Executive summary

University Decision-Making and Data Analytics

Joseph E. Steinmetz, Executive Vice President and Provost,
Professor of Psychology and Neuroscience, The Ohio State University

- The development of standard methods for data collecting and the formal analysis tools to mine the data and make sense out of the information have historically lagged behind our ability to collect and store data. However, administrators now have a variety of powerful tools available to collect, mine, and analyze very large data sets in relatively quick, standard, and reliable ways.
- I chaired a psychology department for nine years; measuring and evaluating the performance of the faculty in the department was by far the toughest part of the job. While many metrics were available, it was always difficult to assign relative weights for each category. Further, how we evaluate scholarship in the arts and humanities is very different from how we evaluate scholarship in the natural and mathematical or the social and behavioral sciences.
- The Ohio State University has begun using Academic Analytics in two areas related to the research productivity of our faculty. First, we have been able to compare the overall productivity of individual scholars with others inside and outside of The Ohio State University to identify areas of strengths and weaknesses in our faculty. Second, we have been able to use the analytics during program reviews to compare the overall productivity of departments and programs with identified peers and aspirational benchmarks, with an eye toward finding areas of strength and weakness relative to these benchmarks.
- No data analytic system is perfect, but in spite of some of the criticism and concerns about the Academic Analytic data set, I believe it is among the best that are now available for our use. I am very comfortable using this approach with the caveat I keep reminding all of my colleagues here—these are only a few of many data points that are available to us for conducting comprehensive evaluations of our faculty.
- It is clear that our increased ability to collect, process and analyze large data sets has enabled us to be much more data driven in making administrative decisions. Academic Analytics has proven to be useful for comparative reviews of the research productivity of individual faculty as well as departments and programs. We have also begun using data analytic techniques to identify collaborators inside the institution as well as those at other universities and in the private sector.
- We have also used data analytic methods to examine how we teach and how our students learn. We have mined large data sets to find out how prepared our students are and where they may need some additional help. We are using data to design classes that integrate traditional teaching methods with available technology. And we are taking advantage of the rich data sets available through MOOCs. MOOCs can be an effective way to reach large numbers of students and provide high quality learning experiences, while generating huge amounts of data that can be used to personalize learning and improve instruction.

- We constantly have to remind ourselves to look at <u>all</u> available data whenever possible. Decisions that impact faculty scholarship and teaching should be informed by more than one data point. Academic administrators need to be mindful of this approach.

Squaring the Circle: Using Analytics to Pursue Institutional Goals

Regina Werum, Associate Vice Chancellor for Research, University of Nebraska, **and** Michael Zeleny, Assistant Vice Chancellor for Research, University of Nebraska

- In recent years, several organizations and software solutions have emerged, designed to provide business and intelligence data solutions for research universities. What do research administration offices need to know in order to pursue institutional goals successfully? What can analytics software actually and potentially tell us? How can we address challenges that remain outside of the scope of these software solutions?

- University offices or research administrators have three needs. First, they need to be able to identify and often quantify *institutionally specific metrics of success*. Second, administrators need to be able to identify intellectual and organizational strengths and weaknesses. Third, research administrators need to be able to track funding trends throughout the institution over time, by unit, and by funding source.

- Academic analytics, in this context, refers to the analysis of research-related data to help educational institutions monitor progress on key institutional goals. Various software packages are available and offer products ranging from business intelligence at levels ranging from the individual faculty member to department/college/university-wide. Each of these providers claims to provide users with a clear and comparative understanding of research performance and/or productivity. Still, it seems clear that a comprehensive, one-stop research productivity software solution does not yet exist.

- It is not clear how data analytics will take into account the dynamics that are currently changing networks and research collaboration patterns across institutions and with non-academic partners. In their current form, analytics are not well suited to help university leadership address the impact of increasing lateral and vertical stratification within the higher educational sector. Software solutions have been designed to help institutions look inward, rather than foster the types of collaborations across institutions likely to mitigate the ever more fierce competition over resources and its effect on the feasibility of long-term institutional goals.

