

An Initiative of NORC AT THE UNIVERSITY OF CHICAGO



Progress and Pitfalls in Tracking U.S. Doctoral Career Paths

May 9, 2019

Tom Hoffer

Senior Fellow, NORC at the University of Chicago

hoffer-tom@norc.org

Debra Stewart

Senior Fellow, NORC at the University of Chicago

stewart-debra@norc.org

Norman Bradburn

Senior Fellow, NORC at the University of Chicago

bradburn-norman@norc.org

Erin Knepler

Research Scientist, NORC at the University of Chicago

knepler-erin@norc.org

INTRODUCTION

In August 2018, NORC at the University of Chicago received a grant from the National Science Foundation to support a spring 2019 workshop and related preparatory work, focused on how doctorate-granting universities and their doctoral programs are collecting and utilizing data on science, technology, engineering, and mathematics (STEM) PhD career pathways. The overriding objectives of the project are to help develop widely shared standards for the kinds of career outcomes that should be measured, the methods for collecting the data, and the ways in which the data should be analyzed and findings disseminated to and utilized by prospective students, faculty, and administrators.

The project has four main activities: a web-based national survey of graduate deans in fall 2018, focus groups of graduate deans in December 2018, a one-and-a-half-day workshop in May 2019, and a multifaceted dissemination of the results of the project. The survey collected data on current practices of monitoring graduates' careers followed by doctorate-granting universities. Questions probed for information on factors facilitating and obstructing development of monitoring systems. Preliminary findings from the survey informed a set of guiding questions that were the subject of the focus groups, which served to highlight, elaborate, qualify, and refine our initial interpretations of the survey findings.

This working paper summarizes and synthesizes the survey and focus group findings and will serve as the guiding framework for the workshop. The workshop will convene a group of 35 graduate deans and research experts to discuss the issues and advance the development of what we refer to as “aspirational standards” for collecting and reporting data on doctoral career pathways. Beyond the workshop, dissemination activities will include publication and broad electronic dissemination of a final project white paper, in-person briefings of government and academic leaders, and conference presentations.

This paper and the research it summarizes were supported by funding from the Division of Graduate Education of the National Science Foundation (Grant No. 1841792). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. The project benefited from the help of a large number of individuals. The authors particularly wish to thank NORC colleagues Karen Grigorian, Bernard Dugoni, Cynthia Simko, Matthew Deihl, Amanda Lynch, Lauren Allensee, Lynette Bertsche, Yajaira Giron, Brian Wagner, Michael Bush, and NORC’s Desktop Publishing Department. We are deeply grateful to the graduate deans and other national leaders in graduate STEM education for their participation in the May 28 and 29, 2019 Workshop on Monitoring Doctoral Career

Paths: Progress and Pitfalls held at NORC in Chicago and to the much larger number of graduate deans who responded to our survey.

Background

A number of factors have likely contributed to the growing interest in doctoral graduate career pathways. Career trajectories for PhD holders typically involve employment in university faculty positions, research positions in the private profit and nonprofit sectors, or administrative or managerial work in higher education, business, or government. While many PhD recipients stay in academia, employment opportunities have changed dramatically with the emergence of federal funding for research, especially from agencies such as the National Science Foundation (NSF), National Institutes of Health (NIH), and Department of Defense. Through this shift, government funds have increasingly been used to support faculty salaries as well as to provide support for graduate students through fellowships and research assistantships. These changes resulted in a steady rise of PhD holders with the annual number of research doctorate recipients increasing 172 percent over the last 30 years (NSF, 2018). While there has been a marked increase in graduates, the proportions of doctorate holders employed in academe, business and industry, and government have changed little. As a result, 60 percent of science, engineering, and health PhDs are employed outside of academe, and only 61 percent of those in academe, or about 25 percent overall, are employed in tenure or tenure-track positions (Bradburn, 2017). The diverse and fluid nature of the career pathways of those with doctoral degrees has provoked considerable discussion about the nature of graduate programs and whether universities are preparing students for the kinds of careers that are open to them in a rapidly changing world.

Emerging Interest in University- and Degree-Program Level Data on Career Paths

Responding to the growing interest of universities and programs in obtaining data on graduates' career paths, a number of coordinated efforts across multiple universities have developed in recent years. Three of the most prominent projects collecting career pathways data are those of the Council of Graduate Schools, Coalition for Next Generation Life Science, and Institute for Research on Innovation and Science. A number of other initiatives underway focus on understanding how doctoral education needs to change in order to better prepare people for the jobs they get and the careers they follow. The American Association of Universities, through its PhD Education Initiative, and the NIH, through its Broadening Experiences in Scientific Training (BEST) program, are examples of these efforts. However, we highlight here the three initiatives that focus on securing data on career outcomes as a prime factor in influencing and improving practice in PhD education.

Council of Graduate Schools Career Pathways Project.

For nearly a decade, graduate schools and individual graduate degree programs have considered the feasibility of collecting information on the career pathways of graduates. Tracking career pathways of graduates was first called for in a report issued by the Council of Graduate Schools (CGS) and Educational Testing Service (ETS) in 2012 (CGS & ETS, 2012). But, according to a 2014 CGS survey of graduate deans, only about one-third of institutions have a formal data collection process. After a series of projects exploring the feasibility of such data collection, the CGS launched the Understanding PhD Career Pathways for Improvement Program in 2017. The program is a three-year project in which 29 institutions are collecting snapshot data from current PhD students and doctoral program alumni in both STEM and the humanities using a mostly standardized data collection instrument. The surveys are collecting data on doctoral students' and recent doctorate

recipients' career aspirations, preparation, and attainments. Notably, the data from this effort will allow descriptions and analysis of career pathways at the doctoral degree-granting program level and promises to better inform decision-making at both the institutional and program levels for those institutions participating. Since a number of universities were also invited to participate as affiliate members of the project, this work could result in collection of data by an additional 30 universities.

Coalition for the Next Generation Life Science project.

The Coalition for Next Generation Life Science (NGLS), launched in 2017, is composed of 30 institutions committed to ongoing collection and dissemination of career data for both graduate and postdoc alumni (Blank et al., 2017; Silva, Mejía, & Watkins, 2019). The Coalition has a mission to provide "...meaningful career exploration and placement support for a broad array of potential career paths, improve mentorship at both the doctoral and postdoctoral stages, and increase and improve recruitment and retention aimed at diversifying the life sciences workforce." Coalition members commit to collecting and disseminating data according to common standards. The scope of the initiative is noteworthy in that it collects information on doctoral life sciences students from point of admission through program completion, and into postdoc and career positions. Dimensions measured and published include:

- Admissions and matriculation counts of PhD students
- Median time-to-degree and completion data for PhD programs
- Demographics of PhD students and postdocs
- Median time in postdoc status at the institution
- Occupations obtained by PhD and postdoc alumni, classified by job sector and career type

Occupational data consists of job titles either reported by graduates on surveys or acquired via web searches (e.g., LinkedIn). A significant feature of the NGLS project is its effort to have participants utilize common employment sector and career type taxonomies to ensure comparability of the data collected on graduates' occupations. The taxonomy was developed collectively in 2017 by representatives of universities with NIH BEST awards, members of Rescuing Biomedical Research, and the founding institutions of NGLS.

Institute for Research on Innovation and Science.

The Institute for Research on Innovation and Science (IRIS) is a consortium of 30 universities hosted at the University of Michigan's Institute for Social Research. Founded in 2015, IRIS has developed an Internal Review Board-approved data repository of individual graduate-level records provided by its members from university human resources and sponsored projects, and supplemented with data on scientific outputs including publications, patents, and dissertations procured from standard sources. These data can in turn be linked to data in the U.S. Census Bureau's Federal Statistical Research Data Center (FSRDC) system which holds economic (employment and earnings) and demographic data. A pilot project

by IRIS is linking student records from two universities with data from the Census Bureau to explore impacts of undergraduate and graduate students' university experiences and degrees on their earnings data five and 10 years postgraduation. IRIS analyzes those outcomes in terms of administrative data from the universities on students' fields of study, enrollment patterns, and demographics. Exploring such data can help institutions better understand outcomes associated with various majors and programs of study, including linking to employment, earnings, and geographic dispersion information.

