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

T he need for metrics that quantify the scholarly productivity of PhD programs at universities has been a topic extensively debated for quite some time. In fact, in 2001 the Merrill retreat focused on *Evaluating Research Productivity.* The keynote speaker at that time, Dr. Joan Lorden, stated[1]: *"In choosing measures for the future, we need to bear in mind our goals. Why are we engaged in a measurement process? Are we asking how to move up in the ranks? Or, do we want to know how we have served the state or advanced our mission?"*

The future is now, but these questions continue to be valid as we are only starting to understand new technological tools to attempt to measure research productivity. 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. We now have easy access to large sets of data and numerous analysis tools that can be put to good internal use too, as universities embark in strategic planning and the improvement of the research profile of their programs. We will briefly describe how we are starting to use such data and tools at the University of Kansas (KU) from the perspective of the Office of Research and Graduate Studies (RGS) and the Office of Institutional Research and Planning (OIRP). In particular, we will present how we have been exploring Academic Analytics (AA), and our plans for its use based on

our analysis and feedback received. This project has been a joint effort with Steven Warren, Vice Chancellor, RGS, and Deb Teeter and Sandra Hannon, Director and Associate Director, OIRP. Previous analysis of Academic Analytics and research programs at KU involved also Joshua Rosenbloom, former Associate Vice Chancellor, RGS.

What has changed in research productivity evaluation since the 2001 conference?

One of the main focuses of the 2001 Merrill conference was on the 1995 National Research Council (NRC) study, *Research-Doctorate Programs in the United States: Continuity and Change*. This study was the major systematic data collection regarding graduate programs, broader than its 1982 predecessor. The 1995 NRC study collected important quantitative information about PhD programs, but the rankings were based on surveys. To some extent one can consider that these rankings were based on the reputation of the

programs among peers. One of the criticisms of the rankings was how much the research reputation of the programs affected the reviewers' opinion of their educational quality. Other weaknesses of the 1995 NRC study, as presented in the executive summary of the analysis conducted by The Committee to Examine the Methodology to Assess Research-Doctorate Programs[2], were related to the taxonomy used to classify programs, the obsolescence of the data, and the poor dissemination of the results and the difficulty to access the data. Nevertheless, the rankings of the 1995 NRC study were widely used in many contexts including statistical reports of professional organization. For example, until 2012 the American Mathematical Society used the 1995 NRC rankings to divide all US PhD programs in mathematics into groups on which it compiled annual statistics regarding faculty salary, PhD production, and other quantitative parameters[3]. In some disciplines, substantial correlation was also observed between the NRC rankings and rankings done by other publications such

US News. This correlation is perhaps not surprising since both sets of rankings have been substantially derived from surveys sent to experts in the different fields.

The NRC attempted to address some of the criticisms of the 1995 review in its next study, which was not published until 2010. The 2010 NRC study drastically changed the methodology used. The rankings were based on two different statistically derived analysis of quantitative measures combined in a weighted fashion. Rather than absolute ranks, an interval of confidence was provided for each program. The immediate issues this time, as it was promptly debated in the media, still included the obsolescence and completeness of the data and the convoluted (at least in appearance) methodology employed. We are not aware of any systematic study done to compare the changes in rankings of departments with respect to the 1995 study, but it would be interesting to see if other rankings based mostly on reputation still correlate well or not with the new NRC ones. The future

Figure 1. A hard copy of the 1995 NRC study vs. AA in a smart phone.

of similar NRC studies in years to come is uncertain and many universities rely now for a quantitative analysis of research productivity on commercial tools such as Academic Analytics and self-collected information. Though the needs for such analysis continue to be similar to those in the past, the access to the data and numerous web-based tools is now literally at our fingertips. While the 1995 NRC study almost preceded the World Wide Web and was only available in hard copy, today we can use AA and other resources even in our smart phones.

