Introduction to Academic Analytics (AA) at CU-Boulder

December 20, 2016

Jason Straub, Institutional Research, Office of Data Analytics (ODA-IR)

Posted at http://www.colorado.edu/oda/institutional-research/employee-data/academic-analytics-aa-cu-boulder

IR ref L:\IR\Reports\GuideGR\AcAnalytics\AA_UCB_Intro.docx

Background
- Academic Analytics (AA) is a private firm providing data on faculty scholarly activity to PhD-granting universities in the US and UK. Founded in 2005 by then-graduate dean Lawrence Martin of Stony Brook University, AA has at least 36 AAU and 108 total subscribers plus over 85 employees working from locations throughout the US and UK. Public website: http://www.academicanalytics.com/. The AA database links over 270,000 individual faculty members in over 400 US PhD-granting institutions to organizational units (e.g., departments) and PhD programs, and to comparable records of journal articles, citations, books, conference proceedings, federal research grants, and honorific awards. These records are collected by AA itself from publishers, granting agencies, and awarding bodies. Units and PhD programs are assigned to a common disciplinary categorization (e.g. history, physics).
- CU-Boulder has held a master, full details, license with Academic Analytics for the 2011-2015 releases. Rob Stubbs and Jason Straub of ODA-IR are CU-Boulder liaisons to Academic Analytics. (Lou McClelland was liaison through 2016 and served on the Academic Analytics consulting committee since its founding.)
See diagrams of what Academic Analytics does, and how CU-Boulder's submission (lists of people, departments, PhD programs, and affiliations with disciplines) fits in following the text here and at http://www.colorado.edu/oda/institutional-research/employee-data/academic-analytics-aa-cu-boulder. Diagrams show the UCB submission for the 2015 release.

Releases
- AA comparative data are released annually and based on faculty as of mid-fall term (November 1, the federal counting date for faculty).
- Releases are labelled with the year of the faculty list
- The 2015 comparative release in use now is based on faculty 2015-16 with work in the comparative time window:
  - Journal Articles: 2012 - 2015
  - Citations: 2011 - 2015
  - Conference Proceedings: 2012 - 2015
  - Books: 2006 - 2015
  - Grants: 2011 - 2015
  - Awards: No Limit
- The 2015 comparative release was delivered mid October 2016; the 2016 release is expected in September 2017.

Data delivery modes
- Web portals for comparative data, with screen displays and downloads, with a major revision, version 2, released late 2016. Both v1 and v2 portals will be in use indefinitely.
- A "data mart" accessed by Institutional Research staff – like a data warehouse of AA data for CU-Boulder. Includes both comparative and full data on units, PhD programs, and

individuals at CU-Boulder, on PhD graduates from UCB, and some data on units and PhD programs at other schools.
- Tableau files on selected topics are under development.
- Custom files and reports by Straub and by Academic Analytics

Using Academic Analytics data for comparisons with other schools

AA's designed use is comparison of PhD programs (especially) or organizational units within a discipline (e.g., history) to programs and units at other US PhD-granting universities, or to a subset thereof, such as the AAU.   This use makes sense if and only if

