Research Analytics Academic Publications   Research Analytics

2021

Best Practices for Research Analytics & Business Intelligence within the Research Domain

Baron Wolf
University of Kentucky, baron.wolf@uky.edu

Terri Hall
University of Notre Dame

Katherine Robershaw
Berea College

Best Practices for Research Analytics & Business Intelligence within the Research Domain

Baron Wolf
University of Kentucky

Terri Hall
University of Notre Dame

Katherine Robershaw
Berea College

ABSTRACT

While an increased number of industries with business operations have been shaping their use of data analytics, the use of such tools and methods within the higher education research domain, specifically research administration, is still in its infancy. This mixed-methods study collected data to identify best practices in how universities and other research organizations use data analytics to drive their strategic agendas, create efficiency in operations, and promote complex research proposals throughout their institutions. Research methods included a survey to collect data on how research offices are using analytics and business intelligence tools, Rasch analysis (Rasch, 1993) to examine survey instrument quality and provide insights into the use of analytics and business intelligence tools in research offices, and interviews with research administrators and stakeholders to identify best practices in using data tools to impact their decisions, processes, and programming. Results from the Rasch analysis showed that except for two recommendations for individual scales, all survey scales exhibited satisfactory reliabilities and rating scale performance. Findings from interviews revealed best practices such as clear ownership and definitions of the data entry process, identified stewards of each of the high-level areas of data, and confirmed understanding of terms after data requests.

Keywords: data analytics, research administration, business intelligence, higher educational institutions, best practices, Rasch analysis

INTRODUCTION

There has been an explosion in the use of data to inform decision making throughout business, industry, and higher education (Brynjolfsson & McElheran, 2016). However, based on a review of sessions offered at professional association meetings such as the National Council of University Research Administrators (NCURA) and the Society of Research Administrators International (SRAI), it appears that the use of such tools and methods within the higher education research domain are still in their infancy as evidenced by our survey and interviews. Yet we have seen a growing number of sessions each year about data, data analytics, and business intelligence, reflecting the need for and interest in the subject. Some institutions are ahead of the game and have deployed sophisticated analytic tools and data science methodologies that are used in their decision-making process, although they are largely focused on student enrollment, predictive models for tuition, and student retention (Picciano, 2012). Between 2010 and 2017, the number of job postings that listed skills related to data analytics, visualization, and the use of analysis grew by over 1,000% and is expected to continue to grow in all fields (Ryan, 2018). As the marketplace and funding opportunities for sponsored research become more competitive, and funders hold research entities more accountable, research administration professionals will need to learn new skills that employ the use of advanced data analytics, data science, and business intelligence tools to remain competitive and sustain research programs and growth.

The focus of this research was to collect primary data to identify best practices in how universities use data analytics and data tools or software to drive their strategic agendas, create efficiency in operations, and promote complex research proposals throughout their institutions. For the purposes of this research, *data analytics* included the use of dashboards, data science, network analysis, bibliometrics, and similar analysis or data-driven approaches to drive strategic priorities and decision making within research organizations.

RESEARCH ADMINISTRATION

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. The use of data also gives us confidence in our strategies. Data scientist and author D. J. Patil believes that "a data driven organization acquires, processes, and leverages data in a timely fashion to create efficiencies, iterate on and

develop new products, and navigate the competitive landscape" (Patil, 2011, p. 2). The use of data within the university is happening every day. Institutions use predictive analytics to project enrollment trends, provide growth incentives, and target specific student populations (Raini & Malaya, 2015). However, the use of analytics, data-driven decision-making, and understanding the role of business intelligence tools within the research administration profession remains an underutilized strategy.

A wealth of vendor-provided products exist to evaluate research success, scholarly output, bibliometrics, and R&D funding trends (Moral-Munoz, Herrera-Viedma, Santisteban-Espejo, & Cobo, 2020). In addition, institutionally produced analytics using institutional data repositories have started to take shape within the research administration profession—however, institutions often lack understanding on how these tools can create impact in the broader field. Even though faculty researchers are often suspicious (Flahert, 2015) of such tools, analytics, business intelligence, and competitive intelligence can be used throughout the research administration profession to create growth strategy, impact efficiency, and increase research output, discovery, and success. Analytics can be a powerful tool for senior leaders across the research organization in their efforts to identify strategic funding

opportunities, make hiring decisions, and identify areas for increased investment, such as new buildings or lab space, and create an overall growth strategy within the research domain. Additionally, research administration offices can use analytics to impact operational efficiency, highlight lack of workload balance, create innovative pathways to improve processes, and help lead the expansion of the research enterprise. When entities improve their operations—in this case, integrating research-related data analytics and tools—those improvements over time will result in increased efficiencies and greater success (Kasim, Haracic, & Haricic, 2018).

OBJECTIVES AND EXPECTED OUTCOMES

This research had two major objectives. The first was to analyze in more detail the survey results collected in spring 2019 using the Rasch Measurement Model. This helped validate the survey instrument constructs. Second, this research sought to gain more insight into how various institutions are using analytics and business intelligence tools in their research offices. Best practices were identified through qualitative interviews (Nehls, Smith, & Schneider, 2015) with research administrators and stakeholders on how they use data tools to impact their decisions, processes, and programming.

The outcomes of this research include a collection of best practices based on both

quantitative and qualitative data collection that can later be tested at institutions across the country. These best practices can be used by practitioners within research administration offices to impact practices on their own campuses. Analysis of the initial survey data helped develop a potential operating model (see Figure 1). The data-analytics operating model includes both research administration units and senior leadership roles. To impact research

operations, growth, or strategic innovation, institutions need to first assess their business context, mission, and goals. This assessment provides a path for increased use of data to inform insights. In addition to data, organizations need to provide staff with the correct skills and training to effectively deploy business intelligence tools that are supported by organizational leadership.

Figure 1. *Example model for an analytics framework*

METHOD

Data Collection & Survey Instrument

In the spring of 2019, an initial survey to collect data on how research offices are using analytics and business intelligence tools was conducted. The instrument was

originally designed and created based on areas related to the development of data centric strategies and approaches that use business intelligence and data analytic tools. The survey instrument, initial data collection, and this current study were

approved by the University of Kentucky Institutional Review Board (IRB). The survey was distributed to colleagues in research administration or in data analytic positions across the country, using a snowball sampling methodology involving community forums and listservs (Goodman, 1961). This current research expanded on this initial survey collection using the Rasch Measurement Model (Rasch, 1960) and additional qualitative data collection funded by NCURA. The survey instrument included items covering a variety of topics. Questions related to the value of data-driven decision-making; and the usefulness to institutions of the data they collected. The survey asked: does your institution foster a data-driven culture, and what are the current relevant challenges you face when using data analytics within the research domain? Survey respondents (*n*=56) came from both central research administration and departmental level staff from a total of forty-eight different institutions. The majority (68.9%) of those who responded had worked in research administration for at least ten years. Figure 2 highlights the geographical locations of respondents and institution size as represented by the total research and development expenditures from the Higher Education Research & Development survey (National Center for Science and Engineering Statistics, n.d.) conducted by the National Center for Science and Engineering Statistics, a division of the National Science Foundation.
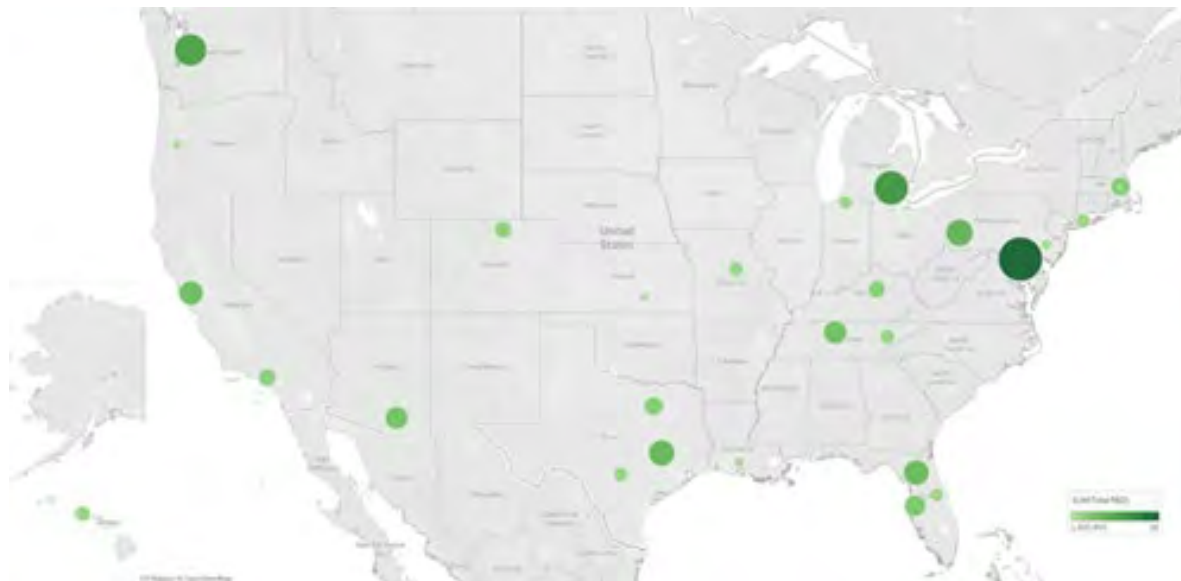

