|  |  |
| --- | --- |
| Title | Developing and Executing an ASA Statistical Collaboration Program with Special Emphasis on Data Scientists |
| Requested Budget | $ 21,630 |
| Year of Implementation | 2016 |
| Contact Name | Erin Tanenbaum |
| Company | NORC at the University of Chicago |
| Address | 8011 Trevor |
| City | Vienna |
| State/Province | VA |
| Zip/Postal Code | 22182 |
| Country | United States |
| Work Phone | (703) 288-0472 |
| Email Address | tanenbaum-erin@norc.org |

## Brief Description

Many statisticians concentrate on their role of doing the numbers. However, collaboration has become increasingly recognized as necessary to successful workplaces and statisticians can greatly contribute. Also with large streams of financial, internet, social media, and other data, the nature of analysis is changing. The attraction of the data science reflects this change. The question becomes “what can we glean from all this information?” rather than “what information do we need to inform our problem?” With collaboration skills, we can proactively seek collaborators, work to develop the research problems, design the analyses, and help write up all, not just the statistical methods and analysis. As with other research areas and scientific cultures, it is important to bridge diverse cultures and terminology to meaningfully collaborate.

However, the best ways to collaborate with others is not something that is taught within Statistics graduate programs and collaborations often do not come easily for a large number of statisticians.

The Committee on Applied Statisticians (CAS) proposes to develop an ASA Collaborations Program. Collaboration initiatives targeted to practicing statisticians have been made in a variety of ASA units, including JSM continuing education courses, as well as the great work being done by the SPAIG Committee. Recent discussions among ASA members have also focused on the topic of collaborations (community.amstat.org/communities/community- home/digestviewer/viewthread?MID=28309 and community.amstat.org/communities/community- home/digestviewer/viewthread?MID=28399).

ASA Executive Director Ron Wasserstein reached out asking if CAS can foster collaborations with others (including data scientists). Thus, we propose to develop a program to improve the collaboration skills of ASA members who self-identify as applied statisticians. If we receive funding, our committee would consider i) developing course materials for use at future JSM programs, ii) drafting a proposal to the National Science Foundation (NSF) and/or other funders to create a collaborations conference iii) creating and deliver webinars on collaboration skills, iv) monitoring the program’s results, and v) making recommendations for continuing the collaboration program (which may include recommendations that could be incorporated into future CPS or JSM conferences).

The scope of our pilot will focus on self-identified, applied statisticians since we believe these individuals frequently collaborate outside of the profession. We will work with them in all areas and no group of applied statisticians will be excluded. To avoid redundancy, we will consult with other units of ASA and assemble a short list of existing materials relating to collaboration efforts within ASA. Because of its value, we will summarize the collaboration efforts we’ve identified to the existing CAS microsite. The hands-on component of this program will focus on training up, reaching across, and reaching out.

The Extraordinary Power of Statistics year in 2016 is a perfect time to empower and enable statistical collaborations for years to come - as a way to promote our profession and expand the boundaries of statistics to further enhance the visibility and impact of our discipline.

Detailed Plan

The Committee on Applied Statisticians (CAS) puts forward this proposal and agrees to be coordinators and mentors for the proposed Applied Statisticians’ Collaboration Clearinghouse Program. With the rise of “team science” in academia and industry and an increasing recognition that data-driven decisions can improve efficiency and policy-making in industry and government, statistical collaboration is perhaps more important now than ever before. Collaboration is clearly very important and has been for a long time. Some statisticians have been collaborating with others for most of their career to their great benefit. Our intention is to aid in the career growth of applied statisticians and enhance the capabilities of industry, academic, government, and business researchers. To that end, the first part of CAS’s overall plan to promote statistical collaboration is to develop a program to support ASA members who self-identify as applied statisticians and seek to improve upon their collaboration skills.

CAS’ track record of partnering within ASA has addressed the career paths of applied statisticians, such as participation in the President’s initiatives and the organizing committee for the Conference on Statistical Practice, regular articles in the AmStat News Masters’ Notebook column; activities at the JSM to help applied statisticians connect with resources, such as the roundtable How Can ASA Support the Applied Statistician?, and the hugely successful Mentoring Pilot and Clearinghouse Program.

