Documenting And Evaluating Data Science Contributions In Academic Promotion In Departments Of Statistics And Biostatistics

Lance A. Waller

Department of Biostatistics and Bioinformatics

Rollins School of Public Health

Emory University

1518 Clifton Road NE

Atlanta GA 30322

lwaller@emory.edu

ABSTACT

The dynamic intersection of the field of Data Science with the established academic communities of Statistics and Biostatistics continues to generate lively debate, often with the two fields playing the role of an upstart (but brilliant), tech-savvy prodigy and an established (but brilliant), curmudgeonly expert, respectively. Like any emerging discipline, Data Science brings new perspectives and new tools to address new questions requiring new perspectives on traditionally established concepts. We explore a specific component of this discussion, namely the documentation and evaluation of Data Science-related research, teaching, and service contributions for faculty members seeking promotion and tenure within traditional departments of Statistics and Biostatistics. We focus on three perspectives: the department chair nominating a candidate for promotion, the junior faculty member going up for promotion, and the senior faculty members evaluating the promotion package. We contrast conservative, strategic, and iconoclastic approaches to promotion based on accomplishments in data science.

Key Words: Tenure, Team Science, Academic Promotion, Professional Development

1. INTRODUCTION

The field of Data Science has generated much discussion, enthusiasm, and investment within colleges and universities in recent years. Within the field of Statistics, Data Science extends and expands concepts such as significance testing and survey sampling to address inference in unstructured, big data settings. At the same time, some high-profile overviews of Data Science fail to mention the field of Statistics at all, despite its fundamental role in data-driven science (Davidian and Louis, 2012). While recent reports examine the need for statistical thinking within Data Science research and training (National Research Council 2013, 2014), few address the value of Data Science concepts within the field of Statistics, particularly with respect to the recognition of Data Science research, teaching, and service activities by faculty members in traditional Statistics and Biostatistics academic departments.

As in the development of any new area of scientific inquiry, leading researchers and instructors in Data Science typically are not themselves trained within official data science curricula, rather they bring training and experience from computer science, mathematics, statistics, and other fields to define scholarly elements of research, teaching, and service for an emerging science. These definitions remain fluid but play a critical role in determining evidence of academic success.

Several points of view are relevant in this discussion. For a junior faculty member, it is invigorating to be involved in the growth of a new field, but also concerning to wonder if contributions to interdisciplinary science will be appropriately valued and appreciated as scholarly contributions when one is evaluated for promotion and/or tenure within traditional disciplinary departments. For a department chair supporting meaningful participation in the development of emerging areas of science relating to Statistics and Biostatistics, it can be

exciting to encourage leadership by department faculty in cutting edge development of a new field of inquiry, but it can be challenging to determine how best to document such accomplishments to ensure full appreciation of an individual faculty member's unique accomplishments within the review process. For a tenured faculty member voting on promotion cases, it can be confusing to evaluate new types of accomplishments that can be quite different from traditional lists of publications and research grants, but, at the same time, clearly provide foundational support for new areas of science.

In the sections below, we begin with a brief and generic overview of the academic promotion process including a review of the typical dossier that provides evidence for review at each step of the process. We outline the usual components of the dossier with suggestions as to where and how to incorporate, document, and highlight contributions to Data Science, noting the specific need to establish interpretable context for these contributions. We identify conservative, strategic, and iconoclastic strategies for both the candidate and the department chair. Like the field of Data Science itself, these recommendations are dynamic and are likely to change (perhaps rapidly) over time. We hope this overview encourages ongoing discussion on the topic among junior and senior faculty members, department chairs, and academic administrators.

2. THE ACADEMIC PROMOTION PROCESS

Roughly speaking, the academic promotion process resembles a pre-Copernican view of the universe centered on the junior faculty member, expressed as a set of concentric administrative layers at the department, school/college, and university level. The promotion review process passes outward from the individual through each of these layers. Generally, the process is similar for both tenure-track and non-tenure-track faculty members where the tenure track often

involves stricter time constraints, a clear up-or-out (tenure) decision, and can involve more comprehensive documentation of accomplishments.

The first step for any promotion begins early, often during the candidate's initial interview for an academic position. It is always beneficial for the candidate, the department chair, and current faculty members to begin an ongoing set of conversations regarding the process, expectations, and timeline associated with promotion. However, it is also wise for a department chair to recommend discussions between a candidate and other senior faculty members within the department who will be evaluating their accomplishments, departmental representatives on the school or college Promotion and Tenure evaluation committee, as well as recently promoted faculty members to gain perspective regarding the promotion process from all points of view. It is also wise for the candidate to seek out and initiate these conversations, even if the department chair has not (yet) done so. These conversations introduce the candidate to the process early and initiate dossier documentation (and informal evaluation) very early in the candidate's career allowing regular updates rather than a flurry of preparation immediately preceding a promotion review.

The formal promotion evaluation process summarizes a candidate's accomplishments in each of three areas: Research, Teaching, and Service. My own institution requires individuals be evaluated as "Excellent" on one of these three areas (with specified criteria for excellence provided in the guidelines for promotion), and evaluated at least as "Very Good" in the other two. Each institution has a variant of this type of criterion and it is important for the junior faculty member to be familiar with the specific language as well as how this is interpreted by faculty and review committees throughout the promotion process.

