

The Faculty Bibliography Project at the NYU School of Medicine

Dorice Vieira, Richard McGowan, Aileen McCrillis, Ian Lamb, Catherine Larson, Theodora Bakker, Stuart Spore

Vieira, D, McGowan, R, McCrillis, A, Lamb, I, Larson, C, Bakker, T, Spore, S. (2014). The Faculty Bibliography Project at the NYU School of Medicine. *Journal of Librarianship and Scholarly Communication* 2(3):eP1161. <http://dx.doi.org/10.7710/2162-3309.1161>

© 2014 by the author(s). This open access article is distributed under a Creative Commons Attribution License, which allows unrestricted use, distribution, and reproduction in any medium, providing the original author and source are credited.

JLSC is a quarterly journal sponsored and published by Pacific University Library | ISSN 2162-3309 | <http://jlsc-pub.org>

The Faculty Bibliography Project at the NYU School of Medicine

Dorice Vieira *Clinical Outreach & Graduate Medicine Education Librarian, NYU Health Sciences Library*

Richard McGowan *Research Librarian, NYU Health Sciences Library*

Aileen McCrillis *Research Librarian/User Experience Librarian, NYU Health Sciences Library*

Ian Lamb *Web Designer, NYU Health Sciences Library*

Catherine Larson *Web Services Librarian, NYU Health Sciences Library*

Theodora Bakker *Assistant Director, Knowledge Informatics, NYU Health Sciences Library*

Stuart Spore *Head, Bibliometrics & Senior Systems Advisor, NYU Health Sciences Library*

Abstract

INTRODUCTION This paper describes the development of the New York University Health Sciences Library's Faculty Bibliography. **DESCRIPTION** Since 2000, the NYU Health Sciences Library's Faculty Bibliography project has systematically tracked publications of the NYU School of Medicine faculty. The project has grown to a significant institutional service making prominent contributions to the School of Medicine's public web presence and to advanced productivity metrics. Migrating from Gopher to EndNote to MySQL, the Faculty Bibliography harvests data from multiple abstracting and indexing resources and uses sophisticated quality assurance methodologies. At present the Faculty Bibliography tracks over 228,000 publications of well over 13,000 faculty, including faculties of the NYU Colleges of Dentistry and Nursing. Both technical and social engineering aspects of the project's success are discussed; the project's role in deepening professional contact between the Library, clinical and research faculty, and School administration is stressed. **NEXT STEPS** The Library currently envisions broadening coverage to include faculty engaged in scientific and medical publishing from other schools and colleges at NYU. We also anticipate significant improvements in the project's methodology once the ORCID initiative takes root.

Received: 04/22/2014 Accepted: 06/10/2014



© 2014 Vieira et al. This open access article is distributed under a Creative Commons Attribution 4.0 Unported License, which allows unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

INTRODUCTION

Academic medical libraries have long been interested in tracking the publications of medical school faculty. Over the years libraries have developed in-house systems to this end. Often beginning as collections of reprints, these efforts enjoyed varying levels of institutional support and usefulness.

Motivations for undertaking faculty publication projects have remained more or less consistent over the years, viz,

- i. to document institutional history by establishing an authoritative record of the institution's scholarly output;
- ii. to contribute to the better understanding of institutional strengths and weaknesses; to support assessment of departmental and individual faculty performance, including justification for promotions, budgets, etc.;
- iii. to facilitate cross-discipline research and identify opportunities for collaboration;
- iv. to support preparation of grants and to assist in the justification of other funding;
- v. to support and stimulate faculty publishing effort, including identification of publishing opportunities, manuscript preparation, etc.;
- vi. to support library services, including collection development, liaison efforts, and general reference; and
- vii. to contribute to library and institutional PR including production of printed bibliographies, online faculty profile pages, etc.

Many difficulties are associated with such projects. It has never been easy to comprehensively identify faculty publications and, in larger institutions at least, it can be nearly as difficult to reliably track the comings and goings of eligible faculty. While most such projects have concentrated on current faculty publishing, a few (Mayo Clinic; Illinois State University) aim for historical completeness (Schwartz & Stoffel, 2008). Whether comprehensive or selective, such projects required considerable investment of scarce time and attention. It is notoriously difficult to get faculty to actively contribute; on the other hand, it is remarkably easy to offend faculty with omissions and other inevitable errors.

Despite the problems, there have been notable successes. One of the oldest sustained efforts is the Mayo Clinic's "Author Catalog," which dates back to 1907 and remains an important effort today.¹ Mayo was also an early adopter of automation, with computerization beginning in 1966. (Key & Sholtz, 1973). Indeed, computerization offered to make such projects more achievable. As far back as the mid-1970s there are published discussions of automating such projects, often describing technologies that seem quaint today, for example a 1976 *MLA Bulletin* article describing a Wake Forest project employing Magnetic Tape/Selectric Typewriters (Lee, Gratz, & White, 1976). Faculty publishing projects were a natural for the successive waves of microcomputer automation in the 1980s and early 1990s; *Endnote* and *Procite* were perennial choices, though many other software packages were also employed (Anderson & Monroe, 1984; Bai & Kelly, 2000; Fenichel, 1990; MacCorkle, 1991; Marsalis & Kelly, 2004, McKee & Feng, 1979; Potter, 1987).

But even with the help of automation, publication harvesting can be tedious and time-consuming, making the effort required difficult to balance against the sometimes theoretical benefits. All of these considerations tend to discourage wide adoption. In a 1995 *Medical Reference Services Quarterly* article, Mansheim and Thompson (1995) describe a survey of 172 academic health sciences libraries which found that only 32 maintained faculty publications databases. The quality and utility of such products no doubt varied widely depending on levels of local support.

BEGINNINGS AT NYU SCHOOL OF MEDICINE

At the NYU Health Sciences Library compilation of monthly faculty publications lists began in the late 1980s. Citations were harvested from MEDLINE, PsycINFO, Embase, and other abstracting and indexing (A&I) databases using a heuristic search algorithm based on institutional affiliation. Citations were managed in *EndNote*. Faculty were encouraged to submit citations not covered by those sources. The end-product was regularly printed lists of bibliographic citations which were distributed with faculty council minutes and in annual compilations. No authoritative list of faculty was used; there was no mechanism by which one could view only the citations authored by a single faculty member or those belonging to a particular department.

¹ <http://www.mayo.edu/research/publications>

After a brief interval as a Gopher resource, the Faculty Bibliography went on the web in 1994. The *EndNote*-generated annual flat files were simply posted on the Library's (then new) website. The web version was still not readily searchable, nor was it possible to generate useful metrics from the reports. Predictably, there were no links from listed publications to full text or other content, licensed or otherwise. It is not possible now to determine how heavily the web-based Faculty Bibliography was actually used during its first six years, but it is safe to assume that usage was not heavy (Vieira & Faraino, 1997).

During this period the Library operated the Faculty Bibliography largely in a vacuum. The Faculty Bibliography was a "silo"; that is, it was completely independent of and incompatible with other institutional efforts tracking faculty activities. There were two such relevant efforts: the "Faculty Resource Catalog" (FRC) and "NYU Biosketch." The FRC was a Library sponsored project to track faculty research interests and expertise (expressed in MeSH) with the intention of facilitating collaboration and sharing of scientific resources; the FRC ran on proprietary Innovative-Interfaces software. "NYU Biosketch" was a School of Medicine/IT sponsored system of online faculty profiles, running on Oracle. The Biosketch system included both a public web presence and a back-end (called "MyInfo") which enabled faculty to correct and update their profile data. NYU Biosketch included an option that allowed faculty to post lists of their publications along with their other data. Like other self-managed systems of this type, data quality varied wildly from one faculty member to another. Unlike the FRC and the Faculty Bibliography, the Biosketch system was closely integrated with the Medical Center's authoritative online directory.²

The flat-file Faculty Bibliography was little noticed by faculty or hospital management. The publication lists were used offline to prepare for the Liaison Committee on Medical Education (LCME) accreditation and annual departmental research assessments, but most faculty were unaware that their publications were being tracked at all. At best the Library was a decidedly minor player in management decision making.

