

## University Decision Making and Data Analytics

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

I am a scientist-academician and therefore have had a rather lengthy career forming hypotheses, collecting data, analyzing those data, and making decisions concerning those hypotheses from a very data-driven perspective. Somewhere along the way I also became an academic administrator and have held administrative positions that have included stints as department chair, associate dean, dean, executive dean, and provost. One would think that given my background as a scientist that my administrative decision-making would be highly data driven. This has not always been the case, especially early in my days as an administrator when decisions were often made on a much more ad hoc basis, largely influenced by the case made by an individual or a group seeking the decision at the time.

Hiring faculty is a case in point—in my days as a department chair and dean, I often heard arguments for hiring based largely on the self-perceived “excellence” of the existing faculty in that department or program. I rarely heard arguments based on objective sources of data, but rather typically heard arguments based on a single ranking, past reputation, or perhaps past hiring history (e.g., we have always had X faculty in our area or in 1968 we once had as many as Y faculty). Absent data, it was hard for me to make a decision whether a given program should recruit and hire additional faculty.

It is not that we had a shortage of data 25 years ago; as computer capacity grew over the years so did the available data sets. In fact, for the last several years we have been able to collect massive amounts of data on measures of performance for our students and faculty and store those data rather cheaply. The data

we now have available can seem overwhelming at times and at other times even conflicting. For example, I seem to always get different numbers when I poll chairs and deans about the average teaching load of their faculty—the source is important. I know one thing for certain: The development of standard methods for data collecting and, perhaps more importantly the formal analysis tools to mine the data and make sense out of the mountains of information that are available to the academic world, have lagged behind our ability to collect and store data.

The data analytics movement has changed this picture dramatically. 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. And administrators are using these tools in growing numbers. Like many institu-

tions, The Ohio State University is turning increasingly to data analytics to aid in decision making and doing so in the four traditional areas that define the institution: research/innovation, teaching/learning, outreach/engagement, and support services (which include finances, recruitment and admissions, and other university offices). While data analytics have proven useful and important in all of these areas, I am going to focus on how we are using data to inform and support decision making with regard to research and innovation, as well as teaching and learning. I will present a few examples here, starting with measuring research and innovation productivity of our faculty.

#### **Academic Analytics: Measuring Research and Innovation Productivity of our Faculty**

I chaired a psychology department for nine years and during that time found that measuring and evaluating the performance of the faculty in the department to be by far the toughest part of the job, especially related to salary setting and retention efforts. It was difficult for several reasons. While many metrics were available (e.g., number of publications, number of citations, number and size of grants, etc.) it was always difficult to assign relative weights for each category. To complicate matters further, my department had a wide variety of disciplines within it, such as social psychology and behavioral neuroscience, and each discipline had different patterns of productivity. For example, the behavioral neuroscientists tended to publish several relatively short articles every year while the mathematical psychologists pub-

lished fewer but longer articles. The relative sizes of laboratories varied greatly, as did the number of coauthors. In addition, the various disciplines published in different sets of journals. When I became a dean, and later provost, the differences became even more pronounced—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.

Faculty evaluations can involve either internal or external comparisons. Department chairs are normally interested in evaluations of faculty *within* their department, so the comparators are faculty members inside that department and often within sub-areas inside the department. Other administrators are often interested in how the productivity of faculty stacks up against departments or area *outside* the university. For example: how does the productivity of the psychology department at Ohio State compare with the productivity of psychology departments at benchmark institutions that are selected? Until recently, making these internal and external comparisons has been relatively difficult. Collecting the data for your own discipline, department, or institution is the easy part. The much harder part has been finding reliable data with which to make comparisons, and then finding the analytic tools to easily and effectively evaluate faculty productivity with internal or external benchmarks. These tools are now available.

At The Ohio State University we have begun using Academic Analytics to help evaluate the research productivity of

our faculty. There are other excellent analytic data sets available for use, including Thomson Reuters' Web of Science and Elsevier's SciVal. I am only going to discuss the use of Academic Analytics since this is the analytics tool we have used most at Ohio State. Academic Analytics was co-founded by Lawrence Martin and Anthony Olejniczak. Martin had served as dean of the Graduate School at Stony Brook University and, like many administrators, he realized there was a need for comparative productivity metrics to be used for assessing performance of programs. Martin and Olejniczak released their first database in 2005, which has been refined in subsequent years with input from an advisory committee that was formed. Ohio State's Julie Carpenter-Hubin, Assistant Vice President for Institutional Research and Planning, has served as a member of this advisory committee.

The comparative database includes information from more than 270,000 faculty members, each given a unique numerical identifier. Those faculty members come from over 9,000 Ph.D. programs and 10,000 departments at more than 385 universities in the United States and abroad. The data set has been created by faculty lists supplied by the participating institutions, as well as data mined from several sources such as public databases, web sources, and government reports. These data fall in four areas: (1) publications of scholarly works (journals, conference proceedings and books), (2) citations to published journal articles, (3) research funding by federal agencies, and (4) honorific awards to faculty members. These data are used to define the Faculty Scholarly Productivity Index (FSPI) for each

faculty member. Institutions also supply data about faculty distribution in departments, so that the individual faculty can be aggregated appropriately for evaluative comparisons.

