An Initiative of NORC AT THE UNIVERSITY OF CHICAGO

WHITE PAPER

Identifying Promising Practices in University-Based Monitoring of Doctoral Career Pathways

October 8, 2019

Thomas B. 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

ACKNOWLEDGEMENTS

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.

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 were to help develop widely shared standards for the kinds of career outcomes that should be measured, the methods for collecting the data, and how the data should be analyzed and the findings disseminated to and utilized by prospective students, faculty, and administrators.

The project had 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 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 the 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. The workshop was held in Chicago on May 28 and 29, 2019, at NORC and was attended by 35 graduate deans and research experts. In advance of the workshop, the NORC project team developed a working paper providing details on the survey findings and identifying several discussion questions for the workshop to address (Hoffer, Stewart, Bradburn, & Knepler, 2019). This white paper summarizes and synthesizes the findings from the survey and focus groups along with the principal discoveries from the workshop.

Background

Many factors contribute to a growing interest in doctoral graduate career pathways. Career trajectories for PhD holders typically follow 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 have 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 the number of 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 of those in academe, only 61 percent, 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, some 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. Several 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 the 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 had a data collection process of any kind. 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 (CGS Pathways Project). 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 humanities fields, 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 results promise to better inform decision-making at both the institutional and program levels for those institutions participating. Since many universities were also invited to participate as affiliate members of the CGS Pathways 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 (CNGLS), 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" (Blank et al., 2017, p. 1389). 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 consist 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 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 five and 10 years after graduation. IRIS analyzes those outcomes using 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 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 (Stewart & Hoffer, 2017). 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?
- What is the role of federal datasets in facilitating institutional understanding of the career paths of graduates?

Motivated by these questions, NORC initiated the phase of work described in this white paper: an assessment of the current status of career pathways data collection in a large set of research PhD-granting universities, followed by a workshop convened to discuss and refine the findings. This effort aimed to advance, and, to the extent possible, articulate widely shared "aspirational standards." The hope was that such standards could address: 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 the proximate stakeholders in quality graduate degree programs: deans, department heads, faculty, and students. Stewart and Hoffer (2017) authored a working paper to capture observations and conclusions of those workshop participants.

The observations and conclusions, especially from the university administrators who attended, pointed to a 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 outcomes data emerged from the 2017 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 would need 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 would be 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 sustainable way. 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. Several deans mentioned the special challenges of tracking career pathways of international alumni (Stewart & Hoffer, 2017).

CURRENT LANDSCAPE OF PATHWAYS RESEARCH: FINDINGS FROM THE FALL 2018 DOCTORATE RECIPIENT CAREER PATHWAYS SURVEY

The 2017 workshop participants communicated that, notwithstanding the barriers and especially the very challenging technical ones, program-based and institution-wide data on outcomes of program graduates at different career stages were needed by applicants to make informed choices and by faculty to improve programs. We recognized that some institutions had been collecting career pathways data for periods ranging from one to four years. But we also knew that a comprehensive national portrait of current practices across a wide range of universities was lacking. We believed that a systematic assessment of what was currently being done, and not done, was prerequisite to determining the optimal ways for institutions to acquire career pathways information going forward. Specifically, the field needed more information on 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. 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 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?

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.

Extent of Pathways Data Collection by the Graduate Institutions

The primary question posed in the survey was whether the institution is currently directing an institution-wide, centralized effort to collect data on graduates' career pathways. Of 175 institutions responding to this primary question, 73 (42 percent) indicated that they are collecting data in this way.(2) The nuanced responses to whether an institution is collecting career pathways data are shown in Table 1.

Table 1. 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%
e. No, neither now nor in the past	18	10.3%

Method of Collecting Career Pathways Data	Frequency	Percent
f. No, not on an institution-wide basis with primarily centralized direction (but no additional information available)	29	16.6%

Methods of Collecting Pathways Data

Focusing on 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 2).

Table 2. Methods for Tracking Doctorate Recipients' Careers with Centralized Collection (n=73)

Method	Frequency	Percent
Surveys of graduates	43	72.9%

Method	Frequency	Percent
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 methods 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.

Our survey found over 80 percent of the 43 survey respondents who indicated 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.

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. As noted above concerning sampling, here we are unable to distinguish between institutions whose engagement is prescribed 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 every year after the CGS project ends. Some of the focus group comments suggested that continuation for some institutions is problematic.