Moreover, some data and tools are publically available and subject to scrutiny by the general public. It is important then 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 both take advantage of the information for strategic planning and bench marking, but also to properly communicate to different stakeholders true measures of research productivity and how they evince the achievements of our institutions of higher education.

Tools, barriers, and objectives for research productivity analytics

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. We gather data related to research funding such as current and pending awards, but we also try to forecast future funding based on past performances and other parameters. Through our Academic Information Management System (AIMS), we have detailed data about PhD production, time to degree, student support, and student placement for our programs, which can also be combined with demographic information. We are also implementing a new system for self-reported data by faculty. The Professional Record Online (PRO) is a web-based product of Digital Measures, which will not only gather information about faculty but will also produce and update vitas, web pages, and a variety of customized reports. In addition to AA, there are several commercial or publically available tools that provide citations reports, citations maps, h-indexes, and journals impact factors. They include Web of Knowledge, Scopus and Google Scholars, among others. We have also recently subscribed to Pivot, which is another web-based tool that provides information about funding opportunities.

Despite the relatively easy access to tools and information, there are commonly encountered barriers that restrict a wider use of research analytics. In particular, there is not enough "buy-in" about the data/analysis from faculty in certain disciplines, which is compounded by the lack of training and expertise in quantitative analysis in some areas. The analysis of the data is sometimes complex and subject to misinterpretations. Equally important is the fact that the type of data analysis needed could be sometimes extremely time-consuming.

To mitigate some of these barriers we are currently developing a *"consulting service"* model. We will have specialists trained in data analysis and with familiarity with our available tools and databases to assist programs and academic

units with specific reports and requests. Our goal is to help academic programs to analyze the data by

- Looking beyond rankings and indexes, understanding how different metrics affect them;
- Identifying additional discipline specific important metrics and combining them with the PRO system, AIMS reports, and other sources;
- Attempting to compensate with local information some of the lag and incompleteness that the global data may have;
- Identifying relative strengths (weaknesses) of different programs and devising ways to further enhance (reduce) them;
- Customizing our data analysis based on specific goals, strategic initiatives, or requests;
- Exploring funding opportunities that have not been substantially tapped;
- Presenting the data in a comprehensive way that can be easily read, allowing for analysis at increasing levels of depth.

A few simple examples

We will briefly illustrate a few features of Academic Analytics that we have been analyzing in combination with other tools. AA collects information on more than 30 different metrics of research productivity divided into 6 categories: Awards, Publications (articles in journals), Conference Proceedings, Books, Citations, and Grants. The data, as numerous talks at the 2013 Merrill retreat presented, can be displayed in a variety of formats, tables and graphs. The access to raw data also allows for customized local usages. 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 by AA 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.

An important issue is the understanding of the effect of the variable weights in different disciplines. We illustrate this with the following example involving two programs in Figure 2 below. The figure shows the typical *summary of variables* radar plot of AA where the dark
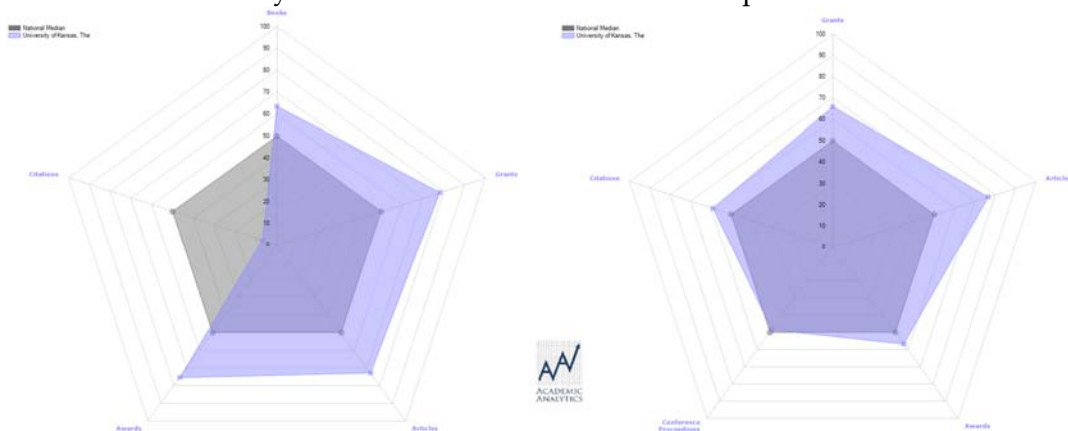