- AA's records of scholarly work and matches to individuals are sufficient – This is the weak link for arts and humanities and law, less an issue for other fields. The lack of book chapters is a special difficulty for humanities.
- Comparison units and programs are indeed comparable.   To ensure this
  - User must review comparison programs/depts *first* – names of programs/depts, associations with disciplines, etc. Sometimes this is simple – history departments and PhD programs are (at least in the AAU) generally stand-alone and associated with AA discipline history. Sometimes it's not at all simple; business, education, visual/performing arts, and media-related programs and units parse into completely different arrays even within the AAU.
  - At the same time, user must review AA's faculty counts in comparison units to see if AA has captured the department/program they know. AA counts and lists sometimes include only "core faculty," sometimes include more. Sometimes departments and PhD programs in a discipline have the same number of faculty, sometimes one or the other is far larger.   Just within the AAU, departments associated with history list from 3 to 90 faculty; those associated with management 3 to 140. PhD programs are no better, and sometimes have fewer, sometimes more, faculty than corresponding departments. User should consider both "departments" and PhD programs
  - User should provide *a priori* peer or comparison-school lists
  - User and Straub/IR together settle on a selection of departments and/or PhD programs held by AA – or, if absolutely necessary, *parts* of those (e.g., Dept. X at School Y minus people A B C).
- User considers measures carefully
  - AA summarizes to two general classes of metrics for departments and programs:
    - Absolute: e.g., number of publications in the time window; books (in the time window) per faculty member
    - Relative: e.g., this program's number of publications relative to number of publications over programs associated with the same discipline, at specified institutions – e.g., AAU only
  - Within each class there are volume measures, per-capita measures, participation measures (e.g., what number or percentage of faculty in a group had any publications), and intensity measures (dollars per grant, citations per publication). Smaller departments/programs generally like per-capita measures; bigger ones like volume measures. Reputation *may* reflect volume rather than per-capita measures.
  - Relative measures *can* be combined across areas of scholarly work. Cross-area measures can be convenient but can hide patterns of difference across areas of work. They should rarely be used alone.

- AA's "Index" (formerly FSPI, Faculty Scholarly Productivity Index) is now calculated
  - For <u>individuals</u> based on counts of their publications, citations, books, grants, grant dollars, awards, and conference proceedings, in the release time window, relative to all individuals in the comparative data in the same discipline. Z-scores for the areas are combined using weights appropriate to the discipline (e.g., for English, books weight heavily, grants not at all).
  - For entire <u>institutions</u> based on averaging the indices of all (unduplicated) people in the comparative data.
  - For <u>programs and units</u>, a weighted average of volume measures, per-capita measures (articles, books, etc. per person), relative measures of participation (pct of people who had articles, books, etc.), and intensity measures (dollars/grant, citations/article), all relative to other programs and units in a selected discipline.
- McClelland's "Index01," developed fall 2012, averaged 7 relative per-capita measures with relative number of faculty and relative citations/article.   This included no volume measures.
- McClelland, October 2015, delivered a measure to Academic Affairs Budget Advisory Committee, for Academic Prioritization, based on the maximum of average per-capita and average volume measures, relative to AAU schools only.
- Many other cross-area measures are possible.

What CU-Boulder submits annually and gets in return
- Each year we submit to AA (see diagram for counts)
  - *People*:
    - All tenured and tenure-track faculty (TTT) for the academic year, including those starting in spring, on leave with or without pay, with any title including administrative.
    - Those with formal HR appointments, paid or not, active as of November 1, with any of these job titles
      - full-associate-assistant research and adjoint professors (except ROTC)
      - Research Associate, Senior Research Associate museum-related, post-doctoral fellow (*as the HR job title, used for a fraction of actual post-docs on campus*)
    - Individuals who have no CU Boulder HR appointment who are formal listed on web as fellows (or equivalent title) in research institutes, and retired faculty with no appointment in selected programs and departments.
    - *Roughly half of all people submitted each year contribute to the comparative data. The others (all RA/SRA and all affiliated solely with a research institute) are in the full data only.*
  - *Departments*: All degree- and tenure-granting units including departments/divisions in A&S, Leeds, CMCI, and Engineering; other tenure-granting programs (e.g. environmental studies); other schools/colleges; all research institutes & museum.
    - *In AA these units are called units, type = department.*
    - *CU Boulder research institutes are "departments" but are not in the comparative data.*
  - All *PhD programs*, including interdisciplinary and joint, per CU Boulder websites for prospective students.
    - *In AA PhD programs are called units, type = program.*