PeerJ Preprints | https://doi.org/10.7287/peerj.preprints.3204v1 | CC BY 4.0 Open Access | rec: 29 Aug 2017, publ: 29 Aug 2017

The candidate works with the department chair to prepare a dossier for review at the following stages, typically in conducted in the following order:

1. An informal review by department faculty currently at or above the level to which the candidate seeks promotion (e.g., tenured faculty in the case of an individual being evaluated for tenure, Professors for those seeking promotion to the rank of Professor). The faculty review a full curriculum vita (CV) and a brief Personal Statement (see below for details) by the candidate regarding past performance, current efforts, and future plans in Research, Teaching, and Service.
2. A set of several (typically six) individual external evaluations, summarized in formal letters, by experts working in areas similar to the candidate's areas of expertise but not working closely with the candidate (i.e., typically not publishing or directly collaborating with the candidate). Depending on the status of open records laws and policies associated with the home institution, these letters may or may not be held as confidential.
3. A formal review and vote by the faculty members in step 1, based on the updated dossier material (the CV, the Personal Statement, and the external letters from step 2). The department chair conducts the vote and summarizes results for the next step along with the dossier materials.
4. A formal review and vote of approval by a Promotions and Tenure committee within the academic unit overseeing the department (often a school or a college);
5. Review and approval of the Promotions and Tenure committee's recommendation by the dean of the school or college;
6. Many Universities include a review by an University-level advisory committee reporting to the President; and
7. Final review and approval (or not) by the University Board of Trustees, Regents, or similar governing body.

At each stage of the review process, the results and documentation of each previous stage are summarized and included in evaluation materials (e.g., the results of votes at the department and school level are summarized in a letter by the department chair, the results at the college/school level summarized by the dean), and included for review at subsequent levels of the process.

Prior to promotion review, many universities also conduct an interim review of the progress of junior faculty, e.g., during year 3 of the six-year probationary period for tenure-track Assistant Professors, or annually for some universities. Such interim reviews incorporate elements of steps 1-4 in the outline above (typically with the exception of the external letters from experts). Interim reviews provide an excellent opportunity for junior faculty members to keep their dossier items up to date and, importantly, for both the candidate and the department chair to gauge the reception of accomplishments to date and the candidate's future plans from the faculty members who will be voting on promotion. Interim review also provides an opportunity to assess whether the candidate's accomplishments are being communicated in a manner that is fully understood and appreciated by reviewers at these critical early steps of the process. This last point is important and provides a mechanism for the candidate, the department chair, and for the evaluating faculty to make sure accomplishments are communicated *and* received in the most effective manner possible, and to make adjustments between the interim and promotion reviews.

It is important to note that evaluations at the first few stages in the process above (e.g., department review) typically focus on the individual's contributions within a disciplinary environment (i.e., addressing the question "is the candidate a good *academic statistician*?"), and later stages by reviewers further removed from the individual's area of expertise who rely more

on general assessments of academic accomplishment and summaries from earlier stages of review (i.e., addressing "is the candidate a good *faculty member*?"). The interdisciplinary nature of Data Science requires a candidate to prepare carefully for this typical (but often underappreciated) feature of the review process, i.e., by preparing strong, documented evidence to provide positive responses to both questions. This is especially important if many of the candidate's accomplishments fall outside of the "typical" Research, Teaching, and/or Service activities familiar to the senior faculty in the candidate's department, school, or university.

It is particularly important for the candidate to continue the conversations mentioned above with the department chair, senior faculty in the department, and the department representatives to the Promotion and Tenure Committee throughout the probationary period. It is also important for these to be two-way conversations providing advice to the candidate from voting faculty and providing advice to voting faculty from the candidate (and peers) to highlight the value of the data science contributions. These conversations provide a continuing opportunity to raise awareness of the value of contributions and gain insight on methods for documenting such accomplishments.

3. STRATEGIES FOR PROMOTION: BUSINESS AS USUAL OR DAMN THE TORPEDOS?

At the risk of being overly simplistic, candidates and department chairs often take one of two extreme strategies for promotion. I briefly review each, then suggest a more strategic approach.

First, candidates and department chairs may seek to minimize risk through a *conservative strategy* where the candidate's dossier follows the standard format and stresses that each of the typical standards in Research, Teaching, and Service are met or exceeded using standard

191 metrics of peer-review publications/external grant funding for Research, teaching
192 evaluations/awards for Teaching, and a list of committee work/editorial duties/etc. for Service.
193 The key in this strategy is to demonstrate excellence in at least one of the three categories,
194 typically in Research or Teaching, using time-honored departmental metrics of success (in my
195 experience, it is very challenging to center a conservative approach to a promotion case around
196 excellence in Service, as Appointment, Promotion, and Tenure Committees encounter very few
197 dossiers structured in this manner). The strength of the conservative approach is that voting
198 faculty members evaluate familiar material (i.e., they feel they know a good dossier when they
199 see it) but the weakness is that the novel aspects of Data Science are hidden at best or
200 overlooked and ignored completely at worst.
201
202 The second strategy is an *iconoclastic strategy* based on the idea that the candidate's
203 excellence at what they do and the very different way in which they do it are so foundational that
204 it is patently clear that they should be promoted. Under this strategy, the dossier, the
205 department chair, and the candidate essentially dare the voting faculty at every level to question
206 the promotion. The emergent state of Data Science and its rapidly expanding skill set merging
207 computational and statistical science can make this strategy tempting (Why should the old rules
208 and standards apply?) and a strength of the approach is its push to question the validity of the
209 existing standards for new elements of scholarship. However, successful implementation
210 requires the department chair to convince the old guard (or at least a majority of the old guard)
211 to acknowledge the value of a brave new world. The approach can be effective for the right
212 candidates presented in the right way, but, as one might expect, the weakness of this strategy is
213 that it is a very risky gamble with very high stakes for the candidate.
214
215 Based on many discussions with both candidates with strong interests in Data Science and
216 department chairs familiar with voting patterns at the department and higher levels at different