Figure 2. Two different programs in the radar plot of Academic Analytics

99

area represents the median of the discipline at the national level and the light area shows the percentage rank of the program.

After a first look at the summary plot, Program B (on the left of Figure 2) may appear to over preform Program A. Program B ranks at or above the 50% mark in all variables, while Program A appears to be very weak in terms of citations. However a look at the *all variables* radar plot in Figure 3, where we have in-

o No differentiation between Editor/Chapter-author in books is yet available
o Only federal funding is counted
• The Social Sciences fall somewhere in between
o There is more diversity from "book based" disciplines to "article based" ones.

Weights correlate well among related disciplines as can be seen in the following samples from the Natural Sci-
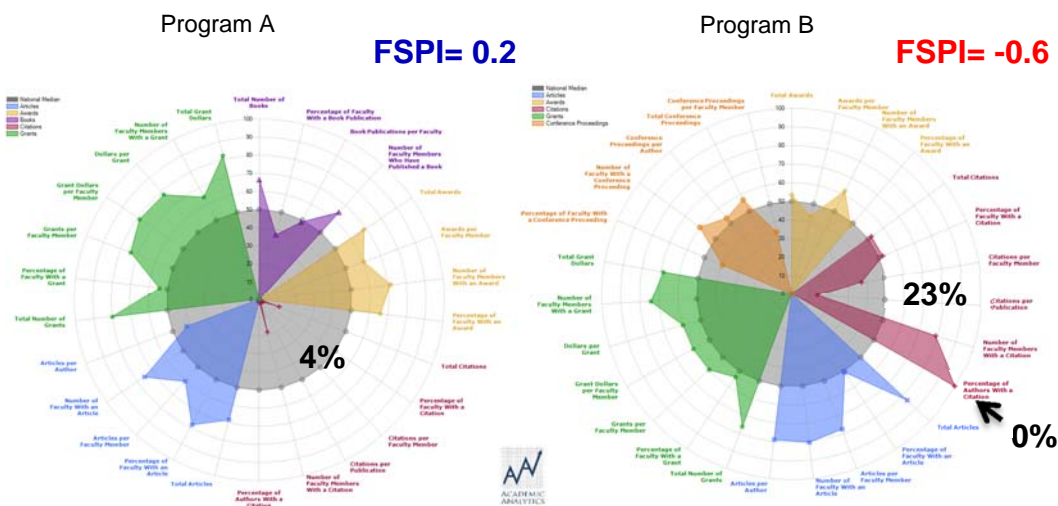

Figure 3. All variables radar plots of two programs in which the relative weights of some variables were added.

cluded the weights of selected variables, reveals why actually Program A has a better profile as indicated by the FSPI.

Some important facts about the variables and their weights to keep in mind are:
• AA metrics best resonate with the STEM disciplines
o Grants and Citations are heavily weighted in STEM fields
• The Humanities have some major criticisms including:
o Citations to/from books are not counted

