

ACADEMIC ANALYTICS

AAD 2019 Database Methodology

OCTOBER 2020



This document contains Academic Analytics' confidential and proprietary, business trade secrets. This document may not be transferred or used by any other person or entity other than your organization. For internal use only.

© 2020 Academic Analytics, LLC. All rights reserved.

Introduction

There are two levels of licensure to the Academic Analytics Database; Comparative/Details and Master License Agreement (MLA). Both of these modules provide different levels of access to the data and tools to uncover and analyze the data.

Content and Coverage

The Academic Analytics Database includes data on the fundamental areas of scholarly research activity and provides users with benchmarking and analysis capabilities. The database is organized by institutions, broad fields, departments, Ph.D. programs, and by individual faculty members. The database includes information on over 227,000 faculty members associated with more than 9,900 Ph.D. programs and 11,000 departments at 472 universities in the United States and abroad. The database is fully customizable, allowing users to create custom peer groups at different levels of analysis as well as custom groups of disciplines based on selected options.

The Academic Analytics Scholarly Research Index (SRI) was developed to express the level of research activity across disciplines. The Scholarly Research Index can be applied at different levels; the level of individual faculty, units (programs or departments) within a university, and the overall performance of universities. The Index uses metrics that are independent of discipline values and of the portfolio of disciplines at universities. The Index is matched to seven separate variables in six areas of activity:

- Total Publications
- Total Citations
- Total Books
- Total Grants
- Total Grant Dollars
- Total Awards
- Total Conference Proceedings

Comparative/Details Database

The Comparative/Details Database is the base level of licensure. Clients subscribing at this level receive two views of national scholarly activity, comparative data which act as a base or ground and details data which is an overlay on top of comparative data, providing enumerations of the base data. The Comparative Data are released annually and reflect scholarly activity in academic research for a given time window – a rolling range of years. The database includes metrics on professional honors and awards, federal grants, book publications, journal articles, conference proceedings, and citations. The coverage period of the current Comparative Data is:

Articles:	2016 - 2019
Citations:	2015 - 2019
Conference Proceedings:	2016 - 2019
Books:	2010 - 2019
Grants:	2015 - 2019
Awards:	No Limit - 2019

The Details Data include everything in the Comparative Data plus greater details regarding the individual pieces of scholarly activity linked to faculty in the Comparative Data. Unlike the Comparative Data, which provide a locked snapshot of activity for a given range of years, the Details Data are updated daily and contain activity outside of the most current Comparative Data time window. The Details Data may be accessed on the downloads screen of our online portal and in faculty-specific portal tools such as Quintiles by Rank, the Career Progression chart, and the Unit Modeling tool. The maximum coverage timeframe within the Details Data is:

Articles:	2004 – present.
Citations:	2004 – present.
Conference Proceedings:	2004 – present.
Books:	2003 – present.
Grants:	2006 – present.
Awards:	No Limit - present
Patents:	1976 – present.
Clinical Trials	1995 – present.
Book Chapters	2003 – present.

The Details Data allow clients to drill down into the particulars of individual faculty research activity to see publication meta-data for journals, books, and conference proceedings in addition to detailed information on awards and grants. The Details Data contain over fifty separate downloadable files which are organized into four segments:

Administration Data
 Comparative Data
 Comparative Details
 Full Data

Master License Agreement (MLA)

The Master License Agreement is the second level of licensure. Clients with a Master License Agreement have access to the entire Academic Analytics production data set. Access includes the entire Comparative and Details databases, as well as access to the larger data warehouse. Further, this level of access offers other tools such as Collaborations and Research Insight. Finally, custom analyses performed by Academic Analytics analysts are provided to answer specific client questions.

Description of data elements

The following sections describe construction of the database (inclusion of faculty and organization of entities), the elements of the database (research activity), calculations of performance comparisons, weighting of metrics, and metric definitions.

Scholars and Academic Units

Each individual faculty member is assigned an Academic Analytics Unique Identifier (AAUID), descriptive attributes and unit affiliations linked to their institution. The AAUID is carried with that person should they move from institution to institution in the database. Whenever possible, this ID number is cross-walked to an institution-assigned ID for each person to facilitate (or expedite) year-over-year updates and maintain accurate matches to publications, grants, etc. An example where this ID structure is helpful is the disambiguation of two individuals with exactly the same name, no middle initials, and where both are in the same college or school of an institution. Without a distinguishing identifier, it is easy to confuse their publications (especially if both are engaged in similar research).

In addition to attaching an institution-level ID, (Open Researcher and Contributor ID) ORCID IDs can be submitted and warehoused alongside AAUIDs and employee IDs. A list of match candidates can be provided if an institution submits faculty groupings without ORCID and would like to coordinate effort to match it with AAUID and internal employee IDs.

Descriptive attributes and unit affiliations are obtained or inferred through two processes:

Submission – Institutions provide information about their faculty members.

Collection – Academic Analytics staff compile information from publicly available web resources.

Unintended changes in the inclusion of faculty between product years may occur. Usually, this is due to one of the following situations.

An institution was collected, and a more complete faculty roster was identified online.

An institution previously submitted a faculty list and what was provided does not coincide with the information that is publicly available online (e.g., only core faculty were submitted for a PhD Program, however, joint faculty from other departments were then identified through the collection process).

If an institution is interested in looking at only “core” or primary appointment faculty, the department level data is recommended as only core members of the departments are represented. All affiliated faculty are grouped into units classified as PhD Programs.