- In its current form, analytics software is not yet designed to help higher education leadership engage in the sort of simulation exercises necessary to determine the intended and unintended consequences of prioritizing specific metrics of success, typically gauged in terms of faculty productivity. Ideally, analytics software of the future could enable the types of simulation exercises needed to help predict the intended and unintended consequences of reaching specific institutional goals for a five, ten or even fifty year trajectory. If so, they should take into account the possibility of fundamental shifts regarding federal, industry, and other research funding opportunities, as those constrain the ability if institutions (and offices of research) to engage in strategic planning.

Transparency in the Age of Scholarly Analytics

Mardy T. Eimers, Vice Provost for Institutional Research & Quality Improvement, University of Missouri

- There is a call for public higher education to be more transparent with the information they share externally, as well as internally. Constituencies want to know how their tax-generated state appropriations are spent and whether they are getting their "money's worth." At the same time, transparency is important *within* the academy. Faculty and staff want access to the same information used by the key decision-makers, and they desire to understand the rationale behind the key decisions that will affect them directly.
- Much of the ability to be truly transparent depends on the audience's ability to understand and interpret the data provided. Organizational leaders can also be transparent by sharing *actions, processes, and/or decisions.* Sharing the rationale behind the decision or actions can be equally if not more important.
- Tremendous progress has been made in assembling quality scholarly data and building web interfaces to capture and use this data in planning and decision-making. Because detailed scholarly productivity information is now available, there are some critical questions: what data do you share? With whom do you share the data? When do you share the data? In what format, and with how much flexibility?
- The questions of "what to share" and "whom to share it with" looms large. Being fully transparent, or knowing the level of transparency that might be most appropriate, is not that simple. It requires sound judgment within the context of your internal and external environment.
- If an institution is choosing to be more transparent, it is not likely to be as simple as "switching a light on." If institutions could develop a set of principles to guide their actions, it would help considerably. By all indications, practicing "measured or tempered transparency" has a tremendous number of benefits to the institution and its constituencies.
- By measured or tempered, I mean that we *intentionally* and *consciously* consider the implications of what may be shared, and then adjust what is delivered accordingly. We need to find better ways to decide how to share data and information for the common good of the institution. I believe that if we can outline universal principles that can serve as a foundation on our campus, tailor them accordingly given the context, it will go a long way to serving our needs and building trust through tempered, transparent actions and exchanges.

"Let's Play Moneyball!": Analytics, Accountability, and the Future of Research Universities

Steven Warren, Vice Chancellor for Research and Graduate Studies, University of Kansas

- Research universities make massive investments in research. Many of these investments are obvious and easily accounted for. Arguably the largest relatively undocumented university investment is the "release time" from teaching provided to most tenure line faculty members. This investment is an excellent one in the majority of cases in which faculty use this “research time” to actively engage in important and measurable scholarship. But what about faculty members who are “inactive scholars”?

- There are at least two reasons that research universities should be concerned about tenure line faculty members who are inactive scholars. First, there may be an ethical issue if these individuals maintain graduate faculty status that allows them to chair or serve on PhD level doctoral student committees. We want our PhD students to be supervised by committees consisting of active scholars. Second, if you receive release time, you are expected to use it as intended unless given explicit permission to do otherwise. If not, this behavior (or lack of it) is in violation of the implicit and explicit employment agreement that exists between a full time tenure line faculty member and their employer.
- In the past, the problems of “inactive” scholars at research universities was most evident to their colleagues. However, in the world of electronic publication we live in, the evidence of this problem is more transparent. Inactive scholarship can now be identified by outside groups that harvest information on the productivity and faculty among other things, and then sell these analyses back to universities. These aggregators can also sell the same data to other groups such as state legislators and boards of regents.
- What is the best course of action to mitigate risk for the university? Get the data on your university and your competitors and develop an in depth knowledge of it. Start using the data to make decisions about hiring, retention, reorganization, etc. Work closely with deans, chairs, and faculty to create a broad understanding of the serious downside of ignoring this type of data. Put in place policies aimed at eliminating problems like unproductive tenured scholars. Use analytics data to make budgeting decisions.
- Caveat: Having a huge amount of data is a separate issue from using data wisely. Having a high publication rate and having high impact and value can be remarkably unrelated states. Nevertheless, the right data, wisely used and qualified can help us identify scholars who are no longer active. It is necessary that we evaluate scholarly productivity within the fields/disciplines it resides in, and against the standards of that field. Publication patterns differ greatly across various disciplines. Finally, the visual and performing arts present significant challenges in terms of evaluating the impact of creative activities in a valid way. We need to take great care and tread lightly in these areas.
- Analytics and big data are already having a significant impact on higher education in all sorts of ways. We need to embrace analytics and big data or we will run over by others that do embrace them. But this is not just about playing defense in an age of rapid change. These new tools present great opportunities for improving the performance and impact of higher education in general and research in specific.