ences in Figure 4, the Social Sciences in Figure 5, and the Humanities in Figure 6. In the programs in the Natural Sciences displayed, the variables Awards (Aws), Citations (Cit), Publications (Pub), Grants (Grts), are all substantially weighted. Aws weight varies from 12% to 18%, Cit varies from 23% to 28%, Pub from 22% to 31%, Grts 30% to 33%. On the other hand in all these programs Conference Proceedings (Cnfp) are only weighted from 1% or 2% and Books (Bks) are weighted 0%.
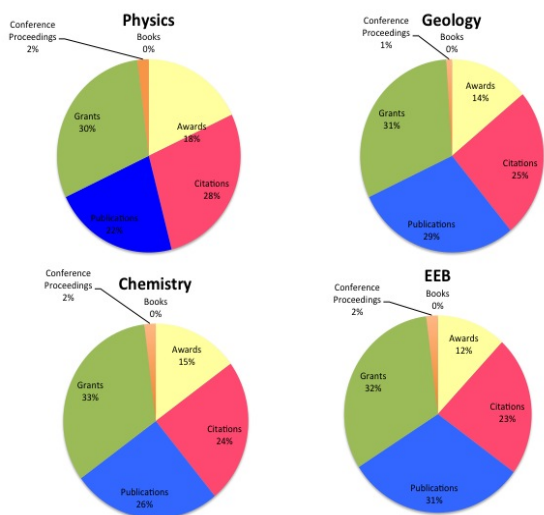
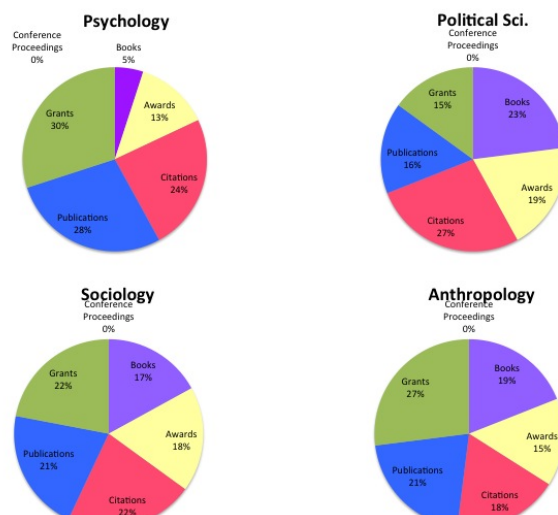**Figure 4.** Sample of AA weights in the Natural Sciences

Figure 5. Sample of AA weights in the Social Sciences.

In the Social Sciences programs displayed, Bks starts to have a more prominent role ranging from 5% to 23% and taking some of the weight from Grts and Pub, but Cnfp remains insubstantial at 0% weight. Cit continues to be close to the 20% range. It is interesting to observe that in Psychology, a discipline with more quantitative aspects but included in the Social Sciences division of KU College of Liberal Arts and Sciences, the weights are distributed more like the Natural Sciences than in the Social Sciences. In particular Grts is again weighted in the 30% range but Bks with 5% weight takes part of the weight in Pub.

Moving into the Humanities, History and Art History exhibit very similar weight distribution, with Bks carrying about half of the total weight. Aws becomes more predominant too, but Pub and Cit are not significant, weighting only from 0% to 4%. Bks remains the variable with the bigger weight in

the Languages but Pub regains more weight, 7%-8%, taking away some weight from Grts and Aws.

As we saw in the earlier comparison between Programs A and B, the visual effect of the radar plots could be misleading if one does not keep in mind the metrics weights. To help in this regard, we
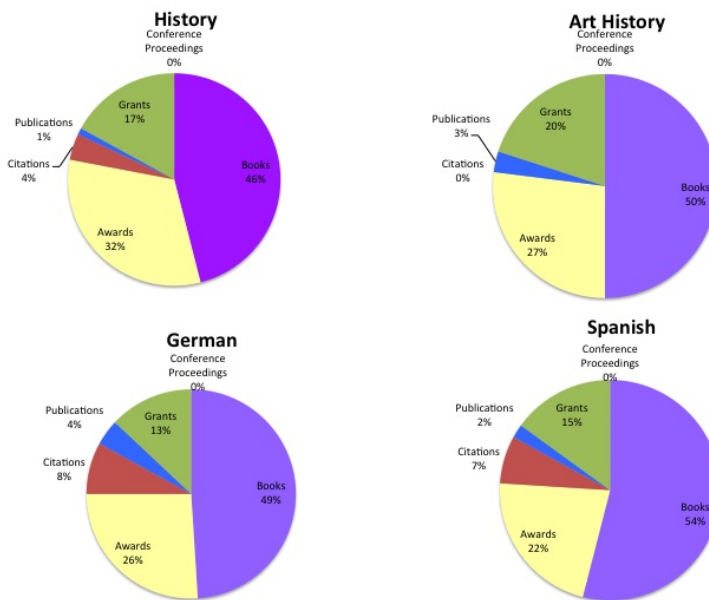