Submission Process

While access to the data is subscription based, participation is not. Academic Analytics encourages institutional involvement to ensure that the groupings represented are as inclusive and accurate as possible. Hence, each year, a letter is sent to every Ph.D. granting institution in the U.S. requesting participation in the identification of people and structure of their university for the purpose of including that information in the comparative database.

The window for submitting faculty is between November 1st and March 30th each year. However, submissions and updates are accepted throughout the year to constantly refine the data.

We request faculty lists include all tenure/tenure-track faculty and non-TTT faculty who are expected to produce research as of November 1 of a given academic year. This includes:

- All tenure/tenure-track faculty, including administrators, faculty who are on sabbatical, faculty on leave without pay but are expected to return, and those who haven't published recently.
- Individuals not paid by the institution but who are affiliated with the institution and are expected to produce research as a condition of the affiliation.
- Others whose job involves scholarly research and are expected to pursue grants and publish. This could include research-track faculty, emeritus faculty who continue to hold a research obligation or professors of practice engaged in research.

A submission template is provided to institutions illustrating the preferred format of submitted faculty lists. In cases where an institution is unable to respond in the format requested, a list that is more convenient for the institution to produce will be accepted and processed into our system. Partial submissions are accepted in the event that an institution cannot easily provide all requested fields. Requested fields are assigned one of the following priorities.

Required – These fields are essential; without them, Academic Analytics is unable to process the submitted list. If unable to provide these fields, we ask that our request be directed to another office on campus.

- Person Name (both Last, First Mid and FML formats accepted)
- Academic Title (from HR or other system)
- Unit Name (Individual's primary academic unit affiliation or additional affiliation for which the individual has a disciplinary focus. In some cases, a person may not have a tenure home. Centers or research institutes may be submitted as their unit affiliation or as an additional affiliation. Individuals with a significant secondary focus may be submitted in more than one departmental unit.

Highly Recommended – These fields, if not provided, are assigned/inferred by Academic Analytics. They have been separated from other recommended fields due to their increased impact on accuracy and processing time.

- Person Identifier (from university system)
- Tenure Status (from HR or other system)

Recommended – These fields, if not provided, are assigned/inferred by Academic Analytics through web research. By providing any of these fields, institutions help to ensure accuracy and improve processing time.

- Faculty Rank Type (Professor, Associate, Assistant, Lecturer, Instructor, Other)
- Tenure Status Type (Tenured, TenureTrack, NotTenureTrack)¹
- Faculty Type (Regular, Research, Clinical, Public, Librarian, Other)

- Is Administrator (university-level administrator, dean of a college, or associate dean)²
- Administrative Title
- Is Emeritus
- Is Primary Appointment (for individuals submitted with more than one unit)
- College
- PhD Program (where individuals may sit on a dissertation committee and mentor Ph.D. students)³
- Hire Date (from HR or other system, date of hire, not date of tenure)
- Terminal Degree (if multiple, include first obtained)
- Terminal Degree Year (if multiple, include first obtained)
- Terminal Degree Institution (if multiple, include first obtained)
- CIP Codes (to units and Ph.D. programs as classified in institution's system)⁴

¹Note that in some cases NTT research track faculty are not easily defined or do not exist at a given institution.

²The expectation of research varies from institution to institution with regard to administrators. In some cases, this can impact unit placement in the comparative data.

³In cases where an institution cannot easily identify Ph.D. Program placement, Academic Analytics takes the list with departmental affiliations and infers Ph.D. Program placement which is later provided back to the institution for approval.

⁴Note that while CIP Codes are referenced in the creation of the Academic Analytics Taxonomy and are requested in the submission instructions, they are currently used for reference only. Reporting by CIP is not available.

Not all requested fields are available in the comparative database. Some data remain incomplete because institutions have not submitted those fields. However, they are archived because they may prove useful for clients and Academic Analytics, particularly in the data matching processes.

Useful Reference – These fields are helpful in the data matching process but are not required to perform data matching.

- Terminal Degree Field
- Publication Alias
- Prior Institution
- Gender
- ORCID

The process by which a submission list is created varies from institution to institution. It also depends on the resources available. The following describes several different paths taken to compile a submission.

- An HR file is sent to Academic Analytics followed by a review of Ph.D. program placement
- An HR file for departments is sent to Academic Analytics; the Graduate School compiles a list for Ph.D. Programs
- The previous year's submission list is distributed to Deans and/or Department Chairs for corrections and then compiled and reviewed
- An IR office identifies people using one or more systems on campus, assigns program affiliations and reaches out for feedback as needed to other offices

Once a list is submitted, Academic Analytics uploads all individuals to the data warehouse and reviews the groupings for consistency with the other submitted and collected groupings. When questions arise processing a submission, they are sent to the Academic Solutions team who follow up with the submitting institution. This often prompts conversations about inclusion since research expectations by rank and appointment can vary. This is where working directly with the institution is

preferable since the definitions of included faculty can be refined to a more granular level not possible through web collection.

Our approach is to cast a wide net and then build comparative groupings applicable to all institutions based on information available in the database. This is why some people who were included on a submitted list may not appear in the data provided back to the subscriber. A common example is clinical medicine departments. Data for these units are currently collected and stored in the warehouse; however, Academic Analytics cannot report back on clinical medicine departments because we do not yet have enough people grouped and matched to research activity metrics to build a useful national comparison.

