiable information is not disclosed, shared, or published. If you do not enter an email address when prompted at the end of the survey, you will not receive a payment.

Confidentiality

No web-based action is perfectly secure. However, reasonable efforts will be made to protect your transmission from third-party access. The records of this study will be kept private. Information used in any written or presented report will not make it possible to identify you. Only the Principal Investigator will have access to the research records. Records will be kept in a locked or password-protected file. Records will be kept for a period of at least three years, then the electronic format(s) will be deleted from the hard drive, and the web-based survey platform.

Taking Part is Voluntary

Participation is Voluntary. Your decision whether or not to participate will not affect your current or future relationship with the investigator. It will not affect your membership with the National Council of University Research Administrators (NCURA). Participation will not affect your relationship with your institution of higher education or Marywood University. You may withdraw at any time until you submit your answers. Because the survey is anonymous, I will not be able to identify your answers after the survey is submitted. There will be no penalty. To withdraw from the survey, simply exit the survey at any time. If you answered less than 75% of the survey questions, then your responses will be excluded from the data analysis.

Contacts and Questions

If you have questions about this study at any time, contact the principal investigator or the advisor. Their contact information appears at the top of this form.

If you have questions related to the rights of research participants or research-related injuries (where applicable), please contact the Institutional Review Board at (570) 961-4782 or irbhelp@marywood.edu.

You may save or print a copy of this form for your records.

Statement of Consent

By proceeding to the survey:

- You understand what the study involves.
- You have asked questions if you had them.
- You agree to participate in the study.

Appendix E

Survey Tool

1. Name of the current institution which you are employed?
Make this a dropdown from IPEDS data or IPEDS ID code, unique identifier
Other [text box]

2a. What is your institution's classification according to the Carnegie Classification of Institutions of Higher Education (2021)?
Doctoral Universities – Very high research activity (R1)
Doctoral Universities – High research activity (R2)
D/PU: Doctoral/Professional Universities

2b. What is your institution's subtype? (select all that apply)
Private Institution of Higher Education
Public Institution of Higher Education
Academic Medical Center
Predominantly Undergraduate Institution (PUI)
Historically Black College/University (HBCU)
Hispanic/Minority-Serving Institution (MSI)
Other [text box]

3a. What is your primary role in research administration?
Pre-award proposal development
Post-award management (non-financial)
Pre-award and Post-award (non-financial)
Post-award finance administration
Pre-award and Post-award finance administration
University Research Compliance
Other [text box]

3b. Where is your role in the research administration structure?
Central office of sponsored programs
College administration
Department administration
Finance and administration
Provost or other leadership office
Research accounting
Research or grant development
University Research Compliance
Other [text box]

3c. What is the primary structure of Research administration at your institution?
Centralized administration
Decentralized administration

Shared service administration
Other, please explain [text box]

4. In the past five (5) years, on average, how many proposals has your institution submitted?

[open text box/number]
I don't have access to this information

5. On average, how many new grant awards, encompassing all types and sources, did your institution receive in Fiscal Year 2023?

[open text box/number]
I don't have access to this information

6. Within the last 3 years (most recently completed FY20 - FY22), what is your institution proposal / award success or yield rate?

Less than 5%
5% - 10%
11% - 15%
16% - 20%
21% - 30%
31% - 40%
41% - 50%
More than 50%
I don't have access to this information

7a. How many Research Administration staff directly report to you?

Less than 5
6 - 10
11 - 20
21 - 30
31 - 40
41 - 50
More than 50

7b. How many Research Administration staff make up your department?

Less than 5
6 - 10
11 - 20
21 - 30
31 - 40
41 - 50
More than 50

8. What is your age range?

18 - 24
25 - 34

35 - 44
45 - 54
55 - 64
65+

9. What is your gender identity?
Male
Female
Non-binary / Third gender
Prefer Not to Answer

10. How many years of experience in research administration do you have?
Less than one year
1 - 5
6 - 10
11 - 15
More than 15 years

11. What is the range of your current salary?
$0 - $30,000
$31,000 - $60,000
$61,000 - $90,000
$91,000 - $120,000
$121,000 - $150,000
$151,000 +
Prefer Not to Answer