Figure 2. *Map of responding institutions, location, and size of their 2018 R&D expenditures*

Rasch Analysis

The quantitative component of the data analysis was the application of the Rasch rating scale modelling to examine the psychometric properties of the *Research Administration Data Analytics* Questionnaire. Rasch modelling provides a series of rating scale diagnostics to determine whether the major scales within the questionnaire are performing as expected. Rasch modelling also provides indexes and analyses to gather psychometric evidence for establishing unidimensionality, individual item fit, and other properties of the scales. Rasch has been used throughout higher education and is being used more frequently to evaluate survey data in a manner that other statistical tests do not address (Boone, Townsend, & Staver, 2010).

This study used Rasch analysis to analyze six scales within the *Research Administration Data Analytics* Questionnaire for survey validation and informing item revision. Rasch allows for the construction of scales, the monitoring of instrument quality, and a mathematically defensible method for measuring growth and change. Rasch is a theoretical and mathematical approach to measuring a variable. Specific to the survey validation purpose, six scales from the questionnaire were selected to examine the measurement properties and rating scale diagnostics. These scales were named *Trouble* (4 items), *Usefulness* (4 items), *Challenge* (13 items),

Skills (5 items), *Trust* (5 items), and *Data Use* (10 items) (see Appendix II for information about these six scales from the questionnaire). The Rasch model was used in this study because (1) it is the only model that meets the requirement of invariant measurement (Engelhard, 2013); (2) it performs non-linear transformation from raw scores constructed from ordinal observations to compute linear interval measures (Bond & Fox, 2007); and (3) unlike most traditional statistical analyses that find the most appropriate model to fit the data, Rasch requires the data to fit the model (Zi, 2010).

Data collected from the survey were analyzed with the Rasch rating scale model using WINSTEPS 4.4.5 (Linacre, 2019), a Rasch computer software. The Rasch rating scale model was employed in this study because the number of response categories within each of the six scales were the same. The formula for the Rasch rating scale model is displayed below (Linacre, 1997):

$$log\left(\frac{P_{nik}}{P_{ni(k-1)}}\right) = B_n - D_i - F_k$$

where P_{nik} is the probability of a person n achieving category k on item i, $P_{ni(k-1)}$ is the probability of a person n achieving category k-1 on item i, B_n is the ability (B) of person n, D_i is the overall difficulty (D) of item I, and F_k is the step difficulty [threshold] of category k.

Using survey data, the Rasch model performs a range of rating scale diagnosis to

determine if the six scales in the *Research Administration Data Analytics* Questionnaire are performing as expected. Rasch analysis generates dimensionality analysis, reliability and separation indices, item measures and fit statistics, rating scale diagnostics, and Andrich thresholds.

Interviews with Research Administrators

Survey respondents who provided their emails were asked to volunteer for a deeper conversation about data. Eight interviews were conducted with twelve people: five with university staff in the central office, a department and a lab, one hospital and one medical center. The interviewees were in varied roles, such as Assistant Vice President for Research Intelligence, Director of Sponsored Projects, Director of Research Analytics in the central office, Agreements Manager in the Compliance & Contracts Office, Senior Financial Manager, Manager of Financial Planning and Analytics, Associate Director for Research Innovation, IT (Information Technology) staff, and departmental Director of Contracts and Grants. Finally, the level of annual research expenditures from responding institutions varied from less than $5 million to more than $800 million. Although small in number, the interview groups were geographically representative and included wide-ranging experience and expertise.

FINDINGS

Respondent Characteristics

The survey instrument was sent to a variety of listservs online through professional organizations. Fifty-six responses were received from forty-eight named institutions:

- 41% (*n*=23) of the respondents primarily work in a pre-award office (central or dept/unit)
- 14% (*n*=8) were senior level administrators which include vice presidents, provosts, and chancellors
- 55% — The remaining 55% were evenly distributed among post-award offices, proposal development, and compliance, and one respondent was from technology transfer and commercialization

Responses came from both central administration staff and departmental-level staff. The majority (68.9%) had worked in research administration for at least ten years. Further, the majority provided contact information for additional survey follow-up.

Analysis with Rasch Rating Scale Model (RSM)

The quantitative data analysis using Rasch rating scale model revealed the following:

1. The Principal Component Analysis (PCA) of Rasch residuals and disattenuated correlations confirmed unidimensionality of all six scales.

2. All six scales exhibited satisfactory item reliability, item separation, and person reliability, but relatively less satisfactory person separation. This could be due to the homogeneity of the survey sample.

3. The individual item fit statistics confirmed that all items on the six scales are productive for measurement, but four items exhibited infit and/ or outfit mean square statistics slightly outside the desired range between 0.6 and 1.4 logits (see Table 1). This finding prompted researchers to revisit the four items showing slight misfits. Possible follow-ups included investigating whether the items aligned with the construct being measured and revising the wording and phrasing of the item. This helped inform questions for the qualitative data collection process. For instance, an item from the "Data Use" scale measured how often respondents predicted possible outcomes when considering changes in practice. The absence of the word "data" in the item may have led some respondents to believe that the tasks described in the item did not necessarily involve the use of data, thus responding to the item in ways that caused misfit to the item.

4. The category functions performed well on five of the six scales. For the "Challenge" scale, Rasch analysis recommends collapsing category 0 and 1, (1.28 logit difference), category 1 and 2 (0.94 logit difference) and category 2 and 3 (1.13 logit difference). Categories 3 and 4 of the "Challenge" met the Andrich thresholds requirement of 1.4 – 5 logits (1.69 logit difference) (see Table 2).

Table 1. The Item Difficulty Measures and Fit Statistics of the Six Scales (only showing those slightly out of the range between 0.6 – 1.4 logits)

Scale	Item No. and Content	Measure	Infit MNSQ	Outfit MNSQ
Challenge	**3.2** How relevant are the following data related challenges to your institution in the area of research administration: overwhelming volume of available data	.36	1.67	1.72
Challenge	**3.6** How relevant are the following data related challenges to your institution in the area of research administration: data ownership issues	.11	1.23	1.42
Data Use	**6.2** How often do you individually or in teams do the following: discuss preconceived beliefs about a data issue	-.54	1.35	1.74
Data Use	**6.8** How often do you individually or in teams do the following: when we consider changes in practice, we predict possible outcomes	1.92	1.25	1.50

Table 2. Category response statistics of the "Challenge" Scale

Category Label	Observed Count	%	Infit MNSQ	Outfit MNSQ	Andrich Threshold	Category Measure
0	30	6	.85	.83	None	(-2.35)
1	36	7	1.06	.96	-.78	-1.07
2	104	21	.97	1.01	-.90	1.13
3	160	32	.98	.99	.28	1.00
4	173	34	1.13	1.09	1.39	(2.69)

Results from the Rasch analysis suggested that (1) unidimensional constructs were established in all six scales in the survey; (2) survey item revision was recommended for four survey items listed in Table 1; and (3) response category revision was recommended for the "Challenge" scale. In terms of survey item revision, researchers can carry out qualitative investigations, such as any issues with phrasing or wording, and unclear terminology. Regarding the response category revision for the "Challenge" scale, researchers may consider reducing the number of response categories, or renaming the response categories to make them more distinguishable for survey respondents.

FINDINGS FROM INTERVIEWS WITH RESEARCH ADMINISTRATORS
How Data are Used

All individuals interviewed already captured and used basic research administration data around research activity such as proposals submitted, awards received, and expenditures; however, some dove deeper than others due to the presence of more staff, better tools, and higher levels of skills.