institutions, I describe a *strategic approach* between the conservative ("be sure to look like an excellent candidate by the existing standards") and the iconoclastic ("exhibit a new type of excellence with new standards"). This approach draws from the strengths of both extremes and tempers the risks associated with each, as outlined in more detail below.

From the candidate's perspective, a strategic approach requires ongoing education of senior faculty colleagues as to the value of the candidate's work and documentation of accomplishments that may not be as familiar as first-authored peer-reviewed publications and principal investigator roles on grants. Elements of this ongoing education include initiating and participating in department discussions on the role of and value of Data Science within broader scientific inquiry. Some effective ways to do this include: suggesting successful data scientists as departmental seminar speakers, volunteering to give departmental seminars, suggesting data science papers for discussion in journal clubs/reading groups/working groups, and offering to teach classes (or guest lectures) on Data Science topics within the departmental curriculum. Most departments of Statistics and Biostatistics are filled with graduate students interested in Data Science topics who are eager to try new ideas. These efforts will ensure that key ideas relating to the content and importance of the candidate's skills are familiar topics, not new ideas first encountered by reviewers during the evaluation process.

In addition to raising the profile of Data Science within department academic life, it is equally essential for the candidate to identify effective ways to describe and document her/his work as research contributions and to identify her/his unique role in each project in which she/he participates. This documentation aims to answer such questions as: What did the candidate contribute? Do these contributions result in creation of the base data set for a series of publications? Do they result in a reproducible analytic pipeline? Do they result in a software

package? Are others using the data set and/or software? Are others citing these data science research products?

As a helpful aside, in discussions relating to the preparation of this article, Hadley Wickham noted: "Academics tend to like warm feet, but they don't appreciate who makes their socks." I will be forever grateful for this metaphor, since my main point in the preceding paragraphs is to create an environment where the hard work such as tool and/or dataset development for reproducible research (the "sock development") is understood to be an important research contribution *in addition to* the publication of the primary manuscript (the "warm feet").

From the department chair's point of view, a strategic approach requires similar efforts and responsibilities. First, it is essential to listen carefully to the candidate's descriptions of her/his contributions and to accurately articulate these regularly in discussions at the department, school, and university levels. It is helpful for the chair to reach out to the candidate for suggested publications relating to the application and value of their work. These may be peer-reviewed publications by others, opinion pieces calling for more efforts in a given area, or online discussions of challenges and creative solutions. A department chair need not become an expert in Data Science (just as a department chair need not be expert in every area of research conducted by department faculty members), but the department chair should be able to outline the candidate's interests and why these are important to deans, visitors, students, and, importantly, potential collaborators across campus and in other institutions. By linking the basics of the candidate's interests and linking these to networks of individuals and ideas (e.g., asking "have you thought about applying this approach to this area?"), the department chair can have a profound and helpful influence, not only as a mentor, but also as a matchmaker for productive collaborations. Finally, discussions with other department chairs from Statistics, Biostatistics, Informatics, or Computer Science can be extremely helpful, especially in outlining how best to document particular research activities. This issue need not be solved *de novo* at

each institution nor in each separate department, but rather the chair can benefit immensely from the lessons of others. To borrow the Wickham Metaphor: the department chair has the opportunity and responsibility to raise the profile of sock-making among potential collaborators, fellow administrators, and disciplinary colleagues, all of whom desire warm feet.

From a senior faculty member's perspective, it is helpful to approach the candidate's areas of interest and productivity with the same level of curiosity as one might a colleague working in another area of Statistics, and it is extremely helpful to have such conversations with the candidate well ahead of the formal review process. It is helpful to discuss with colleagues from other institutions how they evaluate Data Science-related accomplishments in Research, Teaching, and Service. It is also helpful to follow ongoing discussions on Data Science issues by professional associations (e.g., the American Statistical Association, the Royal Statistical Society, and the International Biometrics Society), related workshops and reports by the National Academies of Science, Engineering, and Medicine, and relevant calls for funding by the National Science Foundation, the National Institutes of Health, and other funding agencies. Senior evaluators have the responsibility to stay current (and perhaps ahead of the curve) on evolving trends and directions in their and related fields.