² For a representative example of a NYU School of Medicine Biosketch page see <http://www.med.nyu.edu/biosketch/llinar01>

THE NEW FACULTY BIBLIOGRAPHY

During the year 2000 local institutional and industry changes coincided to influence the future of the NYU Faculty Bibliography.

By 2000 changes to the organization, financing, and delivery of health care had forced leaders of academic medicine to recognize their need for much more objective and verifiable information to support planning and administration. Indeed these changes were so profound that the Medical Education Panel of the AAMC's Mission-based Management Program characterized them as threatening "to undermine the financial status and traditional roles of medical schools," putting their very academic mission into doubt (Nutter et al., 2000). The day of rigorous assessment metrics was dawning.

At the same time medical informatics was maturing rapidly, resulting in the public availability of sophisticated discovery tools and unprecedented ease of access to the medical literature. The NCBI, which had rolled out PubMed three years earlier, was about to unveil its Entrez utilities, which would greatly facilitate automated data harvesting and indexing. Meanwhile, at NYU, the Medical Library had recently established its own in-house systems unit, with a small but skilled staff of systems administrators and integrators, enabling the Library to initiate and maintain its own informatics projects.

The Library systems staff became interested in the Faculty Bibliography as a demonstration project for MySQL technology. During the Spring of 2000 the data was ported to MySQL and the new database went live on the web in July. Library systems handled all technical aspects of the Faculty Bibliography project, cooperating with Medical Center IT staff when necessary for integration. The in-house development process was nimble and protected the effort, small as it was at the beginning, from being stalled by competition with more prominent projects.

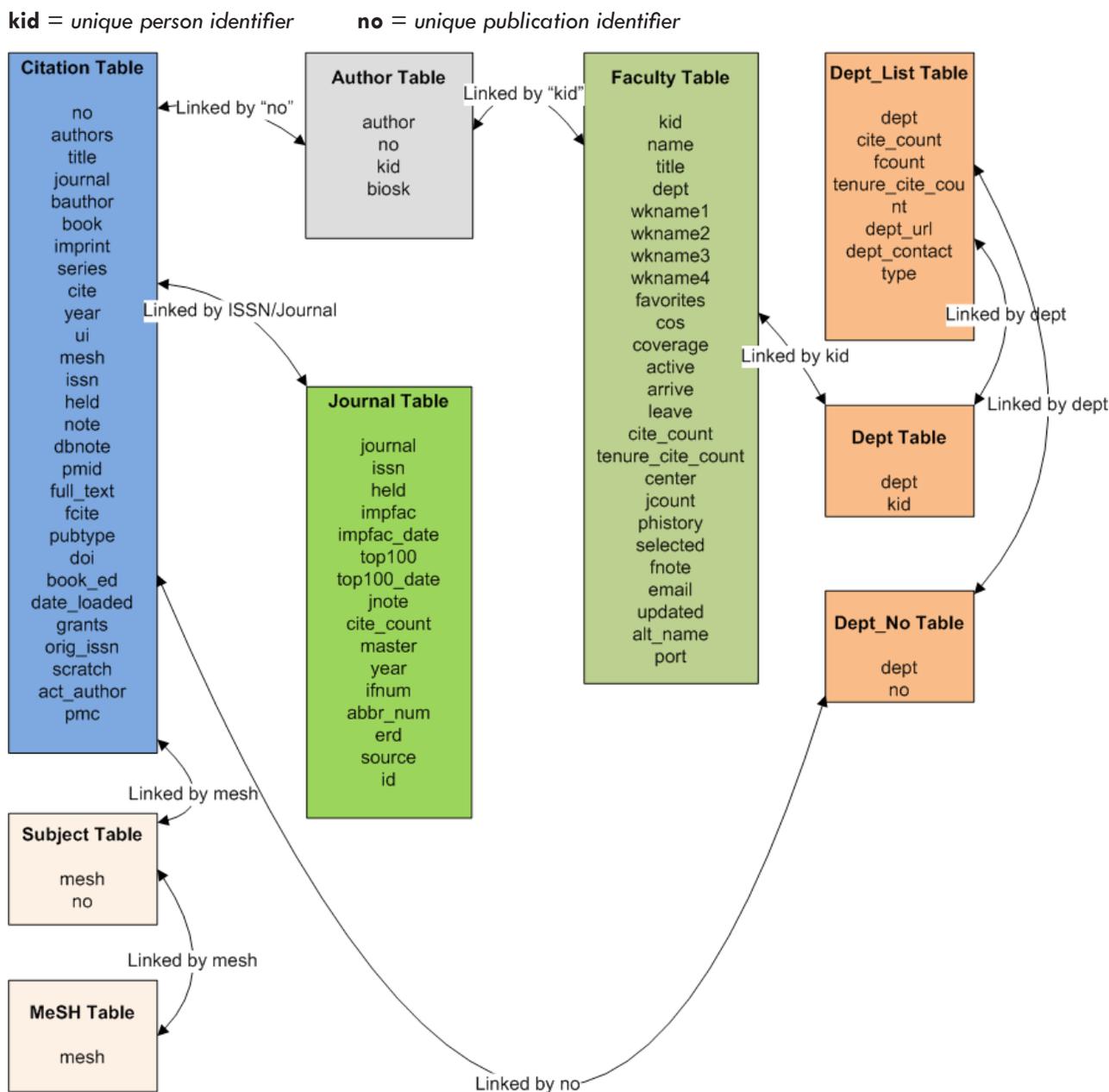
The new database required a complete rethinking of the publication data and how it was managed. Instead of a simple list of citations, the new database consisted of separate, linked tables of faculty and citation data (see Figure 1). This abstraction provided a great deal of power, but it also entailed matching faculty against author names, a challenge in its own right. From the beginning faculty data was carefully coordinated with the NYU Medical

Center’s authoritative online directory, the first step in breaking down silo walls.

As the Faculty Bibliography went live, the Library made a number of policy decisions that were to have far reaching consequences:

1. The Faculty Bibliography would be comprehensive; that is, it would aspire to track all formally published works by NYU School of Medicine faculty current and past, omitting only residents and instructor-level faculty (who were allowed to opt in if they wished).
2. Every effort would be made to integrate the Faculty Bibliography with other institutional systems.
3. There would be a sustained effort to develop systematic publication metrics at the individual, departmental, and school levels.
4. Faculty initiated additions and corrections would normally be handled within 2 business days of receipt. In addition, faculty could submit their current CV and have the listed publications verified and added.
5. The Faculty Bibliography would only track formally published works. It would not include “in press”

Figure 1. Original Faculty Bibliography Data Model, 2000-2012



citations or the like. Nor would it include posters or meeting abstracts unless they had been formally published.

6. Whenever possible the Faculty Bibliography would provide links to the full-text of the works included.³

The Library then undertook a systematic effort to retrospectively harvest relevant works from an expanded set of abstract and indexing sources (Table 1). Search algorithms were carefully reviewed and expanded in an effort to make the results more comprehensive and, in so far as possible, to exclude irrelevant citations. Because authors do not always supply relevant affiliation information (authors may have multiple simultaneous affiliations) and because affiliations are depicted inconsistently (i.e., without any attempts at standardization), such algorithms are inherently unable to capture all relevant publications.