12. Does your position description or current job duties include data aggregation, analysis, or reporting for interdepartmental (inter-college) decision-making? Data could be used for a number of things internal to your sponsored research office(s) including, but not limited to, strategic planning, infrastructure or staffing decisions, proposal development or center/institute focus.
Yes, and it **is** part of my position description
Yes, and it **is not** part of my position description
No
I'm not sure

12.a.<conditional question> **If not you or your position, is there someone within the research enterprise at your institution whose role it is to perform research data aggregation, analysis, or reporting for internal use?**
Yes
No
I'm not sure

12.b.<conditional question> **Does your institution have data analysis tools for use in research administration (e.g., Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau)?**
Yes
No
I'm not sure

12.c.<conditional question> **Which data analysis tools do you use? Select all that apply**
Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau, Other [text box]

12.d.<conditional question> **How often do you use data analysis tools (e.g., Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau)?**
Everyday
Once or twice a week
Once or twice a month
Quarterly
Annually
Never

12.e.<conditional question> **How did you acquire skills in data aggregation, analysis, or reporting? (select all that apply)**
Current employer institution
Previous employer in higher education research administration
Previous employer in industry/for-profit company
Advanced degree or technical training in data analysis (this includes courses during undergraduate education)
Self-taught or self-learner
Online or certification training
Other [text box]

12.f.<conditional question> **What are you using data analysis tools for? (select all that apply)**
Routine reporting (monthly or quarterly)
Budget planning and resource allocation
Calculate proposal volume
Calculate award volume
Calculate institutional or organization unit research expenditures
Predict proposal volume
Review or predict research staff workload
Predict award volume
Calculate or predict success rate for proposals
Anticipate five + year research activity
Summarize organization unit activity for Board of Trustees or other external stakeholders
Summarize organization unit activity for faculty, staff or other internal stakeholders
Explore data by looking for patterns and trends

Other [text box]

12.g.<conditional question> **Who is the main target audience for your data output, analysis or reports? (select all that apply)**
Myself and staff within my organizational unit
College Dean
Department Chairperson
Faculty
Provost or Vice President/Provost of Research
External Partners or Stakeholders
Other [text box]

The *Institutional Data Readiness Assessment Tool* was adapted for this study, specifically offices of sponsored research, and the framework was used for questions 13, 14, and 15. A 5-point Likert Scale of 0 - 4 where No Implementation should be interpreted as "Strongly Disagree" and Full Implementation should be interpreted as "Strongly Agree."

13. PEOPLE One component to ensuring that the institution has the capacity and willpower to act on data. The expertise in place and the ability to develop expertise speak volumes about institutional readiness.

PEOPLE Statements	**No implementation (Strongly Disagree)**	**Under discussion (Disagree)**	**Marginal implementation (Neither Agree or Disagree)**	**Partial implementation (Agree)**	**Full implementation (Strongly Agree)**
The institution is investing in stakeholder discussions and building consensus around data as a means to improve information flow.					
There is someone on my team who is skilled in data analysis tools or who has the expertise to make sense of data collected for internal use.					
The president and senior leadership emphasize the importance of data and actionable information to the health of the Research enterprise at the institution.					
The president and senior leadership work closely with central data analytics professionals and are willing to take actionable					

insights to pilot test operational or improvement ideas.					
Faculty and Research administrative staff see a clear and visible connection between data, institutional planning, and resource allocation.					
The institution is open to shaping a new career path for existing Research administration staff which includes building skills in data analysis.					
There is frequent dialog among Faculty and Research leadership about what information is critical for the institution to know.					
User groups for Research databases have been clearly defined and contain a variety of end-users including administrative staff, faculty, analysts, accountants, and technical people.					
Research administrators at my institution regularly use data analysis tools to support decision-making.					
The institution regularly conducts surveys and focus groups with faculty and Research administrative staff to identify weaknesses in programs and services and opportunities for improvement.					

14. PROCESSES Explore the interactions among people and guidelines necessary to ensure that data are shared widely and processes are in place to produce information that the institution can use.