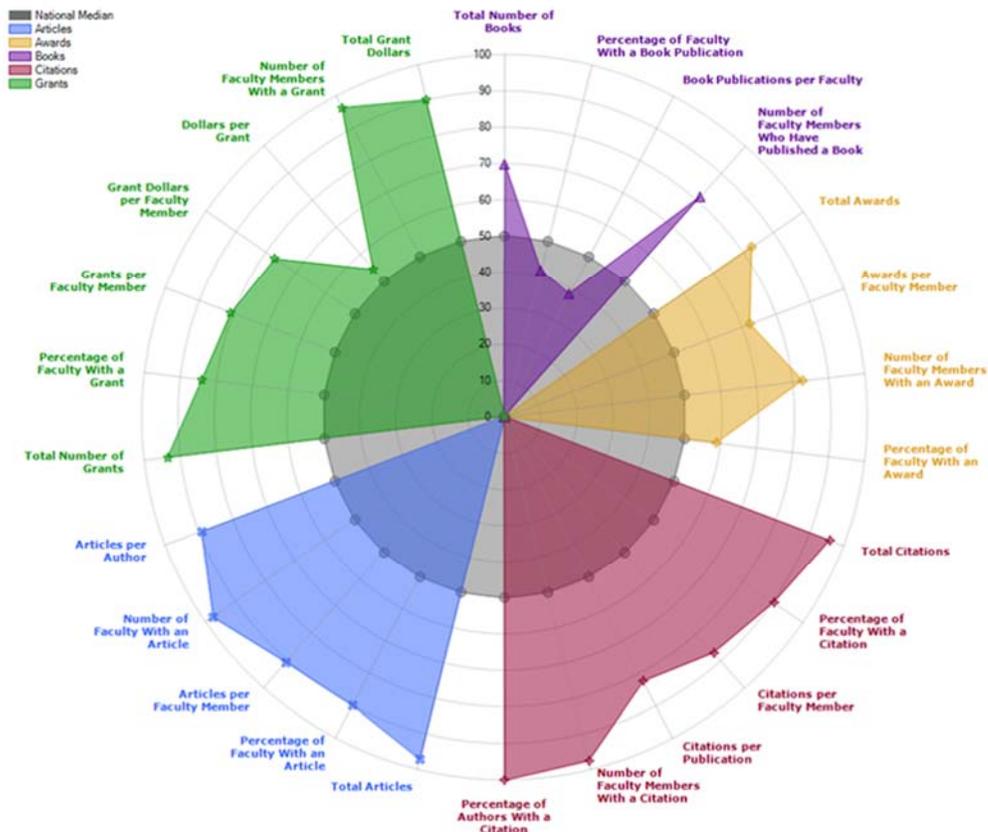
The database is accessible through an online portal that offers a variety of tables, charts, and data cutting tools. An interesting question that frequently comes up is: who within the institution should have direct access to the data set? Should individual faculty members? Chairs and/or deans? Only provosts and other administrators? I know deans who believe that if individual faculty or chairs were allowed access to the data set that they would "play with the weights" until only favorable comparisons emerged—several categories of data go into the analysis and the weights on the categories can be easily manipulated. For example, the analysis could look different if total publications are weighted heavy and total citations weighted light, or vice versa. At Ohio State, we don't give access to the data set to individual faculty members but do give access to department chairs; department chairs can use the data to self-assess their strengths and weaknesses because they are most familiar with their disciplines. At a central level we typically define the benchmarks for the institution; our peers and aspirational peers also use the data set with a more standard set of weights to do trans-institutional comparisons.

I have found the "flower chart" available from Academic Analytics to be an excellent way to get a snapshot look at overall faculty productivity in a unit. An example of one of these flower charts can be seen in Figure 1. This figure shows the

performance of one of our academic departments on 26 different metrics that are color coded in five different categories: articles, awards, book, citations, and grants. Within the grants category, for example, there are seven metrics: total number of grants, percentage of faculty with a grant, grants per faculty member, grant dollars per faculty member, dollars per grant, number of faculty members with a grant, and total grant dollars. The chart is easy to read. The diameter of the gray circle within the concentric rings shows where the median performance is for the benchmark institutions chosen for the analysis. The further out on the concentric rings the better—that is, relative performance in a given category is stronger. The example in Figure 1 is a

strong department at Ohio State whose faculty are well above the median performance in citation, article and grant indices and slightly weaker (though overall still strong) in book and award indices. Performances on only two of the 26 categories were below the median: percentage of faculty with a book publication and book publications per faculty. Note that the total number of books published and the number of faculty who have published a book were well above the median, though. These kinds of charts are good starting points for discussions concerning departments’ strengths and weaknesses.

It is also possible to use Academic Analytics to compare the performance of individual faculty members with other

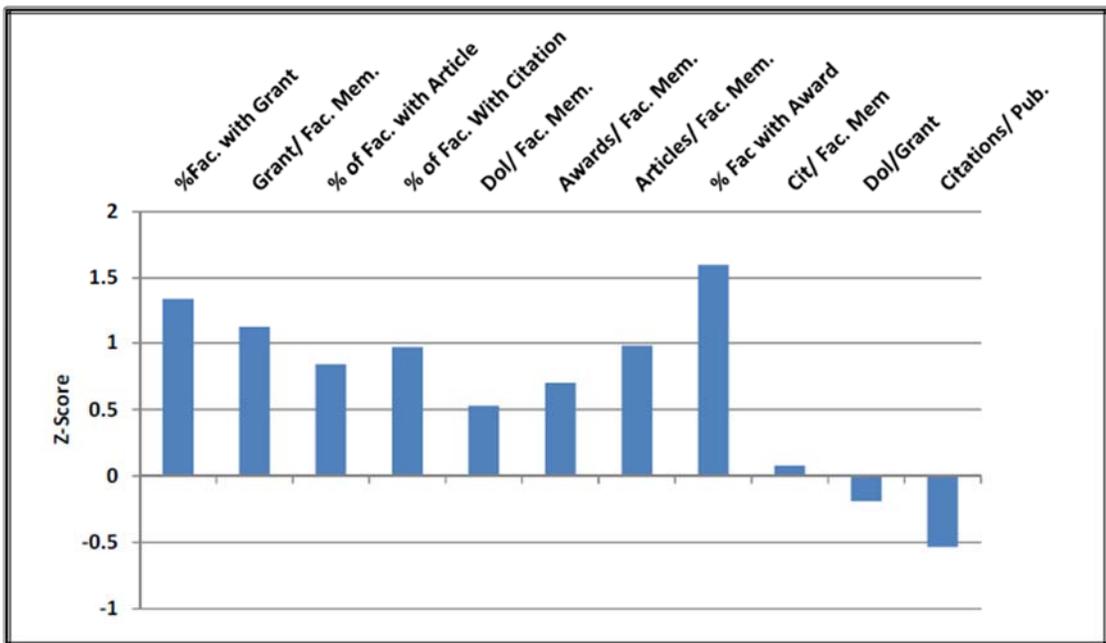


**Figure 1:** Example of a “flower chart” available from Academic Analytics depicting relative strengths

researchers in their field by generating Faculty Count Charts. The Faculty Count Charts show how the performance of a faculty member looks relative to others in her or his department. The reports that are generated can take on several forms, including table of the raw numbers, which are useful for seeing the raw comparative data; a modification of the flower chart showing relative performance on a subset of metrics; or shown as a national quantile.