4. DIFFERENT TYPES OF SCHOLARLY CONTRIBUTIONS

The interdisciplinary nature of Data Science results in multiple types of scholarly contributions and, as noted above, it is important for chairs, candidates, and senior faculty to have a conceptual framework for appreciating and documenting these (i.e., a language for valuing sock-making). In the 1990s, Ernest Boyer of the Carnegie Foundation for the Advancement of Teaching articulated the value of different types of interdisciplinary scholarship in his highly cited *Scholarship Reconsidered: Priorities for the Professoriate* (latest, expanded edition: Boyer et

al. 2016). This report, very familiar to university administrators such as deans, provosts, and presidents, but often less well known by junior faculty, highlights the value of multiple types of faculty contributions within academia, specifically noting four types of scholarship. The first, *scholarship of discovery*, mirrors the standard disciplinary model of original research advancing knowledge within a field, often evidenced by peer reviewed publications in established disciplinary journals and success in obtaining competitive research funding. A second type, *scholarship of integration*, recognizes innovative synthesis of information across traditional disciplines, across subdivisions within a discipline, or across time. Such scholarship creates new knowledge through novel links between specific concepts, tools, and studies from disparate fields of inquiry. Boyer's third type, *scholarship of application* (sometimes called *scholarship of engagement*), goes beyond simply applying existing tools (as would a technician) to value the deep collaborative contributions in creating advances in interdisciplinary studies, particularly within a team science framework. The fourth type, *scholarship of teaching and learning*, values the systematic study of pedagogical methods for the transfer and creation of new knowledge between faculty members, colleagues, and the next generation of scholars.

These four categories provide rich support for many current efforts within the field of Data Science and its link to academic departments of Statistics and Biostatistics and topics for the continuing critical conversations between the chair, the candidate, and senior faculty. The four types of scholarship also provide a context for collecting, presenting, and reviewing scholarly contributions of junior faculty. The concept of the scholarship of integration is immediately extensible to Data Science, particularly with respect to linking heterogeneous data components and developing new analytic tools, hence enabling new lines of inquiry. Documentation of such contributions within a promotion dossier is somewhat non-traditional and may include citable data within repositories and software packages/toolboxes in addition to peer-reviewed publications and funded grants. The scholarship of application is evidenced by interdisciplinary

publishing, the creation of data repositories and complex data sets, and clear contributions unique to the candidate within an interdisciplinary team. Data Science research contributions often are linked deeply to the intersection of Boyer's scholarships of integration and application, and it will be advantageous to highlight the impact of these contributions within this context.

The rapid development of training programs, concentrations, and degree programs within the area of Data Science offers multiple opportunities for the scholarship of teaching and learning. Success in this area extends well beyond simply teaching new courses and advising students, it involves research and discovery on the modes and methods of instruction and learning, an area of clear interest in the statistical education research community, but only just developing in the broader area of Data Science (National Academies 2014).

Boyer's categories provide a valuable framework for organizing and presenting a candidate's scholarly contributions for review. The candidate can organize materials under Boyer's categories in the CV, and mention them in their Personal Statement. A department chair can frame contributions in light of discovery, integration, application, and teaching to external reviewers and when presenting a candidate's promotion for consideration by senior faculty and promotion committees. Senior faculty can use Boyer's categories as a lens through which to view accomplishments and assess impact of a candidate's scholarly work.

Boyer's categories link well with current calls for increased appreciation of success in statistical collaboration and participation in interdisciplinary Team Science. The Mathematical Association of America's 2003 *Guidelines for Programs and Departments in Undergraduate Mathematical Sciences* directly cites the categories and the American Statistical Association's comments on the MAA Guidelines identify the key role of consultation and collaboration within the statistical profession (links to both appear in the references below). More recently, Mazumdar et al.

(2015) draw on the spirit of Boyer's work in their proposed framework for evaluating academic scientists working in team-based environments and lists specific Research, Teaching, and Service examples of scholarship of integration and scholarship of application within an academic health center. These examples readily extend to the general academic setting and are particularly relevant (though not limited to) the interplay of Statistics, Biostatistics, and Data Science.

5. PROVIDING CONTEXT FOR DATA SCIENCE CONTRIBUTIONS

We next review the core set of evidence to be evaluated for promotion, i.e., the elements of the individual's promotion dossier, a packet of information typically including (at least): (1) a full curriculum vitae (CV) summarizing the individual's accomplishments to date; (2) a Personal Statement (or Statements) by the candidate summarizing their contributions to and future plans in the areas of Research, Teaching, and Service; and (3) a set of external evaluation letters. In addition, the dossier often also includes a set of a handful of representative publications illustrating contributions to Research, teaching evaluations and sample syllabi illustrating contributions to Teaching, and a full summary of contributions to Service. Different institutions offer slight variations to this general framework, but the elements listed above are fairly consistent across most universities in the United States. We next review each component of the dossier in detail.