Figure 6. Sample of AA weights in the Humanities.

have developed a new radar plot where the area for each variable is plotted proportionally to the variable. An example of this, analyzing the productivity profile of a program in combination with AA plots, is given in Figure 7. This display gives a propriate venues (conference proceedings as opposed to journal articles). This may help explain in part the relative low citations rates observed. This could translate in a lack of visibility that may also negatively affect the Awards metric. This
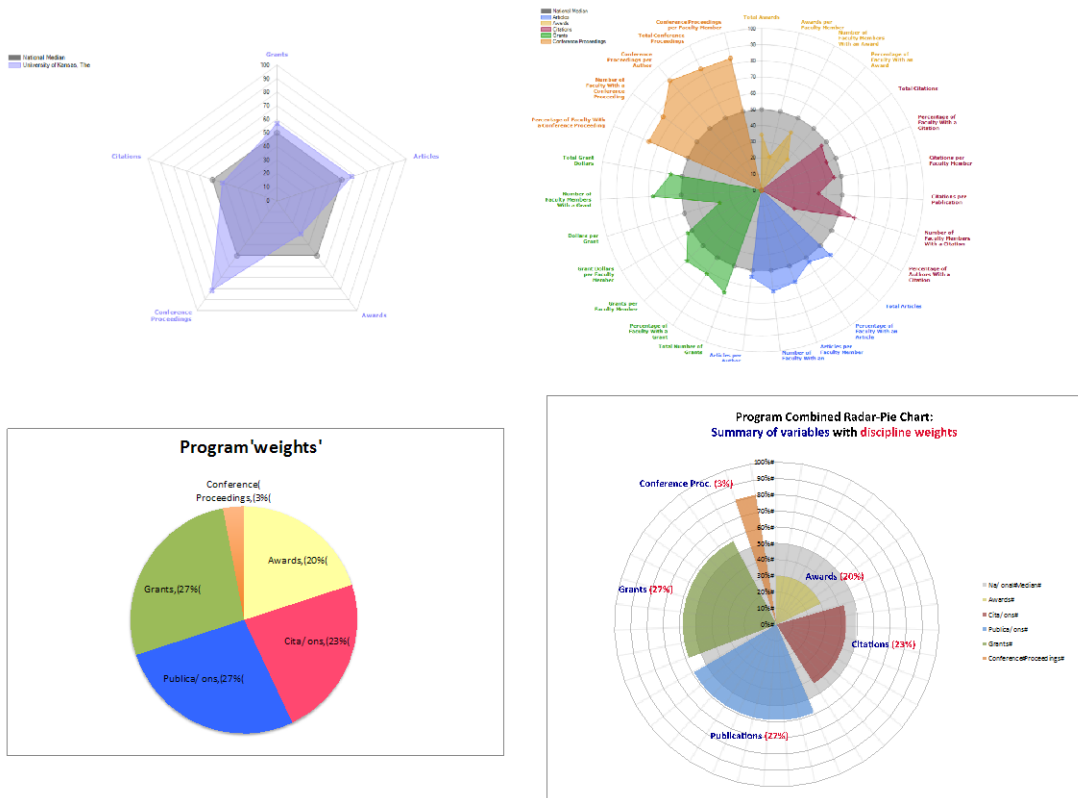

Figure 7. Summary and all variables radar plots of AA combined with custom made plot of variable weights

more complete visualization of the relevance of each metric.