- *Affiliations* of people to units and PhD programs, based on tenure homes, appointments in UCB HR files, and web listings.
- *Associations* of units and programs to one or more of the 172 AA disciplines, such as history.
- The 2015 release added College of Media, Communication and Information (CMCI) programs and departments.
- AA matches data on scholarly work – publications, citations, books, grants, honors and awards – to our people. It does the same for other US institutions with PhD programs.
- At least through the 2015 release, we have a "master license" with AA, which means we can access
    - Counts and records of individual works (e.g., a single journal article, book, or grant), by person, by name, including h-indices. Also summaries or counts by person by year.
    - Lists of departments and programs, by name, at all schools, with what AA discipline(s) each is associated with, count of faculty, and more. We use these to help users to construct appropriate peer/comparison groups, and even to explore how programs and departments are named across the country.
    - "Comparative data" – summaries based on work in defined time windows, generally by TTT and research/adjoint professor-line faculty only -- for departments and programs at CU-Boulder and at other U.S. schools. Schools, departments, and programs are all named.
    - The web portal, with comparative data presented by discipline, plus drill-down detail on journals, grant agencies, and individual faculty plus downloads of both comparative and full data. "Full data" on CU-Boulder, which extend beyond the time windows, people selections, and rules (e.g., peer-reviewed articles only) used in comparative data
    - Data on individuals tracked by AA who received PhD's from CU-Boulder
    - AA beta products such as displays of collaborators and word or phrase searches of article titles and abstracts

Campus users

- College of Arts and Sciences – dean, associate dean research, and some departments
- CIRES
- Institutional Research for various
- Faculty Affairs for honors and awards
- AABAC for Academic Prioritization
- Academic Review and Planning
- School of Education
- VC Research & Innovation office
- Leeds, Engineering

Access

- Access to the web portal, and/or to custom analyses naming other schools, requires agreement and adherence to a non-disclosure agreement. The agreement generally prohibits mention of schools other than CU-Boulder by name in discussing or presenting any results that may become public.
- Campus prospective users should clip the non-disclosure section of http://www.colorado.edu/oda/institutional-research/employee-data/academic-analytics-aa-cu-boulder and send with statement of agreement, and user's full name, title, and phone, to IR@colorado.edu. ODA-IR will then request portal access (if desired) and notify the user.

CU-Boulder uses of Academic Analytics data (also see below for Academic Analytics facilities in development or underused at CU-Boulder as of yet)

A. **Comparison**
- In fall 2012, McClelland used the AA 2010 release to provide comparison data to CU Boulder social science departments for use in Academic Review and Planning self studies. The comparison group was a straight discipline match to almost all schools in AA data. Several measures were displayed, both absolute and relative, including Index01, a cross-area measure combining seven relative per-capita measures plus number of faculty. Most departments did mention AA data in their self-studies. However, this exercise also exposed perils of peer selection and index creation.
- Index01 for all departments (not just social studies) was used in the first campus report on Academic Prioritization (no longer posted), issued 2014.   This was problematic due to variably-appropriate discipline matches, the narrowness of Index01, reliance on a single measure to tap "scholarly accomplishments," and various other issues.
- Since 2013 the School of Education and the dean of the College of Arts and Sciences have found the web portal useful for comparisons.
- In October 2015, ODA-IR used the 2013 release and modified methods to update figures for Academic Prioritization. Nevertheless, problems remain.
- In fall 2015, Academic Review and Planning, and department chairs of the Arts and Humanities units undergoing review, received both Academic Prioritization results and snapshots of "what the dean sees" about them in the Academic Analytics portal. Two departments, English and Asian Languages and Civilizations, collaborated with McClelland throughout 2015-16 to understand, characterize, and improve Academic Analytics data coverage. See an overview of results in the section "Exploring AA data for humanities" of http://www.colorado.edu/oda/sites/default/files/attached-files/2016aaude_cu-boulder_acanalyticsmasked.pdf
- In fall 2016, Academic Review and Planning, and chairs and directors of the physical sciences units undergoing review, received both Academic Prioritization results and snapshots of "what the dean sees" about them in the Academic Analytics portal.
- Fall 2016, credentialing and orienting new users in A&S, VC Research/Innovation, and Engineering

B. **Find CU-Boulder PhD recipients in AA database** (faculty at PhD-granting schools US/UK)
- Example: 25 found with current discipline anthropology, 8 at AAU's.
- The Graduate School, BioFrontiers, and others are interested in these data.
- We are in process of matching AA records to our degree records to get CU-Boulder degree program or counts of those in/not in the AA db. AA is helping with this.

