Technical Memo – Aggregate 2-Page Report

Last edited: April 15, 2015

Overview

This technical memo accompanies the April 2015 UMETRICS aggregate university report titled "*Connected CIC: Report on the Regional Effects of Federal Research Funding*." The report documents aggregate federal research funding and expenditures at nine universities affiliated with the Committee on Institutional Cooperation (CIC), using UMETRICS data.[1]

The report includes data from the following universities:

University	**Time Period**
Indiana University	2012 Q3 – 2014 Q2
Pennsylvania State University	2012 Q3 – 2014 Q2
Purdue University	2012 Q3 – 2014 Q2
The Ohio State University	2012 Q3 – 2014 Q2
University of Chicago	2012 Q3 – 2014 Q2
University of Iowa	2012 Q3 – 2014 Q2
University of Michigan	2012 Q3 – 2014 Q2
University of Minnesota	2012 Q3 – 2014 Q2
University of Wisconsin – Madison	2012 Q3 – 2014 Q2

UMETRICS Data

The UMETRICS[2] program is an effort parallel to the federally supported STAR METRICS[3] program, developed in collaboration with 15 university members of the CIC. The UMETRICS data are provided by university members and describe the expenditures (overhead, employee, vendor and subaward/subcontract) made on federal awards. These data are extracted from administrative systems and then shared with UMETRICS researchers for reporting and data analysis.

See additional information on the individual data elements: http://www.cic.net/docs/default-source/umetrics/umetrics-data-dictionary.pdf?sfvrsn=4

[1] UMETRICS is a university-led initiative to build a scientific framework that will inform research management, enable evidence-based decision making, and support credible advocacy. Universities participating in the UMETRICS initiative submit quarterly micro-data on university payroll, vendor, subaward/subcontract, and overhead expenditures from federal and non-federal grants and contracts. The data submitted are transactional in format and aggregated for all analyses. In the rest of this report, subaward/subcontract data are referred to as subaward.

[2] https://www.cic.net/projects/umetrics

[3] Science and Technology for America's Reinvestment – Measuring the EffecTs of Research on Innovation, Competiveness, and Science. STAR METRICS was the initial effort to develop administrative data for use in reporting and research (https://starmetrics.nih.gov).

Initial Data Processing Steps

Identifying federal funding transactions. The UMETRICS transactional data include a federal agency code or a Catalog of Federal Domestic Assistance (CFDA) code for each transaction. The CFDA provides a full listing of all federal programs available to universities (and other types of organizations) and is captured in the UMETRICS data in order to filter federal award expenditures by federal funding agency. For these reports, only transactions associated with federal funding were kept. No other transactional data were included.

Selecting reporting time periods. The report covers a two-year time period representing the greatest overlap for the reporting institutions. The transactional data include both a period start date and a period end date; the aggregated data are based only on the start date of the period. No adjustment was made to account for periods that cross quarters or years for these reports; future reports will incorporate such adjustments.

National/Regional Distribution of Research-Related Expenditures

These analyses describe the research funding spent on purchasing supplies, equipment, technology, and travel as well as the research funding spent on collaboration with other research universities and private/public organizations. The aggregated data for the United States and individual state maps are from the UMETRICS vendor and subaward files, which are derived from university accounts payable, credit card accounts, and contracts accounts.

ZIP code extraction for expenditure aggregation. After filtering for federal research grants and selecting the relevant time period, the UMETRICS vendor and subaward transactional data files are combined to produce award expenditures.