At Ohio State we conduct periodic formal external reviews of all our academic units. As part of this process, we require our academic units to prepare program review self-studies that include summaries of their performances in research and scholarship. Going forward, with assistance from the Office of Academic Affairs, we are asking departments to use data from Academic Analytics in their program review self-studies. Figure 2 shows a chart that was developed by

the chair of one of our departments for their self-study. The analysis helped the chair understand the strengths and weaknesses of the department as compared to benchmark institutions. In this example, it seems clear that the department has many strengths and some weaknesses, including relatively low numbers of citations per faculty member, low numbers of dollars per grants obtained, and low numbers of citations per publication. The chair looked at the data over a four-year period and the numbers consistently held. It is possible to get an even more fine-grained analysis of these data. In this particular example the chair compared the grant data with five benchmark universities of similar size and scope. The analysis showed that the benchmark institutions had about 30% more funding than did the Ohio State department as measured by both total grant dollars and dollars per grant (data not shown). Not satisfied with these data, the chair went



**Figure 2:** A chart developed by an Ohio State department chair showing the performance of a department’s faculty on several metrics relative to some benchmark institutions.

further to look at what sources of federal support were lower than the five benchmark institutions—relatively low levels of support from three agencies were identified.

The example above is a wonderful illustration of the power of these analyses, especially when in the hands of a department chair interested in identifying the strengths and weaknesses in the department. Armed with the data, the chair was able to have meaningful conversations with the dean's and provost's offices regarding the external review. Chairs are in the best position to interpret the data. In the example above concerning the grant data, a possible explanation for the relatively smaller grants, as well as the low numbers from some of the federal agencies, could be the make-up of the department. For example, if you don't have faculty in your department who do energy-related research it is likely you will not have high levels of funding from the Department of Energy. Lower grant totals might be attributable to having more junior faculty members than the benchmarks (as might be the case for publication numbers and citations). The major point here is that the data set can trigger a meaningful discussion between the department and the college or university regarding the relative productivity of its faculty.

Along these lines, we have identified a weakness in the Academic Analytics data set that affects the interpretation of the analysis. When measuring grant activity, the data do not account for co-PI status on grants: Academic Analytics attributes grants to the first PI listed. So, for highly interdisciplinary work, one or

more faculty members and their programs may not receive credit for grants they are on. This is a very legitimate issue for a chair and department to raise, especially since at Ohio State we are emphasizing interdisciplinary approaches to research and teaching. Returning to the above example regarding grant funding, however, it was clear that a high level of interdisciplinary work could not explain the gap between the department's grant funding and the funding levels of the comparison group; there were relatively few grants with Ohio State co-PIs in the department. This serves as a good example of how these data can be probed further to reveal potential reasons for the strengths or weaknesses in the department.

In summary, we have 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 strength and weakness in our faculty. These analyses should help us make decisions about where we should invest resources to most effectively impact the development of 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. These analyses should help us make better decisions concerning where we should invest central funds to facilitate development of our departments and programs.

While to date we have limited our use of these data sets for the two purposes outlined above, I have had discussions with other deans and academic administrators about how the analytics could be used in other ways. For example, data regarding individual scholars could be used for internal salary equity analysis by identifying a local cohort either inside or outside the department for comparison purposes (e.g., other scholars who have been in rank for a similar number of years and are in a similar discipline or field). Likewise, the data could be used for analyses of faculty retention cases. In these instances, the comparison group might be other institutions of similar size, scope, quality and cost-of-living. Cases for hiring senior targets-of-opportunity would be strengthened by the use of these data. And decisions about distribution of resources across departments could be aided by these analyses.

No data analytic system is perfect. Over the last year or so, I have had several discussions with colleagues inside and outside of Ohio State about the use of data analytics for decision-making, including the strengths, weaknesses and criticisms of Academic Analytics. I present some of these concerns here:

- *Administrators use the wrong benchmarks for comparisons.* Often departments want to choose their own list of institutions against which the performance of their faculty can be compared. To some extent this might be valid, especially when trying to probe for strengths and weaknesses in specific areas. However, to gain the overall perspective it is important that we use a standard set of institutions—after all, our departments and program

exist in the context of the institution and choosing institutions that are like us is important for comparisons.

- *My department is different and the Academic Analytics indices don't capture our strengths.* While this could be true, the fact still exists that the research indices included in the analyses are generally agreed upon as standard measures of faculty performance.
- *The data are not correct.* When confronted with this possibility, I have asked the department to show me how the data are incorrect. We have found the Academic Analytic data set to be in good agreement with our own data sets when we have attempted to check independently. An issue that has received some attention is making sure that a scholar's work is attributed correctly, especially when that person has a relatively common name. Academic Analytics has effectively dealt with this issue.
- *This may be good for research, but faculty also teach and do service locally and nationally.* This is a very valid point and we must keep in mind that only research is being evaluated with this tool. It is but a *single* data point in an overall evaluation of faculty performance.
- *This analysis is largely quantitative and not qualitative in nature.* That is, the relative quality of journals and books is not factored in. To some extent this is true. Other sources need to be used in conjunction with the Academic Analytics data. For example, we often use journal impact factors as well to get an idea of the quality of outlets used by our scholars.