C. **Find depts nationally with same discipline mix**
- Our Chemistry and Biochemistry is in two disciplines. But 105 depts nationally in AA are named "Chemistry and Biochemistry" – those would make better comparators than looking separately within each discipline
- We reported to the Regents on names of units in the AAU associated with gender studies
- We used AA data to explore patterns of organization of mathematics and applied mathematics
- *The AA database of unit and PhD program names is a unique asset.*

D.   Use AA-collected data on individual works
- Through CU Experts https://experts.colorado.edu/ Faculty Affairs publishes AA-collected compilations of the national and international honors and awards received by CU-Boulder faculty. E.g., Fred Anderson, fellow, National Humanities Center, 2012. Faculty Affairs receives a regular data flow on honors and awards from Academic Analytics.
- CIRES (Cooperative Institute for Research in Environmental Sciences) in 2015 and 2016 used AA-collected articles matched to over 300 CIRES-affiliated scientists at CU-Boulder and NOAA. Records from AA were used to supplement, check, and standardize CIRES own records, especially for required annual reporting to NOAA. The result was improved CIRES tracking, improvement to Academic Analytics data, and a request from CIRES director Waleed Abdalati to continue our collaboration. In 2016 this collaboration gave CIRES more than 100 additional publications for the annual report, not found by CIRES' own tracking system.
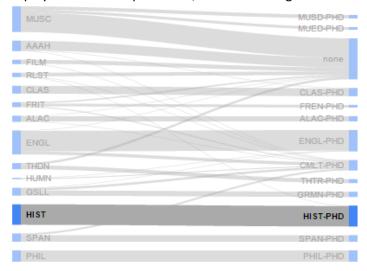
E.   Characterize the work of individuals. Examples:
- All publications, grants, awards, books for one person, plus that person's collaborators (as indicated by co-authorship) inside and outside CU Boulder
- All the people who have received a specified award
- People in a department or school or institute or group of units who have no publications, or high publications, or high h-indices, or high citations, or other specified criterion
- Yet to come: Use of Academic Analytics publication records in CU Boulder's Publication Data Project. Use in promotion and tenure processes, reviewed and augmented by the individual.
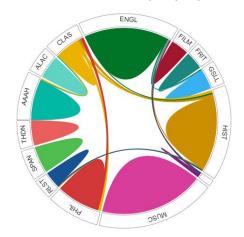
F.   Work with Academic Analytics to improve and understand the data. Examples:
- Match AA publication lists to CV's
- Match AA publications to CIRES' publication records
- Feedback on portal v1, portal v2, and Tableau files
- Developing methods of specifying peer or comparison departments and programs
- Use of unit "demographics"
- Clean-up of degree institution records
- Design of data mart
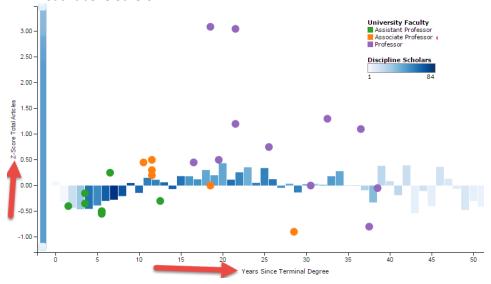- Metrics: Gini coefficients, index alternatives

G. Characterize ourselves using simply our annual submission <u>to</u> AA of people, units, people-to-units, and units-to-disciplines. Examples:

- Org unit-to-PhD program relationship, CU-Boulder music/arts/humanities units, with HIST highlighted. In this diagram, depts are on left, PhD programs on right. CMLT is the comparative literature PhD program, with no corresponding department. Similar displays can show depts on left, institutes on right.