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 pathways 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.

Role of Federal Data in Facilitating Institutional Understanding

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 undergone a positive change 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 Survey of Earned Doctorates. While the SDR data still cannot represent particular doctorate-granting institutions, the survey 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.

In sum, the survey and focus groups served to highlight a number of fundamental ways in which graduate institutions and programs vary with respect to monitoring graduates' career pathways. First, it is important to note that only about half of the graduate institutions are systematically collecting data on career pathways in a comprehensive institution-wide fashion, and only 30 to 40 percent are doing so on a centrally organized and directed basis. While many of those not collecting data institution-wide have at least some programs collecting data, many are largely in the dark about what becomes of their doctorate recipients beyond graduate school, even while recognizing the need for such data. Second, the institutions that are collecting institution-wide data with a centralized system vary with respect to the use of surveys and web scraping, periodicity of data collection and collecting cross-sectional versus longitudinal records. As for reporting the data collected, most disseminate results to institution and program leadership, faculty, and students. Benchmarking of results to national or peer-institution standards is valued by the graduate deans but is to date sharply limited by the availability of comparable external data and resources to process the data for the desired comparisons.

CHALLENGE OF ARTICULATING ASPIRATIONAL STANDARDS AND A PATH FORWARD

The May 2019 workshop with graduate deans and researchers aimed to identify and contribute to building an emerging consensus on what data to collect, how to collect them, and how best to analyze and disseminate the data collected. To achieve this goal, the workshop was organized into five panels, each topic addressing an essential element of a functioning system of career pathways tracking. From this discussion, we sought to extract guidance supporting the articulation of clear aspirational standards for a career pathways data collection and utilization system. While the discussion was rich and deep, it did not yield the kind of clear consensus required to support a statement of standards at this point. But the discussion did serve to refine our understanding of where the most difficult "pressure points" reside and illuminated seven domains of uncertainty that have to be clarified before a standard of best practice can be advanced.

The domains of uncertainty revealed through the survey and the workshop participants' reflections on the findings are the following: deciding what data to collect, how to collect it, how often to collect it and from which graduates, how to classify the career data, how to manage the data collected, how to analyze and disseminate the data, and how to obtain comparative data on graduates from other universities to benchmark against one's own graduates. This section maps out the specific questions and issues that give rise to the "pressure points" and which must be addressed as a foundation for a broadly applicable and sustainable system for tracking and utilizing career pathways data.

Content of Data Collection

There are diverse views on the content that graduate deans want in any data collection efforts. In general, workshop participants expressed the view that, whatever the specific content of the data collected, it should be locally relevant, accessible, actionable, timely, and comparable across units

within the university. In other words, they believed that the content should be shaped by the audiences and purposes. All data should be collected to achieve a specific purpose or to address a specific audience, while recognizing standardization of some core career pathway variables is essential to enable comparisons across peer universities or with national benchmarks.

For example, workshop participants agreed that the basic information captured by both the CGS Pathways Project and the CNGLS protocol on graduate's employment sector, occupation, and main work activities are essential. The CGS Pathways Project establishes specific survey questions about occupations and work activities that all partners must ask of graduates, and the CNGLS protocol establishes a taxonomic system for classifying information collected on graduates' occupations. Participants also agreed that graduates' assessments of how closely related their work is to their doctoral field of study, and how their degree program might have better prepared them for their postgraduation work – both of which are captured by the CGS Pathways Project — are informative but are more difficult to collect well since they require surveys and representative respondents. The one main area of uncertainty appears to be how deeply to probe graduates for retrospective evaluations of how well graduate school prepared them for current work opportunities. Responses to those questions require some elaboration and context in order to provide clear guidance and thus place greater demands on a survey-based system, raising respondent burden and survey administrative costs.

Data Collection Methods

The three main methods of collecting data are surveys, online data abstraction, and linking student records to administrative datasets. Each has important limitations and shortcomings. The main questions here relate to the optimal mix of methods, particularly of surveys and online

abstraction, and to how to obtain adequate coverage of the target populations of graduates via any of the methods.