Table 1. NYU Faculty Bibliography Data Sources

Source	From	number in DB (total=228478)	% in DB
MEDLINE	Beginning	178659	78.20
PsycInfo	Beginning	3261	1.43
Embase	Beginning	2205	0.97
CINAHL	Beginning	222	0.10
Web of Science	Beginning	23034	10.08
INSPEC	2009	143	0.06
Proquest	2008	7066	3.09
WorldCat	Beginning	1767	0.77
Bioabstracts	Beginning	59	0.03
Biosis	Beginning	1615	0.71
CancerLit	Beginning, defunct 1998	24	0.01
Global Health	2005	9	0.00
AIDSLine	Beginning, defunct 1999	32	0.01
Original	Beginning	8533	3.73
DOSS	2013	55	0.02
Scopus	2013	823	0.36

Given the acknowledged shortcomings of the affiliation algorithms, they were supplemented with searches by

³ This was first implemented systematically in the late summer of 2003 with openurls pointing to the main NYU SFX server; late the next summer openurls were repointed to the recently established Medical Center SFX server.

individual faculty names (uniqueness permitting) and by co-author network searches. As a result the database grew quickly; starting with approximately 4,000 citations from EndNote, the database doubled in size within 4 months, doubled again 11 months later, and again 18 months after that (see Figures 2 & 3, following page).⁴

A parallel effort was begun to retrospectively add ex-faculty to the Faculty Bibliography database, working from scans of paper files supplied by the NYU School of Medicine Office of Faculty Records. This effort, which is still ongoing, has resulted in the addition of over 5000 ex-faculty to the database.

In 2001 the Library arranged with Medical Center IT to replace the underutilized self-managed faculty publication stanzas on Biosketch pages with daily feeds from the Faculty Bibliography. The presence of Faculty Bibliography data on Biosketch pages lead to much higher profile for the data. Usage rose dramatically. By Spring 2004 we found that well over 90% of Faculty Bibliography usage was via the Biosketch pages (out of about 11,700 hits a month), a pattern that has held steady over time. This turned out to be the decisive step in making the Faculty Bibliography an institutional success.

To support inclusion of Faculty Bibliography data on the NYU Biosketch pages, the Library created a web service, optimized to facilitate embedding focused lists of faculty publications on web sites. Some academic departments and other units at the School of Medicine adopted the web service to add dynamic feeds of their faculty's publications to their web sites; other units created links from their home pages to the Biosketch publications pages for their faculty. Either way the Faculty Bibliography data were increasingly exposed, and the Library's role in its curation came to be better understood.

Soon after the migration to MySQL the Library began developing canned faculty and departmental reports which graphed publishing activity over time and were available on request to department heads and others. These rather basic reports evolved into a suite of more sophisticated reports created for the 2006 LCME

⁴ Regrettably, much original documentation relating to the Faculty Bibliography project, including database population statistics, were lost in flooding caused by Hurricane Sandy in November 2012. Numbers used in this account have been pieced together from surviving paper and digital documentation.

process. Soon there was modest but steady demand for bespoke reports from various quarters. Often these were motivated by the need for grant documentation and for annual departmental reviews, but there was also demand for one-off studies of particular issues, such as cross-departmental collaboration.

With growth and confidence the project dimensions widened. In January 2007 the database and scripts were cloned to support tracking of the NYU College of Dentistry faculty publishing and in June 2009 the College of Nursing was also supported with its own database.

Meanwhile, the Faculty Resource Catalog stagnated. Policies relating to the FRC did not evolve in parallel with the Faculty Bibliography. Indeed, running on proprietary software, there was no way that the FRC could be made to truly integrate with other systems. Nor was there an easy mechanism for faculty to discover and make corrections to their FRC listings; FRC had to rely on emailed requests for updates, which were often ignored or treated as spam. Lack of integration led to a dismal mind-share situation. By June 2004, the Faculty Bibliography was averaging about 386 hits a day, while the FRC got just over a tenth of that, about 48 hits a day. As time

Figure 2. Publications Tracked

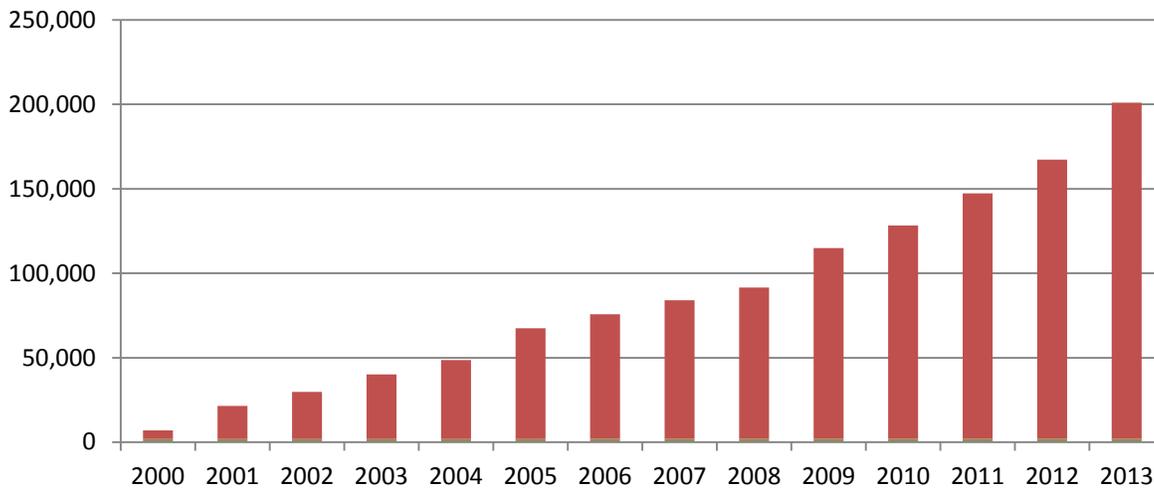
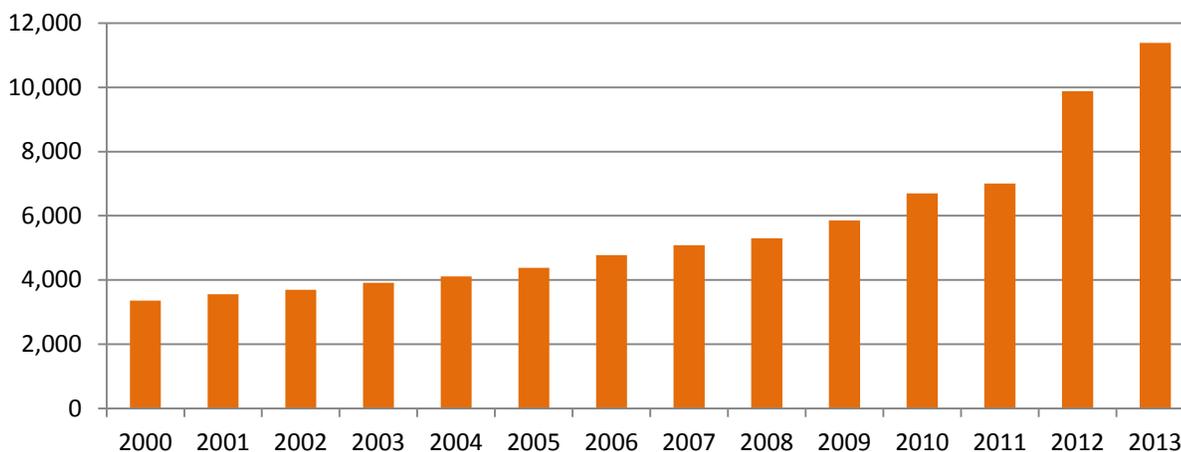


Figure 3. Faculty Tracked



went on usage continued to decline, so that by the next summer, the FRC was hardly used at all (3.8 hits a day) and the decision was taken to discontinue updating it.⁵

Expanding Faculty Bibliography usage led to increasing complaints about missing and mismatched citations. Late in 2000 the Library had added an online form for faculty to submit additions and corrections. In 2001 this form was replaced by integration of Faculty Bibliography data into the institutional Biosketch update tool (“MyInfo”), giving the faculty one stop shopping for maintenance of their online institutional data. The MyInfo tool was also used to enable faculty to pick and choose which of their publications would appear on their NYU Biosketch site.