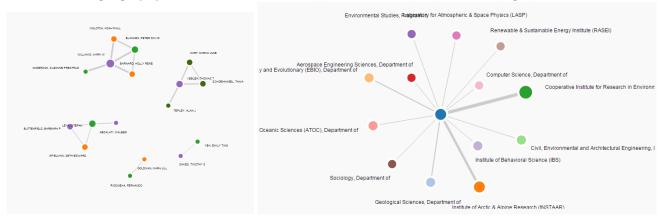- Relative size and overlap of people among music/humanities units

H. **Look at work of people in a unit** (e.g., number of books in 10 recent years) relative to others in same discipline; plot vs. years since terminal degree. Example: CU-Boulder program faculty vs. others in same (unspecified here) discipline. AA calls this "career progression."
- X-axis = Years since terminal degree.
- Y-axis = Z-score (standard deviations) for number of total peer-reviewed articles in 4 recent years, calc'd vs. all faculty in programs mapped to this discipline.   0 = average for whole discipline
- Blue bars = All faculty in the discipline. Those 1-10 years out produced fewer articles in the 4 years than those 10-25 years out.
- Dots = CU-Boulder faculty, color-coded by rank. Two with outstanding publication records "float" above others.

Academic Analytics facilities in portal version 2 and/or Tableau
- Locate individuals throughout the database by searching words and phrases in article titles and abstracts
  - AA sees potential use to identify teams of individuals, within the institution or not, suitable for addressing a funding proposal or other challenge. Current issues in development include characterizing faculty skills/niches, complementary vs. same-skill teams, portraying stronger and weaker records for people from different disciplines
  - Searches can be limited to own school, own state, own graduates, a discipline
  - Searches will potentially return articles themselves with notation of degree of match to search terms, and entire work profiles of individuals
- Identifying honors and awards held by individuals at peer schools in a discipline, in order to nominate or push CU-Boulder people for them. *Coming as a Tableau tool.*
- Studying journal and citation rates for a single discipline and set of institutions, and for CU-Boulder compared to others, with ability to specify a list of journals
- Detailed study of granting agencies for a single discipline and set of institutions. E.g., high grant departments in Asian Languages generally receive 90-100% of funding from the US Department of Education, 0-10% from NEH, but CU-Boulder's funding is 75% DOE, 25% NEH.

- Collaborations among institutions, units, and individuals. E.g., from the portal v2 (beta!), UCB geography collaborations (on articles) of individuals (left) and units (right)

Academic Analytics facilities untapped at CU Boulder
- Identifying scholarly work with CU-Boulder authors (or PIs, or award recipients) not in the Academic Analytics list of people for the release. This could help identify individuals in job classifications not reported to AA, who nevertheless had significant scholarly work.
- Use of Academic Analytics publications data in the CU-Boulder Publications Data Project.
- Use in promotion and tenure processes, reviewed and augmented by the individual.
- Characterizing CU-Boulder PhD recipients in terms of employment and work held in the AA database
- Characterize the work of the whole campus, or parts (e.g., geosciences), either absolutely (number of articles, books) or relative to others.
- Characterize the work of subsets of faculty – TTT, adjoint, assistant prof, etc. E.g., 20% of articles associated with INSTAAR faculty had no INSTAAR TTT on the author list.
- Characterize disciplines.
- Ask AA to form subsets or combinations of units at other schools, for comparison. We cannot do this ourselves with the data available to us.
- Consistencies across CU Boulder units vs. other schools. E.g., low dollars/grant
- Virtual units – e.g., Center Arts and Humanities, Jewish Studies

Academic Analytics gathers, stores, and links data

Scholarly work
Source: AA gathers and stores.
Millions of publications w DOI/citations; thousands of books, honors and awards, and federal research grants.
With details re agency, journal, etc.
In progress: Book chapters, patents.
Plus words & phrases extracted from titles/abstracts

Extracted by AA →

People not in AA db – in development

Matches – Made by AA

People – with name, faculty type, tenure status, rank/title.
Source: AA and/or school.

Metrics
developed by AA (e.g., FSPI (Faculty Scholarly Productivity Index) or standard index and/or users. Generally compare people/units w/i disciplines or broad fields.

Data flows to clients**

Full data
On all entities, ~2004 to present.