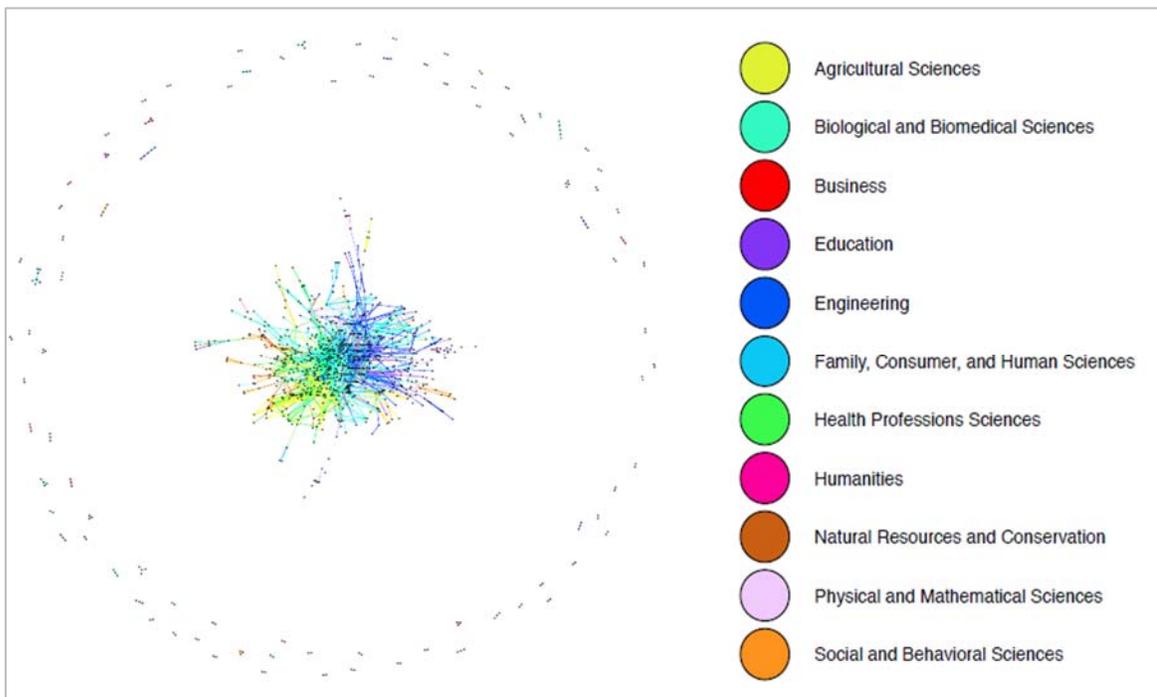
- *The system rewards faculty who write a lot and that doesn't always reflect effort or prominence.* This may be true to some extent as well. However, prominence might be captured in other metrics such as the awards section of the data.
  - *Sub-disciplines within departments can vary greatly and this is not factored into the analysis.* This point has been a concern of mine. For example, psychology departments can vary greatly in their composition (e.g., heavily oriented toward social psychology versus clinical psychology versus behavioral neuroscience—all disciplines that publish in different ways). This is where selecting the benchmarks becomes critical: Finding departments with similar composition for comparisons is very important.
  - *Citation indices are not used for books in the current version of Academic Analytics.* This is true at this time, so it is difficult to gauge the impact of books like one can do for journal articles.
  - *Non-conventional publications such as web-based publications are not included in the database.* This also is true at this time and could be a significant deficiency in the future as more and more of our disciplines turn to more non-traditional means of publishing.
  - *This approach will produce more "publishing to the index" similar to how some teachers are "teaching to the test."* The answer to this will be yes if administrators encourage this by the way they use the results obtained from the analyses. Again, I stress here that this analytic approach is just one of what should be several ways we evaluate performance of our faculty.
  - *The FSPI does not capture interdisciplinary work and collaboration.* This is currently true for the grant data but not the publication data, as detailed above.
  - *Finally, the data set only goes back the last few years and does not include the most recent data.* Hence, we are getting a fairly limited view of faculty performance. In particular, early publications are not captured for our more senior faculty and the most recent publications are not captured, which can limit the data available to evaluate more junior colleagues. With regards to the senior faculty issue, this also would be more problematic if we were investigating how productive the department has ever been, rather than how productive the department has been in the last few years, which is usually the central question. I would argue that if a department has a scholar who wrote the seminal article in the field 20 years ago and then stopped writing, the department has a brilliant scholar who adds value to the department, but is no longer producing scholarship. Academic Analytics measures productivity, and to some degree by including citations and awards also gets at brilliance/quality, but it is advertised as a scholarly productivity tool. Over the years, this picture will change, of course, as more years are added to the database.
- In short, in spite of some of the criticism and concerns about the Academic Analytics data set, I believe it is among the best that are available for our use at this time. I am very comfortable using

this approach with the caveat I keep reminding all of my colleagues here—these are but relatively few of many data points that are available to us for conducting comprehensive evaluations of our faculty.

### Using Data Analytics to Identify Potential Collaborators and Existing Research Networks

We also are using Scholarly Signatures, created by Academic Analytics, as a method to identify where we have concentrations of similar activity at The Ohio State University. The Scholarly Signatures are created by mining journal article abstracts for key words that are used to describe the research presented in the paper. For example, my research over the years has been in the neural bases of learning and memory. More specifically, I have studied how the cerebellum of mammals is involved in a simple form of

motor learning known as classical eyeblink conditioning. My Scholarly Signature would include words like “conditioning”, “eyeblink”, “cerebellum”, “reflex”, “rats”, “conditioned”, and “stimulus.” Using techniques like cluster analysis and factor analysis to create semantic groupings, we can use the Scholarly Signatures descriptors to connect with other scholars who use the same or similar descriptors, such as “learning” and “habituation,” that are conceptually related but not identical. These keyword abstractions also can be used to identify topics that cross disciplines even when there is little or no observable collaboration between researchers. We plan to use these techniques to assist us in locating major themes of scholarship at the university that might be explored further for investment and development.



**Figure 3:** A diagram of co-authorship patterns at Ohio State. The outer ring shows co-authorships involving two authors. The inner cluster shows more complex co-authorship patterns.