Taking Stock of Current Initiatives

In an effort to better understand the initial challenges universities encountered when attempting to track PhD career outcomes, NORC at the University of Chicago (NORC) held a stakeholders' workshop in July 2017, in partnership with CGS and with support from the Spencer Foundation, to explore the particular conceptual and methodological barriers to making progress on tracking PhD career outcomes.

Our conclusion from the 2017 workshop was that practical and technical data challenges posed the greatest obstacles to collecting systematic data on PhD career pathways. The workshop ended with a series of questions that inspired the next phase of this work:

- What specific pathways and outcome data are needed?
- What is the appropriate timeframe for collecting them?
- What are the strengths and weaknesses of the main data collection methodologies?
- How can these data most effectively be used to improve programs and achieve transparency?
- How can institutions improve the sustainability of their data collection initiatives?
- How can benchmarking best be achieved?

Since the Spencer-funded workshop in 2017, NORC has begun the next phase of work in this space. The new effort includes the May 2019 workshop and involves both quantitative and qualitative information-gathering from doctoral degree-granting institutions. These institutions include those that have engaged in the CGS Career Pathways Project, others that are independently collecting data on PhD graduate career outcomes, and institutions that have not elected to engage in this topic to date. The overriding objective of NORC's current effort is to help develop widely shared "aspirational standards" for the kinds of career outcomes that should be measured, best practices for data collection methods, and guidance on the ways in which the data should be analyzed and the findings disseminated to prospective students, faculty, and administrators to improve programs and promote consistency, comparability, and transparency across institutions.

Challenges to Collecting Data

The 2017 workshop sought to identify research needs to inform decisions by those most directly connected with graduate degree programs: deans, department heads, faculty, and students. Coming out of the workshop, Stewart and Hoffer (2017) authored a working paper to capture observations and conclusions of workshop participants and to assist university officials as they attempt to better understand the career pathways of their graduates.

The observations and conclusions, especially from the university administrators who attended, point to strong interest in the graduate education community for collecting information about doctoral career pathways. While existing national datasets, especially the NSF's Survey of Doctorate Recipients (SDR), are enormously useful for providing information about the varied doctoral degree pathways at the national and various sub-national levels of aggregation, these datasets fall short in addressing the information needs of two key stakeholders in the graduate education space: students considering application to particular degree programs, and university faculty and administrators working to improve an existing degree program.

Three barriers to acquiring and utilizing doctoral outcome data emerged from workshop discussions: competition, limited budgets, and technical challenges (Stewart & Hoffer, 2017). The conclusions among the participants were broadly as follows:

- **Competition.**

Discussion focused on how doctoral career pathways and outcomes data should be shared. Deans observed that in cases when career outcome data were made available to faculty, faculty were willing to take action, but with respect to how these data were shared, one dean noted that anything this controversial needs to begin with a conversation. Another dean pointed out that "blaming and shaming" rarely works. Several

participants agreed that since the publication of such data might be misinterpreted in a way that could be harmful to programs and institutions, it is very important to “have a plan for how to tell the story before telling the story.” However, there was an emerging consensus that, while the dissemination of information on career outcomes must be curated in a way to reflect a university’s community culture, fear of competition will not in the long run frustrate the broadening of knowledge about graduate career outcomes. Going forward, top graduate programs and universities nationwide are likely to provide information on career outcomes of graduates and see that as a way to gain rather than diminish competitive advantage (Stewart & Hoffer, 2017).

- **Limited budgets.**

Budgetary constraints and competing claims for investment are part of every graduate school’s reality, but in this collection of universities no participant expressed budgetary constraints as a major barrier preventing their university from tracking career outcomes. Discussion focused more on how much of an investment would be needed and if the information could be garnered in a way that was sustainable. In the words of one dean, “Data collection needs to be feasible, easy, and cost-effective to be sustainable” (Stewart & Hoffer, 2017).

- **Technical challenges.**

Technical challenges emerged as the topic that generated the most extensive discussion in the workshop. First, some university officials noted the difficulty of identifying technically prepared individuals, either from the current staff or through new hires, ready to design and implement an effective data collection effort and subsequently analyze the results. One

dean noted that very highly qualified faculty had full research agendas and, understandably, were not inclined to lend their talents to the graduate school data collection effort. Several specific data collection challenges were noted. A major problem for many institutions is simply finding their graduates, especially if they completed the doctorate some years ago. This became even more challenging when graduates pursued non-academic employment with a more limited public record tied to publication. Closely related to the difficulties of locating individuals is the problem of nonresponse and the risk of nonresponse bias in the data that are collected. A number of deans mentioned the special challenges of tracking career pathways of international alumni (Stewart & Hoffer, 2017).

In summary, workshop participants expressed the view that the information needed by applicants to make choices and by program faculty to improve programs would include degree program-based and institution-wide data that describe in some detail the occupational outcomes of program graduates at different career stages. While less broadly endorsed, some participants also emphasized needs for longitudinal data on the same individuals in order to understand career trajectories and dynamics, and for data that illuminate the graduates' assessments of their graduate school education and how it relates to their career paths.

Focus of this Paper: Describing the current landscape and proposing aspirational guidelines

We are familiar with a number of institutions that have been collecting career pathways data for periods of time ranging from one to four years. However, these institutions are not necessarily representative of the graduate school population, and a comprehensive national portrait is lacking. Specifically, more information is needed on what is being done and, importantly, not done; why institutions are collecting data or why not; how the data

are being collected; how well the data collection has worked; and how the data are being analyzed and shared. To address these shortcomings, NORC researchers developed the Doctorate Recipient Career Pathways Survey in fall 2018 and asked the graduate deans of all research doctorate granting institutions with 20 or more doctorate recipients in 2017^① to complete it. The data collected are the focus of this working paper and will be used to inform an upcoming workshop in May 2019. The main goal of the workshop is to seek consensus on which career pathways data to collect, approaches for collection, and how best to analyze and utilize them. The NORC team will follow up on the workshop with a white paper describing the full project and findings, and summarizing progress toward the aspirational standards for collecting and using data on doctoral career pathways.



CURRENT LANDSCAPE OF PATHWAYS RESEARCH

Findings from the fall 2018 Graduate School Survey

The main questions we set out to answer with the survey were the following:

- To what extent are graduate schools collecting data on doctoral career pathways?
- For those not collecting pathways data, why are they not collecting?
- For those collecting pathways data:
 - How are they collecting those data?
 - How are the data being used?
 - How well is the tracking effort going and how might it be improved?

In order to gain a fuller understanding of the survey responses, we also convened two focus groups of graduate deans in December 2018. Points raised in the focus groups are noted in connection with specific survey findings in the discussion that follows.

Survey Participation

NORC invited 257 graduate school deans or others in a similar role to complete the survey in fall 2018. The survey was conducted by self-administered web instrument. The data collection period began on October 16, 2018, and final outreach efforts ended approximately nine weeks later. After response activity ended, the survey was closed on December 15, 2018. The online survey was completed by 130 institutions for a 50 percent response rate.

Examination of the responses from the survey participants revealed that over 80 percent reported their institution was collecting career pathways data. This was a higher national rate than expected based on CGS findings, and we were concerned that an inadvertent nonresponse bias may be affecting our estimate. In particular, we

were concerned that institutions that were not collecting pathways data were less likely to respond to the survey than those engaged in data collection. To assess this, we conducted a single-question follow-up canvassing of the 127 nonresponding institutions, asking simply whether their institution was collecting pathways data or not. Replies to that question were obtained from an additional 45 deans.