When asked about some of the big questions that they were being asked to answer about research activity, we received a myriad of responses. One respondent indicated that she used only "proposals submitted, and external dollars expended. We do not use awards because they may skew one year over another making us appear way up when it arrives, but way down when it ends and that's not a realistic picture". Others talked about the use of trend data, which can often prove to be an issue—depending on the source of information or availability. One example was: "our Provost has asked me to pull 30 years of our NIH R01 history and she wants to know where our faculty were—at what point they were in their career, what rank they were at when they got their first R01. I don't have a lot of that human resources data and in the last 30 years we've switched systems so I'm having to work with institutional research to get some of that historic data and how faculty move through the ranks." Respondents indicated that they used data for both standard reporting but also may often have "unstructured

conversations" that led them to explore data sources, analysis, or data output. One individual pointed out that "I probably spend more time explaining what the data is not than what the data is". Additional selected interview responses are included in Appendix I.

Responsibility for Reporting on Research Activity

"I am a team of one", we were told a few times. That meant little or no time to investigate tools and methods that are available to make the process more efficient or more impactful. The focus was on getting things done. Another respondent said, "We have a data analytics and reporting group and we started building some pages with tables and dashboards but it really was a project that never took off because of staffing and ownership." Ownership and buy-in seemed to be a central theme in the successful use of a data-informed culture. Investment in transparency, skill building, and dedicated time is critical.

Building an Analytics or Research Business Intelligence Program

Most of those interviewed started their career in sponsored programs and currently still do or report to the Vice President/Chancellor for Research. According to one interviewee, "The Sponsored Projects office has their own team who own and run Tableau." Another shared: "I was in Sponsored Projects until a couple of years ago, now I work entirely in

data analytics and intelligence under the Senior Vice President for Research." There seemed to be a trend in which staff started in sponsored project offices and positions and ended up doing some type of work using internal and external data that transitioned their work into a data-centric role for the institution. In essence, they became a subject matter expert in research administration who now managed data, analysis, and business intelligence using data from the unit. These individuals often had analytic skills and naturally fell into data management for their offices.

This progression in responsibility for data analysis led us to ask further questions related to building analytic teams and hiring new team members. It was acknowledged that "to find a sponsor programs person who has analytics experience is a needle in a haystack"; therefore, some institutions tried to hire data analytics staff to complement their sponsored projects office staff. A data person can learn the research administration body of knowledge, but it was often difficult to learn technical skills required to complete forecasting, dashboard development, and robust data management and quantitative analysis. It was acknowledged, however, that while data management and analytics expertise were needed, it must be paired with the

individual learning at least the basics of research administration, how it works, and what is needed.

Data Tools & Quality

While most of those interviewed said they used Microsoft Excel (one of their most invaluable tools), others had access to a variety of tools. Table 3 provides a common list of data and analytic tools used by research administration professionals. While other tools, software, and vendors exist, this list includes those referenced most often by those interviewed for this study. Additional tools included the use of sponsor-provided online data like NIH RePORTER, the HERD survey data tool, PowerBI, Click, and other internal data tools and systems. Some tools, such as Tableau and other analytic platforms, were described as "game changers". They provided easy ways to explore data, create robust data visualizations, and easily communicate information. Many respondents acknowledged that their offices were immature in their effective use of business intelligence tools, either due to a lack of available systems or to a lack of technical skills and knowledge. One respondent indicated that "as we have grown, there has been a realization...that we need to invest in building out analytics and reporting and back-office tools."

Table 3. Top ten tools used, in alphabetical order

Academic Analytics	GIS	PowerBI	Sponsor websites
Access	InCites MathWorks	SmartSheets	Tableau
Excel			

Many respondents have noticed that report requests are becoming more sophisticated and complex. They involve not only research activity data, but human resource and financial data to answer such questions as "do our assistant professors submit more proposals than those with tenure?" or "if awards are up, why are expenditures down?" Even when armed with sophisticated data tools, staff are continually being asked new questions, or to conduct additional analysis, and this is a trend that is not soon going away. One challenge identified is gaining access to additional data elements that are owned by offices other than sponsored programs. This might include salary data, faculty effort data, or other internal or external data sources. Staff are often asked to benchmark their institution, but good benchmarking data does not exist beyond the National Science Foundation (NSF) HERD survey. Those institutions with a sound, data-informed culture and growing research analytics domain do not struggle as much with these data challenges. They are provided access to the data they need to help make decision, support strategies, and are partners across the institution.

Best Practices

Transcripts from the virtual interviews were analyzed and coded to help identify possible best practices for the use of data analytics within the research administration field. These identified best practices align with two basic groups: the bigger picture of Data Governance, Strategy, and Data Quality; and day-to-day suggestions about Data Analysis and Expertise. Table 4 provides a summary of the two major areas of best practices and lists each item.

Table 4. Big picture of data governance, strategy, and data quality, and day-to-day suggestions about data analysis and expertise

Best Practices: Big Picture Data Governance, Strategy, and Data Quality	**Best Practices: Day to Day** Data Analysis and Expertise
▪ Clear ownership and definitions of the processes when entering the data ▪ Identify stewards of each of the high-level areas of data ▪ Dedicated analytics person(s) in the central research office ▪ Develop a Data Playbook with all data tables, sources, definitions in a single place ▪ Start small and grow. Understand current before adding new ▪ Before access is given, consult the data owner ▪ Invest in transparent tools available for increased efficiency, time-saving	▪ After data request, confirm understanding of terms; ideally, how the data is to be used ▪ Take time to think through the needed data elements not just those requested ▪ Double-check the data from different sources; errors can lead to mistrust ▪ Don't just provide data, also provide context ▪ Only one data element in each data column ▪ For recurring reports, keep the format the same for easier comparison ▪ Paginate multi-page reports ▪ Explain your data sources, any limitations to the data, any anomalies

Best Practices – Bigger Picture around Data Governance, Strategy, and Data Quality

The following is an overview of best practices organized by theme from conducted interviews. The overview includes selected quotes from interviewees who agreed to participate in the study.

Clear ownership and definitions of the process when entering data: "Ownership of the underlying planning and budgeting and forecasting processes are sort of all over the map. There was some that our Information Technology Services owned. There were some that were orphaned. And I think that one of the best practices is that there needs to be clear ownership and it needs to be clear ownership of the processes, as well as clear definition of the processes. And those processes need to be actually stood up and propagated throughout the organization...."

Identify who owns the data, how terms are defined, and processes around how the data are to be entered. Those processes go a long way toward ensuring that data are entered consistently. Start by documenting the current process and why it is what it is. Then from time to time, review it to see if enhancements or efficiencies can be made. And always keep the data in mind. Consider if any change will affect the data, and if so, how that is to be managed. It is a good idea to keep an up-to-date and comprehensive data dictionary.

Identify stewards of each of the high-level areas of data: These stewards are responsible for who has access and what level of access is appropriate. "Having multiple people in charge of different aspects of the data can be a challenge," said one research administrator.

Dedicated analytics person(s) in the central research office: This person or team is responsible for reviewing data quality so reports are accurate; working with business owners to ensure staff are trained in the business process; and holding them accountable when data are not entered correctly and working directly with information technology staff who manage data systems. It also helps to de-silo. It is good to have a centralized vision and to some extent, standard data-related business processes so everybody is using the same data set and reporting on things in the same ways. An enterprise-wide view can lead the way toward breaking down some barriers, whether they are real or perceived.

Develop a data playbook with all data tables, sources, and definitions in a single place: When there are questions, this reference source can get everyone on the same page with the same understanding. "Not knowing where to go sometimes, especially with a large research enterprise" has been a challenge for many in research analytics. It is a good idea for everyone to know where to go for which types of data; which people are experts when you have questions, how to

analyze and use the data, and how to put the data results in layman's terms so that you can translate the data into useful information and conclusions.

The audience for this data playbook or data dictionary is not likely to be leadership, however. "We have an enterprise analytic system that has a data dictionary", explained one administrator, "But I can tell you that none of the people I'm working with who are asking these questions know it exists or would read it if they did. And that's just a function of their time more than anything else. And so that's why in the beginning when I'm being asked something, what I want to do is document, like reframing the question around the data I know we have, and the data I know we don't have. And telling them, you asked for X. I can give you X minus this or X plus this to kind of get at what you're asking. When I get some agreement on what I think the question should be I can reframe it like a librarian would around what's possible in our systems before I proceed. You have to communicate that while they asked for one thing, the way that we would define that thing with the terms we use on this campus might be called something else. So, I'm going to tell them what that something else is and how it matches what they asked me."

