Academic Analytics (AA) at CU-Boulder

Update 2/4/2015 – 3/31/15

Lou McClelland, Planning Budget and Analysis – Institutional Research (PBA-IR)

PBA ref: L:\IR\Reports\GuideGR\AcAnalytics\AA_UCB_Update_201502.docx
Posted from http://www.colorado.edu/pba/peer/acanalytics.htm

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 35 AAU and 90 total subscribers plus over 70 employees working from locations throughout the US and UK. Website: http://www.academicanalytics.com/
- The AA database links over 270,000 individual faculty members in almost 400 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 holds a master license with Academic Analytics. Lou McClelland is CU-Boulder liaison to Academic Analytics.
- Visualizing what Academic Analytics does, and how CU-Boulder's submission (lists of people, departments, PhD programs, and affiliations with disciplines) fits in. *See last 2 pages or http://www.colorado.edu/pba/peer/Diagrams201503.pdf*

User meeting – and pre-and post- meetings – Jan 20 2015-- for past and/or potential campus users –

- Attendees, with PBA code for department
  - Russ Moore (IPHY), provost
  - (Stein Sture (CVEN), VC Research – Tricia Rankin (PHYS), AVC research
  - Jeff Cox (ENGL), AVC faculty affairs – over FIS/VIVO, and data on individuals' activities; over Academic Review and Planning
  - Steve Leigh (ANTH), A&S dean
  - John Stevenson (ENGL), graduate school dean
  - (Rob Davis (CHEN, Engineering dean)) – JoAnn Zelasko, asst dean administration
  - Jim White (GEOL), INSTAAR director, and Shelly Sommer, INSTAAR info/outreach
  - Myron Gutmann (HIST), IBS director, and Jane Menken (SOCY), outgoing director
  - Bill Lewis (EBIO), RAISE interim; CIRES
  - Liz Tomich, Faculty Information System (FIS) director (later: plus Don Elsborg, Nate Prewitt, and Vance Howard)
  - Steve McNally, sr AVC budget, finance, enrollment services
  - Rob Stubbs, director IR Institutional Research in PBA
  - Lou McClelland (LMcC), IR

- Avenues for activity explored at the user meeting
  - A: Maintain, enhance infrastructure
  - B: Characterize ourselves from submission process
  - C: Validate, understand, improve AA collection, matching, metrics
  - D: Characterize and compare individuals
  - E: Incorporate AA data and visualizations into FIS/Elements/VIVO
  - F: Characterize and compare depts., org units, PhD programs, groupings
  - G: Characterize whole campus vs. other schools – Publicize accomplishments
  - H: Grant opportunities
  - I: Explore interdisciplinary work – institutes, PhD programs, virtual, at UCB, not

Visit to campus by five Academic Analytics senior staff – Feb 23-24 2015

- Pete Maglione – CEO
- Mike Rohlinger, CIO and Matt Moericke, CTO – data structure, data mart, faculty profile system
- Dave Ramsey, director product and analytics – portal, reports, SAS user - *Could not attend, snow troubles on Virginia end*
- Matt Horvath, chief production officer - matching, discipline matches

Special sessions during/before campus visit

- **FIS, Faculty Information System**, with Liz Tomich, Don Elsborg, Nate Prewitt, Vance Howard
  - Preview, plans for summer 2015 campus visit on technical details of data flow between AA and FIS/Elements, on individuals and individual scholarly work.
    - Eventually much from FIS to VIVO, including honors and awards
    - Eventually two-way street
  - ORCID IDs, processes, AA use
  - See potential for both AA and UCB in Reach NC, North Carolina, with "fingerprints" and drill-down http://reachnc.org/ -> resource finder http://reachnc.org/
- **Interdisciplinary**
  - Does not fit well within AA initial design, for comparisons within disciplines
  - Identifying people and groups at UCB and elsewhere working on common topics – "Faculty Profile System" promising
  - Is there synergy – do people in formal interdisciplinary structures produce more research and collaboration than if not in such?
  - As much interest in characterizing ourselves as in comparing to other schools
  - More than institutes are interdisciplinary – PhD programs may be, and organizational units such as environmental studies
- **VC Research** – Stein Sture, Tricia Rankin – many topics, and session with economics PhD students working on models for annual internal and external reports on research
- **IR and submission to AA**. 90% of AA use goes thru IR!

Tidbits, activities

A. **Find CU-Boulder PhD recipients in AA database** (faculty at PhD-granting schools US/UK)
- With current discipline Anthro – N=25, 8 at AAU's
- With current discipline Physics – N=65, 14 here
- Have not matched to our degree records yet to get degree program or counts of those in/not in the AA db – AA will do name matching for us

B. Find depts. nationally with same discipline mix
- Our Chemistry and Biochemisty 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
- **The AA database of unit and PhD program *names* is a unique asset –** use with data mart or downloaded files.