A total of 175 individuals responded to the key question of whether their institution is collecting data on doctoral graduates' career pathways for a 68 percent response rate. Of those responding, 70 percent represent public four-year or above institutions, and 30 percent represent private, not-for-profit, four-year or above institutions (see Table 1). Responding deans were predominantly from four-year institutions with "very high" research activity (R1) (58 percent) or "high" research activity (R2) (32 percent). These classes of institutions awarded 75 percent and 16 percent, respectively, of the nation's research doctorates according to the NSF's annual Survey of Earned Doctorates (SED).² About 30 percent of both the R1 and the R2 institutions are private not-for-profit.

Table 1. Responding Institutions, by Carnegie Classification and Sector (n=175)

| Basic Carnegie Classification Description (2018) | Private for-profit, 4-year or above Percent (frequency) | Private not-for-profit, 4-year or above Percent (frequency) | Public, 4-year or above Percent (frequency) |
|---|---|---|---|
| Doctoral universities – very high research activity | 0 | 16% (28) | 42% (73) |

| Basic Carnegie Classification Description (2018) | Private for-profit, 4-year or above Percent (frequency) | Private not-for-profit, 4-year or above Percent (frequency) | Public, 4-year or above Percent (frequency) |
|--|---|---|---|
| Doctoral universities – high research activity | 0 | 9% (16) | 23% (40) |
| Doctoral/professional universities | 1% (2) | 1% (2) | 2% (3) |
| Special focus four-year(a) | 0 | 3% (6) | 3% (5) |

(a) Includes Medical Schools & Centers, Other Health Professions Schools, and Other Special Focus Institutions. Also included here are two private not-for-profit institutions that do not have Carnegie classifications.

Extent of Pathways Data Collection by the Graduate Institutions

The primary question posed in the survey was whether or not the institution is currently directing an institution-wide, centralized effort to collect data on graduates' career pathways. Of the 175 institutions responding to this primary question, 73 or 42 percent indicated that they are collecting data in this way.

This is a fairly high rate given that data collection efforts of this sort were virtually unknown even five years ago. But is that an accurate picture of the national landscape? A definitive answer is not possible without more information about the 83 institutions that did not respond to either the survey or the post-survey follow-up on the

primary yes/no question. If we assume that those 83 institutions are like the 45 that responded to the yes/no question, then about 36 percent of the 83 would be collecting career pathways data on a centralized, institution-wide basis, and the overall rate would drop from the 42 percent to 40 percent. If none of the 83 are collecting on a centralized, institution-wide basis, the estimate would drop to just 28 percent. In any case, we can have some confidence that the national rate is less than 50 percent and likely in the range of 28-40 percent of all graduate institutions.

The nuanced responses to whether an institution is collecting career pathways data are shown in Table 2. Note that the 45 institutions that only answered the primary yes/no follow-up question are included in either category “a” (n=16 of the 73) or category “f” (n=29).

Table 2. Methods of Collecting Career Pathways Data (n=175)

| Method of Collecting Career Pathways Data | Frequency | Percent |
|--|-----------|---------|
| a. Yes, on an institution-wide basis with primarily centralized direction on what kinds of information are collected and how the information is collected | 73 | 41.7% |
| b. Yes, on an institution-wide basis but with decentralized direction on content and methods determined on a department or degree-granting program, school, college, or divisional basis | 13 | 7.4% |
| c. Yes, not institution wide; only by some departments/programs, schools, colleges, or divisions | 38 | 21.7% |
| d. Not currently, but we did in the past | 4 | 2.3% |

| Method of Collecting Career Pathways Data | Frequency | Percent |
|--|-----------|---------|
| e. No, neither now nor in the past | 18 | 10.3% |
| f. No, not on an institution-wide basis with primarily centralized direction on what kinds of information are collected and how the information is collected | 29 | 16.6% |

Institutions Not Collecting Pathways Data

In the presentation that follows, most of our focus will be on the institutions that indicated engagement with a centralized, institution-wide pathways data collection effort. Before turning to that, it is worth noting the information we obtained from the 22 institutions that are not collecting any information on graduate pathways.

Four of these institutions had collected career pathways data in the past but had stopped doing so prior to the 2018-19 academic year. Presented with a list of reasons for stopping, three of the four indicated “lack of staff to collect and analyze the data.” This is consistent with responses from the full set of 22 institutions to the question of why they are not collecting career outcome data, which show that lack of staff to collect and analyze the data along with a lack of accurate contact information on their PhD recipients were the two main reasons identified (Table 3). It is worth noting that no institution cited “political risk of negative findings” as a reason for not collecting career pathways data, and only four institutions indicated that collecting this data was “not currently a priority.”

Table 3. Reasons for Not Collecting Career Outcome Data from PhD Graduates (n=22)

| Primary Reasons | Frequency | Percent |
|---|-----------|---------|
| Lack of accurate contact information | 14 | 63.6% |
| Low response rate from graduates | 9 | 40.9% |
| Not currently a priority | 4 | 18.2% |
| Lack of funding | 6 | 27.3% |
| Political risk of negative findings | 0 | 0% |
| Lack of staff to collect and analyze data | 15 | 68.2% |
| Other | 5 | 22.7% |

Note: Institutions were asked to identify all main reasons for not collecting and could identify more than one reason.

When asked to rate the importance of information on doctoral graduates' career paths, 13 of the 22 schools said it is very important, and another eight said it is somewhat important (Table 4). Asked to rate the likelihood of their institution collecting data on doctoral pathways in the future, 11 of 22 said "somewhat likely," and another five said "very likely" (Table 5). In general, the responses from those not collecting data indicate an interest in getting the data and point more to technical obstacles than political or financial ones, consistent with the themes emphasized by participants in the 2017 workshop.

Table 4. Despite Not Collecting, How Important Is the Information? (n=22)

| Importance of Information on Doctoral Pathways | Frequency | Percent |
|--|-----------|---------|
| Very important | 13 | 59.1% |
| Somewhat important | 8 | 36.4% |
| Not at all important | 1 | 4.6% |

Table 5. Likelihood of Starting to Collect Career Outcome Data from PhD Graduates (n=22)

| Likelihood of Starting to Collect Pathways Data | Frequency | Percent |
|---|-----------|---------|
| Very likely | 5 | 22.7% |
| Pretty likely | 11 | 50.0% |
| Not very likely | 5 | 22.7% |
| Extremely unlikely | 1 | 4.6% |

Methods of Collecting Pathways Data

Turning to the institutions that are collecting pathways data on an institution-wide basis with centralized direction, the first question we address is how they are obtaining the data. Our 2017 workshop discussions suggested that there are three main methods currently being implemented: surveys of doctoral graduates, web scraping^③ data on individuals to identify current activities and achievements from public internet sources, and administrative data linkages.

Surveys are used by almost three-quarters of the institutions, followed by web scraping, which is used by 33 of the 73 institutions. Administrative data linkages are relatively rarely used, with only 12 of the 73 (20 percent) identifying that approach (Table 6).

Table 6. Methods for Tracking Doctorate Recipients' Careers with Centralized Collection (n=73)

| Method | Frequency | Percent |
|--|-----------|---------|
| Surveys of graduates | 43 | 72.9% |
| Web-based data collection to obtain publicly available online information, including automated web scraping or manual review of social media, such as LinkedIn | 33 | 55.9% |
| Administrative data collection through linking university records with various government-maintained databases | 12 | 20.3% |
| Other | 13 | 22.0% |

Note: Respondents indicated all methods that applied and some indicated two or more.

A number of important options are available to each of the three main data collection methods, including whether to collect data from or about all graduates versus from a representative sample of them, which university units should collect the data, how often the data should be collected, whether to collect data longitudinally or cross-sectionally, and whether to collect data from graduates residing outside the United States.

Sample versus population.

For both surveys and web scraping, a key decision is whether to collect data from a sample versus all eligible doctorate recipients. Sampling can be a cost-effective strategy but is only applicable to institutions and degree-granting programs that have relatively large numbers of doctorate recipients. Efficiencies associated with sampling generally are not relevant to administrative data linkage method since the record matching is typically programmed and involves matching all graduate records from an institution dataset, with massive population datasets maintained by the federal government or state governments.