Start small and grow. Understand current before adding new: We would all like to have the gold standard right out of the gate, but if you are going to grow from scratch, that

may not be possible due to funding, data tools, or expertise. That is OK. It takes time and you can work with leadership to set expectations and chart an analytics growth plan.

Functional knowledge is necessary to understand and interpret data accurately – what is missing, what may be wrong and need further review… AND you need someone with data expertise to easily find the stories within the data. Programming skills will be helpful if you wish to build your own online tools.

And finally, someone who can communicate data stories well is needed. You may get lucky and find these skills in a couple of people or be fortunate enough to have the funding to hire a team. Once the people are in place, start with a limited set of metrics that you care about and do a good job on those before you start expanding into new metrics, new data products, and new skills.

"My boss has a great saying: 'let's start with a dirt road before we start paving things.' Don't make it super-permanent and big overhead until you know it's the road you want. Start small, know what you're doing first, with a limited number of metrics. Concentrate on the data quality, and then think about expanding. Because it's really not as much of a technical problem as it is a human problem. You need to get the humans on board with how this

all needs to work before your technical data is going to be of any quality or usefulness."

Before access is given, consult the data owner: When someone starts working with a new set of data, it is important that they understand any nuances it may have, or that they have a good understanding of the terms – what the column "Year" means, for example, so it is clear it means fiscal, not calendar, year. It is good to provide a bit of training or share data definitions so new users understand the context and data owners know how the users plan to utilize the data. Users also need understanding on what they can and cannot do, should and should not do, with data that they typically wouldn't be using. The same goes for those in the central research office who may be asked to pull in new data elements such as academic rank or other data we may not normally use … all users need to confirm they have a good understanding of the data, any nuances, how they may be misinterpreted, as well. Assuming we understand can be a recipe for disaster.

Invest in transparent tools available for increased efficiency, timesaving: Whenever possible, it is helpful to have tools available so that people can view the data or run their own reports when needed rather than relying on a data person or team to do it during a workday. This not only assists the user but frees the data team to work on other projects. Self-service reporting can help enhance stakeholder experience and

help maintain consistent reporting, definitions, and interpretation of data analysis. Whether you develop the tools yourself or hire a vendor to develop them, it is important that the tools have user filters to provide the most popular options but not allow access to the entire data set in case pulling in some fields may lead to duplicate data or incorrect calculations.

Best Practices – Day-to-Day Considerations Around Data Analysis and Expertise

The following additional details from interviews on best practices relate to the data-to-date operations of work in research analytics. Selected quotes are from individuals who participated in the qualitative data collection.

After data are requested, confirm understanding of terms; and ideally how it will be used: It is important to take the time to ask clarifying questions on what the requestor is seeking and to confirm the definitions of terms in a data request to ensure you understand what is being requested and ideally how the data are to be used. Inform the person requesting the information that you need a bit of clarification and will be asking follow-up questions that may seem annoying but will ensure he/she receives the data needed. For example:

"Last year" – does this refer to calendar year, fiscal year, Federal fiscal year—what date range?

"Amount of funding received…" – does this refer to the latest increment, total obligated amount, total project amount (i.e., what's obligated and anticipated)? One of the biggest questions is whether it is awarded dollars or expended amounts—this can often be confusing to those with little knowledge of research administration. Said one analyst, "One of the things that we always have to ask is: Do you want proposals submitted IN a fiscal year or submitted FOR a fiscal year? And understanding that difference. "

"People ask 'What is a project worth'? Okay. What do you mean by project? Is it just this budget year or the full five-year, NIH grant? And renewed… Are you talking about over the 10- or 15-year life of this project? And then dates and amounts… Dates are not simple. Dollar amounts are not simple. And with some [requestors] the attitude is 'just give me a report,' you know, like they do not want to have that conversation with us, because they don't really know. And so, we have moved lately to a model of getting what we can out of the user or the request. We take our best guess based on what they tell us they want to use it for. We try to pick the data elements that we think they need. And then there is a discussion on the results. We show them the report and they say well this is not quite what we are looking for. And we do it again**.** It's very much an iterative process, and an education process."

"Our upper administration has always been supportive and say, 'That's okay, you can be annoying. We understand the value of you having these checks in place.' They don't say, 'You've got to back off. You're irritating the users.' I think they strike a very thoughtful balance between ensuring quality, while minimizing the administrative burden."

Take time to think through the needed data elements not just those requested: Too often a report is set up and its use has begun, only to realize that additional fields are needed, and you need to revise the report or that what they thought the report included was not what they had envisioned. One institution we interviewed gave the example of a request from a senior administrator to talk about funding by department and college or division. She had everything in the report by department. But the report did not include a column for "college", so she had to add that. Know that you cannot think of everything, and needs will likely change over time, but it is good to think about needs from the start – what has been requested in the past and might be needed in the future – to structure the data based on how they will be used over time.

Double-check the data from different sources; errors can lead to mistrust: The goal is always to provide accurate information; by double-checking the data, we mean pulling separate reports or re-calculating it. We are all human and mistakes WILL OCCUR but

overall, we want our customers to be able to trust the data we provide. Errors can lead to mistrust. One institution we interviewed told us, "the last thing we want is to provide a report and somebody says, 'That's not right', because once they see something that's not right they don't trust us anymore. So, we take it very seriously."

What are some ways of double-checking the data? "Whenever we are collecting a new data set, we also ask ourselves 'how do we think about double checking it? Can we create electronic cross checks or validations with other data sets?'"

Another administrator shared her method: "After any report I run, I always take five (data items) and I check whatever I've got, and then I cross check them in a couple different systems, and make sure it looks right. And if I get five that are right, then I trust the whole thing. I just figure statistically the odds are it's right."

Do not just provide data, also provide context: It is good practice to provide a small analysis of what the data are saying rather than asking the customer to do it. The customer may see different stories in the data (which is great when shared with you), but often customers appreciate you not only providing the data, but explaining what it shows – Any trends? Any patterns? Are we up or down from last year, last month? Is a downturn in expenditures due to a loss of funding or is it merely that a big project has ended? Be transparent about what the data

mean and what they do not mean. This transparency can assist value-based decision making.

Only one data element in each data column: Did you separate last name and first name? Do you need to separate month and year in your dates to be able to report on monthly figures? It is good practice to have just one element in each column of your dataset.

For recurring reports, keep the format the same for easier comparison: This goes not only for column headings, but any formatting such as color, look, and flow of the reports. If red means concern in the first report, it should remain as signifying concern on future reports.

Paginate multi-page reports: This helps when there are questions about the data and being able to locate it more quickly so each of those involved in the discussion are looking at the same thing.

Explain your data sources, any limitation to the data, any anomalies: Several of the folks we interviewed brought this up, and the example was about COVID-19 or CARES funding. They told us, "We'll show a spike this year in our awards and expenditures for the fiscal year. While this may look like growth, it is not true growth, it is pandemic relief funding. Similarly, a drop does not mean you are doing worse. It could just be the state decided to give us all those contracts earlier last year than they did this year, or a big project ended." Including these anomalies in your reports or

dashboards is key to understanding the data now and in the future when these anomalies may not be top of mind.

CONCLUSION

Based on the findings of the Rasch analysis, we can see that while all constructs had validity and were unidimensional, it was advisable to re-phrase a few items and change some of the scales. This ensured that respondents would better understand the item(s) being asked. For instance, items concerning *data challenges* and *data use* should be revised on future survey instruments to better capture unidimensionality for those constructs. This is common in survey research development. Since the Rasch analysis did confirm the constructs through the various tests, we are confident with our findings. This analysis allowed us to better prepare and focus our qualitative data collection and interview items.

According to study participants, the biggest challenges in leveraging data analytics were the difficulty in identifying valuable audiences, and the knowledge and skills in data use needed to gain insights that triggered actions. Our participants had mixed perceptions of the usefulness of data analytics and metrics in improving customer satisfaction and office or research community. Some believed the incorporation of data analytics in research administration will create value, while others had mild reservations. The

implementation of data analytics and business intelligence poses considerable challenges in the higher education research domain. Some of the biggest challenges include the cost of data stewardship, corporate culture, and privacy concerns and risk. Availability of data, data quality, and shortage of talent / skills are some of the least important data-related concerns in research administration in higher education institutions. In terms of data-related skills in the workplace, our study participants acknowledged that talents and skills in data use and data management and analytics are not difficult to find, but they currently do not have access or available resources for these positions. Nonetheless, they also acknowledge that their leadership provides enough professional development opportunities about data use, and that staff assist their units change their practices, processes, and policies based on data. Our participants exhibited high levels of trust and respect for colleagues who are experts in their specific areas. Our study findings are consistent with observations from professional association meetings (e.g., NCURA, SRAI) that the use of data tools and methods within the higher education research domain are still in their initial stages for many of our study participants. Most reported having very low frequencies in engaging in tasks that involve the use of data, such as exploring data to look for trends and patterns, devising questions that

they will seek to answer using data, identifying solutions, drawing conclusions, and informing changes in practice using data.