The flow of submissions has been fairly constant from 2000 on, although over time there has been a slight decline in volume as better quality assurance measures have been implemented. Between June 2005 and June 2012 the Library fielded about 2,800 change requests, an average of a little over one a day. At the same time, there was a steady, if small, stream of over-the-transom requests for additions and changes. Prompt attention to all requests, including explicit acknowledgments, made a major difference in building trust with the faculty and, of course, contributed to improved database quality.

At the same time, the Library established a policy of allowing faculty to submit their CV to identify and add missing works, a particularly helpful option in regard to poorly controlled literature such as chapters. It was gratifying to see that, although the numbers of CVs processed was never large (one or two a month), it was a rare CV that did not omit legitimate citations that the Faculty Bibliography already contained.

In 2012, with more than ten years of experience to draw on, the Faculty Bibliography was completely rebuilt from the ground up. The new version, called “FB20” in the library back offices, included a new, sophisticated, data model and a superior and more nuanced name matching algorithm (Figure 4, following page). College of Nursing and College of Dentistry data, which had been in separate

⁵ The FRC was taken down for good in 2008, but the wish for a tool to facilitate collaboration did not die with it. In 2008 the Library and Medical Center IT co-created a tool called “Find-A-Researcher” based on Faculty Bibliography data. This tool remains in use, but cannot be said to have thrived. Discussions are underway on ways and means to recast or replace it.

databases, were consolidated with School of Medicine data, providing a single pool for all tracked data. At the same time, policies were revisited and a decision was taken to systematically include all faculty, not just those of a certain rank. The improved name-matching algorithm reduced the need for some aspects of the earlier quality-control effort, which were phased out in favor of more focused attempts to involve the faculty in the accuracy of their own records. Beginning in October 2012, faculty received nightly emails whenever new citations were added to their listings, providing them with an opportunity to catch and report errors early. Since then more than 10,000 notifications have been sent, averaging about 140 per week. We have received and acted on over 500 corrective responses.

In 2013 the web service that had been in use since 2001 was replaced by a full-featured, publicly available API.⁶ The API supports publication list discovery and retrieval by a wide range of search criteria. In addition to the usual author, title, and keyword searches these included grant number, academic title and department, gender, appointment start and end dates, and others. It also offers a rich set of result limiters (by chronology and publication type) and output formats (multiple XML formats, JSON, Endnote, and RSS).

Later in 2013, the Faculty Bibliography data was incorporated into the Medical Center’s data warehouse (via nightly data dumps), making the raw data available for use in institutionally sponsored projects outside the Library. Simultaneously the Faculty Bibliography was modified to synchronize its faculty data with that in the Medical Center data warehouse, making for more accurate faculty identification and greatly reducing time devoted to faculty record maintenance.

As of this writing, the Faculty Bibliography indexes 228,490 publications. The database is predominantly journal literature (96%) and it tracks publications from nearly 10,000 distinct journals. Over the last 18 years we averaged about 6250 publications per year. There is a distinct latency between the nominal year of publication and entry into the database; as a rule of thumb, we consider that a year’s publications are not “complete” until mid-summer of the following year. The database tracks 13,618 faculty (6,775 of them active faculty). A fairly large portion of faculty have no publications listed at all

⁶ <http://library.med.nyu.edu/api/publications>

(about 20%), most of them clinicians and younger faculty. A public Faculty Bibliography search interface is available at <http://hsl.med.nyu.edu/faculty-bibliography-search>.

IMPLEMENTATION PROBLEMS & ISSUES

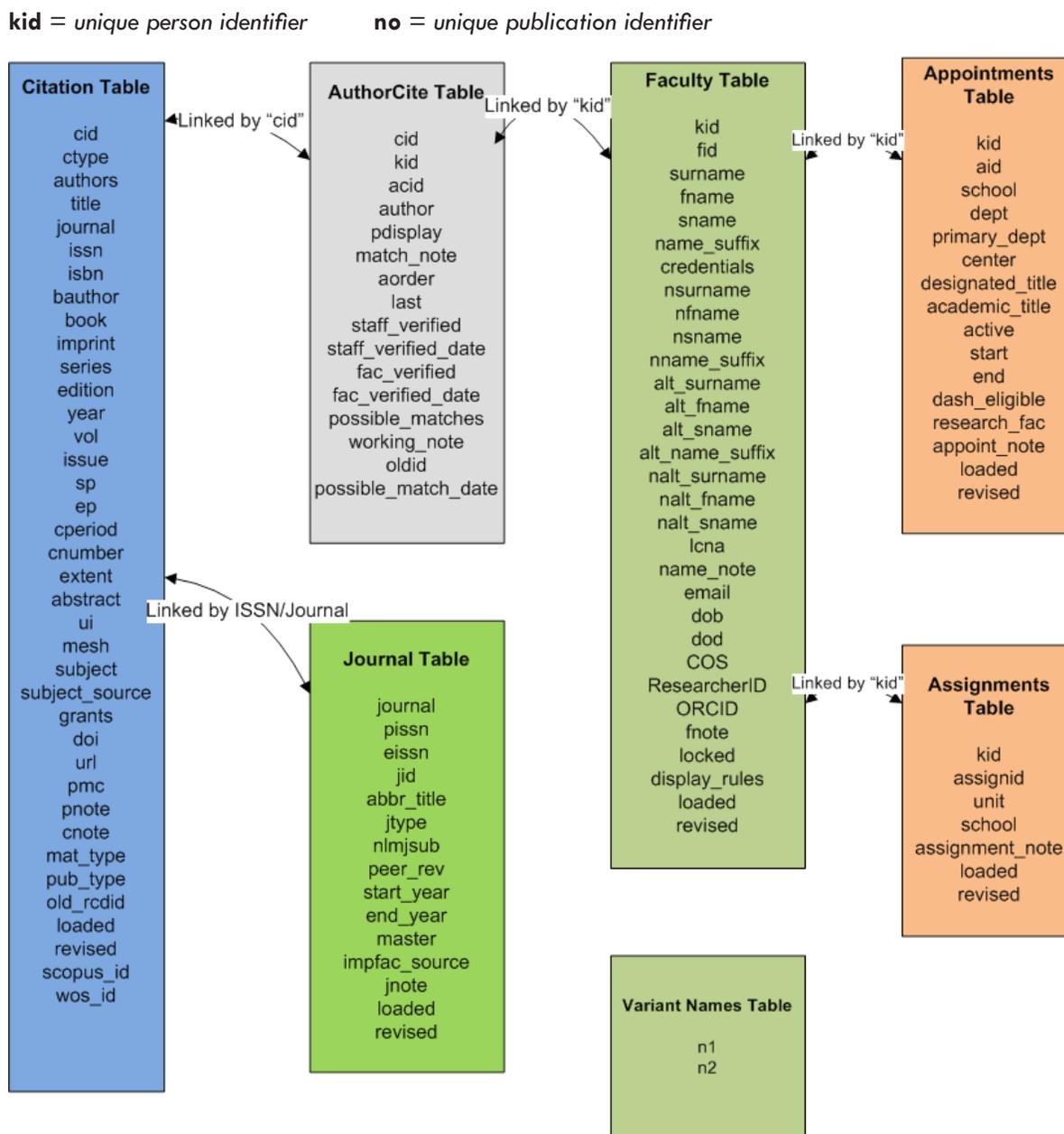
(1) Missing publications

One early lesson learned was that we had been seriously under representing the quantity of faculty publishing. For example, in 1997 the flat files listed just over 500 publications each for 1992 and 1993. By late 2000, the numbers for those years had increased to over 800 pub-