Over 80 percent of the survey respondents indicating use of surveys to collect pathways data collect those data from all doctorate recipients, and only eight of the 43 (19 percent) report surveying a sample.⁴ A similar pattern is evident for those using web scraping, with 73 percent collecting from all doctorate recipients, and only six of 33 institutions using sampling (Table 7).

Table 7. Census versus Sampling of Doctorate Recipients at Institutions with Centralized, Institution-Wide Data Collection

| Were Data Collected from All Doctorate Recipients or from a Sample? | Survey Data Collection (n=43 institutions) | | Web Scraping Data Collection (n=33 institutions) | |
|---|--|---------|--|---------|
| | Frequency | Percent | Frequency | Percent |
| | | | | |

| Were Data Collected from All Doctorate Recipients or from a Sample? | Survey Data Collection (n=43 institutions) | | Web Scraping Data Collection (n=33 institutions) | |
|---|--|-------|--|-------|
| All research doctorate recipients at the institution | 35 | 81.4% | 24 | 72.7% |
| Only a sample | 8 | 18.6% | 6 | 18.2% |
| Don't know | 0 | 0 | 3 | 9.1% |

Which university units collect data?

The universities implementing centralized, institution-wide data collection report that a variety of institutional units are involved in the effort (Table 8). The modal response to the question was the graduate dean's office (79 percent). Paradoxically, 68 percent of the respondents from these centralized institutions noted that sub-centralized individual programs or departments, schools, colleges, or divisions have data collection responsibilities. What seems to be the case here is that data collection responsibilities are delegated by a central authority (e.g., the graduate dean or provost) to institutional sub-units that collect the data using an institution-wide template and common methodology.

Table 8. Institutional Units Collecting Career Outcome Data at Universities with Centralized, University-Wide Data Collection (n=57)

| Institutional Unit | Frequency | Percent |
|--|------------------|----------------|
| Individual programs/departments, schools, colleges, or divisions | 39 | 68.4% |
| Alumni or development office | 21 | 36.8% |
| Graduate dean's office | 45 | 78.9% |
| Institutional research office | 18 | 31.6% |
| An outside organization | 3 | 5.3% |
| Other, please specify | 6 | 10.5% |

Note: Respondents indicated all units that collected data, and some indicated two or more.

As a corroborating note, a much higher proportion, almost 90 percent, of the institutions with a decentralized approach to data collection (i.e., those in categories “b” and “c” in Table 2) identify individual programs or departments, schools, colleges, or divisions as the loci of data collection responsibilities.

How often are data collected?

The periodicity of data collection is an important consideration for institutions. The risks of collecting too frequently are: having more data than are needed to draw the relevant inferences, and wasting time and money in the process. The risk of not collecting frequently enough is that important changes might be missed, leaving the institution and degree-granting programs without the insights they need to make informed decisions. The survey results show that over one-third (36 percent) of the schools with centralized, institution-

wide graduate tracking indicate they are collecting data annually while the rest collect data every two years to every five years (Table 9). As noted above with respect to sampling, here we are unable to distinguish between institutions whose engagement is proscribed by the CGS project requirements and those who, independent of CGS project requirements, would have decided to collect data every year on an ongoing basis. Where yearly data collection was a function of CGS project participation, it remains unclear whether the data collection will continue after year two on a yearly basis. Some of the focus group comments suggested that continuation for some institutions is problematic.

Table 9. Frequency of Data Collection on Graduates' Career Outcomes at Universities with Centralized, University-Wide Data Collection (n=52)

| How Often | Frequency | Percent |
|--------------------------------|-----------|---------|
| Every year | 20 | 36.4% |
| Every two years | 4 | 7.3% |
| Every third year | 2 | 3.6% |
| Every fourth year | 1 | 1.8% |
| Every fifth year | 5 | 7.3% |
| Other interval; please specify | 11 | 20.0% |
| No regular interval | 9 | 16.4% |

The responses from the relatively large number of “other interval” institutions (n=11) indicate some complexity around the question of periodicity. Three of the 11 reported that they collect data only upon graduation, which seems the same as “every year” and which is not really collecting data on their graduates’ career outcomes. It is more akin to an exit survey. Five of the 11 reported collecting data at intervals aligned with or similar to the CGS Pathways Project. For the CGS Pathways Project, affiliate institutions initially agreed to participate for two consecutive years in fall 2017 and fall 2018. Those schools agreed to distribute the Alumni Survey to those earning their doctorate after three, eight, and 15 years post-degree award, which would effectively capture career pathway information from six separate doctoral cohorts.

These responses point to the need for further research to distinguish among the following temporal dimensions:

- Graduating cohorts: Is each new cohort part of the data collection plan, or are some cohorts followed and other cohorts skipped? Which cohorts are included?
- Data collection cycles: For cohorts included in the data collection project, are data collected at just one point in time, or from at least some cohorts at two or more points in time?

How are the data collected: longitudinally, cross-sectionally, or both?

Collecting longitudinal data on the same graduates over time is attractive in that it provides a picture of career stability and change that add important detail to a simple cross-sectional slice.

Longitudinal data collection poses challenges in that it requires maintaining and updating locating information in order to collect the data needed to chart changes over time. However, because updating

work is more difficult and expensive for individuals farther removed in time from when they were last found, the periodic updating of a longitudinal database for follow-up data collections can actually be easier and less expensive compared to locating a new, comparably aged, cross-sectional cohort that has not been updated since earning the doctorate. The survey found that two-thirds of the centralized, institution-wide schools indicate using either exclusive longitudinal (36.4 percent) or mixed longitudinal and cross-sectional (30.9 percent) data collection (Table 10).

Table 10. Longitudinal and Cross-Sectional Data Collection (n=51)

| Institutional Unit | Frequency | Percent |
|---|-----------|---------|
| Longitudinally (same individuals over time) | 20 | 36.4% |
| Cross-sectionally (following different sets of individuals each time) | 9 | 16.4% |
| Both longitudinally and cross-sectionally | 17 | 30.9% |
| Other | 5 | 9.1% |

Try to track outside the United States?

Data from the SED indicate that the numbers of non-U.S. citizens earning research doctorates in the United States while on temporary visas are equal to about one-third of all STEM-field doctorate recipients in recent years. Rates of staying in the United States following receipt of the doctorate vary by country of origin but exceed 70 percent overall and are up to 90 percent for Chinese and

Indian nationals. However, the SDR data show that many of those who initially stay in the United States after earning their doctorate leave the United States after completing a postdoc. And foreign employment for U.S. citizens is also increasingly common. In any case, collecting career pathways data from graduates residing outside the United States can pose challenges, particularly with respect to obtaining current email and other contact information needed to reach them. Reflecting these difficulties, almost half of the centralized, institution-wide institutions indicated that they do not follow up on those living outside the United States (Table 11).

Table 11. Tracking Outside the United States (n=53)

| Tracking Internationally | Frequency | Percent |
|--------------------------|-----------|---------|
| Yes | 27 | 50.9% |
| No | 26 | 49.1% |

Current Practices to Utilize Data to Improve Programs

The survey asked the graduate deans at institutions collecting the data how the data are being used and with whom the data are being shared. Again, we focus on the responses from institutions implementing a centrally directed, institution-wide data collection program, under the assumption that those respondents are more likely to have an accurate reading.

How is your institution using the data?

The two main uses of the data reported by the deans are to inform guidance services (88 percent) and to inform prospective students (84 percent) (Table 12). Both of those uses are likely to entail providing information on the kinds of career paths past doctorate recipients are following, which can be helpful to current and prospective students who otherwise might have very little understanding of non-academic career options.

