INSTITUTE FOR RESEARCH ON INNOVATION & SCIENCE

IRIS

IRIS University Employee Profile Report

Many universities are interested in characterizing the work of research-trained employees when those employees leave their universities. Do these employees work in the education sector, or in the private sector? What are their average earnings one year after leaving the university, or two years, or five years? In what US states do these employees work? The Employee Profile report for your university answers these questions.

Georgraphic Destinations

- Roughly **3,383** research-trained employees at IRIS University found subsequent employment in **49 states and the District of Columbia** between **2009** and **2015 (the most recent year available through federal data)** .
- Approximately **47%** of these research-trained employees at IRIS University found employment in the state of FL.

IRIS University employees find employment in states across the U.S.

Top employee destinations by state with employee count

- Florida (1,600)
- Georgia (1,100)
- California (400)
- Ohio (200)
- Colorado (150)
- Massachusetts (150)
- New York (150)
- Illinois (100)
- Maryland (100)
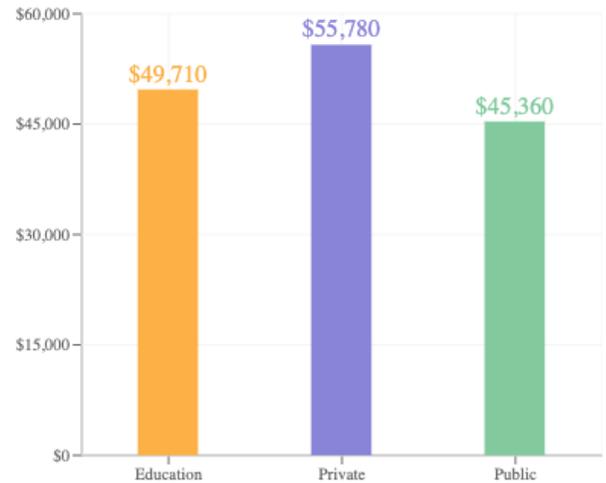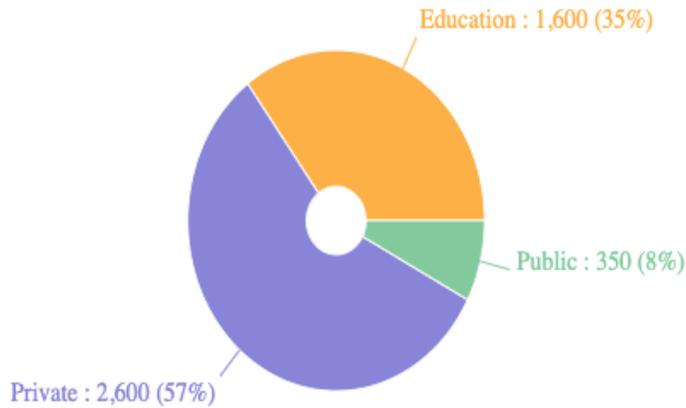- North Carolina (100)

# of Employees staying in-state: 1,600

Employee Count
- 14
- 20 to 40
- 50 to 80
- 100 to 1,100
- 1,600

United States Census Bureau
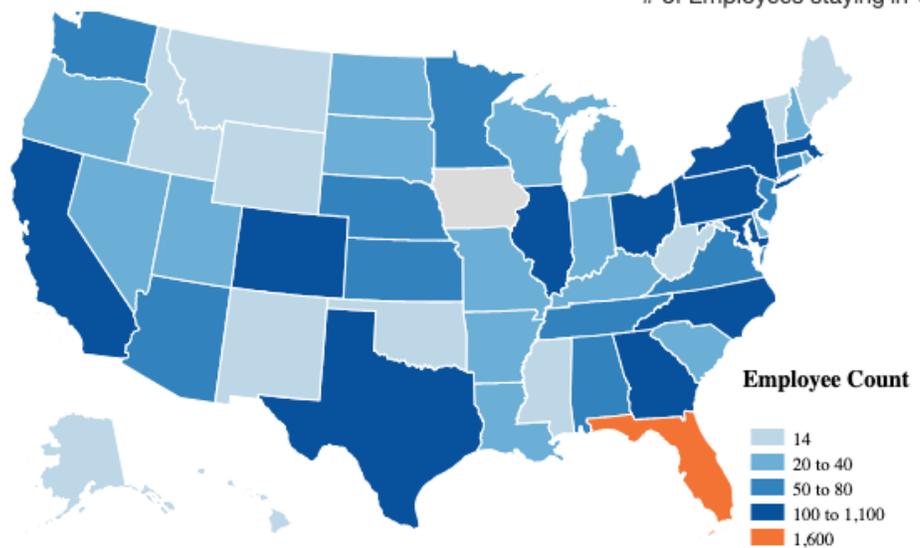
IRIS University Employee Profile Report

Sectors and Earnings

In what sectors do your research-trained individuals work and how much do they earn?