lications each. Since then we have systematically added ex-faculty and fugitive publications, so that today the Faculty Bibliography lists 1955 relevant citations for 1992 and 2189 for 1993. Because there are still departed faculty to add and because even now we occasionally stumble upon omitted publications, we expect those numbers to continue to increase, if not by much.

We concluded that no single data source consistently contains all relevant publications and even a rich mixture of A&I databases will never reveal all faculty publications. It is necessary to provide a means for adding “original” publications (i.e., publications which appear in none

Figure 4. Current Faculty Bibliography Data Model, 2012-present



of the A&I data sources). These are frequently “minor” works (which may nevertheless be important to the author), but not always; some publication types are simply more accessible than others. In particular books and chapters, audiovisual works, and software present marked harvesting difficulties. These inconsistencies may well affect disciplines unequally. For example, in our experience psychiatrists and ethicists are more likely to publish chapters than other faculty; it seems probable that the prestige of such publications in their fields is similarly higher.

(2) Deduplication/Source Management

The health sciences literature is large, but A&I databases have a pronounced tendency to overlap in their coverage. Works published in high-prestige journals are very likely to be indexed in multiple A&I sources. As these works are harvested it is necessary to eliminate duplicative citations. The first step in accomplishing this is to settle on a hierarchy of source preferences. MEDLINE quickly became the primary source, due to its preeminence in the field, but also because the superb suite of Entrez programming utilities which made maintenance of MEDLINE records far easier than those from other sources (Sayers, 2009). As of this writing MEDLINE accounts for about 78% of Faculty Bibliography citations. As citations are harvested from non-MEDLINE sources,⁷ citation metadata is extracted and matched against MEDLINE. Those that match are loaded directly from MEDLINE; only those that do not match are loaded from the relevant source. Because no matching algorithm is ever completely effective, periodic sweeps of the database are run to identify and remove fugitive duplicate citations.

(3) Name Matching

Name matching is easily the single most troublesome problem in managing a faculty bibliography project. Until only a few years ago, the norm for A&I databases was to identify publication authors by last name and initials, often only the first initial. This has the obvious shortcoming of allowing for a great deal of ambiguity. Even full names are notoriously subject to ambiguity, variation, and change over time.

⁷ SDIs are run at regular intervals, with the larger contributors (Web of Science, Embase, Psycinfo) scheduled more frequently than smaller contributors.

At NYU, official faculty names are supplied by Human Resources. But there is no guarantee that a faculty member will choose to publish under the official version of their name. Nor is there any guarantee that the faculty member will consistently use one and only one form of their name, even apart from straightforward name changes. Moreover, the vagaries of the publishing/indexing process provide lots of opportunities for transcription errors and confusion.

Complicating the matching problem further is the existence of name collisions within the roster of faculty. As the number of faculty grows, name collisions become more frequent and with them the risk of troublesome mismatches. Obviously this can be a severe problem in matching the older literature. At this writing, our faculty database has four Kim Ds, four Kim Ss, four Lee Ms, and numerous cases of two and three faculty member collisions. But even with full names, collisions occur; we have, among others, two Robert J Allens (both in plastic surgery), two Richard J MacKools (both ophthalmologists), two Adam D Browns (both in psychiatry), and three David E Cohens.

By way of illustration, when searching for “Robert L Smith” in the literature, how can we ascertain whether the author is (a) the NYU internist, (b) the NYU otolaryngologist, (c) the Australian neurologist, (d) the Stanford orthopaedist, or (e) the LSU biochemist? Publication source, institutional affiliations, and co-author networks can help clarify authorship, as can examining the works themselves (which may or may not include full author names or useful affiliation data), but in some cases it may be impossible to quickly and positively identify who wrote a particular article, particularly in the older literature.⁸

When the Faculty Bibliography first confronted this problem our approach was to build a very crude name matching routine based on name normalization. Both incoming and faculty names were normalized and then matched, with the first match winning the assignment.⁹ It was recognized that this would result in mistakes, but it was accepted as a necessary evil in the absence of a readily attainable alternative.

⁸ For a recent discussion of disambiguation strategies see Johnson et al (2014).

⁹ For example, the incoming name “Smith, John R” was normalized into the match strings “SMITHJOHNR”, “SMITHJOHN”, “SMITHJR”, and “SMITHJ” which were then compared to similarly normalized faculty names in order of specificity.

In fact sledge-hammer name matching did cause plenty of problems. These problems retarded, but did not ultimately prevent, faculty acceptance of the Faculty Bibliography. The chief reason for the ultimate acceptance was the sustained application of very aggressive quality control measures. Early in 2012, the algorithm was replaced by a much more nuanced and less-error-prone matching algorithm which supports AKAs, typical name variants (e.g., “Bob” for “Robert”), and other refinements. The new algorithm nevertheless allows through some mismatches, so quality-control measures remain important.

(4) Quality Control/Faculty Involvement

The problems discussed above all lead to endemic errors in the database of citations, particularly in the nexus of their linking to faculty authors. Since there is at present no way to simply eliminate the introduction of these errors with existing tools and data sources, they must be removed after the fact.

To this end the Library developed a suite of online quality control tools. These tools attack the problem by making it as convenient as possible to identify and remedy errors. They collect a specific set of problematic records in an environment where a knowledgeable user can quickly correct problems or determine that corrections are not necessary. The Library has developed tools to address the following problem sets:

- i. Citations published well after their putative faculty author has died.
- ii. Citations published when the putative faculty author was young enough to make their authorship doubtful.
- iii. Conspicuous gaps in the publishing history of individual faculty authors (e.g., more than 15 years between publications).
- iv. Possible duplicate citations based on year and truncated title.
- v. Citations for which there are no authorship matches.

Before the introduction of improved name matching in 2012 we generated a bimonthly printed list of every faculty member with publications tracing all the citation author strings matched on. A visual comparison of faculty

names with author names revealed mistaken or doubtful matches which were then reviewed and corrected online. This obviously tedious process help us eliminate many, many erroneous matches.

Quality assurance tools effectively identify and eliminate significant numbers of errors from the database, but cannot identify or remove all errors. It is also vital for those working with the data to constantly be on the alert for data problems. For instance, when loading new records, the load scripts generate logs of matches which can reveal specific or even general name matching issues. Once spotted these can be dealt with at the appropriate database level (e.g., eliminating a bad match or modifying a faculty record to include a previously undocumented middle name).

It is even more important for those who know the publishing most intimately to be actively involved; that means the faculty authors themselves. The Library encourages faculty to review their listings regularly and report mistakes promptly. To this end, we notify faculty whenever a new publication is added to their listing. The Library urges department heads, deans, and other influential people to encourage faculty participation. We make it a high priority to always respond to faculty inquiries, corrections and additions promptly and positively. As is only natural some faculty are very attentive to their listings, while others are seldom heard from.