The ZIP code information is derived as follows. The primary source is the field in the vendor and subaward transactions, which provides either a DUNS number, or, if that is not available, a ZIP code[4]. The second source of ZIP code data is the UMETRICS organization list, which contains the organization name and address for each institution identifier in the vendor and subaward transaction data files. The third data source is a precompiled crosswalk table of DUNS numbers and ZIP codes that was extracted from the System for Award Management (www.sam.gov), a government system for contracting entities. ZIP codes are assigned to these transactions, which are then mapped to U.S. counties by a ZIP code to Federal Information Processing Standard (FIPS) code crosswalk (see below). Multiple sources are used when assigning ZIP codes to transactions in order to achieve maximum data consistency: the transaction record, the UMETRICS organization list, and the DUNS number to ZIP code crosswalk. The three data sources are checked in order: the transaction record, the organization list, and the DUNS number crosswalk. If in any step a valid ZIP code is found, then it is assigned to the transaction and no further assignment is made. A valid ZIP code consists of 5 digits, optionally followed by a hyphen, and optionally followed by an additional 4 digits.

[4] A DUNS number is a unique nine-digit identification number issued by Dun & Bradstreet (D&B) that represents a unique physical location for a business. http://www.grants.gov/web/grants/applicants/organization-registration/step-1-obtain-duns-number.html

Matching ZIP codes to FIPS codes. To generate state and county maps, vendor and subaward transactions are aggregated by FIPS Code, a five-digit Federal Information Processing Standard code which uniquely identifies counties and county equivalents in the United States. The UMETRICS transaction data do not include a FIPS code variable. The team has created a ZIP code to FIPS code crosswalk based on data extracted from the CDC's County Cross Reference File[5]. This crosswalk is used to translate ZIP codes from the UMETRICS data to FIPS codes.

Foreign transactions in the university data are not included in this analysis.

The data are then aggregated by the FIPS code to generate total sum of expenditures by county for map generation. Aggregate expenditures less than $1 are dropped for these analyses.

University Workforce Analyses

Aggregating occupational categories. The UMETRICS employee transactional data include the occupation associated with the employee ID. Occupation data fields are submitted as the original university human resources (HR) titles or are aggregated by submitting university staff into the following 15 UMETRICS workforce categories: faculty, postdoctorate researcher, graduate student, undergraduate student, clinician, instructional, research admin, research analyst, research coordinator, research support, staff other, staff scientist, technical support, technician, or other[6]. Original university HR job titles were submitted by eight of the nine universities in this report. The eight sets of original university HR job titles were mapped to the 15 UMETRICS workforce categories by an affiliated research team at The Ohio State University. Their occupation data mapping approach is documented in the recent Research Policy Paper titled "New linked data on research investments: Scientific workforce, productivity, and public value."[7]

Organization of the Scientific Workforce

This analysis shows the occupational distribution of individuals supported by grants from eight federal agencies. The eight agencies chosen were those that support the greatest number of individuals at all nine universities in aggregate during Q3 2012 – Q2 2014.

The data are aggregated from the UMETRICS employee file, which is originally pulled from university payroll data. After filtering for federal grants and selecting the time period in question, the data are filtered by occupational classification. All transactions with an "other" occupation are dropped and are not included in either analysis. All occupations that are not designated as "faculty," "postdoctorate researcher," "undergraduate," or "graduate student" are combined into a new "staff" category. For employees who have multiple occupations during the two-year period, the occupation that the person held at the end of the period is chosen. If a unique person worked more than one occupation at the end of the period, one occupation is selected for reporting based on the following priority order: graduate student, undergraduate student, faculty, and postgraduate researcher. All other occupations are chosen in alphabetical order.

---

[5] http://wonder.cdc.gov/wonder/sci_data/codes/fips/type_txt/cntyxref.asp

[6] Conrad, C., Cheng, W., & Weinberg, B.A. (2014). *University occupation classification: Technical paper*. Columbus, Ohio.

[7] http://www.sciencedirect.com/science/article/pii/S0048733315000025

Individuals that worked on grants from multiple agencies during the reporting period are counted in all the relevant agency classifications. For instance, if a graduate student worked on a grant from the National Institutes of Health (NIH) and the National Science Foundation (NSF) in the same time period, they would be included as both NSF and NIH supported in the research workforce chart.