C. **Characterize work by subgroups within a unit** – here, articles, INSTAAR Institute for Arctic and Alpine Research
- 1445 articles in 4–year period by 31 TTT and 1 research prof assoc'd with INSTAAR by appointment (15) or PI on grant (17) – unduplicated count of articles, 1 with 2 co-authors in INSTAAR counts as 1 not 2
- 360 *additional* articles (a 25% increase in total) in 4-year period with 4 Sr Research Assoc and 28 Research Assocs in INSTAAR as authors – These 360 articles had no INSTAAR TTT or research prof as a co-author. The SRA/RA people are submitted by CU-Boulder, and matched to work by AA, but are not in the AA "comparative database."

D. **Consistencies in metrics across units or disciplines** -- UCB across disciplines consistently has higher number of grants per person, higher grant dollars per person, but lower grant dollars per grant, than other schools.
- Evidently from many small grants, rather than few big ones
- Pinning down would take detailed grants data for other schools
- Several entities at CU-Boulder are addressing

E. **Get AA help on doing our submission** – we will submit list of websites used (e.g. Materials Science and Engineering PhD program faculty, available only on web) and AA will "scrape" and match to our people lists

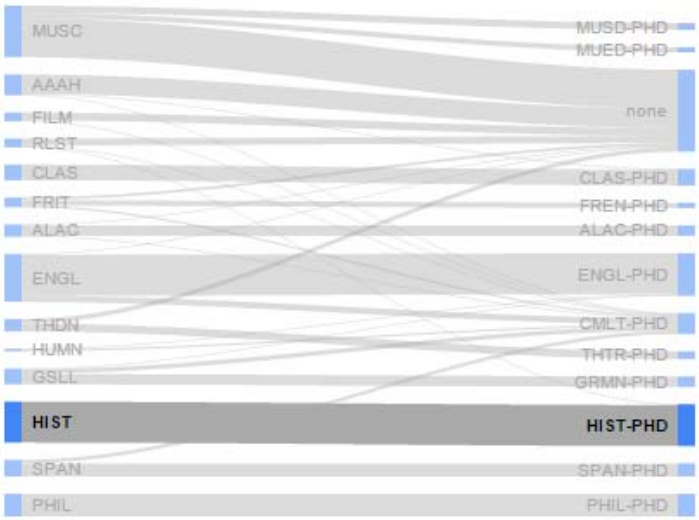
F. Asked AA to produce additional displays
- Distributions, not just percentiles
- "Demographics" of our depts/programs vs. others – size, pct full, years since degree, etc.
- Relationship between depts. and PhD programs, linked by people in common
- Profiles of a discipline; how activity patterns differ by discipline; the saga of some of physics authors with thousands of co-authors on some papers
- Profiles of an institution, depts. and programs, relationships among them

G. Make sure faculty counts make sense to chairs, deans, users *first*

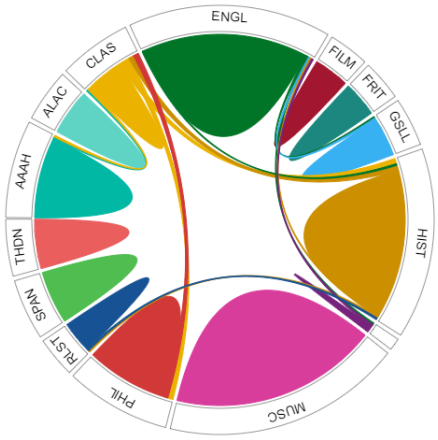
- Story of CU-Boulder Chemical Engineering – chair request – sent 18 files to chair at noon. At 3:00 chair says "Fantastic, will dig in to this." At 6:00: Faculty counts are all wrong, can't use anything, goodbye.
- I'd used programs (not depts) for all data – really wanted departments which HAD PhD programs, but was too lazy to pull. Chair interested in depts.
- In discipline chemical engineering, N of faculty in AA dept *usually* matches N in AA PhD program, or is 1-2-3 different (either way)
- At some schools the numbers differ radically. And at others the numbers matched but differed considerably from what the chair expected. All this invalidated all other data for the chair. Examples (in all, "AA" numbers may be from a submission, or from an AA web pull, or from an AA pull reviewed/revised by school)
  - Montana State, program N = 4 x dept N. Reason: 4 depts all feed one PhD program, "Engineering," associated with multiple disciplines. Can tell this from dept, program names.
  - Cornell, program N = 3 x dept N. Dept N matches web and chair expectation. Reason for high program number – probably from inclusive lists of faculty shown with research areas under Chem-Eng http://www.cheme.cornell.edu/cbe/academics/graduate/phd/research_areas.cfm lists 4:
  - Irvine – AA program = AA dept = 42, but chair says 15-16. Web lists 16 *ChEMS core faculty* plus 15-25 affiliated; AA reflects both. http://www.eng.uci.edu/dept/chems/faculty
  - Northwestern – AA's dept ~= program = core from web. AA picked up only core. http://www.chbe.northwestern.edu/people/faculty/index.html
  - Penn – AA dept = AA program = 13; chair says ~26. From web – chair must be including secondary, adjunct, emeritus. AA reflects only core.
- Lessons
  - User should review demographics of comparison programs/depts. *first* – names of programs/depts., counts. Revise/reconcile to OK before going on.
  - AA counts/lists sometimes include only "core faculty," sometimes include more. Users may want only core, or not.
  - Sometimes (at one school) the program and the dept in a discipline differ substantially
  - Users can review much more easily if they see names of faculty elsewhere
  - Users should also get a URL for a dept or program