Deans, Decisions, Data

Danny Anderson, Dean, College of Arts and Science, University of Kansas

- Here are some suggestions to guide in the use of data for decision-making in the context of a distributed authority model, which is characteristic of a large public research university. While these practices and lessons learned have emerged from work with Academic Analytics, the recommendations can guide in the collective use of a variety of datasets for the purpose of shared decision-making.

- **First, engage department chairs**. By drawing upon the strengths and insights of the department chairs, decisions can be more effective, generate buy-in at all levels, and avoid some pitfalls. **Second, contextualize the datasets with a variety of institutional research information.** Sometimes the unusual detail in one dataset or the anomaly in another is linked to historical changes, policy changes, or personnel practices and the juxtaposition of multiple, related datasets can help draw out these connections.

- Third, make conversations with department chairs and faculty central in the task of understanding complex data and building a shared vision for the future. Chair engagement and contextual information both emerge through collaborative examination of the data. This strategy is essential for owning the process of change. Having conversations with department chairs reminds us that change is "human-driven."

- Fourth, take a deep breath and be prepared to state repeatedly: data informs the decisions we make; data will not make the decisions for us. Fifth, as we emphasize engagement, context, conversation, and human traits, we can begin to see that data are narratives waiting to be told. If we have to go deep into the numbers when telling the story, besides the human faces we portray, we also need to make use of data visualization strategies that promote deep understanding as our audiences rapidly interpret complex statistical information.

- University leaders need to develop a coherent strategy for the effective use of data within their institutional contexts. We must be clear about our responsibility to use tools wisely to <u>inform</u> our decision making. We cannot and should not abdicate our judgment, authority, or responsibility to datasets. We must develop strategies for working on multiple organizational levels. Data and analytics as well as engagement, context, conversation, judgment, and narratives can all be brought together to help us map our way forward and release the energies we need to construct our future.

A Map for Understanding Decision Making

Michael J. O'Brien, Professor of Anthropology and Dean, College of Arts and Sciences, University of Missouri

- How does one make good decisions when faced with an information overload? In view of the different processes and scales involved in decision making, especially decisions about the quality of a behavior or product, how do we determine which one predominates in a given situation? At one extreme, an individual makes an informed decision based on careful analysis, and at the other extreme, people effectively copy one another without thinking about it.

- Big data can be used to "map" decisions along two dimensions: social influence and information. My colleagues and I have developed a simple heuristic map which captures the essential elements of human decision making that should be of concern to businesses, marketers, and even university administrators. The north–south axis of the map represents how well people are informed about their decisions. The east–west axis represents the degree to which agents make their decisions individually or socially. At the far west is one hundred percent individual learning, where agents rely only on their own knowledge of the costs and benefits of a particular behavior. At the far eastern edge is pure social learning, where people do only as others do.

- Why might this matter? Because most policymaking assumes that people all reside in the northwest—people make their own decisions asocially, with their own goals and preferences. The map relates specifically to patterns we can resolve from behavioral data, whether those data come from sales records or citations to scholarly articles and books.