Public exposure of the data does much to increase attention. As faculty come to understand that their Faculty Bibliography listings directly feed into the metrics that the deans use to evaluate performance and allot resources, their motivation to correct mistakes and keep their listings current becomes more pronounced.

NEW OPPORTUNITIES—BIBLIOMETRICS

For better or worse the modern world is metrics hungry and, as the French say, *l'appetit vient en mangeant*. It is widely believed that, whenever possible, management decisions should be justifiable with numbers. If numbers are not currently available, they should be made available soonest. If a thing is not directly measurable, there ensues a mad scramble to find a surrogate that can be measured. Surely measurement has its limits and there are plenty of anecdotes dramatizing the misuse or misunderstanding of otherwise innocent metrics. Still the current focus on

management metrics shows little sign of diminishing and those who run our universities, colleges, schools of medicine, and other institutions increasingly rely on metrics for both day-to-day operations and long-term planning.

As it happens, the Library is uniquely located to feed this appetite, that is, provided the Library is on top of bibliometrics and e-resource usage metrics. Both metrics are direct or indirect indicators of intellectual attention and research effort, which are of crucial importance to the research enterprise. Moreover, management by metrics entails politics. Justifiably or not, people anticipate there being winners and losers in the metrics game. These concerns make it vital that numbers be trustworthy. By extension, those who compile the numbers must also be perceived as trustworthy. The Library is fortunate in that, compared to other players in the academic medical center, it is typically viewed as neutral, a sort of academic Switzerland. Whether a library can fulfill this promise depends to a large degree on its ability to reliably determine what the institution's faculty actually publish. As we have seen, that is not a trivial undertaking.

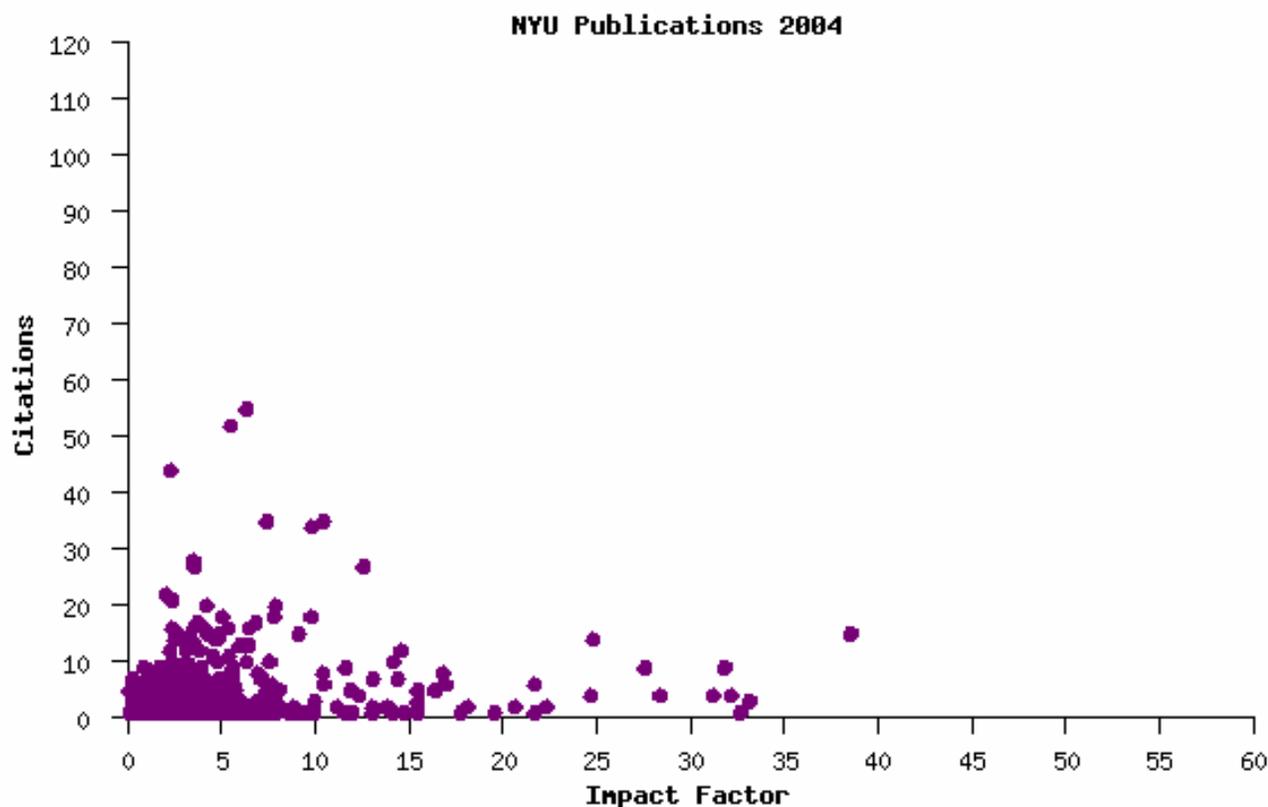
When the Faculty Bibliography project began, the NYU Health Sciences Library had only a slim involvement in

bibliometrics. The Library worked with individual faculty to calculate h-index values and other metrics on an ad hoc basis. The Library also supplied interested parties with journal impact factors and contributed compiled lists of publications for various purposes. With the migration to MySQL the Library was able to expand its offerings of bibliometric support, beginning by developing canned faculty and departmental 10-year reports which graphed publishing activity over time and were available on request to department heads (e.g., see Figure 5). These rather unsophisticated products were eclipsed by a set of reports created for the 2006 LCME process.

Bibliometrics efforts blossomed starting in 2009 when data from the Faculty Bibliography was incorporated as one of the key performance indicators on the “Dean’s Dashboard”, a major institutional management initiative.¹⁰ As part of a “Departmental Excellence” project, NYU School of Medicine’s Vice Dean for Science had asked the Library to include h-index values in relevant bibliometrics reports; in practice that proved unscalable without a local store of citing works data, which was not

¹⁰ For background on what dashboards are and how they work, see (Eckerson, 2006; Few, 2006).