In 2013, CAS won an ASA Member Initiative Developing and Piloting a Partnership ASA Mentorship Program worth $11,700. Our direct involvement resulted in:

* CAS developed a mentorship program. CAS developed resources to 1) help individuals start and successfully develop a mentoring relationship, materials 2) organizations start mentoring programs, and 3) find ways to they could sustain those relationships and programs.
* Clearinghouse, researched past mentoring programs within ASA and disseminated materials in an online clearinghouse.
* 2014 and 2015, ran programs at the Conference for Statistical Practice (CSP) and the Joint Statistical Meetings (JSM) matching over 150 mentor-mentee pairs.
* Creation and dissemination of “Mentoring in a Box,” a how-to guide for organizations interested in creating a mentoring program. Of which two sections have implemented our program to-date:
  + Biopharm Mentoring Program
  + WSS Mentoring Program
* Creation and dissemination of “DIY Mentoring,” at CSP and JSM in 2015 (in every attendee bag). One page guide on ways to create a mentor or mentee relationship on your own
* Special Issue on Mentoring in The American Statistician (planned fall 2016)
* Organized and executed the first ever Mentoring Workshop for New and Seasoned Mentors and Mentees, JSM 2015. Included 3 panelists and 14 table leaders with over 55 attendees.
* Organized and executed two roundtable discussions at JSM 2015 for establishing a mentoring program
  + Kick-off a Chapter Mentoring Program, 12 attendees
  + Kick-off a Section Mentoring Program, 18 attendees
* Organized and executed annual JSM Social Mixers to promote the mentoring agenda.
* Amstat News articles and advertisements
  + The American Statistician Announces Special Issue: Mentoring in Support of Statisticians, October 2015
  + Biopharmaceutical Section News, July 2015
  + Washington Statistical Society Starts Mentoring Program, March 2015
  + Eric Vance Talks Mentoring, February 2015
  + ASA Mentorship Program to Continue, May 2014
  + Conference on Statistical Practice Mentoring Program to Help Promote Statistics, November 2013
  + Developing Successful Career Relationships: Leveraging Mentoring, Coaching, and Sponsorship, November 2013
  + Seeing Is Believing, June 2013
* Creation of the Mentoring Committee, a group currently taking over many of our activities.

Keeping in mind the ASA’s Big Tent metaphor, this initiative will extend that track record by developing programs to:

1. Train up — train statisticians (students and non-students) in collaboration skills essential for success.
2. Reach across — collaboration between data science and statistics including training.
3. Reach out/Invite in — promote collaborative statisticians, train non-statisticians in the value and how-to’s for collaborating with statisticians.

To achieve these goals, CAS envisions a multi-year program to research, educate, and institutionalization best practices for collaboration. We believe this program will be successful because of the strong principles of collaboration already held by CAS members. We approach the work of the committee itself as a collaboration, engaging in face-to-face interactions, promoting positive interdependence, using interpersonal skills, monitoring and processing the group’s functioning, and holding each individual accountable to the committee.

This proposal is partitioned into 1) a development program in 2016 (current request); 2) a proposed full-scale program in 2017, (ASA, NSF, or NIH funding possible); and our 3) final recommendations to institutionalize a long-term program in 2018 and thereafter.

The first year we will continue our research and solidify roadmap details to successfully achieve our goals. The Committee will disseminate the program and roadmap details for feedback in multiple ways. First, directly to other strongly collaborative groups within ASA, such as SPAIG, to build on the excellent work they are doing, as well as the JSM BOD, COC, COS, COM, Membership Council, etc. Second, through the ASA Communities, AmStat News, and Friends of CAS to the wider membership. Third, through a CAS microsite that supports the initiative and a collaboration clearinghouse. Finally, through events at JSM including:

* socializing at business meetings and mixers and seeking support and/or feedback
* an interactive JSM topic-contributed panel that will present the roadmap and seek input from attendees
* an invited JSM Roundtable to begin the process of creating a ‘community of practice’ (working group of experts) around statistical collaboration.