People are classified based on criteria described in the section Collection Process which appears below. Units are organized into three categories: departments, Ph.D. programs, and other units. These categories are defined as follows:

Departments: these units can vary from institution to institution and include departments, divisions, schools (that do not contain departments), colleges (that do not contain departments), programs (degree granting units), and sometimes areas (in cases where they function as a department).

Ph.D. programs: this category includes every accredited Ph.D. program in the United States. In some cases, a program is separated into Ph.D. concentrations/specializations in an effort to provide more granular data (e.g., a Ph.D. in Business Administration with specializations in Accounting and Finance would be separated into two Ph.D. areas. In this case, data on Business as a combined effort of these two areas is available at the broad field level).

Other units: includes academic units that are not yet a part of the comparative data displayed to all subscribers (e.g., clinical medicine departments, research institutes, centers, labs, other program types, and custom groupings created by subscribers).

Once people are grouped and classified and academic units are assigned, a verification file is provided back to the institution for review of faculty placement, unit naming and classification. This allows the institution to make edits and submit them prior to finalizing the groupings that will appear in the comparative data released to all clients.

A cautionary note: There is a common misconception that all people who appear on submitted lists can be included in national benchmarking. While individuals who fall outside of these guidelines are still uploaded to the data warehouse and are available for custom groupings and the details data product, they are not included in the national comparative database. Data on these individuals are collected and stored with the idea that, over time, enough institutions will participate to open up new possibilities for comparisons. Based on client feedback, updates to the collection process will be reviewed and considered for future iterations of the database.

Maintaining standardized groupings across all institutions is a constant focus in database construction. Inclusion/exclusion decisions made by subscribing institutions are reviewed during the construction of each yearly database to ensure that the nuances of research expectation across all institutions is captured without skewing the data. This is most apparent in cases where certain types of individuals are not regularly included in submitted lists (e.g., if a small number of emeritus professors are included in a submitted list from one institution, and the following year the number of emeritus professors submitted doubles or triples, all emeritus professors are excluded by default and noted for later discussions with the client).

If an institution decides to cease submission of a faculty list, person and unit groupings are carried over for one database year in the hope that the institution will reconsider and continue to submit. In some cases, there may be an extended delay in providing a list. If a list is not submitted, the institution will be collected the following database construction cycle.

Collection Process

The first step of the collection process is to review a university's website for changes in college structure and Ph.D. programs. During this process, sources for faculty lists are identified and the best available list for a given unit is identified. In the case of departments, university directories, college directories and departmental faculty lists are reviewed to determine who to collect. Since definitive lists of individuals who sit on dissertation committees and mentor Ph.D. students are difficult to infer from website resources, the following hierarchy is employed to decide what source list to collect:

1. List with a header of Ph.D. program faculty
2. List from a graduate catalog/bulletin
3. List by research area from a departmental website
4. Department faculty list

Faculty members are then collected from the source lists identified during the structure review. Collection and inclusion of individual faculty are based on the person's title that appears on the collected list or their profile page when not available from the collected source list. Individuals with the following titles are collected, assigned an AAU-ID, and grouped in a department and/or Ph.D. program: *Professor, Associate Professor, Assistant Professor*

Based on the information available online, a person is assigned to their primary department and any Ph.D. program where they hold a primary, secondary or joint appointment. Additional appointments are not assigned to people in the department level data. However, there are times when a primary appointment is difficult to identify, and a person is represented equally in more than one department. It is assumed that individuals with a joint or secondary appointment are actively mentoring Ph.D. students and sitting on dissertation committees and that a significant portion of their research is expected in that additional discipline.

Taxonomy

Each entity is classified into our taxonomy of 171 disciplines. While the taxonomy has three levels of aggregation, reports are generated at four levels. The most specific level reports on individual faculty activity. The first and second taxonomic level reports provide comparative data at the unit level (program or department) and on a broader scale. The third level is the broad field level. Faculty names and records of research activity are de-duplicated at each level of analysis for academic units that roll up to the same broad field, so that a professor who is a member of both the *Sociology* and *Political Science* units at level one of the taxonomy appears only once in the *Social and Behavioral Sciences* broad field at level 3. The table below shows the classification of three sample disciplines.

Level 1	Level 2	Level 3
Cell Biology	Biological Sciences	Biological and Biomedical Sciences
Statistics	Mathematical Sciences	Physical and Mathematical Sciences
Anthropology	Social Sciences	Social and Behavioral Sciences

Each Ph.D. program in the United States is assigned one or more disciplinary classifications to identify the discipline(s) the program addresses. Level one categories in the Academic Analytics taxonomy are defined to include at least 20 programs in each category. Where groups are smaller than 20 programs, they are aggregated to the next level of the taxonomy to provide the most specific classification possible while maintaining a sufficiently large number of programs to facilitate statistical comparison. Level one programs are then grouped into level 2 categories that allow comparisons, for example, of activity between universities with finely divided programs in the biomedical sciences with those having broadly based "umbrella" programs. Finally, level 3 of the taxonomy presents 11 broad areas of scholarly endeavor that match those typically used by the National Research Council (NRC),

federal funding agencies, etc. (e.g., social and behavioral sciences, engineering, etc.). Comparisons of faculty research activity may be made at any of these levels.