PROCESSES Statements	**No implementation (Strongly Disagree)**	**Under discussion (Disagree)**	**Marginal implementation (Neither Agree or Disagree)**	**Partial implementation (Agree)**	**Full implementation (Strongly Agree)**
Research enterprise **data input** at the institution is typically viewed as being					

reliable for decision-making.					
Research enterprise **data output** at the institution is typically viewed as being reliable for decision-making.					
There is an understanding of the difference between official data (maintained by the entire institution) and unofficial data (maintained by one or more individual offices or units).					
The institution recognizes that data from various offices and departments (and data collected at different times) may not agree. Accordingly, it has instituted a process for reconciling competing information.					
The institution has acquired a System To System (S2S) software package for proposal submission and/or award management.					
Members of the campus community participate in the planning and priority setting using data to formulate strategies to measure success.					
Institution provides training to faculty and administrative staff on using data and research to improve programs and services.					
Routine training is held to assist end-users in making the best use of existing data systems and dashboards.					

External professional development training on data analysis tools is available to administrative staff in the Research enterprise.					
There is a concentrated effort to improve user knowledge about what data exist and where they can be obtained.					
Policies, processes and procedures related to Research are or have been revised based on insights gained from using data analysis tools and dashboards.					
Data analysis tools and dashboards provide actionable solutions for the Research enterprise.					

15. DATA MANAGEMENT Refers to storing and retrieving information and how information that is critical to the institution becomes transparent with good management.

DATA MANAGEMENT Statements	**No implementation (Strongly Disagree)**	**Under discussion (Disagree)**	**Marginal implementation (Neither Agree or Disagree)**	**Partial implementation (Agree)**	**Full implementation (Strongly Agree)**
The institution has constructed or acquired a database and user-friendly "dashboard" that permits administrators and others to see, at a glance, the status of key indicators.					
The institution has a "data lake" or "data reservoir" for collecting Research data (and other types of institutional data).					
Standard roles and					

responsibilities have been assigned for data management, and it has been agreed who has access to data.					
There are standard written procedures for entering, extracting, editing, auditing, merging, and altering data.					
An official dictionary of Research data terms and definitions is available to all users of data.					
Requests for data are known to other units within the institution so that groups or individuals working on similar questions have the opportunity to share data and expertise.					
Software is available to support statistical analysis and can produce graphical or visual displays of data.					
Software systems allow non-skilled users to create data tables and view reports.					
Data analysis tools and dashboards are used regularly to produce actionable solutions for the Research enterprise.					
Research data are captured using the same categories and codes regardless of who is responsible for collecting and entering those data.					
Databases are regularly monitored to ensure that contents are accurate.					

When problems are found, data is cleaned, edited, and checked to ensure accuracy.					
Procedures exist to ensure that no personally-identifiable data are shared with inappropriate personnel or others outside the institution.					
The institution is developing strategies to mitigate concerns of risk and data security related to Research data.					

Thank you for completing this survey. If you would like to receive compensation for your time, please add your name (first and last) and email address below and a $25 Amazon gift card will be emailed to you within the next 5 -7 business days. (Qualtrics' Anonymized Raffle feature)

Email Address: [text box]

Appendix F

National Council of University Research Administrators (NCURA) Recruitment Approval Letter

NCURA
Supporting Research...together™

7/21/2023

Re: **Factors that Determine the Use of Data Analysis Tools in an Office of Sponsored Research in an Institution of Higher Education within the United States: A Readiness Assessment**

Dear Marywood University, Institutional Review Board Committee:

This letter confirms that as an authorized representative of the National Council of University Research Administrators (NCURA) I am aware of Heather Messina's research project and protocol.

I will allow the investigator to post the survey recruitment materials to NCURA Collaborate and/or contact NCURA members from our membership directory that meet the study criteria as defined in the methodology. **However, activities may commence only after the investigator provides evidence of final approval from Marywood University's IRB or ERC for the proposed project.**

If you have any questions, please contact me at ainsworth@ncura.edu.