Most of these activities are underway and do not require funding. The only one that will require funding is an extra day of hotel pre-conference to prepare for the conference itself. This time will allow the Committee to coordinate activities at the conference and with conference leadership. Where it makes sense, the Committee will be using Collaboration Best Practices to effectively interact with colleagues and leadership. These include the following seven rules:

1. **Look for common ground:** find shared values, personal experiences, receive/give feedback, be willing to accept differences in perception and opinions
2. **Learn about others:** consider their perspectives and needs, appeal to the highest motives, let others express themselves freely
3. **Critique results, not people:** don’t show personal hostility, make others feel good
4. **Give and get respect:** show respect for others' opinions, be responsive to emotions, remain tactful
5. **Proceed slowly:** one idea at a time, seek consensus (understanding and acceptance of decisions) before proceeding. stay organized with agendas.
6. **Be explicit and clear:** share your ideas and feelings, pay attention to nonverbal communication, speak clearly, make eye contact
7. **Remember the five "Cs" of communication:** clarity, completeness, conciseness, concreteness, and correctness

CAS also plans to attend one or more highly collaborative conferences to both learn from them and to reach out to them in preparation for later stages. Our primary suggestion is the Strata+Cultivate conference: http://conferences.oreilly.com/strata/big-data-conference-ny-2015/public/content/cultivate

Strata is the premier conference for Data Scientists. Held around the globe, the conference helps its attendees “put big data, cutting-edge data science, and new business fundamentals to work.” Cultivate is a co-located conference that focuses on the soft skills that are necessary for success.

*Cultivate trains leaders—and those who aspire to lead—in the business practices they need to thrive in the new world. It’s about building a corporate culture that enables design thinking, collaboration, and agility. It’s about understanding people and working with them effectively. It’s about incorporating designers, data scientists, developers, analysts, engineers, and other specialists into multidisciplinary product teams from the beginning. And finally, Cultivate is about building environments where all people, including those from underrepresented groups, can contribute freely and meaningfully.*

Strata is a key conference for data scientists and focuses on corporate participants. From our attendance at the conference, we can bring back ideas to ASA. To increase our participation in the conference, we plan to submit four presentation proposals.

The results of these extensive and interactive activities will be discussed and incorporated into the roadmap at a Face-to-Face meeting of the Committee. This ASA-funded meeting will allow the Committee to engage in a Kaizen event for rapid progress on a large scale project so that we can quickly and efficiently lay out the details to implement its next stages.

One specific future plan is for the Committee to write a proposal to NSF or another funding agency to gather minds at a mini-conference/symposium to determine best practices in statistical collaboration, what are the important unanswered questions, and a strategy for accelerating progress toward answering these questions. This mini-conference/symposium can be accomplished in a variety of ways. One option is to have a mini-conference in conjunction with the Conference on Statistical Practice (CSP), similar to the Strata+Cultivate conference. This would allow CAS to “piggy-back” on the infrastructure already in place for CSP. It could also add an attractive dimension to CSP that would increase attendance.

Another option is to collaborate with an independent conference/symposium such as the ICSA Applied Statistics Symposium and 13th Graybill Conference (see here for the 2015 conference website: http://www.stat.colostate.edu/graybillconference2015/) Again, this would build on the infrastructure already in place in order to minimize cost and maximize the chance for success.

One specific intended outcome of the future mini-conference/symposium would be for the collective community of experts to agree on the types of collaborative, non-technical skills we believe are essential for statistics students (and statisticians and data scientists in general) to become effective statisticians and to develop a plan for creating modules to teach such skills to statistics students (and statisticians and data scientists in general). These modules could be adapted and freely shared by ASA members who educate and train statisticians both in the classroom and on the job. The result would be to train the next generation of statisticians and data scientists who can move from theory to practice to apply statistics and data science to solve problems and make decisions for real-world impact.