We also are experimenting with a beta version of Academic Analytics' "Expert Picker" to identify researchers from across the university who are working in related areas to individual scholars. This features uses analysis of text (e.g., from abstracts of papers) to identify other scholars in the institution who work in similar areas. We hope this tool will help our faculty find on-campus collaborators more easily, something that is important at an institution as big and complex as Ohio State. Specifically, we believe this tool will be valuable for linking scholars on campus as we build out our "Discovery Themes," which are three general research areas in which we are investing heavily: Energy & Environment, Health & Wellness, and Food Production & Security. Also, we are hoping to eventually extend the use of this tool to identify collaborators at other institutions, especially potential research partners in our state and region. Our first two tests of this will

likely be to find scholars from West Virginia and Pennsylvania who are interested in shale gas research, as well as scholars in the states around the Great Lakes who are interested in algae bloom research.

We have found that the Academic Analytics data also can help identify and better understand the informal research networks that already exist on our campuses. Figure 3 is a diagram of the co-authorships that exist among our Ohio State faculty gleaned from the Academic Analytics data set. The outer ring gives us information about faculty who coauthor with another Ohio State faculty member, but only with one other faculty member. The cluster in the middle shows co-authorships that are more complex. The different colors in the diagram represent different broad fields such as agricultural sciences, business, engineering, humanities, etc. Figure 4 provides a zoomed-in

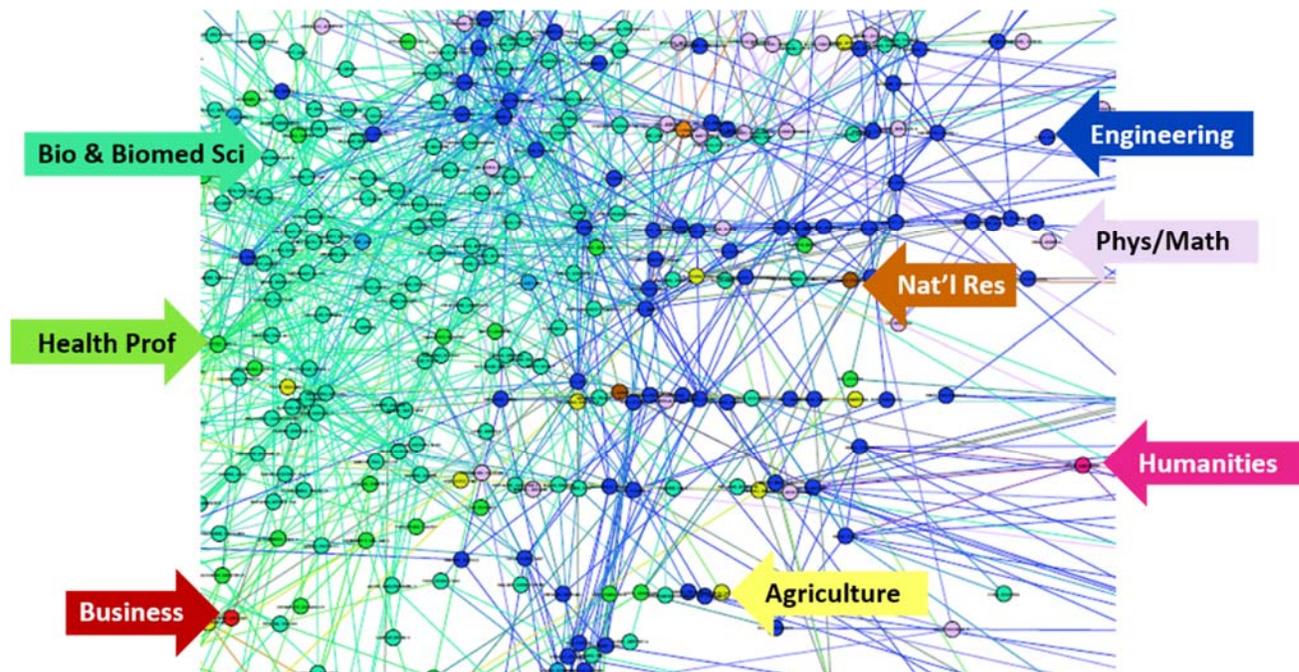
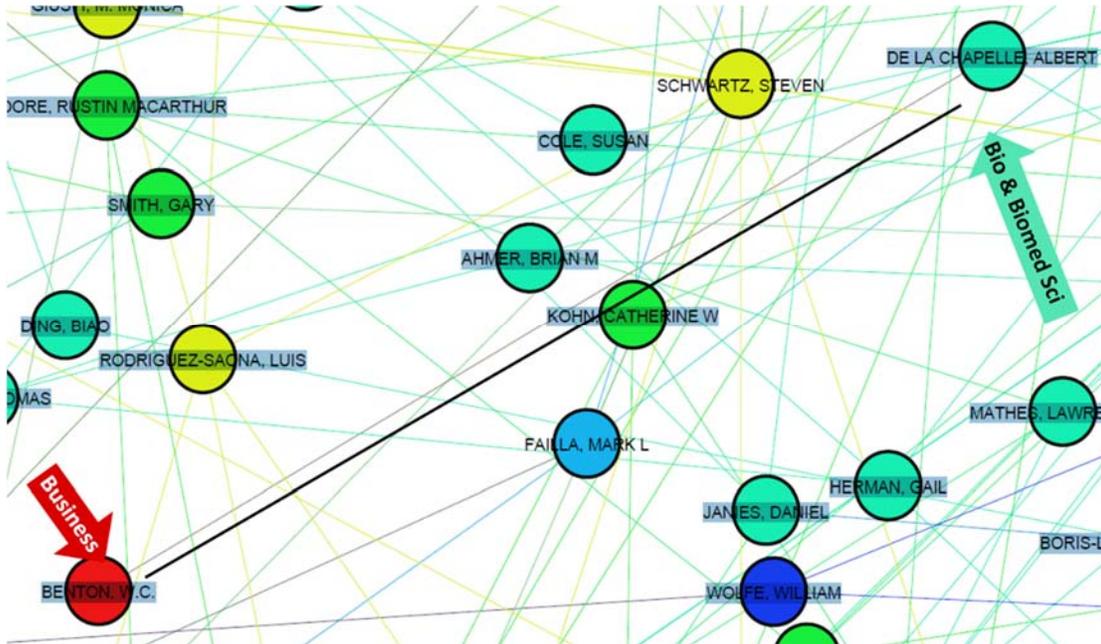


Figure 4: A zoomed-in view of the complex patterns of co-authorships shown in Figure 3.