Our quantitative and qualitative analysis both supported our suggested analytics framework and model as presented in Figure 1. Institutions must begin by initiating a business context—look to data quality, governance, and resources—and then build an expert team with skills in both research administration and data analytics. The context is very important. Some small institutions may not have a need for robust research analytics due to their mission and organization vision. While findings on this are mixed—it was clear that to have a rich data-informed culture, an institution needs to invest in dedicated data analytics and business intelligence teams and tools. While those individuals do not necessarily need to be trained in analytics—there is a benefit to having technical data staff as a dedicated resource.

Growing an analytics culture takes teamwork, a clear vision from senior leadership, and resources. While this area is growing within the research administration domain, there are clear competitive advantages to building such a culture within higher education and other research-intensive organizations. Best practices are focused on general themes related to data governance, strategy, and data quality

while others are focused more on day-to-day operations which relate to how to use data for effective decision making. Having a framework and clear analytics strategy is key to having effective use of the rich information and data institutions have related to sponsored grants and contracts. It can be a journey—but one that can confirm existing strategies and help drive research forward.

ACKNOWLEDGMENTS

LITERATURE CITED

Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the human sciences* (2nd ed). Mahwah, NJ: Lawrence Erlbaum Associates.

Boone, W. J, Townsend, J. S., & Staver, J. (2010). Using Rasch theory to guide the practice of survey development and survey data analysis in science education and to inform science reform effort: An exemplar utilizing STEBI self-efficacy data. *Science Education, 95*(2), 258–280.

Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making, *American Economic Review, 106*(5), 133–139.

Engelhard, G. (2013). *Invariant measurement: Using Rasch models in the social, behavioral, and health sciences.* New York: Routledge. Retrieved from: http://site.ebrary.com/id/10691755

Flaherty, C. (2015). Refusing to be evaluated by a formula. *Inside Higher Education* (December 11; online).

Goodman, L. (1961). Snowball sampling. *The Annals of Mathematical Statistics, 32*(1), 148–170.

Kasim, T., Haracic, M., & Haracic, M. (2018). The improvement of business efficiency through business process management. *Journal of Economics and Business, 16*(1), 31–43.

Linacre, J. M. (2019). Winsteps (4.4.5) [Computer software]. Retrieved from: https://www.winsteps.com/winsteps.htm

Linacre, J. M. (1997). Guidelines for rating scale and andrich thresholds. MESA Research. (Chicago: Midwest Objective Measurement Seminar). Retrieved from: https://www.rasch.org/rn2.htm

Moral-Muñoz, J. A., Herrera-Viedma, E., Santisteban-Espejo, A., & Cobo, M. J. (2020). Software tools for conducting bibliometric analysis in science: An up-to-date review. Profesional De La Informacion, *29*(1).

National Center for Science and Engineering Statistics, Higher Education Research and Development Survey. (n.d.). Retrieved from: https://www.nsf.gov/statistics/srvyherd/

Nehls, K., Smith, B. D., & Schneider, H. A. (2015). Video-conferencing interviews in qualitative research. Pp. 140–157 in S. Hai-Jew (Ed.), *Enhancing qualitative and mixed methods research with technology.* New York: Information Science Reference.

Patil, D. J. (2011). *Building data science teams*. Sebastopol: O'Reilly Media.

Picciano, A. (2012). The evolution of big data and learning analytics in American higher education. *Journal of Asynchronous Learning Networks, 16*(3), 9–20.

Rajni, J., & Malaya, D. B. (2015). Predictive analytics in a higher education context. *IT Professional, 17*(4), 24–33.

Rasch, G. (1993). *Probabilistic models for some intelligence and attainment tests* (ERIC).

Rasch, G. (1960). Probabilistic models for some intelligence and achievement tests. Copenhagen: Danish Institute for Educational Research.

Ryan, L. (2018). Visual storytelling with Tableau. Boston: Addison-Wesley.

Zi, Y. (2010). Objective measurement in psychological science: An overview of Rasch model. *Advances in Psychological Science, 18*(8), 1298–1305.

APPENDICES

APPENDIX I
SELECTED QUOTATIONS FROM VIRTUAL INTERVIEWS

How Data are Used

All of the people interviewed capture the basic data around research activity: proposals submitted, awards received, expenditures; however, some dive deeper than others due to more staff, better tools, and higher levels of skills.

When asked about some of the big questions that they are asked to answer about research activity we received a myriad of answers:

"Just two. It's proposals submitted and external dollars expended. We do not use awards because they may skew one year over another making us appear way up when it arrives, but way down when it ends and that's not a realistic picture. I also use the NSF HERD survey to do a little bit of comparison with our peers, and I track submissions vs. funded proposals for a success rate."

"Our Provost has asked me to pull 30 years of our NIH R01 history and she wants to know where our faculty were-- at what point they were in their career, what rank they were at when

they got their first R01. I don't have a lot of that human resources data and in the last 30 years we've switched systems so I'm having to work with institutional research to get some of that historic data and how faculty move through the ranks."

"I find that it's often an unstructured conversation that leads to going down rabbit holes to figure things out. They understand the conventional wisdom, based on anecdotes around certain things. Like we always engage with industry sponsors in this way. And we always engage with donors in that way, but they're only talking about that in terms of stories they know, people they know, anecdotes they've heard. What I want to do is see how-- or if-- the data supports those stories. Sometimes it does, and sometimes it doesn't but in the end, they know the story they are telling is based on data.

We are also looking at how we can increase partnerships with other kinds of sponsors beyond federal money. This requires looking at what our engagements are now. And not just 'I feel like the College of X does it this way and the College of Y does it another way'. But what does the data actually say? What can we learn from it and what can we infer from the data?

So we're looking at things like data we collect every single year e.g., the HERD survey, but also what the specific slices tell us about whether or not these anecdotes are true. And if they are true, what do we want to do, if anything, to change things? Is it a growth opportunity?

And I think that's where the decision-making process comes in is knowing what the facts are, knowing what it looked like the past five to 10 years and saying okay we do see a trend? We either want to reverse it or we want to take advantage of this trend continuing. And what do we need to do to achieve that?

But a lot of times what we're finding is not material to the conversation they want to have, or it's actually the reverse of what the conventional wisdom is. And then we decide whether or not we need to do anything about the way that's communicated.

> *I probably spend more time explaining what the data is not than what the data is.*

I probably spend more time explaining what the data is not than what the data is." The actual work I'm doing is more about partnerships and different engagements with certain kinds of sponsors…what proposals are looking like trend-wise. I don't mean the number of proposals, but more:

- Are we submitting more to certain kinds of sponsors than others?
- Is our average proposal size growing?
- Are we engaged in partnerships with companies that want to engage on multiple levels?

So, I'm looking at those kinds of things and comparing them over time."

"For internal reporting, we have a handful of canned reports in Business Objects that anybody with permissions can run. There are more than a dozen reports out there that would tell them a five-year history of their department, as far as counts, dollar amounts, dates, or they could identify a single investigator and it would give their activity, either today or over a period of

time. They can run a snapshot of today with what's pending and what's active and what's kind of hanging out there, or they can get a report looking at trends over the last five years. Anybody with permissions can run those reports at any given time.

But we also do an increasing number of ad hoc reports to support proposal applications. For example, the sponsor wants to know how many of this type of award we've had in the past. And so, we would gather that data. We also create reports for internal retreats looking at the history for the department to plan for the future."

"On a broader, institutional level, we try to use data to make our decisions as much as possible. As part of our accreditation, we have to demonstrate that we're using some data driven decision making methods, and one of the things we're trying to work toward is getting a little more granular.

We put together the scorecards for each of our colleges that incorporates data from Academic Analytics -- bibliometric data, books, articles citations etc., but they're also pulling in our research grant expenditures into these scorecards. For each of our academic departments, they are putting together a snapshot for each year. It's frozen in time at the end of the fiscal year."

"We have developed relationships with Google and Amazon and other big data industry providers and have tracked everything from where cell phones are so that we can see where students are congregating and if this space will allow social distancing. Big data and data analytics is something that we promote, encourage, and support."