Outreach to graduates is affirmed by 63 percent and likely consists of efforts to build or maintain bridges that can be helpful to both the institution and the graduate. For the institution, graduates can be helpful references for current students or recent graduates for information about career options in particular universities, firms, and broader labor market sectors. For the graduates, maintaining connections to the institution can provide useful recommendations on new graduates for alumni involved with recruiting and hiring, but may also be useful for gaining information from the institution on new career options for those graduates looking to pursue new opportunities. ⁵

Internal accountability (59 percent) and informing curriculum (54 percent) are also significant uses. Accountability pressures for information on graduates' career outcomes have increased in recent years, particularly at the undergraduate level but also at the graduate degree program level. The main thrust of this type of accountability effort is to assess the extent to which graduates are able to find employment in jobs related to their degree programs. Variants on that theme may include the extent to which graduates find such employment and have an economic and scientific impact in the same state as the degree-granting university (measured by such outcomes as income, publications, and patents).

Informing curriculum seems likely to take the form of indicating content that aligns better with actual career lines graduates are pursuing, and this could be elaborated with feedback from graduates on how the curriculum they followed could have been better matched to the knowledge and skills they have found to be most important in their respective career paths.

Table 12. Institutional Uses of Data

| Institutional Uses of Data (n=56) | Frequency | Percent |
|---|-----------|---------|
| To inform curriculum (course content) | 30 | 53.6% |
| To inform instruction (how classes are taught) | 13 | 23.6% |
| To inform career guidance services | 49 | 87.5% |
| To inform prospective students | 47 | 83.9% |
| To inform outreach to graduates | 35 | 62.5% |
| For internal accountability purposes (resource allocations, program review, etc.) | 33 | 58.9% |
| For fundraising | 23 | 41.1% |
| Other | 6 | 10.7% |
| It is not used in any systematic way | 3 | 5.4% |

Note: Respondents indicated all institutional uses of the data, and some indicated two or more.

To whom are the data disseminated?

University administrative leadership are the main recipients of the data currently being collected, with survey respondents indicating the central administration and deans (both 69 percent), followed by department heads (64 percent) as the most frequently named recipients. Department heads are likely to share with their faculties, but the faculties appear to not necessarily be direct recipients in about half of the institutions.

Interestingly, prospective and current students are only named as recipients by 42 percent and 33 percent of the institutions, respectively, which is less than what the intended use responses (Table 13) would suggest. This may simply be another instance of departments and career guidance services handling the distribution to the students rather than the students receiving the information in a direct way from the central administration. In a similar vein, the relatively large number of “other” responses (n=15) were almost all clarifications that the information is posted on one or more university websites and is available to the public.

Table 13. Recipients of Disseminated Information (n=55)

| Recipients of Disseminated Information | Frequency | Percent |
|--|-----------|---------|
| Program faculty | 26 | 47.3% |
| Department heads | 35 | 63.6% |
| Central administration | 38 | 69.1% |
| Deans of schools or colleges | 38 | 69.1% |

| Recipients of Disseminated Information | Frequency | Percent |
|--|-----------|---------|
| Board of trustees | 12 | 21.8% |
| Current students | 18 | 32.7% |
| Prospective students | 23 | 41.8% |
| Alumni or development offices | 19 | 34.5% |
| Other | 15 | 28.8% |
| Don't know | 1 | 1.9% |

Note: Respondents indicated all recipients, and some indicated two or more.

Role of federal data in facilitating institutional understanding (from focus groups).

One topic that was not asked about in the survey but that emerged as a noteworthy subject in the focus groups we conducted at the annual CGS meeting in December 2018 was the potential utility and availability of federal data on graduate career pathways, particularly as collected by the NSF SDR. In response to a question of whether they are familiar with those data, most of the focus group deans affirmed they are but that their use is sharply limited by difficulties accessing the data for analysis and, once the data are obtained, by the relatively high level of aggregation at which the data are reported.

The level of aggregation issue has recently changed dramatically due to a major redesign and increase of the SDR sample, such that the survey now supports reporting at most of the fine fields of

doctoral study that are used in the annual SED. While the SDR data still cannot represent particular doctorate-granting institutions, they can provide useful data on career paths of graduates in a broad array of doctoral fields with further breakdowns by broad classes of doctoral institutions (e.g., Carnegie classification categories). These data have the potential to be useful benchmarking tools for universities and degree programs, enabling comparisons of their local data with high-quality national data.

The strength of benchmark comparisons critically depends on commonality of measures and data collection methodology, particularly response rates and coverage of the intended population. Focus group respondents expressed interest in exploring further ways to better align career pathways instruments and methods to capitalize on the rich national resources of the SDR.

Successes and Shortcomings of Current Efforts

The survey questionnaire concluded with a number of questions about how well the data collection efforts are working and how they might be improved. An important standard question about any data collection effort concerns the extent to which data were successfully collected from or about the population of interest. The populations are typically defined in terms of all doctorate recipients from the institution after a certain period of years (e.g., one year, five years, 10 years ago). For institutions conducting surveys of doctoral graduates, the standard measure of success is the response rate defined as the percentage of all eligible doctoral graduates from those cohorts that completed the pathways survey. For institutions collecting data via web scraping, the analogous response rate is defined as the percentage of all eligible doctorates for whom relevant career information could be found and acquired.

Satisfaction with current efforts.

The modal response rates reported by the graduate deans at the centralized, institution-wide institutions were 50 percent for their surveys and 73 percent for the web scraping. These rates are fairly high for university-based data collection efforts but still raise important questions about how representative the respondents are and whether the data support generalizations to the target populations. Two issues require consideration. The first is nonresponse bias. Just as we have concerns about nonresponse biases in the data we have collected from the graduate deans, the institutions need to take a critical stance toward the data they have collected, and particularly consider whether the nonrespondents are likely to systematically differ from the respondents in their career pathways experiences and outcomes. The second is consistent and standardized use of methods for calculating response rates. For example, it is essential that standard statistical practices be followed in defining the denominator on which the calculation is based. Focus group discussions suggested that there are challenges with respect to both of these issues.

The satisfaction questions pertained to three aspects of the institutions' career pathways data collection efforts: the institution's ability to collect, analyze, and utilize the data. Slight majorities of the deans reported being very or generally satisfied with their institution's abilities to collect (56 percent) and analyze (54 percent) the data, but only 46 percent are very or generally satisfied with the utilization of the data (Table 14). The overall evaluation of how successfully they are tracking graduate career paths found 67 percent rating it very or somewhat successful (Table 15).

Table 14. Satisfaction Levels with Institution's Ability to Collect, Analyze, and Utilize PhD Career Outcome Data

| Rating | Satisfaction with Collecting Data (n=55) | Satisfaction with Analysis of Data (n=52) | Satisfaction with Utilizing Data (n=57) |
|---------------------|---|--|--|
| Very satisfied | 12.7% (7) | 17.3% (9) | 10.5% (6) |
| Generally satisfied | 43.6% (24) | 36.6% (18) | 35.1% (20) |
| Somewhat satisfied | 29.1% (16) | 23.1% (12) | 26.3% (15) |
| Very dissatisfied | 14.5% (8) | 1.9% (1) | 5.3% (3) |
| Too soon to say | N/A | 23.1% (12) | 22.8% (13) |

Table 15. Evaluation of Institution's Success in Tracking Careers

| How Successful in Tracking Careers? (n=57) | Frequency | Percent |
|---|------------------|----------------|
| Very successful | 11 | 19.3% |
| Somewhat successful | 27 | 47.4% |
| Not successful | 8 | 14.0% |
| No opinion yet – just started to track this academic year | 11 | 19.3% |

Areas of improvement.

The final question on the survey asked the deans in an open-ended format to note how the tracking of graduate careers might be improved at their institution (see list of verbatim responses in the Appendix). We received comments from 71 respondents. Responses clustered around four main themes, described in Table 16. By far the most numerous (n=44) were general calls for greater institutional support in the form of increased staffing support, greater centralized direction, and increased resources. A second large cluster (n=23) focused on study design issues, calling for longitudinal follow-ups on graduates, following graduates for longer periods (e.g., 10 instead of 5 years), more uniformity in the periodicity of data collection, and improvements in the questions asked of the graduates. Ten respondents noted the need to do a better job of locating graduates, including those outside the United States. Finally, four respondents noted the need to improve analysis and reporting of the data being collected.