**Figure 5:** An example of a “non-traditional” co-authorship pattern between a business faculty member and a cancer researcher.

view of this middle cluster. Not surprisingly, the figure shows that faculty in engineering, the health professions, and biological and biomedical sciences have lots of connections. What I find more interesting are the non-traditional connections, as seen in Figure 5. In this example, we can see a publishing collaboration between business and health and biological sciences faculty members—this collaboration involves W. C. Benton, a distinguished research professor from our Fisher College of Business who studies health care performance issues and the economics of cardiovascular surgery, and Albert de la Chappelle, who is a prominent cancer researcher. An example of one of their papers is one that looks at the feasibility of universal screening for a type of colorectal cancer. How might we use this tool? First, we think that highlighting these kinds of collaborations—that is, using them as demonstration projects in a sense—will serve to increase our

cross-disciplinary faculty partnerships. Second, these analyses may tell us where fruitful collaborative investments should be made in the future. And third, by looking at these nodes of activity we can identify who our most collaborative faculty are—faculty who might be good candidates for committees or advice for developing collaborations on campus.

#### **Data Analytics: Connecting the University with Industry Research Partners**

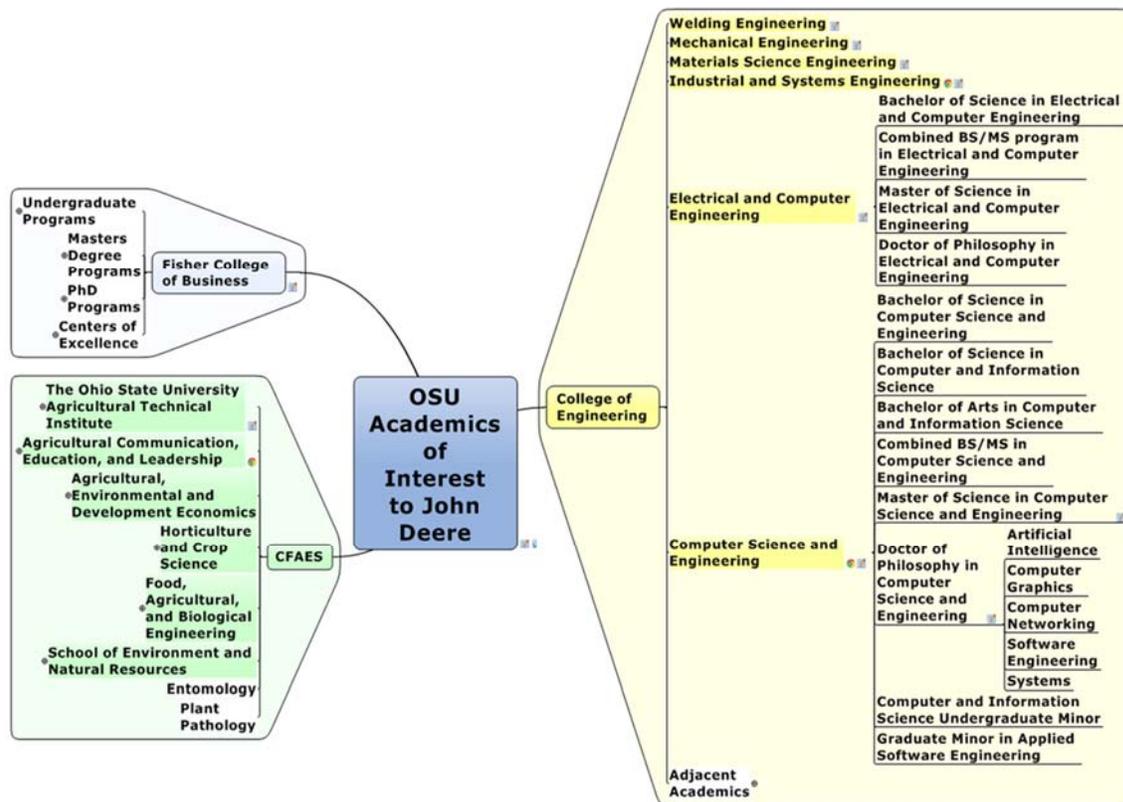
Up to this point I have restricted my presentation to how we use Academic Analytics to look at faculty performance, department reviews, and scholarly collaboration. This is not the only data analytic tool used at Ohio State, however. We are using available data to connect the university with potential industry partners. Alba McIntyre and Bryan Kinnamon from the Industry Liaison Office of the College of Food, Agricultural, and Environmental Sciences (FAES) have

used key words to map our academic programs, as well as faculty research capabilities, to industry needs. We believe this mapping is potentially useful in helping industries see the opportunities for collaboration at Ohio State and vice versa.

Figure 6 shows one of these mappings. In this example, Alba and Bryan reviewed one company's website, which included information about the type of students they wanted to attract for internships and positions in the company. Using this information and including information from FAES as well as our colleges of Business and Engineering, they created a map of all of the Ohio State degrees that would be of potential interest to the

company. Providing this kind of information lets the company know about the programs we have in the university that are specifically related to the core business of their company. We believe such mapping could be very useful for building internship opportunities and convincing companies that coming to Ohio State to recruit students would be productive and fruitful. This type of mapping strategy also has been used to match research interests between faculty members and industry. At this point, the mapping process is very labor-intensive—each map is unique, rather college-specific, designed to fit rather narrow objectives, and with data sets that come from multiple sources. But this approach has great potential, especially if somehow

**Figure 6:** An example of a mapping done by our College of Food, Agricultural, and Environmental Sciences linking Ohio State academic programs with personnel and research needs of a company.



combined with data sets like Academic Analytics and when a more institution-wide perspective is taken. The benefits to the institution are at least two-fold. First, in addition to providing these maps to industry we can use them to locate industry partners for our faculty, thus increasing our opportunity for commercialization and licensing activity. Second, these maps could help us better understand where industry hiring needs are, which could be valuable for curriculum and program development.