5.1 Documenting Data Science Contributions in the CV

The most familiar element of the promotion dossier is the curriculum vita (CV), which summarizes the candidate's background education, employment history, awards and honors, as well as a full list of peer-reviewed publications and grants, conference and seminar

372 presentations, teaching/advising/mentoring activities, and service to the department,
373 school/college, university, and profession. In the CV, senior faculty and external reviewers
374 typically seek evidence of research success through peer-reviewed publications, competitive
375 grant funding, and invitations to speak at conferences and seminars. For teaching success,
376 reviewers typically wish to see good and continually improving teaching evaluations, growing
377 course responsibilities, innovations and new ideas in the classroom, and self-awareness
378 evidenced through the development and articulation of an overall teaching philosophy by the
379 candidate. For service, reviewers look for participation on committees at the department,
380 school/college, and university level as well as participation in activities associated with various
381 professional organizations, refereeing and membership on editorial boards, and participation in
382 research and grant review panels (e.g., for the National Science Foundation or the National
383 Institutes of Health, etc.).
384
385 As in many fields, the traditional academic CV in the fields of Statistics and Biostatistics
386 highlights research contributions through peer-reviewed publications, especially those in
387 journals with strong reputations in the field, and grant funding, often as Principal Investigator.
388 Many times, research contributions in Data Science result in the creation of novel software
389 tools, merged and curated data sets, and contributions to research teams where the candidate
390 plays an essential role on an interdisciplinary team, but may not serve as the lead investigator.
391 It is critical for the candidate to work with her/his department chair to provide context for these
392 contributions and to raise awareness of the value of such contributions among the senior faculty
393 in the department, on the Promotion and Tenure Committee, and at higher levels of the
394 promotion process. The ongoing conversations mentioned above aid in this effort, but it is
395 equally important to document the accomplishments within the CV to make sure they are
396 noticed and appreciated.
397

As a faculty member in a department of Statistics/Biostatistics, there may be an expectation by senior faculty of some publications in traditional journals in these fields, an expectation that is infuriating to junior faculty member taking the iconoclastic promotion approach, stifling to a successful Data Scientist taking a conservative promotion approach, but an important consideration for a junior faculty member and chair taking a strategic approach to promotion. Early conversations can involve identification of journals in Statistics/Biostatistics that value Data Science contributions, and Data Science journals that value statistical contributions. A junior candidate need not narrowly focus on only a few statistical journals (unless the department is very rigid and traditional), rather the strategic candidate seeks outlets highlighting both aspects of their research contributions. This is not necessarily a straightforward task, but should be part of the ongoing discussions between junior faculty, senior faculty, the department chair, and their colleagues at other institutions.

In addition to publications in statistical or related literature, the candidate will likely also have publications in the area of Data Science. Publications particular to Data Science likely will appear in newer, electronic journals rather than long-standing established journals. The candidate should note the relevance and reputation of the journals, and include new metrics of impact for her/his publications such as “most downloaded” or “highly cited” over a period of time. The candidate should be careful to distinguish peer-reviewed publications from non-peer reviewed publications. It is fine to list non-peer-reviewed publications, but these should be listed in a separate, clearly marked section.

It is important for department chairs and senior faculty in Statistics/Biostatistics to be aware that many computational fields place higher value on refereed meeting proceedings than on traditional journal publications, due to timeliness and competitiveness of review and due to historical culture of the field. Since Data Science contributions often occur in computational as

well as statistical fields, the candidate may have several such publications listed in her/his CV. If so, adding a parenthetical note identifying the acceptance rate can be helpful for senior reviewers who may otherwise (incorrectly) view meetings proceedings as a non-peer-reviewed publication. The strategic department chair will raise awareness of this issue among senior faculty well before the promotion evaluation, keep this in mind in selecting external evaluators and assessing external evaluations, and will clearly make this point in the chair's cover letter to the dossier as it is presented to the later stages of the promotion process.

Software development and data wrangling will likely be a key aspect of any candidate involved in Data Science so these should be clearly identified and it is important to place such accomplishments prominently within the CV. However, historically, traditional CVs in Statistics/Biostatistics do not consistently list such contributions so senior evaluators may not know to look for them. One approach is to have separate sections for software tools and possibly for curated data, but the notion of peer-review and citations for software and data are not clear, nor are they consistently used. Fortunately, new options are rapidly becoming available even though their rate of adoption is slower than many would like. Software (e.g., R packages) may be accompanied by peer-reviewed publications in journals such as *Journal of Statistical Software*, but may also be closely tied to peer-reviewed journal or proceedings articles already appearing in the candidate's publication list (e.g., an R package associated with the statistical methodology proposed in a publication in a statistical journal, or an analytic pipeline associated with a large, team-science paper appearing in the biomedical literature). As a way to link the development of curated data, software packages, and analytic pipelines to traditional lists of peer-reviewed publications, I suggest adding parenthetical comments linking the software package/data set/pipeline to the associated publication (or publications). Such comments allows the candidate (and the department chair) to highlight citations as well as downloads in a strategic manner to provide a link between the software tool and the peer-

reviewed publication that can be helpful for reviewers scanning a CV with traditional lenses. Such comments also clarify the unique contributions of the candidate to large, team-based projects.

While some areas of science have well documented data repositories and several mandates exist for making federally funded research project data publically available, two challenges remain in the general recognition of data sets, particularly complex, linked, and curated data as citable, scholarly research output. The first is the establishment of a peer-review equivalent to publications for quality control, and the second is the absence of a standard method to cite complex data sets (including verifiable attribution and date/version labels). Some citation standards certainly exist, but, in general, these are not (yet) as universally applied as, say, citation of publications, and the true impact of a data set likely falls somewhere between the full number of downloads and the current number of formal citations. These challenges present an obstacle for the clear recognition by senior evaluators (both internal and external) of the impact of the contributions of data wrangling and curation to the advancement of Data Science as well as Statistics/Biostatistics.