- We evolved in a world of few but important choices, but we live in a world of many, largely interchangeable ones. Just as we feel adapted to the new order of the world, new fashions and technologies wash over us, new buzzwords enter our conversation.
- These elements—flux, learning, selection, and random events—bring about a new age of models of human behavior. Probability distributions, population size, invention rate, interaction networks, and time span become the key parameters. Marketing becomes less about satisfying "the" archetypal consumer and more about how many interconnected consumers affect each other's behavior. Old ideas, such as the sanctity of the "brand," have to be recast in terms of this bigger, more anthropological map. To do all this, it pays to have data analysts schooled in evolutionary theory.

"Big Data" Projects in High Energy Physics and Cosmology at Kansas State University

Glenn Horton-Smith, Associate Professor, Department of Physics, Kansas State University

- Present-day experimental high energy physics has been characterized as having three frontiers: an Energy Frontier, explored by experiments requiring the highest energies achievable; an Intensity Frontier, explored by experiments requiring the highest intensities achievable; and a Cosmic Frontier, explored using naturally-occurring cosmic particles and observations of the cosmos. Research at these frontiers naturally requires the analysis of vast amounts of data.
- The High Energy Physics (HEP) group at K-State engages in research on all three frontiers. On the Energy Frontier, the primary effort is the CMS experiment at the Large Hadron Collider (LHC). On the Intensity Frontier, we work on multiple neutrino experiments. On the Cosmic Frontier, the emphasis is on developing and testing models of dark energy with the goal of understanding the nature of the phenomenon driving the observed acceleration of the expansion of the universe.
- In HEP, we tend to use open-source software as much as possible. The ability to inspect source code, and correct and contribute to it if necessary, is important. Two examples of commonly used software are Geant4 and ROOT.
- Intermixed with this development process is a process of presentation of ideas and intermediate results to individual colleagues and groups of various sizes within the experimental collaboration, invariably leading to suggestions and corrections based on the colleagues' knowledge of relevant aspects of the experiment. The design of ROOT allows the researcher to quickly modify and repeat analyses as needed.
- A particularly useful data-driven method for measuring efficiency is the "tag-and-probe" method. It is especially useful when the new particles or interactions are detected solely through the observation of known particles whose properties are well understood. The known particles are also produced in simpler, well-understood reactions. The tag-and-probe method "tags" known interactions in which a particle of a particular type must be produced, then uses the particle known to be produced in that interaction as a "probe" to determine efficiency and an estimated uncertainty for the efficiency estimate.

- In academic analytics, data from “peer” and “aspirational peer” institutions and programs can be used to enable a kind of closed-boxed analysis in which metrics are developed in a data-driven way without using any data from the analyst's own institution. Insisting on such an approach to academic analysis could be a way for top research administrators to address concerns about releasing detailed program data to individual program heads or researchers for their own analyses.

Evolution of Research Reporting – From Excel to QlikView

Matthew Schuette, Principal Research Analyst, Enterprise Analytics, University of Kansas Medical Center

- In the last ten years, the University of Kansas Medical Center (KUMC) has experienced strong growth in its academic, clinical and research enterprises. The importance of accessible data, high-quality reporting, and analytics for both research and financials escalated during this time, shaped by enterprise growth and leadership focus. The lead partner in business intelligence (BI) and institutional research (IR) at KUMC is the Office of Enterprise Analytics (EA). Starting in 2009 the institution began looking heavily into comprehensive financial tracking and an appropriate BI tool for this venture.

- The primary source of research and financial data is PeopleSoft (PS) Enterprise Financial, Grants, and Human Capital Management systems. One of the vital roles of Enterprise Analytics is to mine, massage, and join tables from PS, and to use internal business practice rules to create consolidated tables. Prior to the implementation of QlikView (QV) on campus, most research data tables and reports were created on-demand using SAS data steps, procedures, and SQL queries. The use of SAS as a data mining and consolidation tool remains high, specifically for ad-hoc reporting and areas where development in a BI tool would not be cost- or time-effective.

- Up until the BI-era at KUMC, nearly all research reports were delivered with Excel. SAS provides easy exporting and importing of Excel files, and most staff on campus has familiarity with its features. The use of Excel for ad-hoc reporting will continue for the foreseeable future. BI tools require moderate training in the use of developed applications, as well as security access being granted. QlikView is a business intelligence tool which is highly flexible, has a rich, visual user interface, and allows users to clearly see associations between data.