Our taxonomy is based on the National Center for Education Statistics (NCES) Classification of Instructional Programs (CIP) code classifications, with some adjustments to include disciplines not represented with a six-digit CIP code and some omissions where 20 or more PhD Programs could not be identified. While CIP codes are requested in the submission template, reports by CIP code are currently unavailable due to the difficulty in making assignments for un-submitted faculty and the inconsistency across institutions in their assignment of CIP codes to their faculty.

Journal articles

All publishers that assign a digital object identifier (DOI) to article content and submit to CrossRef are included. The completeness and accuracy of authors collected from CrossRef relies on the information that is submitted to CrossRef. While a DOI is a requirement for inclusion, it is possible for a journal that assigns DOIs to its content to not appear in the Academic Analytics warehouse in cases where a publisher has opted out of sharing content with third-party members or due to the timing of their submission to CrossRef. If a publisher submits incomplete information, we inherit the omission. For example, some publishers only submit an article's first author. As issues are identified, corrections are pursued to the extent that resources allow. Academic Analytics encourages faculty to request that publishers participate with CrossRef DOI.

Author affiliations listed in the bibliographic record are not separated and linked to each individual author, thus it is currently not possible to create custom reports that uses author affiliations. Author position may be known for a subset of journals but is not considered in the comparative data due to the incompleteness of those data.

To prevent "double-counting" co-authored journal publications when two or more authors are in the same academic program, each of them is credited with having written a publication. However, if two or more authors of the same article are in the same Ph.D. program or academic unit, this publication counts only once toward the number of unique publications for the academic unit (likewise, the citations garnered by this publication are counted only once for the program, although each of the co-authors is credited as having written a publication that has garnered these citations when faculty members are considered individually).

Matching authors to scholars is an elaborate process that relies greatly on a manual validation effort. Probability scores help reviewers prioritize verification order and are based on keywords in abstracts/titles, co-authorship, publication patterns, statements of affiliation, and other descriptive metadata. Unsourced match candidates are also reviewed in an effort to create the most complete article profile possible using articles with a DOI. The primary focus is to not attach false matches to individuals, this makes common names the most difficult to complete since there is often over 10,000 potential unsourced articles to a single person. Scoring methods are under constant refinement to bring more probable match candidates to the top of the research queue. A small percentage of very highly scored articles are matched without manual review, this process is done with great caution to remain in line with the primary focus of not creating false author-to-scholar links.

Since it is often the case that all entries from a journal issue are assigned a DOI, Academic Analytics classifies DOIs to exclude non-article content. The article classifications currently assigned include: Article, Book Review, Other (e.g., Letters to/from the Editor, Editorials, Interviews, Working Papers).

It is assumed that the majority of journals participating in DOI are peer-reviewed. However, there is no indication that exists in the CrossRef data that identifies a journal as peer-reviewed. Unless we are pointed to a specific journal that is not peer-reviewed it is counted by default. Some publications, such as professional magazines, do participate in DOI and are warehoused but excluded from the comparative data. Academic Analytics understands that non-peer-reviewed content exists in the literatures of most disciplines with varying overall "importance" to specific disciplines. Non-DOI article

content is not warehoused by Academic Analytics, but is under consideration as a future collection project pending resource availability.

A complete list of journal coverage is available on the client portal main page under “Home > Documentation.”

Conference proceedings

Proceedings that are assigned DOIs and submitted to CrossRef are included. The lag time between publication and appearance of publication records in CrossRef is greater than with journals. Hence, completeness of matches to scholars in subsequent comparative releases will increase as more data become available. It is estimated that 80% of conferences expected to submit proceedings with DOIs will be represented in the first release of AAD each year. Data matching processes are like articles, but conference proceedings present additional difficulties in that only the first author’s affiliation is sometimes listed, and abstracts aren’t as populated.

Disciplines in engineering and physical/mathematical science devote a substantial amount of their overall publishing activity to the publication and dissemination of conference proceedings, and enough is captured that they are weighted by default in the comparative data. Though proceedings are not weighted in all disciplines, they are matched to all scholars. These numbers will appear in Faculty Counts and in the Faculty Details downloads. Some disciplines will have limited coverage based on conference participation with CrossRef. A complete list of proceedings is available on the client portal under “Home > Documentation.”

Citations

The CrossRef citation-linking network is the source for citations. According to their website (as of September 2018) “Our citation-linking network today covers over 100 million registered content records (journal articles, book chapters, data, theses, technical reports) from thousands of scholarly and professional publishers around the globe.” Citations to journals and conference proceedings in “other content items” are included in the citation count. Currently, we do not have the ability to separate citations based on source.

Books

The Academic Analytics data warehouse includes book publication data obtained from Baker & Taylor and The British Library. A book title is reported once per author/editor/translator; all published works are weighted equally (distinctions between authors, editors and translators are available in the Academic Analytics database portal). Series editors are not included in the books metrics and do not appear in the data feeds from our providers. As with journal articles, co-authored/edited books duplication is removed as the data are aggregated to broader categories.

Introductions, forewords, afterwards, and citations in and to books are not currently captured. If a book has been published in multiple printings during the ten-year rolling comparative window, only one instance is included. If a book has been updated with each publication, all editions/volumes are included. If a book was originally published before the comparative window and was published in a different format within the observed period, it is included (i.e., a book published originally in 2003 and released as a paperback in 2017 would be included).

Books are included and weighted in the comparative data for the following broad subject areas:

- Business
- Education
- Family, Consumer and Human Sciences
- Health Professions Sciences
- Humanities
- Social and Behavioral Sciences

Though books are not weighted in all disciplines, they are matched to all scholars. These numbers will appear in Faculty Counts and in the Faculty Details downloads. Books are not weighted based on publisher, but publishers are attached to the book records and are available through the Academic Analytics database portal.