Table 16. Categories of Responses to Question of How Career Tracking Efforts Could Be Improved

Grouping of Open-Ended Responses to the Question: How might the doctoral career tracking efforts at your institution be improved? (n=71)

Group 1. Institutional support – increasing staffing support, data collection project organization including greater centralized direction, use incentives for graduates to respond and departments to collect, and more resources (n=46)

Group 2. Study design – longitudinal vs. cross-sectional, periodicity of data collection, question content (n=21)

Group 3. Tracking – improving the locating of graduates, tracking non-U.S.-resident graduates (n=9)

Grouping of Open-Ended Responses to the Question: How might the doctoral career tracking efforts at your institution be improved? (n=71)

Group 4. Analysis and reporting – make analysis more systematic, integrate data into data warehouse (n=5)

ASPIRATIONAL GUIDELINES FOR CONSIDERATION

A main goal of our current effort is to identify and, through the data we are collecting and discussions about it with the graduate deans, contribute to building an emerging consensus on 1) what data graduate career tracking efforts should collect, 2) how best to collect them, 3) how often to collect them, and 4) how best to analyze and disseminate the data collected. The forthcoming workshop goal is to ascertain the extent to which a level of consensus can be achieved with respect to these four topics. At this project's conclusion, we hope to posit a set of guidelines or recommendations that can be followed and further tested across a carefully selected set of universities. In order to reach this set of recommendations, we will structure each segment of the workshop to address a set of questions embedded in each broad segment, designed to focus and catalyze discussion. For purposes of facilitating the discussion, we also advance sample aspirational guidelines. Note that these are examples, suggested only to illustrate the kinds of concrete directional statements that could guide institutions as they create systems and processes for collecting and utilizing career pathways information. The final guidelines or aspirational statements, published in the fall of 2019, will represent to the extent possible a consensus view from the May 2019 meeting's participants.

Content: What data should be collected?

Our survey of deans did not ask about the kinds of data being collected or what kinds they would most like to collect. We did pursue this question in our focus group sessions with responses that ranged widely: some deans expressing a desire for minimal information relating to job title to others desiring much more information, including respondent's reflection on their graduate school experience. We have also collected a number of surveys that illustrate the range of outcomes of which universities desired information. Here, we simply inventory the range of topics covered across all of the multi-university initiatives currently underway and highlight the areas of convergence and divergence across those initiatives (CGS, NGLS, IRIS, and SDR).

The CGS Career Pathways Project collects a broad range of data on graduates' employment and occupations, including job title, employment sector, type of employment (e.g., regular vs temporary, postdoc, tenure-eligible and tenured, self-employed, part-time vs full-time), main work activities and skill requirements, how closely related one's job is to one's field of doctoral study, and job satisfaction. In these respects the CGS survey is very close to the SDR, and in fact seeks to build in extensive overlap in order to use SDR findings for national benchmarking against the institutions' program-specific graduates. The CGS survey goes beyond the scope of the SDR to also include questions about the graduate school experience and how well it prepared the surveyed graduates for the workforce. Whereas the SDR builds career history or trajectory data on sample members through the project's longitudinal panel design, the CGS survey includes batteries of retrospective questions about prior employment in an effort to build individual employment histories within a cross-sectional design.

This extent and level of detail in the CGS survey contrasts with the more closely-focused approaches of the NGLS and IRIS projects, which aim strictly for employment sector and occupation descriptors for each graduate. NGLS collects data through a very brief four-question email or web-based survey asking about current employment (job title and name and address of one's employing organization/ institution/ company). Data on survey nonrespondents are gathered via extensive web scraping or "cyber-sleuthing" search activities (Silva, Mejia, & Watkins, 2017). The job-title and employer data are then linked to taxonomies and job-content databases that allow standard classifications of the data for reporting and comparison purposes.

IRIS members provide individual graduate-level administrative records compiled by university human resources and sponsored projects, and supplemented with data on scientific outputs including publications, patents, and dissertations procured from standard sources. These data can in turn be linked to data in the U.S. Census Bureau's Federal Statistical Research Data Center system (see <https://www.census.gov/fsrdc> <<https://www.census.gov/fsrdc>> for more information) to link to economic (employment and earnings) and demographic data. The main purpose of the IRIS system is to document impacts of research-trained individuals on science and innovation in the public interest. While these outcomes can be linked back to graduate programs, the aim of IRIS is more toward producing institution-level outward-facing documentation and dissemination on specific types of scientific and economic impacts graduates have in their careers.

Reviewing the collections of possible data elements currently in use raises the following questions:

- What are the most important things to know about career experiences of doctoral graduates? (In answering this question ask yourself what are the essential items without which collecting data of these kind would not be worth the effort? And then in addition to these essential items how would you rank the other important data elements required?)
- To what extent do the data elements need to be adapted to the field of doctoral study?
- On which data elements is standardization most important as it would be required for benchmarking across programs within the university or across programs across universities?
- Are either the NGLS taxonomy or long-established taxonomies from the NSF's National Center for Science and Engineering Statistics and U.S. Department of Education National Center for Education Statistics viable starting points for adopting a STEM PhD occupations taxonomy?

To help frame the forthcoming discussion, we present **examples of aspirational guidelines that could emerge from this discussion:**

- *Irrespective of data collection methods, career pathways data collection should include information on principal job, employer, and employment changes over the early career period (up to about 15 years out).*
- *Graduates' assessments of the value of their doctoral programs and how they might be improved are valuable and need to be obtained.*
- *Taxonomies utilized to classify raw response data on principal job should represent the full range of doctoral-level occupations and work activities to accommodate the diverse range of outcomes across doctoral fields, while simultaneously allowing for essential field-specific classification.*

Methods: How best to collect the data?

Our survey found that surveys and web scraping are the most widely used methods of data collection. Each has its strengths and weaknesses. Surveys allow standardization of the data collected by means of presenting identically worded questions and response formats to all targeted individuals. However, they require accurate information on how to contact the target individuals, and, even with that information, most individuals will not complete a survey without considerable work on the part of the data collector to make contact, explain the project, issue several reminders, and, for many, provision of some sort of monetary incentive to complete. Even at that, response rates above 60 percent are difficult to obtain even on relatively brief questionnaires taking 10-15 minutes.

Web scraping sacrifices the standardization of a survey in favor of the online convenience of collecting data from social media such as LinkedIn and Facebook, and personal and employer website postings. Web-based information can be extraordinarily rich for many doctorate recipients, especially those who purposively build their online presence through regular posting of publications, project descriptions, and Twitter feeds and blogs. While the numbers with web profiles sufficient to support a career tracking effort are growing, many doctorate recipients do not maintain those profiles or never establish them in the first place. Consequently, response rates to web scraping efforts are not likely to exceed coverage for 60 percent of the doctorates—that is, a response rate that stronger survey efforts are able to realize.

Whether one uses surveys or web scraping or some combination of the two as the mechanism for collecting career pathways data, the problem of nonresponse and the associated potential for nonresponse bias should be addressed. Nonresponse bias occurs when graduates responding to your survey, or who you find in cyberspace, differ in meaningful ways from graduates who did not respond or could not be found. In our work thus far, we've identified two potential sources of nonresponse bias. The first is straightforward. Assuming that you have reached out to collect information on all of your defined population of graduates, the individuals on whom you have failed to collect information, either because they failed to respond to your survey or you were unable to secure information through web scraping, may differ in some systematic way from those on whom you have information. The second source of nonresponse bias comes from survey-based data collections that proceed with incomplete contact information on the targeted population. In these cases, by reaching out only to those for whom contact information is available, nonresponse bias enters even in the face of very high response rates since it is not known how those without available contact information would have responded.