Sincerely,

Emily Ainsworth
Director, Regional and Membership Services

National Council of University Research Administrators | www.ncura.edu
1015 18th Street, NW, Suite 901 | Washington, DC 20036 | T: +1 202.466.3894 | F: +1 202.223.5573

Appendix G

Marywood University IRB Approval Letter

MARYWOOD UNIVERSITY
INSTITUTIONAL REVIEW BOARD
Liberal Arts Center, 2300 Adams Avenue, Scranton, PA 18509

DATE:	September 15, 2023
TO:	Heather Messina, MPA
FROM:	Marywood University Institutional Review Board
STUDY TITLE:	[2079272-5] *Factors that Determine the Use of Data Analysis Tools in an Office of Sponsored Research in an Institution of Higher Education within the United States: A Readiness Assessment*
MUIRB #:	2023-002
SUBMISSION TYPE:	Amendment/Modification to Revision #1
ACTION:	APPROVED
APPROVAL DATE:	September 15, 2023
CHECK IN DUE DATE:	**August 7, 2024**
REVIEW TYPE:	Expedited Review
EXPEDITED REVIEW TYPE:	45 CFR 46.110 (b)(1)(i)(7)

Dear Heather Messina:

PLEASE READ THIS LETTER CAREFULLY IN ITS ENTIRETY.
IT CONTAINS IMPORTANT INFORMATION ABOUT YOUR RESEARCH PROPOSAL AND YOUR RESPONSIBILITIES AS AN INVESTIGATOR. THE IRB IS REQUIRED BY FEDERAL LAW TO REPORT ALL SERIOUS OR CONTINUING NONCOMPLIANCE WITH THESE REQUIREMENTS TO FEDERAL AGENCIES.

Thank you for your submission of Amendment/Modification materials for Revision #1 to this research study. Marywood University's Institutional Review Board has **APPROVED** your submission. This approval is based on an appropriate risk/benefit ratio and a study design wherein the risks have been minimized. All research must be conducted in accordance with this approved submission.

Please remember that informed consent is a process beginning with a complete description of the study and assurance of subject understanding, followed by a signed informed consent and/or assent form,

unless a waiver of documentation of informed consent was granted. Informed consent must continue throughout the study via a dialogue between the investigator and each participating research subject. Federal regulations require each subject to receive a written copy of the consent document, unless waived. **A waiver of documentation of informed consent was reviewed and previously approved for both the focus group and the research administrator subjects. Therefore, no signatures are required.**

No new documents required the IRB approval's stamp as part of this revision. We had previously applied the IRB's approval stamp to the following documents, as well as to the waiver of documentation of informed consent request, which are still available in IRBNet (While viewing Project Overview for your study > Reviews tab on left > Under Board Documents at the bottom > Next to package #3). The stamp must appear on versions shared with subjects wherever possible. If it is not feasible to use the stamped versions online (e.g. some email systems or survey platforms), please ensure that the language in the transmitted versions is identical to the stamped versions.

1. Email Recruitment Messages and Flyer
2. Poster
3. Informed Consent Form for Focus Group (Already Completed)
4. Informed Consent Form for Research Administrators

Please also note that:

- **CLOSURE REPORTING:** Upon completion of the research, you must file a closure report form via IRBNet.
- **CHECK IN REPORTING OR CONTINUING REVIEW:** If activities will continue beyond your approval's one-year anniversary of **August 7, 2024**, file a check in report form by that date.
- **RECORDS RETENTION:** You must retain records for a minimum of three years after the official closure date in IRBNet.
- **DEVIATION, UNANTICIPATED PROBLEM OR SERIOUS ADVERSE EVENT REPORTING:** If any of these events occur, you must file the appropriate form immediately via IRBNet.
- **REVISION REQUESTS:** If you decide to make procedural or document changes to your approved project, you must file a revision request form for review and approval prior to implementation, except when necessary to eliminate apparent, immediate hazards to the subjects. In hazardous situations, you must file the form immediately afterward.

The appropriate forms for any of the reports mentioned above may be found at irbnet.org. After logging in, click the Forms and Templates button on the left menu, or find the library after you begin a follow up package within your existing project (Designer button on the left menu, followed by the blue "Need forms" link on the main screen, which opens the library under Step 1).

If you have any questions, please contact the Research Office at 570-348-6211, x.2418 or irbhelp@marywood.edu.