H. **Characterize ourselves** using our annual submission to AA – eventually AA to do characterizations

- People, units, people-to-units, units-to-disciplines
- Examples
  - Org unit-to-PhD program relationship, CU-Boulder music/arts/humanities units, with HIST highlighted. In diagram, depts. are on left, PhD programs on right. Can do with depts on left, institutes on right.

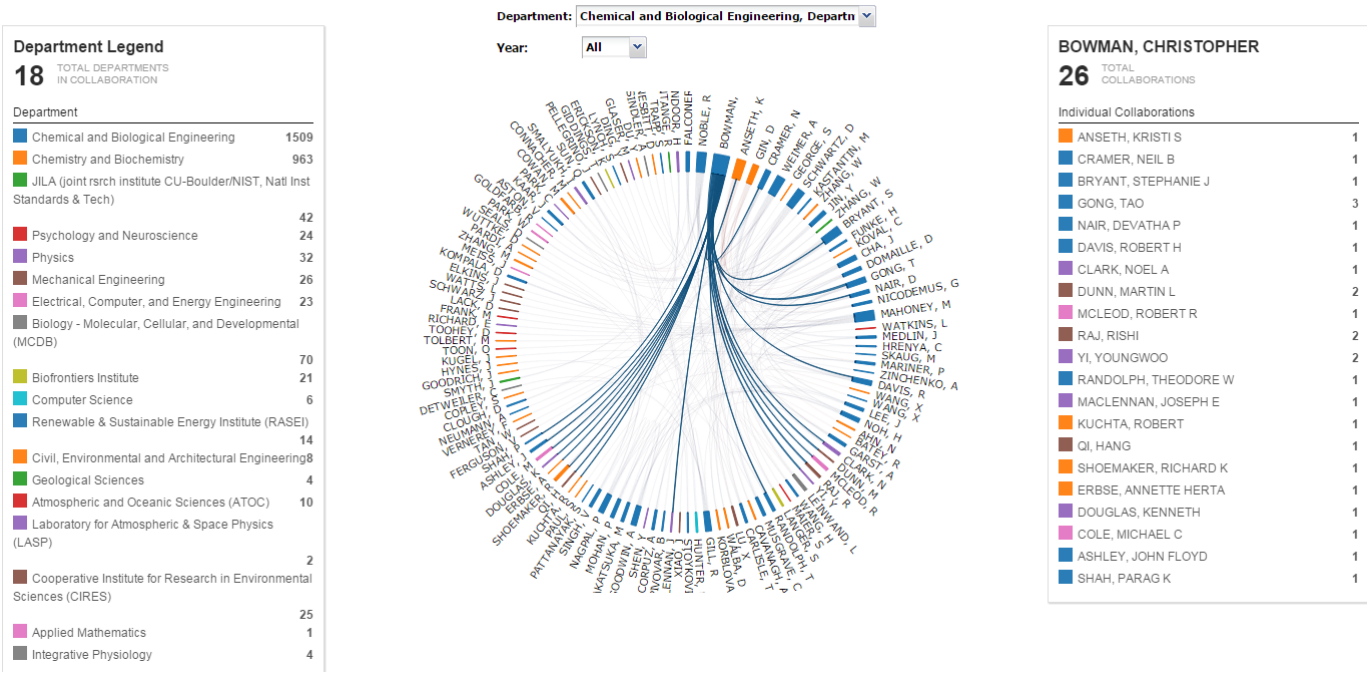


  - Relative size and overlap of people among music/humanities units

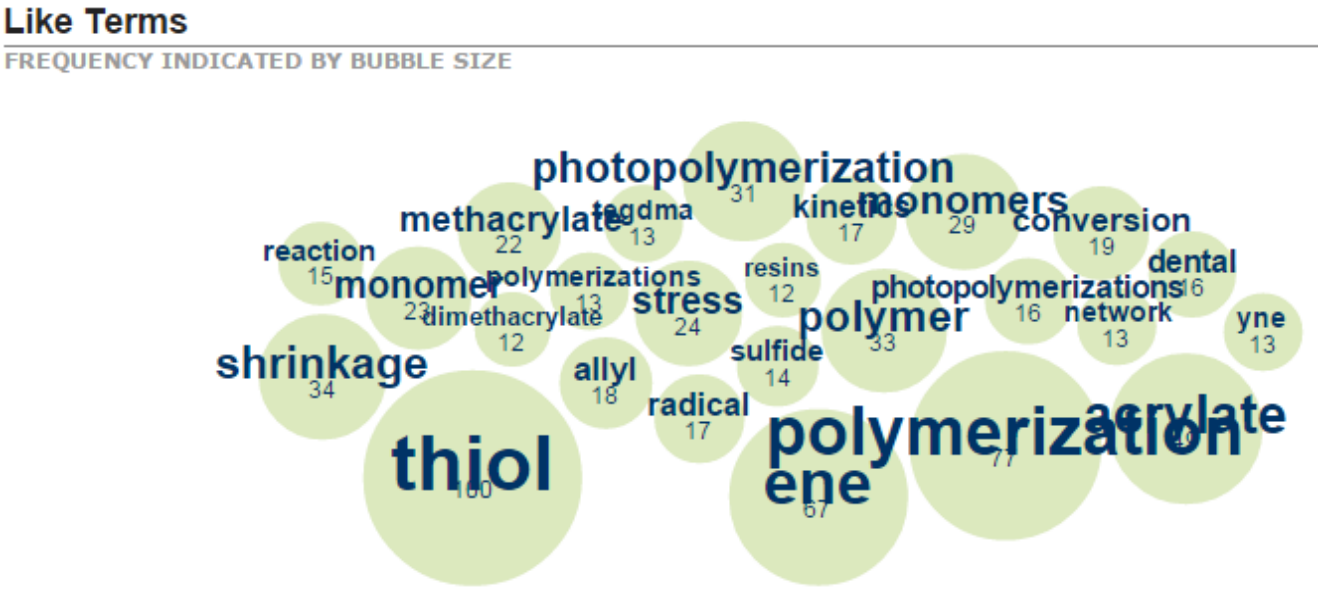


I. **Characterize a person's work.** Here, Chris Bowman, chem engineering.

- by depicting co-authorships. Are problems with people in multiple units (e.g. Anseth)