Matching books to authors/editors/translators is less difficult than journal articles in the sense that there are less possible records to match. However, they are more difficult in that they do not have all the additional descriptive data such as author institutional affiliation at time of publication. Academic Analytics uses a combination of name uniqueness and book subject to level one classification to present reviewers with more probable matches. Common names involve a more intensive review.

Grants

The database includes grants data from 17 federal agencies matched to the principal investigator and Co-Pis (when the data is available). Funding is attached to principal investigators, so a grant received while at one institution is carried to the next institution as people move between universities. Grants data are obtained through a combination of online search engines and through Freedom of Information Act (FOIA) requests.

Data are presented as annualized amounts based on the awarded or estimated amount of the grant divided by the number of years derived from either the start/end dates of the project or of the budget period where transactions have been reported. The duration in years is calculated by subtracting the start date from the end date, total days are divided by 365.25 to come up with the annualized dollar amount. When the full history of a grant is not known, funding is calculated based on the budget start and end dates rather than the project start and end dates.

For renewals and extensions (including no-cost), we update the total dollars and the start/end dates and re-calculate the dollars if the Federal Award Id remains the same. Renewals that are assigned a new ID by a federal agency are counted as a new grant. For example, NSF renewals compete with all other pending proposals and are then assigned a new Federal Award Id. Since we count grants based on the ID, we count that renewal as a new grant.

No-cost extensions are currently included as they have come through grants reporting as updates. Filtering dollars per year based on the original duration of the grant is currently not available.

Due to the dynamic nature of the online grant search systems, Academic Analytics provides frequent updates to the dollars and durations as new information on transactions and extensions are available. This will impact the dollars per year variable between releases of the database. In cases where the funding is consistent YOY to a project, the dollars per year will show little or no change between releases, however, it is often the case that there is an increase or decrease in funding over the lifespan of a grant which will cause changes between each release.

Federal agencies captured in AAD:

- National Aeronautics and Space Administration (NASA)
- National Endowment for the Arts (NEA)
- Department of Defense Air Force Office of Scientific Research (AFOSR)
- Department of Defense Army Research Office (ARO)
- Department of Defense Office of Naval Research (ONR)
- Department of Education Institute of Education Sciences (ED)
- Environmental Protection Agency (EPA)
- Department of Energy Office of Science (DOE)
- Federal Aviation Administration (FAA)
- National Institute of Food and Agriculture (USDA)*
- Department of Health and Human Services (HHS/NIH)*
- Institute of Museum and Library Services (IMLS)
- National Endowment for the Humanities (NEH)*
- National Science Foundation (NSF)*
- National Oceanic and Atmospheric Administration (NOAA)

This document contains Academic Analytics' confidential and proprietary, business trade secrets. This document may not be transferred or used by any other person or entity other than your university. For internal use only.

*CoPIs also matched (distribution of funds unknown)

Agency	Funding Type	Coverage	Note
AFOSR	Grants	2005-2014	Included in Comparative Data by Default
ARO	Grants	2007-2013	
DOE	Grants	2005-	
ED	Grants	2005-	
EPA	Grants	2005-2016	
FAA	Grants	2005-	
HHS/NIH	Grants	2005-	
IMLS	Grants	2005-	
NASA	Grants	2005-	
NOAA	Grants	2005-	
NSF	Grants	2005-	
ONR	Grants	2007-2016	
USDA	Grants	2005-	
NEA	Grants/Fellowships	2005-	
NEH	Grants/Fellowships	2005-	
ARO	Cooperative Agreements	2006-2016	Excluded from Comparative Data by Default
DOE	Cooperative Agreements	2005-	
HHS/NIH	Cooperative Agreements	2005-	
NASA	Cooperative Agreements	2005-	
NOAA	Cooperative Agreements	2005-	
NSF	Cooperative Agreements	2005-	
NIH	Fellowship	2005-	
NSF	Fellowship	2005-	
NIH	Grant (Career)	2005-	
NIH	Grant (Center)	2005-	
HHS/NIH	Grant (Training)	2005-	
NIH	Sub-Grants	2005-	

Awards

The database includes over 16,000 honorific awards from over 2,500 governing societies.

The general principle for the inclusion of an award in the database is that the award must be open to all people in a (sub-) discipline or to a large subset (i.e., age, gender) at the national and/or international level. State and local awards are not currently captured by Academic Analytics. Expansion of the awards list is ongoing and largely based on feedback from client institutions.

A complete list of awards is available on the client portal main page under Home > Documentation.

Patents

Patents are collected directly from the USPTO on a yearly basis and Academic Analytics has collected this data back to 1976. The database includes over 156,000 patents from the USPTO

This document contains Academic Analytics' confidential and proprietary, business trade secrets. This document may not be transferred or used by any other person or entity other than your university. For internal use only.

matched to current scholars. Patents are not currently included in the comparative tools but are utilized within Research Insight.

Clinical Trials

Clinical Trials are collected directly from ClinicalTrials.gov on a quarterly basis and the data reaches back to 1995. The database includes over 64,000 Clinical Trials matched to current scholars. Clinical Trials are not currently included in the comparative tools but are utilized within Research Insight.