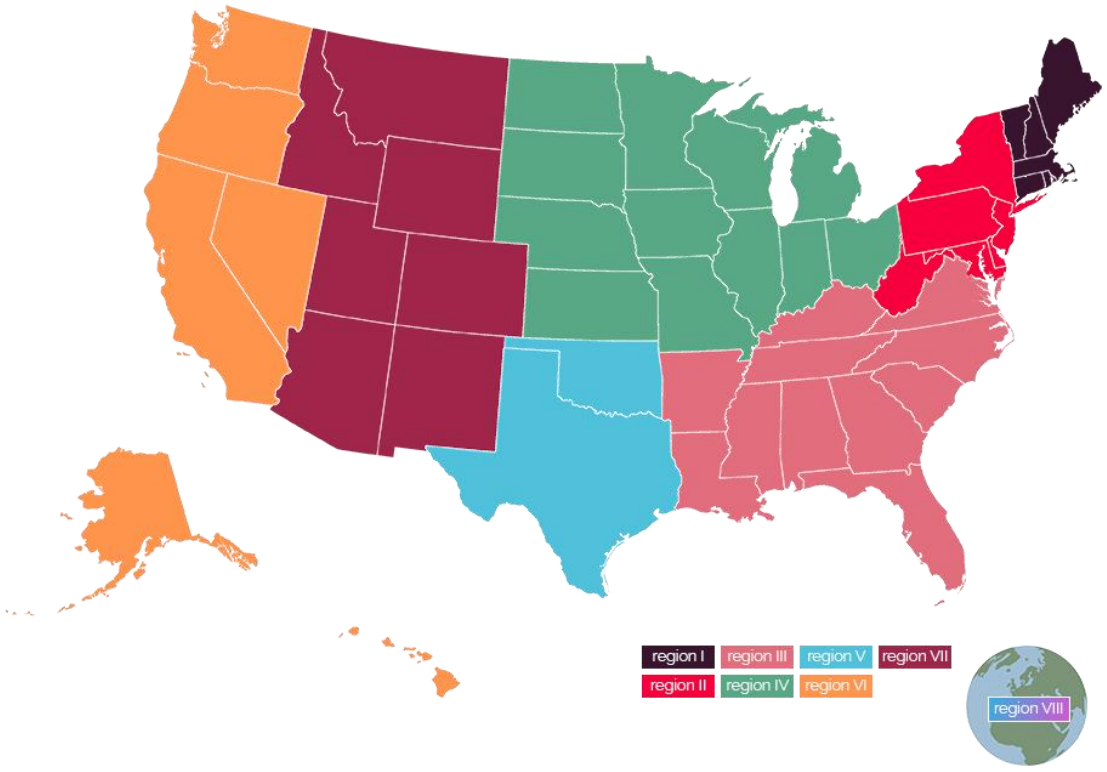
Please include your study title and IRBNet ID number in all correspondence with this office.

Thank you and good luck with your research!

Regards,
Institutional Review Board

Appendix H

Regional Map, National Council of University Research Administrators (NCURA)



Appendix I

SPSS Variables and Coding

Question	Description	SPSS Variable	SPSS Coding
1.	Name of the current institution in which you are employed. (IPEDS dropdown)	INSTITUTION	Text
2a.	What is your institution's classification according to the Carnegie Classification of Institutions of Higher Education (2021)?	Independent Variable; CLASS	Doctoral Universities-Very high research activity (R1); Doctoral Universities-High research activity (R2); D/PU: Doctoral/Professional Universities
2b.	What is your institution's subtype? (select all that apply)	SUBTYPE	Private Institution of Higher Education; Public Institution of Higher Education; Academic Medical Center; Predominantly Undergraduate Institution (PUI); Historically Black College/University (HBCU); Hispanic/Minority-Serving; Institution (MSI); Other [text box]
3a.	What is your primary role in research administration?	Independent Variable; ROLE_1	Central office of sponsored programs; College administration; Department administration; Finance and administration; Provost or other leadership office; Research accounting; Research or grant development; University Research; Compliance; Other [text box]
3b.	Where is your role in the research administration structure?	Independent Variable; ROLE_2	Central office of sponsored programs; College administration; Department administration; Finance and administration; Provost or other leadership office; Research accounting; Research or grant development; University Research; Compliance; Other [text box]