Figure 5. Number of Citations vs. Source Impact Factor, 2004



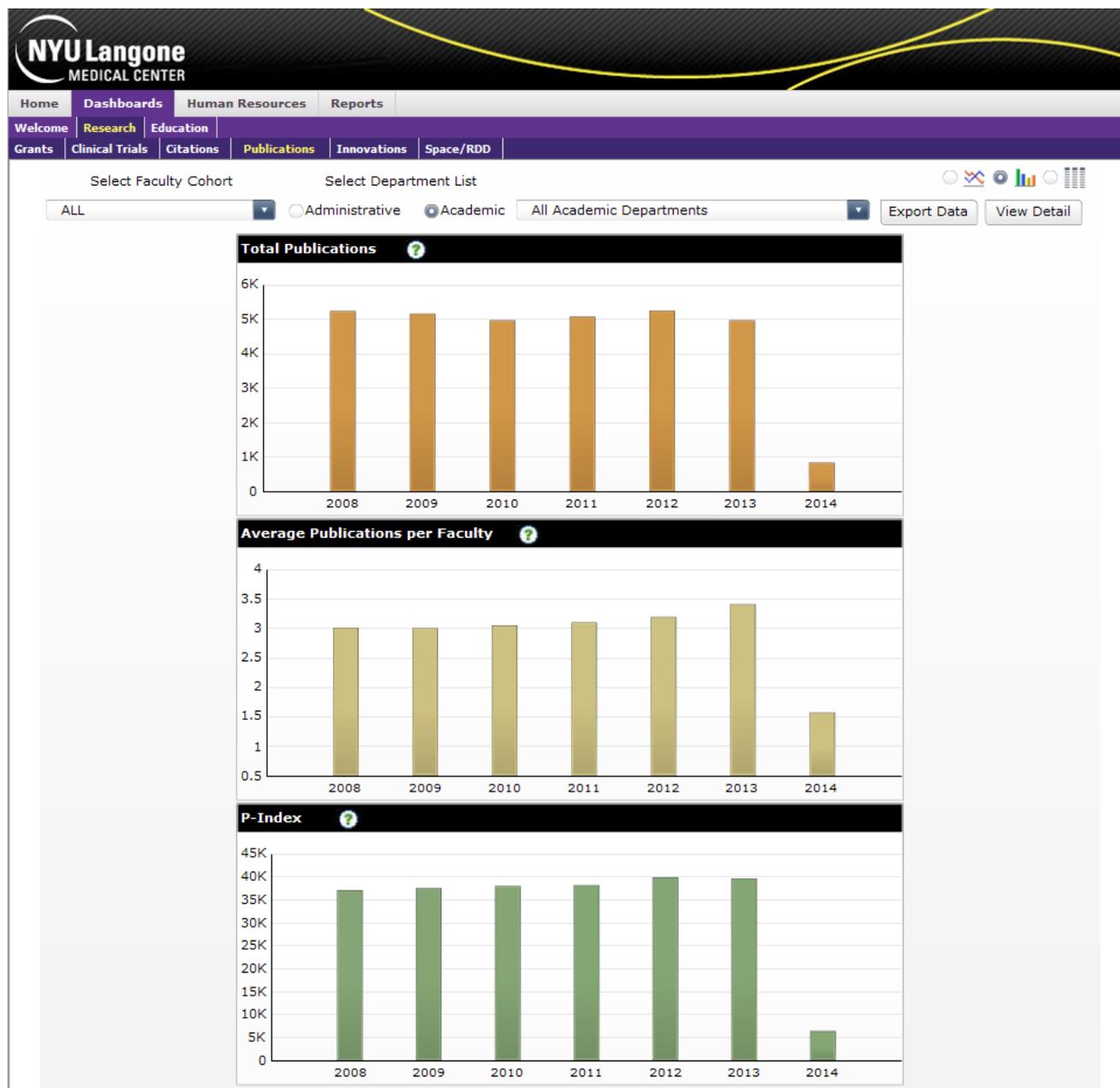
available. To meet this need, the Library developed a new metric called the “P-Index.” P-Index values patterned roughly with h-index values, but were easier to calculate for large author populations (Spore, 2010). Dashboard publications metrics (Figure 6) included the P-index, along with more basic measures. At first bibliometric data was supplied to the dashboard on a batch basis, giving rise to serious concerns with data staleness. In 2013 the batch process was replaced with dynamic feeds from the Faculty Bibliography via the enterprise data warehouse.

centering around the experience of being measured and can be as important as the generation of actual numbers. For instance, we learned to exercise discretion, avoiding potentially divisive uses of our work (for instance by warning against making comparisons across disciplines without careful consideration of what numbers actually represent), to be honest and transparent (for instance by being frank about limitations of the data and the inevitable presence of errors and omissions), and finally, by learning to speak the language of the audience, not just Library argot.

The Dean’s Dashboard lent legitimacy and prestige to Library bibliometrics work. It also taught us a good deal about “soft bibliometrics.” These relate to human concerns

The Library has continued to expand its bibliometrics efforts, focusing particular attention on collaboration and social networking analysis. For example, in the fields of

Figure 6. Publications Dashboard



medicine and biology the position of one's name in the list of authors customarily implies information about one's role in the preparation of the paper, last author generally being the most prestigious. Using Faculty Bibliography data we can chart authorship position (Figure 7) over a career, making it possible to visually identify patterns of prestige and influence. Similarly, using co-authorship

(Figure 8) it is possible to illustrate how much a particular researcher collaborates, both intra- and extramurally.

FINAL WORDS/ASSESSMENT

For the NYU Health Sciences Library, the Faculty Bibliography project has been a long term effort which has

Figure 7. Authorship Position Report

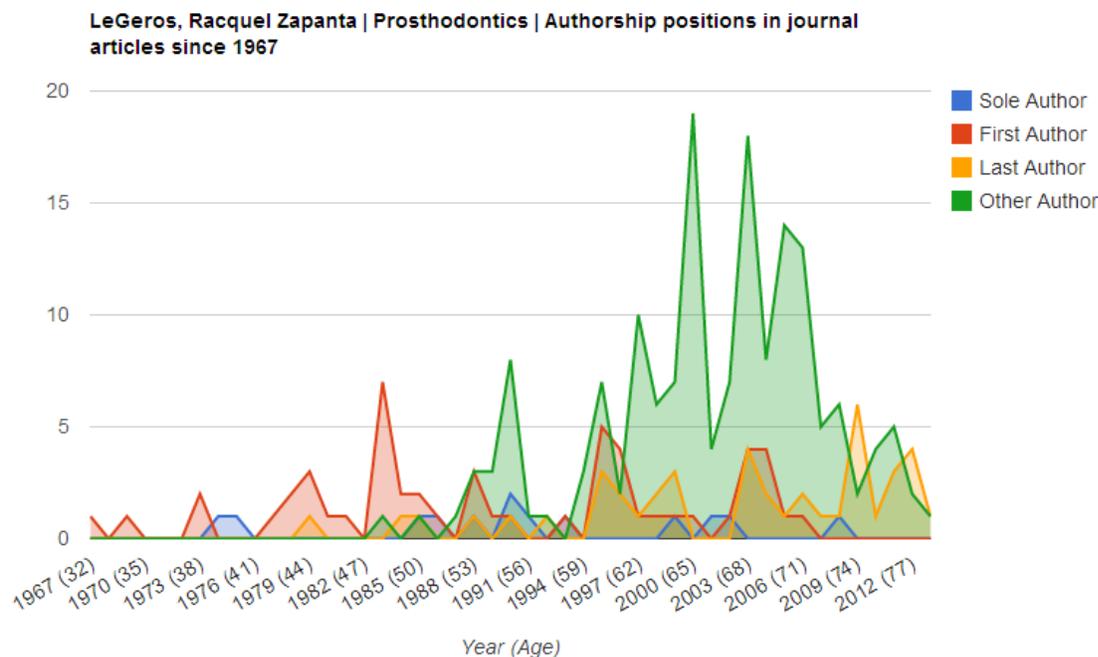
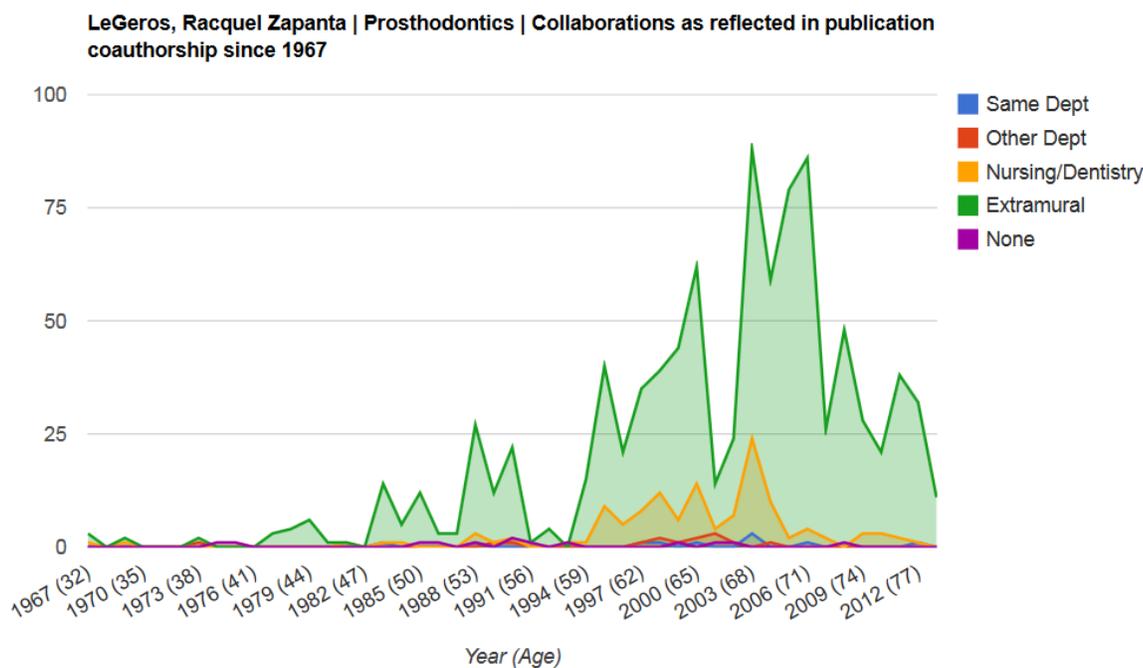