Understanding how the different metrics affect the program profile and how they may relate to each other is of crucial importance. In the above example we see how this program ranks very high in Conference Proceedings (about 80%) but that metric is only weighted by 3% in the discipline at the national level. It may be possible that the faculty members in this program are not publishing in the ap-

observation can be presented to the program for further consideration and possible remedy actions.

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. In Figure 8 we
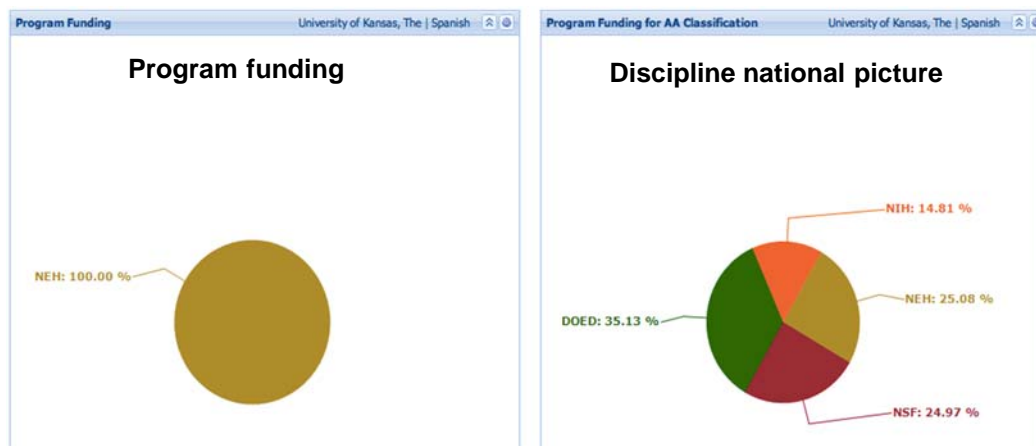
Figure 8. Program funding compared to funding available in the discipline

see a program that is receiving all of its funding from the National Endowment for the Humanities (NEH). This is a well-funded program at KU, yet a comparison with the discipline national picture reveals how the program may be missing on about 75% of the available opportunities. Such opportunities include support in the discipline from the National Institute of Health (NIH), Department of Education (DOED) and the National Science Foundation (NSF). When potential untapped resources have been identified, the already mentioned web-based tool Pivot could be then used to seek specific funding opportunities suitable for the program. Such information could become very valuable for a program trying to increase their external funding.

Other important measures that should be incorporated into a program evaluation and strategic planning are related to data on students' performance and success. The tables in Figure 9 give a snapshot of part of the academic profile of a program. At KU we have such information, which is generated from our AIMS system, reported on the Office of Graduate Studies website for all PhD programs. It would be of interest to explore any correlation of these student performance and demographic metrics with the research productivity metrics in AA.

The few examples presented illustrate how a more detailed analysis of the metrics used in AA, beyond the computation of the FSPI, and the use of additional data resources and tools can help units make decisions to improve their research profile. When some weakness is identified we can use some of the additional data and tools mentioned earlier to look deeper into the sources of such weakness in a multiple level analysis fashion.