3c.	What is the primary structure of Research administration at your institution?	Independent Variable; STRUCTURE	Centralized administration; Decentralized administration; Shared service administration; Other, please explain [text box]
4.	In the past five (5) years, on average, how many proposals has your institution submitted?	Independent Variable; PROP_VOL	Survey respondents had the choice between a Text Entry; I don't have access to this information. This variable was recoded to "Access proposal data" = 1 and "No access proposal data" = 0
5.	On average, how many new grant awards, encompassing all types and sources, did your institution receive in Fiscal Year 2023?	Independent Variable; AWARD_VOL	Survey respondents had the choice between a Text Entry; I don't have access to this information. This variable was recoded to "Access award data" = 1 and "No access award data" = 0
6.	Within the last 3 years (most recently completed FY20 - FY22), what is your institution proposal / award success or yield rate?	YIELD_RATE	Less than 5%; 5% - 10%; 11% - 15%; 16% - 20%; 21% - 30%; 31% - 40%; 41% - 50%; More than 50%; I don't have access to this information
7a.	How many Research Administration staff directly report to you?	STAFF_1	Less than 5; 6 – 10; 11 – 20; 21 – 30; 31 – 40; 41 – 50; More than 50
7b.	How many Research Administration staff make up your department?	STAFF_2	Less than 5; 6 – 10; 11 – 20; 21 – 30; 31 – 40; 41 – 50; More than 50
8.	What is your age range?	Independent Variable; AGE	18 – 24; 25 – 34; 35 – 44; 45 – 54; 55 – 64; 65+
9.	What is your gender identity?	Independent Variable; GENDER	Male; Female; Non-binary / Third gender; Prefer Not to Answer
10.	How many years of experience in research administration do you have?	Independent Variable; YRS_EXP	Less than one year; 1 – 5; 6 – 10; 11 – 15; More than 15 years
11.	What is your current salary?	Independent Variable; SALARY	$0 - $30,000; $31,000 - $60, 000; $61,000 - $90,000; $91,000 - $120,000; $121,000 - $150,000; $151,000 +; Prefer Not to Answer
12.	Does your position description or current job	ROLE_3	Possible survey answers were Yes, and it is part of my position

	duties include data aggregation, analysis, or reporting for interdepartmental (inter-college) decision-making? *Data could be used for a number of things internal to your sponsored research office(s) including, but not limited to, strategic planning, infrastructure or staffing decisions, proposal development or center/institute focus.*	became the Dependent Variable; USAGE	description; Yes, and it is not part of my position description; No; I'm not sure Variable was recoded into a new variable during data cleaning to form dichotomous variable USAGE. Recoded as Tools Used; Tools Not Used
12a.	<conditional question> If not you or your position, is there someone within the research enterprise at your institution whose role it is to perform research data aggregation, analysis, or reporting for internal use?	NO_PERSON	Yes; No; I'm not sure
12b.	<conditional question> Does your institution have data analysis tools for use in research administration (e.g., Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau)?	NO_TOOLS	Yes; No; I'm not sure
12c.	<conditional question> Which data analysis tools do you use? Select all that apply.	TOOLS became a new scored variable TOOLS_SCORE	Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau, Other [text box] The highest possible tools use score was 12.
12d.	<conditional question> How often do you use data analysis tools (e.g.,	FREQ_TOOLS	Every day; Once or twice a week; Once or twice a month; Quarterly; Annually; Never

	Access, Argos, Apache Spark, Jupyter, Microsoft Excel, PeopleSoft, PowerBI, Python, SAS, SPSS, Tableau)?		
12e.	<conditional question> How did you acquire skills in data aggregation, analysis, or reporting? (select all that apply)	LEARN became a new scored variable LEARN_SCORE	Current employer institution; Previous employer in higher education research administration; Previous employer in industry/for-profit company; Advanced degree or technical training in data analysis (this includes courses during undergraduate education); Self-taught or self-learner; Online or certification training; Other [text box] The highest possible learn score was 7.
12f.	<conditional question> What are you using data analysis tools for? (select all that apply)	USE_CASE became a new scored variable USECASE_SCORE	Routine reporting (monthly or quarterly); Budget planning and resource allocation; Calculate proposal volume; Calculate award volume; Calculate institutional or organization unit research expenditures; Predict proposal volume; Review or predict research staff workload; Predict award volume; Calculate or predict success rate for proposals; Anticipate five + year research activity; Summarize organization unit activity for Board of Trustees or other external stakeholders; Summarize organization unit activity for faculty, staff or other internal stakeholders; Explore data by looking for patterns and trends; Other [text box] The highest possible use case score was 15.
12g.	<conditional question> Who is the main target audience for your data output, analysis or	AUDIENCE became a new scored variable	Myself and staff within my organizational unit; College Dean; Department Chairperson; Faculty