With respect to collecting data on career pathways, the workshop will consider the following questions:

- How do you assess the relative value of surveys versus web scraping versus some combination of the two?
- What is the appropriate response rate for career pathways survey? What is the appropriate success rate for a web scraping data collection?
- How does one verify that web scraped data is contemporary?
- How does one verify that web scraped data actually belong to the graduate you are searching for?
- What strategies have you utilized for dealing with nonresponse bias?

Examples of aspirational guidelines that could emerge from the discussion:

- *All universities should aim to collect basic information on graduates' current employment using a combination of brief email or online surveys followed by a strong secondary effort utilizing web scraping to collect information on survey nonrespondents.*
- *For universities desiring more information than their graduates' current employment, more in-depth web-based surveys should be used with strong follow-up prompting efforts, including a paper-and-pencil questionnaire option to improve survey response.*

Periodicity: How often should data be collected and what mix of cross-sectional and longitudinal data?

Our survey found considerable variability and complexity in the responses to the question of how often the institutions are collecting pathways data. The modal response was “every year,” but it was not clear what those efforts entail or the extent to which the respondent had in mind every year for the two-year duration of a funded project. On the one hand, annual efforts to locate individuals and update employment and occupational information via web scraping and even very brief surveys can establish valuable ongoing connections and provide useful pathways data. Locating graduates becomes increasingly difficult the longer the time between locating efforts. On the other hand, annual or even biennial data collection efforts may carry significant costs and may yield more data than is needed or is manageable, and thus waste resources.

It is crucial to consider tradeoffs among frequency, respondent burden, data collection costs, data processing and analysis costs, and how the information collected should be employed by the university and its degree-granting programs.

Discussion questions for the workshop:

- How often should graduates be asked to provide information on their careers?
 - What are the most useful cycles for longitudinal data collection (e.g., 1, 3, 5, 7, 9, etc. years after degree completion)?
 - Is annual web scraping feasible? Is an annual survey feasible?
 - Should all or only a sample of graduating cohorts be included in longitudinal follow-up efforts?
- For which particular career pathways questions are longitudinal data needed?

Examples of aspirational guidelines that could emerge from the discussion:

- *Institutions should make a strong effort to obtain contacting information on all doctorate recipients within the first year following degree completion and do so for each graduating cohort.*
- *New doctorate recipients should be strongly encouraged to join LinkedIn or to provide routine updates on their employment to the university to improve efficiency and coverage of whatever efforts are taken by the degree-granting institution to follow their careers.*
- *Institutions should periodically survey graduates for feedback on how well their degree program prepared them for their career to date and how the program might be improved to better prepare future graduates.*

Findings: How best to analyze and disseminate the data collected?

Our survey could not go into any depth on how data are being analyzed and disseminated. The workshop will provide an opportunity to share information on what participants have found to be useful and effective as well as strategies and tools that have not worked well.

Discussion questions for the workshop:

- What are the most effective strategies for sharing career pathways data with programs and with faculty?
- Are there particular approaches to dissemination of career pathway information that are especially effective in shaping curriculum?
- What kind of information is appropriate to share with faculty, with deans and department heads, and with students (i.e., inward-facing reports)?
- What kind of information is appropriate to share with the public at large, including prospective students (i.e., outward-facing reports)?
- What media are best for outward-facing reports? Are online dashboards helpful for internal reporting?
- How should these data be used in standard graduate program reviews?

Example of aspirational guidelines that could emerge from the discussion:

- *Institutions should provide a public-facing information outlet on the university's general website or departmental websites to provide prospective students with information on graduates' career paths and outcomes.*
- *Institutions should provide internal-facing data dashboard facilities for their doctoral degree-granting programs. This should provide comparisons with peer institutions and national data at the appropriate field of study level within the institution's Carnegie class as well as for all institutions.*

CONCLUSIONS

Next Steps

As we approach the May 2019 workshop, we have learned many things from our research to date. We know, for example, that a substantial number of PhD degree-granting universities are making a serious effort to collect and utilize data on PhD career outcomes. While a notable number of universities are not currently engaged in this work, most of them still believe it is important to do, and many would do so if resources were available. We know that while there are a number of units across the university with an interest in centralized data collection, the main locus of responsibility for this work in many universities remains in the graduate school.

With respect to how data are collected, we found a wide range of practices on both the methodology and the periodicity of data collection. There is also a wide range of satisfaction among the deans on current practices. While a slight majority of deans are satisfied or very satisfied with the efforts underway, many deans feel there is significant room for improvement, and many provided specific information on what needs to be improved.

Finally, our work so far revealed three areas that, while not strong themes in our 2017 workshop, clearly require attention in 2019. The first is a host of challenges associated with documenting career outcomes for our graduates who pursue their careers outside of the United States. These include both international students who return to their home countries to pursue their careers and domestic students who elect to work abroad. The second challenge is associated with the relatively low survey response rates and web scraping locating rates that most institutions realize. Low rates of successful data collection raise the prospect of nonresponse bias, which occurs when nonrespondents differ significantly from respondents on career pathways outcomes, such that inferences about the target population of doctorate recipients based on the data from respondents are inaccurate and potentially misleading with respect to practical implications. The possibility of nonresponse bias must be addressed in order to improve confidence in any inferences for program improvement based upon the data gathered. The third challenge is to develop career outcome information systems that ultimately lend themselves to meaningful benchmarking. Universities inevitably will want to compare the outcomes of their graduates to the outcomes of particular peer comparison groups that are meaningful to them. But in order to do this there needs to be a strategy in place early on to coordinate and standardize measures of the data elements upon which benchmarking would be most valuable and, equally important, to identify the meaningful peer comparison group.

This paper has summarized findings around tracking doctoral career pathways from a survey of graduate deans and focus groups of deans. The May 2019 workshop will enable graduate deans, in collaborations with researchers, to discuss further these findings and to begin developing aspirational guidelines for collecting and using the data. The workshop is designed to create a space where those individuals on our campuses responsible for thinking deeply about PhDs careers will come together to share their thoughts, their questions, their concerns, their hopes, and their aspirations. Our final report will build on this working paper to outline the emerging aspirational guidelines and a plan for piloting a system of implementing them.

REFERENCES

Blank, R., Daniels, R.J., Gilliland, G., Gutmann, A., Hawgood, S., Hrabowski, F.A., Pollack, M.E., Price, V., Reif, L.R., & Schlissel, M.S. (2017). A new data effort to inform career choices in biomedicine. *Science* 358 (6369), 1388-1389.

<http://science.sciencemag.org/content/358/6369/1388/tab-pdf>

<<http://science.sciencemag.org/content/358/6369/1388/tab-pdf>>

Bradburn, N.M. (2017). *Thoughts on the future of graduate education*.

Chicago: NORC. https://reports.norc.org/white_paper/thoughts-on-the-future-of-graduate-education/ <https://reports.norc.org/white_paper/thoughts-on-the-future-of-graduate-education/>

Council of Graduate Schools & Educational Testing Service. (2012). *Pathways through graduate school and into careers. Report from the Commission on Pathways through Graduate School and into Careers*. Princeton, NJ: Educational Testing Service.