Book Chapters

Book Chapters are collected on a monthly basis from our Book provider, Baker & Taylor. The database includes over 730,000 chapters matched to current scholars, with data reaching back to 2003. Book Chapters are not currently included in the comparative tools but are utilized within Research Insight.

Calculations and Definitions

Calculating Program, Broad Field and Institutional Index Scores

Person Based Calculations

We calculate the aggregate Scholarly Research Index based on the mean person-level scores of individual faculty members. Thus, the SRI of each unit or other level of aggregation – program, department, broad field, institutions, etc.) is the average of the faculty who comprise that unit. Using a “person based” SRI score has a number of advantages. First, it allows us to expand comparisons beyond units to groups of individuals aggregated under different rubrics. We are now able to compare research activity of individuals by faculty rank, comparing faculty holding different academic ranks and comparing faculty within a single academic rank. Further, we can look at all faculty in a discipline and view their research activity compared to years since terminal degree, viewing their career progression within the discipline. In addition, we have found that this method compensates for the “portfolio effect” found in the data for many institutions. For example, our previous unit-based method of calculation for an entire institution took the average of all the Ph.D. programs at that institution, which means a program with only 5 faculty members influenced the university Scholarly Research Index as much as a program with 100 faculty members. Now, the institutional Scholarly Research Index is the result of the average SRI scores of all individual faculty members, effectively eliminating this issue.

Metrics for Person-Based Scholarly Research Index (Default Metrics)

Metrics for the person-based Scholarly Research Index are:

Total Journal Articles	Total Grant Dollars
Total Citations	Total Awards
Total Books	Total Conference Proceedings
Total Grants	

By capturing national data on all faculty research activity within a specific discipline and attaching those data to individual faculty, Academic Analytics is able to organize and rank individual faculty based on their individual research activity within their discipline. We have organized faculty into 20% bands, or quintiles, for presentation of these data. A tool in our online portal provides a picture of research activity in the discipline through a description of each of the quintiles in terms of:

Average Citations	Average Awards
Average Journal Articles	

This document contains Academic Analytics' confidential and proprietary, business trade secrets. This document may not be transferred or used by any other person or entity other than your university. For internal use only.

Average Books
Average Number of Grants

Average Annual Grant Dollars
Average Conference Proceedings

For clients at the Details level of licensure, Academic Analytics provides a second view of research activity by academic rank. For this view, all discipline faculty have been separated out by academic rank. They have been organized into five quintiles according to their individual within-rank SRI scores. The first quintile is the most active; the fifth quintile is the least active. Each academic rank is identified and the total number of individuals at that rank in the discipline is given. Each quintile is described in terms of:

Average Citations
Average Journal Articles
Average Awards

Average Books
Average Number of Grants
Average Annual Grant Dollars
Average Conference Proceeding

Weights for Person-Based SRI

Our weighting scheme is calculated as follows:

1. The weights presented in the 2010 National Research Council Survey of Faculty Opinion were applied to the 59 disciplines that Academic Analytics and the 2010 NRC study had in common, and then modified as described in point #2, below. For the remaining 113 disciplines measured only by Academic Analytics (total 172), the median value of the weights for the disciplines in the broad field (AA Level 3) were calculated based on the NRC study's "known" disciplines, and those weights were applied to the discipline (AA Level 1), and then modified as described in point #2, below.
2. A proprietary series of thresholds is applied to the data for each academic discipline to determine whether to include each of the categories of variables in the calculation of SRI (e.g., in Physics, books are not included in SRI; in English, grants are not included in SRI, etc.). Following this procedure, steps are taken to assign weights to each variable:
 - If honorific awards are included in the calculation of SRI, they are given a weight of either 20%, 10% or 5% based on their frequency in the discipline.
 - If federal research grants are included in the calculation of SRI, they are given a weight of either 30%, 20% or 10% of SRI based on their frequency in the discipline. The total weight applied to the general category "Research Grants" is then divided by the number of Research Grants metrics that are used in the weighting scheme and that percentage is applied to each Research Grants metric (e.g., if the Research Grants weight is 10% and there are two Research Grants metrics, then each metric is weighted at 5%).
 - After the percentages for awards and grants is known, the remaining percentage that can be allocated to books, journal articles, conference proceedings and citations is calculated (e.g., if awards are worth 10% and grants are worth 20%, then there is $100\% - 10\% - 20\% = 70\%$ remaining to be distributed among books, journal articles, conference proceeding and citations).
 - If published books are not included in the calculation of SRI, the remaining percentage is distributed between publications variables and citations variables based on the NRC survey results.
 - If published books are included in the calculation of SRI, then the weights for books, journal articles, and conference proceedings from the NRC survey are applied.

Finally, it is important to note that simply multiplying the weight for a given metric and the value for that metric (or of its z-score), then summing all metrics and averaging all metrics, will not result in SRI. Due to long-tailed (i.e., non-normally distributed) data and differing scales of measurement, we rank each metric before multiplying by the weight to that rank and then summing, to arrive at SRI.

Unit Aggregations

The Academic Analytics database is organized according to a taxonomy of 171 disciplines. Each program or department is classified into our taxonomy with one or more disciplinary classifications. Reports are generated at four levels of aggregation; the base level is the most specific, the individual faculty. The first and second level reports provide comparative data on a broader scale. The third level is the Broad Field. Faculty names and records of research activity are de-duplicated at each level of analysis for academic units that roll up to the same broad field, so that a professor who is a member of both the *Sociology* and *Political Science* units at level one of the taxonomy appears only once in the *Social and Behavioral Sciences* broad field at level 3. Thus, the SRI score of each unit or level of aggregation; program, department, Broad Field or institution is the average of the faculty who comprise that unit.