- One year after departing IRIS University, roughly **1,430** research-trained graduate students found employment across the U.S.
- Within one year after leaving IRIS University, **650** research-trained graduate students go on to find employment in the private sector and earn, on average, **$69.9 thousand** annually.
- Within one year after leaving IRIS University, **550** research-trained graduate students go on to find employment in the education sector and earn, on average, **$51 thousand** annually.

Methodology

General information about the process

The Employee Profile report is built on the employee data your institution provides. Job titles, pay dates, and personally identifiable information (PII data) for employees paid by federal and other grants are submitted to IRIS by your university. IRIS applies a unique employee ID to every university employee. IRIS also rolls up job titles into broad job categories, including graduate student and postdoctoral researcher as separate categories. IRIS then sends employee PII data, as well as these broad job categories for each employee, to the US Census Bureau. Payroll dates are used to select the time window for which employee transactions can be matched to federal data. Currently, this window is 2005 through 2017; additional years will be added in the future.

At the Census Bureau, PII data are used to assign each employee a unique individual identifier called the Protected Identification Key(PIK).This identifier in turn is used to match the employee to their yearly earnings, as recorded in their W2s for that year and in two other databases, the Longitudinal Employer - Household Dynamics(LEHD) dataset, which contains information on Unemployment Insurance - eligible jobs in 49 states, and the integrated Longitudinal Business Database, which contains information on self - employed individuals.In addition to providing earnings information, data from the LEHD are used to provide geocoding.North American Industrial Classification System(NAICS) codes in the Longitudinal Business Directory, Business Register, and LEHD are used to categorize each firm - level employer as belonging to the private, public, or education sector.In addition, Federal Office of Personnel Management data were used to improve the categorization of employees in the public sector.

Disclosure review

Once all matches have been made and geographic and earnings information has been calculated, results are rigorously screened by the Census Bureau in a process called "Disclosure Review." Disclosure review is intended to prevent identification of a particular individual or organization from released information. In the employee data, this means that an employee earnings measure can only be released if the number of cases is high enough that no single person can be identified. Disclosure review also involves comparison of successive reports, so that a small change in the number of cases from one report to the next does not result in the identification of any one individual. Applying these tests can sometimes mean that certain data points must be excluded from your reports. Disclosure review of results from analyses of restricted Census data is required by federal law (Titles 13 and 26, U.S.C.). The data in this report were merged with federal data in December 2017 and passed disclosure requirements through the Census Disclosure Review Board receiving the approval number CBDRB-FY18-411.

Geographic Destinations

The Geographic Destinations tab shows a map of the United States and a list of the top states in which your research-trained employees work after leaving your university. This report is summed across cohorts and across calendar years, so that any employee moving to a given state is counted in the total for that state, regardless of the year (from 2005 to 2017) that the move took place. The logarithmic scale used for the maps improves the shading of state destinations making the map more readable.

Sector and Earnings

The Sectors and Earnings tab displays the average yearly earnings for certain groups of research-trained employees in the years following their departure from your university. These earnings are displayed separately for each year after leaving the university. You can opt to view information separately for graduate students and for postdoctoral trainees, as well as for all employees. By selecting employee types and years, a user can see how average education, public, or private-sector earnings vary over time for specific groups of research-trained employees.

Methodology

Data in the employee sector and earning report is based on analytic cohorts, rather than calendar years.This means, for example, that a measure of earnings at one year after departure can be based on averaging individual records from a number of different calendar years; their common feature is that they measure one year after departure from the university.

The sector and earnings measures are best understood as a series of yearly cohort - based snapshots: one year after leaving may be assessed at different calendar years for different members of the group, and the group for one given year after leaving may contain a somewhat different mix of individual members than the group in another year.The earnings measures should also be regarded as approximate: each average is rounded to the nearest $100, and is based on earnings reported in real 2018 dollars.

Limitations

One current limitation of these reports results from the classification of each employee as working in a single industry and sector in each calendar year: the industry and sector that provides the greatest yearly earnings, controlling for partial-year employment, to that employee. While this rule allows us to account for higher-paying jobs where an employee only works for a small part of the year, it may lead to issues when an employee switches between the private and public sectors within the same year. That is, in cases where an employee has earnings in the public sector and greater earnings in the private sector in a given year, this rule sums all the employee's earnings and attributes them to the private sector. We are conducting research with collaborators with Census to develop appropriate methods to address this limitation.

A second limitation is that a small number of employment records in our data sources include earnings but no NAICs codes, and so cannot be classified as to their sector.Earnings for these employees are not included in these calculations.