Some final comments

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. The tools we have now are only the beginning of better technology in research analytics yet to come. Further databases will be created and aggregated by tools like AA. More accurate and complete sets of data

Doctoral Program Profile:

Department Faculty	Fall 2011
Total Faculty	43
Tenured and Tenure Track Faculty	37
Tenured Faculty	29
Non-tenured Faculty on Tenure Track	8
Other Faculty/Instructors/Lecturers	6
Total Faculty FTE	30.5
Median Age of Ten/TenTrk Faculty (for N>4)	46.0

Doctoral Enrollment	Fall 2008 - 2010 Average	Fall 2011
Total	91.7	89
% Minority	13.5%	11.2%
% International	8.0%	10.1%
% Female	71.6%	69.7%
% Doctoral in Graduate Program	94.8%	91.8%
Average Hours for Doctoral Students	8.3	8.8

Doctoral Admissions: Summer / Fall 2009, 2010, and 2011				
	3-Year Average	Mean GPA	Mean Verbal GRE	Mean Quantitative GRE	Mean Analytical GRE
Applications	220.0	3.7	554.9	657.9	4.6
Admissions	30.3	3.8	601.4	716.7	4.7
Matriculations	16.7	3.7	596.2	698.7	4.6

GPA and GRE data not available for all applicants. See program website for specific admissions requirements.

Financial Aid Awarded to Doctoral Students		
Aid Type	Fall 2011	Mean FY Amount
Inst Scholarships, Fellowships, and Grants	13	$6,568
Need-based Loans	32	$7,632
Non-need-based Loans	10	$8,375
Distinct Student Count / Mean Amount	41	$10,082

Financial aid data for aid year 2012 as of April 12, 2012.

Doctoral Student Support		
Type	Fall 2011	FTE
GTA	51	25.5
GRA	15	7.0

Some assistantships may be outside the program.

Doctoral Degrees Completed			% of Degrees Completed		
Year of Completion	Count	Median Elapsed Years to Degree	Within 5 Yrs	Within 7 Yrs	Within 10 Yrs
FY 2006 - 2008	39	5.9	15.4%	76.9%	92.3%
FY 2009 - 2011	37	5.9	29.7%	78.4%	94.6%

Median years to degree calculation may include both full- and part-time students, and those also obtaining a master's degree during the time interval.

Figure 9. An example of a program profile.

will become available, which hopefully will help us assess disciplines for which some of the current metrics are not significant. For example, it is not hard to imagine that a database tracking performances at major artistic venues could become available in the future, providing a valuable component missing from current metrics in disciplines in the performing arts. Another interesting development is the potential use of Altmetrics[4], which can provide a measure of the impact of scholarly work on social media, blogs and new forms of communications. All these tools add new dimensions to the evaluation of research productivity and should be further explored. While *quality* is not always quantifiable, there are metrics that are indicators of good quality programs. More importantly they could be used to demonstrate to the non-experts why a program is of good quality.

We would like to conclude by citing again some of the words of the main speaker at the 2001 Merrill retreat[1]:

"We will not always agree on what to measure and not everything that we value will be easily captured in quantitative measurements. But as members of the academy, we are in the best position to develop valid measures that will promote our values and apply them in ways that sustain and enhance our mission."

The task continues to be difficult but we now have much better tools at our disposal. 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.

References

1. *Keynote Address* by Joan Lorden, in *Evaluating Research Productivity*, no. 105 - June 2001. A Merrill Center publication on the Research Mission of Public Universities. http://www2.ku.edu/~masc/publications/2001whitepaper/intro.html

2. *Assessing Research-Doctorate Programs. A Methodology Study.* Edited by Jeremiah P. Ostriker and Charlotte V. Kuh. National Research Council (US) Committee to Examine the Methodology for the Assessment of Research-Doctorate Programs. Washington (DC): National Academies Press (US); 2003. ISBN-10: 0-309-09058-X. http://www.ncbi.nlm.nih.gov/books/NBK43475/

3. *Annual Survey of the Mathematical Sciences. Old Annual Survey Groupings of Doctoral Departments (1996 to 2011).* American Mathematical Society. http://www.ams.org/profession/data/annual-survey/groups_des

4. *Almetrics: A Manifesto,* Jason Priem, Dario Taraborelli, Paul Groth, Cameron Neylon. http://altmetrics.org/manifesto/