The emerging consensus from the NORC 2019 workshop was that surveys of graduates provide the most accurate information on employment sector, occupation, and work activities, and are the only method for obtaining attitudinal data and evaluation feedback on how graduate school prepared or failed to prepare them for their career pathway. The primary weakness of survey methods is that many graduates do not complete the survey, and sustained efforts to locate nonrespondents and gain their cooperation are often time-consuming and expensive.

Online data abstraction can be relatively fast and inexpensive and often accurately captures sector and occupation. Unfortunately, online abstraction can have significant error because many graduates do not regularly update their records. Detailed online information on work activities is often not available or is edited to present the graduate in particular ways for particular purposes, introducing presentation bias and, to that extent, may be inaccurate. Further, many graduates do not maintain any online employment information and are thus missed completely by the online data abstraction methodology. One dean participating in the workshop summed it up this way: “In collecting career pathway information I want to ask what I want to know. But with web-scraping you can’t ask what is not there!” Finally, there is the problem of “binning” or classifying the information that is found. Workshop participants agreed that while there is sometimes general agreement on how to classify job information when abstracting from a source like LinkedIn, often the information provided is ambiguous and defies confident classification. All these limitations notwithstanding, online data abstraction is still recommended as a complement to surveys and particularly as a means of obtaining career pathways information about those who do not complete the survey.

Periodicity of Data Collection and Sampling Graduates

Collecting, analyzing, and reporting data on graduate pathways takes time and resources from the university, and responding to a survey represents some level of burden on the cooperating graduates. A key question is how often the data should be collected to achieve the informational needs of program improvement and transparency for students. The question is more complicated than it appears at first glance because, while many informational needs do not require annual updates to convey the important points, annual updates are likely to result in more accurate information on graduates, who often make frequent job and place of residence changes. Some deans participating in the workshop emphasized that annual data collection was ultimately better because only in this way would it become part of the standard operating procedure at an institution, deeply integrated into the normal data collection and dissemination process. Other deans noted that for very large institutions, some with over 50,000 living graduates, comprehensive annual data collection may simply not be feasible. These deans suggested that a carefully drawn sample of graduates could provide the trend data needed to inform students and programs. The modal position of the participants seemed to be that an institution would update locating information on all graduates out to some limit of years since receipt of the PhD, but it would collect occupation and other pathways information from only a select subset of those graduates (e.g., those who are three, 8, and 15 years since receipt of the PhD) and/or would collect occupation and other pathways data less frequently than the annual location data.

The optimal periodicity of collecting pathways data has not yet been resolved into anything like a consensus at this point. There is a general recognition among workshop participants that the further out the lower the response rate, especially at 15 years and beyond. The view was also expressed that, in terms of both length of tracking period and comprehensiveness of the coverage, the ideal should not become the enemy of the good for individual institutions. One dean

in the workshop summarized the view of many when he commented that, however the periodicity issue is resolved, the important thing is to establish some system: “We need to make some data collection part of the standard operating procedure for our universities, so that you can’t chose not to do it.”

The SDR is the one national survey that does achieve very high response rates decades post-graduation. It is administered to a large sample of all STEM doctorate recipients from U.S. institutions every two years and does extensive advance work before each round to update contacting information. The CGS surveys are administered to graduates three, 8, and 15 years since receipt of the PhD. Locating the graduates targeted to receive the CGS survey is a task the partner institutions individually assume and presumably accomplish with different levels of success. An open question is how useful are annual snapshots of graduates at the three time points and whether less frequent snapshots would be adequate. The model for the NGLS recommended by Silva, Meija, and Watkins (2019) advocates for annual data collection, primarily to maintain current contacting information and with minimal respondent burden by using a very brief (four-item) survey and relying heavily on gathering data through online data abstraction.

Taxonomies for Variables Related to Career Pathways

Information on career pathways variables such as sector, occupation, and work activities can take a wide variety of forms. To make sense of the data, they must be located within some sort of standard schema or taxonomy encompassing the full range of responses. And to make valid comparisons across programs and institutions, the taxonomies must be shared and used by all.