Strategic junior candidates, department chairs, and senior faculty should be aware of recent developments in regard to data citation, including the work of the Research Data Alliance (and their Data Citation Working Group, https://rd-alliance.org/groups/data-citation-wg.html, providing guidelines for data citation of evolving data sets), among other groups, and the establishment of the Joint Declaration of Data Citation Principles (Data Citation Synthesis Group, 2015, https://www.force11.org/group/joint-declaration-data-citation-principles-final). The Data Citation Principles include: importance (i.e., “data should be considered legitimate, citable products of research”), credit and attribution (e.g. are all individuals involved credited with the data?), evidence (i.e., when results rely on specific data sets, these data should be cited), unique

identification (which version of the data are you using?), access (i.e., are the data available?), persistence (i.e., past data should be available as well as updates), specificity and verifiability (e.g., does the citation clearly identify which version of a data set was used?), and interoperability and flexibility (e.g., data access should work across platforms). These are important steps, but equally important to their definition will be the adoption of the principles in practice by the scientific community, including the fields of Statistics and Biostatistics. (As a side note, it is and will continue to be important for statisticians to be involved in activities and committees related to continued developments in developing data citation guidelines and practice.)

Currently, many details regarding the process of creating and curating complex data sets are deferred to the supplemental information section of a publication, often receiving much lighter review than the main body of the paper but containing information critical for the reproducibility of the results. In response, some novel publication outlets now provide researchers the opportunity to submit their data *and* details on its construction for peer review and publication, as separate manuscripts but parallel to the original research report (Waller and Miller, 2016). These opportunities provide a unique, citable, digital object identifier (DOI) for (i) the original paper, (ii) the detailed description of data development, and (iii) the data themselves. Two examples of this approach include the Dryad Digital Data repository and an online journal by the Nature Publishing Group, *Scientific Data*. These are simply two examples and similar outlets are also available and in development. The Dryad Digital Data repository (http://datadryad.org/) provides a digital repository for data that meet most (if not all) of the Data Citation Principles listed above, again providing a DOI associated with the data and a link to the original publication. *Scientific Data* (http://www.nature.com/sdata/) takes things a step further and publishes peer-reviewed “data descriptors”, full-length papers about the data development process spelling out the details that are often relegated to supplemental information but are

essential for documenting the elements of Data Science involved in the creation of the data set, again published with a DOI. These two examples and others like them, provide an outlet for data-oriented research output (sock making) that are more similar to traditional publishing measures (peer-reviewed and citable) and more familiar to reviewers from traditional disciplinary areas.

Newer sources of research writing such as social media posts and blogs currently do not fit neatly into the traditional promotion dossier, but can have a documented impact on the field. Such activities should be noted in the CV in a clearly identified section along with documentation of impact. A key strategic element will be to provide evidence of the influence by the candidate on the field through such activities. The strategic candidate and department chair should discuss ways to document and demonstrate such impact through reports of reposts or through identification of external reviewers who can be asked to specifically comment on the impact of these efforts in their evaluation letters.

Moving past Research, the CV also documents accomplishments in Teaching. Such evidence includes lists of courses developed and taught, lists of enrollment, sample syllabi, and summarized teaching evaluations. Senior evaluators typically examine this information for evidence of training the next generation of scholars within the candidate's own discipline, but also in training these scholars to be effective collaborators. The emerging nature of Data Science provides opportunities for generating new courses or revised curricula for existing courses weaving in elements of computation, data management, markup reports, and data wrangling based on research trends and demands of future employers (National Research Council 2014). In addition to these standard elements, the emphasis on computation and technology within Data Science also opens the door for online instruction ranging from YouTube channels devoted to instruction in specific software packages, to massive, open, online courses

(MOOCs) enrolling thousands of students. These new opportunities for Teaching may be unfamiliar to traditional evaluators (internal or external), and may require additional context and metrics to document impact and influence on the field. For example, MOOCs are notorious for having very large initial enrollments with a low completion rate, and it is important for candidates and strategic department chairs to be up-front about the full picture of such activities to stem potential skepticism by senior evaluators. Linking YouTube instruction to particular software packages or publications (cited in the CV) can also help. As with novel sources of publication (e.g., blogs) mentioned above, the candidate and department chair should work closely to document impact on the field, identify potential senior evaluators who can speak direction to this impact, and communicate these accomplishments in the department as they happen, not just as part of a promotion dossier.

Finally, the CV highlights Service contributions through lists of committee memberships, refereeing activities, editorial board service, and participation in meeting organization and professional organizations. Leadership roles within these are important to highlight. Candidates with Data Science activities may demonstrate refereeing or associate editor duties for journals in both Data Science and Statistics/Biostatistics or referee activities for data publishing outlets. Many schools, colleges, and universities are in the midst of strategic planning activities relating to Data Science, Big Data, etc., and department chairs are often looking for representatives with knowledge in both Statistics/Biostatistics and Data Science to serve on such committees. These committees provide valuable networking and collaborative opportunities for junior faculty, and often provide leadership opportunities in both the short and long term. Strategic candidates and department chairs alike should be vigilant for such opportunities but also careful to weigh the potential benefits to the candidate against the associated time, effort, and likely outcomes involved. These activities may be aggregated within the CV as “Data Science Service” or simply interspersed among other Service contributions.