- Enterprise Analytics assists KUMC’s Research Institute with annual reporting and there are many ad-hoc requests that we receive each year. Monthly reports provide administrators with an overall look at grant and clinical trial activity at the end of each month, while showing year-to-year trends. The raw data were produced with SAS, exported to Excel templates, and then further formatted. Requests for reports on investigator percent effort as well as formatted NIH Other Support documentation are currently in the form of Excel tables and Word files, and EA receives 300-400 of these requests per year. QlikView development of these same reports is in finishing stages, and the convenience will be provided to the RI and other department administrators to get the information whenever they need it.

- Historically, EA provided rankings of NIH awards to medical schools, based on total dollars awarded during the federal fiscal year, to KUMC research or department administrators, and also produced summary reports for our website. In QlikView, the NIH Rankings report is available to all users, and provides both yearly detail and trending information. The advantage with QV is that the user can select any institution/school/department, one or multiple years, and to view public or overall rankings.

- Departmental "Scorecards" are delivered to the Vice Chancellor of Research and provide a complete fiscal year listing of projects by department as well as information on paid effort vs. committed effort for individual faculty in the School of Medicine. All reports are Excel formatted. There is no intention to integrate these reports into the QlikView environment._The advent of BI tools, quicker and relatively cheaper computing memory and power, and enhanced institutional focus, has led KUMC into a newer world of data mining, intelligent and self-service reporting, along with data and analytically-driven decision making.

A Rational Approach to Funding Your Research Enterprise

Douglas A Girod, Executive Vice Chancellor, University of Kansas Medical Center; Executive Dean, KU School of Medicine

Paul Terranova, Vice Chancellor for Research, University of Kansas Medical Center

Clay Tellers, Principal, Academic Healthcare Division, ECG Management

- This paper outlines the efforts of the University of Kansas School of Medicine to develop a rational and reproducible funding model for the allocation of Institutional resources for the defined purpose of supporting the Research Mission. This effort was undertaken as an element of a more comprehensive funding model project that also including funding allocations for the Education and Service mission areas.
- A transparent and collaborative process was utilized to engage institutional and departmental leaders in the development of the model. Through the course of the process this input was critical in identifying elements of the model or unique situations in the institution that needed to be incorporated or modified to be truly representative of the research efforts. This process has also facilitated the "buy in" of the leaders in the model.
- A first pass high level simulation of the model would suggest a level of funding at about 47% of the amount of salary currently placed on grants for research faculty effort. In other words, this does seem to model roughly 50% of the faculty research effort as envisioned by the model. Thus it would appear to achieve the targeted goal.
- The funding allocation model is developed at the Departmental level. Since the model is based on Associate Professor AAMC salary benchmarks, the actual distribution of the faculty in a given department may differ.
- The successful implementation of the model will require a complete understanding of the key elements by Chairs and faculty alike. A result of developing the model at the Departmental level allows for the Chair to manage the Department budget to account for the idiosyncrasies of a given Department yet sets clear accountability to the Institution for meeting all the required missions with the given funding allocations.

- Once the model is run at the Departmental level there will likely be variations between the funding allocation dictated by the model and the current funding allocations which are largely historical in nature. It is anticipated that if variations of more than 10% occur a staged adjustment over a few years will be necessary to avoid major programmatic disruptions. These adjustments will need to occur in the course of the normal institutional budget cycle.

Understanding, Evaluating, and Reporting Research Productivity and Impact

Julienne M. Krennrich, Assistant Director of Research Initiatives, Engineering Research Institute, Iowa State University

Arun K. Somani, Associate Dean for Research, College of Engineering, Iowa State University

Martin H. Spalding, Associate Dean for Research, College of Liberal Arts and Sciences, Iowa State University

- The peer review system has historically played a large role in measuring impact. As scientific research has matured, the growth and fine-tuning of sub fields has made it increasingly difficult to compare impacts across disciplines and departments.