With respect to career pathways variables, the NSF's National Center for Science and Engineering Statistics has developed and refined taxonomies on employment sectors, occupations, and work activities that are designed to represent all doctoral career paths, including both STEM and non-STEM. The CGS surveys borrow SDR items for sectors, occupations, and work activities, but the coding operations needed to translate responses into the SDR taxonomies are not standardized across partners, and the coding decisions are left to the partners to resolve. The NGLS coalition has made strong efforts to develop standard taxonomies for sector and occupation and is developing guide materials detailing coding decisions to firmly establish comparability of data across institutions, but only for the institutions in their purview: life sciences. The current situation concerning the use of taxonomies in classifying career pathways information is that, while several institutions are using taxonomies in classifying information acquired either from surveys or from cyber scraping, there is no consensus on the right taxonomy. The NGLS taxonomy is used by many institutions for the biomedical sciences, but some institutions are using multiple taxonomies each for different purposes.

Database Management

The large and ever-increasing scope of the pathways information institutions are collecting poses challenges of documenting and maintaining records on graduates, including who has been located and contacted, and who has provided data or had information online and when. There are also significant challenges of storing the survey and online abstracted data in ways that make retrieval efficient for case management during data collection and for analysis purposes. But this discussion of how to manage and store career pathways data stimulated some of the participating deans to call for a wholly integrated system of information that covers the lifecycle of a graduate student from initiation as a student through graduation and into and through their careers.

Analysis and Dissemination of Data

As the pathways data collection efforts by most institutions are in early stages, methods of data analysis and dissemination have not yet been standardized across institutions and typically are not yet articulated at all. The NGLS has gone furthest in these respects, providing specific reporting formats that coalition members should use. CGS conducts some analysis of aggregated data provided by the partners, and those reporting formats can provide models for the partners' work, but there are not recommended or required reporting formats analogous to the NGLS guidelines.

Recognizing that, for most universities, analysis and dissemination are both in the early stages of development, workshop participants converged on two points. First, there was strong support for the view that audience must be a central consideration in developing an analysis and dissemination plan. Deans argued that a robust plan should address internal and external audiences, including the president, provost, fellow deans, and broader publics in addition to current and future students as well as program faculty. In the words of one participant, "…effective dissemination of career outcomes data must be accompanied by an appropriate narrative that fits the target audience." Second, dean participants agreed that whoever the audience might be, the data must be made to "come alive" for that target audience, with well-developed data visualizations and succinct narratives of the main points.

Using Comparative Benchmarks to Help Interpret the Data

Throughout the workshop, participants discussed different kinds of benchmarking activity, categorized by one dean as performance, comparative, and evaluative benchmarking. These referred to comparing a degree program's current performance against prior performance

(performance benchmarking), comparing all programs within a university on a fixed set of indicators (comparative benchmarking), and comparing across peer universities on a selected group of indicators (evaluative benchmarking). An important function of collecting pathways data is to compare programs and institutions with peer group and national aggregations, i.e., evaluative benchmarking. Evaluative or peer benchmarking was deemed important and valuable by the deans present though some participants contested its feasibility. The national benchmarks can, in most cases, be obtained from analysis of the SDR data, provided the institution has access and analysis resources to develop the comparisons. Peer group comparisons are potentially more complicated because they require standardization of methods and measures among the peer group, and they require a peer agreement to share data for comparison purposes.

NEXT STEPS

This paper has summarized findings around tracking doctoral career pathways from a survey of graduate deans, focus groups of deans, and a convening of graduate education leaders and researchers to reflect together on the findings. While the intention was to establish a set of aspirational guidelines for tracking career pathways information and utilizing it, the experience in the field did not support achieving this goal at this time. Participants voiced enthusiasm for expert guidance to inform the development of a model system for collecting and utilizing data on career paths of PhD recipients. However, substantial uncertainties were indicated by the deans around seven critical dimensions of any system that might be developed. This white paper summarizes emergent recommendations on each of those dimensions. The next step in this

research is to resolve these “domains of uncertainty” in the process of building a prototype system that can be rigorously tested across a variety of institutions. The information and understanding generated in this NSF-supported project provides the foundation for that next phase of work.

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.

Hoffer, T.B., Stewart, D., Bradburn, N.M. & Knepler, E. (2019). *Progress and Pitfalls in Tracking U.S. Doctoral Career Paths.* Chicago: NORC. https://reports.norc.org/white_paper/progress-and-pitfalls-in-tracking-u-s-doctoral-career-paths/ <https://reports.norc.org/white_paper/progress-and-pitfalls-in-tracking-u-s-doctoral-career-paths/>

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/>

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
at the
University of
Chicago