"We send monthly reports to faculty and our grant accountant has a spreadsheet that is a crosswalk between our budget categories here at the institution, and the funders categories because as we all know there's almost no relationship between the two. It's a many to many, and she sends that to all of the PI's once a month after the previous month closes, and then it's up to each PI - who's also the financial manager of their account - to review that budget recap and if changes are needed to let us know. Some of them are fabulous researchers but can't read a spreadsheet to save their lives."

"In some of our bigger research centers, and our upper administration, there's targeted, strategic thinking that is very intentional around –

- What kind of grants have we been successful with in the past?
- What areas do we want to move into?
- Where are our strengths, where are our weaknesses?
- What are our success rates on all of these different kinds of proposals?
- What is the submission rate by PI, and
- Based on this PI submitting this many proposals, what is the rate at which the PI is getting funded vs. the PI's that are only submitting once a year

In addition to faculty research data, several are monitoring staff workload as well. "We also have teams that are reporting on operational items, and monitoring and benchmarking research administrators to figure out what does workload look like? At what point do we say we need to add more lines in our budget to hire more people?"

"Our data is used to support or deny whether they really do need extra people. So I think we have both very thoughtful strategic work being done with the analytics, down to, you know, 'I just need some numbers'. It's a real wide variety of applications for the data." We asked if they have any insight into criteria for the hiring decision or the formula that is used? For example, if funding at the medical center increases by this amount or percentage then that means we should add another staff member? "They don't share that with us. They ask us for the information, but we don't know what they're using as the hiring criteria. My educated guess is that volume is one component of the request. But I don't think it's a linear equation model where, if there's an X percent increase you get an extra person. I don't think it's that regimented."

Another administrator explained her process to ensure workloads are balanced:

"For our unit every six months, I do workload metrics for workload balancing. And I use proposal volume and review:

- proposals submitted,
- Direct research expenditures
- number of awards with sub-awards,
- number of awards with cost share. And then
- number of sponsored awards

for each research administrator. It actually helps us quite a bit with our workload and when we get new faculty we know how to assign them. We can look at that and see who's under-subscribed or who's over-subscribed if we need to pull something off."

Responsibility for Reporting on Research Activity

"I'm a team of one", we were told a couple of times. And that means little or no time to investigate tools and methods that are available to make the process more efficient. The focus is on getting things done.

Another said, "We have a data analytics and reporting group and we started building some pages with tables and dashboards but it really was a project that never took off because of staffing and ownership. We're mostly just taking data from our eRA and financial systems,

dumping it into Excel and doing pivot analysis on our own. I wish we would be able to leverage Power BI more and move to reports that are more 'set it and forget it' with better transparency and access to those reports. It's just been a struggle to get the right people doing that. We don't have a dedicated person to manage the inputs from our eRA and financial systems, and that's been the struggle."

Goals may be set, but not always shared. "There's no investment in a transparent way for all to see where things are in relation to goals – like you'd have in a dashboard", shared one administrator.

Another analyst is monitoring the bigger picture at her university with one area of focus being on collaborative efforts. "My work is about the bigger picture research administration data," she said, "the data that comes out of proposals, awards, expenditures and publications. We've built a couple of applications on our campus that leverage things that are available, like Google Scholar. They basically scrape all of this data that's openly available for publications online, plus they scrape our enterprise systems, our research administration system that has proposals and awards in it, and our faculty profile system which has other publications that aren't necessarily in Google Scholar although there's quite a bit of overlap there. And the intent of the application is to figure out how our faculty collaborate with one another, how often, and figuring out what that output is."

Building an Analytics or Research Business Intelligence Program

Most in the group we interviewed started in sponsored programs and currently still do or report to the Vice President/Chancellor for Research. "The Sponsored Projects office has their own team who own and run Tableau, and they produce these reports that are for the fiscal year, and publish a report." Another shared: ""I was in Sponsored Projects until a couple of years ago now I work entirely in data analytics and intelligence under the Senior Vice President for Research."

"For 11 years I was the Associate Director of Post Award, which was my last job within sponsored programs, and part of that role over time became pulling data on our research expenditures, our awards, our proposals, doing some trend analysis, finding out what researchers are doing, and who's working in what space. About a year and a half ago, our Vice Chancellor for Research asked me if I would be interested in doing the data analytics work full time. I took the leap and here I am."

"I started in research administration straight out of college. Currently I'm the Senior Project Manager. I have been in this role for about eight years now. Prior to this role I was the Director for the Sponsored Programs Office, and was in that role for about three years. And then prior to that had just been sponsored program staff so on all I've been here for about 15 years now. Prior to that I was a research administrator at another university. "

"I came from industry where things were a little bit more black and white. There was upper management that said, 'you do it this way'. And when I came into the education world it was like 'well, it depends'. And I can't program 'It depends'. So I had to learn to figure out how to program "it depends'."

> *I had to learn to figure out how to program 'it depends'*

"I have an affinity for making things simpler. And I feel like a lot of people don't have math skills. I've always like tried to create tools to make research administration work simpler. I created the budget template. I created a faculty salary look-up in one go-to place. I created a tuition and fees calculator so people just have to pick the semester and resident or non-resident and it tells them exactly the portion of the fees for that month. So, I'm always creating. I love Excel, I like numbers, that's kind of my background.

For several, data is just one part of the job. "My job is about the data that's related to research, and the systems that generate it and collect it. And also, there's a bit of forecasting and competitive intelligence and business intelligence. I have grown up in research administration. But I think my strengths in data, and communicating it around decision making, are the reasons why I'm in the job I'm in now."

"I am part of an IT team within the Office of Sponsored Programs, and we directly support the campus eRA system that tracks proposals going out and awards coming in. We are not integrated in with the finance system for expenditure data. We do some combined reporting but that is not an automated thing. It's not something where we are linked up to the financial system.

My primary responsibility in the group is interfacing with the research areas of the departments who are doing the research and, in the sense that I do a lot of training. I do a lot of listening to hear what their needs are for our IT system, what problems they're encountering. Our whole team does help desk functions.

We're also responsible for all the development. We do the web development, the database maintenance, we're in charge of the Linux servers, so we are responsible for the whole thing which is good and bad. One of the good things is that it's all in our control. We don't have to reach out to someone else. We are responsible. We can do things on our timeline, when we feel like it's necessary. We do our own testing, and training."

These answers made us curious about hiring new team members, and we asked the group: "When it comes to hiring, do you lean toward hiring someone with research administration experience and teach them data or a data analytics person and teach them about research administration?

> *To find a sponsor programs person who has analytics experience is a needle in a haystack*

"I think, to find a sponsor programs person who has analytics experience is a needle in a haystack. So I look for a data analytics person because the analytics people that I have worked with are much better at using the tools. And I think you can teach sponsored programs to an analytics person. I would take a data person. Definitely."

"I think to a certain extent, we need a certain technical skill level, the skills that require years to get good at. But our team members also have to care about learning the research administration side. We would be very wary of hiring somebody who just wanted to be a programmer. That doesn't work for us. We need somebody who's willing to also pick up the subject matter expertise. We look for somebody who has the capacity or the desire to learn the functional aspects as well."

"One thing that I see that is a problem is when there are research analytics people who do not understand research administration. They're not going to be of much value because they're not going to know the questions to ask of the data, and they're not going to know the data that's important, or even something as simple as just using direct expenditures. So I think it's important that the person has that research administration background, or if they don't, they have somebody next to them, that they vet everything with and have a real good working relationship with. "

Building an Organization-Wide Data Strategy

When it comes to a campus or organization-wide data strategy, the group is in various stages:

"We have an institutional effectiveness office that is forming. Institutionally, an effort has begun to standardize tools and methods, also data governance. Our goal is to have common data definitions, get us all to using the same tools, and using the same processes to get the data we need. We also want to make sure that we have proper documentation of all of our tables and that we are going through the appropriate channels to get access to our data so there is some system-wide coordination, that we are laying the groundwork for it right now. We want all of the campuses within our system to report consistently and to be using the same source, the single source of truth."

"We have data governance which is a higher-level decision-making and stewardship group comprised of the people who are the analysts on each of the main systems. And then you have the folks who are overseeing the data warehouse and the reporting solution. They all meet regularly, just to talk about what's going on, what's changed, if there's anything changing about some of the transactional nature of the work they do. We're in the middle of upgrading our research administration system and part of their conversation is 'this is what's coming, this is

what's changing about some of the data collection.' So, they are regularly having those conversations."