### **Using Data Analytics to Facilitate Teaching and Learning**

The use of data analytics methods also has impacted faculty teaching and student learning at Ohio State. Here I present four examples: (a) student advising, (b) analyzing pre-college experiences, (c) new approaches in the classroom, and (d) analytics related to Massive Open Online Course (MOOC) experiences.

#### Student Advising

We are always looking for ways to improve our undergraduate student advising experiences and to this end have created a data-intensive tool we call "AdvisingConnect." AdvisingConnect is a student-adviser relationship management tool that we use to encourage students to be proactive in their own advising. AdvisingConnect is essentially a very extensive database that creates a record of all of the students' advising contacts. Every advising contact is documented, such as the issue on which each consultation is based and the advice that was provided by the staff member. This large "institutional memory" summarizes each student's past. Access to this information database serves to improve

efficiency, increase the continuity and coherence of conversations the advisors have with the students, and perhaps most importantly provides a very personal sense of contact, whether he or she is seeing the same or a different advisor. The system also tracks missed versus kept appointments, the number of appointments, and the timing of the appointments. With about 66,000 students at Ohio State, collectively this is a massive database, which we are just beginning to mine together with information regarding registration and fee payment. For example, we believe this data set might be useful for exploring patterns that predict success in courses for individual students, as well as for groups of students formed on the basis of major, gender, or ethnicity. We believe we can use the database to track the performance of the advisors as well.

#### Analyzing pre-college experiences

We all know that student learning does not start when freshman first enroll at Ohio State. Students' pre-college experiences in high school and before are very important determinants of their success once they arrive on campus. Given this fact we are very interested in looking at data accumulated from many students over a number of years to determine those pre-enrollment factors that lead to success in our general education courses, such as English and mathematics. For example, we have examined the performance of students in our calculus and analytic geometry classes as a function of their previous mathematics preparation. To do this, we separated students in four groups: (1) students who took high school Advanced Placement (AP) courses, (2) freshman students who came

in with transfer credits, mainly from dual-enrollment high school courses, (3) students with other sources of pre-college mathematics credits, and (4) students who took regular high school mathematics courses. Our analysis showed that, not surprisingly, those students who came in with AP credits had the highest grades in our calculus course—AP mathematics courses, after all, are targeted for those who are good in mathematics. What was a little surprising was the relatively poor performance of students who came in with transfer credits; that is, students who took college mathematics courses while in high school. We are probing the data further to explore at least two explanations for this finding. First, it is possible that the dual-enrollment courses taught in high schools are not rigorous enough or not taught well—more data must be accumulated to look at this possibility. Second, it is possible that some students complete their dual enrollment math courses earlier than their senior year in high school thus leaving a gap between their high school dual enrollment course and their first college course. Our data will eventually shed some light on this.

Almost all students who enter college these days do so with credits earned through Advanced Placement courses taught in our high schools. Recently, the State of Ohio mandated that a score of 3 on this test instead of a score of 4 was sufficient to earn course credit at our state institutions of higher education in several general education courses, such as English, history, and mathematics. After this change was made, several of our faculty indicated that they believed students who earned a 3 instead of a 4 on the AP test were encountering difficulty in

courses once they arrived at Ohio State. This is another example of where the use of data analytics is valuable. We are examining the rather large data set we have from students who have come to Ohio State with AP credits. We are most interested in looking at their performance in subsequent general education courses. If our faculty members' beliefs are correct, we should see poorer performance from students who earned a 3 when compared to those students who earned a 4. We think this same technique can be used to look at the relative performances of transfer students from our community colleges and other institutions with a goal of making sure they are fully prepared for taking courses at Ohio State.

#### Evaluating the Introduction of Technology into the Classroom

Because we are such a big institution—66,000 students distributed across six campuses—we have the numbers to do all kinds of analyses regarding teaching and learning. This is why the data analytics movement is extremely important for Ohio State. Another important area we are looking at is how we can determine whether or not new modes of instruction are at least as effective as traditional methods as technology enters our classrooms at an unprecedented rate. Nationally, a survey by EDUCAUSE in 2011 revealed that 80% of undergraduates said that a laptop computer is “extremely valuable for success.” Of the sample, 37% reported using their smart phone for academic purposes. Not enough tablets were available in 2011 to assess their use, but I suspect the majority of today's students would report using a tablet in their courses. Interestingly, however, only 20%

of the students agreed that their instructors used technology frequently enough and only 22% strongly agreed that their institution effectively used the technology. From these data, we have concluded that students want to use technology to engage with their coursework and they have very high expectations for how our institutions use technology in teaching. Given the high likelihood that more and more technology will be integrated into the students' learning experiences, an important question can be raised: When we integrate technology into the classroom do students learn at the same level or better than with more traditional approaches? Data analytic approaches can be used to answer this question.

For example, in 2011 Jackie Miller and Mark Risser from our statistics department worked with Ohio State instructional designers to convert an undergraduate statistics class to the HyFlex method of lecture delivery. The HyFlex method is a course design model that presents the components of hybrid learning in a flexible course structure. In this statistics class students were given the choice to attend class sessions in person or online (either through a live broadcast or via recorded class sessions). Their choice could be made daily for individual class sessions. Test, quizzes, and assignments were synchronous for all methods. The goal of the instructors was to provide a similar experience for all students across attendance choices. Additionally, backchannel communication and polling were used to enhance student engagement. Backchannel communication is a web-based forum that facilitates real-time conversation among students, as

well as between students and the instructor, during a traditional lecture. Via the internet, students can report and comment on the course content publicly, which provides the presenter direct and immediate feedback and in a manner that is less threatening for students to make contributions to the discussion. Polling is done via standard devices commonly called "clickers" or via text messaging and mobile devices. Polling increases the active involvement of the students, attention levels and interactions, and student comfort in answering questions. Like backchannel communication, it also provides instructors with immediate feedback. Our data suggest more student satisfaction in the course when polling is used.