Taxonomy level 01, 02, and 03 are comprised of both volume and per capita metrics. Volume metrics are the de-duplicated faculty counts of research activity. Because volume metrics are often a function of how many faculty are in a unit (more faculty leads to higher totals) we derive per capita metrics such as, “per faculty” and “percentage of faculty.” The per capita metrics reduce the faculty size effect and allow smaller units to compare to larger units. Unit level metrics are:

1. **Books metrics**

- Total books published
- Number of faculty with a book publication
- Percentage of faculty who have authored a book
- Books published per faculty member

2. **Journal publication metrics**

- Total journal articles
- Number of faculty with a journal article
- Percentage of faculty who have authored a journal article
- Journal articles published per faculty member
- Articles per author

3. **Conference proceedings metrics**

- Total conference proceedings
- Number of faculty with a conference proceeding
- Percentage of faculty who have a conference proceeding
- Conference proceedings per faculty member
- Conference proceeding per author

4. **Citations metrics**

- Total citations
- Number of faculty with a citation
- Percentage of faculty with a citation
- Citations per faculty member
- Citations per journal article
- Citations per author

5. **Federal funding metrics**

- Total federal grants
- Total federal grant dollars
- Number of faculty who have a federal grant
- Percentage of faculty who have won new and competitive federal research funding
- Federal grants per faculty member
- Federal grant dollars won per faculty member
- Dollars per grant

6. **Honorific awards metrics**

- Total honorific awards
- Number of faculty who have won an honorific award
- Percentage of faculty who have won an honorific award
- Honorific awards per faculty member

The following descriptive statistics are provided for each metric and are relevant to the comparison group chosen.

1. Rank
2. Z-score (metric value - μ of the metric value) / σ of the metric value
3. Percentile rank

University Level Aggregation

The institutional SRI score is the average of all individual faculty at the institution. Faculty are de-duplicated at each level of analysis for academic units that roll up to the same broad field.

Metrics at the university level are provided as de-duplicated volume totals and not as per capita metrics. Due to the diversity of metrics that are considered relevant or important among disciplines, the faculty denominator in the calculation becomes less meaningful when calculating per capita metrics.

Unit Based Calculation and Metrics

Historically, the Academic Analytics databases have contained two basic types of metrics: “Full data metrics” and “Per capita metrics.” Full data metrics represent raw data on faculty research activity. This includes total counts of journal publications, citations, number of grants and grant dollars, awards, book publications, the number of conference proceedings published, and the number of faculty who have produced or received each of the above. For AAD 2011 and prior databases, these full data metrics were not used in their raw form to understand the research activity of a given program or department relative to other programs or departments, because they did not take into account the size of the program in terms of total number of faculty. Larger programs typically have more total publications, awards, books, and other forms of research activity simply because they have more people contributing to overall scholarship. Per capita metrics solved this problem by converting raw metrics into per capita and percentage form. For example, “total journal article publications” for a given program became “journal article publications per faculty member.”

In AAD 2011 and prior databases, the z-scores for each metric serve to facilitate comparisons between unlike data types (e.g., journal article publications versus honorific awards, or “Department of Chemistry” versus “Department of History”). A program’s z-score for a given metric represents the research activity of this program relative to the average program in the same discipline. In this calculation method, z-scores are calculated by subtracting the average score of the metric for the entire discipline from the program’s score, and then dividing this difference by the standard deviation for that metric in the discipline. The standard deviation is simply a measure of the overall variation of the scores in the discipline, and this division is a way to put all scores on the same scale. For example, if the average number of journal publications per faculty member in discipline “A” is 6.0, and the standard deviation for this discipline is 2.0, then a program with 8.0 publications per faculty has a z-score of 1.0; i.e., $[(8.0 - 6.0)/2.0] = 1.0$, or “one standard deviation about the mean.” For more information on z-scores, please visit the following URL: http://en.wikipedia.org/wiki/Standard_score.

As you can see from the formula above, programs with scores above the national average have positive z-scores, and programs with scores below the average have negative z-scores. The absolute value of the z-score represents how far the program is from the comparative group average in either the positive or negative direction. Thus, a score of 2.0 represents a program that is very research active relative to the average program in its discipline (two standard deviations above the comparative group average), while a z-score of -2.0 represents a fairly inactive program (two standard deviations below the comparative group average). Z-scores are a convenient way to represent the relative standing of a program within its discipline, compared to standard or user-specified peer groups.

Z-scores are also useful for comparing the performance of a program across metrics. If we wanted to compare the research activity of a given program in terms of grant dollars to its activity in terms of journal publications, we would not want to look at grant dollars per faculty member and publications per faculty member, because these two metrics are on different scales (i.e., dollars and articles). For example, if a given program has \$10,000 per faculty member in grant funding and 6.0 articles per faculty member in terms of journal publications, we are unsure about which score represents greater activity due to different scaling. We can, however, directly compare the program’s z-scores. If the z-score for grant dollars is 1.0, and the z-score for publications is 2.0, then we know that the program is relatively more active in publishing journal articles than in obtaining grant money when compared to the rest of the comparative group.

The z-score representation is also advantageous because it allows for the construction of a single summary measure of the overall research activity of a program, relative to the rest of its discipline, by combining its scores on many metrics. Since z-scores are all on the same scale, we can combine a program’s scores on all fifteen metrics to create the summary rank or percentile in the discipline.