554

555 5.2 Documenting Data Science Contributions in the Personal Statement

556

557 The Personal Statement(s) (either a single statement covering Research, Teaching, and
558 Service or three separate statements) provide the candidate an opportunity to both summarize
559 academic accomplishments, and, importantly, an opportunity to place accomplishments in the
560 specific context of her/his broader professional progress. That is, the Personal Statement offers
561 the strategic candidate a forum to highlight why and how her/his contributions matter to the
562 candidate and to the field(s) of inquiry (Statistics/Biostatistics *and* Data Science). A strong
563 Personal Statement illustrates how the candidate has built on past work in the field, illustrates
564 the candidate's participation in present initiatives, and illustrates the candidate's vision of future
565 directions for themselves and their area of inquiry. The Personal Statement offers the best
566 place to define the strategic candidate's contributions within the framework of Boyer's different
567 types of scholarship, particularly with respect to the scholarship of integration and the
568 scholarship of application. Finally, the Personal Statement offers a the strategic candidate the
569 opportunity to discuss accomplishments from two important perspectives: the candidate's
570 *contributions*, i.e., what the candidate has accomplished; and (equally important, but often
571 overlooked) the *candidate's* contributions, i.e., how these accomplishments reflect the unique
572 abilities of this particular person's skills, abilities, and insights. This second perspective clearly
573 identifies why *this candidate* is unique, why *this candidate* stands out, and why *this candidate*
574 should be promoted. This message is particularly relevant for interdisciplinary contributions
575 within Data Science. Framing accomplishments in this way offers the strategic candidate the
576 opportunity to highlight personal contributions to projects shared within broader research teams,
577 making the case that the success of the projects required the participation of *this candidate*, not
578 simply the presence of anyone with statistical skills.

579

The Personal Statement offers an excellent opportunity for the candidate to define a personal view of what constitutes Data Science, what requires innovation, what the candidate has done in this regard to date, and what opportunities present themselves for future work. The Personal Statement is the perfect place for the strategic candidate to identify how Data Science links to but differs from many traditional paths in academic Statistics and Biostatistics, and to articulate why this matters. A clear general definition with specific examples mentioned in each of the Research, Teaching, and Service sections will frame the discussion for reviewers at all stages of the process, since the external reviewers, senior voting faculty members, and upper level administrators may have varying personal perspectives on emerging areas of inquiry at varying levels of specificity. The Personal Statements are a strategic opportunity for the candidate to define the discussion rather than hope accomplishments are self-evident at all levels of review.

5.3 Documenting Data Science Contributions in the External Letters

The external letters are an essential element of promotion review and consist of three separate components: the letter writer, the letter content, and the letterhead. We consider each of these in turn.

The *letter writer* should be an established and successful academic who can write knowledgeably about the candidate's accomplishments and evaluate overall success in Research, Teaching, and Service. The letter writers are typically "arm's length" evaluators and have had limited direct collaboration with the individual. It is sometimes a challenge for department chairs to identify individuals who know the candidate's work well but have not (yet) collaborated with the candidate. Candidates can provide suggestions of individuals they feel well suited to evaluate accomplishments, similarly, candidates can request that particular individuals with perceived conflicts-of-interest not serve as external reviewers. Institutions

606 typically also require additional external reviewers identified independently of the candidate's
607 suggestions. For strategic candidates with a Data Science focus, it is important to clearly
608 communicate to the department chair both names of potential reviewers as well as the *type* of
609 reviewer (e.g., individuals who make use of large-scale distributed computing) who would best
610 be able to appreciate and evaluate contributions to Data Science, to aid in identifying the full set
611 of reviewers.
612
613 The *letter content* provides detailed assessment of the candidate's accomplishments and career
614 trajectory. Some universities require explicit comparison to others in the candidate's general
615 stage of career as well as an evaluation as to whether the candidate would be competitive for
616 promotion at the letter writer's home institution. Other universities prohibit such statements and
617 it is helpful for strategic candidates to know the rules of their own institution. Letter writers are
618 often experts in areas related to the candidate's work and can provide disciplinary details
619 relating to the quality of journals, competitiveness of grants, success in instruction and
620 mentoring, and service activities, as well as broader contributions to the candidate's
621 department, school/college, university, and profession.
622
623 Including the *letterhead* as component of the evaluation letter may seem a bit cynical, but the
624 reputations of both the letter writer and the letter writer's home institution carry weight in the
625 evaluation, particularly in the latter stages of review. As the candidate's dossier advances, the
626 individuals reviewing the materials will be less familiar with the candidate's particular field of
627 study, target journals, and funding agencies and will rely more on broader measures of
628 assessment and give weight to qualifications of the external reviewers such as titles (e.g.,
629 Distinguished Professor, Department Chair) and the reputation of their home institution (e.g., is
630 it a "peer institution"?). With respect to Data Science, it can be very helpful for letter writers to