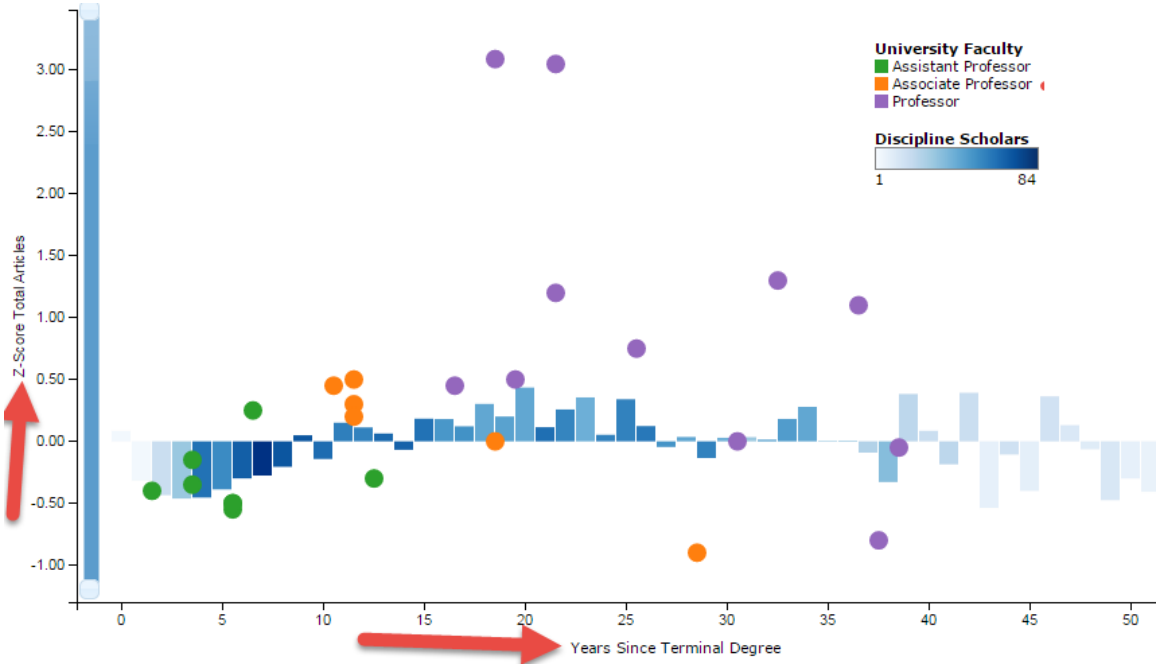


- with word clouds based on words in article titles and abstracts



J. **Look at work of people in a unit** (e.g., N of articles in 4 recent years, or citations) relative to others in same discipline; plot vs. years since terminal degree. Example: CU-Boulder program faculty vs. others in same (unspecified) discipline.

- X-axis = Years since terminal degree.
- Y-axis = Z-score for N of total peer-reviewed articles in 4 recent years, Z calc'd vs. all faculty in programs mapped to this discipline. 0 = average for whole discipline
- Blue bars = All faculty. 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. A couple with outstanding publication records "float" above others.



K. **The beta "faculty profile system" holds potential** for exploring people and groups whose work

- Examples:
  - "particle physics" (Rankin says results are reasonable)
  - Paleoclimatology
  - "ultracold molecules" "ultracold atoms"
- Can limit to own state, own institution, own PhD graduates
- VIVO research statements can give overviews
  - Tim Curran (psych/neuroscience) "brain electrical activity"
- Useful if very specific words characterize the work
- Future: Drill down to publications; more on units; find groups

Academic Analytics gathers, stores, and links data

Scholarly work

Source: AA gathers and stores.
Millions of publications/citations; thousands of books/chapters, honors and awards, and federal research grants.
With details re agency, journal, etc.

Plus words & phrases extracted from titles/abstracts

Extracted by AA

People not in AA db – in development

Matches –

Made by AA

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

Source: AA and/or school.

Full data

On all entities, ~2004 to present.

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

Metrics

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

Comparative data

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

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 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/UK

People: Employed (at one of the schools), with expectation of scholarly work. Exact criteria vary by school.
-- 230k people

PhD programs
-- 9,800 total
-- 1 to >100 per U.

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

Departments
AA term for organizational unit, including departments, schools, colleges, institutes, etc.
-- 12,700 total
-- 1 to >100 per U.

CU-Boulder in Academic Analytics, with totals for 2013 release

Key: AA work. CU-Boulder annual submission. Work involving both.

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

Matches –
All made by AA

Full data, comparative data
Comparative includes all Nov 2013 tenured-tenure-track, research & adjoint professor-line. Full adds retired, left, RA, SRA, museum, & activity outside time window.

People – with name, tenure status, rank/title.
Criterion: Employed with expectation of scholarly work
Must have formal HR appt, paid/not, as of 11-1-2013.
From these, select
- All tenured and tenure-track (TTT) regardless of current title – 1,188. 45 are administrators, 38 in libraries.
- Research and adjoint professors (assistant, associate, full), minus ROTC adjoints. Total 68: 33 research, 35 adjoint
- Specified research positions: Research associate (813), senior RA (112), museum curator adjoint (32), museum associate (17). RA's not rpt'd as affil'd w PhD programs.
- Grand total 2,230 . . . Full data has 3,015.

Affiliations –
People x unit combinations,
from appointments, tenure locus,
PI status/FRPA release to selected units for TTT only,
and web lists of faculty in PhD programs.
Average/person = 1.9; range 0-9

Metrics/Use
developed by UCB and by AA

To UCB
via portal, data mart, feeds, and custom files.
2013 release received Nov. 2014.

Units = Groupings of people
- **PhD programs** – 58 programs operating 11-1-2013 per UCB websites for prospective students. Includes interdisciplinary and joint programs, with multiple in Business, Education, and some departments
- **Departments** – 63 org units including 38 A&S and Engineering departments, 11 research institutes, 6 current/former schools/colleges, 4 Business divisions, and 4 tenure-granting programs (e.g. environmental studies)
- **Other** – Libraries and Herbst Humanities (in Engineering). Not in comparative data.

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. Research institutes are generally associated with over 10 disciplines each.