Definitions of Metrics

In the following explanation of metrics 'μ' refers to the mean of the population while 'σ' refers to the standard deviation of the population.

Books Metrics

Percentage of Faculty with a Book Publication – pfacbp

hasbook / numfac

Z-Score of Percentage of Faculty with a Book Publication – b1

(pfacbp - μ of pfacbp) / σ of pfacbp

Rank on Percentage of Faculty with a Book Publication – Rb1

Book Publications Per Faculty - bppfac

totbooks / numfac

Z-Score of Book Publications Per Faculty – b2

(bppfac - μ of bppfac) / σ of bppfac

Rank on Book Publications per Faculty – Rb2

Publication Metrics

Percentage of Faculty with a Journal Publication – pfacjp

hasjpub / numfac

Z-Score of Percentage of Faculty with a Journal Publication – p1

(pfacjp - μ of pfacjp) / σ of pfacjp

Rank on Percentage of Faculty with a Journal Publication – Rp1

Journal Publications per Faculty Member – jppfac

totpubs / numfac

Z-Score of Journal Publications per Faculty Member – p2

(jppfac - μ of jppfac) / σ of jppfac

Rank on Journal Publications per Faculty Member – Rp2

Conference Proceedings Metrics

Percentage of Faculty with a Conference Proceeding – percentFacultyConfProc

HasConfProc / numfac

Z-Score of Percentage of Faculty with a Conference Proceedings – zScore_percentFacultyConfProc

(percentFacultyConfProc - μ of percentFacultyConfProc) / σ of percentFacultyConfProc

Rank on Percentage of Faculty with a Conference Proceedings –

rank_percentFacultyConfProc

Conference Proceedings per Faculty Member – perFacultyConfProc

TotConfProc / numfac

Z-Score of Conference Proceedings per Faculty Member – zScore_perFacultyConfProc

(perFacultyConfProc - μ of perFacultyConfProc) / σ of perFacultyConfProc

Rank on Conference Proceedings per Faculty Member – rank_perFacultyConfProc

Citations Metrics

Percentage of Faculty with a Citation – pfacc

hascit / numfac

Z-Score of Percentage of Faculty with a Citation – p3

(pfacc - μ of pfacc) / σ of pfacc

Rank on Percentage of Faculty with a Citation – Rp3

Citations per Faculty Member – cpfacc

totcits / numfac

Z-Score of Citations per Faculty Member – p4

(cpfacc - μ of cpfacc) / σ of cpfacc

Rank on Citations per Faculty Member – Rp4

This document contains Academic Analytics' confidential and proprietary, business trade secrets. This document may not be transferred or used by any other person or entity other than your university. For internal use only.

Citations per Publication – cpp
 $\text{totcits} / \text{totpubs}$

Z-Score of Citations per Publication – p5
 $(\text{cpp} - \mu \text{ of cpp}) / \sigma \text{ of cpp}$

Rank on Citations per Publication – Rp5

Grant Metrics

Percentage of Faculty with a Grant – pfacgrnt
 $\text{hasgrnt} / \text{numfac}$

Z-Score of Percentage of Faculty with a Grant – g1
 $(\text{pfacgrnt} - \mu \text{ of pfacgrnt}) / \sigma \text{ of pfacgrnt}$

Rank on Percentage of Faculty with a Grant – Rg1

Grants per Faculty Member – gpfac
 $\text{totgno} / \text{numfac}$

Z-Score of Grants per Faculty Member – g2
 $(\text{gpfac} - \mu \text{ of gpfac}) / \sigma \text{ of gpfac}$

Rank on Grants per Faculty Member – Rg2

Grant Dollars Per Faculty Member – gdpfac
 $\text{totgdol} / \text{numfac}$

Z-Score of Grant Dollars Per Faculty Member – g3
 $(\text{gdpfac} - \mu \text{ of gdpfac}) / \sigma \text{ of gdpfac}$

Rank on Grant Dollars Per Faculty Member – Rg3

Dollars per Grant – dolpgrnt
 $\text{totgdol} / \text{totgno}$

Z-Score of Dollars per Grant – g4
 $(\text{dolphgrnt} - \mu \text{ of dolpgrnt}) / \sigma \text{ of dolpgrnt}$

Rank on Dollars per grant – Rg4

Award Metrics

Percentage of Faculty with an Award – pfacawd
 $\text{hasawd} / \text{numfac}$

Z-Score of Percentage of Faculty with an Award – a2
 $(\text{pfacawd} - \mu \text{ of pfacawd}) / \sigma \text{ of pfacawd}$

Rank on Percentage of Faculty with an Award – Ra2

Awards per Faculty Member – awdpfac
 $\text{Totawds} / \text{numfac}$

Z-Score of Awards per Faculty Member – a1
 $(\text{awdpfac} - \mu \text{ of awdpfac}) / \sigma \text{ of awdpfac}$

Rank on Awards per Faculty Member – Ra1

Weight Metrics

Weight of Awards Metrics in this Discipline – awdwt

Weight of Grants Metrics in this Discipline – grntwt

Weight of Books Metrics in this Discipline – bookswt

Weight of Citations Metrics in this Discipline – citswt

Weight of Journal Publication Metrics in this Discipline – pubswt

Weight of Conference Proceedings Metrics in this Discipline – confprocwt