National Academies of Science, Engineering, and Medicine. (2018a). *Graduate STEM education for the 21st century*. Washington, DC: National Academies Press. <https://doi.org/10.17226/25038> <<https://doi.org/10.17226/25038>>

National Academies of Science, Engineering, and Medicine. (2018b). *Measuring the 21st century science and engineering workforce population: Evolving needs*. Washington, DC: National Academies Press. <https://doi.org/10.17226/24968> <<https://doi.org/10.17226/24968>>

National Science Foundation, National Center for Science and Engineering Statistics. (2018). *Doctorate recipients from U.S. universities: 2016*. Washington, DC: National Science Foundation. <https://www.nsf.gov/statistics/sed/2018/nsf18304/> <<https://www.nsf.gov/statistics/sed/2018/nsf18304/>>

Silva, E.A., Mejía, A.B., & Watkins, E.S. (2019). Where do our graduates go? A toolkit for retrospective and ongoing career outcomes data collection for biomedical PhD students and postdoctoral scholars. <https://www.biorxiv.org/content/biorxiv/early/2019/02/11/539031> <<https://www.biorxiv.org/content/biorxiv/early/2019/02/11/539031>>

Stewart, D., & Hoffer, T. (2017). *Career pathways for graduate degree holders: What do we know and where do we go from here*. Chicago: NORC. https://reports.norc.org/white_paper/career-pathways-for-graduate-degree-holders-what-do-we-know-and-where-do-we-go-from-here/ <https://reports.norc.org/white_paper/career-pathways-for-graduate-degree-holders-what-do-we-know-and-where-do-we-go-from-here/>



APPENDIX

Ideas for How to Improve Tracking Efforts

Group 1. Institutional support: increasing staffing support, data collection project organization including greater centralized direction, use incentives for graduates to respond and departments to collect, and more resources (n=46)

- Assign a specific staff member to this role, and have this be their primary responsibility
- Have other units collect information more systematically
- More resources
- Centralized data base updates – moving toward this
- Engage all graduate programs to use the data
- They should become a resourced (e.g., funded) University Effort that includes the Alumni office, Graduate College and individual programs working as a team
- Dedicated staff to undertake and accomplish this effort on a routine basis
- The ability to hire staff who can focus on this as a major job responsibility
- Dedicated resources and staffing would ensure that this task occurred
- Needs to be assigned to one specific office and resources need to be deployed to handle work load
- Create a national database
- Better coordination between central and local efforts

- Having in house system (that does not require significant human capital) to update annually rather than having to rely on outside vendors
- Provide incentives to departments to track
- Centralized instead of decentralized
- Increased and more streamlined communication between systems of record
- Better internal system for gathering contact information
- Centralized system and frequent communications
- Dedicated FTE or GA support, better integration with graduate programs and administrative units, improved data systems and IT infrastructure
- More collaboration amongst units throughout the university
- Centralized system at the university coordinated with the alumni association to track
- A more centralized approach, but all the accreditation and external organizational reporting requirements make it near impossible as each needs slightly different data, time-frames, etc.
- Increased resources – person-power, technology

- We are very resource limited, primarily in terms of staff time that can be devoted to the tasks.
- More staff, more visibility, could use more uniformity
- More resources
- This would best be centralized in IR rather than relying on individual programs to track their own graduates
- Better staff support
- Systematic approach headed by our IR office
- Centralization
- Make it a priority and provide necessary resources
- Resources devoted to concerted effort in collaboration with departments
- If every program would track their own graduates; they would get far better responses as students tend to be loyal to their individual programs
- Centralized guidelines and system for collecting information
- It needs to be centralized
- Better coordination between decentralized and centralized efforts. Also need more personnel to assist with effort

- More resources for more staffing or accessible and affordable subscription services
- We need to get all units to participate, we need central support (staff time) of these efforts
- Funds for hiring full time staff to collect data continually
- More centralized collection/sharing of data
- By incentives. Incentive to academic programs and incentives to survey recipients
- More resources to track, better database management, more refined tracking questions (Groups 1 and 2)
- Additional resources to have dedicated staff to support effort. Better means to track alumni who are outside the U.S. (Groups 1 and 2)
- More thorough efforts, longitudinal data, help from alumni and career services (Groups 1 and 2)
- Dedicated staff for analysis; better integration into data warehouse (Groups 1 and 4)
- Needs to be more systematic – regular schedule, processes, use of information (Groups 1 and 4)
- Ongoing contact with alumni (Groups 1 and 4)

**Group 2. Study design: longitudinal vs. cross-sectional,
periodicity of data collection, question content (n=21)**

- Expanding beyond email addresses
- Longitudinal data collection
- A longitudinal approach
- Developing institution-wide protocol and sampling paradigm and involve Institutional Research
- More web scraping tools to cut down on individual time doing web research; more participation in and complete data from the graduate alumni survey (beyond our control)
- Automated collection; deeper understanding of the relationship between doctoral training and career path; better data on non-professional outcomes and achievements; satisfaction with training – “would do it again” issue; genuinely comparable peer data
- Tracking students for a longer period of time to gather career trajectory. We currently track them for 5 years
- Having a process by which the Graduate School will collect data every 2-3 years
- It would be great if the alumni office, which is supposedly the custodian of this information, responded to my inquiries. I collect information just for my college
- Continued tracking beyond the first survey

- More directed tracking – not institutional research questions
- We don't centrally track career outcomes in Graduate Division. Efforts are uneven at the program or school/college level. So there is much room for improvement and I believe much interest in and utility in pursuing a systematic method of career tracking
- More use of online information rather than relying on faculty member for information about graduate career pathways
- By having an online repository
- Institutional support & infrastructure for tracking, including a staff position and support from systems and data administrators
- More resources to track, better database management, more refined tracking questions (Groups 1 and 2)
- Additional resources to have dedicated staff to support effort. Better means to track alumni who are outside the U.S. (Groups 1 and 2)
- More thorough efforts, longitudinal data, help from alumni and career services (Groups 1 and 2)
- Better external data sources; Better internal tracking, outreach, and survey strategy (Groups 2 and 3)
- Systematic data collection and analysis (Groups 2 and 4)

- Connect to more programs, do more systematic analysis, additional follow-up with doctoral graduates, implement longitudinal study, more web-based research on current status of graduates (Groups 2 and 4)

Group 3. Tracking: Improving the locating of graduates, tracking non-U.S-resident graduates (n=9)

- Setting expectations of students earlier that they will provide; more staffing to devote time to the task
- Alumni office does not have much information on our grad school alums. Fairly common problem
- More consistent effort to update as people move
- Need to work better with alumni relations
- We will expand upon what we are already doing as part of the CGS career tracking grant and using LinkedIn to locate people
- Better updates of addresses
- Better records and system for updating records (e.g., email addresses missing for large proportion of recent graduates)
- Retention of contact with recipients on a regular basis
- Better external data sources; Better internal tracking, outreach, and survey strategy (Groups 2 and 3)

Group 4. Analysis and reporting: Make analysis more systematic, integrate data into data warehouse (n=5)

- Dedicated staff for analysis; better integration into data warehouse (Groups 1 and 4)
- Needs to be more systematic – regular schedule, processes, use of information (Groups 1 and 4)
- Ongoing contact with alumni (Groups 1 and 4)
- Systematic data collection and analysis. We are great on the research side (Groups 2 and 4)
- Connect to more programs, do more systematic analysis, additional follow-up with doctoral graduates, implement longitudinal study, more web-based research on current status of grads (Groups 2 and 4)

About The Higher Education Analytics Center

The Higher Education Analytics Center at NORC leverages our more than 50-year commitment to and extensive experience with higher education to offer effective, affordable research and data collection to institutions of higher education and other related organizations. Specifically, we deliver the data and insights necessary for these institutions' continued

functioning and improvement. We approach all work with deep technical expertise, a spirit of collaboration, and a commitment to scientific integrity. Our services range from large-scale cross-sectional and longitudinal data collection and analyses, to individualized consulting on discrete issues that can be addressed by data analytics. We also provide consultation on evaluation methods and data collection design. Learn more at **heac.norc.org** <<http://www.norc.org/about/departments/pages/higher-education-analytics-center.aspx>>

About NORC

NORC at the University of Chicago is an objective and non-partisan research institution that delivers reliable data and rigorous analysis to guide critical programmatic, business, and policy decisions. Since 1941, NORC has conducted groundbreaking studies, created and applied innovative methods and tools, and advanced principles of scientific integrity and collaboration. Today, government, corporate, and nonprofit clients around the world partner with NORC to transform increasingly complex information into useful knowledge. **www.norc.org**

<<http://www.norc.org>>

© NORC 2019