Variety of In-Use Data Tools

While most of those interviewed said they use Excel, others have access to a variety of tools:

Top Ten Tools Used, in Alphabetical Order	
ACADEMIC	MATHWORKS
ANALYTICS	POWERBI
ACCESS	SMARTSHEETS
EXCEL	SPONSOR WEBSITES
GIS	TABLEAU
INCITES	

"I use Excel the most. I know there's a lot of other tools and we kind of looked at them but with the pricing and given our volume they just don't make sense. The most important or invaluable data tool that I use is Excel. Whether it's pivot tables and other kinds of filtering or some other thing to get an answer quickly, without question, Excel is my favorite thing."

"We subscribe to Smart Sheets in my office, and we have automated email setup in the system for those types of report reminders we need. For all projects, I have my assistant coordinator build all the final report due dates and a 30 day reminder for whenever it's 30 days from when it's reviewed for all of the projects. We also use that to track our effort reporting because PI's have to be retrained in effort certification every three years, and we use that to track when their three years comes up."

"I use all of the sponsor portals such as NIH RePORTER and NSF's NCSES tool."

"I use Tableau primarily to push out grants and contracts data to our colleges, schools, and campus leadership. I have built dashboards in Tableau to help them predict their volume, identify who among their staff has expertise with certain sponsors and to look at their workload distribution. Tableau was a game changer. We've had Tableau now for about five years. I love it, it's so, so great."

"Our methods are very archaic because we do not have an eRA system. The proposals and awards are entered into an Access database and reports are only as good as what gets entered in there. I think we are somewhat immature in data analytics and BI reporting. Currently the data is mainly in Excel. I've taken some steps to improve that a little bit and make it more of a data-driven approach but what we don't have are business forecasting tools, integrated budgeting tools, project portfolio management tools.

We have this entire digital enterprise transformation effort underway that is a multi-year effort that we're focused on to look at business planning and forecasting tools that would link in with portfolio management tool. So basically, we're thinking about the life cycle of any funding to plan out expected funding and research dollars expenditures and whatnot. We want

to set the overall funding strategy, the budgeting tool would then inform our operating budgets, and forecast our five-year expense plans. We're also looking at Tableau and Power BI.

And once we curate that data then we have our BI analytics tools, Power BI, Click, SAP Analytics cloud, or whatever tool we pick would latch on to that for purposes of dashboard reporting analytics, either ad hoc or ongoing. So that's kind of the vision.

As we've grown, there has been a realization, in the last probably handful of years that we need to invest in building out analytics and reporting and back office tools. They have identified that we need more support. All of our research teams are adding an extra person and have two research administrators versus one."

> *"As we've grown, there has been a realization…that we need to invest in building out analytics and reporting and back office tools."*

Data Cleanliness – Garbage In, Garbage Out

Accurate reporting is dependent on accurate. "There's been some inconsistency on how the inputs are going in" said one project manager, "And we've had a lot of staff turnover in the last year. So, we've done a lot to make sure that the data that's going in is clean so there's a clean data pull of years and effort and everything. Just having someone who has the time to QA the data sometimes is important' she added referring to Quality Assurance. And "Having an individual who knows the data enough to recognize when something is amiss. There'll be times where I'll ask for someone to run a report and I get it, and I'm like, 'Did you look at this? There's duplicates in this data.' So, I think it's having the staff that has the knowledge of the type of work we do, and then having that Excel expertise that is very important.

Another institution feels the same and expressed: "It's a constant struggle with quality control and making sure everybody's on the same page. It's just human nature where if you're keying data, and you know that nobody's ever looking at it, you just kind of get sloppier over time. Other things are going on, you have other priorities, and the quality of that data entry can kind of diminish. But if you know that it's getting reviewed on a regular basis, that it's being used for important things, and there's another set of eyes looking at it you tend to be more careful. I think that's always our goal - to make sure that one way or another, the person who keyed it isn't the person reviewing it."

Many in the interview group keep in mind that the decisions that are being made based on their numbers are very consequential.

> *Bad data is worse than no data*

"Whether it's hiring and firing people based on these numbers or deciding which areas you're strong in and which areas your weak in, if you're basing your strategy on these kinds of things and the data is not accurate, you're making incorrect assumptions that can really have dire consequences if you make the wrong call. It's not only people's jobs and their career on the line, but the organization. If data is used to conclude 'We're really strong here, we're going to double down', and all that data is based on a coincidence or

on a short timeframe or incorrect keying… It can be more damaging than it's worth. Bad data is worse than no data."

Data Access is Often a Challenge

Many have noticed that report requests are becoming more sophisticated and complex. They involve not only research activity data, but HR and financial data to answer such questions as "do our assistant professors submit more proposals than those with tenure?" or "if awards are up, why are expenditures down?"

"One challenge has been getting hold of actual salaries to try to do proper budgeting. At the place that I just left, they had made an executive decision to tie those salaries into the grants management system to use. But it's a matter of integrity and trust. You've got to safeguard that information. Getting the salaries into the grants system really streamlined the process by no longer being held up waiting to get a salary figure. It's very sensitive information so I can see the hesitation. At the same time, we should be a trusted business partner. I mean, we're the people that are helping to guide and plan and forecast for the PI. We should be trusted to work with that information, and it just hasn't panned out that way in the past."

Another shared: "Right now if I need something I go directly to our data warehouse team and say I need access to this and they open up access to the table to me in Tableau. Of course, I make sure I understand what data I'm pulling and what some of the quirks are with that data. Sometimes I need to find the data owner and ask what something means, and I'm told 'Oh, that's actually incorrect data, we've got a different field that we're using to get what you want'. And some are willing to sit with me and talk about what I'm getting access to and how I'm to access it, and then others, it's just been a pretty uphill climb to understand what I'm looking at."

Explaining further, she said: "If I pull a number, I want to make sure that it's right so I may send it to the data owners and ask them to validate it. And if it's not correct, they'll give me the correct data but I also want to know what I was doing wrong there, and I never hear back which is frustrating."

For others, access is not a problem. "I have super user access to our finance system and can go in and see whatever I need to see."

However, even if access is there, the data needed may not be. "I almost never have an access problem. When I say that I don't have the data it's because it's data that we're not aware has been collected ANYWHERE. We're only at the very beginning of figuring out how we could collect what we need in the first place without putting additional burden on faculty researchers. You know, it's great to know what happened to grad students after they left the University. But the person who knows is not necessarily going to fill out a bunch of forms so that I have that data. Our leadership knows the faculty have a lot of administrative burden. And so, every piece of data that we ask above and beyond just adds to their pile. So as much as we'd like to know, we also have to understand what the limits are."

Other Challenges

We heard a couple of times that the level of data expertise for those receiving the data can also be a challenge. Some recipients have more experience than those developing the reports and expect more, while others told us some recipients have little to no experience with Excel and don't understand what is provided. Unless questions are asked, the answers and training for future understanding may not be provided.

"We have these new hires coming from outside organizations where there was stronger reporting or analytics available to them, and they come here and there's nothing but Excel. Or they're sent data in Excel, with some pivots, and if they don't know Excel, it's terrible for these folks."

"If we would have had a dedicated analytics person, maintaining the dashboards that we initially built it'd be much better for everyone (report creator and recipient) but the dashboards were completely abandoned due to other priorities" bemoaned another.

Understanding how someone plans to use the data is an additional challenge, especially when the user is combining it with other institutional data and coming up with erroneous conclusions. Said one analyst: "We're always kind of discovering some new place that people are getting data from, and it's sometimes stressful because of the way they're doing it… They're taking our data and combining it with somebody else's data and the conclusions drawn are not always accurate. We're like, 'Whoa, what are you doing?' because it comes back to us to explain why that conclusion was drawn."

Opportunities – Dreaming Big

We asked each person we interviewed: "If you had an unlimited budget, what is your ideal infrastructure around business intelligence and data analytics?" Here are some of the answers:

"I think if you have a dedicated person who can understand the research administration work and what is happening, they can set up a great visual tool for leaders to make decisions. Something that shows the volume of proposals going out, the award volume, the expenditure data, the different sponsors. Right now, we're just really using Excel to do that, and when you're using Excel, it lives with one person, it gets updated irregularly, where I think if we could use Power BI or another analytics tool, and have those reports available to people who are in leadership positions, and leaders use them, I think would be beneficial."

"You need someone with an analytics perspective, a data visualization perspective, and from a process perspective. We've also recently hired someone with other skillsets such as data scientists and expertise in data structure."