Wage Cohorts

To create the wage cohorts shown in your Employee Sectors and Earnings tab, average earnings are aggregated at specific amounts of elapsed time since employees left employment at your university. For example, to make the "one year" cohort for your university, IRIS adds one year to the most recent year in which each employee worked at your university prior to leaving employment there.We then collect and average wages at that one-year mark. This measure aggregates over several calendar years, because employees leave their university over a number of calendar years, but the elapsed time since being an employee at the university is kept the same for each cohort. The attached diagram displays how these cohorts are created for a sample set of calendar years.
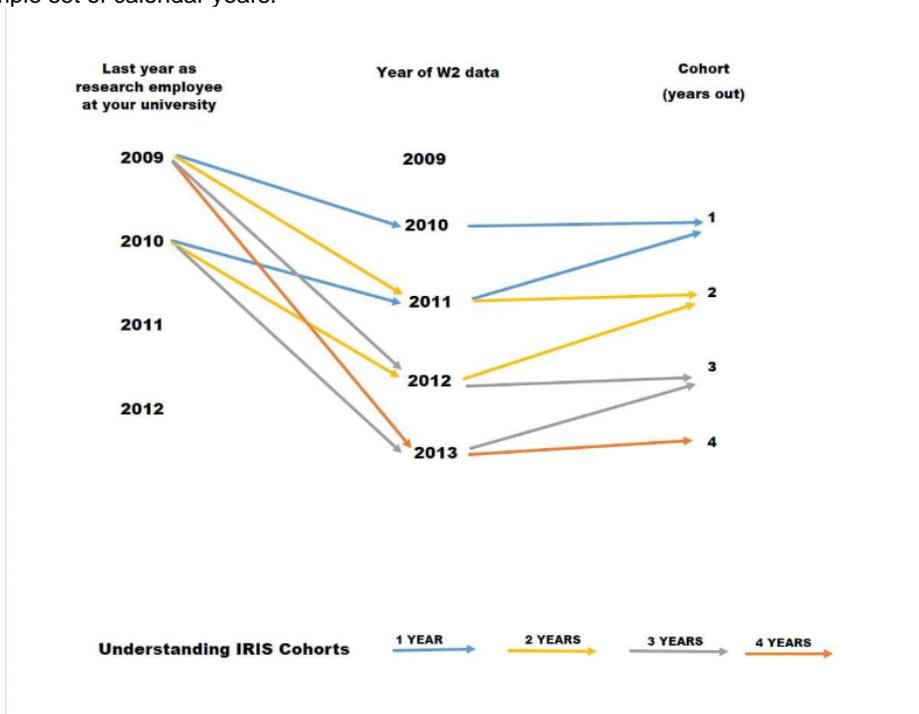
Matching Rates & Process for IRIS University Hot Report Data

Overview

This memo presents details and outcomes of the matching process conducted by the U.S.Census Bureau to integrate your IRIS data, submitted on or before November 2017, with the Census data resources. The memo provides context for those matching outcomes, which we hope will enhance the interpretability and utility of the data presented in your university hotreports. We also highlight potential issues to be addressed as IRIS continues to develop products basedon linked IRIS-Census data. At each stage we highlight workthat IRIS, the Universities, and Census cando to improve the overall quality.

Employee Match

The QA reports are organized around two different views of the data. The first describes the process by which research employees at your university are matched to Census Data. There are three key steps inthis process:

1. Assigning a Census Protected Identification Key(PIK): Census Bureau data uses a unique individual identifier called the PIK. The matching process uses name, month and year of birth to identify unique PIKs for individuals employed in research on your campus. **IRIS University's PIK rate was 86.3%**.

2. Identifying "Leavers": The report we shared with you is based on earnings information for research employees who left your institution. In our process an individual's last year at a university is the last year they receive a W2 from the institution or the last year they appear on a UMETRICSgrant. **Of the employee names submitted by IRIS University and matched to Census PIKs, 46.45% left the university**.

3. Collecting Earnings Data: Earnings data reported here are based on W2 information for leavers. In cases where an individual departs after 2016 (the last year of Census data), we would capture their departure but cannot determine their earnings. Census is currently updating their W2 data to include 2017 information to account for this issue. **4,284 (87.02%) of IRIS University "leavers" have W2 earnings information. We are working with Census to identify means to improve this match rate**.