Figure 8. Collaboration (Co-Authorship) Report



been a positive contribution to the Medical Center and has consequently enhanced the Library's prestige. The project served to generally improve the Library's relationship with faculty and provided the Library with a sound foundation from which to participate in institutional management metrics.

The fourteen years' work is not without its regrets, however. In particular, better name-matching should have been implemented far earlier than it was; it was perfectly apparent that it was needed early on, but the press of other work discouraged earlier action.

Aside from choices we regret, the project taught us certain important lessons:

- Exposure can lead to improved data quality;
- Quality requires sustained attention to detail and constant review;
- Responsiveness builds trust; and
- Don't SILO, integrate.

The Library expects the Faculty Bibliography project to continue to grow. In the near future we hope to begin including non-Medical School NYU faculty, particularly those in neuroscience, psychology, public health, biotechnology, and similar disciplines. Doing so will require significant progress in faculty identity management across the various NYU campuses. We also hope to begin generating mass h-index values and other metrics soon.

We are greatly encouraged by the rapid success of ORCID.¹¹ ORCID's scheme to supply all publishing researchers with unique identifiers promises to finally dry up the swamp of name-matching and make identifying the authors of published works quick and easy—a consummation devoutly to be wished. At NYU we are in the process of developing an opt-out workflow which will ensure that all NYU faculty are established in ORCID and will use the Faculty Bibliography API to synchronize ORCID publication listings.

As faculty acquire ORCID IDs as a matter of course and as tracking ORCID IDs becomes a normal part of the publishing process, tracking faculty publishing will

become more readily achievable than it has been; one can imagine a day when most academic medical centers will have consistent and convenient access to their faculty's publications, making it possible to produce credible comparative cross-institutional publishing data. Accreditation exercises always generate requests for such comparative data; it may eventually be possible to meet such requests in a satisfactory way.

Looking ahead, there are other issues and challenges as well; bibliometrics and related disciplines are evolving rapidly. In particular there are questions as to how the Faculty Bibliography will relate to institutional repositories, to propriety, commercial bibliometrics tools (e.g. InCites, Scopus), and to research networking infrastructure (e.g. VIVO, Harvard Profiles).

REFERENCES

- Anderson, S. & Monroe, A. (1984). Joys of an automated faculty publications system. *Proceedings of the Annual Symposium on Computer Application in Medical Care*. Nov. 7, 1984, 991-994. Retrieved from <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2578650/>
- Bai, S., & Kelly, P. (2000). Development of a web-based faculty publications database. *Bulletin of the Medical Library Association*, 88(2), 189-192.
- Eckerson, W. W. (2006). *Performance dashboards: Measuring, monitoring, and managing your business*. Hoboken, N.J.: John Wiley & Sons.
- Fenichel, C. H. (1990). Combining reference with collection development: Using Pro-Cite to produce a faculty bibliography and as an aid in journal selection. *Medical Reference Service Quarterly*, 9(1):69-75. http://dx.doi.org/10.1300/J115v09n01_06
- Few, S. (2006). *Information dashboard design: The effective visual communication of data*. Cambridge [MA]: O'Reilly.
- Johnson, S. B., Bales, M. E., Dine, D., Bakken, S., Albert, P. J., & Weng, C. (2014). Automatic generation of investigator bibliographies for institutional research networking systems. *Journal of Biomedical Informatics*. [In Press, Corrected Proof]. <http://dx.doi.org/10.1016/j.jbi.2014.03.013>
- Key, J. D. & Sholtz, K. J. (1973). The Mayo Clinic author catalog: A living repository of medical knowledge. *Bulletin of the Medical Library Association*, 61(2), 228-237.

¹¹ <http://orcid.org/>

Lee, S., Gratz, P., & White, J. (1976). Computer-stored faculty publication file using the MT/ST in a medium-sized medical center library. *Bulletin of the Medical Library Association*, 64(1), 25-31.

Mansheim, R. & Thompson, J. L. (1995). Faculty publications databases in academic health sciences libraries. *Medical Reference Services Quarterly*, 13(4):21-34.
http://dx.doi.org/10.1300/J115V13N04_03

MacCorkle, L. (1991). Publishing an annual faculty bibliography at the University of Miami. *Information Technology and Libraries*, 10(2):121-127.

Marsalis, S. & Kelly, J. (2004). Building a RefWorks database of faculty publications as liaison and collection development tool. *Issues in Science and Technology Librarianship*, 40(Summer).
<http://dx.doi.org/10.5062/F4QZ27WK>

McKee, A. M. & Feng, C. C. H. (1979). Using computerized literature searches to product faculty publications lists. *Bulletin of the Medical Library Association*, 67(3), 333-335.

Nutter, D. O, Bond, J. S., Collier, B. S., D'Alessandri, R. M., Gewertz, B. L., Nora, L. M., . . .Watson R.T. (2000). Measuring faculty effort and contributions in medical education. *Academic Medicine*, 75(2), 199-207.
<http://dx.doi.org/10.1097/00001888-200002000-00025>

Potter, L.A. (1987). Creating a faculty publications database using Sci-Mate(tm). *Medical Reference Services Quarterly*, 6(3), 43-62.
http://dx.doi.org/10.1300/J115v06n03_05

Sayers, E. (2009). Entrez Programming Utilities Help. Retrieved from <http://www.ncbi.nlm.nih.gov/books/NBK25497>

Schwartz, V. & Stoffel, B. (2007). Building an online faculty publications database. *College and Undergraduate Libraries*, 14(3), 1-25. http://dx.doi.org/10.1300/J106v14n03_01

Spore, S. (2010). P-Index, a novel algorithm for bibliometric analysis [Poster]. *MLA 2010*, Washington DC. Retrieved from <http://hdl.handle.net/2451/33600>

Vieira, D. & Faraino, R. (1997). Analyzing the research record of an institution's list of faculty publications. *Bulletin of the Medical Library Association*, 85(2),154-157.

CORRESPONDING AUTHOR

Stuart Spore
Head, Bibliometrics & Senior Systems Advisor

NYU Health Sciences Libraries
577 First Avenue, Room 117
New York, NY 10016

stuart.spore@med.nyu.edu