As a starting point, we have identified five broad categories of skills that could be further developed into modules and sub-modules to improve the education and training of statisticians worldwide. These categories would be expanded upon and refined at our future mini- conference/symposium. The categories and subcategories are:

* Structuring meetings and projects
* Communication
  + In-meeting communication (e.g. asking good questions; explaining statistics to non-statisticians; listening, paraphrasing, summarizing; non-verbal communication)
  + Written communication
  + Oral presentations
  + Statistical graphics
* Reproducible research practices, statistical workflow, and statistical ethics
* Will collaborate with the Statistical Leadership Initiative to work on communication with management and other leadership skills
* Data science skills (e.g. Big Data computing, database management, web scraping, API, programming)

Barriers to effective collaboration, and ways to overcome those barriers, will also be an important point of discussion. The committee understands the need to explore the specific challenges we have within the analytical community. For example, according to Peter Blau, a primary barrier to collaboration is reconciling diversity in viewpoints. Each member comes with their particular work background, culture, relationship management style, criteria for success and personal goals. These viewpoints must be considered by the collaborating team when working toward agreement, but at the same time they make effective decision-making more difficult. These perspectives mean that even if agreement is reached it may not be truly collaborative or truly an agreement. “This challenge, often called a cultural boundary,” can be found within groups in the analytical community and even within the community at large. The committee understands that change is hard and part of their focus will be how to overcome these barriers in ways that are inclusive, supportive and affirming to those struggling with change.

This potential mini-conference/symposium would accelerate research on the best practices of statistical collaboration and, in addition to its direct outcomes of papers, training modules, and identifying a community of practice on the topic of statistical collaboration, it could evolve in several ways:

1. be repeated once or twice in subsequent years to grow and disseminate research on best practices in statistical collaboration,
2. potentially be incorporated into a new or existing track of the Conference on Statistical Practice, and/or
3. be incorporated into another conference or workshop of ASA's choosing.

The ideas we will propose will be innovative, fun, collaborative, and, if incorporated well, will likely encourage new attendees to our conference and/or increase retention year over year. This increase will not be only because the attendees will be able to improve their collaborative ability as applied statisticians in their personal career, but also because their experience at the conferences will be even more fruitful, engaging, and fulfilling than otherwise.

CAS will also consider creating webinars throughout the year that focus on a particular skill within collaboration. These webinars can be recorded and made available in the clearinghouse for future members.

Finally, CAS will reach out, in particular, to colleges and universities teaching statistics courses (service, undergraduate and postgraduate) providing them with tools to help them teach collaboration within their curriculum.

## Team Members and Roles:

### CAS Members:

* Erin Tanenbaum (Chair), Senior Statistician at NORC at the University of Chicago. Worked as a Lean Six Sigma Black Belt at the Nielsen Company. Her efforts have led to collaborations in establishment surveys across 80 countries and data presentation/visualization improvements within Rural Health, the Census Bureau, and Social Media.
* Mark Otto (Vice-Chair), Biometrician at the U.S. Fish and Wildlife Service. His Division uses Structured Decision Making to solve complex resource issues, a valuable tool to broaden statisticians’ collaboration over all aspect of projects. He also spearheaded the Washington Statistical Society mentoring program.
* Eric Vance, Director of LISA (Virginia Tech’s Laboratory for Interdisciplinary Statistical Analysis). He has trained more than 220 students in statistical collaboration both in the classroom in his course “Communication in Statistical Collaborations” and on the job. The LISA statistical collaborators have helped over 9000 clients answer their research and business questions via statistical collaboration.
* John Lin, Senior Vice President with Epsilon. John leads a global team of data scientists that provides world-class quantitative analytic solutions for Epsilon’s Fortune 500 clients. He has more than 20 years of experience consulting with clients to maximize the efficiency of their marketing efforts through implementation of statistical tools and techniques.
* Chuck Kincaid, Engagement Director for the Business Intelligence and Analytics Practice of Experis. Chuck has learned the necessity of collaboration with colleagues (technical and sales) as well as with clients from almost every industry, including Academe and Government with over 14 years’ experience.
* Robin Mogg, applied statistician in the pharmaceutical industry with over 15 years’ experience. She supports therapeutic areas across all phases of drug development. She's implemented adaptive clinical trials, and has promoted innovative statistical approaches to enable efficient decision making.
* • Nicole Miller, manager in the Quantitative Economics and Statistics practice of Ernst & Young. Through managing a team of statisticians and collaborating with tax professionals to satisfy client demands, Nicole has developed an ability to work and communicate effectively with statisticians and non-statisticians alike.