Action	Range of Data Maches	Steps to Improve the Data
University uploads data to IRIS Upload includes employee transactional file and employee name file	The data for this report is based on submissions as of **November 2017** # of unique employees found in the Employee Transactional file: **12,482** # of unique employees found in the Employee Name file: **12,278**	University: Make sure that every employee ID in the transactional file has a matching record in the name file. IRIS: Provide feedback on employee ID match rates between the transactional file and the name file. Provide feedback on row counts of data and unique employee IDs that are included in the data. Census: N/A
IRIS reformats the data to meet Census specifications	# of unique employee IDs in the transaction file: **12,482** # of unique employee IDs in the name file: **12,278** % match rate between the two files: **98.37%**	University: N/A IRIS: Make sure that the number of unique IDs in each file remains consistent and match ratio remains consistent with the original uploaded data. Census: N/A
Data transfers from IRIS to Census	No data transformations take place in this step No comparisons or matches are made	University: N/A IRIS: N/A Census: N/A
Employee PIK Matching at Census	# of matching employees to census data for PIK: **10,597** PIK(Protected Identification Key) used by Census to match records across different datasets. **Match rate: 86.3%**	University: Provide year of birth and month of birth in the original data uploads to IRIS to ensure the highest possible match rate. IRIS: Provide a reminder to universities that missing names, YOB, or MOB will significantly impact the match rate at this step. Census: Provide a quality assurance report of match rates between the university/IRIS supplied data and the data available at Census.
Positive hits in the Business Registry are then matched to the LBD (Longitudinal Business Database)	**Match rate: % of employees who left the university: 46.45%** **% of those leavers for whom there are wage data: 87.02%**	University: Provide as many years of data as possible to improve "leaver" tracking and provide better match opportunities for Census. IRIS: N/A Census: Explore additional sources of data to improve wage data matching.
Place identified "leavers" in categories to allow for reporting.	Minimum numbers of employees must make it to this step in order for categories to pass disclosure. Each category that does not contain enough employees to prevent reidentification will be suppressed in final disclosure.	University: Provide as many years of data as possible to improve "leaver" tracking and provide better match opportunities for Census. IRIS: Work with Census to develop disclosure proofed match reports. Census: Work with IRIS to develop disclosure proofed match reports.
Disclosure Review	Final disclosure of aggregate data by Census.	University: N/A IRIS: N/A Census: See the disclosure policies part below.
Populate Hot Report	Census sends aggregated data to IRIS for inclusion in Hot Report. As a result of this matching process, your final report is based on data from **4,284** employees.	University: N/A IRIS: N/A Census: N/A

Matching Rates & Process for IRIS University Hot Report Data

Ensuring Privacy and Confidentiality

Once all matches have been made, results generated from matched data are rigorously screened by the Census Bureau in a process called "Disclosure Review". Disclosure review procedures are designed to prevent re-identification of an individual or organization from the reports you receive. Disclosure review of results from analyses of restricted Census data is required by federal law. Some specific details of the disclosure review process are also protected by law and cannot be shared. With that said, the disclosure process influences the reports you receive in several ways.

1. Individual Data. Information on individual employees can only be released if it is based on a sufficient number of cases to insure that no single person can be identified in the data. Low match rates combined with fine categories in some panels of your reports mean data points cannot be disclosed.

2. Vendor and Subaward Data. Information on business establishments can only be released when it is based on a large enough number of establishments and when the concentration rate is low. In the kinds of spending data we report here, a concentration rate would be high if a small number of vendors accounted for a large percentage of spending in a particular category. Hypothetically, if your institution has a contract with an airline that means most travel expenses are charged to that vendor, then even if your employees use many airlines, we may not be able to disclose spending information because of high concentration rates. Spending in some NAICS codes cannot be reported because of these issues.

3. Secondary Disclosure and Implicit Samples. Effective disclosure review requires attention to the possibility that data in a single report, or data disclosed in more than one report can be used to infer information about individuals or organizations. For instance, if we were to issue an updated report that differed from the data in your prior reports by only one or two data points, comparing reports would allow you to identify information about those data points, compromising privacy and confidentiality.

Likewise, the size and (for businesses) concentration of unreported categories must be taken into account in disclosure review. If we break information about individuals into three categories and report data on two of them along with information about the total number of individuals on which the data are based, the size of the omitted category can be easily inferred. Concerns about secondary disclosure and implicit samples can sometimes lead to suppression of information for categories that would pass disclosure when reported alone. Suppose, for instance, we wish to report information on an industry, 1, and on its four sub industries 1.a, 1.b, 1.c, and 1.d. If 1.d is too small (or too concentrated) to pass disclosure, its suppression from the report will also require suppression of 1.a, 1.b, or 1.c because those three data points along with total information about 1 would allow information about the suppressed category (1.d) to be inferred. This is why effective privacy protection requires that the Census Bureau and IRIS also scrutinize contextual information such as the number of cases associated with a particular report or figure and the match rates for individuals and organizations.

IRIS is committed to responsible use of restricted data. We continue to work with our University partners and with the Census Bureau to produce themost detailed and valuable reports thatconscientious attentionto privacy and confidentiality concerns will allow.