"I would create a core group of a few individuals, a few experts to really pull together all the information resources from across the research enterprise and create a data library of sorts as a central resource for everyone. Kind of like a core facility around data. And the experts in this core would be experts in different types of data. This core group would be in the Office of Research and connect to our technology transfer people, connect to our clinical research people, so that they can then help, whether it's proposal side or if it's a new tech transfer, the patent license, that they can really help move things forward and facilitate more interdisciplinary and progressive thinking."

"I wish I had more of the data science background. I feel like that's where I'm lacking. Not just statistics, but a bit more of the technical side - being able to use R, use Python, know what statistical information is important and helpful. That is actually my goal in hiring as we move forward with hiring more analysts. I would like to (for one of them at least) to have a bit more of that data science background."

"We're moving towards this ultimate, holy grail of a dynamic dashboard that's drillable, and people can dive into data at various levels. Our interim step is we took our monthly operations report which used to be kind of a long word document and converted it into more of a PowerPoint presentation. And this isn't the end game for us but what it's doing is, it's challenging the way people organize and present data. It's getting them used to a centralized operating model. I'm no longer the conduit by which this report gets updated. The legacy report was a bunch of charts and graphs interrupted by bullets and sentences. We've replaced them with all these Power BI visuals."

"I would have more people trained on Tableau and generating visualizations aside from data analysts who are good at pulling data out and getting answers out of data. My next real wish would be actual data scientists and programmers who are good at AI and natural language processing and machine learning. We've hired someone who's experienced in machine learning and can figure out how to make that process faster**.** It's made a world of difference in just combing through data from our grants management system which includes proposal abstracts and narratives. What he can do with thousands and thousands of pages of text is unbelievable. So I would want people like that, because data is more than numbers. It's also the words we use.

One of my problems with data is that we're counting things that have already happened not things that are going to happen, because sometimes that's the invisible piece. And I think machine learning can help us figure out what's going to happen when we have the right inputs. He's using natural language processing. Identifying keywords is one thing. You could get a couple of students to find out how often a certain term shows up in a ton of text, but the natural language processing actually helps you figure out words that are related to that term but wouldn't come up in a search. It is the difference between looking for the word 'house', and the word 'home' comes up…you or I may only find house because we weren't looking for home. But natural language processing would consider them the same or similar or related. Does that

make sense? It is really interesting and I can't pretend to understand how he does it. He's part of the research office. It came out of a collaboration with our computer science department on campus. They had an idea and we decided to fund it. And we've basically decided to keep the position because he's so talented. And I'd also want a statistician. So, that's what I'd want. "

"It's that mix of having database people who really have the skills to manage the tables, manage the views, manage all the interactions with other systems. We need analysts and report writers who are savvy. And we need people who are doing the education and outreach, the communication around that stuff because you need all of those levels to do it well. And what we have said is, the more ambitious we want to be with development, if we want to build more web tools, we need more of all of those people: more analysts, more programmers, and more educators… those same elements have to be there as you expand. Your capacity goes directly in line with how many of those people you have.

I think we're very wary of anybody who tries to solve a problem with a new app. The software's rarely the bottleneck. It's the people and understanding that are the Bible. And it's a balance between having people who are 100% focused on one particular job function, versus having people who are a mixture of all of them so that they understand the entire life cycle and can use that knowledge to inform their particular task. And I think it has been very helpful for us that all of us work the help desk, because then when we design reports, or when we design the application, we understand what the problems will be with the end user, where if we had a bigger staff and someone said, 'I am a programmer only', they would miss out on something. So it's finding enough people in the department and keeping that balance."

"I'm interested in learning more forecasting, forward-looking tools and analytics. For instance, I know I can run a report and see what we've gotten, but then to be able to take that and project that into the future in terms of the workload that data will represent in an easy simple way. That would be great. I think it would really benefit this whole campus, if there were a way to have more licenses (for Tableau). I'm one of only like eight people on the campus that have a license to use Tableau and the data. I treasure that and I also pay for it because I'm asked to produce all these reports for many other people, but on the other hand, I wouldn't want to relinquish it. I know why they don't have more of them, they're really expensive."

Another wants "a tool to help our faculty to build collaborations, and to really catalyze on shared expertise, and shared interests. Faculty are very dispersed throughout the campus. Another is that we don't have a good handle on their publication records. Right now, once a year, the president asks for some information on their publications. The ability to see what all is going on in the realm of research beyond just the funding environment would be absolutely lovely. "

Getting back to reality, we asked the group about some best practices they'd offer. Their answers fell into two basic groups: the bigger picture of Data Governance, Strategy and Data Quality plus day-to-day suggestions about Data Analysis and Expertise.

Appendix II
Overview of the Six Scales and the Corresponding Item Descriptors in *Research Administration Data Analytics* Questionnaire

Scale 1: *Trouble (4 items)*

In your role within research administration, what is your level of agreement with the following: leveraging data analytics can be troublesome because...

Item no.	Descriptors	Strongly agree	Agree	Disagree	Strongly disagree	Unsure
1.1	no unified view of customers	○	○	○	○	○
1.2	inability to anticipate behavior	○	○	○	○	○
1.3	cannot identify valuable audiences	○	○	○	○	○
1.4	unable to gain insights that trigger action	○	○	○	○	○

Scale 2: *Usefulness (4 items)*

In your role within research administration, what is your level of agreement with the following: the use of data analytics and metrics...

Item no.	Descriptors	Strongly agree	Agree	Disagree	Strongly disagree	Not sure
2.1	helps me organize and visualize strategic data to improve our office or research community	○	○	○	○	○
2.2	lifts campus customer satisfaction	○	○	○	○	○
2.3	increases return on investment in research	○	○	○	○	○
2.4	allows me to respond faster to customer needs	○	○	○	○	○

Scale 3: *Challenge (13 items)*

How relevant are the following data-related challenges to your institution in the area of research administration:

Item no.	Descriptors	Very important	Important	Moderately important	Of little importance	Unimportance (not a challenge)
3.1	Timeliness	○	○	○	○	○
3.2	Overwhelming volume of available data	○	○	○	○	○
3.3	Data quality	○	○	○	○	○
3.4	Availability of data	○	○	○	○	○
3.5	Access rights to data	○	○	○	○	○
3.6	Data ownership issues	○	○	○	○	○
3.7	Cost of data stewardship	○	○	○	○	○
3.8	Lack of infrastructure	○	○	○	○	○
3.9	Lack of technology or tools	○	○	○	○	○
3.10	Shortage of talent/ skills	○	○	○	○	○
3.11	Privacy concerns and risk	○	○	○	○	○
3.12	Security	○	○	○	○	○
3.13	Corporate culture	○	○	○	○	○

Scale Four: *Skills (5 items)*
Indicate your agreement with the following:

Item no.	Descriptors	Strongly agree	Agree	Disagree	Strongly disagree
4.1	my staff/ coworkers are adequately supported in the effective use of data	○	○	○	○
4.2	my staff/ coworkers are adequately prepared to use data	○	○	○	○
4.3	There is someone who answers my questions about using data	○	○	○	○
4.4	There is someone who helps my Department change their	○	○	○	○

	practices, processes, policies based on data				
4.5	leadership provides in the professional development opportunities about data use	○	○	○	○

Scale 5: *Trust (5 items)*
Indicate your agreement with the following:

Item no.	Descriptors	Strongly agree	Agree	Disagree	Strongly disagree
5.1	members of my team trust each other's data work	○	○	○	○
5.2	it's OK to discuss feelings and worries with the data work of other members of my team	○	○	○	○
5.3	members of my team respect colleagues who need data improvement efforts	○	○	○	○
5.4	members of my team respect colleagues who are experts in their specific areas	○	○	○	○
5.5	our environment fosters trust when discussing data	○	○	○	○

Scale 6: *Data Use (10 items)*
How often do you individually or in teams do the following:

Item no.	Descriptors	Very often (at least daily)	Often (weekly)	Not very often (monthly or less)	Never
6.1	approach an issue by looking at data	○	○	○	○
6.2	discuss preconceived believes about a data issue	○	○	○	○
6.3	identify questions that we will seek to answer using data	○	○	○	○
6.4	explore data by looking for patterns and trends	○	○	○	○
6.5	draw conclusions based on data	○	○	○	○
6.6	identify additional data to offer a clearer picture of an issue	○	○	○	○

6.7	use data to make links between outcomes	○	○	○	○
6.8	when we consider changes in practice, we predict possible outcomes	○	○	○	○
6.9	revisit predictions made in previous meetings	○	○	○	○
6.10	identify actionable solutions based on our conclusions	○	○	○	○