### Other Members:

* • Holly Shulman (liaison)
* • Jeri Mulrow (liaison)
* • Keith Schleicher (former CAS member and part of the first CSP organization team http://magazine.amstat.org/blog/2011/05/01/presinvitedmay1/)
* • Friends of CAS - we maintain a list of people we call on for help. The group is quite responsive and include past CAS members, as well as those who attended our annual JSM social mixer, responded to requests for Amstat News articles, or were previous participants of our very successful Mentoring Program.

### Roles:

* • Chair (1 person)
* • Logistics and Messaging (1 to 2 people)
* • Conference Proposal detailed planning team (3 to 5 people)
* • Hands-on Collaboration detailed planning team (2 to 4 people)
* • Webmaster (1 person)
* • Treasurer (1 person)
* • Evaluation creation and analysis team (2 people).

## Target Timelines/Milestones:

Timelines are based on the member initiative being accepted by April 30.

| Funding Year | Month/Year | Details |
| --- | --- | --- |
| Pre-funding | March/April 2016 | * Advertise that we are hosting an interactive panel discussion at JSM 2016. * Continue efforts to collect resources and make connections to recent/ongoing collaboration efforts in ASA. * Request time with JSM BOD, COC, COS, COM, Membership Council, etc. to advertise program and seek feedback. |
| 2016 | April 2016 | Kick-off 2016 Funding Project.   * Start to develop consolidated ASA materials framework. * Identify (and assign) conferences for CAS team members. * Start to develop materials for ASA members (panel discussion, continuing education, social mixer materials, evaluation materials). * Start to develop sub-site of CAS microsite to support clearinghouse and overall initiative. |
| 2016 | May - July 2016 | Kick-off Program outside of CAS:   * Microsite goes live to advertise and support our efforts. Update microsite monthly with progress updates. * Communicate the new initiative via AmStat News, ASA Communities, and friends of CAS * Interview/partner with ASA leadership/staff/units regarding collaboration efforts and coordination. * Develop Panel Materials in preparation for JSM 2016. |
| 2016 | Pre-JSM 2016 | CAS members meet for organizing panel discussion, advertising, and planning for remaining 2016 activities. Members also attend business meetings as appropriate to discuss progress and seek support and/or feedback. |
| 2016 | JSM 2016 | * Advertise program at COC, COS, Membership Council, COM, CAS business meetings and mixers. * Inviting more partners and gather information for ongoing efforts. * Host panel discussion on Collaborations |
| 2016 | August-December 2016 | Continually update microsite to meet needs for hands-on mentoring and collaboration |
| 2016 | Fall | Kaizen event (face-to-face meeting) at ASA headquarters for rapid progress. |
| 2017 | January - February 2017 | Update ASA leadership on the efforts. Give update to COS, COC, Leadership Council Executive Committee meetings and solicit ongoing information on collaboration efforts |
| 2017 | February 2017 | * Analyze evaluation results * Make recommendations for a continued program. |
| 2017 | April 2017 | Kick-off 2017 goals |
| Long-term | Annual JSM | Annual update to BOD, COC, COS, COM, and other partners. |
| 2017+ | May 2017+ | ASA Committee on Applied Statisticians to make recommendations to ASA to continue the program. This may include a cost-free option, a low-cost option, and a higher cost option. The pro’s and con’s of each option will be included along with a recommendation for the program. |
| 2017+ | JSM 2017+ | Update stakeholders at in person meetings as well as CAS Social Mixer to receive formal and informal feedback on the program. |

## How will you evaluate the success of the project?

CAS will also measure the success of the initiative through the following avenues:

* Survey attendees of the JSM 2015 panel discussion about our collaboration roadmap and clearinghouse. Each attendee will answer a brief paper survey (no-cost option). The survey will measure satisfaction with 1. The overall plan, 2. Their expected ease of participation in the long term program, 3. Their excitement with the concept (increased collaboration skills), 4. The overall need as perceived by the attendees, and 5. How they heard about the panel discussion.
* Survey attendees of the CAS JSM 2015 social mixer to receive similar feedback as above
* Ask for feedback from ASA leadership to assess the potential impact of the initiative on member (and nonmember) statisticians and data scientists
* Funding of our initiative through grant proposals would underscore the importance of professional collaboration for statisticians and data scientists especially those working across government, academia and industry sectors.

Survey results and feedback from the JSM as well as from ASA leadership will be summarized and analyzed as appropriate. All comments and recommendations will be reviewed and potentially implemented to improve proposed plans and to provide direction for future efforts. The subcommittee will summarize results, including successes, weaknesses, and ways to improve for future plans. The report will be shared with ASA Membership Council Committee and to the ASA Board of Directors.

In the long run, the success of the initiative might be informally reflected in the number of work- products that result from new collaborations. Our initiative not only intends to improve the collaboration skills of statisticians and data scientists but also to remove any pre-existing barriers to collaboration. If our initiative is successful in this endeavor, we would hope to see an increase in professional output such as collaborative publications, new inter-disciplinary research, and other programs and initiatives teaching collaboration skills in student curricula.

## Budget Justification

| Description | Cost |
| --- | --- |
| Face-to-face, Saturday before JSM (just hotel | $250 X 5 = $1,250 costs) to finalize activities (improve JSM = better outcome) |
| Face-to-face Two Day Meeting to formalize plans: Kaizen event for rapid progress on a large scale project. | (6 X $900) + ( 2 local attendees) (1 hotel X $200) = $5,600 |
| Funding to attend Strata or similar conference(s) | (Bronze fees $1,595 X 3 people) + (Diamond fees $3,995) + (travel & hotel $1,500 X 4) = $14,780 |
| TOTAL | $21,630 |

The proposed budget is minimal. The costs outlined will ensure a high quality and successful long-term plan. CAS has a track record of successful leveraging all-day, in-person meetings to allow for rapid progress. A formal agenda will be created to ensure substantial progress on i) coordination with selected stakeholders, ii) disseminating findings, iii) scaling-up, iv) logistics, and v) other game-changing details. It is for this reason that we propose meeting as a team before JSM as well as at ASA headquarters since the JSM attendees would make substantial progress with additional planning.

CAS evaluated various options when requesting funding. We believe we can learn and leverage from other technical conferences. For example, Strata is the premier Data Scientists conference, with the largest attendance, the broadest topics and a substantial focus on collaboration. Attending such a conference will allow us to see how and/or why these conferences have become successful so quickly. The registration price for the Strata ranges from $1,595 (Bronze Pass: all sessions, all keynotes, 3 months of Safari Pro, Expo Hall, all networking events and lunch) to $3,445 (Diamond Pass: Bronze + Cultivate). CAS respectfully requests passes and travel expenses for attendance at Strata or a similar conference with three attendees at the Bronze level and one at the Diamond Level as this combination will provide CAS a broad exposure to conference collaborations at a responsible price.

The costs proposed are small, yet the impact to ASA members and to the profession could be substantial. We believe strongly that increasing ASA member’s collaborations will benefit ASA members. If ASA members are happy with the outcome, they will likely continue their membership and encourage others to join to receive similar benefits. A collaboration program may increase member participation in ASA events, committees, and sections. For example, our recent mentoring program created so much buzz that eleven individuals applied for one spot on the committee in the fall of 2015.

The collaboration guidelines will also encourage applied statisticians to propose involvement in ASA activities such as i) Publication, ii) Joining a Section or Committee, iii) Attending conferences), iv) Attending local ASA events. Thus, those impacted may become more involved in the ASA, offering additional volunteer services and fees for conferences and membership. Thus, the collaborators may offset the costs in the long-run as well with loyalty to the ASA.