One of the major goals of the faculty who were piloting this method of teaching was to be certain that those options other than the "in person" attendance option resulted in student performance and learning that was at least equal to that of those who attended "in person." Data collected when the course was offered over three consecutive terms allowed that analysis to be conducted. The instructors looked at student grades based on the primary mode of attendance. They found no significant difference for all three grade categories for students who primarily attended face-to-face versus those who attended and used the distance technology. There was also no significant difference between live attendance and recorded video attendance for two of the three terms; during one term, however, students who attended "in person" or through live streaming did have higher grades.

And there is more good news: We have lots of data from these courses that could shape additional improvement in student learning. For example, we know when the recorded lectures were viewed, and we can look at those data and their relationship with test and quiz performance. We can ask questions like: Does watching the recorded lectures at 3 a.m. lead to poor grades? What is the optimal number of viewings? Which lectures were watched the most times? We also know from student comments that a number of students who attended class in person also watched the recorded lecture as a review; we can assess the value and timing of the reviews. Overall, responses to a survey about the HyFlex model were very positive, with students clearly expressing their preference that the model be available for other classes. And having the data available to analyze and confirm that the HyFlex model produces at least equivalent learning outcomes was critical to a decision we made to expand this methodology across the university.

#### Massive Open Online Courses (MOOCs) and Data Analytics

Everyone seems to be talking about massive open online courses (MOOCs) these days. These are online courses that have been created by several institutions to offer learning opportunities to a very large-scale audience of students. The best MOOCs out there are highly interactive, offering the students several activities in addition to the traditional course materials that are created. MOOCs have the potential of having a significant impact on how we teach and how we reach students. At present they are offered free or at a low cost and they are available literally in all corners of the globe. As of this

writing, at Ohio State we have completed three MOOCs that are offered via the Coursera platform: Calculus One, Writing II, and TechniCity. These courses have a combined enrollment of more than 101,000 students from over 150 countries. To date there have been over 3.3 million interactions with Ohio State lecture content in our MOOCs. These interactions are the reason I am discussing MOOCs here—MOOCs provide us with a tremendous opportunity to understand how to enhance and improve the way teaching and learning happens at Ohio State. We believe these courses are impacting how we think about how we teach and how our students learn.

Jim Fowler from our mathematics department teaches one of these MOOCs; Calculus One, which is an introductory calculus course. He has created a platform for the course called MOOCulus. From a data analytics perspective, the great thing about MOOCs like Fowler's Calculus One is that the enrollments can be very large, which means there are a lot of student and course data with which to work. In this case, more than 35,000 students have enrolled. Perhaps the best way to get a feel for the world of data associated with MOOCs is to hear from Jim Fowler. Here is a link to a short video by Fowler that gives a clear and excellent explanation of how he is using data from the course to improve both his teaching and the students' learning:

<http://www.youtube.com/watch?v=pj-C0JY6mY>

The large data sets collected during the Calculus One MOOC experience provide data analytic opportunities in at least two major areas: assessing overall

course performance, as well as “personalizing” the learning experiences. For example, all student responses are stored from homework and quizzes, such as whether responses were correct or incorrect and the time it took to complete the assignment. A hidden Markov model can then estimate the student’s understanding. And here is what I think is most powerful: Additional practice can be created until the student reaches a threshold, thus “personalizing” the student’s learning. But there is much more here. As the students participate the data are used to refine the Markov model and the database continues to grow. These data can be mined further to answer a number of questions. For example, is there a strong correlation between success on earlier problems in the course and later problems? Can “just in time” recommendations be made regarding previous sections of the course that should be revisited to help the mastery of new material? Can performance in all or a portion of a course predict success in subsequent courses that are taken? It is our hope that, above all, student learning and the student experience is improved significantly by examining the huge data sets that are available through MOOCs.

### **Some Final Thoughts**

I have provided here a few examples of how we are using data analytics at The Ohio State University to help us better understand how well we as faculty and students are performing in the activities that are core to the university: research, teaching, and service. It is clear that our ability to more effectively collect, and more importantly, process and analyze large data sets has enabled us to be much

more data driven in making administrative decisions. Academic Analytics, one of a growing number of data sets that are available, has proven to be useful for comparative reviews of the research productivity of individual faculty as well as departments and programs. At Ohio State we have also begun using data analytic techniques to identify collaborators inside the institution, as well as collaborators outside at other universities and in the private sector. This process should help our faculty connect with others inside and outside of Ohio State, thus enhancing discovery and innovation. We have also used data analytic methods to examine how we teach and how our students learn. For example, 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. But they have another value: MOOCs generate huge amounts of data that can be used to personalize learning and at the same time improve instruction.

For me, the underlying principle for all of these examples is to use the data at hand whenever possible. But we constantly have to remind ourselves to look at all available data whenever possible. Decisions that impact faculty scholarship and teaching should be informed by more than one data point. Indeed, the whole data analytics movement is based on the premise that there are great volumes of data now available and there is

power in the systematic and thorough analysis and interpretation of those data. Scientists have known of the power of data for centuries. Academic administrators need to be mindful of this approach.

### **Acknowledgements**

Over the last couple of years I have been involved in many discussions concerning the use of data analytics in universities in performance evaluations of faculty scholarship, the evaluation of faculty teaching, and enhancement of the student experience. Many of these discussions have taken place at meetings involving my former arts and sciences dean

colleagues, such as at the annual AAU deans meeting. I thank my many colleagues for their insights, with a special thank you to Larry Singell at Indiana University who has presented formally on this topic. At The Ohio State University I wish to thank Julie Carpenter-Hubin, our Assistant Vice President for Institutional Research and Planning, for her huge contributions to this presentation. Julie is truly a data analytics expert. I also thank Jennifer Belisle, Jim Fowler, Alba Clivati-McIntyre, Tom Evans, Liv Gjestvang, and Rob Griffiths.