631 come from institutions that have demonstrated success in the areas relating to the candidate's
632 accomplishments and goals.
633
634 External reviewers are encouraged to provide their view of the candidate based on the dossier
635 as well as any personal experience they have interacting with the candidate or the candidate's
636 work. The context provided by the Personal Statement and the organization of the CV will be
637 critical in presenting the contributions of the candidate in light of their impact on the candidate's
638 discipline (Statistics/Biostatistics) as well as on Data Science. the organization of material can
639 be extremely helpful in documenting the vision, experience, and trajectory of the candidate in
640 each of Research, Teaching, and Service. The strategic candidate has the opportunity to
641 frame material to illustrate connections and the overall career path, and the strategic
642 department chair has the opportunity to request specific comments from the evaluators relating
643 to the integration of novel Data Science elements within traditional disciplinary
644 accomplishments. Following a conservative strategy can minimize Data Science contributions
645 within lists of more traditional disciplinary accomplishments, and an iconoclastic strategy can
646 assume links that are obvious to the candidate are clearly visible to an external reviewer. A
647 strategically prepared dossier provides clear indications of how Data Science contributions are
648 relevant and contribute to the candidate's success in expanding the frontiers of her/his
649 discipline.
650
651 ## 6. SUMMARY
652
653 In summary, it is in the best interest of the strategic candidate, the department chair, and our
654 field to create an accurate and organized assessment of accomplishments in Research,
655 Teaching, and Service within a dossier where the elements intertwine to make the strongest
656 case possible for promotion. The strategic CV documents accomplishments in both traditional

areas (e.g., publications, grant support, classroom teaching, committees), as well as emerging outlets for scholarly productivity (e.g., DOIs for data, data descriptions, blogs, online education). The strategic Personal Statement provides personal context of past work, vision, and future plans, allowing the candidate to place Data Science accomplishments in perspective within an individual's professional development in a traditional disciplinary environment. Strategic external letters provide the perspective of the broader academic community and reflect the letter writer's perception of the candidate's accomplishments within current and future trends of academic success. Together, the elements of the dossier provide the basis for review of the candidate's progress by senior faculty and administrators in the candidate's department, school/college, and university.

While promotion considerations should not be an obsession for junior faculty, I do feel it is very helpful to be familiar with the local steps of the process and the local rules governing promotion at the candidate's institution. Early and frequent discussions between the candidate and department chair (e.g., during annual reviews) are essential, to document accomplishments and establish context for activities, especially when including an interdisciplinary focus in Data Science. Effective communication goes in both directions with the strategic candidate learning the local processes and the strategic department chair learning the value and context of the candidate's accomplishments. Ongoing discussions between the candidate and fellow faculty members, colleagues, and senior administrators are also critical. These provide informal, but important, milestones for progress toward promotion and allow the candidate to build a coherent collection of documentation of accomplishments, context for these accomplishments, and a clear trajectory forward in their career. The promotion review should not be the first time senior faculty members encounter the candidate's interests and accomplishments.

In closing, the fields of Statistics/Biostatistics and Data Science are both evolving and dynamic. An academic environment should encourage the integrated growth and development of both. Academia often experiences a tension between seeking new knowledge and holding on to established disciplinary distinctions (perhaps a bit longer than necessary). Managing a healthy and beneficial tension between novelty and establishment requires creativity, patience, collaboration, and experimentation, especially in fostering interdisciplinary excellence in both the current and the next generation of scholars in our own field.

REFERENCES

American Statistical Association discussion of the Mathematical Association of America's *Guidelines for Programs and Departments of Mathematical Sciences.* Available at http://www.amstat.org/asa/files/pdfs/POL-ASAEndorsementMAA2000Guidelines.pdf

Boyer, E., Moser, D., Ream, T.C., and Braxton, J.M. (2016), *Scholarship Reconsidered: Priorities for the Professoriate. Expanded Edition.* Jossey-Bass: San Francisco.

Data Citation Synthesis Group (2014), Joint Declaration of Data Citation Principles. Martone M. (ed.) San Diego CA: FORCE11.

Davidian, M. and Louis, T.A. (2012), Why Statistics? *Science* **336**, 12. DOI: 10.1126/science.1218685

Mazumdar, M., Messinger, S., Finkelstein, D.M., Goldberg, J.D., Lindsell, C.J., Morton, S.C., Pollock, B.H., Rahbar, M.H., Welty, L.J., and Parker, R.A. for the Biostatistics, Epidemiology

708 and Research Design (BERD) Key Function Committee of the Clinical and Translational
709 Science (CTSA) Consortium (2015), Evaluating Academic Scientists Collaborating in Team-
710 Based Research: A Proposed Framework. *Academic Medicine* **90** (10): 1302–1308.
711 DOI:10.1097/ACM.0000000000000759
712
713 Mathematical Association of America (2003), *Guidelines for Programs and Departments in*
714 *Undergraduate Mathematical Sciences.* Available at
715 http://www.maa.org/sites/default/files/pdf/guidelines/Dept-Guidelines-Feb2003.pdf
716
717 National Research Council (2013), *Frontiers in Massive Data Analysis.* Washington, D.C.: The
718 National Academies Press.
719
720 National Research Council (2014), *Training Students to Extract Value from Big Data: Summary*
721 *of a Workshop.* Washington, D.C.: The National Academies Press.
722
723 Waller, L.A. and Miller, G. (2016), More than manuscripts: Reproducibility, rigor, and research
724 productivity in the big data area (Editorial). *Toxicological Sciences* **149**, 275-276. doi:
725 10.1093/toxsci/kfv330
726