- The traditional measures of impact are: publications, citations, student and postdoc involvement, funding profile and technology transfer. The h-Index[2], a measure combining publications with citations, was developed as a way of measuring individuals' career achievements, but depending on the completeness of the publication-tracking system, faculty-to-faculty comparisons *within the same discipline* are difficult to compare. Another question is whether a citation implies a positive or negative impact. There is a bias toward reporting only positive impacts and with an additional pressure that more is always better.

- Over the past 30 years, the research enterprise in the United States has seen amazing growth in the competition for research dollars (state, private and federal). In many areas (particularly mature ones such as physics and chemistry), growth in the scientifically-trained workforce has continued, but the trend in available research dollars is decidedly negative.

- It is possible to gauge the impact of a single grant by tracking publications enabled by the funding, intellectual property enabled by the funding, student/postdoctoral training enabled by the funding, impacts on the discipline and outside the discipline. Taken together, these metrics can provide a qualitative measure of the grant, but it may be years before an accurate measure can be made. There is an inherent time lag in achieving outputs after dollars are allocated.

- We propose a topic-based evaluation model, grouping publications by researcher-defined topics and computing an equivalent h-index for an entire topic. This would alleviate the time lag in the system by collecting publications on a topic and not just as a result of a single grant. Using appropriate weight factors we would include citations, intellectual property and follow on, such as news articles. This would enable multiple papers with low-medium citations to be weighted more, thereby more accurately measuring a researcher's contribution to a topic over a lifetime.

- To accomplish this, researchers would register with a publication-tracking service, e.g., Google Scholar.
- Topic keywords or topic numbers would need to accompany publications in the profile so that the publications can be grouped according to topic. We envision that faculty would report such data to their Department Chairs annually. Tagging grants with data such as number of graduate students being supported and number of degrees conferred will be more time intensive. Over time, the data produced would be very valuable, so it is worth investing in the effort up front. On the whole, we are moving forward; we are beginning to understand how technology and metrics can help us perform better evaluations, but we are still in the experimentation stage.

Data Mining and Neurocomputational Modeling in the Neurosciences

Kimberly Kirkpatrick, Professor, Department of Psychological Sciences, Kansas State University

- The era of "big data" and the increasing focus on analytics is impacting most scientific disciplines, including research in cognitive and behavioral neuroscience. The growth of complexity of experimental data sets has led to the need for increased emphasis on data reduction and data mining techniques. An important companion to data mining is neurocomputational modeling, which is increasing in importance in the neurosciences.
- Such techniques such as data mining and modeling require the use of technical computing applications such as MATLAB, which can create barriers for incorporating students into the research process. The present paper discusses the challenges faced in the big data era of neuroscience and provides some ideas for tools than can promote success by researchers, and their students, in facing such challenges.
- The overarching mission of modern behavioral and cognitive neuroscience research is to pinpoint the neurobiological mechanisms of that underlie complex cognitive processes and the resulting behaviors. Cognitive neuroscientists typically focus on studying human populations, whereas behavioral neuroscientists typically focus on animal models of human behavior. There have been a number of exciting breakthroughs in the neurosciences that have led to the expansion of the complexity and size of data sets that are now typically collected in experimental studies.
- The growth of the collection of increasingly large and more complex data sets in the neurosciences is leading to the need for the development of new tools to promote capabilities for data mining. Technical languages such as MATLAB can serve as an excellent source for developing customized scripts and functions, and these can be made accessible to students involved in research through the use of GUIs.
- The future of neuroscientific research would be greatly benefited by increased availability of archived data for mining and computational modeling, increased sharing of tools for analysis, and the development of standards for approaches to mining neuroscientific data. An important companion to data mining is computational modeling, which provides a means of understanding complex patterns in data.
- Computational modeling is increasingly informed by neurobiology and this is leading to increased developments in neurocomputational modeling, which explicitly incorporate

neurobiological evidence in the development of process models of behavior. Here, too, the use of technical computing languages coupled with GUIs can provide powerful tools for model development and implementation.

What Does It Mean?