Comparative data
on volume, intensity, and per-capita amounts of scholarly work.
Annual release for 4-5 years of work (10 for books).
For people, PhD programs, org units, schools; comparable across all.

Rules – years, peer-reviewed, people, etc.

Affiliations –
People x unit combos
Source: AA and/or school

Disciplines or **taxonomy**
172 "level 1" disciplines
Source: AA, set ~2005
E.g. History, Civil Engin.
Rolled into 14 broad fields
- - - -
Why? Because patterns of scholarly work vary widely by discipline

Associations –
Unit x discipline combos
Source: Mostly by AA, some by school

Units = Groupings of people within a school, with local unit name
- **PhD programs**
- **Departments** (org units)
Source: AA and/or school

Scope of entities rep'd in AA data -- not restricted to AA clients

Schools: Universities offering one or more PhD programs
-- Over 400 U's, US

People: Employed at/ assoc'd w one of the schools, with expectation of scholarly work. Exact criteria vary by school.
-- 270k people

PhD programs
-- over 10,000 total
-- 1 to >100 per U.

Both called "**units**" in AA data

Departments
AA term for organizational unit, including departments, schools, colleges, institutes, etc.
-- 12,500 total
-- 1 to >100 per U.

30-40% of depts, PhD programs, and people are AAU

**Data are provided to clients only. AA releases no public data.
At present clients get name-identified data from other schools only rarely.

CU-Boulder (UCB) PBA, LMcC

AA_UCB_Ovv_Diagrams.vsdx -- 12/21/2016

CU-Boulder in Academic Analytics, with totals for 2015 release (Oct. 2016)
Key: AA work. CU-Boulder annual submission. Work involving both.

Scholarly work

Source: AA gathers and stores. Millions of publications w DOI/citations; thousands of books, honors and awards, and federal research grants. With details re agency, journal, etc. In progress: Book chapters, patents.
Plus words & phrases extracted from titles/abstracts

Matches –
All made by AA

Full data, comparative data

<u>Comparative</u> includes all 2015-16 tenured-tenure-track w/o libraries; research & adjoint professor-line & selected others not solely in institutes. <u>Full</u> adds RA, SRA, museum, institutes, some retired, & activity outside time window.

People – with name, tenure status, rank/title, fac type.

Criterion: All TTT (including start-in-spring) plus others with expectation of scholarly work including a) formal HR appt, paid/ not, as of 11-1-2015, plus b) additional fellows/faculty w no HR appt in institutes & some depts/programs. **Submit**
• All tenured and tenure-track (TTT) regardless of current title – 1,213. 52 are administrators, 38 in libraries.
• Research and adjoint professors (assistant, associate, full), minus ROTC adjoints. Total 91: 52 research, 39 adjoint.
• Specified research positions: Research associate (867), senior RA (115), museum-related (40), post-doc (HR position, 34).
• Fellows and equiv & faculty w no HR appt in institutes & some depts/programs (N=63)
Grand total 2,424 . . . 1,219 in comparative data

Metrics/Use
developed by UCB and by AA

To UCB
via portal, data mart, feeds, Tableau, and custom files. 2015 release received Oct. 2016.

Affiliations –
People x unit combinations, from appointments, tenure locus, and web lists for institutes & some depts, PhD programs. Average/person = 1.2 org units, 0.7 PhD program

AA disciplines (taxonomy)
N = 172 "level 1" disciplines. Set by AA ~2005. E.g. History, Civil Engin. Rolled into 14 broad fields.

Associations –
of PhD programs and organizational units with Academic Analytics disciplines. Initially listed by Academic Analytics; later augmented by LMcC to improve comparability to peer listings.

Units = Groupings of people
- **PhD programs** – 63 programs operating fall 2015 per UCB websites for prospective students. Includes interdisciplinary and joint programs, with multiple in Business, Education, and some departments. CMCI added for 2015.
- **Departments** – 67 org units including depts/divisions in A&S, Leeds, CMCI, and Engineering; other tenure-granting programs (e.g. environmental studies); other schools/colleges; all research institutes & museum.
- **Other** – Libraries and Herbst Humanities (in Engineering)