	reports? (select all that apply)	AUDIENCE_SCORE	Provost or Vice President/Provost of Research; External Partners or Stakeholders; Other [text box] The highest possible audience score was 6.
The *Institutional Data Readiness Assessment Tool* was adapted for this study, specifically offices of sponsored research, and the framework was used for questions 13, 14, and 15. A 5-point Likert Scale of 0 - 4 where No Implementation should be interpreted as "Strongly Disagree" and Full Implementation should be interpreted as "Strongly Agree."			
13. People Statements	One component to ensuring that the institution has the capacity and willpower to act on data. The expertise in place and the ability to develop expertise speaks volumes about institutional readiness.	PEOPLE_SCORE	No implementation (Strongly Disagree) = 0; Under discussion (Disagree) = 1; Marginal implementation (Neither Agree or Disagree) = 2; Partial implementation (Agree) = 3; Full implementation (Strongly Agree) = 4. The highest possible raw score in the People matrix table was 10 x 4 = 40
14. Processes Statements	Explore the interactions among people and guidelines necessary to ensure that data are shared widely and processes are in place to produce information that the institution can use.	PROCESS_SCORE	No implementation (Strongly Disagree) = 0; Under discussion (Disagree) = 1; Marginal implementation (Neither Agree or Disagree) = 2; Partial implementation (Agree) = 3; Full implementation (Strongly Agree) = 4. The highest possible raw score in the Processes matrix table was 12 x 4 = 48
15. Data Management Statements	Refers to storing and retrieving information and how information that is critical to the institution becomes transparent with good management.	DATA_SCORE	No implementation (Strongly Disagree) = 0; Under discussion (Disagree) = 1; Marginal implementation (Neither Agree or Disagree) = 2; Partial implementation (Agree) = 3; Full implementation (Strongly Agree) = 4. The highest possible raw score in the Data Management matrix table was 13 x 4 = 52

Scored Variable	This variable is the sum of the total scores of People Score, Processes Score, and Data Management Score.	TOTAL_SCORE	The highest possible total raw score for the combined matrix tables was 140.
Recode Variable	This variable is the categorical quartiles of TOTAL_SCORE.	Dependent Variable; DATA_SCORE_QUAD	This was recoded to a categorical variable No implementation; Under discussion; Partial implementation; and Full implementation
Scored Variable	Question 12c. Number of software tools used; Question 12e. Number of courses taken or ways learned; Question 12f. Number of different reports generated; and Question 12g. Number of audiences with whom reports are shared.	SKILLSET	One point was given for each option checked in the "select all that apply" questions. Therefore, the highest possible total sum score for skillset was 40. Skillset sub-questions were conditional based on participants' response to Question 12.
Recode Variable	This variable is the categorical levels of SKILLSET	SKILLSET_CATEG	This was recoded to a categorical variable with three levels: Foundational skills; Intermediate skills; and Advanced skills
Added in dataset	Region, as defined by the National Council of University Research Administrators (NCURA)	REGION	Region 1; Region 2; Region 3; Region 4; Region 5; Region 6; Region 7 *(See Appendix H)*
Added in dataset	2021 NSF HERD Survey data for institutional rank	RANK_HERD	Numeric where the lower number indicates a higher rank of the institution on a scale. This variable was recoded to quartiles: Top 25% of Rank, 50% of Rank, 75% of Rank, Bottom 25% of Rank
Added in dataset	2021 NSF HERD Survey data for institutional research expenditures	Dependent Variable; EXPEND_HERD	Numeric in Dollars ($) where a higher dollar amount indicates a higher rank of the institution on a scale. This variable was recoded to quartiles: Top 25% of Expenditures, 50% of Expenditures, 75% of

			Expenditures, Bottom 25% of Expenditures