Susan Kemper, Roberts Distinguished Professor, Psychology, and Senior Scientist, Gerontology, University of Kansas

- The great promise of analytics is that benchmarking – faculty members, departments, universities, - will lead to wise strategic decision-making. My question is "what does it mean" to see "every variable in each academic discipline …[and] national quartile, quintile, decile, and vigintile summaries.." (Academic Analytics, 2013)?"

- The real challenge is to move beyond descriptive analytics. Even comparative analytics don't really answer the right questions. The data and its visualization must be coupled with an explanatory theory. Knowing how individual faculty members, departments, or universities stack up on various metrics – those "quartile, quintile, decile, and vigintile" comparisons - doesn't really provide answers to how productivity can be enhanced or sustained. And I think we are distracted by the logistics of compiling all this data and generating the fancy graphics, apps, and visualizations.

- At the 2001 Merrill Retreat on "evaluating research productivity," I turned to some sage advice from 1897: Cajal (1999) recognized 6 impediments to faculty productivity – what he termed "diseases of the will: the dilettantes or contemplators; the erudite or bibliophiles; the instrument addicts; the megalomaniacs; the misfits; and the theory builders (p. 75)."

- Cajal cautions that independent judgment, intellectual curiosity, perseverance, and concentration are the keys to productivity. Beyond these prerequisites, Cajal emphasizes that research productivity results from a "passion for reputation, for approval and applause," and a "taste for originality, the gratification associated with the act of discovery itself". These are the real determinates of faculty productivity. Analytics, no matter how aesthetically plotted as "quartile, quintile, decile, and vigintile summaries" do not assess this "passion for reputation" and this "taste for originality." That's what it means – to be productive, to have an impact.

Research Analytics: Facilitating the use of metrics to improve the research profile of academic programs

Rodolfo H. Torres, Associate Vice Chancellor, Research and Graduate Studies, University of Kansas

- The increase in external requirements of accountability faced by academic institutions and the need to convey to diverse non-expert audiences the contributions that the research enterprise provides to society, make it important that we find simple ways to put in evidence what we do.

- Some data and tools are publically available and subject to scrutiny by the general public. It is important that we conduct a serious analysis within our academic institutions to provide a solid understanding of what we can measure and what we cannot, to properly communicate to different audiences some true measures of research productivity and how they demonstrate the achievements of our institutions of higher education.

- The data sources and tools available today for quantitative analysis are sophisticated and diverse. At KU, like at most research universities, we systematically track institutional data that relates to our programs scholarly productivity in different forms.

- Despite the relatively easy access to tools and information, there are commonly-encountered barriers that restrict a wider use of research analytics. The analysis of the data is sometimes complex and subject to misinterpretation. Equally important is the fact that the type of data analysis needed can be extremely time-consuming. To mitigate some of these barriers we are currently developing a *"consulting service"* model. Our goal is to help academic programs to analyze the data.

- Academic Analytics collects information on more than 30 different metrics of research productivity. Using 15 of the metrics, which are typically "per faculty" counts meant to account for different program sizes, a *Faculty Scholarly Productivity Index* (FSPI) is computed using z-scores for each metric and weights similar to those used by the last NRC study. While the FSPI provides a snapshot number that could be used for a quick comparison with peers, looking in more detail at the data on which the index is based is often a lot more revealing. Understanding how the different metrics affect the program profile and how they may relate to each other is of crucial importance.

- A common need of programs in the current economic environment is the search for new funding sources. The *program market share* tool of AA can be used to aid in this regard. The analysis is limited to funding from Federal Agencies, which can present a quite incomplete picture in some disciplines, but it is still of value and shows potential opportunities not tapped by a program. Such information could become very valuable for a program trying to increase their external funding.

- As imperfect as the current metrics and data may be, they still provide tremendous amount of information that we did not have before. The key is to focus on what we can tell from such metrics and data and what we cannot. A careful use of technology and the availability of data could prove to be a big aid in the important engagement of our academic institutions in the planning and assessing of our research mission.

Research Excellence in the Era of Analytics: Considerations for Information Technology

Gary K. Allen, CIO, University of Missouri-Columbia; VP-IT, University of Missouri System

- The intended outcome of applying analytics to student and enrollment management data is to identify students who are at risk and provide interventions to help retain the students. Applying business intelligence tools to the task of helping students succeed is a natural extension of data-driven guidance.

- Successfully applying analytics to research and other faculty activities is likewise predicated on clear and feasible outcomes. Application of analytics to the research enterprise might well be as productive if focused on how to support researchers' data analytics activities rather than trying to measure a given faculty member's research productivity.

- The quality of research activities is particularly difficult to measure. Clear, comprehensive sets of relevant measures and approaches to compare those measures are not universally agreed-upon and are currently unavailable. Several hundred research universities are clients of Academic Analytics, LLC. For a subset of scholarly disciplines, this group has defined variables and will generate and manipulate structured data related to the productivity and quality of research. The primary data comparisons use the following data: (1) the publication of scholarly work as books and journal articles, (2) citations to published journal articles (3) research funding by federal agencies, and (4) honorific awards bestowed upon faculty members.
- For the foreseeable future, institutions will face increasing pressure to assess and optimize their research enterprises in response to diminished research grant funding, reduced financial support from state and federal governments, and pressure from the general public and university boards to limit increases in tuition revenues.
- Analytics must be thoughtfully and carefully applied to higher education. To be accepted, research analytics must be conceived and used as a mechanism for improvement. As higher education struggles to balance openness and data security, identity management to control access privileges and protect intellectual property will be increasingly critical. Clearly intentional choices will be necessary to optimize an IT infrastructure that can be sufficiently flexible and nimble to meet demands not yet known or fully understood.
- To be worthwhile, research analytics must support planning and illuminate decisions. The data being analyzed must be relevant to the question at hand and needs to be studied within the context of the strategic decisions. Analytics cannot take the place of leadership. While IT can contribute to a successful data analytics program, the technology is not what is vital - rather it is the leadership and the ability to make difficult choices.

Student Training in the Era of Big Data Physics Research

Amit Chakrabarti, William and Joan Porter Professor and Head, Department of Physics, Kansas State University

- Availability of Big Data is having a major impact on research and student training in all sub disciplines of physics. High Energy Physics and Cosmology are at the forefront of Big Data Physics. How do we train undergraduate physics majors and graduate students in this era of Big Data physics research? All physics students must be encouraged to view physics as both a fundamental and foundational science that provides an effective background for a diversity of career paths. Many of the problems that will need to be solved in the coming decades will occur on the interface between physics and related areas.
- Of foremost importance is to train students in the physical models that have been so successful in explaining Nature. This is essential to provide the students with Big Data interpretation skill. Early involvement in research is a must. Research experience lets students put to use theories they learn in class and acquaint themselves with the faculty, post-docs and other students. These experiences help students make good career decisions, and involvement in research is fun.
- Another essential component of student training in this new era is the introduction of specialized computational skills early in their career. On one hand, this will teach them to apply tailor-made computational algorithms based on understanding the specific physics

of the problem at hand. On the other hand, introduction to Open Source and Visual programming skills will help them with their career decisions. Training in both oral and written technical communication skills and the ability to translate from Techie language to English will be critical for success in a wide variety of situations.

- Once new opportunities for physics faculty are identified, their research programs can be broadened by systematically engaging companies in the research work. This will bring industrial support to research and create a culture of solving practical problems. Such experience in "producing products" will have a profound impact on professors and students equally. K-State will be a powerful economic driver for growth and development by generating new knowledge and producing graduates who will impact Kansas, the nation and the world.

- Finally, a brief discussion of assessment of student achievements in the Big Data era is warranted. The K-State Physics Education Research Group is in the forefront of creating a large database of nationally representative data with support from the American Association of Physics Teachers and the National Science Foundation. Once the database is created, faculty will be able to visualize and compare their students' performance to huge national database of results from 50+ research-based assessment instruments.

- Curriculum development and student training must be undertaken in view of these recent developments. Topics on student training and Big Data Physics projects discussed here are in the context of the physics department at Kansas State University. Their implications, however, go beyond the borders of one physics department or one University.